Lecture Supplements

Plant Bacteriology

Bacterial Classification

Compiled by N. Hassanzadeh

Version 4.25

January 1, 2025

Website Address:
http://www.phytobacteriology.com

Contact address

- Department of Plant Protection, Faculty of Agricultural Sciences and Food Industries, Science & Research Branch, Islamic Azad University, Tehran-Iran.
- P.O. Box: 14155/775, Postal Code: 1477893855
- Branch website: www.srbiau.ac.ir
- e-mail addresses:
- hasanzadehr@srbiau.ac.ir
- hasanzadehr@yahoo.com

Table of Contents

Table of Contents

Table of Contents

Table of Contents

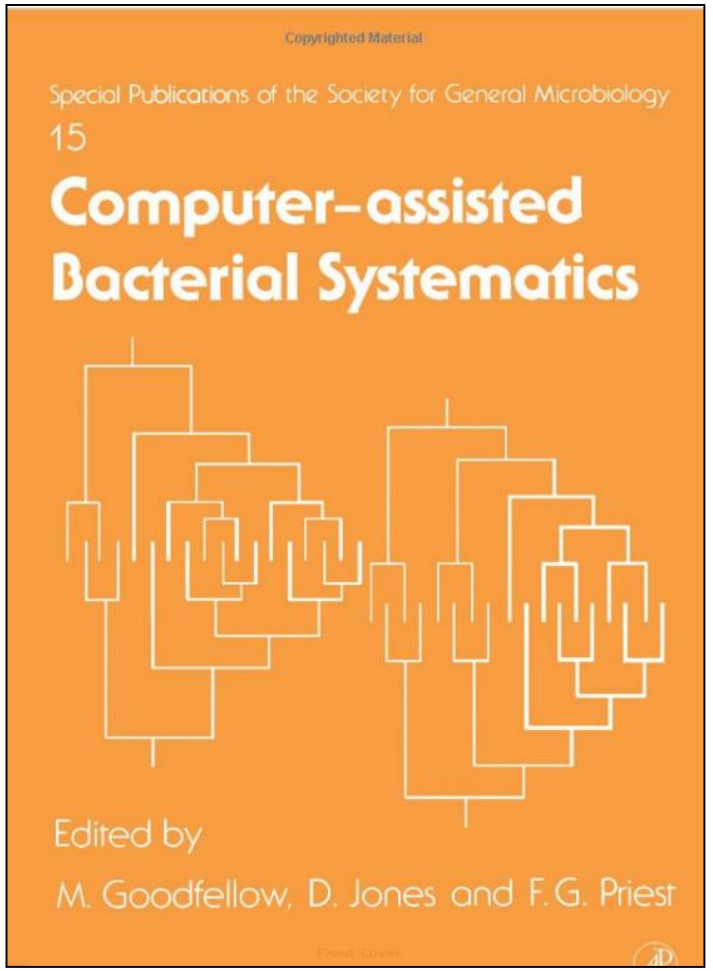
Table of Contents

Books on Bacterial Systematics

Plant Pathogenic Bacteria

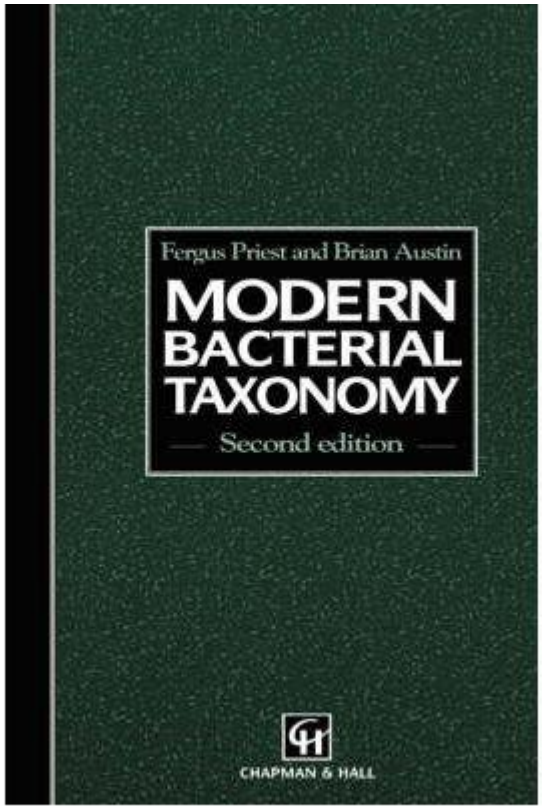
Computer-Assisted Bacterial Systematics

- **Computer-Assisted Bacterial Systematics.**
- Society for General Microbiology
- Edited by: M. Goodfellow, D. Jones and F. G. Priest
- Elsevier Ltd.
- 1985
- 443 pp.



Modern Bacterial Taxonomy

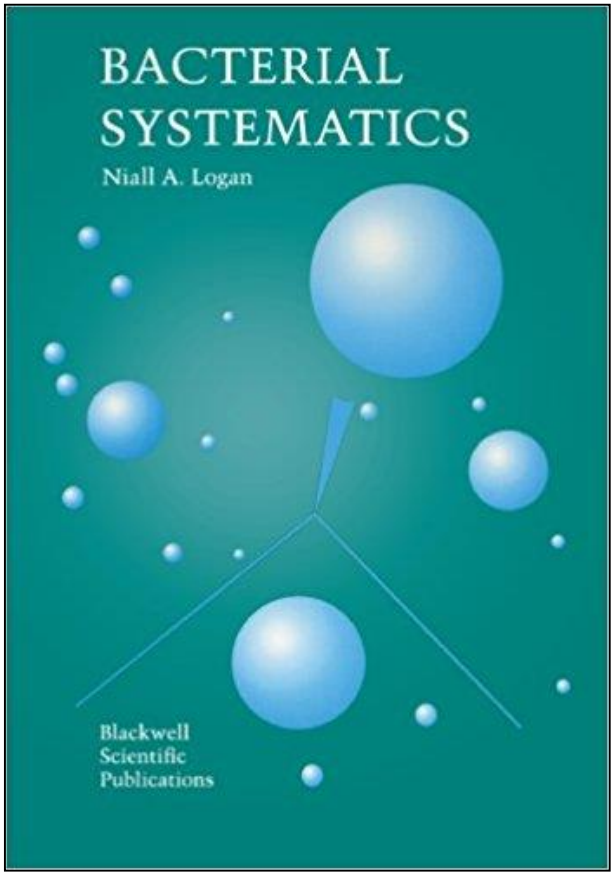
- **Modern Bacterial Taxonomy**
- Author: Kazuo Tsubota
- Chapman and Hall
- 1994
- 228 pp.



Bacterial Systematics

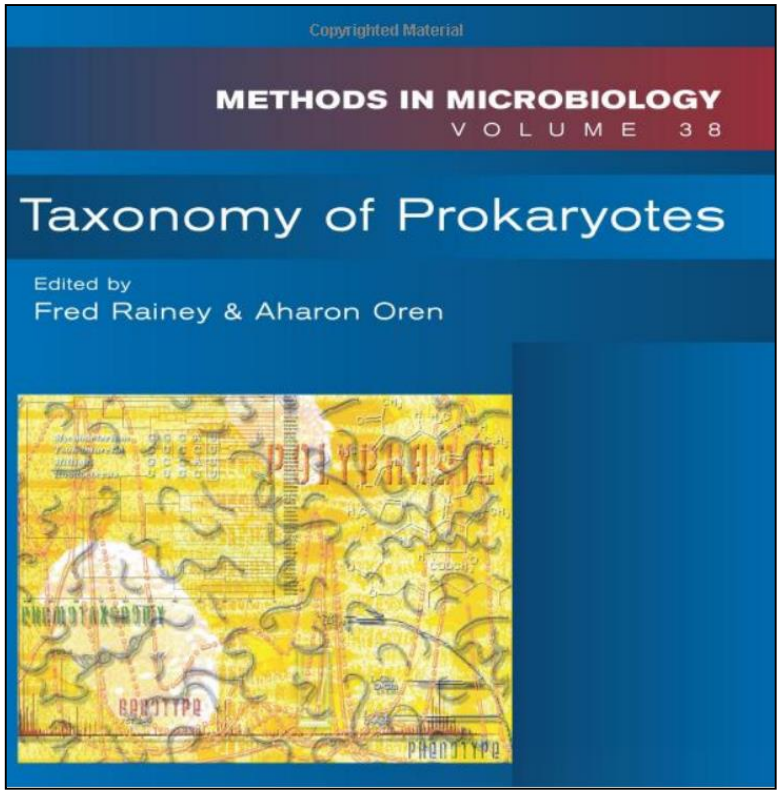
The first book on bacterial systematics at the undergraduate level

- **Bacterial Systematics**
- Author: Niall A. Logan
- Blackwell Scientific Publications
- 1994
- 263 pp.



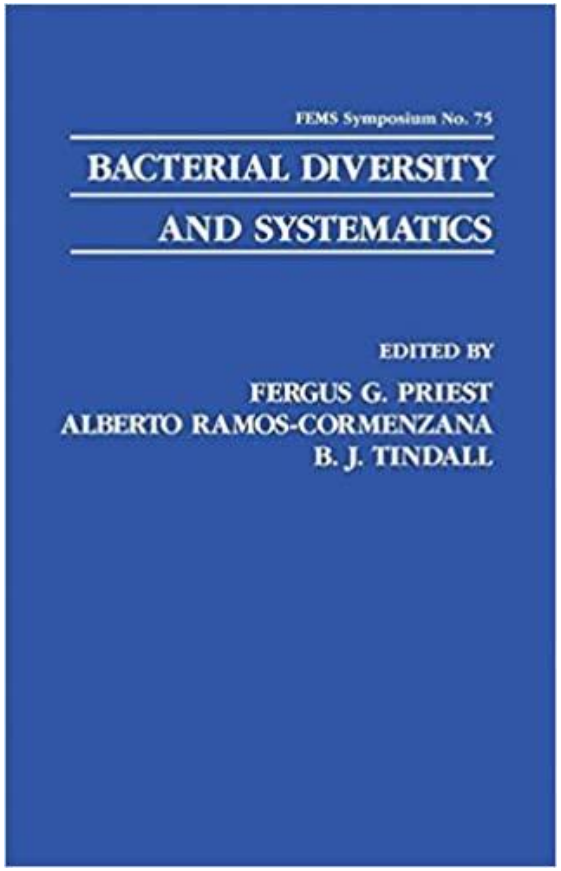
Taxonomy of Prokaryotes

- **Taxonomy of Prokaryotes**
- Volume 38 (Methods in Microbiology)
- by Fred Rainey and Aharon Oren
- Academic Press
- 2011
- 484 pp.



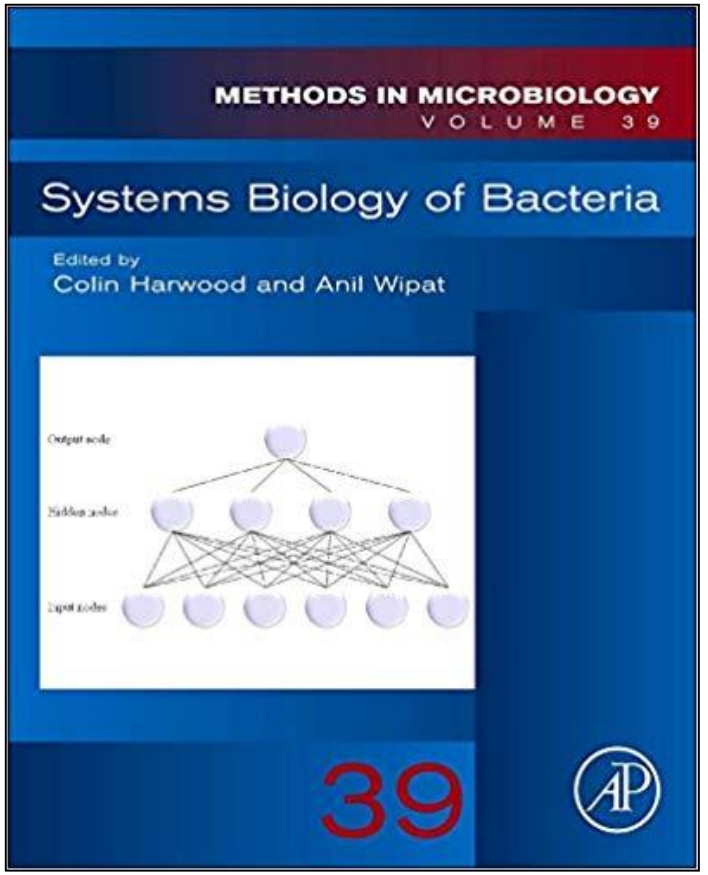
Bacterial Diversity and Systematics (F.E.M.S. Symposium Series Book 75)

- **Bacterial Diversity and Systematics (F.E.M.S. Symposium Series Book 75)**
- by y F.G. Priest, Alberto Ramos-Cormenzana and B.J. Tindall (Editors)
- Springer
- 2012
- 556 pp.



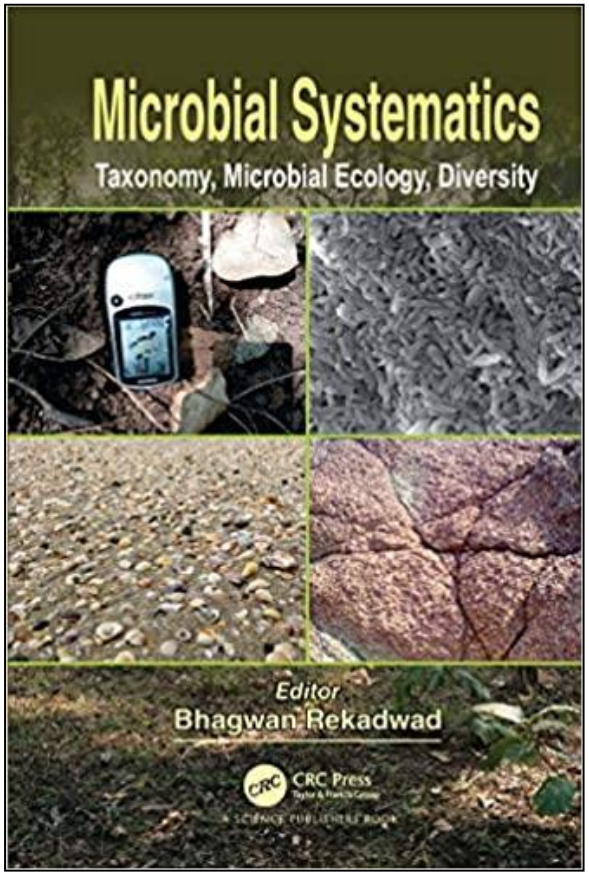
Taxonomy of Prokaryotes

- **Taxonomy of Prokaryotes**
- Volume 39 (Methods in Microbiology)
- by Colin Harwood and Anil Wipat
- Academic Press
- 2013
- 200 pp.



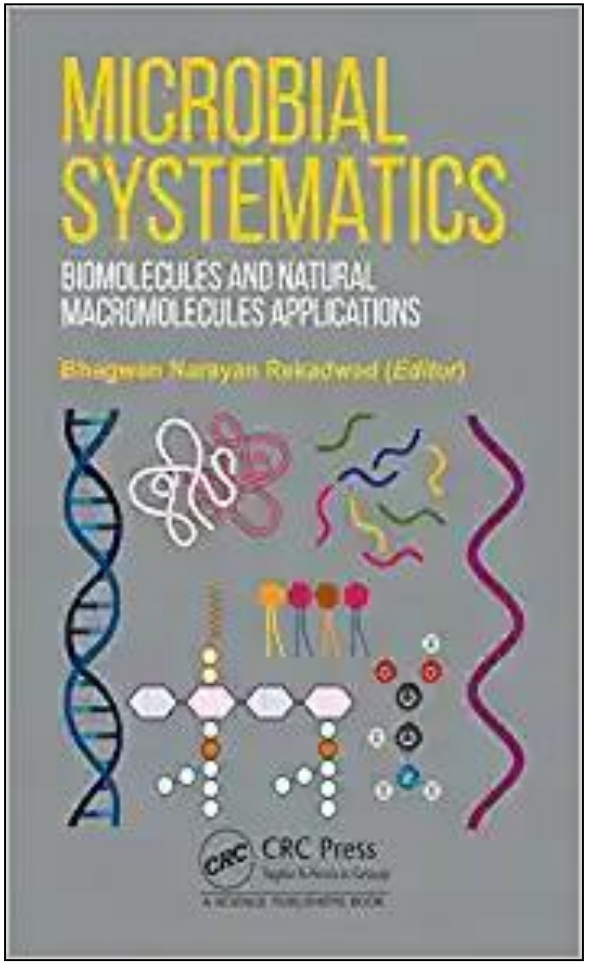
Microbial Systematics: Taxonomy, Microbial Ecology, Diversity

- **Microbial Systematics: Taxonomy, Microbial Ecology, Diversity**
- by Bhagwan Rekadwad
- CRC Press;
- 2020
- 218 pp.



Microbial Systematics:
Biomolecules and Natural Macromolecules Applications

- **Microbial Systematics: Biomolecules and Natural Macromolecules Applications**
- by Bhagwan Narayan Rekadwad
- CRC Press
- 2023
- 230 pp.



Proceedings/Reviews/Monographs/Book chapters/PowerPoints/PDF files

Plant Pathogenic Bacteria

PowerPoints/Monographs/pdf files

- Abedon, S. 2011. Scope and history of microbiology. Chapter10 ID and classification of prokaryotes(Nester10) 2.51 MB. 33 slides.
- Bedone, S.T. Power point presentations and lecture notes. The Ohio State University.
- Britigan, B. E. 2002. Iron: Mechanisms of Pro-oxidant Behavior, Cellular Uptake, and Organism Survival Skills. 979 KB.
- Bull, C.T. 2013. Bacterial taxonomy for the unintentional taxonomist. USDA-ARS, Salinas, CA USA. 27 slides.
- Carro, L. And I. Nouioui.2017. Taxonomy and systematics of plant probiotic bacteria in the genomic era. AIMS Microbiology, 3 (3): 383-412.
- Chester R. Cooper, Jr. 2004. BIOL 3702 Lecture Outline Chapter 19 - Microbial Taxonomy and Phylogeny. 7pp.
- Coenye, T. 2003. Modern bacterial systematics in practice : Polyphasic taxonomy of the *Burkholderia cepacia* complex. 818 kB.
- Coenye, T. 2003. Microbial genomics & the study of prokaryotic diversity. Ghent University, Belgium Laboratory for Pharmaceutical Microbiology. 113 slides.
- De La Fuente, Leonardo.2009. Introducción. Auburn University. 82 slides.
- DuBois,J.2011. Introduction to bioinorganic chemistry of the d-block. Bio Iron. d-block chemistry in action Lecture 11. 1.99 MB.
- El-Safey Mohmad El-Safey,2011. Introduction of bacterial taxonomy. 73 pp.
- Ghenghesh ,K.S. 2015. bacterial taxonomy. 26 slides.

PowerPoints/Monographs/pdf files

- Hawksworth, D.L. 2010. Terms used in Bionomenclature: The naming of organisms (and plant communities). Web application at http://www.gbif.org/communications/resources/print-and-onlineresources/bionomenclature/218 pp.
- Johnson, N. C. *Agrobacterium*, Plant-Microbe Interactions, Lecture 22. 0.98 MB.
- Karki, G. 2017. Classification of bacteria. www.onlinebiologynotes.com.
- Parkinson, N. Identifying Relatedness Between Bacterial Plant Pathogens. Parkinson_MolID_CSL_1.pdf.
- Portier, P. 2009. French Collection for Plant-pathogenic Bacteria. COST 873 practical workshop Sept 24th 2009. 1.90 MB.
- Priest, F.G., Alberto Ramos-Cormenzana and B. J.Tendall. 2013. Bacterial Diversity and Systematics (F.E.M.S. Symposium Series). 331pp.
- Rescott *et al.*, 2005. Microbiology- Unit Four - Phylogeny of Bacteria, Prokaryotes: Bacterial Genetic Systems.
- Vandamme, P. ?. Molecular taxonomy: the bacterial species concept and polyphasic taxonomy revisited. September 12, Lyngby , Denmark. 48 slides.
- Vandamme, P. 2013. Introduction to polyphasic taxonomy. Third European Congress on Microbial Biofilms. 64 slides.

PowerPoints/Monographs/pdf files

- Hawksworth, D.L. 2010. Terms used in Bionomenclature: The naming of organisms (and plant communities). Web application at http://www.gbif.org/communications/resources/print-and-onlineresources/bionomenclature/218 pp.
- Johnson, N. C. *Agrobacterium*, Plant-Microbe Interactions, Lecture 22. 0.98 MB.
- Karki, G. 2017. Classification of bacteria. www.onlinebiologynotes.com.
- Parkinson, N. Identifying Relatedness Between Bacterial Plant Pathogens. Parkinson_MolID_CSL_1.pdf.
- Portier, P. 2009. French Collection for Plant-pathogenic Bacteria. COST 873 practical workshop Sept 24th 2009. 1.90 MB.
- Priest, F.G., Alberto Ramos-Cormenzana and B. J.Tendall. 2013. Bacterial Diversity and Systematics (F.E.M.S. Symposium Series). 331pp.
- Rescott *et al.*, 2005. Microbiology- Unit Four - Phylogeny of Bacteria, Prokaryotes: Bacterial Genetic Systems.
- Vandamme, P. ?. Molecular taxonomy: the bacterial species concept and polyphasic taxonomy revisited. September 12, Lyngby , Denmark. 48 slides.
- Vandamme, P. 2013. Introduction to polyphasic taxonomy. Third European Congress on Microbial Biofilms. 64 slides.

PowerPoints/Monographs/pdf files

- National Institutes of Health. 2020. Classification of 16S rRNA reads is improved using a niche-specific database constructed by near-full length sequencing. 28 pp.
- Berman, H. 2021. Bacterial species in the age of next-generation sequencing. American Society for Microbiology.
- Janda, J. M.2021. Taxonomic Classification of Bacteria. In: Lorrence H Green, Emanuel Goldman, eds. Practical Handbook of Microbiology, CRC Press. 975 pp.
- Jian-Yu Jiao, Rashidin Abdugheni et al., 2024. Advancements in prokaryotic systematics and the role of Bergey's International Society for Microbial Systematicsin addressing challenges in the meta-data era. National Science Review, Volume 11, Issue 7, nwae168.

Bacterial Taxonomy Websites

Plant Pathogenic Bacteria

The old front page of the ISPP website
International Society for Plant Pathology

- The old front page of the ISPP website that listed the names of the all plant pathogenic bacteria.

International Society for Plant Pathology

Promoting World-Wide Plant Health and Food Security

ISPP INTERNATIONAL SOCIETY FOR PLANT PATHOLOGY

Home Congresses Commission on Global Food Security Newsletter Archive Subject Matter Committees Resources Administration

- TPPB Home
- Acetobacter - Curtobacterium
- Enterobacter - Nocardia
- Pantoea - Pseudomonas
- Ralstonia - Xanthomonas bromi
- Xanthomonas campestris - Xylophilus
- 'Candidatus'
- References

Names of Plant Pathogenic Bacteria, 1864-2004

Acetobacter - Curtobacterium

Acetobacter Beijerinck 1898

Acetobacter aceti (Pasteur 1864) Beijerinck 1898

ATCC 15973; ICMP 8807; LMG 1261

Acetobacter pasteurianus (Hansen 1879) Beijerinck & Folpmers 1916

= [*Pseudomonas pomi* Cole 1959] not valid – Dhanvantari *et al.* (1978) refers

Pathogenic reference strain: ATCC 19877; ICMP 3878; LMG 1805; NCPPB 463

Acidovorax Willems, Falsen, Pot, Jantzen, Hoste, Vandamme, Gillis, Kersters & De Ley 1990

Acidovorax anthurii Gardan, Dauga, Prior, Gillis & Saddler 2000

CFBP 3232, ICMP 13404

Acidovorax avenae (Manns 1909) Willems, Goor, Thielemans, Gillis, Kersters & De Ley 1992

= *Pseudomonas avenae* Manns 1909

= *Pseudomonas avenae* subsp. *avenae* Manns 1909

= *Pseudomonas rubrilineans* (Lee, Purdy, Barnum & Martin 1925) Stapp 1928

ATCC 19860; CFBP 2425; ICMP 3183; LMG 2117; NCPPB 1011

Acidovorax avenae subsp.***avenae*** (Manns 1909) Willems, Goor, Thielemans, Gillis, Kersters & De Ley 1992

= *Pseudomonas avenae* Manns 1909

= *Pseudomonas avenae* subsp. *avenae* Manns 1909

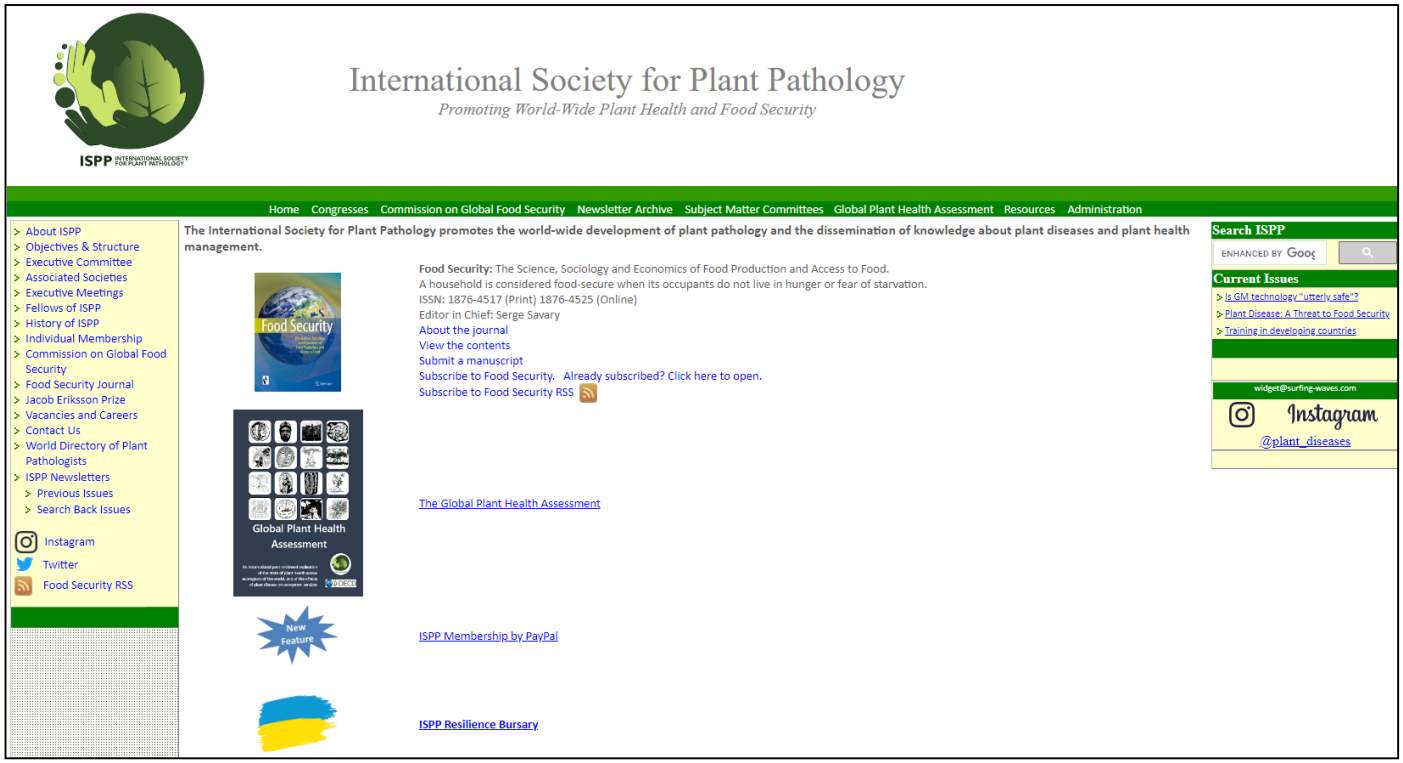
= *Pseudomonas rubrilineans* (Lee, Purdy, Barnum & Martin 1925) Stapp 1928

Show de

Old version of the ISPP website

The new first page of the ISPP website
International Society for Plant Pathology

- Currently, the first page of the ISPP website that has removed the names of the bacteria from its site.



New homepage of ISPP website,2023

LPSN is the most complete and authoritative source for prokaryotic nomenclature
History of LPSN

1. Jean P. Euzéby era (1997-2013). Information on the original author of LPSN, Jean P. Euzéby, is found in his CV in the archive, which also contains a copy of the first version of this website. LPSN as well-known internet site was launched by Jean in 1997.
2. Aidan C. Parte era (2013-2020). Information on the history of LPSN during this period is found on the Changelog page. The entire website has been archived as of November 2019, including an overview.
3. LPSN-DSMZ era (since 2020). In autumn 2019, Aidan C. Parte joined forces with staff members of the Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH to create a website based on a completely new database implementation, which was originally developed for the Type-Strain Genome Server or TYGS.

- Data from LPSN were imported into that database together with data from Prokaryotic Nomenclature Up-to-date or PNU. PNU served essentially the same purpose as LPSN but covered fewer taxonomic categories. Please see the Changelog page for ongoing modifications of LPSN.

LPSN is the most complete and authoritative source for prokaryotic nomenclature

History of LPSN

- With the rise in actual and anticipated new prokaryotic names, the LPSN had become increasingly difficult to maintain, **so a better technical and funding basis was sought**. To that end, in November 2019, the LPSN was acquired by Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures GmbH, Germany.
- The DSMZ already had its Prokaryotic Nomenclature Up-to-date (PNU) service since 1993, and because there is significant overlap in the content of both sites, it was decided to merge the two services into a completely new one, still under the LPSN name, which has a higher profile in the community.
- On 17 February 2020, the new LPSN was launched at https://lpsn.dsmz.de/under a modified Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license (Parte *et al.*,2020).

List of Prokaryotic names with Standing in Nomenclature (LPSN)
Creator by J.P. Euzéby

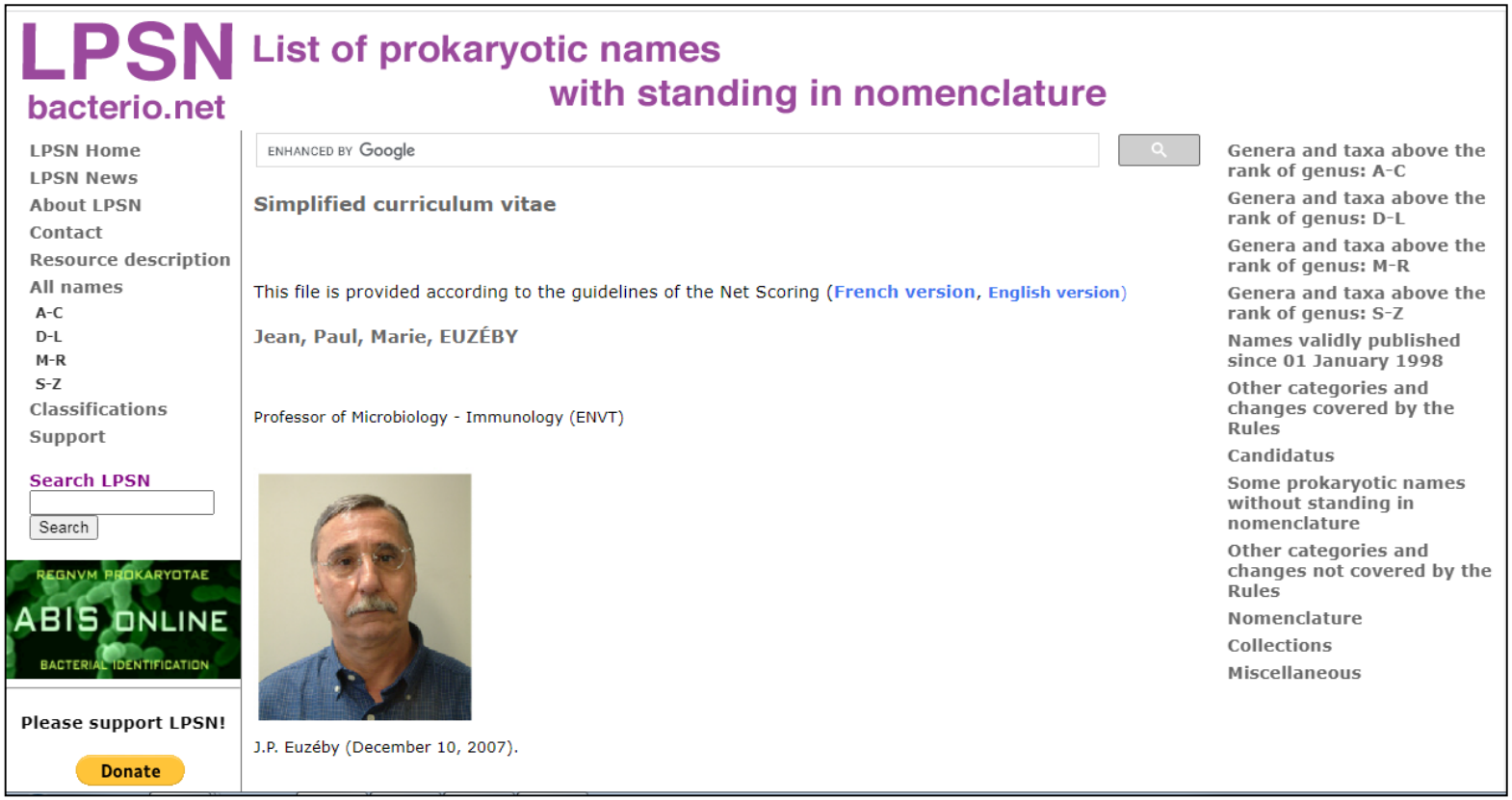
- The great gratitude and honor to Prof. Dr Jean Euzéby for his tireless work in providing the 'List of Prokaryotic Names with Standing in Nomenclature (LPSN)'.
- Since 1997, he has manually checked all issues of International Journal of Systematic and Evolutionary Microbiology (IJSEM) to extract the taxonomic information (such as new species, new combinations and emendations), classify it in an orderly manner and make it electronically available in LPSN.
- Furthermore, LPSN compiles information provided by:

1. the Taxonomic Outline of Bacteria and Archaea (TOBA),
2. the NCBI taxonomy,
3. the Taxonomic Outlines of the Bergey's Manual of Systematic Bacteriology, and
4. Silva and 'The All-Species Living Tree Project (LTP)'.

Yilmaz *et al.*,2014

List of Prokaryotic names with Standing in Nomenclature (LPSN)

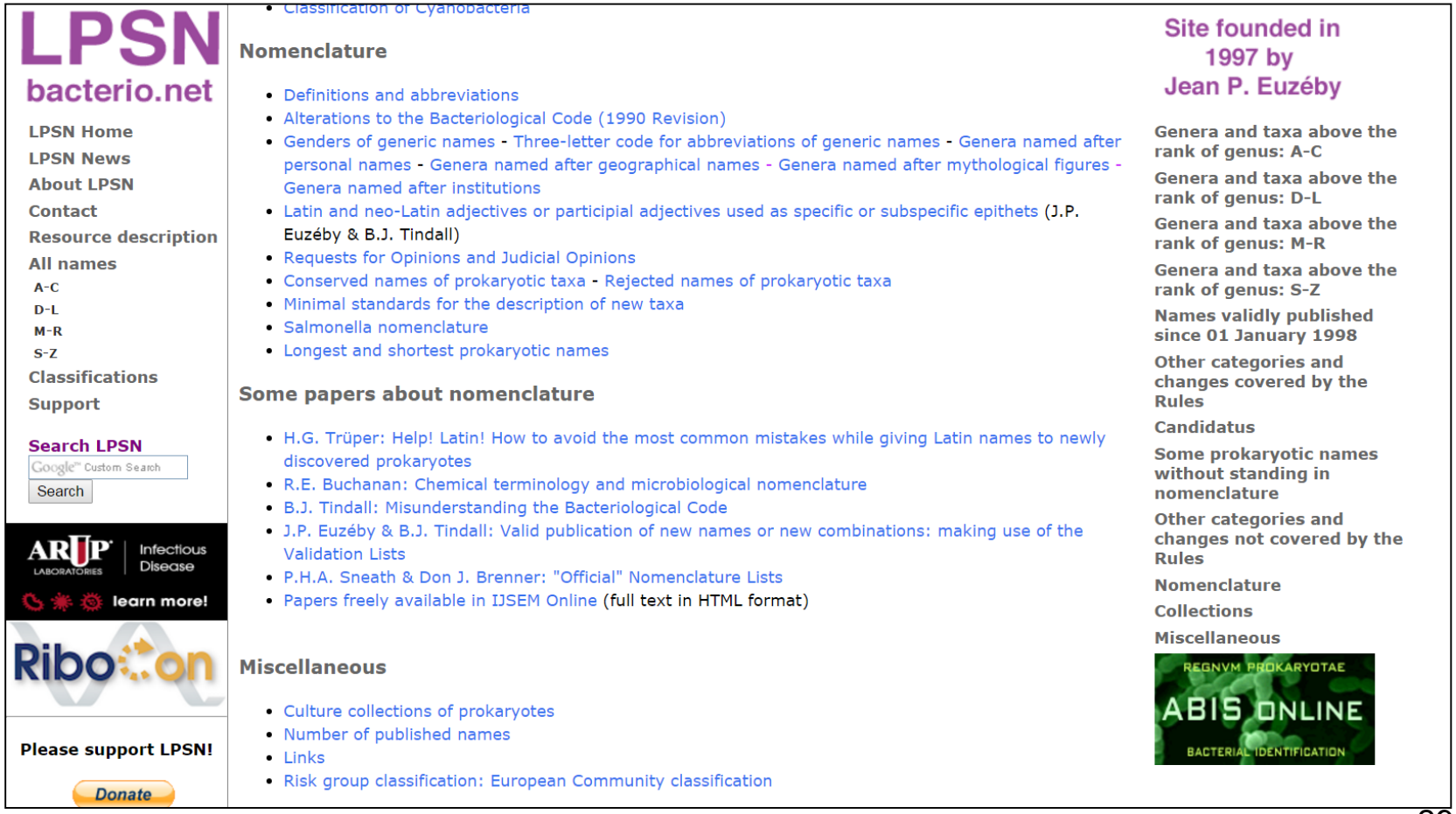
The old front page of the Euzéby website



Jean Paul Euzéby, founder of LPSN, is retired at the end of June 2013

List of Prokaryotic names with Standing in Nomenclature (LPSN)

The old front page of the Euzéby website



LPSN
bacterio.net

LPSN Home
LPSN News
About LPSN
Contact
Resource description
All names
A-C
D-L
M-R
S-Z
Classifications
Support

Search LPSN
Google™ Custom Search
Search

ARUP LABORATORIES | Infectious Disease
learn more!

Ribocon

Please support LPSN!
Donate

- Classification of Cyanobacteria

Nomenclature

- Definitions and abbreviations
- Alterations to the Bacteriological Code (1990 Revision)
- Genders of generic names - Three-letter code for abbreviations of generic names - Genera named after personal names - Genera named after geographical names - Genera named after mythological figures - Genera named after institutions
- Latin and neo-Latin adjectives or participial adjectives used as specific or subspecific epithets (J.P. Euzéby & B.J. Tindall)
- Requests for Opinions and Judicial Opinions
- Conserved names of prokaryotic taxa - Rejected names of prokaryotic taxa
- Minimal standards for the description of new taxa
- Salmonella nomenclature
- Longest and shortest prokaryotic names

Some papers about nomenclature

- H.G. Trüper: Help! Latin! How to avoid the most common mistakes while giving Latin names to newly discovered prokaryotes
- R.E. Buchanan: Chemical terminology and microbiological nomenclature
- B.J. Tindall: Misunderstanding the Bacteriological Code
- J.P. Euzéby & B.J. Tindall: Valid publication of new names or new combinations: making use of the Validation Lists
- P.H.A. Sneath & Don J. Brenner: "Official" Nomenclature Lists
- Papers freely available in IJSEM Online (full text in HTML format)

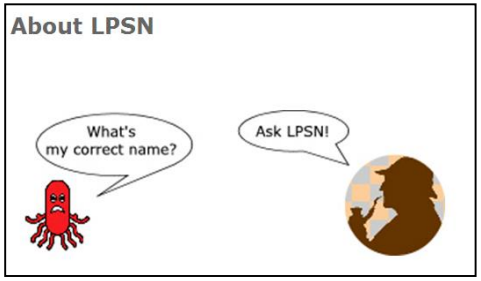
Miscellaneous

- Culture collections of prokaryotes
- Number of published names
- Links
- Risk group classification: European Community classification

Site founded in 1997 by Jean P. Euzéby

Genera and taxa above the rank of genus: A-C
Genera and taxa above the rank of genus: D-L
Genera and taxa above the rank of genus: M-R
Genera and taxa above the rank of genus: S-Z
Names validly published since 01 January 1998
Other categories and changes covered by the Rules
Candidatus
Some prokaryotic names without standing in nomenclature
Other categories and changes not covered by the Rules
Nomenclature
Collections
Miscellaneous

REGNVM PROKARYOTAE
ABIS ONLINE
BACTERIAL IDENTIFICATION



LPSN: Creator by J.P. Euzéby

The old front page of the Euzéby website

LPSN
bacterio.net

LPSN Home
LPSN News
About LPSN
Contact
Resource description
All names
A-C
D-L
M-R
S-Z
Classifications
Support

Search LPSN
Google™ Custom Search
Search

ARUP LABORATORIES | Infectious Disease
learn more!

Ribo On

Please support LPSN!
Donate

- Classification of Cyanobacteria

Nomenclature

- Definitions and abbreviations
- Alterations to the Bacteriological Code (1990 Revision)
- Genders of generic names - Three-letter code for abbreviations of generic names - Genera named after personal names - Genera named after geographical names - Genera named after mythological figures - Genera named after institutions
- Latin and neo-Latin adjectives or participial adjectives used as specific or subspecific epithets (J.P. Euzéby & B.J. Tindall)
- Requests for Opinions and Judicial Opinions
- Conserved names of prokaryotic taxa - Rejected names of prokaryotic taxa
- Minimal standards for the description of new taxa
- Salmonella nomenclature
- Longest and shortest prokaryotic names

Some papers about nomenclature

- H.G. Trüper: Help! Latin! How to avoid the most common mistakes while giving Latin names to newly discovered prokaryotes
- R.E. Buchanan: Chemical terminology and microbiological nomenclature
- B.J. Tindall: Misunderstanding the Bacteriological Code
- J.P. Euzéby & B.J. Tindall: Valid publication of new names or new combinations: making use of the Validation Lists
- P.H.A. Sneath & Don J. Brenner: "Official" Nomenclature Lists
- Papers freely available in IJSEM Online (full text in HTML format)

Miscellaneous

- Culture collections of prokaryotes
- Number of published names
- Links
- Risk group classification: European Community classification

Site founded in 1997 by Jean P. Euzéby

Genera and taxa above the rank of genus: A-C
Genera and taxa above the rank of genus: D-L
Genera and taxa above the rank of genus: M-R
Genera and taxa above the rank of genus: S-Z
Names validly published since 01 January 1998
Other categories and changes covered by the Rules
Candidatus
Some prokaryotic names without standing in nomenclature
Other categories and changes not covered by the Rules
Nomenclature
Collections
Miscellaneous

REGNVM PROKARYOTAE
ABIS ONLINE
BACTERIAL IDENTIFICATION

Dr. Aidan C Parte, Curator

List of Prokaryotic names with Standing in Nomenclature (LPSN) is a free to use service founded by Jean P. Euzéby in 1997 and later on maintained by Aidan C. Parte. Thank Dr. Aidan C. Parte for taking over Prof. Dr. Euzeby's tasks, and continuing the LPSN Resource (Yilmaz *et al.*,2013).

The new front page of the Euzéby website

LPSN was integrated into the DSMZ's Prokaryotic Nomenclature Up-to-date (PNU) database as an all-new service

LPSN-DSMZ's website (https://lpsn.dsmz.de)

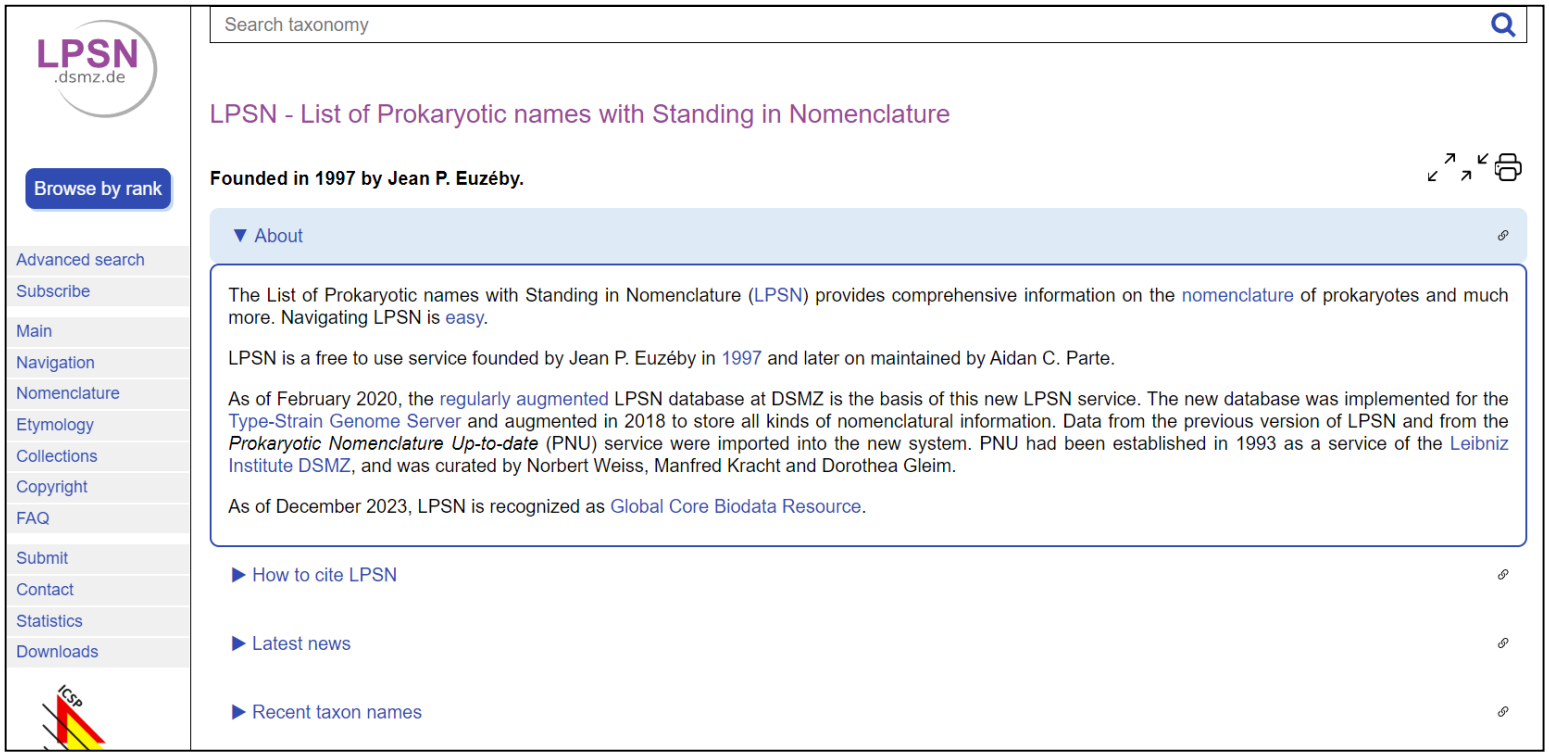
- **Prokaryotic names**
- Since 2020, LPSN is maintained in its own platform.
- In February 2020, LPSN move from the previous platform to the DSMZ and now is located at https://lpsn.dsmz.de at the Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures.
- Since decades the Leibniz Institute DSMZ offers the service Prokaryotic Nomenclature up-to-date (PNU), which provides a compilation of all names of Bacteria and Archaea that have been validly published according to the International Code of Nomenclature of Prokaryotes.

DSMZ website,2022;..

The new front page of the Euzéby website

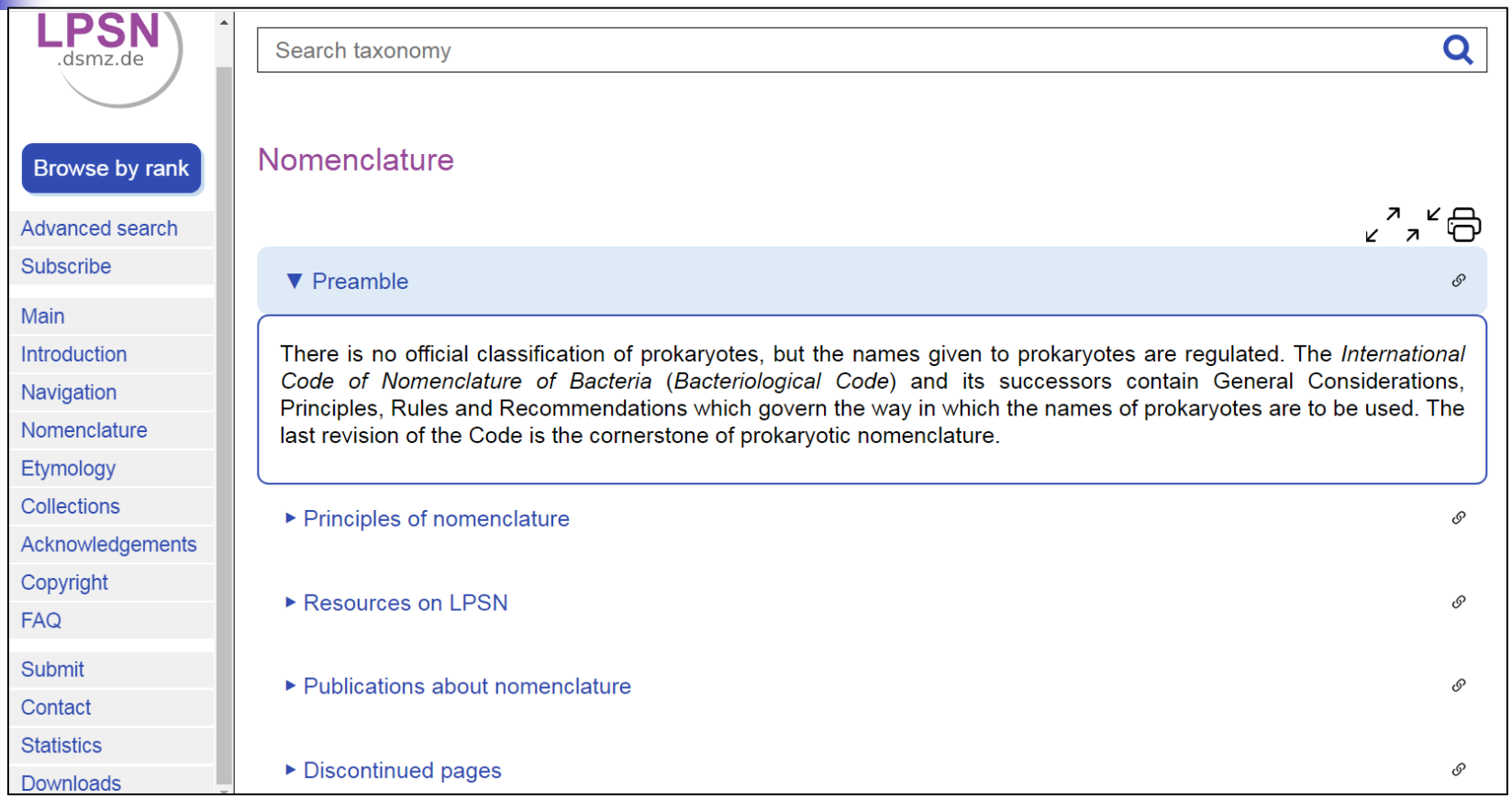
LPSN was integrated into the DSMZ's Prokaryotic Nomenclature Up-to-date (PNU) database as an all-new service

LPSN-DSMZ's website (https://lpsn.dsmz.de)



In February 2020, a new version of LPSN was published as a service of the Leibniz Institute DSMZ, thereby also integrating the Prokaryotic Nomenclature Up-to-date service. New LPSN-DSMZ's website is: https://lpsn.dsmz.de.

The new front page of the Euzéby website
LPSN-DSMZ's website (https://lpsn.dsmz.de)
Nomenclature

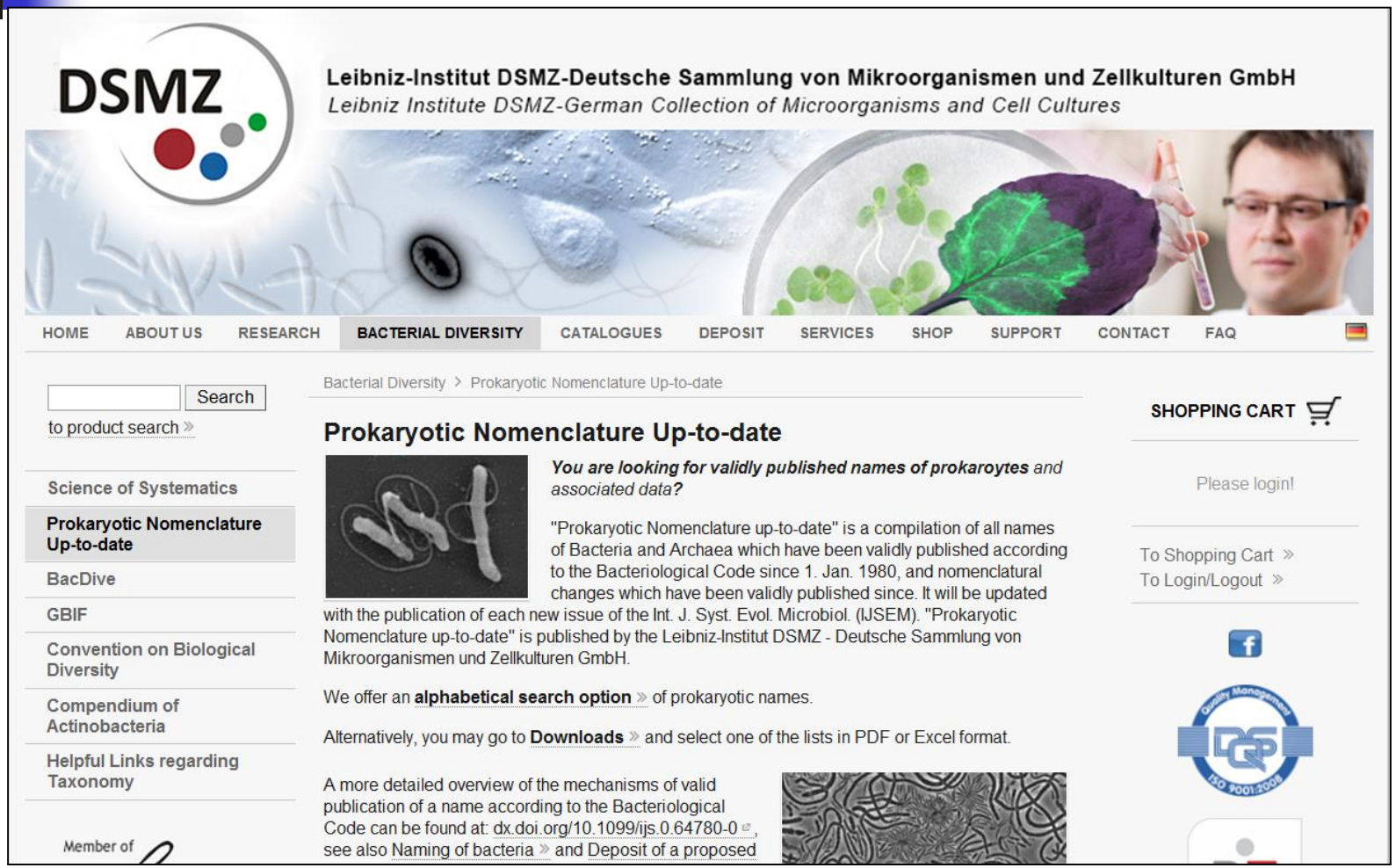


LPSN includes information on names, synonyms, publications, nomenclatural types and sequences.

DSMZ

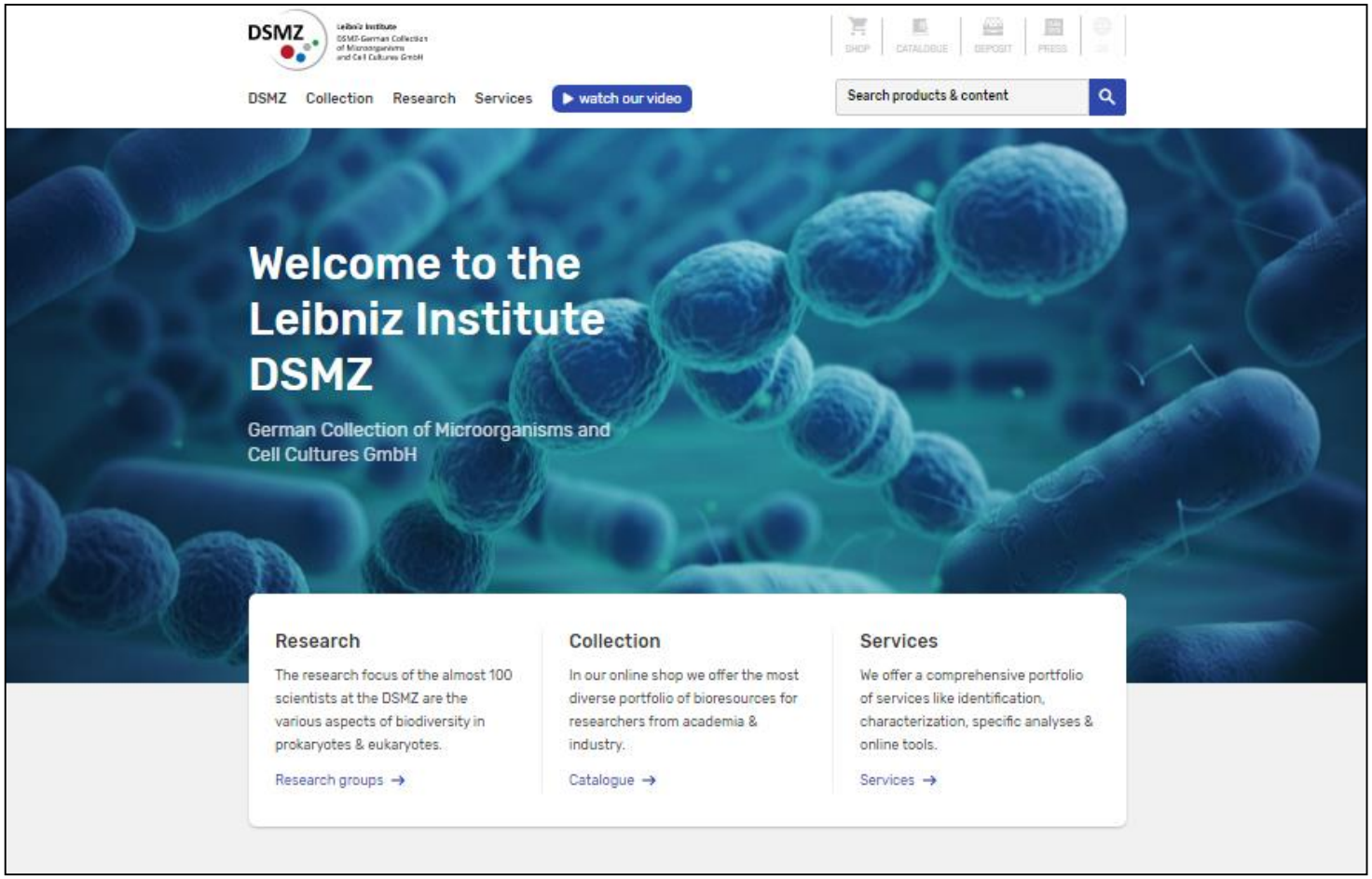
Prokaryotic Nomenclature Up-to-Date

The old front page of the DSMZ website



The new front page of the DSMZ website

German Collection of Microorganisms and Cell Cultures GmbH



DSMZ

German Collection of Microorganisms and Cell Cultures GmbH

The three main activities of DSMZ

1. Research

- The research focus of the almost 100 scientists at the DSMZ are the various aspects of biodiversity in prokaryotes & eukaryotes.

2. Collection

- In our online shop we offer the most diverse portfolio of bioresources for researchers from academia & industry.

3. Services:

- We offer a comprehensive portfolio of services like identification, characterization, specific analyses & online tools.

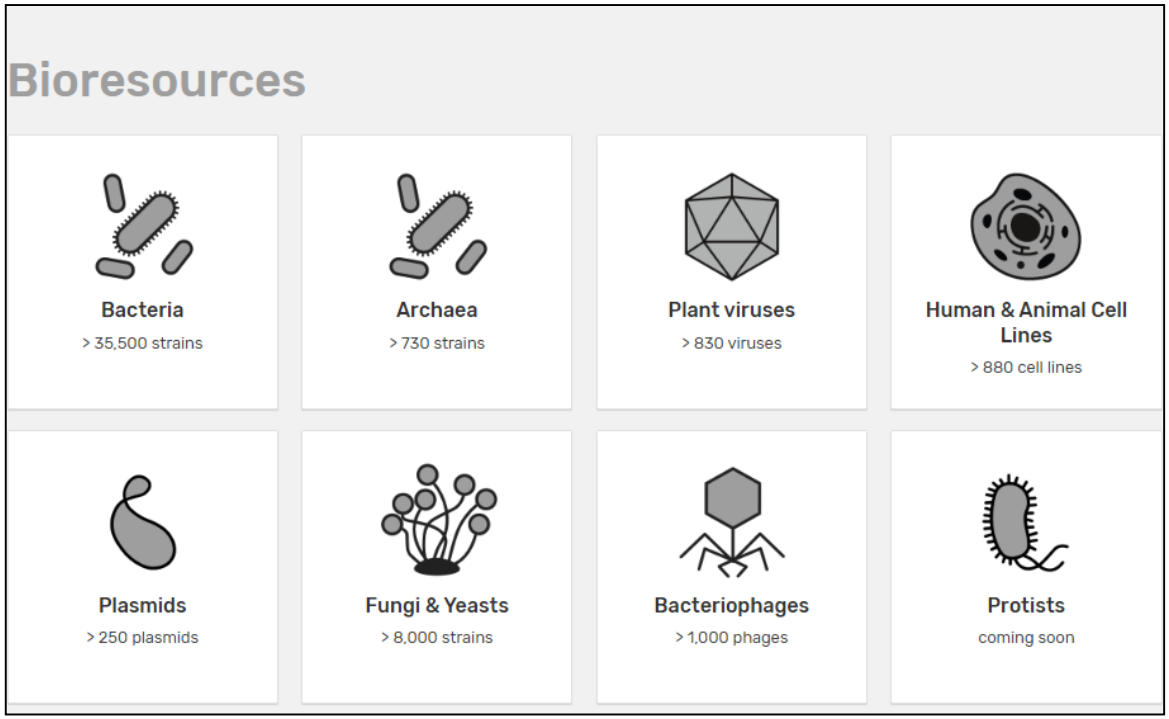
GmbH is an abbreviation of the German phrase "Gesellschaft mit beschränkter Haftung," which means "company with limited liability". portfolio: a range of investments held by a person or organization.

DSMZ

German Collection of Microorganisms and Cell Cultures GmbH

Bioresources

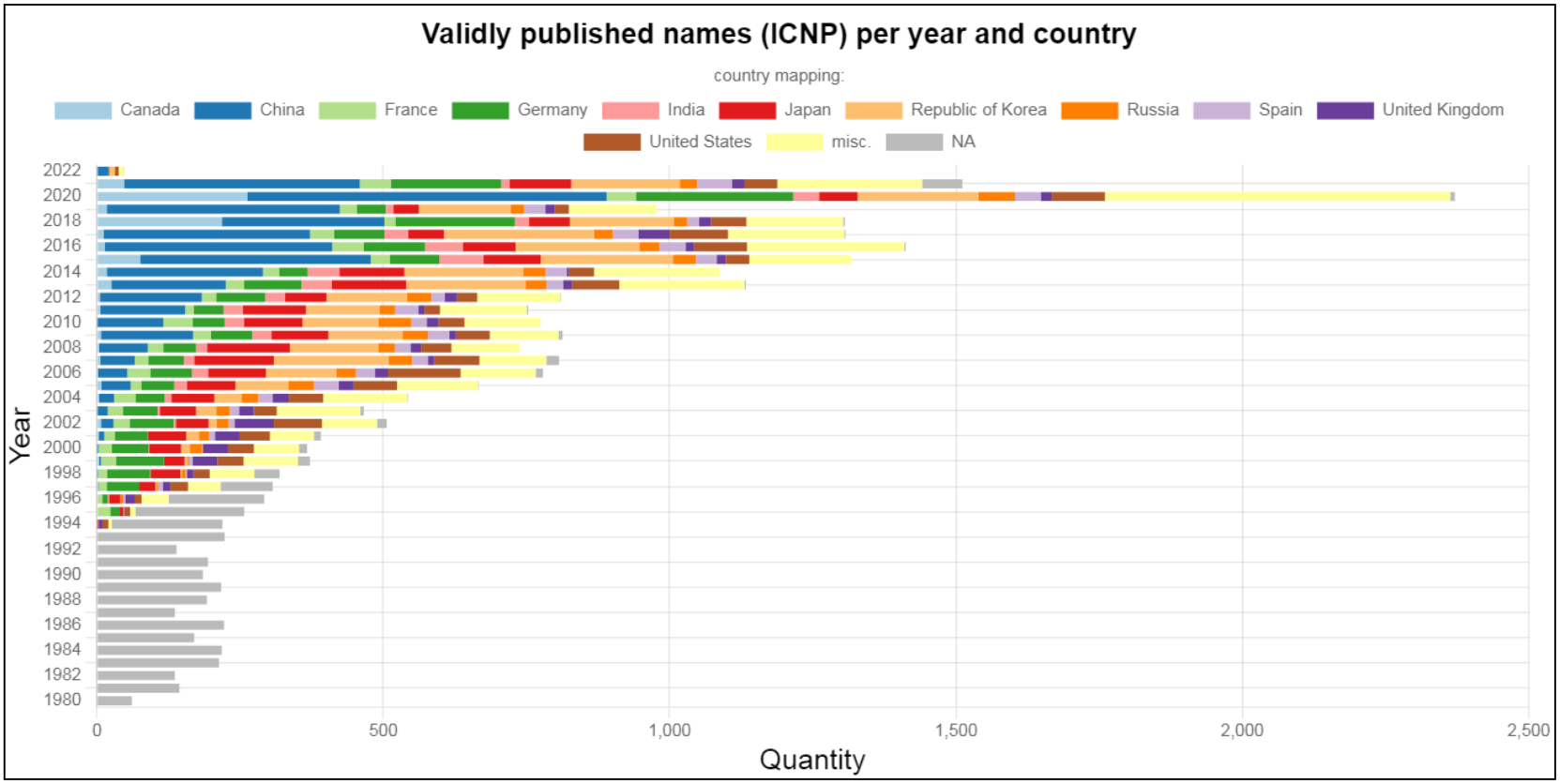
- The Leibniz Institute DSMZ is the world's most diverse collection of biological resources (bacteria, archaea, protists, yeasts, fungi, bacteriophages, plant viruses, genomic bacterial DNA as well as human and animal cell lines).



LPSN: Creator by J.P. Euzéby

Validly published names (ICNP) per year and country

The International Code of Nomenclature of Prokaryotes (ICNP)

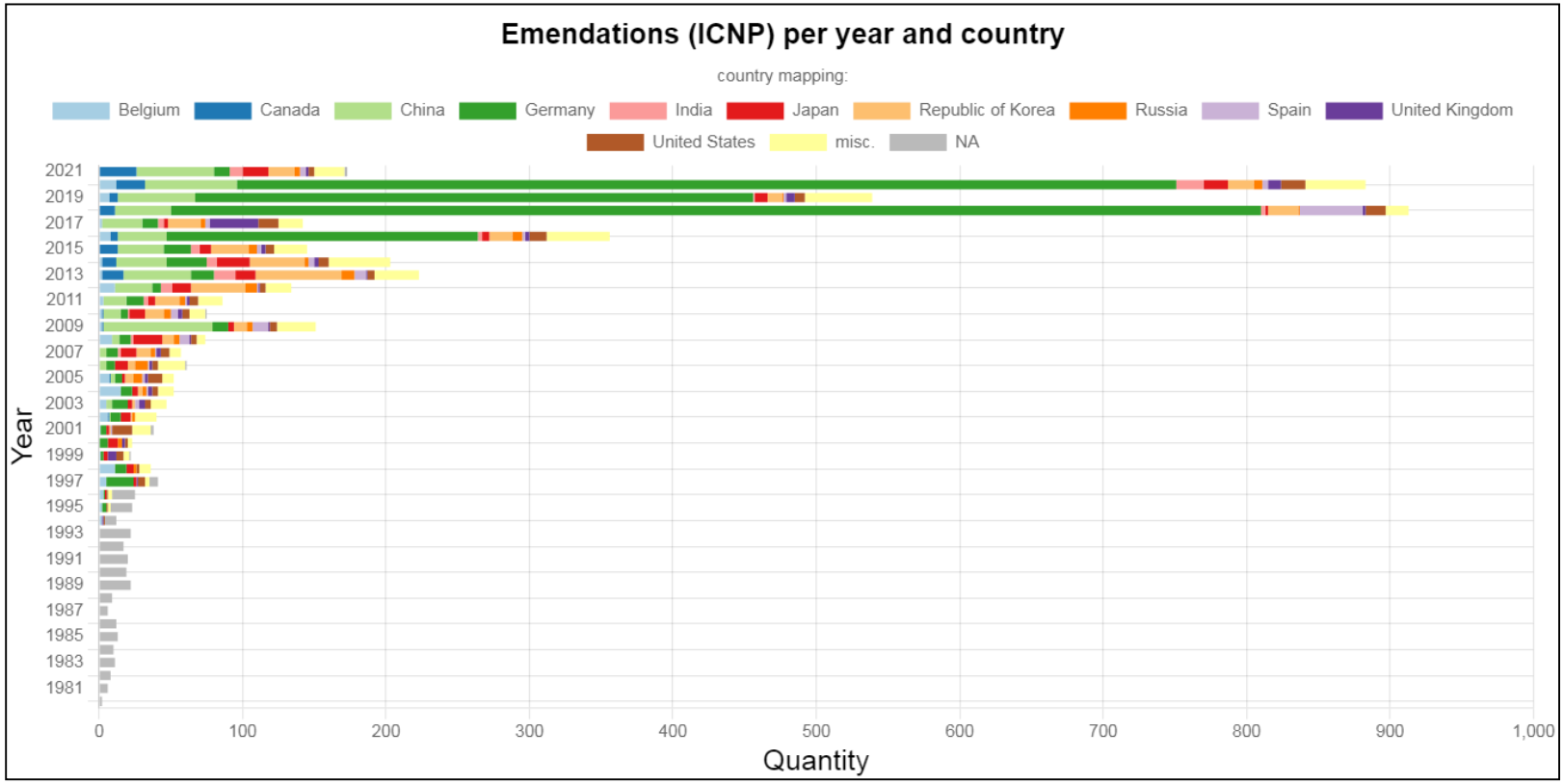


LPSN-DSMZ's website (https://lpsn.dsmz.de)

LPSN: Creator by J.P. Euzéby

Emendations (ICNP) per year and country

The International Code of Nomenclature of Prokaryotes (ICNP)

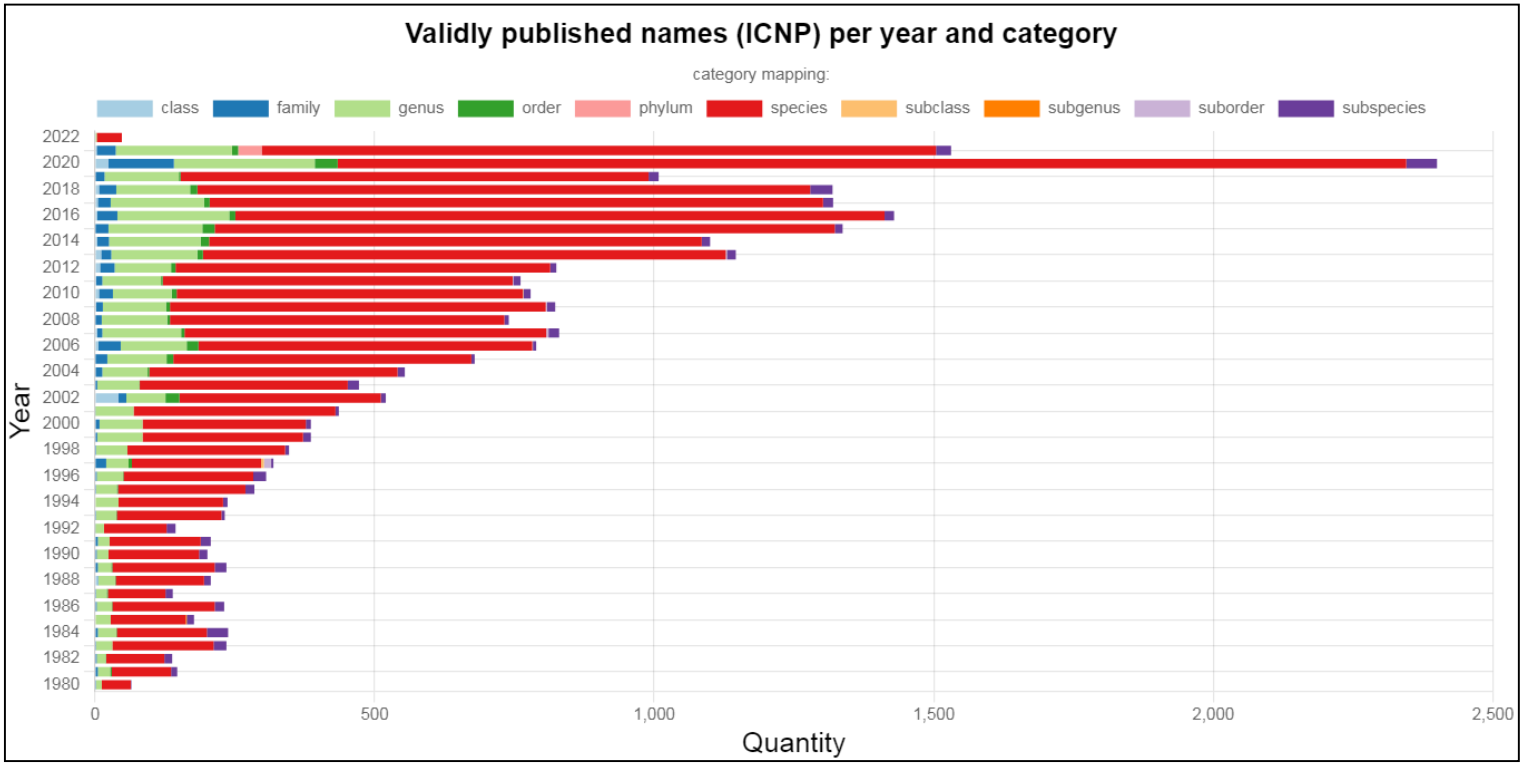


LPSN-DSMZ's website (https://lpsn.dsmz.de)

LPSN: Creator by J.P. Euzéby

Validly published names (ICNP) per year and category

The International Code of Nomenclature of Prokaryotes (ICNP)

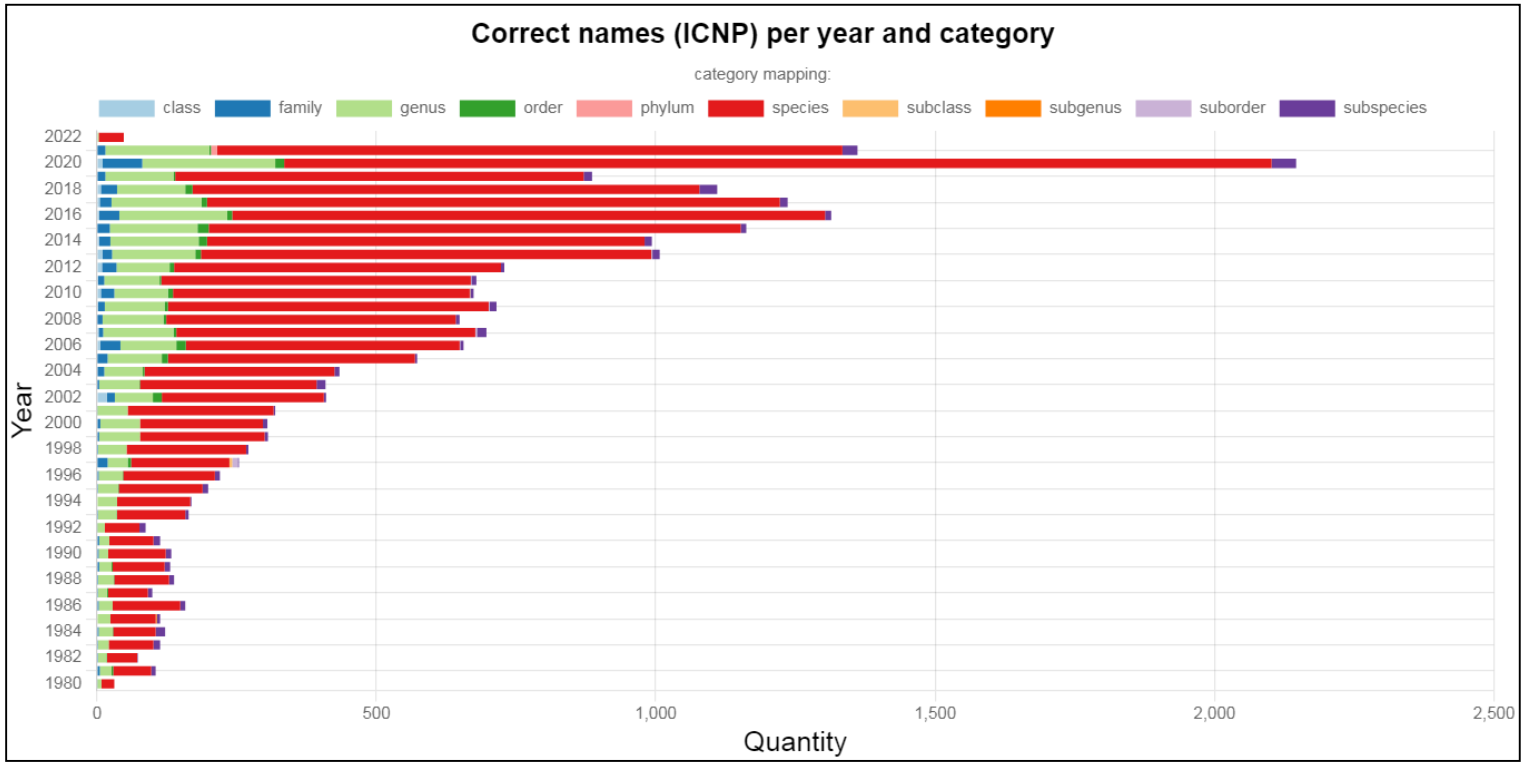


LPSN-DSMZ's website (https://lpsn.dsmz.de)

LPSN: Creator by J.P. Euzéby

Correct names (ICNP) per year and category

The International Code of Nomenclature of Prokaryotes (ICNP)



LPSN-DSMZ's website (https://lpsn.dsmz.de)

Bergey's Manual Trust Home Page
Online encyclopedia of Systematics of Archaea and Bacteria (BMSAB)

- David Hendricks Bergey was an American bacteriologist, born December 27, 1860.
- He was chairman of the Editorial Board for the first edition of Bergey's Manual of Determinative Bacteriology, published in 1923.
- The Determinative Manual has subsequently been published in a further eight editions, and Bergey's Manual Trust is currently publishing the second edition of Bergey's Manual of Systematic Bacteriology.

Bergey's Manual Trust Home Page
Online encyclopedia of Systematics of Archaea and Bacteria (BMSAB)

- Bergey's Manual Trust and John Wiley & Sons, Inc., are pleased to announce a new partnership to produce an online encyclopedia entitled Bergey's Manual of Systematics of Archaea and Bacteria (BMSAB).
- This electronic work will provide:
- Up-to-date descriptions of the taxonomy, systematics, ecology, physiology and other biological properties of all named prokaryotic taxa.
- Release is anticipated in the fall of 2014.
- BMSAB will be available via Wiley Online Library.
- Individual volumes of BMSAB will also be made available in e-book format.
- Visit Bergey's Manual Trust at http://www.bergeys.org/for more information on the Trust and our other publications.

Bergey's Manual Trust Home Page

The Bergey Award

The Bergey Award is given in recognition of outstanding contributions to bacterial taxonomy

- **Awardees:**
- 1979 - Roger Y. Stanier
- 1980 - John L. Johnson
- 1981 - Morrison Rogosa
- 1982 - Otto Kandler
- 1983 - Carl R. Woese
- 1984 - W. E. C. (Ed) Moore
- 1985 - Josef De Ley
- 1986 - William H. Ewing
- 1987 - Patrick A. D. Grimont
- 1988 - Lawrence G. Wayne
- 1989 - Hubert A. Lechevalier
- 1990 - M. David Collins
- 1991 - Erko Stackebrandt
- 1992 - Wolfgang Ludwig
- 1993 - Wesley E. Kloos
- 1994 - Friedrich Widdel
- 1995 - Michael Goodfellow
- 1996 - Karel Kersters
- 1997 - Rosmarie Rippka
- 1998 - David A. Stahl

- **Awardees:**
- 1999 - Barry Holmes
- 2000 - William B. Whitman
- 2001 - Lindsay I. Sly
- 2002 - Peter Vandamme
- 2003 - Peter Kämpfer
- 2004 - Rudolf Amann
- 2005 - Jean Paul Euzéby
- 2006 - David P. Labeda
- 2007 - Jürgen Wiegel
- 2008 - Hans-Jürgen Busse
- 2010 - Antonio Ventosa
- 2014 - Hans-Peter Klenk
- 2016 - Elizaveta Bonch-Osmolovskaya
- 2017 - Ramon Rossello-Mora
- 2018 - Jongsik Chun(creator of the EzBioCloud database)

Stop awarding the Bergey Medal from 2018 onwards

Bergey's Manual Trust Home Page
The Bergey Medal

- The Bergey Medal is awarded in recognition of outstanding and life-long contributions to the field of systematics of Bacteria and Archaea.
- Nominations may be submitted at any time to the Bergey's editorial office at bergeys@uga.edu.
- Nominations should include a cover letter

1. summarizing the candidate's qualifications,
2. the significance of contributions to the systematics of prokaryotes, and
3. their curriculum vitae.

Bergey medal front

Bergey's Manual Trust Home Page

The Bergey Medal

- **Recipients:**
- 1994 - R.G.E. Murray
- 1994 - Riichi Sakazaki
- 1994 - V.B.D. Skerman
- 1994 - Eyvind A. Freundt
- 1995 - Norbert Pfennig
- 1995 - Dorothy Jones
- 1995 - Norberto Palleroni
- 1996 - Thomas D. Brock
- 1996 - Marvin P. Bryant
- 1996 - John G. Holt
- 1996 - Emilio Weiss
- 1997 - Lillian H. Moore
- 1997 - Ralph S. Wolfe
- 1997 - George A. Zavarzin
- 1998 - Kjell Bøvre
- 1998 - Holger Jannasch
- 1998 - Julius P. Kreier
- 1998 - Peter H. A. Sneath
- 1999 - Wilhelm Frederiksen
- 1999 - James W. Moulder
- 1999 - Karl O. Stetter
- 1999 - Hans Trüper
- 2000 - Peter Hirsch
- 2000 - Hans Reichenbach

- **Recipients:**
- 2000 - Eiko Yabuuchi
- 2001 - Joseph G. Tully
- 2001 - Floyd E. Dewhirst
- 2001 - E. Imre Friedmann
- 2002 - Don J. Brenner
- 2002 - Rita R. Colwell
- 2002 - Noel R. Krieg
- 2003 - Monique Gillis
- 2003 - Hans Hippe
- 2004 - David R. Boone
- 2005 - Richard W. Castenholz
- 2005 - Kazau Komagata
- 2006 - Klaus P. Schaal
- 2008 - Fergus Priest
- 2008 - James T. Staley
- 2009 - Karl Heinz Schleifer
- 2009 - Erko Stackebrandt
- 2011 - Michael Goodfellow
- 2011 - Zhiheng Liu
- 2011 - Ji-Sheng Ruan
- 2011 - James M. Tiedje
- 2018 - Fritz Widdel
- 2018 - Niall Logan

Stop awarding the Bergey Medal from 2018 onwards

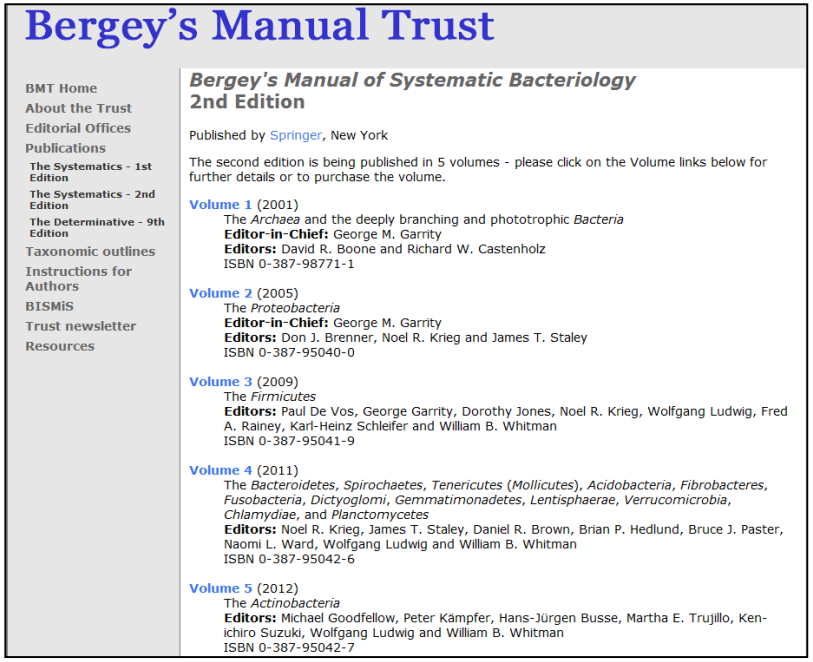
Bergey's Manual Trust
Publications

Bergey's Manual of Systematic Bacteriology
1st Edition

John G. Holt, Editor-in-Chief
Williams & Wilkins, Baltimore, MD

Published in 4 volumes:

Volume 1 (1984)
Gram-negative *Bacteria* of general, medical, or industrial importance
ISBN 0-683-04108-8
Volume 2 (1986)
Gram-positive *Bacteria* other than *Actinomycetes*
ISBN 0-683-07893-3
Volume 3 (1989)
Archaeobacteria, *Cyanobacteria*, and remaining Gram-negative *Bacteria*
ISBN 0-683-07908-5
Volume 4 (1989)
Actinomycetes
ISBN 0-683-09061-5



Bacterial taxonomy

SILVA

On-line databases on bacterial, archaea and eukaryotic taxonomy

SILVA logo

- Most life on earth is microbial, belonging to the 'Bacteria' and 'Archaea' domains, and to numerous lineages of microbial 'Eukaryota' (e.g. protists).
- Less than 1% of microbes are cultivable, and therefore diversity was vastly underestimated by traditional microbiological methods.
- The known extent of microbial diversity has grown and continues to grow rapidly as sequence-based methods are used to characterize microbes.

Yilmaz *et al.*,2013

Bacterial taxonomy

SILVA logo

SILVA

Databases on bacterial, archaea and eukaryotic taxonomy

- One of the major breakthroughs in the study of the diversity of microbes was the use of the ribosomal rRNA (rRNA) gene sequences, particularly of the small subunit (SSU; also called 16S rRNA for Bacteria and Archaea and 18S rRNA for Eukaryota).
- Appropriate taxonomic classification in sequence databases is crucial for organizing and cataloging microbial diversity.
- The SILVA rRNA gene databases use a phylogenetic tree-guided manual curation approach for the taxonomy of Bacteria, Archaea and Eukaryota.

Yilmaz *et al.*,2013

Bacterial taxonomy

SILVA

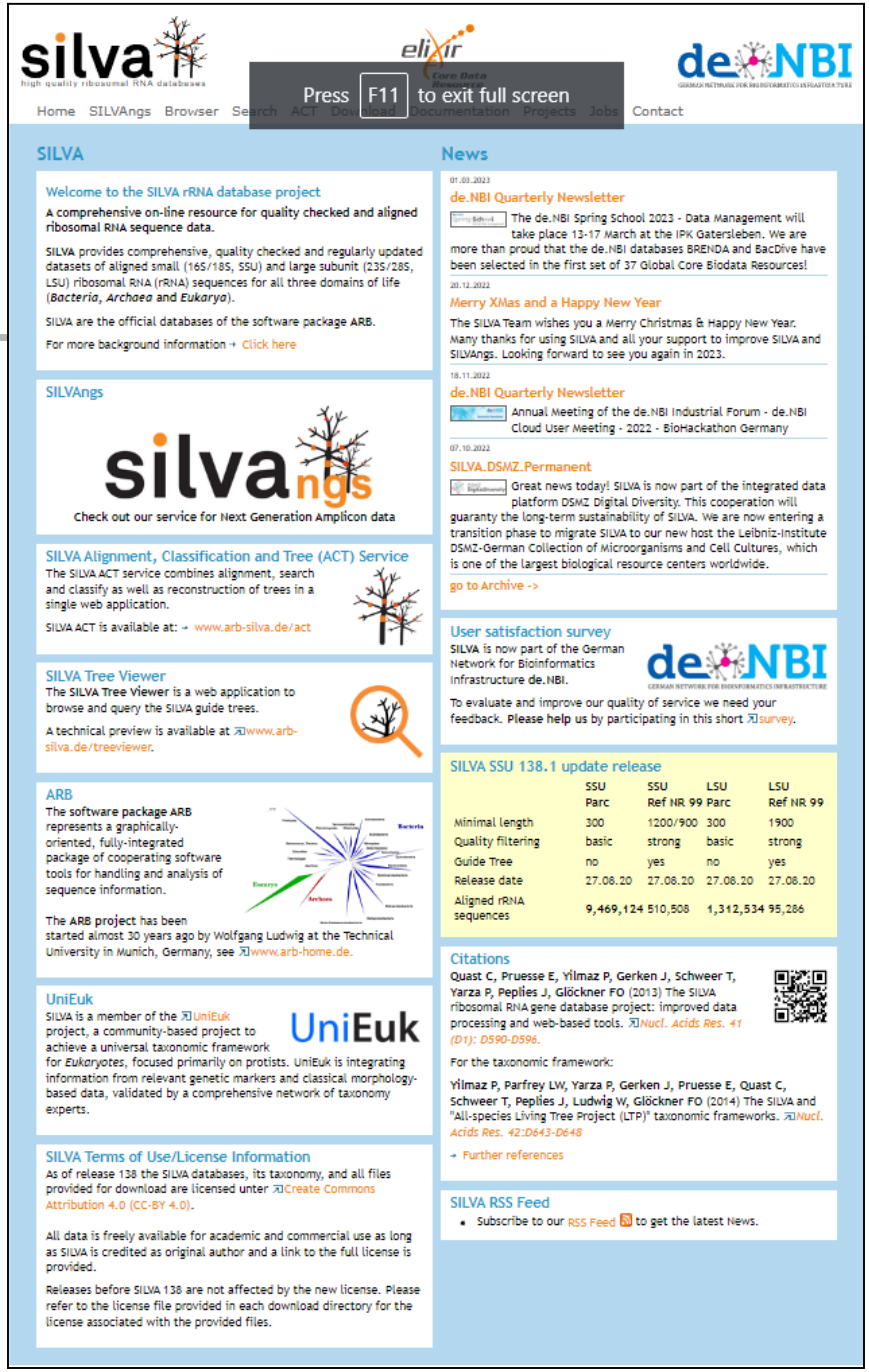
Databases on bacterial, archaea and eukaryotic taxonomy

- SILVA (from Latin silva, forest, http://www.arb-silva.de) is a comprehensive web resource for up-to-date, quality-controlled databases of aligned small(16S/18S, SSU) and large subunit (23S/28S, LSU) ribosomal RNA (rRNA) sequences for all three domains of life (Bacteria, Archaea and Eukarya).
- SILVA as a comprehensive web resource provides:

1. a manually curated(accurate) taxonomy for all three domains of life, based on representative phylogenetic trees for the small and large-subunit rRNA genes.
2. It is also an authorized provider of the LPSN(List of Prokaryotic Names with Standing in Nomenclature) taxonomy.

SILVA the rRNA Gene Databases

- SILVA represents the world's leading public database for ribosomal RNA (rRNA) gene sequences for all three domains of life (Bacteria, Archaea, and Eukarya).
- It also provides comprehensive, quality checked, and regularly updated data sets of aligned small (16S/18S, SSU) and large subunit (23S/28S, LSU) sequences for for all three domains of life (Bacteria, Archaea, and Eukarya).



Bacterial taxonomy

SILVA

Databases on bacterial, archaea and eukaryotic taxonomy

Bacterial taxonomy

SILVA

Databases on bacterial, archaea and eukaryotic taxonomy

- How big is the SILVA database?
- The uncompressed version is ~6.8 GB and the compressed version is 241 MB. recreated seed database (5736 bacteria, 81 archaea, and 1824 eukarya sequences).
- The actual reference alignment that SILVA uses with their SINA aligner is called the SEED alignment.

What is alignment seed?
Seeded alignment is the dominant technique for large-scale genomic sequence comparisons, and BLASTN is the most popular implementation of it.
SINA Aligner: will align your rRNA gene sequences according to the global SILVA alignment for rRNA genes.

Bacterial taxonomy

SILVA

Databases on bacterial, archaea and eukaryotic taxonomy

- With the newly developed SILVA-NGS pipeline, we will grant easy access to the SILVA taxonomy for the classification of rRNA gene amplicon data.
- SILVA-NGS accepts any kind of short- and long-read sequence rRNA gene data in FASTA format and performs quality control, alignment and classification of rRNA genes based on the curated SILVA taxonomy.
- All steps (upload, progress monitoring, visualization of results and download of data) can be geared via the SILVA-NGS web-interface.
- The system is available at www.arb-silva.de/ngs.
- Finally, we would like to emphasize that the **SILVA taxonomy curation is an open and transparent process, and input from users and experts is highly appreciated**.

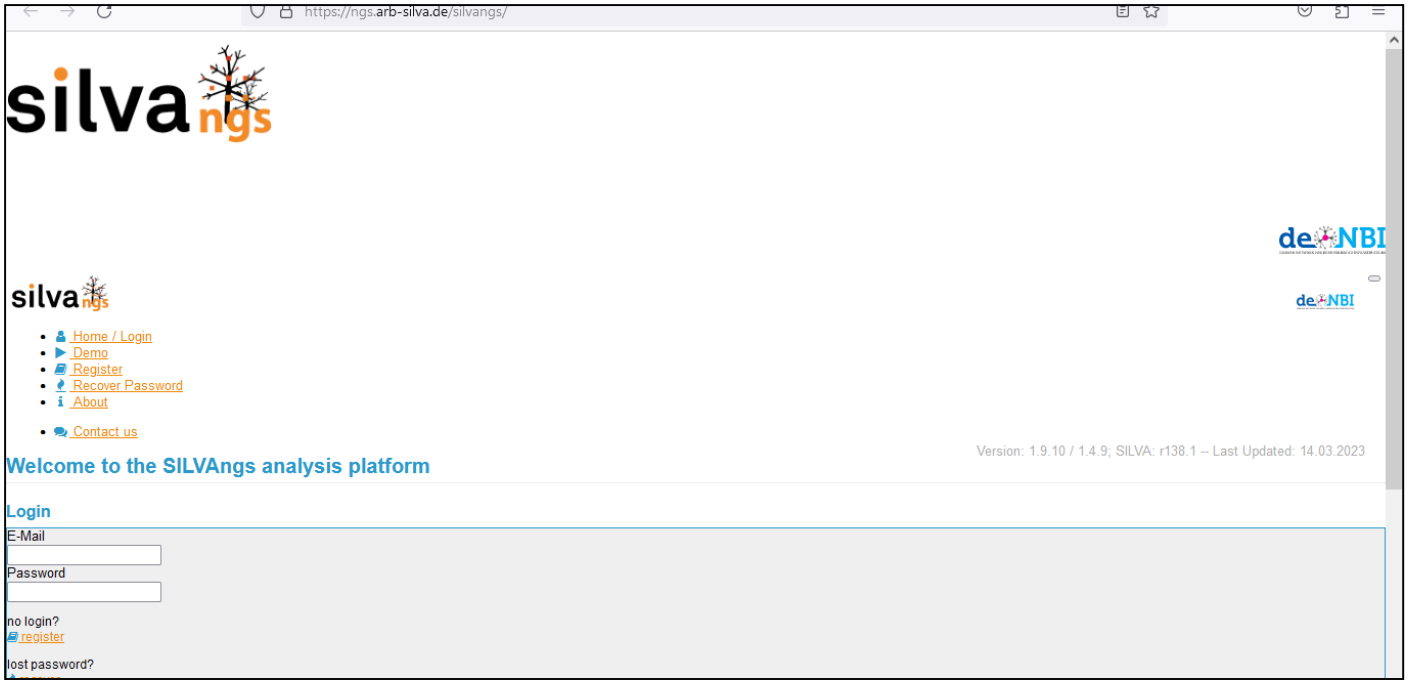
Yilmaz *et al.*,2013

Bacterial taxonomy

SILVA

Databases on bacterial, archaea and eukaryotic taxonomy

- The system is available at www.arb-silva.de/ngs.
- **Welcome to the SILVAngs analysis platform**



Bacterial taxonomy

'The All-Species Living Tree' Project logo

LTP

Databases on bacterial and archaea taxonomy

- The All-Species Living Tree Project (LTP) is an international initiative for the creation and maintenance of highly curated 16S rRNA gene sequence databases, alignments and phylogenetic trees for all the type strains of Archaea and Bacteria.
- At one stage, 23S sequences were also collected but this has since stopped.
- Like SILVA, LTP is also an authorized provider of the LPSN taxonomy.

LPSN: List of Prokaryotic Names with Standing in Nomenclature

Wikipedia,2023;..

Bacterial taxonomy

'The All-Species Living Tree' Project logo

LTP

Databases on bacterial and archaea taxonomy

1. 'The All-Species Living Tree' Project is a collaboration between various academic groups/institutes, such as ARB, SILVA rRNA database project, and LPSN, with the aim of assembling a database of 16S rRNA sequences of all validly published species of Bacteria and Archaea.
2. Similar (and more recent) projects include the Genomic Encyclopedia of Bacteria and Archaea (GEBA), which focused on whole genome sequencing of bacteria and archaea.

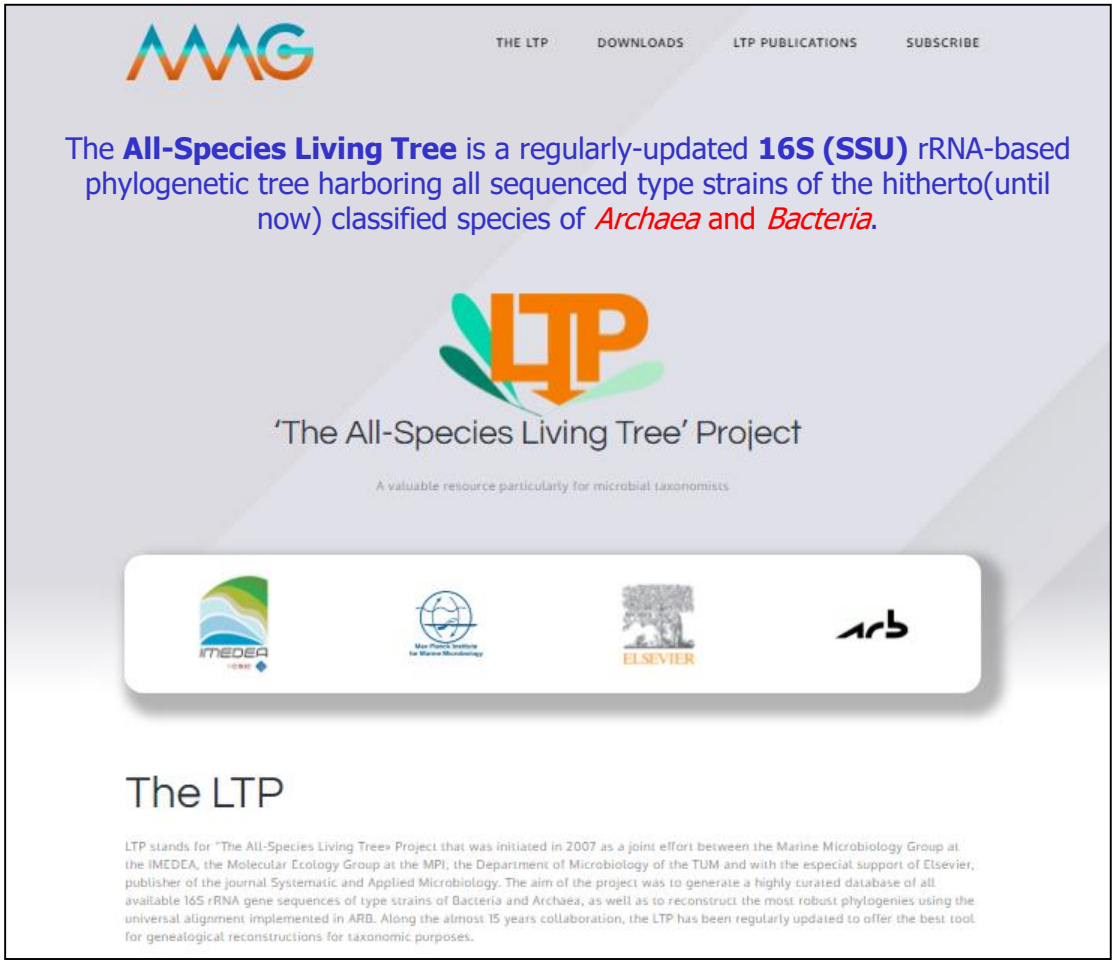
The ARB Project: It is a free software package for the phylogenetic analysis of rRNA and other biological as sequences such amino acids.

Wikipedia,2023

Bacterial taxonomy

The' All-Species Living Tree' Project (LTP)

Databases on bacterial and archaea taxonomy

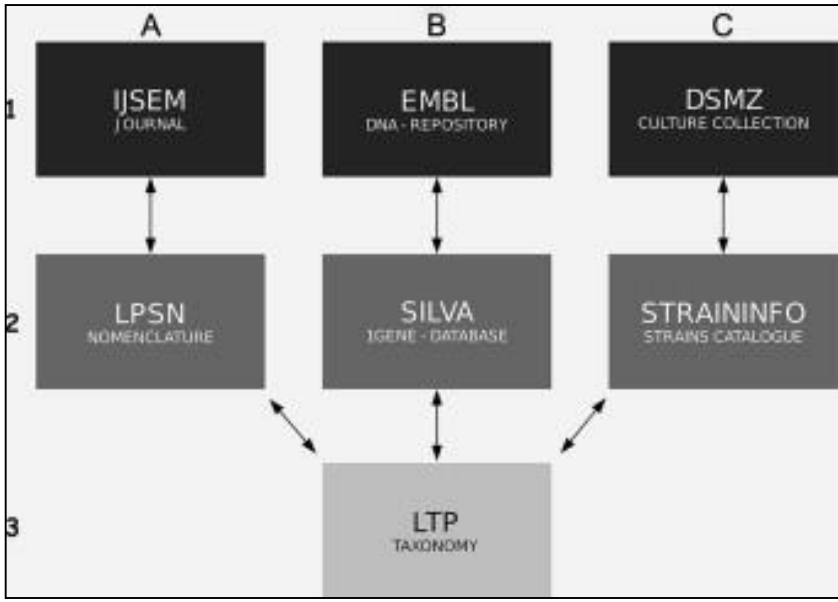


Bacterial taxonomy

The' All-Species Living Tree' Project (LTP)

Databases on bacterial and archaea taxonomy

- 'The All-Species Living Tree' Project is a collaboration between various academic groups/institutes, such as ARB, SILVA rRNA database project, European Molecular Biology Laboratory(EMBL), LPSN and some more.



Yarza and Maunoz,2014

The' All-Species Living Tree' Project (LTP)

List of species that are susceptible to being reclassified due to a clear distant affiliation with the *sensu stricto* members of their taxon

- The current release LTP_2020 (Ludwig *et al.*,2021) introduces

A. important changes over the former as a large set of sequences had been substituted,

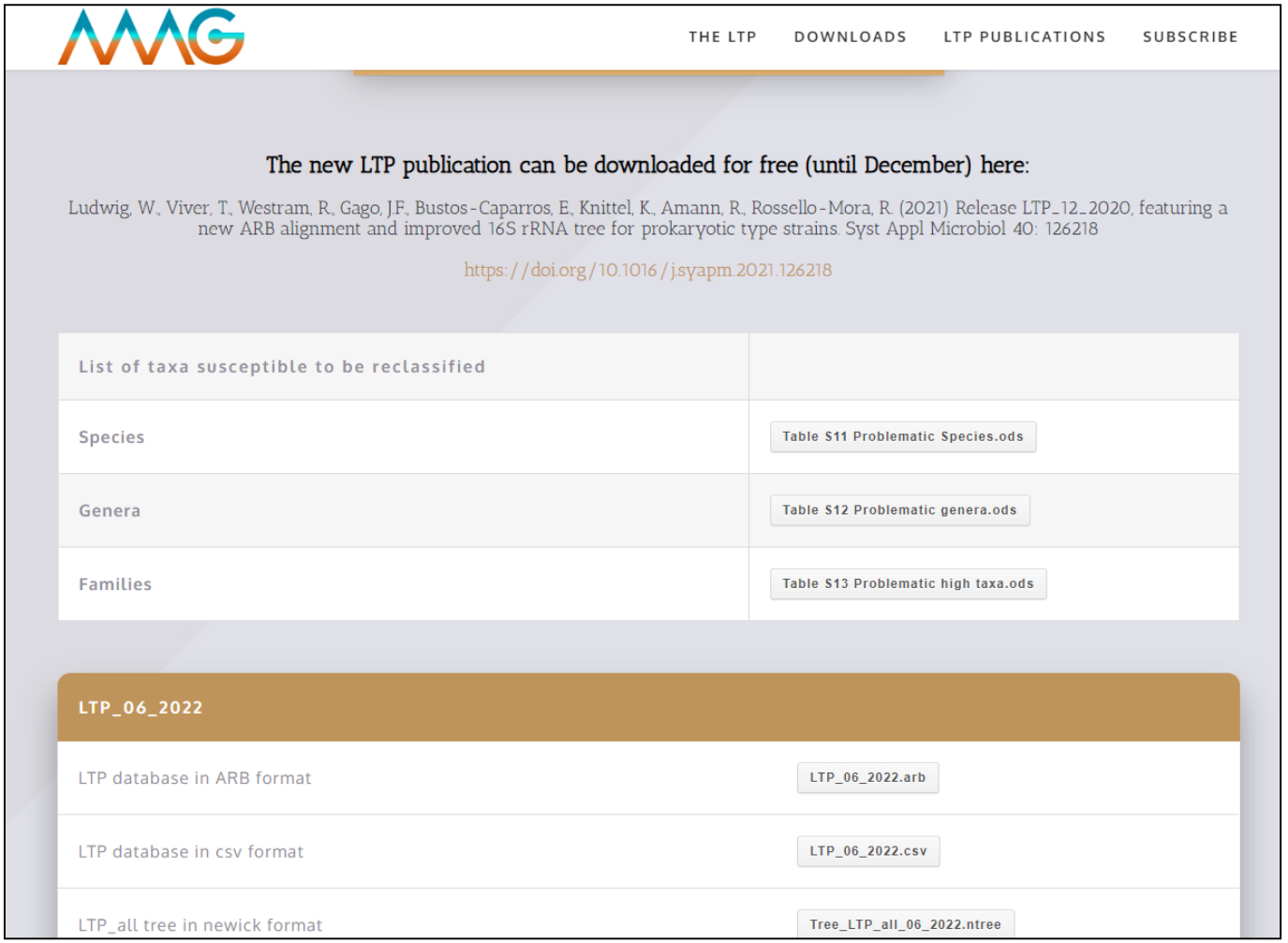
B. the universal alignment has been thoroughly improved to accommodate the new sequences that contain:

1. insertions,
2. as well to improve the helix structures, and
3. has increased from 42,283 to 98,863 homologous positions.
4. We also provide the most robust phylogenetic tree currently achievable with the available dataset.

sensu stricto= in a narrow or strict sense. In taxonomy(of taxon).

The' All-Species Living Tree' Project (LTP)

List of species, genera and families that are susceptible to being reclassified due to a clear distant affiliation with the *sensu stricto* members of their taxon



The' All-Species Living Tree' Project (LTP)

List of species that are susceptible to being reclassified due to a clear distant affiliation with the *sensu stricto* members of their taxon

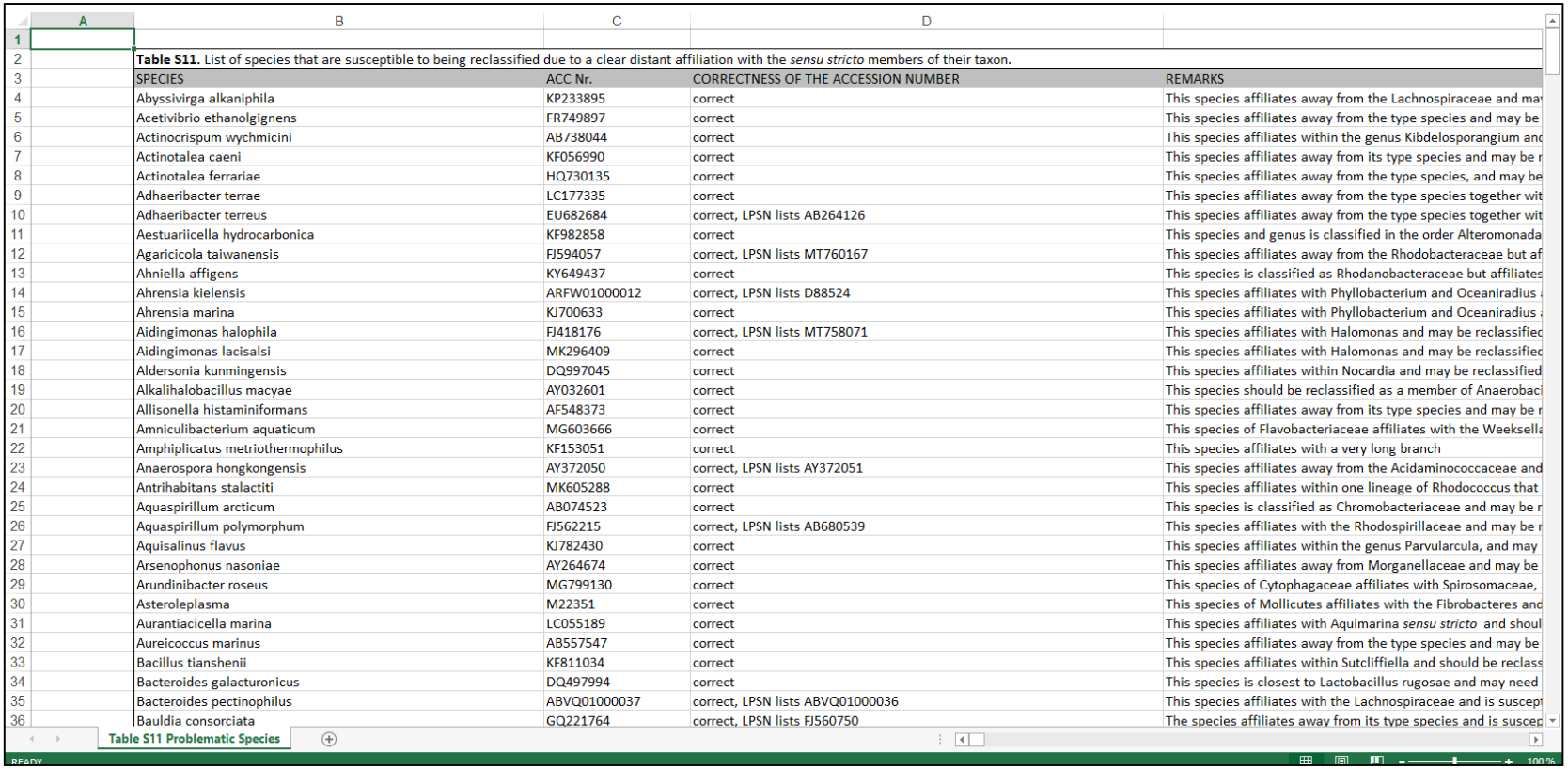


Table S11. List of species that are susceptible to being reclassified due to a clear distant affiliation with the *sensu stricto* members of their taxon.

SPECIES	ACC Nr.	CORRECTNESS OF THE ACCESSION NUMBER	REMARKS
Abyssivirga alkaniphila	KP233895	correct	This species affiliates away from the Lachnospiraceae and ma
Acetivibrio ethanolgignens	FR749897	correct	This species affiliates away from the type species and may be
Actinocrispum wychmicini	AB738044	correct	This species affiliates within the genus Kibdelosporangium and
Actinotalea caeni	KF056990	correct	This species affiliates away from its type species and may be r
Actinotalea ferrariae	HQ730135	correct	This species affiliates away from the type species, and may be
Adhaeribacter terrae	LC177335	correct	This species affiliates away from the type species together wit
Adhaeribacter terreus	EU682684	correct, LPSN lists AB264126	This species affiliates away from the type species together wit
Aestuariicella hydrocarbonica	KF982858	correct	This species and genus is classified in the order Alteromonada
Agaricicola taiwanensis	FJ594057	correct, LPSN lists MT760167	This species affiliates away from the Rhodobacteraceae but af
Ahniella affigens	KY649437	correct	This species is classified as Rhodanobacteraceae but affiliates
Ahrensia kielensis	ARFW01000012	correct, LPSN lists D88524	This species affiliates with Phyllobacterium and Oceaniradius
Ahrensia marina	KJ700633	correct	This species affiliates with Phyllobacterium and Oceaniradius
Aidingimonas halophila	FJ418176	correct, LPSN lists MT758071	This species affiliates with Halomonas and may be reclassifiec
Aidingimonas lacisalsi	MK296409	correct	This species affiliates with Halomonas and may be reclassifiec
Aldersonia kunmingensis	DQ997045	correct	This species affiliates within Nocardia and may be reclassified
Alkalihalobacillus macyae	AY032601	correct	This species should be reclassified as a member of Anaerobaci
Allisonella histaminiformans	AF548373	correct	This species affiliates away from its type species and may be r
Amniculibacterium aquaticum	MG603666	correct	This species of Flavobacteriaceae affiliates with the Weeksella
Amphiplicatus metriothermophilus	KF153051	correct	This species affiliates with a very long branch
Anaerospora hongkongensis	AY372050	correct, LPSN lists AY372051	This species affiliates away from the Acidaminococcaceae and
Antrihabitans stalactiti	MK605288	correct	This species affiliates within one lineage of Rhodococcus that
Aquaspirillum arcticum	AB074523	correct	This species is classified as Chromobacteriaceae and may be r
Aquaspirillum polymorphum	FJ562215	correct, LPSN lists AB680539	This species affiliates with the Rhodospirillaceae and may be r
Aquisalinus flavus	KJ782430	correct	This species affiliates within the genus Parvularcula, and may
Arsenophonus nasoniae	AY264674	correct	This species affiliates away from Morganellaceae and may be
Arundinibacter roseus	MG799130	correct	This species of Cytophagaceae affiliates with Spirosomaceae,
Asteroleplasma	M22351	correct	This species of Mollicutes affiliates with the Fibrobacteres and
Aurantiacicella marina	LC055189	correct	This species affiliates with Aquimarina *sensu stricto* and shoul
Aureicoccus marinus	AB557547	correct	This species affiliates away from the type species and may be
Bacillus tianshenii	KF811034	correct	This species affiliates within Sutcliffiella and should be reclass
Bacteroides galacturonicus	DQ497994	correct	This species is closest to Lactobacillus rugosae and may need
Bacteroides pectinophilus	ABVQ01000037	correct, LPSN lists ABVQ01000036	This species affiliates with the Lachnospiraceae and is suscept
Bauldia consorciata	GQ221764	correct, LPSN lists FJ560750	The species affiliates away from its type species and is suscep

Table S11 Problematic Species

sensu stricto= in a narrow or strict sense. In taxonomy(of taxon).

The' All-Species Living Tree' Project (LTP)

List of genera that are susceptible to being reclassified due to a clear distant affiliation with the *sensu stricto* members of their taxon

	A	B	C
1			
2		**Table S12.** List of genera that are susceptible to being reclassified.	
3		Genus	Remark
4		Acidovorax	The genus is polyphyletic and needs revision
5		Actinobacillus	The genus is polyphyletic and needs revision
6		Actinokineospora	The genus is polyphyletic and needs revision
7		Actinomadura	The genus is polyphyletic and needs revision
8		Actinomyces	The genus is polyphyletic and needs revision
9		Actinoplanes	The genus is polyphyletic and needs revision
10		Actinosynema	This genus affiliates within Lentzea and may be susceptible to reclassification
11		Afipia	The genus is polyphyletic and needs revision
12		Alkalihalobacillus	The genus is polyphyletic and needs revision
13		Allorhizobium	The genus is polyphyletic and needs revision
14		Altererytrhobacter	The genus is polyphyletic and needs revision
15		Alteromonas	The genus is polyphyletic and needs revision
16		Ancylobacter	The genus is polyphyletic and needs revision
17		Aquibacillus	The genus is polyphyletic and needs revision
18		Archaeoglobus	The genus is polyphyletic and needs revision
19		Arthrobacter dispersed	The genus is polyphyletic and needs revision
20		Aureimonas	The genus is polyphyletic and needs revision
21		Azohydromonas	The genus must be reclassified within the Comamonadaceae
22		Bacillus miscanthi	The genus is polyphyletic and needs revision
23		Beijerinckiaceae dispersed	The genera Pseudochelatococcus, Methylocapsa, Methylocella, Methylorosula and Methylovirgula may need to change the family as they do not affiliate with Beijerinckia
24		Bizionia	The genus is polyphyletic and needs revision
25		Bizionia dispersed	The genus is polyphyletic and needs revision
26		Brenneria	The genus is polyphyletic and needs revision
27		Brucella	The genus is polyphyletic and needs revision
28		Caloramator	The genus is polyphyletic and needs revision
29		Celerinatantimonas	The genus is not monophyletic and could be susceptible to reclassification as Aliagarivorans, both genus and family
30		Cellulophaga	The genus is polyphyletic and needs revision
31		Cereibacter	The genus is polyphyletic and needs revision
32		Chelatococcus	The genus is polyphyletic and needs revision
33		Chryseolinea and Ohtaekwangia	Both genera are classified within the Fluvivirgaceae, however, together with Chryseotalea they could be a new family within Cytophagales
34		Clostridium	The genus is polyphyletic and needs revision
35		Comamonas	The genus is polyphyletic and needs revision
36		Cryobacterium	The genus is polyphyletic and needs revision

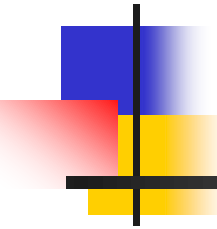
Table S12 Problematic genera

sensu stricto= in a narrow or strict sense. In taxonomy(of taxon).

eLMSG (eLibrary of Microbial Systematics and Genomics)

- eLMSG (eLibrary of Microbial Systematics and Genomics) is a web microbial library that integrates not only taxonomic information, but also genomic information and phenotypic information (including morphology, physiology, biochemistry and enzymology).
- The taxonomic system of eLMSG is manually curated and composed of all validly and some effectively published taxa.
- For each taxon, the Latin name, taxon ID (NCBI taxonomy), etymology, rank, lineage, the dates of effective and/or valid publication, feature descriptions, nomenclature type and references for the proposal and emendations during the history of the taxon are presented.
- Besides these data, the species taxa contain information about 16S rRNA gene and/or genome sequences. All publicly available genome data of each type species including both type and non-type strains were collected, and if needed, re-annotated using the standardized analysis pipeline.
- Furthermore, pan-genomic data analyses were conducted for species with ≥5 genome sequences available. Finally, for all type species, taxonomically relevant phenotypic data were extracted and curated from literatures, which were further indexed into eLMSG as searchable and analyzable data records.
- Taken together, eLMSG is a comprehensive web platform for studying mi- crobial systematics and genomics, potentially useful for better understanding microbial taxonomy, natural evolutionary processes and ecological relationships.

re3data.org: eLibrary of Microbial Systematics and Genomics; editing status 2022-03-24; re3data.org - Registry of Research Data Repositories. http://doi.org/10.17616/R31NJN4V last accessed: 2024-12-24.



An Introduction to Bacterial Taxonomy

Plant Pathogenic Bacteria

Taxonomy
The oldest biological sciences

1. Science of Classification of organisms.
2. Hopes to show relationships among organisms.
3. Is a way to provide universal identification of an organism?
4. Why do we care things are related?

- It is perhaps one of the oldest biological sciences, having been in existence for at least 2,400 years, since Aristotle devised the first hierarchy based on creationist and essentialist tenets/ideas (Ereshefsky,1994).

Rosselló-Mora,2005; www.smccd.edu

Taxonomy
History of bacterial taxonomy

- 1675: Discovery of bacteria by Antonie Van Leeuwenhoek.
- 1880s: Introduction of "pure culture" techniques (Koch & Petri).
- 1900-1950: Taxonomy based on morphology.
- 1950s: Numerical taxonomy & chemotaxonomy.
- 1960s: Introduction of genotypical methods.
- Present: Polyphasic taxonomy

Coenye,2003

Taxonomy

The development of prokaryote taxonomy

- Prokaryote classification is the youngest and most dynamic among the different classifications of living organisms.
- Prokaryotes were not even known to exist until a few centuries ago, due to their small size and the fact that they can not normally be seen with the naked eye.
- The development of a reliable classification based on morphological traits as these for higher eukaryotes has been difficult because of the relative simplicity of the prokaryotes.
- The lack of a useful fossil record, together with the difficulties in identifying diagnostic characteristics from these small organisms have contributed to the instability of the prokaryote classification system.

Taxonomy
Prokaryote taxonomy

- Due to the simpler nature of the prokaryotes, taxonomists understood that in the technological developments of other disciplines, there were profitable parameters that would help to reflect natural relationships among organisms.
- The search for and establishment of novel criteria have been recurrent themes throughout the history of prokaryote taxonomy.

Rosselló-Mora,2005

Taxonomy

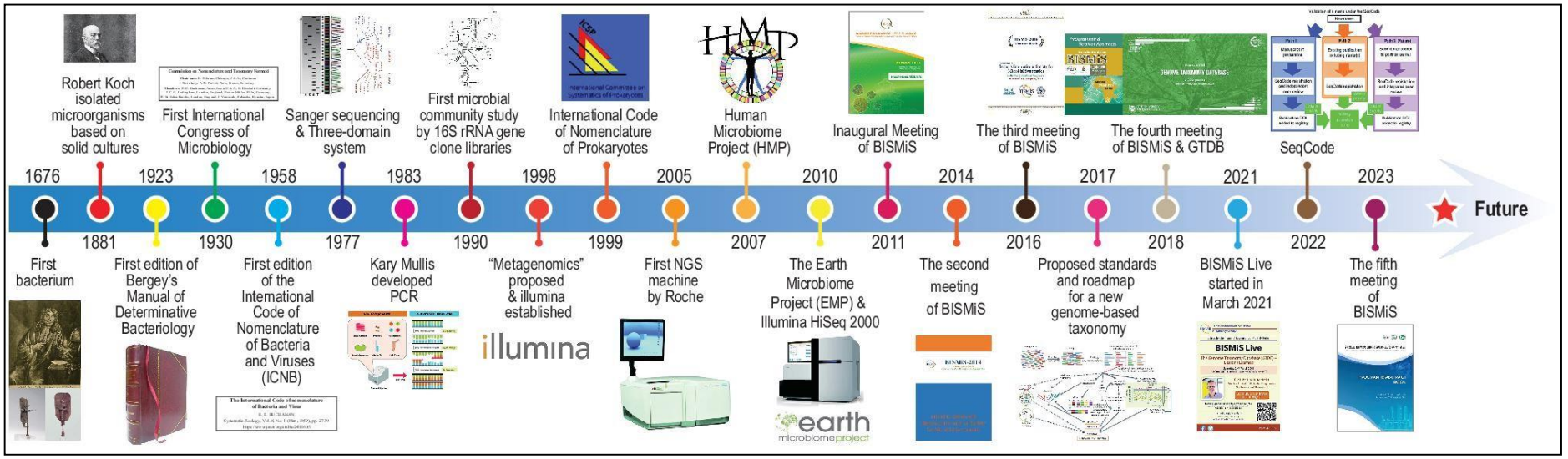
Ferdinand Julius Cohn (1828-1898)

A founder of modern microbiology; Pioneer of bacteriology

- Ferdinand Julius Cohn (24 January 1828–25 June 1898) was a German biologist.
- He is one of the founders of modern bacteriology and microbiology.
- Ferdinand J. Cohn was borne in the Jewish quarter of Breslau in the Prussian Province of Silesia (which is now Wroclaw, Poland).

Milestones of prokaryotic systematics

Recent decades have seen remarkable progress with the advent of whole-genome sequencing and bioinformatics, which allowed the development of a wide array of quantitative similarity criteria, such as average nucleotide identity (ANI), digital DNA-DNA hybridization (dDDH), average amino acid identity (AAI), protein ortholog clusters percentage (POCP) and core genome-based phylogenetics. Together, these approaches have improved the resolution and reliability of prokaryotic classifications and assignment at all taxonomic.



Jiao *et al.*,2024

Milestones of prokaryotic systematics

Significant advancements in prokaryotic systematics

Year	Name of event	Significance
1684	First observation of bacterial cells by Antonie van Leeuwenhoek	Origin of phenotype-based microbial taxonomy
1773–1786	Introduction of the first bacterial description by O.F. Müller	Introduction of morphological descriptions
1809	Nomination of *Polyangium vitellinum* by Johann Heinrich Friedrich Link	Beginning of nomenclature of bacterial species
1875	Purification of *Bacillus anthracis* by Robert Koch	Beginning of the pure-culture based bacteriology
1875	Attempt to establish formal rules for microbial nomenclature by Ferdinand Cohn	First attempt at formal microbial nomenclature
1896	First report and description of bacterial genera by K.B. Lehman and R. Neumann	The first report of bacterial genera
1901	The first manual of bacterial taxonomy was published by Professor Frederick Dixon Chester	Publication of the first handbook of bacterial taxonomy
1923	Bergey's *Manual* was published by the American bacteriologist, David Hendricks Bergey and American Society for Microbiology members	The birth of *Bergey's Manual of Determinative Bacteriology*
1977	Archaea were first classified as a separate group of prokaryotes based on phylogenies derived from 16S rRNA catalogs	The birth of the three-domain system and rRNA-based taxonomy
1984	Introduction of culture-independent studies of prokaryotic diversity in natural environments based on rRNA	The birth of the culture-independent study of prokaryotic diversity
1977–2000	Inclusion of DNA-DNA hybridization, GC content analysis	Beginning of the genome-based taxonomy of prokaryotes
2000–2023	Inclusion of dDDH, ANI, core/conserved gene-based phylogenetics, AAI, POCP, MALDI-TOF-MS	Comprehensive phenotype, genotype and protein-based taxonomy of prokaryotes

Jiao *et al.*,2024

Taxonomy

Ferdinand Julius Cohn (1828-1898)

A founder of modern microbiology; Pioneer of bacteriology

- About 1868 Cohn started to study bacteria.
- He was the first to recognize and study bacteriology as a separate science.
- He discovered the formation and germination of spores (called endospores) in certain bacteria, particularly in *Bacillus subtilis*. He was also the first to note endospores' resistance to high temperatures.
- His initial classification of bacteria consisted of four groups based on shape:

1. Sphaerobacteria (sphere-shaped),
2. Microbacteria (rod-shaped),
3. Desmobacteria (filamentous), and
4. Spirobacteria (screw-like shaped).

World of Microbiology and Immunology,2003, Britannica.com

Taxonomy
The Science of Classification

- The field of taxonomy is divided into three disciplines:

1. Classification, which means the orderly arrangement of units into groups.
2. Nomenclature, which means the labeling of units defined by classification.
3. Identification of unknowns with the units defined by classification and labeled by nomenclature (i.e., identification is the practical application of the arts of classification and nomenclature).

- Taxonomy can be viewed these three separate but interrelated areas.

Taxonomy
Classification vs. Nomenclature

1. Bacteria have no "official" classification scheme. Because, classification is mediated by the scientific method and there are no rules for it. In another word, taxonomy is not regulated by the code.
2. But they do have formal nomenclature, which is regulated by internationally accepted rules.
3. The International Committee on Systematics of Prokaryotes (ICSP) is responsible for updating and implementing the rules of the Code.

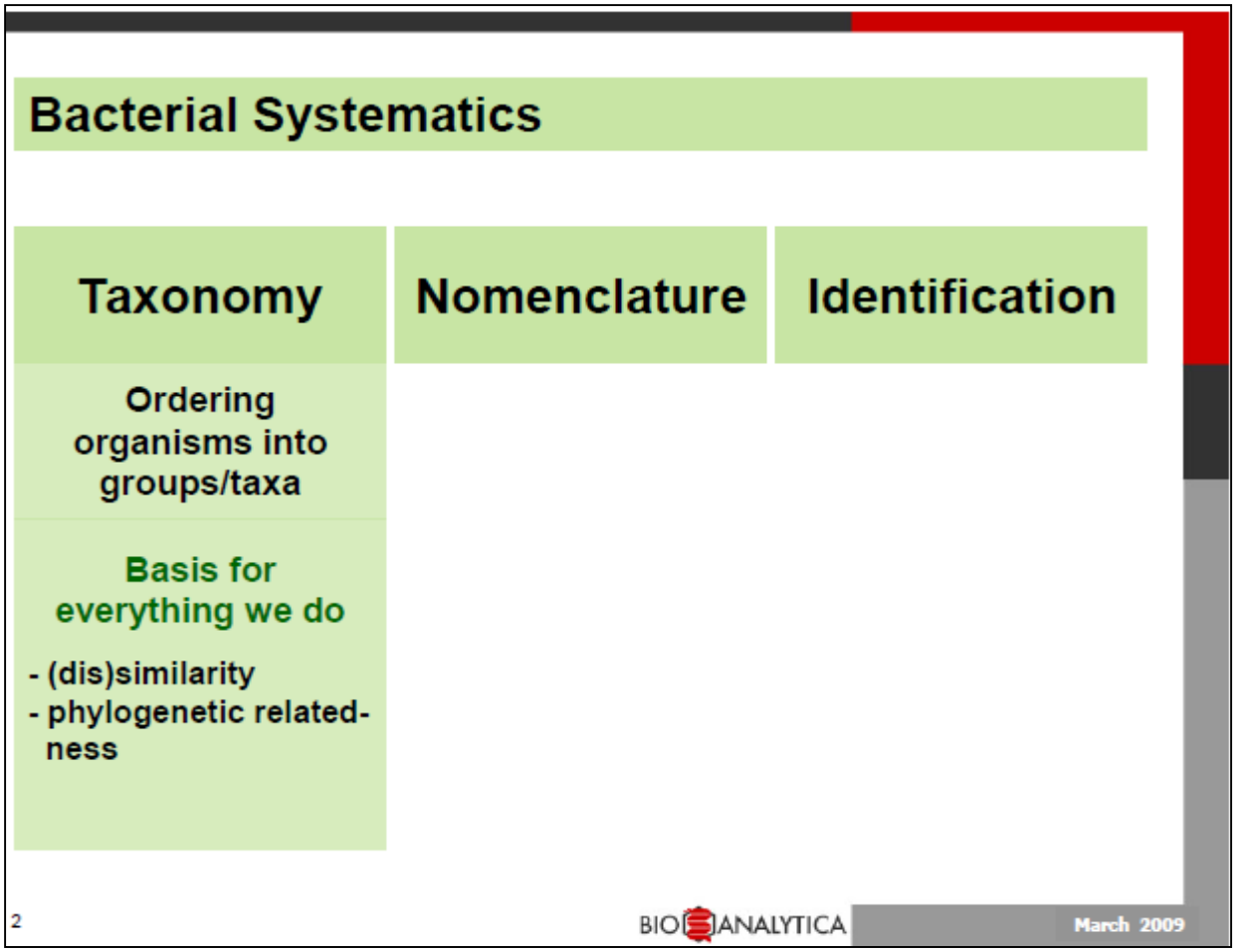
Parte,2013; Oren,2020

Taxonomy
Classification vs. Nomenclature

- Principle 1(4) of the Code states, **"Nothing in this Code may be construed (interpret) to restrict the freedom of taxonomic thought or action"**. What does that mean? Essentially, that anyone is free to design their own system of classifying bacteria.
- Code only deals with the way species, genera, and higher taxa of prokaryotes are named.

Oren,2020

Taxonomy
The Science of Classification

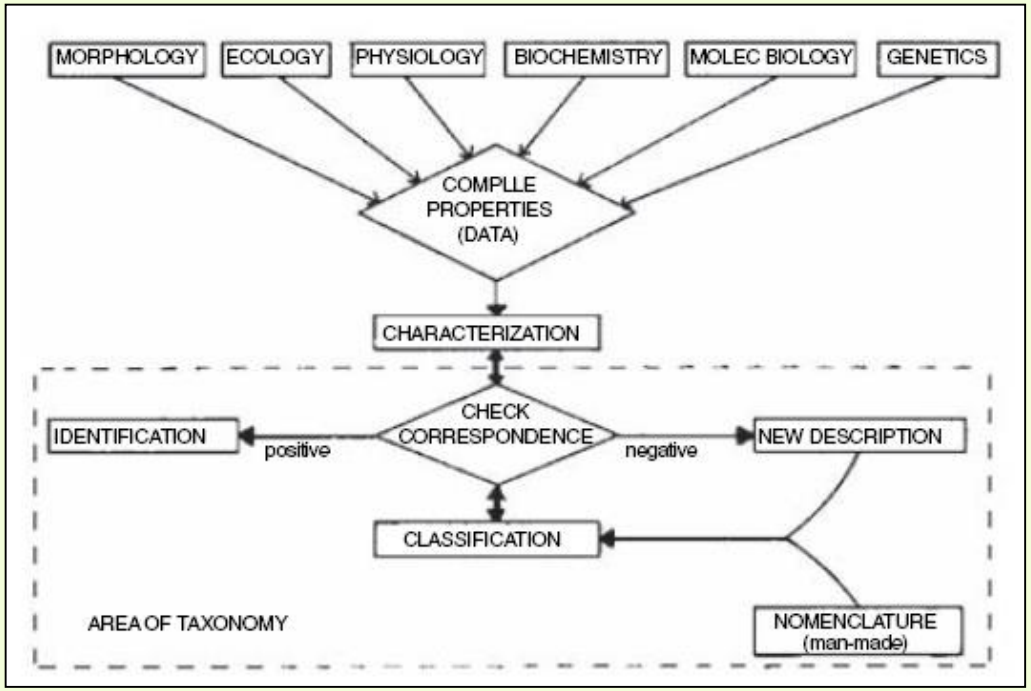


Altwegg,2005

Information flow diagram

Relationships between the characterization and the classification

- Information flow diagram indicating the relationships between the characterization and the classification of a bacterial strain.
- The two important working steps (indicated by diamond-shaped boxes) are compiling properties and checking for correspondence between characterization and classification.



The Prokaryotes (chapter 1.4),2006

Evolution of taxonomy

1. Classification dominated by phenotype.
2. Classification dominated by genotype.

- ("...classification must be a reflection of the natural relationships of bacteria, i.e. their degree of DNA similarity...")

Coenye,2003

Taxonomy

The development of prokaryote taxonomy

- During the last decades bacterial taxonomy has undergone remarkable changes.
- New categories of information of potential taxonomic value have become available:
- e.g.

1. Chemotaxonomy,
2. DNA base composition,
3. DNA-DNA hybridization,
4. Comparative sequence analysis of conserved macromolecules like 16S rDNA.

- Make possible very fine distinctions between organisms and reveal dissimilarities not detected before.

1. Identification

Basically, bacteria are identified/classified according to phenotypic differences, or by genotypic differences

Characteristics used for ID

- Phenotype (the set of observable characteristics of an individual resulting from the interaction of its genotype with the environment): Size, shape, metabolism, serology, FA analysis.
- Genotypes (the genetic constitution of an individual organism): Genetic makeup-PCR, DNA probes, Whole-genome sequencing using NGS technology (Next Generation Sequencing), etc.

2. Nomenclature

The handmaid of taxonomy

International Code of Nomenclature of Prokaryotes (The Code/Nomenclature Code).

Bacterial nomenclature

Some bacteriologists view

- The Code only deals with the way species, genera, and higher taxa of prokaryotes are named.
- Scientific Neutrality dictate that there is no official taxonomy or nomenclature.
- Don't just use the latest published name as it may not represent the best classification.
- Each individual researcher must choose the best classification and use the associated nomenclature in support of the classification.

Carolee T. Bull; Oren,2020

Bacterial nomenclature

Some bacteriologists view

- Yet a small number of plant bacteriologists appear to believe that, except for nomenclatural constructions regulated by the Code, they are free to use any names of their choosing and, unwittingly (without knowing) perhaps, breach (break)these basic linguistic conventions.

Young,2008

Bacterial nomenclature

What name should be used

What name should I use

Any **valid name** that correspond to the classifications that most accurately describe the relationships of the organisms according to your scientific judgement

Agrobacterium tumefaciens

vs

Rhizobium radiobacter

Erwinia carotovora

vs

Pectobacterium carotovorum

Carolee T. Bull

Bacterial nomenclature

The primary objective of Code of Nomenclature of Bacteria (now Prokaryotes)

- The Bacteriological Code governs names of prokaryotes in the ranks of:
- Class, Subclass, Order, Suborder, Family, Subfamily, Tribe, Subtribe, Genus, Subgenus, Species and Subspecies.
- Taxa above the rank of Class (Phylum, Kingdom, Division and Domain) are not covered by the Code.

Domain: The highest of taxonomic rank ('80s)
Kingdom (not used by most bacteriologists),1969
Phylum or division of the kingdom
Class
Order Family(related genera) Genus(related species) plural: Genera Species(related strains) both singular & plural Subspecies

Class also is covered by Code.

Parte,2013;..

Bacterial nomenclature
The primary objective of the Code

- The nomenclature of bacterial plant pathogens, like that of many other life forms, is constantly changing in response to:

1. New insights, and
2. Our understanding of relationships among bacteria.

Bull *et al.*,2010, 2012 & 2014

Bacterial nomenclature

Some more purposes

- Bacterial nomenclature provides the foundation from which:

1. Host-parasite disease relationships are defined,
2. Therapeutic regimens (plans) are developed, and
3. Epidemiological investigations (e.g., comparative analysis of bacterial strains involved in outbreaks) are instigated.

The primary objective of the Code
Key function of the Code

- The key function of the Code is to ensure that names are:

1. legitimate,
2. effectively published, and
3. validly published.

- The mechanism for determining which name is correct is based on:

1. legitimacy,
2. valid publication, and
3. priority of publication (Principle 6).

Tindall, 1999;Bull *et al.*,2008

Bacterial nomenclature
The primary objective of the Code

- The naming of bacteria is controlled by the International Code of Nomenclature of Prokaryotes (The Code/Nomenclature Code).
- The mechanism for determining which name is correct is based on:

1. Legitimacy
2. Effectively and validly publications
3. Priority of publication

Note: Readers should monitor IJSEM and the List of Prokaryotic Names with standing in Nomenclature (www.bacterio.net) to determine if new names have been validly published (Bull *et al.*,2012;2014).

Bacterial nomenclature
The primary objective of the Code

1. **Legitimacy:** The first step in the creation of a correct name is the proposal of a legitimate name. i.e. to be ensure that the proposed name fulfills the Rules and therefore is legitimate.
2. **Effectively publication:** The name must be effectively published in a suitable publication medium. i.e. it is the responsibility of the reviewers and editors of the publication in which the name is effectively published to confirm that the name meets the criteria governing legitimacy.
3. **Validly publication:** An effectively published name must be validly published in IJSEM before it can be considered to have a standing in nomenclature.
4. **Priority:** Priority in nomenclature is determined by the date of valid publication.

Legitimate names proposed in IJSEM are effectively and validly published in a single step.

Young,2008

The primary objective of the Code
Creation and determination of correct names
Legitimacy

- A name published following the Rules and/or the Standards.
- Names and epithets may be:
  - Legitimate: in accordance with the Rules;
  - Illegitimate: contrary to the Rules [Rule 23a Note 5].
- A legitimate name is in accordance with the Rules and only names in accordance with the Code may be validly published (Tindall,2008).

Note:
1. Legitimate names proposed in IJSEM are effectively and validly published in a single step, but the
2. Legitimate proposal, valid publication, and determination of priority, are distinct processes.

The primary objective of the Code

Creation and determination of correct names

Legitimacy

- The primary objective of the Code is to regulate the creation of a nomenclature of correct names.

1. Names should be unambiguous;
2. Names should be stable;
3. Names should be universal.

The primary objective of the Code
Validation list
Validly Published

- Each year the Validation Lists serve as an important source of information for phytobacteriologists.
- **Validly Published:** According to the Code, a proposed name is validly published when the name appears in IJSEM, an official publication of the ICSP (International Committee on Systematics of Prokaryotes) either as:

1. an original manuscript proposing the name, or
2. through listing on the Validation Lists.

The primary objective of the Code
Effectively published

- **Effectively Published:**
- Not everyone may wish to publish in that journal – and, of course, authors are free to publish wherever they wish.
- The Prokaryotic Code allows scientists to validate names effectively published in journals other than IJSEM.
- Once a newly proposed name is published, it is called 'effectively published'.

Oren,2020;..

The primary objective of the Code
Effectively published

- **Effectively Published:**
- Not everyone may wish to publish in that journal – and, of course, authors are free to publish wherever they wish.
- The Prokaryotic Code allows scientists to validate names effectively published in journals other than IJSEM.
- Once a newly proposed name is published, it is called 'effectively published'.

Oren,2020;..

The primary objective of the Code
Effectively publication

1. Publication of the name and description of a taxon in a recognized scientific printed and/or electronic publication [Rule 23a Note 5 and Rule 25a].
2. The publications must be conform to requirements laid down in the Bacteriological Code (1990 Revision).

- No other kind of publication (communications at meetings, minutes of meetings, abstracts of papers presented at meetings, catalogues of collections, microfilms, nonscientific periodicals, newsletters, patents...) is accepted as effective [Rule 25b].

Euzeby,2010; EzBioCloud,2019

The primary objective of the Code
Validation list
Effectively published

- To obtain the status of validly published, the authors must then take the second route:

1. a copy of the publication must be sent to the IJSEM editorial office with the request to include the names in the journal's bimonthly Validation List.
2. Such requests must be accompanied by further documentation – in particular, proof that the type strain of the new species and any subspecies are available from at least two culture collections in different countries.
3. The validation list editors of the journal will check the documents and, if all conditions for valid publication are met, the names will be listed in the next Validation List.

Oren,2020

The primary objective of the Code
Validation list
Effectively published

- Thus, scientists wishing to have new names and/or combinations included in a list should send the pertinent reprint or a photocopy or a PDF file thereof to the IJSEM Editorial Office.
- Any effectively published names be validated by sending the publication pdf to IJSEM with proof that (for species and subspecies) the designated type strain is available without restrictions from at least two public culture collections located in different countries.

1. Unfortunately, many more names of new pathogenic bacteria are effectively published in the literature, but never submitted to the IJSEM for validation.
2. Consequences of failing to do this can be damaging, as the name is 'effectively published' but never 'validated'.

www.ezbiocloud.net; Oren,2020

Bacterial nomenclature

Valid publication vs. Invalid publication

Valid publication

1. The name or epithet (species names) is published in the body of an article in the IJSEM or does appear in a Validation List [Rule 27].
2. A nomenclatural type (type strain) is designated for a new taxon.
3. The name or epithet accepted at the time of publication by the author who published it [Rule 28b].

Invalid publication

1. The name or epithet is not published in the body of an article in the IJSEM or does not appear in a Validation List [Rule 27].
2. A nomenclatural type (type strain) is not designated for a new taxon.
3. The name or epithet was not accepted at the time of publication by the author who published it [Rule 28b].

Validation Lists are lists published in the IJSB/IJSEM, serving as a mechanism for validly publishing prokaryotic names which have been effectively, but not validly published.

Bacterial nomenclature
Invalid publication

- Unfortunately, there are several cases where a name could not be validated.
- A name was used widely before 1980, but not included in the Approved Lists of Bacterial Names in 1980.
- A name was published outside of IJSEM, but never validated as either:
  - No one sent the pdf and further required documentation to IJSEM for validation.
  - The publication does not meet the minimal requirements of the Code. These minimal requirements will be explained later.
- Names not validly published are given in quotation marks, e.g. 'Selenomonas massiliensis' to differentiate from validly published names.

EzBioCloud,2019

Bacterial nomenclature
Valid publication/valid names

- The minimal and only requirements for validation of names of new species and subspecies according to the Prokaryotic Code are:

1. A type strain must be designated. Sometimes, a paper describing a new species does not mention which strain is the type strain. In such cases, the name cannot be validated.
2. The type strain should be deposited to two culture collections in two different countries. This ensures that the type strain remains available also for example when a culture collection discontinues its activity or loses the strain.
3. The type strain must be available to anyone through the culture collections. Patent strains cannot serve as type strains.

- If you want to patent a strain and restrict the distribution, it cannot serve as the nomenclatural type for a species or subspecies.
- A proper etymology and description should be given.

Nomenclatural type
Type strain

- In order for a name to be validly published, the proposal for the name must be accompanied by a number of criteria.
- Among these criteria - is the designation of a nomenclatural type.
- Just as physics has reference points for the meter or the kilogramm - so too does biology, with the nomenclatural types being the reference points for a taxon.
- In the cases of the species and subspecies, these reference points are represented by type strains.

Nomenclatural type
Importance of type strains

- Given the central importance of type strains, it is important that they be made available as widely as possible.
- The best course of action would appear to be to deposit type strains in a suitable culture collection, which should be able to maintain the distribution of the strain in the future.
- Clearly, depositing a type strain in such a way that it is not easy to access is counterproductive (causing problems) to the principle behind the deposit of type strains, that of making them widely and easily available for comparative purposes.

Validation Lists

List of new names and new combinations previously effectively, but not validly, published

Old version

LPSN Home
LPSN News
About LPSN
Contact
Resource description
All names
A-C
D-L
M-R
S-Z
Classifications
Support

Search LPSN
Google
Search

Please support LPSN!
Donate

Google Custom Search

Names validly published by announcement in Validation Lists

Contents

Introduction

Validation Lists are lists published in the IJSB/IJSEM, serving as a mechanism for validly publishing prokaryotic names which have been effectively, but not validly published.

Announcement in a Validation List is primarily the responsibility of the author(s) of the new names or new combinations. However, other individuals may also submit a new name or new combination for valid publication, provided it conforms to the rules of the code. Scientists wishing to have new names and/or combinations included in a list should send the pertinent reprint or a photocopy or a PDF file thereof to the IJSEM Editorial Office.

IJSEM Editorial Office, Microbiology Society, Charles Darwin House, 12 Roger Street, London WC1N 2JU, UK.
e-mail: info@microbiologysociety.org
Tel: +44 (0)20 7685 2400

2 February 2017 - The names published in the July 2016 issue (66/7) of IJSEM have been added to LPSN.

Note: From February 2020, LPSN was integrated into the DSMZ's Prokaryotic Nomenclature Up-to-date (PNU) database as an all-new service. LPSN-DSMZ's website (https://lpsn.dsmz.de).

Validation Lists

List no. 194,2020

Oren and Garrity

Int. J. Syst. Evol. Microbiol. 2020;70:4043-4049.

The names given in these lists have priority according to the issue of the IJSEM which they were published. (Rule 27).

Microbiology Society,2020

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

VALIDATION LIST NO 194

Oren and Garrity, *Int. J. Syst. Evol. Microbiol.* 2020;70:4043–4049

DOI 10.1099/ijsem.0.004244

MICROBIOLOGY SOCIETY

List of new names and new combinations previously effectively, but not validly, published

Aharon Oren[1,*] and George Garrity[2,*]

The purpose of this announcement is to effect the valid publication of the following effectively published new names and new combinations under the procedure described in the *International Code of Nomenclature of Prokaryotes* (2008 Revision). Authors and other individuals wishing to have new names and/or combinations included in future lists should send an electronic copy of the published paper to the IJSEM Editorial Office for confirmation that all of the other requirements for valid publication have been met. It is also a requirement of IJSEM and the ICSP that authors of new species, new subspecies and new combinations provide evidence that types are deposited in two recognized culture collections in two different countries. It should be noted that the date of valid publication of these new names and combinations is the date of publication of this list, not the date of the original publication of the names and combinations. The authors of the new names and combinations are as given below. Inclusion of a name on these lists validates the publication of the name and thereby makes it available in the nomenclature of prokaryotes. The inclusion of a name on this list is not to be construed as taxonomic acceptance of the taxon to which the name is applied. Indeed, some of these names may, in time, be shown to be synonyms, or the organisms may be transferred to another genus, thus necessitating the creation of a new combination.

Name/authors	Proposed as	Nomenclatural type	Priority§	Reference
Acetobacter ascendens (De Ley and Frateur 1974) Hördt *et al.* 2020, 42[1]	sp. nov. [elevation in rank from *Acetobacter pasteurianus* subsp. *ascendens* De Ley and Frateur 1974 (Approved Lists 1980)][a]	CCM 3612 (=LMG 1590=NCCB 51001)	32	[1]
Acidithiobacillus ferrianus Norris *et al.* 2020, 336	sp. nov.	MG (=DSM 107098=JCM 33084)	13	[2]
Actibacterium lipolyticum (Park *et al.* 2019) Hördt *et al.* 2020, 43[1]	comb. nov. [basonym: *Confluentimicrobium lipolyticum* Park *et al.* 2019]	SSK1-4 (=CECT 8621=KCTC 42136)	32	[1]
Actibacterium naphthalenivorans (Jeong *et al.* 2015) Hördt *et al.* 2020, 43[1]	comb. nov. [basonym: *Confluentimicrobium naphthalenivorans* Jeong *et al.* 2015]	NS6 (=DSM 105040=JCM 30828)	32	[1]
Actinomadura physcomitrii Zhuang *et al.* 2020, 684	sp. nov.	LD22 (=CCTCC AA 2018050=JCM 33455)	33	[3]
Afifellaceae Hördt *et al.* 2020, 38[1]	fam. nov.	*Afifella*	32	[1]
Afipia carboxidovorans (Meyer *et al.* 1994) Hördt *et al.* 2020, 43[1]	comb. nov. [basonym: *Oligotropha carboxidovorans* (ex Meyer and Schlegel 1978) Meyer *et al.* 1994]	OM5 (=ATCC 49405=DSM 1227)	32	[1]
Ahrensiaceae Hördt *et al.* 2020, 38[1]	fam. nov.	*Ahrensia*	32	[1]
Albibacillus Hördt *et al.* 2020, 41[1]	gen. nov.	*Albibacillus kandeliae*	32	[1]
Albibacillus kandeliae (Zhang *et al.* 2018) Hördt *et al.* 2020, 43[1]	comb. nov. [basonym: *Rugeria kandeliae* Zhang *et al.* 2018]	DSM 104293 (=MCCC 1K03284)	32	[1]
Allgaiera Hördt *et al.* 2020, 41[1]	gen. nov.	*Allgaiera indica*	32	[1]
Allgaiera indica (Jiang *et al.* 2014) Hördt *et al.* 2020, 43[1]	comb. nov. [basonym: *Defluviimonas indica* Jiang *et al.* 2014]	20V17 (=DSM 24802=JCM 17871)	32	[1]
Allorhizobium taibaishanense (Yao *et al.* 2012) Hördt *et al.* 2020, 43[1]	comb. nov. [basonym: *Rhizobium taibaishanense* Yao *et al.* 2012]	CCNWSX 0483 (=DSM 100021=HAMBI 3214)	32	[1]
Alloseboldiniviria Hördt *et al.* 2020, 41[1]	gen. nov.	*Alloseboldiniviria pacifica*	32	[1]
Alloseboldiniviria pacifica (Wu *et al.* 2013) Hördt *et al.* 2020, 43[1]	comb. nov. [basonym: *Roseivivax pacificus* Wu *et al.* 2013]	22DY03 (=DSM 29329=JCM 18866)	32	[1]
Allosphingosinicella Hördt *et al.* 2020, 41[1]	gen. nov.	*Allosphingosinicella vermicomposti*	32	[1]

Author affiliations: [1]The Institute of Life Sciences, The Hebrew University of Jerusalem, The Edmond J. Safra Campus, 9190401 Jerusalem, Israel; [2]Department of Microbiology & Molecular Genetics, Biomedical Physical Sciences, Michigan State University, East Lansing, MI 48824-4320, USA.
***Correspondence:** Aharon Oren, aharon.oren@mail.huji.ac.il; George Garrity, garrity@msu.edu

Validation Lists
Priority of names

- Priority in nomenclature is determined by the date of valid publication.
- Thus, since 1980, names effectively published outside IJSEM can only make a claim to priority if they are subsequently published in Validation Lists in issues of IJSB/IJSEM.
- For example, in revisions involving *Clavibacter xyli*, Suzuki *et al*., in 1999, and Evtushenko *et al*., in 2000, both proposed the generic name, *Leifsonia*.
- However, the proposal of Suzuki *et al*.,1999 was made directly in IJSEM and therefore took priority.

Young,2008

Notification Lists

Notification that new names and new combinations have appeared in the IJSB or IJSEM

- The names and citations to appear in the Notification Lists are those that are validly proposed in full articles in the IJSB (now IJSEM), an official publication of the ICSP.
- This listing of names published in a previous issue of the IJSEM is provided as a service to bacteriology to assist in the recognition of new names and new combinations.
- This procedure was proposed by the Judicial Commission [Minute II(ii), *Int J Syst Bacteriol* 41 (1991), p. 185].

Notification Lists

Notification that new names and new combinations have appeared in the IJSB or IJSEM

- The Notification Lists are published regularly in the:
- International Journal of Systematic Bacteriology (IJSB), or
- From January 2000, in the International Journal of Systematic and Evolutionary Microbiology (IJSEM).
- The first Notification List was published on July 1991.

Notification Lists

Notification that new names and new combinations have appeared in the IJSB or IJSEM

- They are provided as a service to bacteriology to assist in the recognition of new names and new descriptions.
- These lists have no formal status in prokaryote nomenclature except to allow for orthographic corrections (using proper spelling, capitalization and grammar. E.g. *Streptomyces scabies* to *S. scabiei*) to be made.

Young,2008;Euzeby,2010

Notification Lists

Oren and Garrity,2021.
Int J Syst Evol Microbiol 2021;71:004816

Microbiology Society,2021

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

LIST OF CHANGES IN TAXONOMIC OPINION
Oren and Garrity, *Int. J. Syst. Evol. Microbiol.* 2021;71:004816
DOI 10.1099/ijsem.0.004816

MICROBIOLOGY SOCIETY

Notification of changes in taxonomic opinion previously published outside the IJSEM list of changes in taxonomic opinion no. 34

Aharon Oren[1,*] and George M. Garrity[2,*]

The International Code of Nomenclature of Prokaryotes (Prokaryotic Code) deals with the nomenclature of prokaryotes. This may include existing names (the Approved Lists of Bacterial Names) as well as new names and new combinations. In this sense the Code is also dealing indirectly with taxonomic opinions. However, as with most codes of nomenclature, there are no mechanisms for formally recording taxonomic opinions that do not involve the creation of new names or new combinations. In particular, it would be desirable for taxonomic opinions resulting from the creation of synonyms or emended descriptions to be made widely available to the public. In 2004, the Editorial Board of the *International Journal of Systematic and Evolutionary Microbiology* (IJSEM) agreed unanimously that it was desirable to cover such changes in taxonomic opinions (i.e. the creation of synonyms or the emendation of circumscriptions) previously published outside the IJSEM, and to introduce a List of Changes in Taxonomic Opinion [Notification of changes in taxonomic opinion previously published outside the IJSEM; Euzéby *et al.* (2004). *Int J Syst Evol Microbiol* **54**, 1429–1430].

Scientists wishing to have changes in taxonomic opinion included in future lists should send a PDF file of the relevant publication to the IJSEM Editorial Office or to the Lists Editor.

It must be stressed that the date of proposed taxonomic changes is the date of the original publication, not the date of publication of the list. Taxonomic opinions included in the List of Changes in Taxonomic Opinion cannot be considered as validly published nor, in any other way, approved by the International Committee on Systematics of Prokaryotes and its Judicial Commission. The names that are to be used are those that are the 'correct names' (in the sense of Principle 6) in the opinion of the bacteriologist, with a given circumscription, position and rank. A particular name, circumscription, position and rank does not have to be adopted in all circumstances. Consequently, the List of Changes in Taxonomic Opinion must be considered as a service to bacteriology and it has no 'official character', other than providing a centralized point for registering/indexing such changes in a way that makes them easily accessible to the scientific community.

Name/authors:	Proposed as:	Reference
Algibacter lectus Nedashkovskaya *et al.* 2004 emend. Liu *et al.* 2021, 1037	emend.	[1]
Algibacter wandonensis Yoon and Park 2013 pro synon. *Algibacter lectus* Nedashkovskaya *et al.* 2004	synon.	[1]
Alkanindiges Bogan *et al.* 2003 emend. Yadav *et al.* 2021, 381	emend.	[2]
Bacillus ciccensis Liu *et al.* 2017 pro synon. *Cytobacillus solani* (Liu *et al.* 2015) Patel and Gupta 2020*	synon.	[3]
Catenulispora rubra Tamura *et al.* 2017 emend. Świecimska *et al.* 2021, 8†	emend.	[4]
Crateriforma Peeters *et al.* 2021 emend. Kumar *et al.* 2021, 351	emend.	[5]
Crocinitomicaceae Munoz *et al.* 2016 emend. Bowman 2020, 9†	emend.	[6]
Cryomorphaceae Bowman *et al.* 2003 emend. Bowman 2020, 9†	emend.	[6]
Cytobacillus solani (Liu *et al.* 2015) Patel and Gupta 2020 emend. Guo *et al.* 2021, 847	emend.	[3]
Gallionellaceae Henrici and Johnson 1935 (Approved Lists 1980) emend. Kojima *et al.* 2021, 322	emend.	[7]
Gimesia Scheuner *et al.* 2015 emend. Wiegand *et al.* 2020, 2011‡	emend.	[8]
Gimesia chilikensis Kumar *et al.* 2020 emend. Wiegand *et al.* 2020, 2011	emend.	[8]
Muribaculum intestinale Lagkouvardos *et al.* 2016 emend. Park *et al.* 2021, 284§	emend.	[9]

Author affiliations: [1]The Institute of Life Sciences, The Hebrew University of Jerusalem, The Edmond J. Safra Campus, 9190401 Jerusalem, Israel; [2]Department of Microbiology & Molecular Genetics, Biomedical Physical Sciences, Michigan State University, East Lansing, MI 48824-4320, USA

The Nomenclature Codes

Rule books

- The Nomenclature Codes (or the Codes of nomenclature) are the rulebooks that govern biological nomenclature.
- After the successful introduction of two-part names for species by Linnaeus it became ever more apparent that a detailed body of rules was necessary to govern scientific names.
- From the mid-nineteenth century onwards there were several initiatives to arrive at worldwide-accepted sets of rules.

The Bacteriological Code
History

- For many years, the International Code of Botanical Nomenclature (Botanical Code) was the basis for bacterial nomenclature.
- This is because the versions of the bacterial Code already published (Anonymous,1958;1966 and Buchanan *et al*.,1948) appear not to have been widely read or acted upon, and they came too late to act as a corrective measure.

Young,2008

The Bacteriological Code

History

1. The early Code for the nomenclature of Bacteria was approved at the 4th International Congress for Microbiology in 1947, but was later discarded.
2. International Code of Nomenclature of Bacteria (ICNB) which regulates the nomenclature of archaea and bacteria, first published in 1966.
3. The botanical Code was replaced with bacterial Code in 1975.

Bergey's Manual of Determinative Bacteriology (8th ed.) was published in 1974.

Young,2008; Hawksworth,2010

The Bacteriological Code
History

- Buchanan, R.E.; R. St. John-Brooks and R.S. Breed.1948. International bacteriological code of nomenclature, *J. Bacteriol.* 55: 287-306.
- Anonymous.1958. International Code of Nomenclature of Bacteria and Viruses, Iowa State College Press, Ames, IA, USA .
- Anonymous.1966. International Code of Nomenclature of Bacteria, *Int. J. Syst. Bact.* 16: 459-490.

The First International Microbiological Congress (Paris, 1930)

The Second International Congress for Microbiology (London, 1936)

The Third International Microbiological Congress (New York, 1939)

The Fourth International Microbiological Congress (Copenhagen, 1947)

The Fifth International Microbiological Congress (Rio de Janeiro, 1950)

The Sixth International Microbiological Congress (Rome, 1953)

The Seventh International Congress for Microbiology (Stockholm,1958)

The Eighth International Microbiological Congress (Montreal, 1962)

The Ninth International Congress for Microbiology (Moscow, 1966)

The Tenth International Congress for Microbiology (Mexico City, 1970)

First International Congress of Bacteriology (Jerusalem, 1973)

The Twelfth International Congress for Microbiology (Munich, 1978)

Thirteenth International Congress of Microbiology (Boston, Massachusetts, 1982)

Fourteenth International Congress of Microbiology (Manchester, 1986)

Fifteenth International Congress of Microbiology (Osaka, 1990)

Additionally, the history of the Code is available in recent publications (Sneath, 2003;Knapp *et al.*,2004).

The International Nomenclature Codes

- In the course of time these became the present Nomenclature Codes governing the naming of organisms such as:
- International Code of Zoological Nomenclature, ICZN.
- International Code of Nomenclature for Cultivated Plants, ICNCP.
- International Code of Botanical Nomenclature, ICBN.
- International Code of Nomenclature of Bacteria, ICNB (now ICNP).
- International Code of Phytosociological Nomenclature, ICPN.
- International Code of Virus Classification and Nomenclature, ICTV.

Hawksworth,2010

The nomenclature of plant pathogenic bacteria (now Prokaryotes)

- The bacterial Code was released in 1975.
- The nomenclature of bacteria including plant pathogenic bacteria is regulated by:

1. The Code: International Code of Nomenclature of Bacteria (ICNB), now Prokaryotes (ICNP), and
2. The Standards: The International Standards for Naming Pathovars of Phytopathogenic Bacteria.

Young,2008;..

The Bacteriological Code

1976 Revision of the Code (Lapage *et al.*,1975)

- The Bacteriological Code (BC) was a third set of rules (After plants and animals rules) first developed in 1953 and published in 1958.
- The 1976 revision of the International Code of Nomenclature of Bacteria had two purposes:

1. To legislate the new start for bacterial nomenclature at a specified date i.e. January 1, 1980;
2. To provide a comprehensive text that would clearly specify the criteria for legitimate publication of correct names, and
3. To introduce a novel concept of valid publication to ensure strict priority of all names in future.

The Bacteriological Code

1976 Revision of the Code (Lapage *et al.*,1975)

- The Bacteriological Code governs the scientific names for bacteria.
- The official "Nomenclatural Starting Date" for the current International Code of Nomenclature of Bacteria (ICNB) is January 1,1980.
- At present the name of Bacteriological Code is changed into International Code of Nomenclature of Prokaryotes (ICNP), or 'the Code'.

The Bacteriological Code

1976 Revision of the Code (Lapage *et al.*,1975)

- Some minor editorial changes have been required where the 1975 edition referred to actions in the future.
- Three important reforms were introduced by the revision of the Code published in 1975:

1. A new starting document and starting date were achieved with the publication of the Approved Lists of Bacterial Names on 1 January 1980 (Skerman *et al.*,1980), containing about 2,300 names….
2. All new names are validly published only in the *International Journal of Systematic Bacteriology* (IJSB), although they may be effectively published elsewhere and then validated by announcement in Validation Lists in the IJSB.
3. For valid publication, nomenclatural types must be designated.

Sneath,1992

The Bacteriological Code

1976 Revision of the Code (Lapage *et al.*,1975)

Approved Lists

- Central to the revision of nomenclature was the development of the:

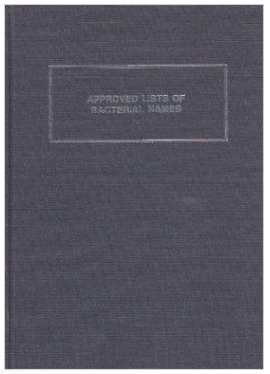
1. Approved Lists (Skerman *et al.*,1980),
2. Amended version (Skerman *et al.*,1989) which were to include only names that conformed fully to the revised Code.

Young,2008

The Code

The 1976 revision of the Code

Approved Lists of Bacterial Names



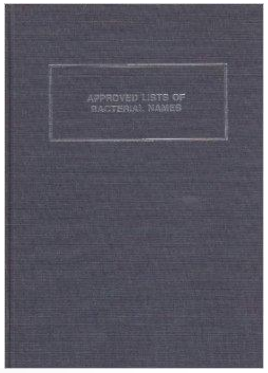
- The Approved Lists of Bacterial Names was published in the International Journal of Systematic Bacteriology (IJSB 30: 225-420,1980) and reprinted in book form to provide for the requirements of the Bacteriological Code (1976 Revision) in initiating a new starting date for bacterial nomenclature, 1 January 1980.

Ad Hoc Committee of the Judicial Commission of the ICSB. First Draft Approved Lists of Bacterial Names. Int. J. Syst. Bacteriol. 1976;26:563-599.

The Code

The 1976 revision of the Code

Approved Lists of Bacterial Names

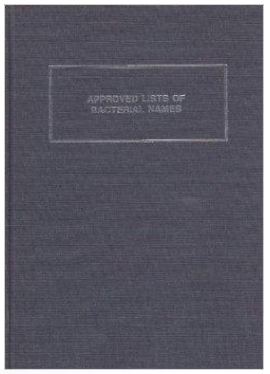


- According to the 1976 revision of the International Code of Nomenclature of Bacteria (Lapage *et al.*,1975) after 1 January 1980, priority of publication shall date from 1 January 1980 and on that date:

1. All names published prior to 1 January 1980, and
2. Included in the Approved Lists of Bacterial Names of ICSB (International Committee on Systematic Bacteriology) shall be treated.

Approved Lists of Bacterial Names

First edition (Skerman *et al.*,1980)



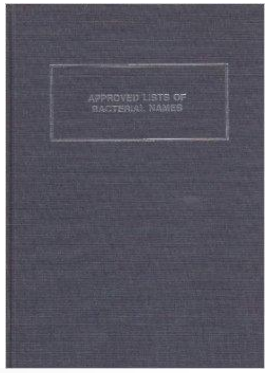
- In 1973, *ad hoc* Committee was appointed by the Judicial Commission of the International Committee on Systematic Bacteriology (ICSB):

1. To compile these names under the title of Approved Lists of Bacterial Names, and
2. To publish the lists in the International Journal of Systematic Bacteriology, to become effective on January 1,1980.

In 1966, the International Bulletin of Bacterial Nomenclature and Taxonomy (1951-1965) was continued with new official journal International Journal of Systematic Bacteriology and from January 2000, it was renamed as IJSEM.

Approved Lists of Bacterial Names

First edition (Skerman *et al.*,1980)



Approved Lists of Bacterial Names

edited by

V.B.D. SKERMAN,[1] VICKI McGOWAN,[1] AND P.H.A. SNEATH[2]

Department of Microbiology, University of Queensland, St. Lucia, Queensland 4067, Australia[1] and MRC Microbial Systematics Unit, University of Leicester, Leicester LE1 England 7RH,[2]

on behalf of

The Ad Hoc Committee of the Judicial Commission of the ICSB

INTRODUCTION

At the meeting of the Judicial Commission of the ICSB held in Jerusalem on the 29th March, 1973 an Ad Hoc Committee was appointed (Minute 22) to organize a review of the currently valid names of bacteria with the object of retaining only names for those taxa which were adequately described and, if cultivable, for which there was a Type, Neotype or Reference strain available; to compile these names under the title of **Approved Lists of Bacterial Names** and to publish the lists in the International Journal of Systematic Bacteriology, to become effective on January 1, 1980. This date would then replace May 1, 1753 (International Code of Nomenclature of Bacteria and Viruses. Rule 10) as the new date for determining priorities for names of new taxa.

The members of the Ad Hoc Committee originally appointed to oversee the task were S.P. Lapage, H.P.R. Seeliger and V.B.D. Skerman (Chairman). Following his election as President-Elect of IAMS, H.P.R. Seeliger resigned from the Committee and was replaced by J.G. Holt. P.H.A. Sneath, as Chairman of the Judicial Commission, was coopted to the Committee. This Committee was responsible for the editing of the first draft of the Approved Lists of Bacterial Names published in the IJSB (26, 1976, 563-599).

The Ad Hoc Committee was reconstituted by the Judicial Commission during its meetings at the International Congress of the IAMS in Munich in September, 1978, with V.B.D. Skerman (Chairman), P.H.A. Sneath (newly elected Chairman of the ICSB) and L.G. Wayne (newly elected Chairman of the Judicial Commission) as Committee members. This Committee was assigned wide powers to complete the work on the Lists and arrange publication.

The initial Committee agreed that, as a basis for inquiry, all names which had been included in the eighth edition of Bergey's Manual of Determinative Bacteriology, whether listed as recognised species, synonyms, *species incertae sedis*, or under other headings should be circulated; that all Subcommittees on Taxonomy of the ICSB should be asked for advice on the retention of taxa for which the Subcommittees were responsible and that specialists be approached for advice on other taxa - only a small list of taxa remained to be considered by the Ad Hoc Committee itself. Advice was also sought on additional names of taxa which had been validly published since the publication of the Manual. Special provision had been made in the Code of Nomenclature for the inclusion of names validly published after the 1st January, 1978 (Rule 24a).

A list of specialists was drawn up and approved by the Ad Hoc Committee. Each specialist was asked to associate two others in reaching a decision on name retention. The Chairman accepted the responsibility for coordination of work associated with the selection of generic names and specific and subspecific epithets and P.H.A. Sneath for the names of higher taxa.

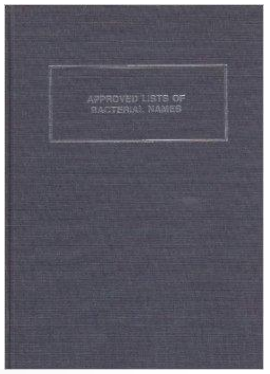
Names of Genera, Species and Subspecies:

The Lists of names of species and subspecies were circulated by airmail in April, 1976 and a draft list of names recommended for retention published in the IJSB (26, 1976, 563-599) together with the list of those people associated with that stage of the project. Reference may be made to the draft list for details of the procedures adopted in processing the information.

The object of publishing the draft list was to enable microbiologists, in general, to submit

Approved Lists of Bacterial Names

First edition (Skerman *et al.*,1980)



- The Approved Lists of Bacterial Names were edited by V.B.D. Skerman, V. McGowan, and P.H.A. Sneath on behalf of the *Ad Hoc* Committee of the Judicial Commission of the International Committee on Systematic Bacteriology (now, the International Committee on Systematic of Prokaryotes, ICSP).
- Names not on the Approved Lists had "no further standing in nomenclature" (Rule 24a).
- Despite every care, errors in the text have been detected and these errors have now been corrected in this Amended Edition.

Approved Lists of Bacterial Names

Two lists

About 2,300 were retained in the Approved Lists

- When bacteriologists agreed to make a new start in bacteriological nomenclature, they were faced with tens of thousands of names in the literature of the past.
- Of these, about 2,300 names were retained in the Approved Lists.
- The Approved Lists of Bacterial Names consist of two lists which were published on 1 January 1980 in the International Journal of Systematic Bacteriology (IJSB):

1. **Approved List 1**: A list for names of taxa above the rank of genus (family and above) which contains 124 names.
2. **Approved List 2**: A list for names of genera, species and subspecies (2212 names).

Euzéby,2014;..

Approved Lists of Bacterial Names
First edition (Skerman *et al.*,1980)

- This reduced the number of named species from ~ 28000 (Buchanan *et al.*,1966) to ~ 2000 (Skerman *et al.*,1980).
- Adoption of the Approved Lists resulted in numerous names of plant pathogenic species being abolished.
- These included such names as:
- *Pseudomonas fabae*, a pathogen on broad beans,
- *Pseudomonas tectonae* on teak,
- *Pseudomonas gardneri* on tomato (the only valid name), and
- *Pseudomonas adzukicola* on adzuki beans.

Bull *et al.*,2008

Approved Lists of Bacterial Names

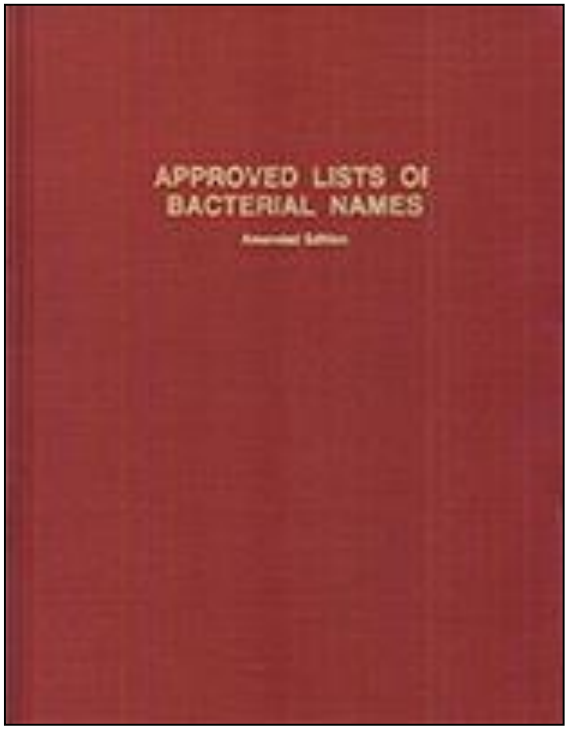
Amended edition (Skerman *et al.*,1989)

- Despite every care, errors in the text have been detected and many of these errors have been corrected in the Approved Lists of Bacterial Names (Amended Edition) published in 1989.
- This work contains all names that were acceptable at that time and since then there has been one major update (Skerman *et al.*,1989).
- The purpose of the Approved List is to ensure stability and consistency in bacterial taxonomy.
- Names not appearing in the List can still be used but normally are enclosed within inverted commas (i.e. ‘ ’) when appearing in written text.

Plant Pathologist’s Pocketbook,2002;Euzeby,2014

Approved Lists of Bacterial Names
Amended edition (Skerman *et al.*,1989)

- Approved Lists of Bacterial Names (Amended) and Index of the Bacterial and Yeast Nomenclatural Changes.
- **Author:** V.B.D. Skerman, V. McGowan, and P.H.A.. Sneath and W.E.C. Moore and L.V.H. Moore.
- **Number of Pages:** 298
- **Publication Date:** 1989
- **Publisher:** ASM Press



The Code

Revised editions of the Code

1990 Revision

- The Bacteriological Code (1990 Revision) was published in 1992, and it is freely available on the Internet: http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=icnb.
- The proposal to rename the International Code of Nomenclature of Bacteria (ICNB) to International Code of Nomenclature of Prokaryotes (ICNP) was unanimously accepted by International Committee on Systematic Bacteriology (ICSB) (now ICSP).
- According to Rule 1b, alterations to this Code have been made by the International Committee on Systematic Bacteriology (ICSB) from January 2000 as the International Committee on Systematic of Prokaryotes, ICSP).

The Code

Revised editions of the Code

1990 Revision

- The latest version to be printed in book form is the 1990 Revision (which may be referred to as the Bacteriological Code (1990 Revision):
- Lapage S.P., Sneath P.H.A., Lessel E.F., Skerman V.B.D., Seeliger H.P.R., Lark W.A. (eds).1992. International Code of Nomenclature of Bacteria (1990 Revision). American Microbiological Society, Washington DC., USA.
- Sneath, P.H.A. and Brenner, D.J.1992. Official Nomenclature Lists. ASM News, 58, 175.
- Young J.M., Bradbury J.F., Davis R.E., Dickey R.S., Ercolani G.L., Hayward A.C., Vidaver A.K. 1991. Nomenclatural revisions of plant pathogenic bacteria and list of names 1980-1988. Review of Plant Pathology 70: 211-221.

The Code

Revised editions of the Code

1990 Revision

- The 1990 revision supersedes all previous editions.
- It shall be cited as Bacteriological Code (1990 Revision) and will apply from the date of publication (1992) [Rule 1a].
- **The Bacteriological Code (1990 Revision) applies to all procaryotes.**
- The procaryotes include groups known by such names as:
- Bacteria, Eubacteria, Archaea, Archaeobacteria, Cyanobacteria, Cyanophyceae, Schizomycetes, Schizophycetes.

Tindall, B.J.1999. Proposals to update and make changes to the Bacteriological Code. *Int. J. Syst. Evol. Microbiol.*,1999 (49)1309-1312.

The Standards
Pathovar system of nomenclature

- In order to formalize the use of pathovar names, in *ad hoc* meeting in February 1978, the Committee on Taxonomy of Plant Pathogenic Bacteria of the ISPP (ISPP-CTPPB) was decided to publish a list of pathovar names known as the International Standards for Naming Pathovars of Phytopathogenic Bacteria (The Standards).

ISPP(International Society for Plant Pathology) website;Thind,2013

The Standards
Pathovar system of nomenclature

- At the introduction of the Approved List many old bacterial names were abandoned and strict rules for publication of new ones came into force.
- Many distinct plant pathogens were rejected from the Approved List, a physiologically too similar to warrant separation as species or subspecies.
- Plant pathologists therefore adopted a 'pathovar nomenclature' outside the jurisdiction of the Code.

Plant Pathologist's Pocketbook,2002

The Standards

Pathovar system of nomenclature

Pathovar concept

- Those names that did not meet these criteria but were important for plant pathologists it was hoped that the ICSB (now ICSP, International Committee on Systematic Prokayotes) would give some recognition to the latter by acceptance of pathovars as an infrasubspecific rank within the Code.

ISPP (International Society for Plant Pathology) website;Thind,2013

The Standards

Pathovar system of nomenclature

The preferred names of infrasubspecific subdivisions

- Domain: The highest of taxonomic rank (‘80s)
- Kingdom (Not used by most bacteriologists),1969
- Phylum or division of the kingdom
- Class
- Order
- Family (related genera)
- Genus (related species) plural: Genera
- Species (related strains) both singular & plural
- Subspecies
- Biovar (usual abbreviation: bv.),
- Chemoform/chemovar,
- Morphovar,
- Pathovar (usual abbreviation: pv.),
- Phagovar,
- Serovar.

Names not covered by Code

Names covered by Code

Names covered by Standards

The Standards

Pathovar system of nomenclature

Dye *et al.*,1980

- **Lists of pathovars and standards for publication of new pathovars have been published (Dye *et al.*,1980).**

Dye D.W., Bradbury J.F., Goto M., Hayward A.C., Lelliott R.A., Schroth M.N., 1980. International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotypes. *Review of Plant Pathology* 59:153-168.

The Standards

Pathovar system of nomenclature

Revisions for standards

- The revisions (Young *et al.* 1991) accepted at the ISPP Conference on Plant Pathogenic Bacteria, Versailles, 1992.
- **Revisions for standards were made by Young *et al.*,2001c).**
- Administration and interpretation of the Standards is by the International Society of Plant Pathology (hereafter called 'the ISPP Bacterial Taxonomy Committee.

Young J.M., Bull C.T., De Boer S.H., Firrao G., Gardan L., Saddler G.E., Stead D.E., Takikawa Y., 2001c. International standards for naming pathovars of phytopathogenic bacteria. http://www.isppweb.org/about_tppb_naming.asp. The list is updated annually and is available on www.bspp.org.uk/ispp/nppb.html

The Standards

Pathovar system of nomenclature

Naming of a pathovar is free and not determined by rules of the code

- The role of the Standards is similar to that of the Code except that the Standards deal specifically with the naming of pathovars of plant pathogens.

Bull *et al.*,2008

The Standards

Pathovar system of nomenclature

Dye *et al.*,1980

- The Standards are derived in general from the Principles which form the basis of the Code.
- They are not intended to replace or modify in anyway the existing Rules and Recommendations of the Code.
- These Standards will apply from 1 January 1980 (Dye *et al.*,1980).

Janse,2006

The Standards

Pathovar system of nomenclature

1. The principal requirements of the Code for legitimate names are essentially the same for pathovars.
2. Requirement of submission of the new pathotype strains at least to two international culture collections located in two different countries to ensure stability and success.
3. Priority for pathovar names published after January 1, 1980 is determined according to the date of effective publication of legitimate pathovar names (list of pathovars names published with the standards).
4. CTPPB of ISPP (Committee on Taxonomy of Plant Pathogenic Bacteria of the ISPP) maintains a list of pathovar names and pathotype strains.

Thind,2013

The Standards

Pathovar system of nomenclature

Pathovar type strain

- Naming of a pathovar is free and not determined by rules of the code.
- This also holds true for subdivision of a bacterial species or subspecies into:

1. Pathological races and biochemical, serological or phage varieties (resp. biovars, serovars or phagovars), and
2. Genetic groups.

- For an example of subdivision of *Ralstonia solanacearum* into races, biovars and genetic RFLP groups.
- When new pathovars are described generally the above-mentioned international standards (including the indication of a pathovar type strain) are followed.

Supplementary Lists of Bacterial Names
Lists of plant pathogenic bacteria
Comprehensive lists

- The base-line for bacterial names are the Approved Lists (Skerman *et al.*,1980), which may not be added to.
- The International Code of Nomenclature of Bacteria (1990 Revision) is the cornerstone of prokaryotic nomenclature.
- New bacterial names has standing in nomenclature, if one of the following criteria is met:

1. The name is cited in the Approved Lists of Bacterial Names;
2. The name is published in papers in the IJSB or in the IJSEM and conforms to requirements laid down in the bacteriological Code.
3. The name is validly published by announcement in a Validation List.
4. Amendments can also be found directly online at the website of Euzéby/different comprehensive lists.

Supplementary Lists of Bacterial Names
Lists of plant pathogenic bacteria
Comprehensive lists

- Subsequently, a supplementary list of names of pathogens was produced (Young *et al.*,1991).
- A complete list of all plant pathogenic bacteria was published (Young *et al.*,1996) and after 2000, the 1996 list and annual updates (Young *et al.*,2004 and 2005) were maintained on the ISPP (International Society for Plant Pathology) Bacterial Taxonomy Committee website, in anticipation of further hard-copy publications.
- The last list was updated by Bull *et al.*,2014.

Supplementary Lists of Bacterial Names
Lists of plant pathogenic bacteria
Comprehensive lists

- Young, J.M., J.F. Bradbury, R.E. Davis, R.S. Dickey, G.L. Ercolani, A.C. Hayward and A.K. Vidaver.1991. Nomenclatural revisions of plant pathogenic bacteria and list of names 1980-1988. Rev. Plant Pathol. 70:211-221.
- Young, J.M., G. Saddler, Y. Takikawa, S.H. De Boer, L. Vauterin, L. Gardan, R.I. Gvozdyak and D.E. Stead.1996. Names of plant pathogenic bacteria 1864-1995. Rev. Plant Pathol. 75:721-763.
- Bull, C.T., T.A. Coutinho, T.P. Denny, G. Firrao, M. Fischer-Le Saux, X. Li, G.S. Saddler, M. Scortichini, D.E. Stead and Y. Takikawa.2014. List of New Names of Plant Pathogenic Bacteria (2011-2012). Journal of Plant Pathology, 96 (2):223-226.

Euzéby website

LPSN: Creator by J.P. Euzéby

LPSN lists all validly published names of prokaryotes, from the Approved Lists and culled from IJSB/IJSEM and its Validation Lists

- The List of Prokaryotic Names with Standing in Nomenclature (LPSN; http://www.bacterio.net) is a database that lists the names of prokaryotes (Bacteria and Archaea) that have been validly published:

1. In the International Journal of Systematic and Evolutionary Microbiology (IJSEM) directly, or
2. By inclusion in a Validation List, under the Rules of International Code of Nomenclature of Bacteria.

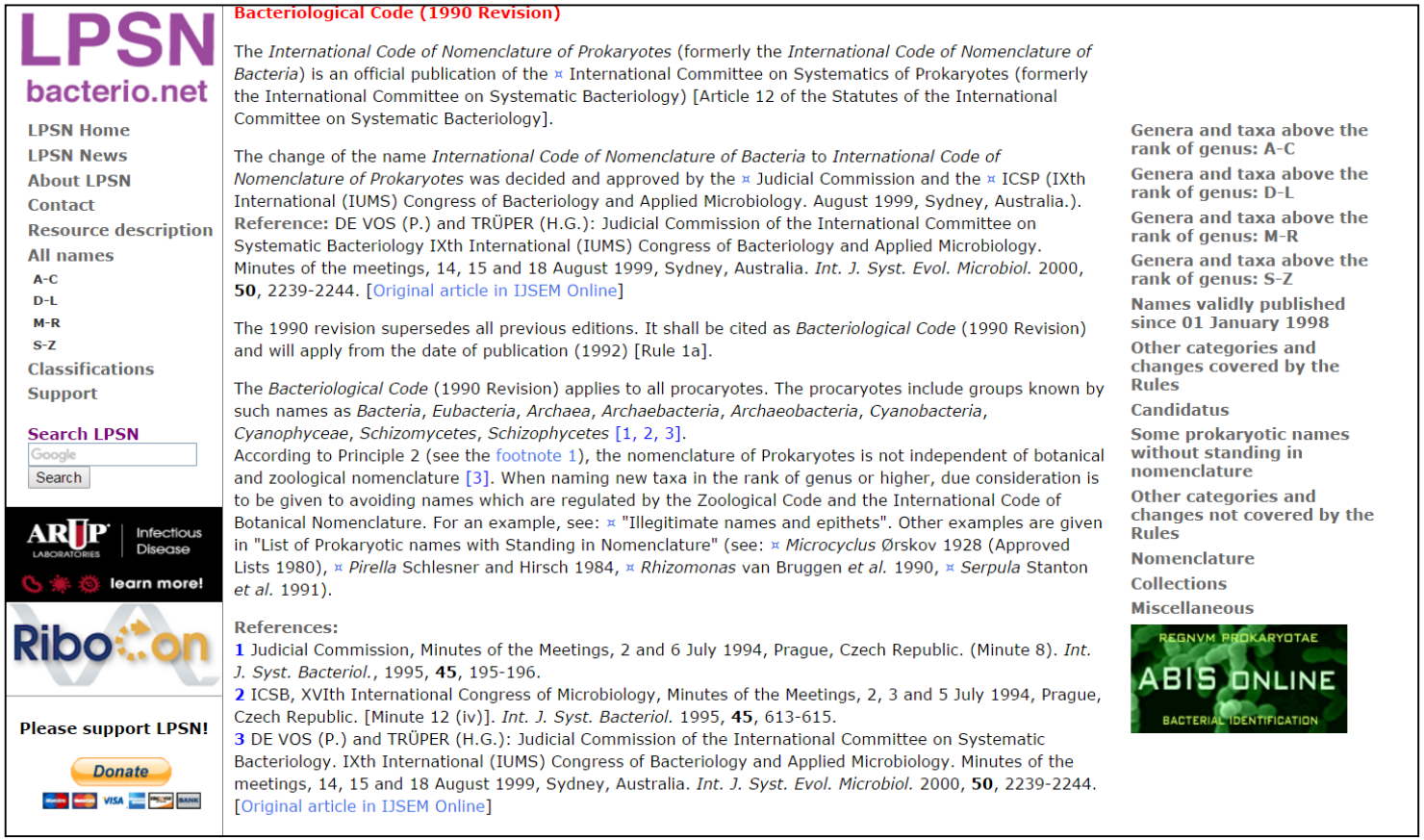
- Currently there are 15974 taxa listed.
- In addition, LPSN has an up-to-date classification of prokaryotes and information on prokaryotic nomenclature and culture collections.

Jean, Paul, Marie, EUZÉBY
Retired professor of microbiology, immunology (École Nationale Vétérinaire de Toulouse).

Euzéby website

LPSN: Creator by J.P. Euzéby

LPSN lists all validly published names of prokaryotes, from the Approved Lists and culled from IJSB/IJSEM and its Validation Lists



LPSN
bacterio.net

LPSN Home
LPSN News
About LPSN
Contact
Resource description
All names
A-C
D-L
M-R
S-Z
Classifications
Support

Search LPSN
Google
Search

ARUP LABORATORIES Infectious Disease learn more!

Ribo on

Please support LPSN!
Donate

Bacteriological Code (1990 Revision)

The *International Code of Nomenclature of Prokaryotes* (formerly the *International Code of Nomenclature of Bacteria*) is an official publication of the ¤ International Committee on Systematics of Prokaryotes (formerly the International Committee on Systematic Bacteriology) [Article 12 of the Statutes of the International Committee on Systematic Bacteriology].

The change of the name *International Code of Nomenclature of Bacteria* to *International Code of Nomenclature of Prokaryotes* was decided and approved by the ¤ Judicial Commission and the ¤ ICSP (IXth International (IUMS) Congress of Bacteriology and Applied Microbiology. August 1999, Sydney, Australia.). **Reference:** DE VOS (P.) and TRÜPER (H.G.): Judicial Commission of the International Committee on Systematic Bacteriology IXth International (IUMS) Congress of Bacteriology and Applied Microbiology. Minutes of the meetings, 14, 15 and 18 August 1999, Sydney, Australia. *Int. J. Syst. Evol. Microbiol.* 2000, **50**, 2239-2244. [Original article in IJSEM Online]

The 1990 revision supersedes all previous editions. It shall be cited as *Bacteriological Code* (1990 Revision) and will apply from the date of publication (1992) [Rule 1a].

The *Bacteriological Code* (1990 Revision) applies to all procaryotes. The procaryotes include groups known by such names as *Bacteria*, *Eubacteria*, *Archaea*, *Archaebacteria*, *Archaeobacteria*, *Cyanobacteria*, *Cyanophyceae*, *Schizomycetes*, *Schizophycetes* [1, 2, 3].
According to Principle 2 (see the footnote 1), the nomenclature of Prokaryotes is not independent of botanical and zoological nomenclature [3]. When naming new taxa in the rank of genus or higher, due consideration is to be given to avoiding names which are regulated by the Zoological Code and the International Code of Botanical Nomenclature. For an example, see: ¤ "Illegitimate names and epithets". Other examples are given in "List of Prokaryotic names with Standing in Nomenclature" (see: ¤ *Microcyclus* Ørskov 1928 (Approved Lists 1980), ¤ *Pirella* Schlesner and Hirsch 1984, ¤ *Rhizomonas* van Bruggen *et al.* 1990, ¤ *Serpula* Stanton *et al.* 1991).

References:
1 Judicial Commission, Minutes of the Meetings, 2 and 6 July 1994, Prague, Czech Republic. (Minute 8). *Int. J. Syst. Bacteriol.*, 1995, **45**, 195-196.
2 ICSB, XVIth International Congress of Microbiology, Minutes of the Meetings, 2, 3 and 5 July 1994, Prague, Czech Republic. [Minute 12 (iv)]. *Int. J. Syst. Bacteriol.* 1995, **45**, 613-615.
3 DE VOS (P.) and TRÜPER (H.G.): Judicial Commission of the International Committee on Systematic Bacteriology. IXth International (IUMS) Congress of Bacteriology and Applied Microbiology. Minutes of the meetings, 14, 15 and 18 August 1999, Sydney, Australia. *Int. J. Syst. Evol. Microbiol.* 2000, **50**, 2239-2244. [Original article in IJSEM Online]

Genera and taxa above the rank of genus: A-C
Genera and taxa above the rank of genus: D-L
Genera and taxa above the rank of genus: M-R
Genera and taxa above the rank of genus: S-Z
Names validly published since 01 January 1998
Other categories and changes covered by the Rules
Candidatus
Some prokaryotic names without standing in nomenclature
Other categories and changes not covered by the Rules
Nomenclature
Collections
Miscellaneous

REGNVM PROKARYOTAE ABIS ONLINE BACTERIAL IDENTIFICATION

Jean, Paul, Marie, EUZÉBY

Retired professor of microbiology, immunology (École Nationale Vétérinaire de Toulouse).

ISPP
International Society for Plant Pathology

- **Affiliations**
- The ISPP (founded in 1968) is a member of:

1. The International Union of Biological Sciences (IUBS);
2. The International Union of Microbiological Societies (IUMS);
3. ISPP has representation at ICSB(International Committee on Systematic Bacteriology (now ICSP).
4. In liaison (co-operation) with the UN Food and Agriculture Organization (FAO).

IUMS ⟶ ICSP ⟶ ISPP⟶ ISPP-CTPPB Committees

Administration and interpretation of the Standards is by the ISPP Bacterial Taxonomy Committee.

ISPP
International Society for Plant Pathology

- ISPP has representation at ICSB (International Committee on Systematic Bacteriology (now ICSP).
- The Executive Committee of the International Society for Plant Pathology (ISPP) set up a Committee on Taxonomy of Phytopathogenic Bacteria (ISPP Committee).

ICSP
(International Committee on Systematic Prokayotes)

⇩

ISPP
(representative of ICSP)

⇩

Executive Committee of ISPP set up:

⇩

ISPP-CTPPB
(Committee on Taxonomy of Phytopathogenic Prokayotes)

ISPP
International Society for Plant Pathology

- The general objective of the ISPP:

1. To promote the worldwide development of plant pathology;
2. The dissemination of knowledge about plant diseases and plant health management;
3. Sponsoring an the International Congress of Plant Pathology (ICPP), at intervals of 5 years, and sponsoring other international meetings on plant pathology and related subjects.
4. The Society establishes committees to consider and report on special fields or problems in plant pathology.
5. The Society organizes other activities including the publication of journals and newsletters, websites, as approved by the Executive Committee.

ISPP
ISPP website

- The ISPP Committee on the Taxonomy of Bacterial Plant Pathogens has created a website at: <http://www.isppweb.org/about_tppb.asp>.
- The major items are posted:

1. The names and addresses of the ISPP Committee on the Taxonomy of Plant Pathogenic Bacteria.
2. An up to date list of names of plant pathogenic bacteria;
3. The International Standards for Naming Pathovars;
4. A discussion news group.

ISPP
ISPP website

International Society for Plant Pathology

Promoting World-Wide Plant Health and Food Security

Home | Congresses | Commission on Global Food Security | Newsletter Archive | Subject Matter Committees | Global Plant Health Assessment | Resources | Administration

- TPPB Home
- Membership of the Committee
- International Standards
- Bacterial Names
- Submit a New Name
- Committee Business

Taxonomy of Plant Pathogenic Bacteria News:

Please contact the convener marion.le-saux@inrae.fr with questions about any of the published lists (below).

Publications:

Bull, C.T., Coutinho, T.A., Denny, T.P., Firrao, G., Fischer-Le Saux, M., Li, X., Saddler, G.S., Scortichini, M., Stead, D.E., Takikawa, Y., 2014. List Of New Names Of Plant Pathogenic Bacteria (2011-2012). Journal of Plant Pathology 96:223-226.
<http://sipav.org/main/jpp/index.php/jpp/article/view/3127>

Bull, C. T., S.H. De Boer, T.P. Denny, G. Firrao, M. Fischer-Le Saux, G.S. Saddler, M. Scortichini, D.E. Stead, Y. Takikawa, 2012. List of New Names of Plant Pathogenic Bacteria (2008-2010). Journal of Plant Pathology 94 (1): 21-27.
<http://sipav.org/main/jpp/index.php/jpp/article/view/2437>

Bull, C. T., S.H. De Boer, T.P. Denny, G. Firrao, M. Fischer-Le Saux, G.S. Saddler, M. Scortichini, D.E. Stead, Y. Takikawa, 2010. Comprehensive List of Names of Plant Pathogenic Bacteria, 1980-2007. Journal of Plant Pathology 92:551-592.
Download PDF <http://www.isppweb.org/SMC_Files/Bull%20et%20al.%202010%20JPP%20List.pdf>

Bull, C. T., S.H. De Boer, T.P. Denny, G. Firrao, M. Fischer-Le Saux, G.S. Saddler, M. Scortichini, D.E. Stead, Y. Takikawa, 2008. Demystifying Nomenclature of Bacterial Plant Pathogens. Journal of Plant Pathology 90:403-417.
Download PDF <http://www.isppweb.org/SMC_Files/Bull%20et%20al%202008%20InvitedReview_403.pdf>

Taxonomy of Plant Pathogenic Bacteria - Committee members:

M. Fischer-Le Saux (Convener), UMR de Pathologie Vegetale, INRA, BP 60057, 49071 Beaucouze Cedex, France e-mail:

T.A. Coutinho (Vice convener), Department of Microbiology and Plant Pathology Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa e-mail:

C.T. Bull , Professor & Department, Head Department of Plant Pathology and Environmental Microbiology, Pennsylvania State University, 212 Buckhout Lab, University Park, PA 16802 USA e-mail:

G. Firrao, Dipartimento di Biologia Applicata alla Difesa delle Piante, Universita, via Scienze 208, 33100 Udine, Italy e-mail:

X. Li, Canadian Food Inspection Agency, Charlottetown Laboratory, 93 Mount Edward Road, Charlottetown, PE C1A 5T1, Canada email:

M. Scortichini, C.R.A. - Istituto Sperimentale per la Frutticoltura, Via di Fioranello, 52, 00134 Roma, Italy e-mail:

Y. Takikawa, Faculty of Agriculture, Shizuoka University, 836 Ohya, Shizuoka 422-8529, Japan e-mail:

ISPPWeb,2023

ISPP

ISPP website

Promoting World-Wide Plant Health and Food Security

International Society for Plant Pathology

Promoting World-Wide Plant Health and Food Security

Home | Congresses | Commission on Global Food Security | Newsletter Archive | Subject Matter Committees | Resources | Administration

- About SMC's
- Crop Losses
- Chemical Control
- Environmental Effects on Plants
- Epidemiology
- Forest Pathology
- Fusarium
- Grapevine Trunk Diseases
- Oomycetes
- Plant Pathogenic Bacteria
- Plant Virus Epidemiology
- Plant Quarantine and Biosecurity
- Postharvest Pathology
- Rhizoctonia
- Sclerotinia
- Seed Pathology
- Taxonomy of Plant Pathogenic Bacteria
- Teaching

ISPP Subject Matter Committees

Current Status of the ISPP Subject Matter Committees

The ISPP Subject Matter Committees (SMCs) are specialist committees established (or endorsed) by the International Society for Plant Pathology to address and report on special fields or problems in plant pathology (ISPP Statute 5 (b)).

An ISPP Subject Matter Committee generally aims at the following objectives:

1. Consider or establish in relation to their function within a particular field of plant pathology key norms, benchmarks and recommendations for their peers as well as national and international bodies,
2. Provide advice to ISPP, associated societies, or other national and international bodies,
3. Liaise as necessary with other ISPP Committees and national and international interest groups to optimize collaborations,
4. Assist and advise the ISPP Executive and the organizers of the International Congress of Plant Pathology (ICPP) in the development of (ICPP) technical sessions, workshops and field visits relevant to their specialist field,
5. Organize specialist workshops, conferences and publications relevant to the field of plant pathology addressed by the Subject Matter Committee,
6. Develop and maintain web related materials relevant to the subject, and
7. Provide news items of relevance for the ISPP Newsletter.

Additionally, an ISPP Subject Matter Committee must:

8. Provide a report to ISPP Council at the International Congress of Plant Pathology.

Frequently Asked Questions (FAQ)

How do I enquire about a Subject Matter Committee?

Subject Matter Committees are small groups of individuals variously chosen by a Committee Chair, the ISPP Executive and/or the Committee with a view to geographic diversity and breadth of expertise. Generally a Committee might consist of six to 24 members.

Some Committees also have a larger mailing list of interested individuals and any enquiries about a Committee should be addressed to the Committee Chair, the ISPP Business Manager or the Vice President responsible (Dr Serge Savary in 2013-2018).

For established Committees, the ISPP Business Manager can also setup a mailing list form on the web page for the Committee to allow other

Current activities of ISPP

International Newsletter on Plant Pathology

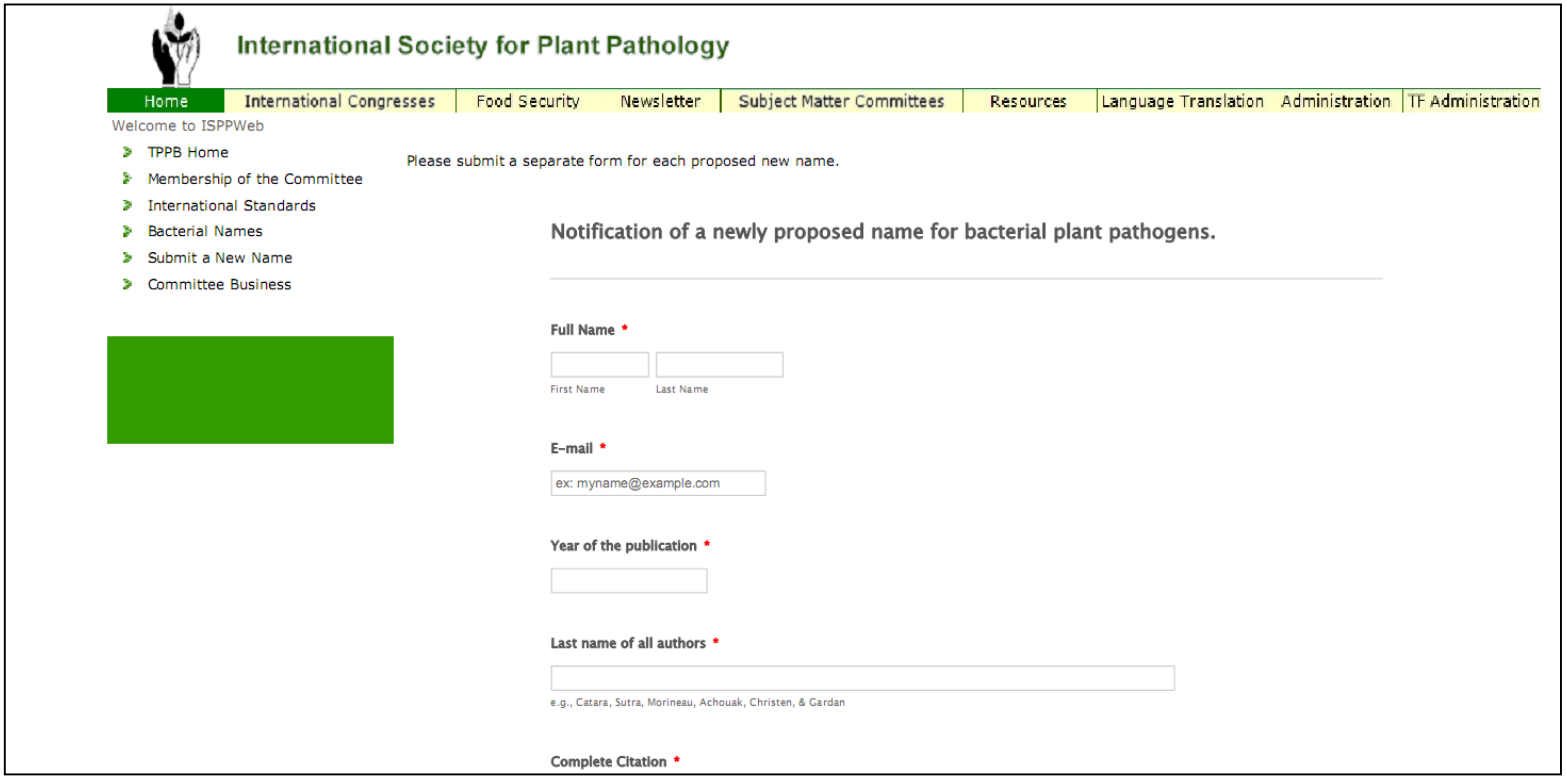
- International Newsletter on Plant Pathology is an official publication of the Society:

1. To inform members of the Society, and others, of the activities of the Society and of related matters.
2. An archive of the Newsletter covering issues published from February 1998 to date are available on this website.
3. Up dating new names list;
4. Developing strategy for on-line submission of new names;
5. Discussing merit and strategies for publishing comprehensive and annual names lists;
6. Writing paper on bacterial nomenclature.
7. Contributions to the International Congress of Plant Pathology (ICPP).

ISPPWeb,2008

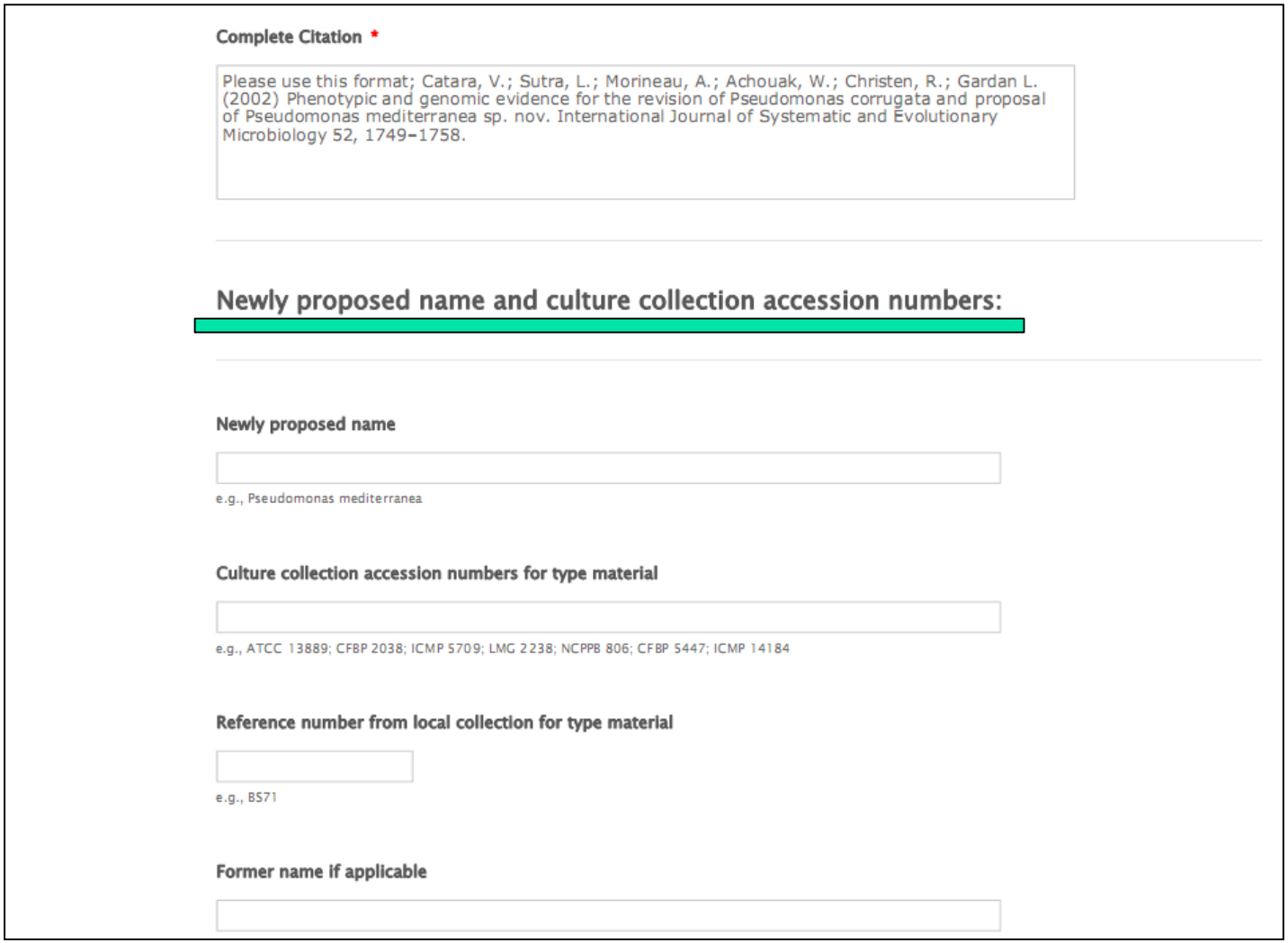
ISPP & Notification of a newly proposed names for bacterial plant pathogens

Submit a New Names



Note: This option is currently disabled on the site.2021

ISPP & Notification of a newly proposed names for bacterial plant pathogens (Continued)



Complete Citation *

Please use this format; Catara, V.; Sutra, L.; Morineau, A.; Achouak, W.; Christen, R.; Gardan L. (2002) Phenotypic and genomic evidence for the revision of Pseudomonas corrugata and proposal of Pseudomonas mediterranea sp. nov. International Journal of Systematic and Evolutionary Microbiology 52, 1749–1758.

Newly proposed name and culture collection accession numbers:

Newly proposed name

e.g., Pseudomonas mediterranea

Culture collection accession numbers for type material

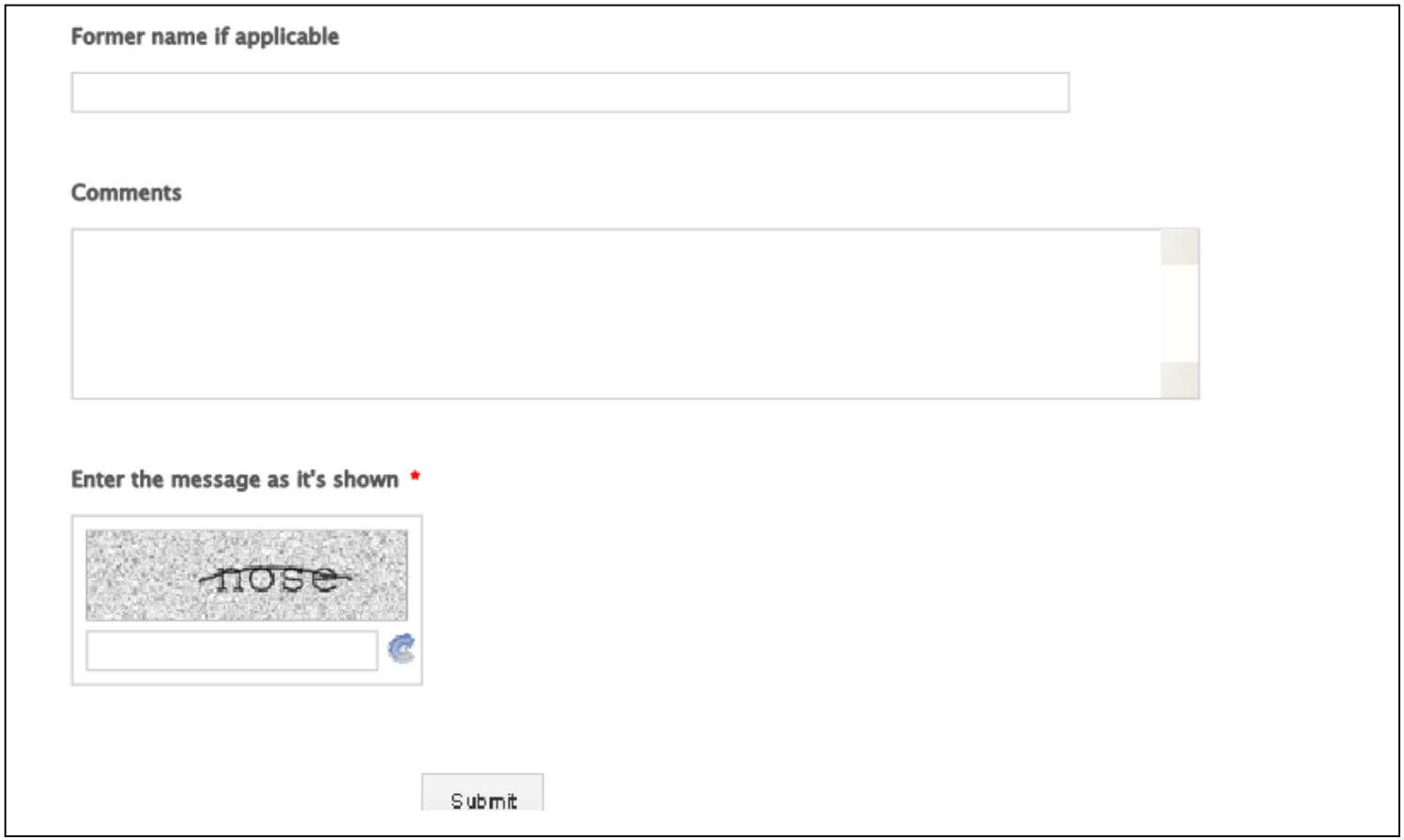
e.g., ATCC 13889; CFBP 2038; ICMP 5709; LMG 2238; NCPPB 806; CFBP 5447; ICMP 14184

Reference number from local collection for type material

e.g., BS71

Former name if applicable

ISPP & Notification of a newly proposed names for bacterial plant pathogens (Continued)



The other codes

A uniform nomenclatural system for all organisms

The other codes

BioCode

Biological nomenclature

- A term proposed in 1990 for any system of letters or numbers used with or in place of the name of a taxon.
- The IUBS/IUMS International Committee on Bionomenclature (ICB) presented the first Draft BioCode in 1997, followed by the Draft BioCode 2011.
- A more radical approach is to replace all existing Codes by a BioCode, basically a synthesis of the existing Codes.
- The draft of BioCode seeks simple rules to provide.
- It has yet to be formally accepted.

BSPP News 30 Spring 1997-Online Edition

The other codes

BioCode

A uniform nomenclatural system for all organisms

- The BioCode attempts to:

1. produce a unified code for the nomenclature of all life.
2. It includes all the major bionomenclatural codes such as bacteriological, botanical and zoological, with special provision for viruses and cultivated plants);
3. Governs the formation and choice of scientific names;
4. But not the definition of the taxa themselves.

The other codes

BioCode

BioCode draft (2000) and revised BioCode draft (2011)

- The originally planned implementation date for the BioCode draft was January 1, 2000, but agreement to replace the existing Codes was not reached.
- In 2011 a revised BioCode was proposed that, instead of replacing the existing Codes, would provide a unified context for them, referring to them when necessary.
- Changes in the existing codes are slowly being made in the proposed directions.

Wikipedia,2017

The BioCode

Not cover taxa below the rank of subspecies

- The BioCode:

1. will not cover taxa below the rank of subspecies and therefore will not directly affect the working of the ISPP Sub-committee for Taxonomy of Plant Pathogenic Bacteria (ISPP-CTPPB), and
2. their published guidelines for the naming of infraspecific taxa such as pathovars.

BioCode
References

1. Biological Nomenclature for the 21st Century- Proceedings of a symposium on the draft BioCode.
2. Greuter, W., Hawksworth, D.L., McNeill, J., Mayo, M.A., Minelli, A., Sneath, P.H.A., Tindall, B.J., Trehane, P., & Tubbs, P. (1996). Draft Biocode: the prospective international rules for the scientific names of organisms. Taxon 45: 349-372.
3. Greuter W; Hawksworth DL; McNeill J; Mayo MA; Minelli A; Sneath PHA; Tindall BJ; Trehane P; Tubbs P (eds) (1998). Draft BioCode (1997): the prospective international rules for the scientific names of organisms. Taxon 47: 127-150.

Other codes

PhyloCode

PhyloCode first draft (2000) and revised PhyloCode draft (2006)

- Another Code in development since 1998 is the phylocode.
- The PhyloCode appears to get more publicity than support.
- The first draft of PhylCode was published in April 2000.
- The current version of the PhyloCode was posted in June 2006 and includes many substantive modifications of the previous version.

Cantino, PD and de Queiroz K.2009. PhyloCode: A Phylogenetic Code of Biological Nomenclature. Version 4c. http://www.ohiou.edu/phylocode/index/html. International Society for Phylogenetic Nomenclature.

The other codes

PhyloCode

Rules governing phylogenetic nomenclature

- Some authors encountered problems in using the Linnean system in phylogenetic classification.
- The phylocode was proposed to regulate what their creators called phylogenetic nomenclature instead of the traditional Linnaean nomenclature (that is, it requires phylogenetic definitions as a "type" attached to every name, and does not contain mandatory ranks).

The other codes

PhyloCode

Rules governing phylogenetic nomenclature

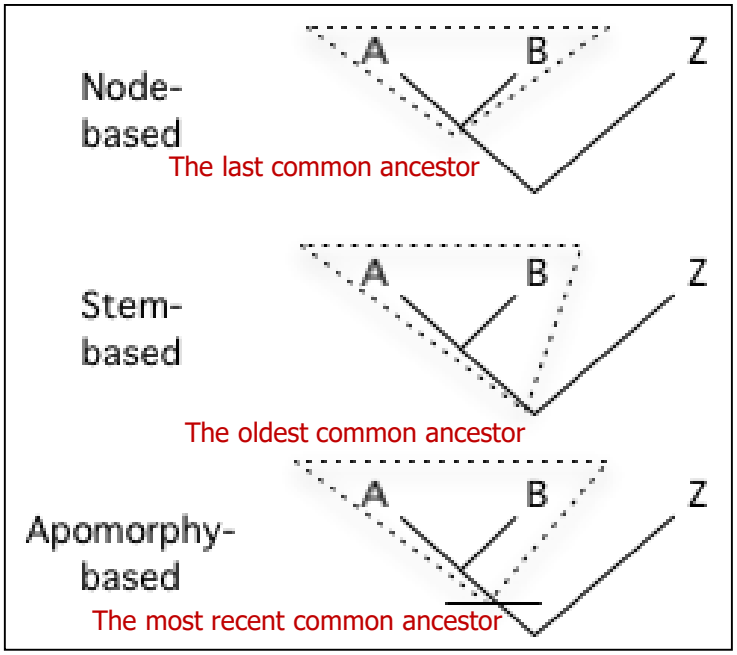
- A phylogenetic Code of biological nomenclature.
- An integrated set of principles and rules governing the naming of taxa and the application of taxon names that is based on the principle of common descent.
- There is also debate concerning development of a phylocode to name clades of phylogenetic trees, rather than taxa.

Cantino, PD and de Queiroz K.2009. PhyloCode: A Phylogenetic Code of Biological Nomenclature. Version 4c. http://www.ohiou.edu/phylocode/index/html. International Society for Phylogenetic Nomenclature.

The PhyloCode

- Three ways to define a clade for use in a cladistic taxonomy.

1. **Node-based**: The last common ancestor of A and B and all its descendants.
2. **Stem-based**: All descendants of the oldest common ancestor of A and B that is not also an ancestor of Z.
3. **Apomorphy-based**: The most recent common ancestor of A and B possessing a certain apomorphy (derived character), and all its descendants.



Types of Clade

Plesiomorphic = "primitive" character (found in common ancestor).
Apomorphy = "advanced" character (a character in a new state, not the primitive one).

The Code
Statutes

Organization

The Code

Statutes

Organization

- The International Committee on Systematic Bacteriology, ICSB (now The International Committee on Systematic of Prokaryotes (ICSP) was established by the International Association of Microbiological Societies (now IUMS).
- IUMS is the organization responsible:

1. The International Congresses on Microbiology (IUMS Congresses), and
2. The International Committee on Systematic of Prokaryotes (ICSP).

- Organization relationships:
- ICSB/ICSP of the International Union of Microbiological Societies (IUMS).
- *Ad hoc* Committee of the Judicial Commission of ICSB/ICSP.

IUMS> ICSP> Judicial Commission> *ad hoc* Committees

The International Union of Microbiological Societies

History and Statutes

- The International Union of Microbiological Societies was founded in 1927 (over 80 years) as the International Society for Microbiology.
- In 1980, the International Association of Microbiological Societies (IAMS) changed its name to the International Union of Microbiological Societies (IUMS) and acquired independence as a Union.
- It was accepted as a member of the International Council of Scientific Unions (ICSU) in 1982.
- IUMS is one of the 29 Scientific Unions of International Council of Scientific Unions (ICSU), a non-governmental organization.
- IUMS acts as an umbrella organization for these scientific research activities.

The International Union of Microbiological Societies (IUMS)

The current IUMS statues

- The IUMS currently comprises:
- 117 Microbiological societies and national committees in 67 countries;
- 1 International organization as full members;
- 7 Microbiological societies, and
- 11 Multinational organizations as associate members.
- Because of its broad contact with thousands of microbiologists worldwide, the IUMS can be regarded as the global voice of microbiology.

Schleifer,2008

The International Union of Microbiological Societies (IUMS)

IUMS scientific activities

- Most of the work under the umbrella of IUMS is conducted by the three divisions within IUMS and, in particular, by their committees, commissions and federations:

A. Three Divisions of:

1. Bacteriology & Applied Microbiology (BAM);
2. Mycology, and
3. Virology,

- And by:

1. six specialist international committees,
2. eight international commissions, and
3. two international federations.

IUMS

Organization: IUMS committees, commissions and federations (COMCOFs)

- The major scientific activities of IUMS take place within the international COMCOFs.
- These *ad hoc* committees are established with one of the three divisions (Bacteriology & Applied Microbiology (BAM), Mycology and Virology).
- Some of the currently exist committees, commissions and federations are:

1. International Committee on the Systematic of Prokaryotes (ICSP);
2. International Committee on Food Microbiology and Hygiene (ICFMH);
3. International Commission of Microbiological Specification for Food (ICMSF);
4. World Federation of Culture Collections (WFCC);
5. International Commission on Bionomenclature and Bioinformatics (ICBB);
6. International Commission on Antigens and Molecular Diagnostics (ICAMD).

The International Union of Microbiological Societies (IUMS)

IUMS scientific activities

- Their major activities include:

1. The classification and nomenclature of bacteria, fungi and viruses,
2. Food microbiology,
3. Medical microbiology and diagnostics,
4. Culture collections,
5. Education, and
6. Biological standardization.

IUMS
IUMS Divisions

- The Union is composed of three Divisions, each concerned with a major microbiological science.

1. Division of Bacteriology and Applied Microbiology (BAM): The aims of the Division of Bacteriology and Applied Microbiology are to further bacteriology and related sciences throughout the world in research, education and services.
2. Division of Mycology: To further the discipline of mycology,…
3. Division Virology: To further the discipline of virology,..

- The Union provides financial assistance to Divisions for operating expenses and scientific purposes.

IUMS
Objectives

- The objectives of the International Union of Microbiological Societies are to:

1. Promote the study of microbiological sciences internationally;
2. Initiate, facilitate and coordinate research and other scientific activities which involve international cooperation;
3. Promote the publication of international study and research;
4. Promote the organization of international conferences, symposia and meetings and assist in the publication of their reports;
5. Represent microbiological sciences in the International Council of Scientific Unions and maintain contact with other international organization.

IUMS
Affiliation to the Union

- Affiliation to the Union is in three categories:

1. Member Societies (National and International);
2. Associate Societies (National and International);
3. Supporting Organizations.

- At present, IUMS also has 96 national member societies and 26 associate member societies adhering to the Union.
- These societies are spread throughout the world and IUMS represents an important link for instance by organizing International Conferences at regular intervals.

Statutes and activities of ICSP

The subcommittees deal with the taxonomy of different groups of bacteria and archaea

- The International Committee on Systematic Prokaryotes (formerly the ICSB) was established by the International Association of Microbiological Societies (IAMS presently as IUMS).
- The Committee was founded in 1930 and is responsible for all matters related to nomenclature and taxonomy of prokaryotes.
- The ICSP consists of:

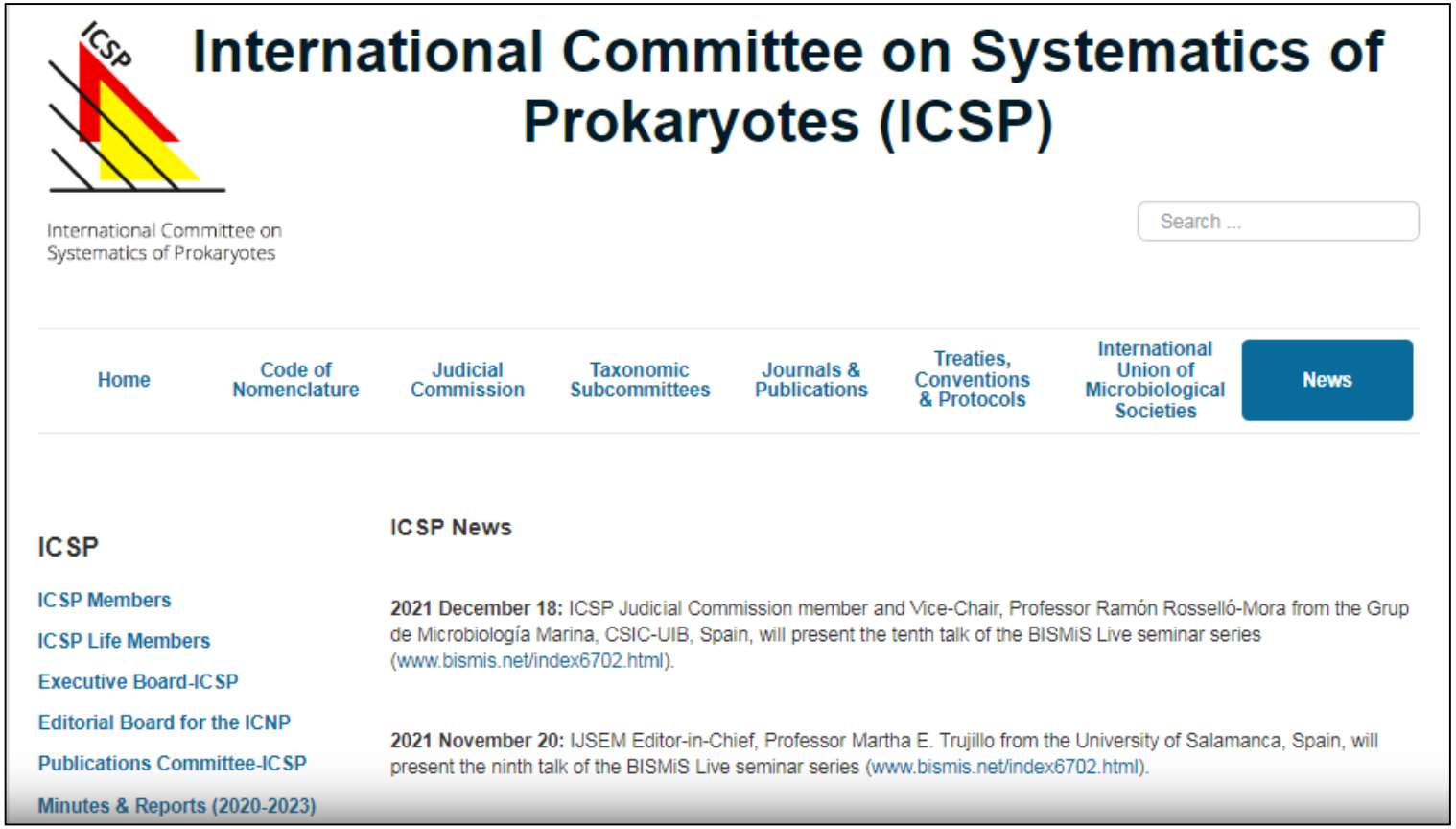
1. An executive board,
2. The members of the Judicial Commission, and
3. Members elected from member societies of the IUMS.

- In addition, the ICSP has a number of subcommittees which deal with matters relating to the nomenclature and taxonomy of specific groups of prokaryotes.

Trüper and Tindall,2005; Schleifer,2008

Statutes and activities of ICSP

The subcommittees deal with the taxonomy of different groups of bacteria and archaea



ICSP

The role of ICSP

- **The functions of the ICSP include the following:**

i. To hold meetings as part of the sessions of the International Congresses and to sponsor a session on taxonomy at the Congresses.

ii. To approve the changes recommended by the Judicial Commission.

iii. To receive and approve the Opinions issued by the Judicial Commission.

iv. To set up special Subcommittees on Taxonomy to study and make recommendations on the classification and nomenclature of bacterial taxa.

v. To publish the International Code of Nomenclature of Prokaryotes, certain official lists of bacterial names, Statutes of the ICSP, and the IJSEM.

Statutes and activities of ICSP

ICSP has 28 subcommittees which deal with the taxonomy of different groups of bacteria and archaea

- The current nomenclature of Bacteria and Archaea is officially regulated by the International Committee on Systematics of Prokaryotes (ICSP).
- There are 28 subcommittees dealing with the taxonomy of different groups of bacteria and archaea.

1. ICSP (International Committee on Systematic Prokaryotes) determines the rules by which prokaryotes are named, and whose
2. Judicial Commission issues Opinions concerning taxonomic matters, revisions to the Bacteriological Code, etc.

Trüper and Tindall,2005; Schleifer,2008; Yilmaz *et al.*,2014

ICSP

Organization and functions of taxonomic subcommittees

- Members of a Subcommittee.
- The Chairman of the ICSB (now ICSP), with the other members of the Executive Board, will select the members of the new Subcommittee on Taxonomy from names submitted and other names.
- New members may be elected to a Subcommittee at any time.
- Each Subcommittee on Taxonomy shall meet at least twice at each Plenary Meeting of the ICSB.
- Subcommittees may make recommendations regarding taxonomic procedures, changes in nomenclature, recognition of types of various taxa, or classification.

International Union of Microbiological Societies,1992

ICSP

List of Some ICSP Taxonomic subcommittees

- ICSP-Subcommittee on the taxonomy of *Rhizobium* and Ag*robacterium*;
- ICSP-Subcommittee on the taxonomy of *Bacillus* and related organisms;
- ICSP-Subcommittee on the taxonomy of *Burkholderia*, *Ralstonia* and related organisms;
- ICSP-Subcommittee on the taxonomy of *Pseudomonas*, *Xanthomonas* and related organisms;
- ICSP-Subcommittee on the taxonomy of Family *Enterobacteriaceae*;
- ICSP-Subcommittee on the taxonomy of Family *Streptomycetaceae*
- ICSP-Subcommittee on the taxonomy of Class *Mollicutes*

Trüper and Tindall,2010. Available in website maintained by the Society for General Microbiology on behalf of the ICSP.

ICSP

Functions of ICSP Taxonomic subcommittees

1. To encourage and to undertake research on the relationships of the organisms in the taxa under study…
2. To use any or all of the techniques of the several branches of science in the recognition of characters useful in distinguishing the bacteria under study…
3. To make recommendations in relation to the classification of the taxon under study…
4. To make recommendations in relation to the nomenclature of the organisms in the taxon under study…
5. To offer advice upon, and to request the Judicial Commission to adjudicate between, conflicting proposals for type species of genera and neotype strains of species and subspecies.

International Union of Microbiological Societies,1992

ICSP

Functions of ICSP Taxonomic subcommittees

6. To study the International Code of Nomenclature of Bacteria (now ICSP) and make recommendations to the Judicial Commission regarding emendations.
7. To recommend to the ICSB, through the Judicial Commission, the official designation of subdivisions of species and subspecies below the category of subspecies, e.g., phagovars of *Staphylococcus aureus*.
8. To recommend to the ICSB through the Judicial Commission minimal standards for the description of new taxa for the purpose of establishing validity of publication.

International Union of Microbiological Societies,1992

The Judicial Commission
Organization and Functions

- The Judicial Commission is a subcommittee elected by the International Committee on Systematics of Prokaryotes, ICSP (formerly the ICSB) to regulate and advise on nomenclatural matters on its behalf.
- The Judicial Commission consists of:

1. Seventeen members,
2. Twelve elected by the members of the International Committee on Systematics of Prokaryotes, the Chairman of the ICSP, and the three Secretaries.

- The Editor of the IJSEM (formerly IJSB) is, *ex officio*, a member of the Judicial Commission.

Euzeby,2010

The Judicial Commission
Revision of old code and ICSP names

- Administration and interpretation of the Code is through the Judicial Commission of the ICSP.
- Within the ICSP:

1. The Judicial Commission is responsible for ruling on any problems which arise in the nomenclature of prokaryotes, and
2. The editorial board of the Bacteriological Code is responsible for overseeing and publishing changes to the published version of the Code.

- The Judicial Commission of the ICSP is a subcommittee of the ICSP charged with the responsibility of dealing with nomenclatural matters.

Trüper and Tindall,2005;....

The Judicial Commission
Revision of old code and ICSP names

- Therefore, ICSP (including the Judicial Commission and the Lists Editor) is responsible for the working of the current Bacteriological Code and its system of registering/indexing names.
- In order to update the Code and to adjust it to modern requirements, the Judicial Commission proposed a number of changes and amendments.
- Among these changes, the Judicial Commission decided to replace the term Bacteria by the term Prokaryotes.

Trüper and Tindall,2005;....

The Judicial Commission
Revision of old code and ICSP names

- According to the proposals of the Judicial Commission (August 1999, Sydney, Australia), the words "Bacteriological Code should be changed to "Prokaryotic Code (1990 Revision).
- As a logical consequence, the complete text of the Code has to be revised accordingly:

1. The name of the International Code of Nomenclature of Bacteria is changed to International Code of Nomenclature of Prokaryotes, and
2. The name of the International Committee on Systematic Bacteriology (ICSB) is changed to International Committee on Systematics of Prokaryotes (ICSP).

Euzéby,2010

The Judicial Commission

Other functions

(i) To considers all proposals for emendation of the International Code of Nomenclature of Prokaryotes.

(ii) To consider all Requests for Opinion.

(iii) To make recommendations to the International Committee on Systematics of Prokaryotes relative to the official designation of subdivisions of species and subspecies below the category of subspecies.

(iv) To review lists prepared by the Editorial Board for the International Code of Nomenclature of Prokaryotes (*nomina conservanda*, *nomina rejicienda*...).

(v) To request Subcommittees on Taxonomy to suggest minimal standards for the description of new taxa.

(vi) To consider recommendations from Subcommittees on Taxonomy for the acceptance of a list of names as valid and applicable to recognizable taxa.

Euzeby,2010

The Judicial Commission functions
Requests for Opinions

- **Rule 4**
- In the absence of a relevant Rule or where the consequences of a Rule are uncertain, a summary in which all pertinent facts are outlined should be submitted to the Judicial Commission for consideration for preparation of a Request for an Opinion.

Opinion: A decision of the Judicial Commission of the International Committee on Prokaryote Systematics (ICPS).

The function of Judicial Commission
Requests for Opinions

- In those cases where strict adherence to the Rules of nomenclature would produce confusion or would not result in nomenclatural stability, exceptions to the Rules may be requested of the Judicial Commission of the International Committee on Systematic of Prokaryotes (ICSP).

Euzéby,2010

The function of Judicial Commission
Requests for Opinions and Judicial Opinion

- Request for an opinion should be published in the IJSB (now IJSEM).
- Request for an opinion must be accompanied by a fully documented statement of the relevant facts [Appendix 8].
- A request is considered first by the Judicial Commission and if approved by 10 or more members is then submitted to the ICSP (International Committee on Systematics of Prokaryotes) for final approval.
- Opinions and actions of the Judicial Commission on requests for opinions, are provided in the file Request for Opinion and Judicial Opinions in the IJSEM.

Judicial Opinion is an official decision taken by the Judicial Commission in favour of a proposal (published in the *International Journal of Systematic Bacteriology* or in the *International Journal of Systematic and Evolutionary Microbiology*) for nomenclatural change or for interpretation of the Principles, Rules, and Recommendations of the *Bacteriological Code* (1990 Revision)

The function of Judicial Commission
Requests for Opinions

- Opinions and actions of the Judicial Commission on Requests for Opinions, not included in the Bacteriological Code (1990 Revision), are provided below under three lists.

1. Opinions 64 to 87;
2. Some actions denied or not considered or withdrawn;
3. Opinions pending (awaiting conclusion or confirmation).

Euzéby,2010

Request for Opinions

Example: Opinion 87

Request for considering *Arthrobacter ilicis* as plant-pathogenic species

- Opinion requested by Young *et al.*,2004.
- Reconsideration of *Arthrobacter ilicis* (Mandel *et al.*,1961) Collins *et al.*,1982 as a plant-pathogenic species.
- **Judicial Commission action:**
- Mandel *et al.*,1961 considered to be pathogenic on American holly (*Ilex opaca*), this has proven not to be the case.
- Further work has shown that the species *Arthrobacter ilicis* represented by the type strain DSM 20138= ATCC 14264= NCPP B1228, is not pathogenic for American holly and should not be listed as a plant pathogenic species.
- At present *Arthrobacter ilicis* is excluded from ISPP list of plant pathogenic bacteria. *Curtobacterium flaccumfaciens* pv. *ilicis* *has* been proposed for the pathogen of American holly.

Euzéby,2014;Bull *et al.*,2010

Request for Opinions

Some actions denied or not considered or withdrawn

Type species of *Agrobacterium*

- Opinion requested by Sawada *et al.*,1993.
- Sawada *et al.*,1993 proposed that the name *Agrobacterium tumefaciens* should be rejected because it is a later synonym of *Agrobacterium radiobacter*.
- Bouzar reviewed the situation, noting that all prior commission actions had the effect of conserving *Agrobacterium tumefaciens* as the type species of the genus *Agrobacterium*.
- **Judicial Commission action:**
- The commission accepted Bouzar's position that *Agrobacterium tumefaciens* is still the type species; no further action is needed.

Note: *A. tumefaciens* and *R. radiobacter* both were used in Bull *et al.*,2010 comprehensive list as well as DSMZ validly published bacterial names (2014) list.

Euzéby,2018

Request for Opinions

Some actions denied or not considered or withdrawn

Proposal of Enterobacteraceae as a substitue for Enterobacteriaceae

- Opinions upon request:
- Lapage (Approved Lists 1980) proposed that the type genus of the family *Enterobacteriaceae* be changed from *Escherichia* to *Enterobacter* and that the family name be changed to *Enterobacteraceae*.
- **Judicial Commission action:**
- The Judicial Commission has reviewed this question and concluded that the family name *Enterobacteriaceae* retains its standing (as does *Escherichia* as the type genus) and should have been incorporated in the body of the Approved Lists.

Euzéby,2018

Request for Opinions

Some actions denied or not considered or withdrawn

Proposal to reclassify *Brenneria quercina* and *Dickeya dadantii*

- Based on 16S rRNA gene sequencing and multilocus sequence analysis (MLSA) and some chemotaxonomic methods:
- Proposal to reclassify *Brenneria quercina* (drippy nut of oak) into a new genus, *Lonsdalea* gen. nov., as *Lonsdalea quercina* comb. nov., descriptions of:

1. *Lonsdalea quercina* subsp. *quercina* comb. nov.,
2. *Lonsdalea quercina* subsp. *iberica* subsp. nov., and
3. *Lonsdalea quercina* subsp. *britannica* subsp. nov., emendation of the description of the genus *Brenneria*.

- Also reclassification of *Dickeya dadantii* (soft rot diseases on many crops)as *Dickeya dadantii* subsp. *dieffenbachiae* comb. nov., and emendation of the description of *Dickeya dadantii*.

Lonsdalea named for David Lonsdale in honour of his contributions to British forest pathology.

Brady *et al.*,2012

Request for Opinions

Opinions pending

References

1. YOUNG (J.M.), PENNYCOOK (S.R.) and WATSON (D.R.W.): Proposal that *Agrobacterium radiobacter* has priority over *Agrobacterium tumefaciens*. Request for an Opinion. *Int. J. Syst. Evol. Microbiol.*, 2006, **56**, 491-493. Original article in IJSEM Online.
2. AN (S.Y.) and YOKOTA (A.): The status of the species *Leifsonia rubra* Reddy *et al*. 2003. Request for an Opinion. *Int. J. Syst. Evol. Microbiol.*, 2007, **57**, 1163. Original article in IJSEM Online.
3. LAMBERT (D.H.), LORIA (R.), LABEDA (D.P.) and SADDLER (G.S.): Recommendation for the conservation of the name *Streptomyces scabies*. Request for an opinion. *Int. J. Syst. Evol. Microbiol.*, 2007, **57**, 2447-2448. Original article in IJSEM Online.

Euzéby,2010

The Code

Statutes

Ad hoc committees

- *Ad hoc* is a latin phrase meaning for this particular purpose, specially.
- In science *ad hoc* means the addition of extraneous (external) hypotheses to a theory to save it from being falsified (to represent falsely).
- *Ad hoc* committees, and commissions are created at the national or international level for a specific task.
- Usually these committees are used on a temporary basis, such as temporary oversight of an issue, or review of the standing rules or the constitution of that organization.
- An *ad hoc* organization may have, in some cases, a long-term or indefinite duration of existence. e.g. The Code.
- International Association of Microbiological Societies (IUMS), *Ad hoc* Committee of the Judicial Commission of ICSP (Int. Committee of Systematic of Proaryotes).

IUMS ⟶ ICSP ⟶ Judicial Commission ⟶ *ad hoc* Committees

Ad Hoc Committee of the Code
Initial functions

- At the meeting of the Judicial Commission of the ICSB held in Jerusalem on the 29th March, 1973 an Ad Hoc Committee was appointed (Minute 22):

1. To organize a review of the currently valid names of bacteria with the object of retaining only names for those taxa which were adequately described and, if cultivable, for which there was a Type, Neotype or Reference strain available;
2. To compile these names under the title of Approved Lists of Bacterial Names, and
3. To publish the lists in the International Journal of Systematic Bacteriology, to become effective on January 1, 1980.

- The Ad Hoc Committee has repeatedly stressed that its task was nomenclatural and not taxonomic.

Ad Hoc Committee of the Code
Initial functions

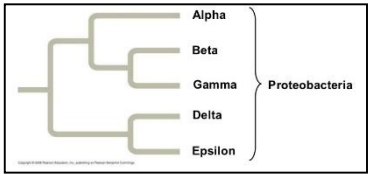
- The ad hoc Committee in 1978 was assigned wide powers to complete the work on the Lists and arrange publication.
- The ad hoc Committee asked:

1. All Subcommittees on Taxonomy of the ICSB for advice on the retention of taxa for which the Subcommittees were responsible, and
2. That specialists be approached for advice on other taxa;
3. Only a small list of taxa remained to be considered by the Ad Hoc Committee itself.

Note that Arthur Kelman as Chairman of the ISPP, convened an ad hoc meeting of interested scientists at Berkeley, in early 1978, and at this meeting, chaired by Milt Schroth, the Committee on Taxonomy of Plant Pathogenic Bacteria of the ISPP was founded.

The *ad hoc*
Functions

1. To make recommendations to the ICSP relative to the official designation of subdivisions of species and subspecies below the category of subspecies, and
2. To request revision of these recommendations when necessary, the Judicial Commission shall invite proposals from the relevant Subcommittee on Taxonomy, or in the absence of such a Subcommittee from an *ad hoc* committee of experts appointed to submit such proposals (including species definition).



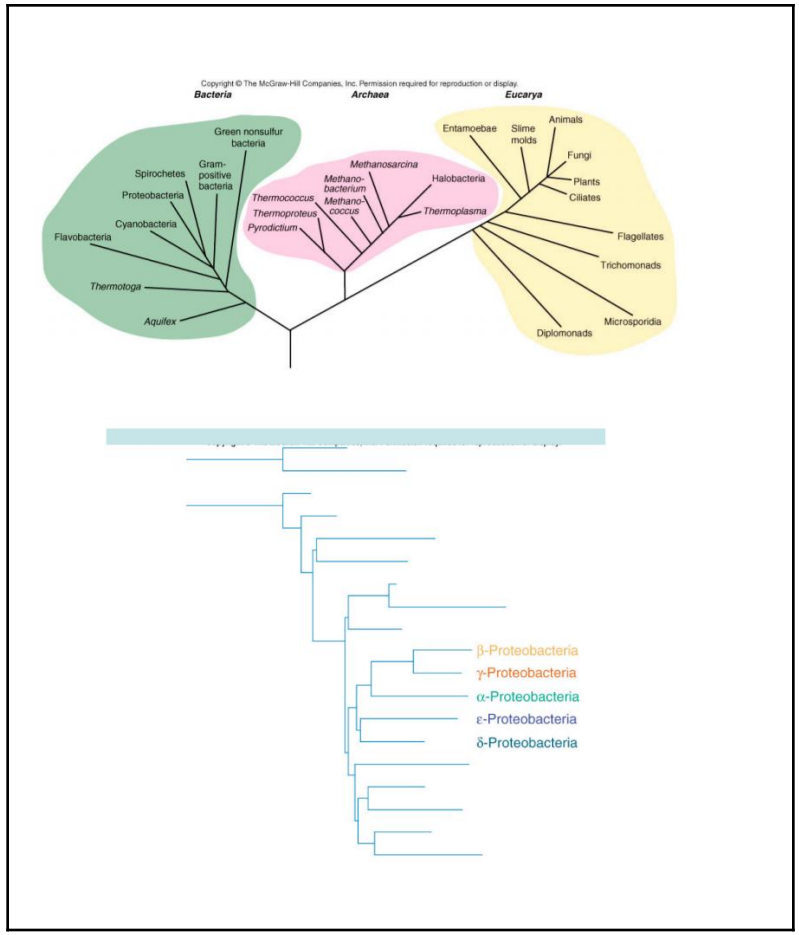
Ad Hoc Committee on Approaches to Taxonomy within the Proteobacteria

- This ad hoc committee was formed following the workshop on reconciliation (the process of making consistent or compatible) of approaches to bacterial systematics (Wayne *et al.*,1987) because it was realized that the "purple bacteria and their relatives" are a major phylum of Gram-negative bacteria.
- They are classified into five classes, alpha (a), beta(b), gamma (g), delta (d) and epsilon (e), distinguished by small differences in their rRNA sequences.
- The gammaproteobacteria are the largest class in terms of species with validly published names.

Ad Hoc Committee on Approaches to Taxonomy within the Proteobacteria

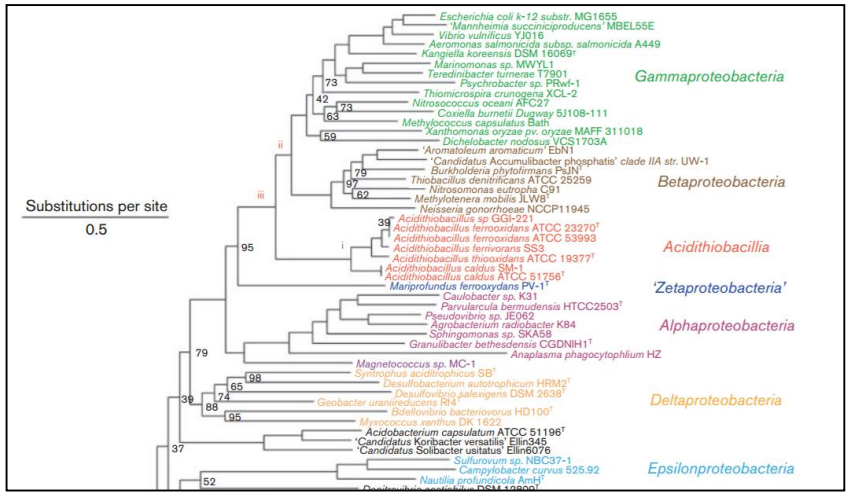
- The Proteobacteria:
- largest cultured group of bacteria - 380 genera & at least 1500 species!
- 5 major groups ("classes"):

1. Alphaproteobacteria
2. Betaproteobacteria
3. Gammaproteobacteria
4. Deltaproteobacteria
5. Epsilonproteobacteria.



Domain: Bacteria
Phylum: Proteobacteria

- A new class (a sixth class) within the phylum Proteobacteria, Acidithiobacillia classis nov.,was proposed by Williams and Kelly,2013 and replaced by the 'Zetaproteobacteria', a sixth class was proposed earlier by Emerson *et al.*,2007 and McAllister *et al.*,2011).



Zetaproteobacteria was excluded by Williams and Kelly,2013.

Williams and Kelly,2013

Domain: Bacteria
Phylum: Proteobacteria

Some selected plant diseases caused by alpha, beta and gamma subdivisions of the Phylum Proteobacteria.

The gammaproteobacteria are the largest class in terms of species.

Proteobacterial class and species	Family [a]	Disease (symptoms)
"Alphaproteobacteria"		
Agrobacterium rhizogenes	Rhizobiaceae	Hairy root
Agrobacterium tumefaciens	Rhizobiaceae	Crown gall
"Candidatus Liberibacter asiaticus"	in cluster of Rhizobiaceae, Bartonellaceae, etc.	Greening disease on citrus (a phloem-restricted disease)
"Betaproteobacteria"		
Acidovorax anthurii	Comamonadaceae	Leaf-spot on *Anthurium*
Burkholderia cepacia	"Burkholderiaceae"	Soft rot (sour skin on onion)
Burkholderia glumae	"Burkholderiaceae"	Sheath necrosis on rice
Ralstonia solanacearum	"Ralstoniaceae"	Moko disease on banana (vascular wilt)
Xylophilus ampelinus	Comamonadaceae	Necrosis and canker on grapevine
"Gammaproteobacteria"		
Brenneria (Erwinia) salicis	Enterobacteriaceae	Watermark disease on willow
Brenneria nigrifluens	Enterobacteriaceae	Bark canker on Persian walnut (*Juglans regia*)
Erwinia amylovora	Enterobacteriaceae	Fire blight on pome fruit (vascular wilt)
Erwinia stewartii	Enterobacteriaceae	Stewart's wilt on corn (vascular wilt)
Pectobacterium (Erwinia) carotovorum	Enterobacteriaceae	Soft rot
Pseudomonas agarici	Pseudomonadaceae	Spots on mushrooms
Pseudomonas marginalis	Pseudomonadaceae	Soft rot (pink eye) on potato
Pseudomonas savastanoi	Pseudomonadaceae	Galls on olive trees
Pseudomonas syringae	Pseudomonadaceae	Wildfire on tobacco, haloblight on beans, spots on tomato and pepper (blights and spots)
Pseudomonas syringae	Pseudomonadaceae	Canker on stone fruit
Xanthomonas campestris	"Xanthomonadaceae"	Black rot on crucifers (vascular wilt)
Xanthomonas citri	"Xanthomonadaceae"	Canker on citrus
Xanthomonas oryzae	"Xanthomonadaceae"	Blight on rice
Xanthomonas populi	"Xanthomonadaceae"	Canker on poplar trees
Xanthomonas translucens	"Xanthomonadaceae"	Blight on cereals
Xanthomonas vesicatoria	"Xanthomonadaceae"	Spots on tomato and pepper
Xylella fastidiosa	"Xanthomonadaceae"	Pierce's disease (e.g., on grapevine)

[a] According to *Bergey's Manual of Systematic Bacteriology* (Garrity and Holt, 2001). See also Fig. 1. Quotation marks are used for names which have not yet been validated (as of mid 2002).

The function of *Ad Hoc*

Tips and suggestions for the practising rhizobial taxonomist

- Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology:

1. **How many strains should be used to describe a new rhizobial species?**

- Answer: New species descriptions based on a single isolate are strongly discouraged!

2. **How many and which molecular markers should be used?**

- Answer: A good compromise would be to generate full-length 16S rDNA sequences for a few carefully selected strains, along with the partial sequencing of two protein-coding core loci (e.g. *rec*A and *rpo*B) and at least one sym (symbiotic im rhizobium) locus.

See also Minimal standards

The function of *Ad Hoc*

Tips and suggestions for the practising rhizobial taxonomist

3. **What phenotypic tests should be performed?**

- Answer: Of particular interest are phenotypes and chemotaxonomic markers that are relevant or expressed in the niches potentially occupied by the target organisms (i.e. ecological adaptive traits). When coupled with clear genetic determinants, such ecological traits are very informative and of extraordinary taxonomic value.

4. **What about DNA-DNA hybridization or reassociation experiments?**

- We favour the use a thorough multilocus sequence analysis (MLST/MLSA) combined with key phenotypic analyses for species demarcation. This evidence should be ideally further extended with 16S rDNA sequencing of the type strain and two other strains, combined with DNA-DNA hybridization data for these reference strains and the relevant phylogenetic relatives.

ad hoc
References

- Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, 0. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore,R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Truper. 1987. Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol 37: 463-464.
- Murray, R. G. E., D. J. Brenner, R. R. Colwell, P. DE VOS, M. Goodfellow, P. A. D. Grimont, N. Pfennig, E. Stackebrandt, and G. A. Zavarzin. 1990. Report of the Ad Hoc Committee on Approaches to Taxonomy within the Proteobacteria. 1990. Int. J. Syst. Bacteriol. 40: 213-215
- Report of the ad hoc committee on the reconciliation of approaches to bacterial systematic.1998. Int. J. Syst. Bacteriol. 37:463-464.
- Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology.2002. Int J Syst Evol Microbiol. 52(Pt 3):1043-7.

The Code
Principles, Rules and Recommendations

- The Code consists of:

1. Principles
2. Rules
3. Recommendations (Lapage *et al.*,1992).

- The Principles form the basis of the Code, and the Rules and Recommendations are derived from them.

The Code

Principles, Rules and Recommendations

- The Principles are general concepts that form the basis and overall guidance for the Code.
- The Rules are derived from these Principles and provide the detailed methods for practical application of the Principles in the naming of bacteria.
- The Rules must be followed when proposing names for bacteria.
- The Recommendations provide additional guidance, but in contrast to the Rules are not obligatory.

Bull *et al.*,2008

The Code

Principles, Rules and Recommendations

Some Principles

- **Principle 1**
- The essential points in nomenclature are as follows:
- Aim at stability of names,
- Avoid or reject the use of names which may cause error or confusion,
- Avoid the useless creation of names.

Names that have been changed to accord with correct Latin are listed in Bull *et al.*,2010 comprehensive bacterial names as 'corrig.' Since these corrections were made, the Code has been amended to take account of the need for stability of names.

The Code

Principles, Rules and Recommendations

Some Principles

- **Principle 2**

  According to Principle 2, the nomenclature of Prokaryotes is not independent of botanical and zoological nomenclature.

- When naming new taxa in the rank of genus or higher, due consideration is to be given to avoiding names which are regulated by the Zoological Code and the International Code of Botanical Nomenclature.

The Code
Principles, Rules and Recommendations

- **Principle 3**
  The scientific names of all taxa are Latin or latinized words treated as Latin regardless of their origin.
- They are usually taken from Latin or Greek.

Nomenclature: "Bacterial names may originate from any language, but they must be given a Latin suffix." (i.e., they must be Latinized).

The Code

Principles, Rules and Recommendations

Some Rules

- **The Rules:**
- The Rules only apply:
- To ranks at the level of subspecies and above, **but** not to lower taxonomic ranks.

Bull *et al.*,2008

The Code

Principles, Rules and Recommendations

Some Rules

- Rules for nomenclature ensure that changes to names of taxa follow a logical progression that maintains the integrity of the previous nomenclature while replacing it with new proposals based on additional scientific injury into relationships among organisms.

Bull *et al.*,2010

The Code

Principles, Rules and Recommendations

Some Rules

- **Section 1. General**
- **Rule 1a**
  This revision supersedes all previous editions of the International Code of Nomenclature of Bacteria. It shall be cited as Bacteriological Code (1990 Revision) and will apply from the date of publication (1992).
- **Rule 1b**
  Alterations to this Code can only be made by the ICSB at one of its plenary sessions. Proposals for modifications should be made to the Editorial Secretary in sufficient time to allow publication in the IJSB before the next International Congress of Bacteriology. For this and other Provisions, see the Statutes of the ICSB, pp. 137 158.
- **Rule 2**
  The Rules of this Code are retroactive (influencing), except where exceptions are specified.

See binomial nomenclature for more rules

The Code

Principles, Rules and Recommendations

Some Recommendations

- **Section 1. General**
- **Rule 3**
- Names contrary to a Rule can not be maintained, except that the International Committee on Systematic Bacteriology (now ICSP), on the recommendation of the Judicial Commission, may make exceptions to the Rules (see Rule 23a and the Statutes of the ICSB).

The Code

Principles, Rules and Recommendations

Some Recommendations

- The Recommendation provide additional guidance, but in contrast to the Rules are not obligatory.
- It is recommended that names be reasonably short and easy to pronounce (Recommendation 6).
- Names may be chosen to recognize individual bacteriologists or pathologists, but names honoring individuals quite unconnected with bacteriology are discouraged (Recommendations 6, 10a, 12c).
- For example, a novel species of *Agrobacterium* from *Ficus benjamina* was proposed and named *Agrobacterium larrymoorei* (of Larry Moore) to honor the "renowned plant pathologist who spent his career studying the genus *Agrobacterium*" (Bouzar and Jones,2001).

Bull *et al.*,2008

The Code

Principles, Rules and Recommendations

Some Recommendations

- **Longest and shortest prokaryotic names:**
- Recommendattion 6 (1) of the *Bacteriological Code* (1990 Revision) states the following:
- Avoid names or epithets that are very long or difficult to pronounce.
- In spite of Recommendation 6 there are some long genus names and specific epithets.

Euzeby,2016

The Code

Some Recommendations

Longest prokaryotic names

Longest prokaryotic names

Names	Genus names	Specific epithets
Forty two letters *Thermoanaerobacterium thermosaccharolyticum*	Twenty four letters *Hydrogenoanaerobacterium*	Twenty six letters *saccharoperbutylacetonicum*
Forty letters *Thermoanaerobacterium polysaccharolyticum*	Twenty three letters *Allocatelliglobosispora*	Twenty four letters *nitratireducenticrescens* *phaeogriseichromatogenes*
Thirty nine letters *Methanothermococcus thermolithotrophicus* *Thermoanaerobacter thermohydrosulfuricus* *Thermoanaerobacterium thermosulfurigenes*	Twenty two letters *Caldicellulosiruptor* *Clostridiisalibacter* *Desulfonatronovibrio* *Methanomethylovorans* *Phascolarctobacterium* *Pseudoflavonifractor* *Pseudosphingobacterium* *Selenihalanaerobacter* *Sulfurihydrogenibium* *Thermoanaerobacterium* *Thermodesulfobacterium* *Thermodesulforhabdus* *Thermoflavimicrobium* *Thermosediminibacter* *Thermoterrabacterium*	Twenty three letters *arabinogalactanolyticum* *phaeoluteichromatogenes* *purpurogeneiscleroticus*
Thirty eight letters *Hydrogenoanaerobacterium saccharovorans* *Streptoverticillium griseoverticillatum* *Thioalkalivibrio thiocyanodenitrificans*		Twenty two letters *psychrosaccharolyticus* *tetrahydrofuranoxydans* *thiocyanodenitrificans*
Thirty seven letters *Aureobacterium arabinogalactanolyticum* *Clostridium saccharoperbutylacetonicum* *Methanothermobacter thermautotrophicus* *Microbacterium arabinogalactanolyticum* *Streptoverticillium luteoverticillatum* *Streptoverticillium olivoverticillatum* *Streptoverticillium rectiverticillatum* *Streptoverticillium roseoverticillatum* *Thermodesulfobacterium hydrogeniphilum*		Twenty one letters *actinomycetemcomitans* *alkalidiazotrophicus* *hydrocarbonoclasticus* *thermocarboxydovorans* *thermohydrosulfuricum* *thermohydrosulfuricus* *thermosaccharolyticum*

Euzeby,2017

The Code

Some Recommendations

Shortest prokaryotic names

Shortest prokaryotic names

Names	Genus names	Specific epithets
Ten letters	Five letters	Three letters
Dyella soli	*Asaia*	*uli*
Vibrio xuii	*Bosea*	*uda*
Yania flava	*Dorea*	
	Iamia	Four letters
Eleven letters	*Leeia*	*agri*
Afipia felis	*Orbus*	*alba*
Blautia luti	*Yania*	*alni*
Deleya aesta		*alvi*
Devosia limi	Six letters	*apis*
Devosia soli	*Afipia*	*arvi*
Dickeya zeae	*Agreia*	*boum*
Erwinia alni	*Asanoa*	*cati*
Frankia alni	*Basfia*	*ceti*
Hafnia alvei	*Deleya*	*cibi*
Haliea rubra	*Derxia*	*coli*
Kaistia soli	*Dongia*	*cyri*
Leeia oryzae	*Dyella*	*doii*
Nevskia soli	*Faenia*	*dura*
Rothia aeria	*Fangia*	*dyei*
Runella zeae	*Gelria*	*equi*
Stappia alba	*Hafnia*	*etli*
Vibrio casei	*Haliea*	*fimi*
Vibrio logei	*Hellea*	*gari*
Vibrio rarus	*Hippea*	*humi*
Vibrio ruber	*Hongia*	*lari*
	Jejuia	*lata*
	Kordia	*limi*
	Labrys	*lini*
	Nisaea	*loti*
	Olleya	*luti*
	Orenia	*mali*
	Pragia	*mays*
	Pricia	*mira*
	Rothia	*nova*
	Ruania	*oeni*
	Rudaea	*olei*
	Salana	*oris*
	Shimia	*ovis*
	Stella	*peli*
	Vibrio	*pini*
	Yangia	*pisi*
	Zhouia	*poae*
		rapi
		rosa
		rubi
		sedi
		sera
		soli
		suis
		tofi
		ulmi
		vaga
		vini
		xuii
		xyli
		yeei
		zeae

Euzeby,2017

Rejected names

nomen rejiciendum: nom.rej., or nom. rejic.

- Only the Judicial Commission can reject names (*nomina rejicienda*). [Rules 23a Note 4 (i) and 56a,b].
- The Judicial Commission may place on the list of rejected names a name previously published in an Approved List [Rule 24c].
- A name may be placed on this list (list of rejected names) for various reasons, including the following [Rule 56a]:

1. An ambiguous name (*nomen ambiguum*), i.e., a name which has been used with different meanings and thus has become a source of error.
2. A doubtful name (*nomen dubium*), i.e., a name whose application is uncertain.

Euzeby,2010

Rejected names

nomen rejiciendum: nom.rej., or nom. rejic.

3. A name causing confusion (*nomen confusum*), i.e., a name based upon a mixed culture.
4. A perplexing name (*nomen perplexum*), i.e. a name whose application is known but which causes uncertainty in bacteriology (see Rule 57c).
5. A perilous(dangerous) name (*nomen periculosum*), i.e., a name whose application is likely to lead to accidents endangering health or life or both or of serious economic consequences.

- Example: *Yersinia pseudotuberculosis* subsp. *pestis* is placed on the list of *nomina rejicienda* because of practical concerns about human welfare.

Euzeby,2010

Rejected names

Assignment of the genus *Sphingomonas* as *Sphingomonas suberifaciens*

- Selected names must not have been rejected previously for specific reasons (Rule 56a; see Euzéby,2005).
- For example, the genus name *Rhizomonas* is listed as a "rejected name of genera and subgenera of bacteria" (*nomina generum et subgenerum bacteriorum rejicienda*) because this genus name had earlier been assigned to a taxon of protozoa (*Rhizomonas* Kent 1880).
- According to Rule 51b, *junior homonyms* of a taxon of bacteria, fungi, algae, protozoa, or viruses are illegitimate.
- Thus, the proposal to use *Rhizomonas* as the genus name for the pathogen causing corky root of lettuce (*Rhizomonas suberifaciens*; Van Bruggen *et al.*,1990) violated the Code.
- This bacterium has since been assigned to the genus *Sphingomonas* as *Sphingomonas suberifaciens* (Yabuuchi *et al.*,1999).

Homonymy is the term applied when the same name is given to two or more different taxa of the same rank based on different types. The first published name is known as the senior homonym and any later published name as a junior homonym.

Bull *et al.*,2008

Revived names

nomen revictum: nom. Rev.

- A revived name is a name which was published prior to 1 January 1980 but not included in the Approved Lists of Bacterial Names and which is proposed by an author for a different or for the same taxon (in the author's opinion) [Rules 28a and 33c].
- If an author wishes to indicate the names of the original authors of a revived name, he may do so by citation of the name of the taxon, followed by the word "*ex*" and the name of the original author and the year of publication, in parentheses, followed by the abbreviation "nom. rev" (*nomen revictum*) [Rule 33c Note 2].
  Example:
- *Corynebacterium iranicum* (ex Scharif 1961) nom. rev., and *Corynebacterium tritici* (ex Hutchinson 1917) nom. rev.
- Present name is *Rathayibacter iranicus*
- Note: A rejected name can not be revived [Rule 28a].

Euzeby,2010

Conserved names

nomen conservandum : nom. Cons.

- A conserved name (*nomen conservandum*) is a name which must be used instead of all earlier synonyms and homonyms [Rules 23a Note 4, and 56b].
- Only the Judicial Commission can place names on the list of conserved names [Rules 23a Note 4 (i) and 56b Note 2].
- Example:
- **Conserved family names (*Nomina familarum conservanda*):**

1. *Enterobacteriaceae* Rhan 1937

- **Conserved names of genera of bacteria (Nomina *generum bacteriorum conservanda*):**

1. *Agrobacterium* Conn 1942
2. *Pseudomonas* Migula 1894

- **Conserved specific epithets in names of species of bacteria (*Epitheta specifica conservanda*):**

1. *sporogenes* in *Clostridium sporogenes* (Metchnikoff 1908) Bergey *et al.*,1923 (Approved Lists,1980)

Homonymy is the term applied when the same name is given to two or more different taxa of the same rank based on different types. The first published name is known as the senior homonym and any later published name as a junior homonym.

Contentious names

- Some proposals are contentious (controversy, contrary) names. e.g. *Xanthomonas fuscans* proposed by Schaad *et al.*,2006 is considered to be a synonym of *X. citri*.
- In these cases it may be adequate to use either:
- The latest validly published combinations, such as those indicated in the ISPP lists, or to use traditional names that no longer necessarily reflect relationships accurately, e.g. *Agrobacterium* (Young,2008).

Proposed name
Xanthomonas citri
**X. citris* subsp. *citri*
**X. citris* subsp. *malvacearum*
Xanthomonas fuscans
**X. fuscans* subsp. *fuscans*
**X. fuscans* subsp. *aurantifolii*
Xanthomonas alfalfae
**Xanthomonas alfalfae* subsp. *alfalfae*
**Xanthomonas alfalfae* subsp. *citrumelonis*

*** Valid names. For more details refer to Bull *et al.*,2010 list.**

Name causing confusion
nomen confusum

- A name based upon a mixed culture.
- Such a name should be rejected [Rule 56a].
- Example:
- *Methanosarcina methanica* nom. rejic. (Opinion 63).

Euzeby,2010

Orthography of names

The system for developing correctly constructed names

- The names of all taxa are Latin or latinized words treated as Latin regardless of their origin [Principle 3, Rule 6].
- Although standardization of orthography produces elegant (beautiful) names in a regularized form, bacteriologists are usually untrained in Latin grammar and many mistakes have been made.
- Some example of the names need correction:
- *Streptomyces scabies* to *S. scabiei*,
- *Erwinia carotovora* to *E. carotivorax*,
- *E. tracheiphila* to *E. trachiiphila*,
- *Pseudomonas savastanoi* to *P. savastanonii* or *P. savastanonis*,
- *Rhizobium leguminosarum* to *R. leguminum*,
- *Xanthomonas axonopodis* to *X. axonopi*

Citation style guide
Author/Editor APA/MLA Format

- When the new name of a taxon is validly published under more than two authors, and when there is not definite designation of a single individual as the author of the name, the citation may be made by listing the names of all the authors or by giving the name of the first author, followed by the abbreviation "*et al*." (*et alii*) [Chapter 4. Advisory notes. B. Quotations of authors and names (1)].
- Example:
- *Afipia felis* Brenner, Hollis, Moss, English, Hall, Vincent, Radosevic, Birkness, Bibb, Quinn, Swaminathan, Weaver, Reeves, O'Connor, Hayes, Tenover, Steigerwalt, Perkins, Daneshvar, Hill, Washington, Woods, Hunter, Hadfield, Ajello, Kaufmann, Wear and Wenger 1992, or
- *Afipia felis* Brenner *et al*. 1992.

Euzeby,2014

Multiple authorship

Comparing MLA and APA Citation Styles

- Here are some ways in which MLA (Modern Language Association) and APA (American Psychological Association) Citation styles are different:

	MLA (Modern Language Association)	APA (American Psychological Association)
Authors/editors	Spells out available names. If more than 3 list first and "et al."	Uses last name and first initial. Lists all authors
Titles	Capitalizes every important word	Capitalizes first word
Publisher	Shortens name	Uses full name
Place of publication	Only lists city	Lists state abbreviations when city is lesser known
Pages	Uses + for pages after the first one	Lists first page and additional pages
Dates	Placed at end of citation	Follows authors name
Indentions	First line is flush with left margin, second and subsequent lines are indented	First line of an entry is indented, second and subsequent lines are flush with the left margin
Parenthetical Citation in Text	Uses author name and page	Uses author name, date of publication, and page number

http://students.msbcollege.edu/library/guides/citecompare.html

Multiple authorship

Comparing MLA and APA Citation Styles

- Here are some ways in which MLA (Modern Language Association) and APA (American Psychological Association) Citation styles are different:

	MLA	APA
Books	Gibaldi, Joseph. MLA Handbook for Writers of Research Papers. 6th ed. New York: MLA, 2003.	Gibaldi, J. (2003) *MLA handbook for writers of research papers*. 6th ed. New York: MLA.
Journal articles	Donaldson, Scott. "Protecting the Troops from Hemingway: An Episode in Censorship." The Hemingway Review 15 (1995): 87-93.	Donaldson, S. (1995). Protecting the troops from hemingway: an episode in censorship. The Hemingway Review, 15, 87-93.
Article from databases (library subscription service)	Anderson, John. "Superego." Journal of Psychology October 2003: 41. Academic Search Elite. EBSCOhost. MSB Brooklyn Center Campus Library, Brooklyn Center, MN. 10 November 2003 <http://search.epnet.com>.	Anderson, J. (2003, October) Superego. Journal of Psychology, 41. Retrieved November 10, 2003, from EBSCOhost.
Quotes and in-text citations	In the paper you must cite the source in parentheses with the author's last name and page number of the source used at the end of the sentence. Examples: ❑ While learning to do research for your academic classes, you will gain skills that will prepare you to do research in your career within business, government, or other professional areas (Gibaldi 5). ❑ Gibaldi describes…and he also writes "this handbook is a guide for the preparation of research papers" (5).	In the paper you must cite the author and date in the appropriate place within the text or at end of the sentence. Examples: ❑ Anderson (1984) urges roommates to settle differences... ❑ In her list of rules for roommates (Anderson, 1984)... ❑ She stated, "The list of rules… in this manner" (Anderson, 1984, p. 198), but she still was upset.
Citation List	This page must be titled Works Cited.	This page must be titled Reference(s).

http://students.msbcollege.edu/library/guides/citecompare.html

Multiple authorship
Comparing MLA and APA Citation Styles

- **APA Citation Style:**
- Donaldson, S. (1995). Protecting the troops from hemingway: an episode in censorship. The Hemingway Review, 15, 87-93.
- **MLA Citation Style:**
- Donaldson, Scott. "Protecting the Troops from Hemingway: An Episode in Censorship." The Hemingway Review 15 (1995): 87-93.

http://lib.radford.edu/tutorial/VII/comp.asp,

Multiple authorship

Comparing MLA and APA Citation Styles

MLA style

- In MLA style, use "and" before the final author's name.
- Reverse the first and last name for the first author only, and separate each author's name with a comma.
- If two authors have the same last name, you should list each full name separately.
- When there are four or more authors, use only first author's name followed by "et al."
- **MLA Example**
- *Psychology* by Douglas A. Bernstein, Louis Penner, Alison Clarke-Stewart, Edward Roy.
- Cited as an example:
  Bernstein, Douglas, et al. *Psychology*. 8th ed. Boston: Houghton, 2008. Print.

Multiple authorship
Comparing MLA and APA Citation Styles
APA style

- APA (American Psychological Association) style is most commonly used to cite sources within the social sciences.
- In APA style, reverse the first initials and last name of each author, and separate each author's name with a comma.
- Use an ampersand before the last author's name.
- When there are more than seven authors, list the first six, followed by an ellipsis, and then the last author.
- Otherwise, list every author by name.
- **APA Example**
- *Psychology* by Douglas A. Bernstein, Louis Penner, Alison Clarke-Stewart, Edward Roy.
- Cited as an example:
  Bernstein, D., Penner, L., Clarke-Stewart, A., & Roy, E. (2008). *Psychology* (8th ed.). Boston, MA: Houghton Mifflin.

Multiple authorship
Three, Four or Five authors
APA style

- **In-text:**
- According to Hall, Timothy, and Duval (2007) ... OR Research indicates ... (Hall, Timothy, & Duval, 2007)
- Cite all authors the first time in-text. In subsequent citations, include only the surname of the first author followed by et al. and year of publication, e.g. (Hall et al., 2007)
- **Reference List:**
- Hall, C. M., Timothy, D. J., & Duval, D. T. (2007). *Safety and security in tourism: Relationships, management and marketing*. Mumbai, India: Jaico Publishing House.
- Separate the final name from the rest with a comma and an ampersand.

Multiple authorship

Six or seven authors

APA style

- **In-text:**
- According to Roeder et al. (1967) ...
- In-text cite only the surname of the first author followed by et al.
- **Reference List:**
- Roeder, K., Howdeshell, J., Fulton, L., Lochhead, M., Craig, K., & Peterson, R. (1967). *Nerve cells and insect behaviour*. Cambridge, MA: Harvard University Press.
- Include all names in the Reference List.

Multiple authorship

Eight or more authors

APA style

- **In-text:**
- Yoon et al. (2001) argue OR
  Research indicates ... (Yoon et al.,2001)
- **Reference List:**
- When authors number eight or more, include the first six authors' names, then insert three dots and add the last author's name:
- Yoon, P. W., Chen, B., Faucett, A., Clyne, M., Gwinn, M., Lubin, I. M., ... Muin, J. (2001). Public health impact of genetic tests at the end of the 20th century. *Genetics in Medicine*, *3*, 405-410.

Multiple authorship

More than 10 authors

Sample International Journal of Systematic and Evolutionary Microbiology references

1. For references with ten or fewer authors, give the names of all authors in the form "Surname, Initials".
2. For references with more than ten authors, list the first nine followed by "& other authors". e.g.

- **In-text:**
- Williamson *et al.*,1998 or Hackett *et al.*,1996
- **Reference List:**
- Williamson, D. L., Whitcomb, R. F., Tully, J. G. & 10 other authors (1998). Revised group classification of the genus *Spiroplasma*. *Int J Syst Bacteriol* 48, 1-12.
- Hackett, K. J., Whitcomb, R. F., Clark, T. B. & 12 other authors (1996). *Spiroplasma leptinotarsae* sp. nov., a mollicute uniquely adapted to its host, the Colorado potato beetle, *Leptinotarsa decemlineata* (Coloeoptera: Chrysomelidae). *Int J Syst Bacteriol* 46, 906-911.

Multiple authorship

More than 10 authors

Sample International Journal of Systematic and Evolutionary Microbiology references

- The total number of the authors is less than 10:
- Ruggiero MA, Gordon DP, Orrell TM, Bailly N, Bourgoin T, Brusca RC, et al. (2015) Correction: A Higher Level Classification of All Living Organisms. PLoS ONE 10 (6): e0130114. doi:10.1371/journal.pone.0130114

Two or more works by the same author in the same year

- If you are using more than one reference by the same author (or the same group of authors listed in the same order) published in the same year, organize them in the reference list alphabetically by the title of the article or chapter. Then assign letter suffixes to the year. Refer to these sources in your essay as they appear in your reference list:
- Berndt, T. J. (1981a). Age changes and changes over time in prosocial intentions and behavior between friends. *Developmental Psychology*, *17*, 408-416.
- Berndt, T. J. (1981b). Effects of friendship on prosocial intentions and behavior. *Child Development*, *52*, 636-643.

Harvard citation style

- **In-text:**
- in response to institutional priorities and community or user needs and interests" (Johnson 2009, p. 1). According to Johnson (2009, p. 1),
- **Reference List:**
- Carlson, A & Pope, BM 2009, 'The "Big Deal": A survey of how libraries are responding and what the alternatives are', *The Serials Librarian*, vol. 57, no. 4, pp. 380-398. Available from: Taylor & Francis Online. [28 September 2015].
- Johnson, P 2009, *Fundamentals of collection development and management*, 2nd edn, ALA Editions, Chicago.
- University of Western Australia Library 2015, *Collection management principles and policies*. Available from: http://www.library.uwa.edu.au/information-resources/collections/management. [14 October 2015].

A digital object identifier (DOI)
The DOI system was introduced it in 2000

- A DOI, or Digital Object Identifier, is a string of numbers, letters and symbols used to permanently identify an article or document and link to it on the web.
- The publisher assigns a DOI when your article is published and made available electronically.
- All DOI numbers begin with a 10 and contain a prefix and a suffix separated by a slash.

1. The prefix is a unique number of four or more digits assigned to organizations;
2. The suffix is assigned by the publisher and was designed to be flexible with publisher identification standards.

Morey, C. C., Cong, Y., Zheng, Y., Price, M., & Morey, R. D. (2015). The color-sharing bonus: Roles of perceptual organization and attentive processes in visual working memory. *Archives of Scientific Psychology*, *3*, 18–29. https://doi.org/10.1037/arc0000014

A digital object identifier (DOI)

- The reference list entry for a chapter in an edited e-book should be written as follows:

Author, A. A. (Year). Title of chapter. In B. B. Editor (Ed.), Title of book [E-reader version, if applicable] (pp. xxx–xxx). Retrieved from http://xxxxx
Author, A. A. (Year). Title of chapter. In B. B. Editor (Ed.), Title of book [E-reader version, if applicable] (pp. xxx–xxx). doi:xxxxx

Species concept for *Spiroplasma*

- Although the species is the basal category for taxonomic classification, the definition of a bacterial species is relatively vague and artificial.
- The basic species concept for *Spiroplasma* is rooted in the species concept for other prokaryotic taxa, which requires genomes exhibiting less than 70% similarity for separate species designations.

Redgassa and Gasparich,2006

Candidatus

New category of indefinite rank

Not covered by the Rules of the Bacteriological Code

- According to the "Ad Hoc Committee for the re-evaluation of the species definition in bacteriology", microbiologists are encouraged to use the "*Candidatus*" concept for well characterized but as yet uncultured organisms.
- It is important to note that the category *Candidatus* is not covered by the Rules of the Bacteriological Code.
- Consequently, a name included in the category *Candidatus:*

1. Cannot be validly published, and it also
2. Cannot be designated sp. nov., gen. nov., etc.

Euzeby,2014

Phytoplasma nomenclature

Candidatus nomenclature

- *Candidatus* species are differentiated primarily on the basis of the comparative analysis of 16S rRNA sequences, other sequence data, ecological information, and metabolic data where available.
- The names included in the category *Candidatus* are usually written as follows:
- *Candidatus* (in italics), the subsequent name (s) in roman type (with an initial cap for the genus name) and the entire name in quotation marks.
- For example
- "*Candidatus* Phytoplasma allocasuarinae"
- "*Candidatus* Liberibacter asiaticus".

Young,2008; Euzeby,2014

Phytoplasma nomenclature
Candidatus nomenclature

- *Candidatus* species are differentiated primarily on the basis of the:
- Comparative analysis of 16S rRNA sequences, other sequence data, ecological information, and metabolic data where available.

1. Most plant pathogenic *Candidatus* are members of the genus '*Candidatus* Phytoplasma'.

- Others are in different taxa such as:

1. '*Candidatus* Liberibacter',
2. '*Candidatus* Phlomobacter'
3. 'Candidatus Burkholderia hispidae'- the bacterial endosymbionts in leaf nodules of *Pavetta* (Rubiaceae).
4. '*Candidatus* Streptomyces philanthi'- an endosymbiotic streptomycete in the antennae of *Philanthus* digger wasps.

Young,2008; Euzeby,2014

Phytoplasma nomenclature

Candidatus nomenclature

Some Rules

- In 1992, Subcommittee on the Taxonomy of Mollicutes proposed the use of the name Phytoplasma in place of the use of the term MLO (Mycoplasma-like organism) "for reference to the phytopathogenic mollicutes".
- In 2004 the genus name Phytoplasma was adopted and is currently at *Candidatus* status which is used for bacteria that can not be cultured (IRPCM,2004).
- According to the "IRPCM Phytoplasma/Spiroplasma Working Team - Phytoplasma taxonomy group" the abbreviation for *Candidatus* should be *Ca*.

The International Committee on Systematics of Prokaryotes (ICSP)-Subcommittee on the Taxonomy of Mollicutes. Minutes of the meetings, 7 and 12 July 2002, Vienna, Austria.

Phytoplasma nomenclature

Rules for the description of '*Ca*. Phytoplasma'

<97.5% 16S rRNA gene sequence similarity to any previously described '*Ca*. is considered as a novel '*Ca*. Phytoplasma' species

- Rules for the description of organisms as novel taxa within '*Ca*. Phytoplasma', as follows:

a) A '*Candidatus* (*Ca*.) Phytoplasma' species description should refer:

1. To a single, unique 16S rRNA gene sequence (>1200 bp), and
2. That a strain can be recognized as a novel 'Ca. Phytoplasma' species if its 16S rRNA gene sequence has <97.5% similarity to that of any previously described 'Ca. Phytoplasma' species.

- The strain from which this sequence was obtained should be named the 'reference strain' and not the 'type strain'.
- Strains in which even minimal differences in the 16S rRNA gene sequence from the reference strain are detected do not 'belong' to the *Candidatus* species, but are 'related' to it.

IRPCM Phytoplasma taxonomy group,2004

Phytoplasma nomenclature

Rules for the description of '*Ca*. Phytoplasma'

Exceptional cases

b) Because of the highly conserved nature of the 16S rRNA gene, many biologically or ecologically distinct phytoplasma strains may warrant designation as new taxons, but these may fail to meet the requirement of sharing <97.5% sequence similarity with existing 'Ca. Phytoplasma', and may not be readily differentiated and classified.

- In that case, additional unique biological properties such as:
- antibody specificity,
- host range, and
- vector trans-mission specificity, as well as
- other molecular criteria (genes) need to be included for speciation (Seemüller and Schneider,2004).

IRPCM Phytoplasma taxonomy group,2004;Bertaccini and Duduk,2009

Phytoplasma nomenclature

Rules for the description of '*Ca*. Phytoplasma'

Exceptional cases

- For such cases, description of two different species is recommended only when all three of the following conditions apply:

i. The two phytoplasmas are transmitted by different vectors;

ii. The two phytoplasmas have a different natural plant host (or, at least, their behaviour is significantly different in the same plant host);

iii. There is evidence of significant molecular diversity, achieved by either hybridization to cloned DNA probes, serological reaction or PCR-based assay.

IRPCM Phytoplasma taxonomy group,2004

Phytoplasma nomenclature

Rules for the description of '*Ca*. Phytoplasma'

d) The rank of subspecies should not be used.

e) The reference strain should be made available to the scientific community from the authors of the *Candidatus* species description paper and it should be deposited (unless *in vitro* micropropagation proves impossible) in the micropropagated collection of Dr Assunta Bertaccini, DiSTA, Patologia Vegetale, Universita`di Bologna, Italy.

f) Manuscripts that describe a novel '*Ca.* Phytoplasma' species should preferably be submitted to the International Journal of Systematic and Evolutionary Microbiology (IJSEM).

g) The abbreviation for *Candidatus* is *Ca.* (e.g. '*Ca.* Phytoplasma japonicum' stands for '*Candidatus* Phytoplasma japonicum').

Minimal standards

Minimal standards for the description of new taxa

- Minimal standards: A series of tests or other features that are recommended to be included in a description when validly publishing a new taxon.
- These are drawn up from different groups of bacteria under the auspices (guidance) of the Judicial Commission and published in the IJSEM.

IUMS> ICSP> Judicial Commission> *ad hoc* Committees

Hawksworth,2010

Minimal standards

Minimal standards for the description of new taxa

- One of the functions of the Subcommittees on Taxonomy is to recommend to the ICSB (now ICSP) through the Judicial Commission minimal standards for the description of new taxa for the purpose of establishing validity of publication.
- According to Recommendation 30b, before publication of the name and description of a new species, the examination and description should conform at least to the **minimal standards** (if available) required for the relevant taxon of bacteria.
- According to the Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology, "Minimal characteristics should be provided and follow the guidelines set forth by various subcommittees of the ICSP.

Euzeby,2010

International Committee on Systematic Bacteriology

Subcommittee on the taxonomy of *Pseudomonas* and related organisms

Minutes of the meeting, 17 August 1999, Sydney, Australia

- **Minute 7. Minimal standards.**
- Minimal standards need to be proposed for the genus *Pseudomonas* and if possible also for other taxa.

Subcommittee on the taxonomy of *Pseudomonas*, and related organisms is changed into Subcommittee on the taxonomy of *Pseudomonas*, *Xanthomonas* and related organisms.

Minimal standards

Bacterial identification for publication: when is enough enough?

- All systems used to identify bacteria, whether phenotypic or genotypic, have limitations, because no single test methodology will provide results that are 100% accurate.
- Reliance on a single identification system, phenotypic or genotypic, to identify an organism provides more opportunity for misidentifying bacterial species.
- Ideally, identification of any taxon is based upon a polyphasic approach that includes a combination of:

1. Phenotypic testing methods (e.g., biochemical testing, cellular fatty acid analysis, and numerical analysis), and
2. Genotypic testing methods (e.g., DNA-DNA hybridization, analysis of G+C content [in moles percent], and 16S rDNA gene sequencing).

Janda and Abbott,2002; See also The function of *Ad hoc*

Minimal standards

Limitations of biochemical tests (traditional tube methods or kits)

- Biochemical properties do not accurately reflect the entire extent of the genomic complexity of a given species.
- Furthermore, phenotypic properties can be unstable at times and expression can be dependent upon changes in environmental conditions, e.g., growth substrate, temperature, and pH levels.
- The drawbacks with some commercial systems is their constant number of the tests.
- For instance, the tests included on the API 20E strip in 1975 were still the same tests on the strip in 2001. Yet the number of newly described taxa increased substantially between 1975 and 2001.
- One of the consequences of all of these limitations is that some commercial systems have great difficulty identifying certain groups of bacteria.

Janda and Abbott,2002

Minimal standards

Molecular biology and molecular techniques

- The use of molecular biology and molecular techniques as an aid to bacterial taxonomy and identification was in its infancy in the 1960s.

1. DNA base composition:

- Early studies using DNA base composition could clearly distinguish between genomes that were unrelated based upon differences in G+C content (in moles percent).

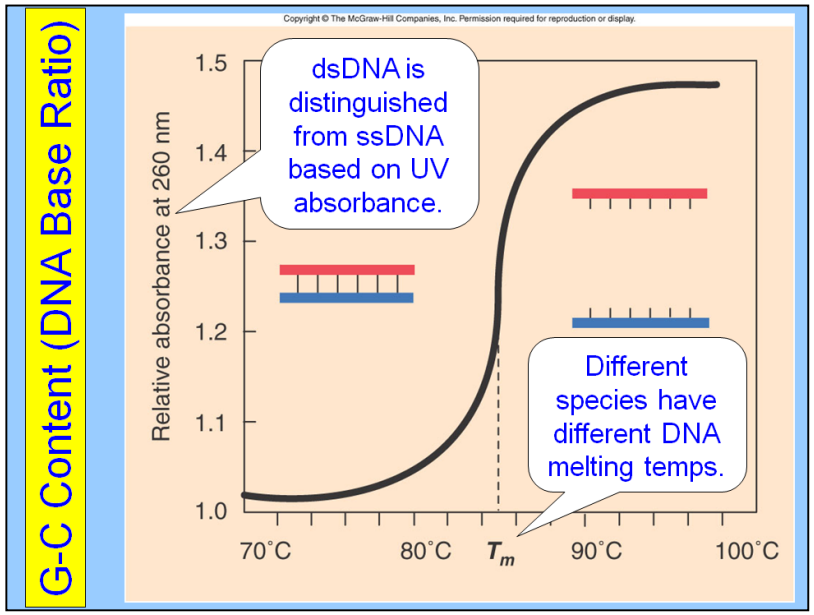
2. DNA-DNA hybridization:

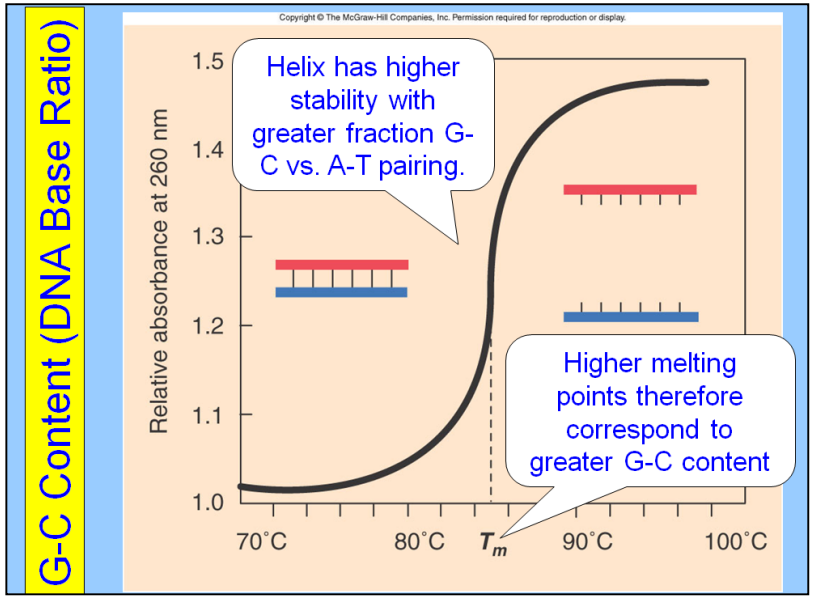
- Pioneered by Don Brenner and his colleagues (1977). It provides a quantitative definition of what constituted a species, ~70% or greater DNA-DNA relatedness with a T_m of 5°C or lower.

Janda and Abbott,2002

GC content and DNA denaturation

G+C content

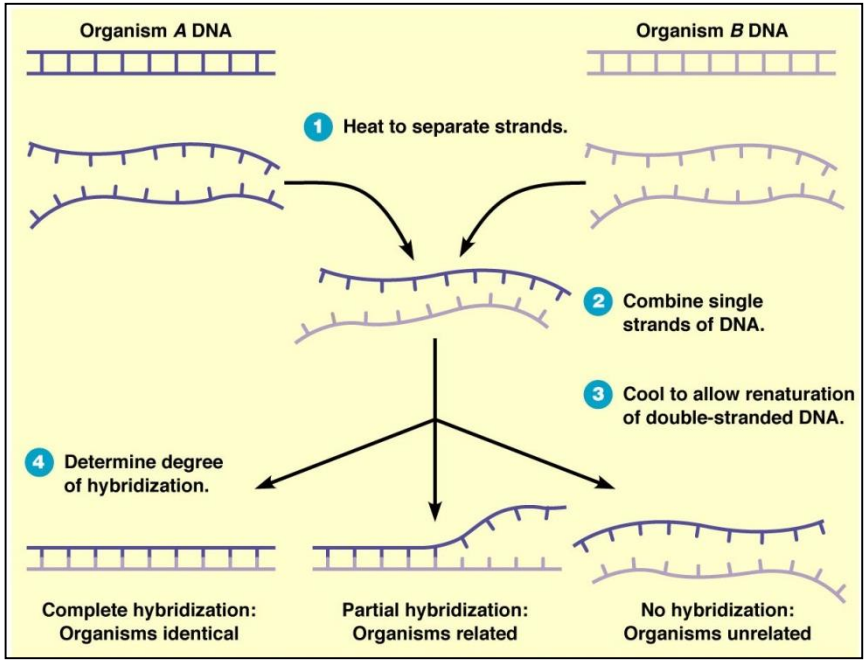




Abedon,2011

Nucleic acid hybridization
DNA/DNA hybridization

- In the 1970s: DNA-DNA hybridization was introduced.
- Isolates that showed >70% DNA homology were considered to belong to the same species.



Karlm,2004;..

Minimal standards

Molecular biology and molecular techniques

3. 16S rDNA gene sequencing:

- Sequencing of the small subunit (16S) rDNA genes is a powerful tool and by far the single most common molecular technique presently used for bacterial species identification.
- It is developed by the mid-1990s and is considered as a standard tool of microbial taxonomists not only for elucidating phylogenetic relatedness but also as a means of bacterial identification.
- Today, bacterial strains that defy (resist) identification by conventional commercial methodologies are often subjected to 16S rDNA sequence analysis so that a useful label can be placed on the isolate in question.

Janda and Abbott,2002

Minimal standards

Limitations of Molecular biology and molecular techniques

- **16S ribosomal RNA:**

1. ubiquitous in bacterial and archeal genomes;
2. matches to 98% of the 70% cut-off method;
3. High-throughput (output)method.

- **Limitations of DNA hybridization:**

1. you cannot distinguish between some phenotypically distinct species (e.g. *Bacillus thuringiensis* and *B. anthracis*);
2. organisms were the 'universal' primers do not fit are not detected;
3. pour method for resolving sub-populations within species.

Medini *et al.*,2008

Minimal standards

Limitations of Molecular biology and molecular techniques

- **Limitations of 16S rDNA gene sequencing:**
- Although this technique relies on sequencing of the DNA that encodes the 16S rRNA subunit, like phenotypic tests, it surveys only a small portion of the microbial genome.
- Since ribosomal genes are highly conserved, sequence variation between strains belonging to different species from distinct genera is less evident with rDNA gene sequencing than with DNA pairing studies, where DNA relatedness values are used to compare strains.
- It is generally accepted that an unidentified isolate whose rDNA gene sequence is <97% similar to those of the isolate's closest phylogenetic neighbors constitutes a new taxon.
- Many attempted were made to establish cutoff values for strain relatedness based upon sequencing of 16S rDNA.

Janda and Abbott,2002

Minimal standards

Limitations of DNA sequences deposited in microbial genome databases

- The accuracy of a partial or complete sequence of a 16S rDNA gene or of any other bacterial gene (s) deposited in GenBank or other databases is dependent upon how extensively the bacterial strain from which the sequences were derived has been characterized.
- Since biochemical characteristics are still the touchstone for bacterial identifications, the choice in the use of commercial versus conventional methodologies for the identification of strains is important.

Janda and Abbott,2002

Minimal standards

Limitations of DNA sequences deposited in microbial genome databases

- There are many more examples in the literature of the misidentification of bacteria by commercial systems.
- Thus, while the sequence data may be correct, if it is incorrectly associated with the wrong taxa, major errors in the literature can result.
- Sequence heterogeneity in the literature can result from strain misidentification.

Janda and Abbott,2002

Proposed guidelines for identification of bacterial species for publication purposes

Minimum standard for description of new species, case reports and publication studies

- Christensen and colleagues,2001 (Ref. 4) made a formal proposal in this area.
- They proposed that Recommendation 30b of the Bacteriological Code (1990 Revision) be revised so that proposals to recognize new species are based:

1. not upon a single strain, but
2. rather upon a minimum of 5 to 10 strains from
3. geographically and epidemiologically unrelated areas.

Janda and Abbott,2002

Proposed guidelines for identification of bacterial species for publication purposes

Minimum standard for description of new species, case reports and publication studies

- We agree with this proposal and believe that similar (and perhaps expanded guidelines) should be applied to case reports or a limited series of case reports involving unusual (rare) bacterial species or infrequent biotypes (genotypes) of established (traditional) pathogens.
- In addition, under these guidelines, proposals to recognize new species would require confirmation of the bacterial species or unusual phenotype or genotype by two independent laboratories (Table 1).

Janda and Abbott,2002

Proposed guidelines for identification of bacterial species for publication purposes

An alternative proposal that should be both technically and financially feasible and would help to reduce the number of publications with misidentifications

Type of study	Circumstances	Minimum requirements
Description of new species	Description of an organism and proposal of a new bacterial species	Identification based upon at least five strains (i) that have been demonstrated to be different by at least one molecular technique (e.g., pulsed-field gel electrophoresis, PFGE) and (ii) that are not related temporally, geographically, or epidemiologically.
Case report (single)	Isolation of a species identified as unusual or rare (e.g., unusual biotype of a common species or a common species associated with a new disease, disease syndrome, or anatomic site of isolation)	(i) Species or biotype identification confirmed by two independent laboratories (preferably, one of the two serves in a reference capacity); (ii) identification methods must differ (e.g., a commercial system and a traditional [tube] method) and must generate results indicating very good to excellent identification likelihood; and (iii) relevant phenotype(s) or genotype(s) used for identification and method(s) of detection (e.g., AP120E, septyl code, and identification probability [results showing excellent, very good, or good likelihood], and/or percent similarity or divergence from a published 16S rDNA gene sequence already in a database) must be reported.
Case report (series)	Isolation of a single agent or multiple agents, some or all of which are species that are uncommon or at least rarely identified in the clinical laboratory.	(i) For a single species, identification of at least two isolates confirmed by two independent methods; (ii) for multiple species, identification of at least one isolate of each species confirmed by two or more methods; and (iii) laboratory data indicating that isolates of the same species do not represent the same strain must be provided.
Population studies	Isolation of multiple strains belonging to multiple taxa (e.g., in epidemiological investigations or validation studies of identification systems).	Identification of each strain by at least a single universally available (i.e., commercial) method.

Janda and Abbott,2002

Minimal standards

Minimal requirements for a description of a new species

- Sneath (1977) suggested the minimum number for description of a new species is about 10 strains.

1. A group of bacteria derived from a single cell is called a strain;
2. Closely related strains constitute a bacterial species.

- If a prokaryote strain is considered as a new species, authors should at first explain why they consider this as a new species;
- They should carry out a thorough comparison between the "new" and the known species (and genera) of the bacterial group with which the new organism shows most similarities.

The Prokaryotes (chapter 1.4),2006

Minimal standards

Minimal requirements for a description of a new species

- If, as is usually the case, the new organism can be allotted to one of the major groups of prokaryotes, the description of a new species (and/or genus) should contain at least the information presented in the next Table.
- The requirements demanded and recommended by the International Code of Nomenclature of Bacteria (now prokaryotes) when a new species is described were discussed in "Nomenclature.

Minimum data required for the description of a new species

	Required data	Desired/required data, if applicable
Cell morphology	Cell shape[a] Cell size (diameter, length) Motility Visible internal or external structures[a] Formation of typical cellular aggregate[a] Occurrence of cell differentiation[a] Ultrastructure (general)[b]	Color Flagellation type[a] Spores[a], appendages[a], capsules[a], sheaths[a] Life cycle[a], heterocysts[a], hormogonia[a] Ultrastructure of flagella, envelope, cell wall[b]
Colonial morphology	Appearance of cell suspensions Appearance of colonies	Color of suspension (absorption spectra) Color of colonies Motility of colonies Formation of fruiting bodies[a] Formation of mycelia[a]
Staining behavior **Cell constituent**	Gram stain DNA base ratio Reserve materials	Acid-fast stain, spore stain, flagellum stain Nucleic acid homology; rRNA sequences Cellular pigments Cell wall and membrane constituents Typical enzymes
Physiology	Temperature range and optimum pH range and optimum Modes of energy metabolism (phototrophy, chemotrophy, lithotrophy, organotrophy) Relation to oxygen List of electron acceptors List of carbon sources List of nitrogen sources List of sulfur sources	Salinity or osmolarity requirements Vitamin requirements, Typical metabolic products formed (acids, osmolytes, pigments, antibiotics, toxins, antigens) Tolerances and susceptibilities
Ecology	Natural habitat(s)	Pathogenicity, host range Antigen formation Serology Phage susceptibility Symbiosis

[a]To be demonstrated by light microscopy.
[b]To be demonstrated by electron microscopy.

Hormogonia are motile filaments of cells formed by some cyanobacteria. Many multicellular cyanobacteria produce specialized nitrogen-fixing heterocysts.

The Prokaryotes (chapter 1.4),2006

Minimal standards

Revised minimal standards for description of new species of the class *Mollicutes*

- **The mandatory (law) requirements are:**

1. Deposition of the type strain into two recognized culture collections, preferably located in different countries;
2. Deposition of the 16S rRNA gene sequence into a public database, and a phylogenetic analysis of the relationships among the 16S rRNA gene sequences of the novel species and its neighbours;
3. Deposition of antiserum against the type strain into a recognized collection;
4. Demonstration, by using the combination of 16S rRNA gene sequence analyses, serological analyses and supplementary phenotypic data, that the type strain differs significantly from all previously named species; and
5. Assignment to an order, a family and a genus in the class, with an appropriate specific epithet.

Brown *et al.*,2006

Minimal standards

Revised minimal standards for description of new species of the class *Mollicutes*

- The publication of the description should appear in a journal having wide circulation.
- If the journal is not the *International Journal of Systematic and Evolutionary Microbiology*, copies of the publication must be submitted to that journal so that the name may be considered for inclusion in a Validation List as required by the *International Code of Bacteriological Nomenclature* (the *Bacteriological Code*).
- Updated informal descriptions of the class *Mollicutes* and some of its constituent higher taxa are available as supplementary material in IJSEM Online.

Brown *et al.*,2006

Description of new species

Tips for writing a taxonomic paper describing new species

- Anyone who is working on isolation of bacteria should have a good chance of finding a new species.
- Using genome sequencing, I would say that your chance of finding new species from one isolation study from soil or water should be almost 100%.
- It is our duty, as good microbiologists, to make them known to the scientific world by describing them as new species.

1. Assuming that you have isolated multiple strains, designate the type strain of your new species.
2. If you have a single strain, that strain will become the type strain of the new species.

EzBioCloud,2019

Species Description

Tips for writing a taxonomic paper describing new species (Continued)

- **Prove it is a new species by 16S rRNA gene (16S) and genome sequencing**
- Start with 16S sequencing and obtain a full-length, high-quality raw sequence.
- Don't trust the assembled sequence from sequencing firms, as these service companies almost never edit the final sequence.
- Always check the chromatograms (Sanger *ab1* files) manually. I have seen terrible 16S sequences used for creating a new species, which are not really new species.
- Search the sequence against the EzBioCloud 16S database at www.ezbiocloud.net.

EzBioCloud,2019

Species Description

Tips for writing a taxonomic paper describing new species (Continued)

- If you find any hit lower than 98.7%, you probably have a new species given that your sequence is of good quality.
- Let's say you have ~99% 16S similarity, don't be disappointed.
- You still have a very good chance of finding a new species.
- I suggest you go for whole genome sequencing. For the taxonomic purposes, a guideline was published here.
- Calculate Average Nucleotide Identity (ANI) values between your isolate and type strains of closely related species in the 16S searches using EzBioCloud database.
- The ANI cutoff is 95~96% for species.

EzBioCloud,2019

Species Description

Tips for writing a taxonomic paper describing new species (Continued)

- If you have indeed confirmed that you have a new species, try to obtain diagnostic phenotypic data. This can vary depending on the taxa.
- Obtain the general morphological and physiological characteristics.
- Obtain the differential characteristics that can be used for separating your new species from others.
- This data is usually given as a Table (e.g. see Table 1 of this paper as an example).
- Bergey's Manual is a great resource to obtain phenotypic data of related species.

Species Description

Tips for writing a taxonomic paper describing new species (Continued)

- **Please add the whole genome sequence of type strain in your study.**
- Like the 16S gene, the genome sequence is becoming the taxonomic basis for all disciplines of microbiology.
- If you do not sequence your new genome, you are leaving a hole in the system which must be filled by somebody else.
- If you have a case for a new species and lack the budget for genome sequencing, ChunLab would be happy to help you free of charge.
- Currently, Syst. Appl. Microbiol. and IJSEM require the genome sequence of the type strain at least for any new classification.

Species Description

Tips for writing a taxonomic paper describing new species (Continued)

- **Deposit the type strains to two culture collections in two different countries.**
- To find suitable culture collections, use the WFCC site.
- My recommendation is to find culture collections that at least have an English website.
- Make sure that you get the certificate of the deposition from two culture collections in two countries.
- If they don't know what the certificate of the deposition is, don't bother and try to deposit to other culture collections.
- This process takes several months, so start as soon as you find that you have a new species.

Species Description

Tips for writing a taxonomic paper describing new species (Continued)

- **Write the taxonomic proposal as a manuscript.**
- As with your usual content in any paper, a proper etymology and description should be given (see below as an example).

Description of *Halostagnicola bangensis* sp. nov.

Halostagnicola bangensis (bang.en'sis. N.L. fem. adj. *bangensis* pertaining to Bange, Tibetan soda lake in China, from which the strain was isolated).

Cells are motile and pleomorphic, with coccoid to bacilar shapes. The size of the single cells is 1.0–1.2 × 2.0–4.0 μm and they grow separately without forming groups. Gram-stain-variable; in young cultures most cells are Gram-stain-negative, while few cells are observed as Gram-stain-positive. Colonies on solid medium after incubation at 37 °C for 30 days are circular, regular edged, smooth, convex and

⋮

The type strain is T26^T (=CECT 8219^T=IBRC-M 10759^T= JCM 18750^T), isolated from sediment of the soda saline lake Bange located in Tibet, China. The genomic DNA G+C content of the type strain is 60.1 mol% (T_m).

EzBioCloud,2019

Species Description

Tips for writing a taxonomic paper describing new species (Continued)

- **Submit the manuscript to IJSEM with two certificates of deposition of the type strain.**
- It is OK to submit to other journals as long as they publish the papers in English.
- If it is published outside of IJSEM, please make sure that you submit the pdf of the publication plus the two certificates of deposition to IJSEM for validation.
- Without validation, your new species will not be properly acknowledged by other scientists. In fact, in many cases, it adds more confusion than good.

Species Description

Tips for writing a taxonomic paper describing new species (Continued)

- **Resources for up-to-date nomenclature**
- Please note that only the Approved Lists of Bacterial Names of 1980, and the Notification and Validation Lists in the IJSB and IJSEM represent the official nomenclature.
- However, IJSEM or other officially recognized bodies do not provide an organized, up-to-date list of validly published names at present.
- Fortunately, the following resources are maintained properly and are therefore useful. Please bear in mind that these websites may contain human errors.

1. EzBioCloud.net
2. List of Prokaryotic names with Standing in Nomenclature.

EzBioCloud,2019

Species Description

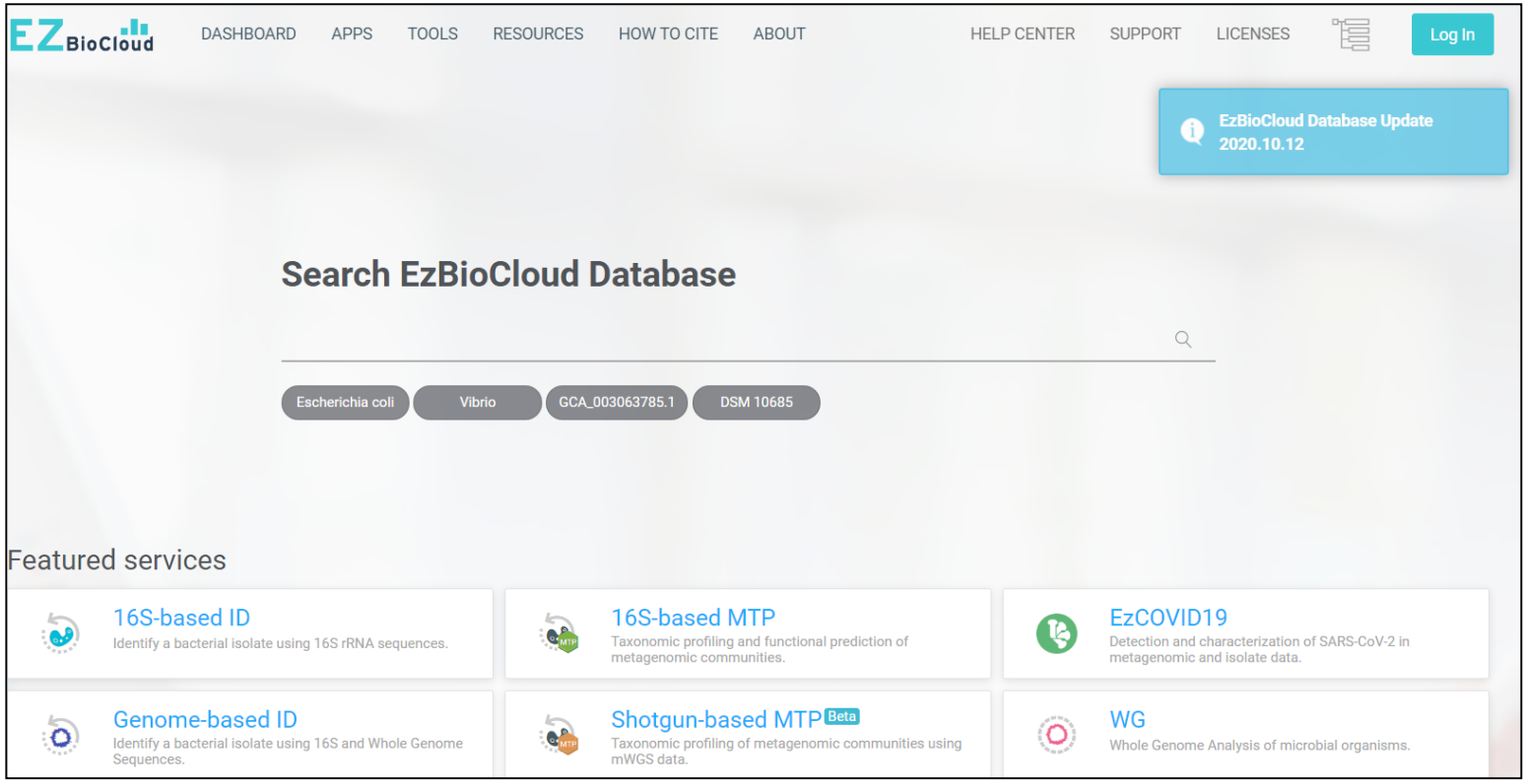
Tips for writing a taxonomic paper describing new species (Continued)

- EzBioCloud.net
- Jon Jongsik Chun, Ph.D.
- Jon is a scientist & entrepreneur dedicated to developing bioinformatics related to bacterial systematics, genomics, and microbiome.
- He is a professor at Seoul National Univ. and founder of ChunLab, Inc.
- He is best known as a creator of EzBioCloud (formerly EzTaxon) database, and recipient of the Bergey Award 2018.

EzBioCloud,2019

Species Description

Tips for writing a taxonomic paper describing new species (Continued)



EzBioCloud,2021

The Code

Binomial nomenclature

The Rules

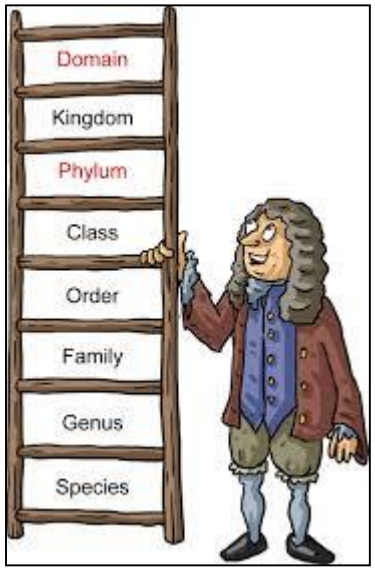
- The scientific name, or more specifically the species name, is a binomial (**two names**).
- The bacteriological code also uses the binomial system.
- The specific epithet is unique, having only one correct name each taxon above species, from genus up to and including order can only have one correct name, which is the earliest name published in accord with the Code (Rule 23a).
- The correct name of a species is a binary combination of a generic name and species epithet (Genus name+species name), applied in a particular classification.
- The term 'epithet' is applied specifically to species names (Principal 8, Note 1).

Young,2008;...

Binomial nomenclature

The Linnaean Binomial Nomenclature

- The Swedish naturalist, Carolus Linneaus developed a scientific system of naming organisms.



Carolus Linnaeus (1707-1778)

Binomial nomenclature

Some more Rules

- According to scientific nomenclature, each organism is assigned two names, or a binomial:
- A genus and a specific epithet, or species
- Examples:
- *Erwinia amylovora*

Binomial nomenclature

- He assigned two Latinized names to each organism:
- A genus consists of a group of similar species.
- Similar genera are grouped into a family.
- The species name or "specific epithet" is unique to the new species.
- The genus name is indicated by a capital letter whereas, the species name starts with a lower case letter.
- By convention both names are italicized (or underlined).
- Example: *Streptomyces scabies* - Once a scientific name has been used in entirety it can subsequently be abbreviated as follows:
- *S. scabies*

Binomial nomenclature

Italic type (or underlining) *versus* roman type

- According to Chapter 4 (Advisory Notes) of the *Bacteriological Code* (1990 Revision), scientific names of taxa should be preferably indicated by a different type face, e.g., italic or by some other device to distinguish them from the rest of the text.
- The *Bacteriological Code* (1990 Revision) sets no binding standard in this respect, as typography is a matter of editorial style and tradition not of nomenclature.

1. The name of genera, species, and subspecies are generally printed in italics (or underlined), but
2. for higher categories conventions vary: in Britain they are often in ordinary roman type (e.g.Times New), but in America or in France they are often in italics.

Euzeby,2010

Binomial nomenclature

- The first name beginning with a capital indicates the genus, e.g. *Xanthomonas*, the second name the species epithet.
- e.g. *citri* (no capital).
- Species name is never abbreviated.
- When the exact name has to be mentioned, the author (s) giving the first description of the bacterium with the prefix 'ex', and those who made the last taxonomic (re-).
- e.g. *X. gardneri* (ex Sutic 1957) Jones *et al*.,2006, nom rev., comb. nov.

Binomial nomenclature

Relative to habit, habitat, geography, growth, morphology

- Scientific names should be unique, unchanging and descriptive (i.e. relative to habit, habitat, geography, growth, morphology).
- **Examples of genus/species names:**

1. The name of the person describing the organism (*Erwinia amylovora*, the name in honor of Erwin Smith).
2. The habitat of the organism(the disease on citrus).
3. The appearance of the organism (X*antho-* produces a yellow pigment).
4. Some names may reflect a disease or infectious process caused by an organism (e.g., *scabies* describes the ability to produce scab).

Binomial nomenclature

Sphingomonas sp.

Scientific Binomial	Source of Genus Name Scientific	Source of Specific Epithet
Klebsiella pneumoniae	**Honors Edwin Klebs *Klebsiella***	**The disease**
Erwinia amylovora	**Name in honour of Erwin Smith *Erwinia***	**The disease**
Pectobacterium carotovorum	**A pectolytic bacterium**	**carrot-devouring**
Streptomyces scabiei	**Chains of cells (*strepto-*)**	**Causes scab**
Xanthomonas citri	**Produces a yellow pigment (X*antho-*)**	**The disease on citrus**
Sphingomonas melonis	***Sphingos* of sphinx** A mythical creature with the head of a human and the body of a lion	**Causal agent of brown spot of melon fruits**

Klebsiella is a type of Gram-negative bacteria that can cause different types of healthcare-associated infections, including *pneumonia*, bloodstream infections, wound or surgical site infections, and meningitis.

Binomial nomenclature

Relative to habit, habitat, geography, growth, morphology, discoverer or a famous person in the field of microbiology

- List of bacterial genera named after geographical names:

1. Aegyptianella: Aegyptus (the Latin name of Egypt)
2. Indibacter: India
3. Koreibacter: Korea

- List of bacterial genera named after personal names:

1. Erwinia: Erwin Frink Smith, an American bacteriologist.
2. Euzebya: Jean P. Euzéby, a French bacteriologist.
3. Euzebylla: Jean P. Euzéby, a French bacteriologist.
4. Ewingella: William H. Ewing, an American bacteriologist.
5. Sneathia: P. H. A. Sneath, a British bacteriologist.

Wikipedia,2013

Binomial nomenclature

Valid names

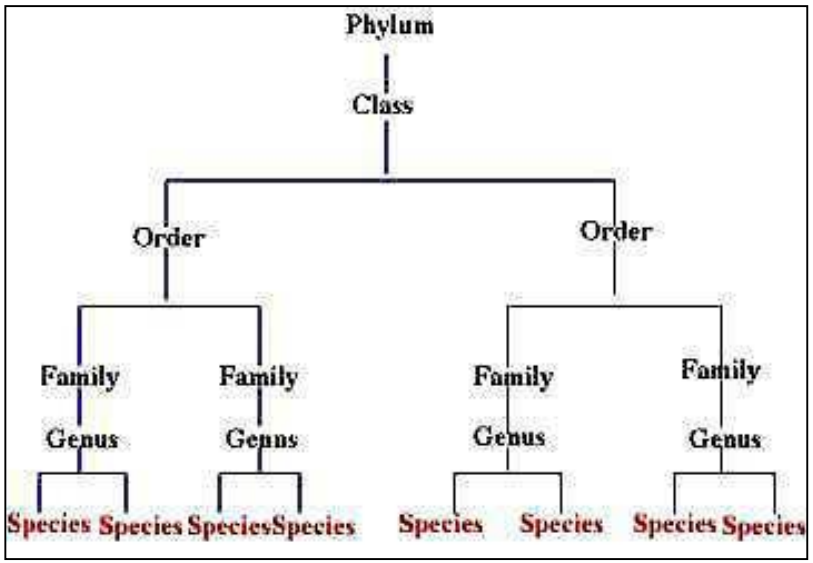
- The epithet *paradisiaca*, first published as *Erwinia paradisiaca* (Brown-black root rot of banana) in the Approved Lists.
- Reclassified in *Brenneria paradisiaca* based solely on a comparative analysis of 16S rDNA sequences.
- And then recently as *Dickeya paradisiaca* based on further comparative molecular analyses.
- Bacteriologists must therefore choose binomials from validly published names.

Young,2008

Linnaean hiearchy

The ranking of taxonomic categories

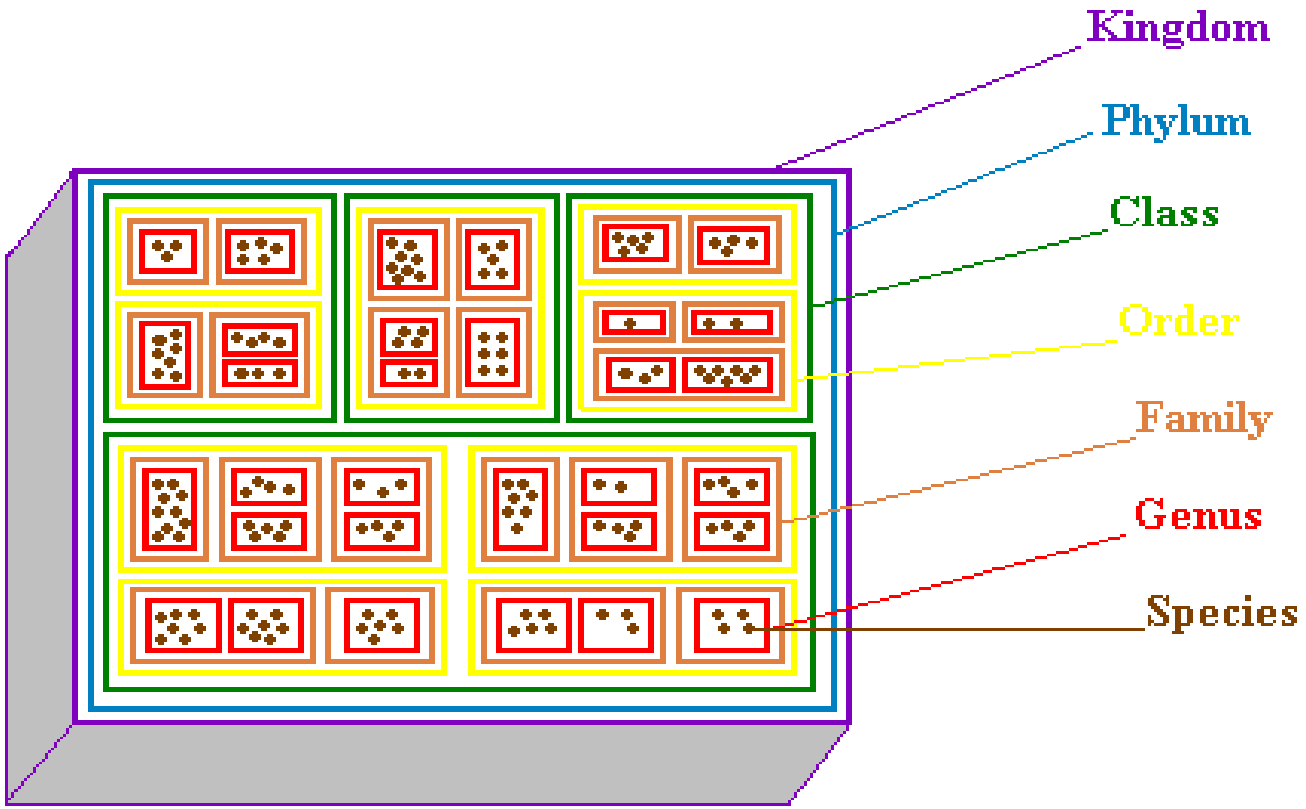
- Taxon: A group or "level" of classification.
- Hierarchies: Broad divisions are divided up into smaller divisions.
- Bacterial classification follows Linnaean hierarchical system that starts with a few categories at the highest level, and further subdivides them at each lower level.
- Species is basic taxonomic unit.
- Subspecies is the lowest phylogenetically supported taxonomic rank that has official standing in nomenclature.



Taxon, plural Taxa, any unit used in the science of biological classification, or taxonomy. Names of taxa above the rank of species are single words.

Linnaean system

The ranking of taxonomic categories



Linnaean hiearchy

Levels of Classification

- Domain: The highest of taxonomic rank ('80s)
- Kingdom (Not used by most bacteriologists),1969
- Phylum or division of the kingdom
- Class
- Order
- Family (related genera)
- Genus (related species) plural: Genera
- Species (related strains) both singular & plural
- Subspecies

Note: Division rank is equivalent to phyla.

It has been estimated that ~1,300 bacterial phyla exist. As of May 2020, 41 bacterial phyla are formally accepted by the LPSN, 89 bacterial phyla are recognized on the Silva database, dozens more have been proposed, and hundreds likely remain to be discovered.

Linnaean hierarchy

Example:

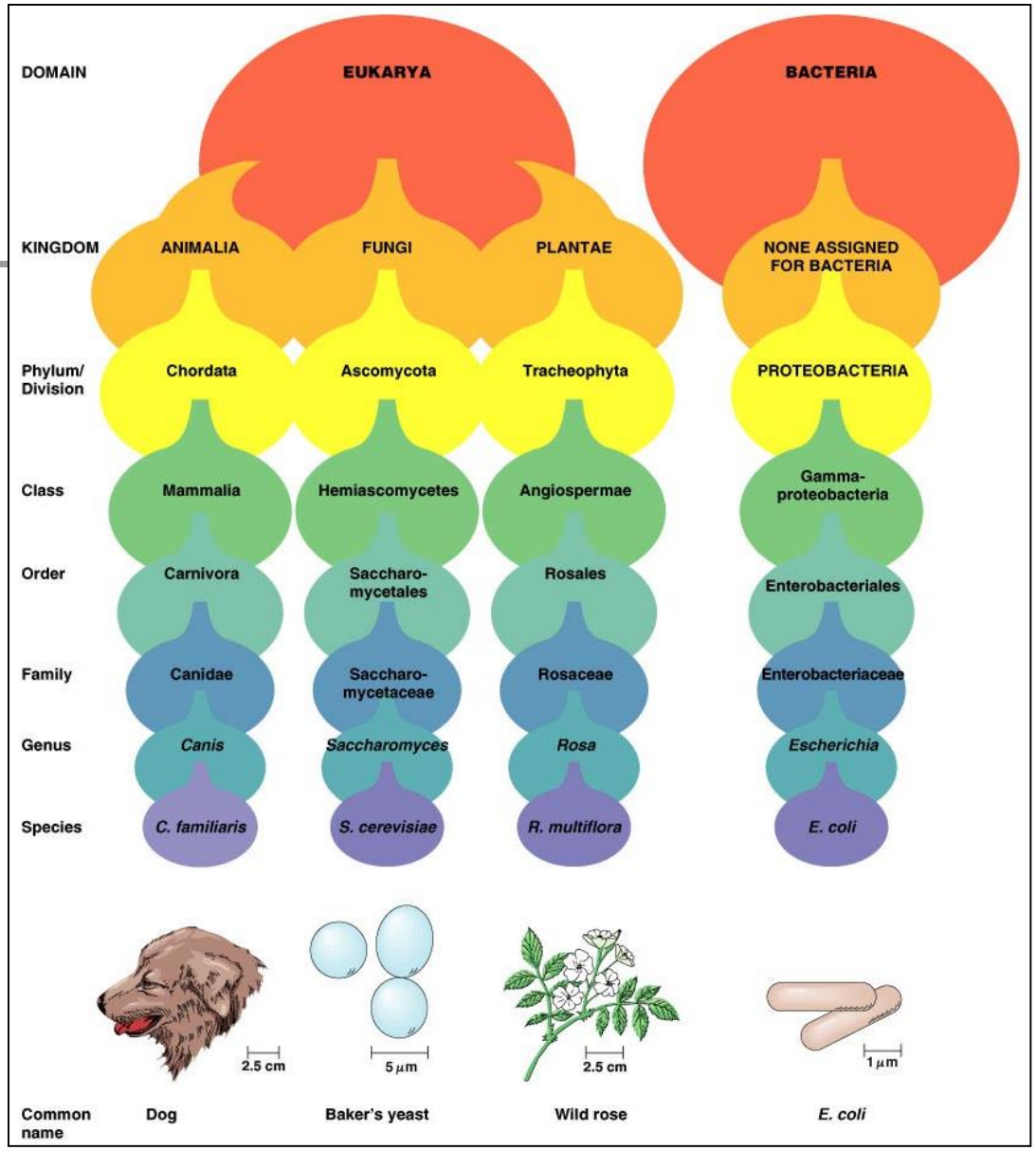
Division: Gracilicutes

Class: Scotobacteria

Family: Enterobacteriaceae

Genus: *Erwinia*

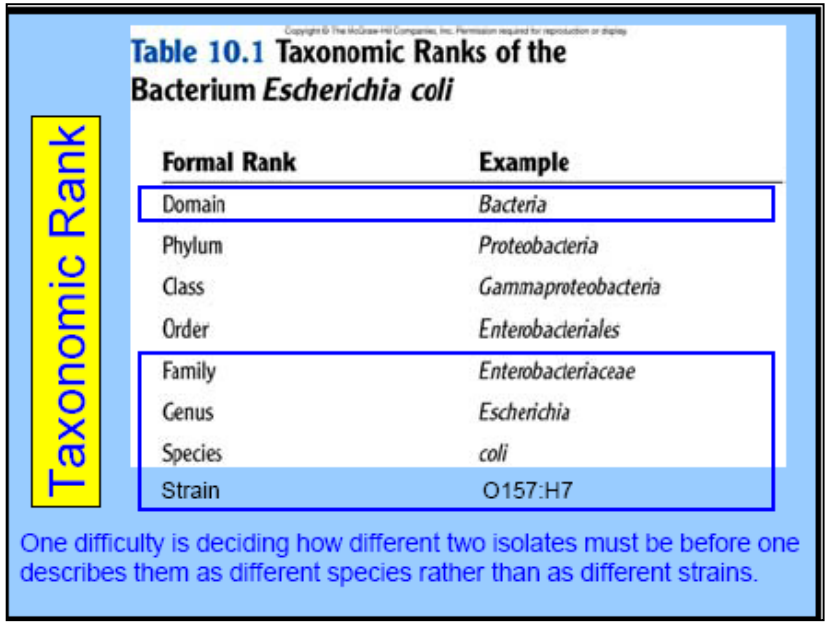
Species: *amylovora*



Karlm,2004

Nomenclature of strains

- In microbiology, the strain designation is often related to:
1. Pathogenicity (Pathogenic *E. coli* like O157:H7);
2. Antagonistic activitity (*A. radiobacter*, K84);
- But otherwise these are very similar to other related strains.
- Also strain may be designated in any manner, e.g.
1. by the name of an individual,
2. by a locality, or
3. by a number.



Abedon,2011

Nomenclature of strains

P. syringae pv. *tomato* strain T1 vs. Pst DC3000

- A comparison of *P. syringae* pv. *tomato* strain T1 against strain DC3000 found that the genomes were highly similar except for the T3SE repertoires (stock of skills), which had diverged significantly.

1. Pst DC3000 is a pathogen of both tomato and *Arabidopsis thaliana*, whereas
2. Pst T1 is pathogenic on tomato but not on *A. thaliana*.

Arnold and Jackson,2011

Nomenclature of Infrasubspecific Taxa
Infrasubspecific designation

- Taxa below the rank of subspecies (infrasubspecific subdivisions) are not covered by the Rules of the *Bacteriological Code* (1990 Revision) [Rules 5d and 14a].
- The preferred names of infrasubspecific subdivisions are:
- Biovar (usual abbreviation: bv.),
- chemoform, chemovar,
- cultivar (usual abbreviation: cv.),
- morphovar,
- pathovar (usual abbreviation: pv.),
- phagovar,
- serovar.

- Chemovar, chemotype and chemoform: An infrasubspecific category for bacteria characterized on the basis of its chemical constitution.
- Cultivar: Cultivated variety of plant from which sequence was obtained.

Definition of Infra-subspecific ranks

- The genetic variability of microbes is further subdivided into subspecies or types:
- **Serovars** are antigenically distinct organisms.
  - For example, over 2,000 serovars of *Salmonella* have been identified which are typed according to their flagella (H),capsule (K) and lipopolysaccharide (O)antigens.
  - The antibodies are generally directed against O-serogroup1 of *D. chrysanthemi*, recognizing only 68% of the strains (Samson *et al.*,1990).
- **Biovars** are organisms which can differ biochemically. e.g. they may possess differing forms of enzymes.
- **Pathovar**: Pathogenic properties for certain hosts.
- **Phagovar**: Ability to be lysed by certain bacteriophages.

Nomenclature of Infrasubspecific Taxa
Infrasubspecific designation
Phagovar

- An infraspecific classification in bacteria based on sensibility against bacteriophages:
- An infrasubspecific taxon is designated or cited by the name of the species followed by the infrasubspecific term used to designate this infrasubspecific subdivision followed by the infrasubspecific designation.
- Example:
- *Staphylococcus aureus* phagovar 81
- *Xanthomonas citri* phagovars Cp1, Cp2 and Cp3
- *Pseudomonas syringae* pv.*morsprunorum* phagovar A15

Suffixes

The suffix "-bacteria" for phyla

56 phyla in the domain "Bacteria"

- As of September 2012, there are 30 phyla in the domain "Bacteria" accepted by LPSN (now LPSN-DSMZ's website, https://lpsn.dsmz.de).
- There are no fixed rules to the nomenclature of bacterial phyla.
- It was proposed that the suffix"-bacteria" be used for phyla, but generally the name of the phylum is generally the plural of the type genus.
- E.g. *Actinobacteria* from *Actinomyces*.
- Exception:
- the *Firmicutes*, *Cyanobacteria*, and *Proteobacteria*, whose names do not stem from a genus name.

Wikipedia,2014; Euzeby,2020

Suffixes

The name of a taxon between genus and class is formed by the addition of an appropriate suffix to the stem of the name of the type genus

- Nomenclature: Bacterial names may originate from any language, but they must be given a Latin suffix (i.e., they Latinized).

Taxonomic rank	Suffixe	Example
Subtribe	-*inae*	
Tribe*	-*eae*	
Subfamily	-*oideae*	
Family	-*aceae*	Enterobacteriaceae
Suborder	-*ineae*	
Order	-*ales*	Pseudomonadales
Subclass	Proposed suffix -*idae*	
Class	Proposed suffix -*ia*	Actinobacteria
Division or phylum**	-	
Domain or empire**	-	

*Taxonomic category not in current use; ** Taxonomic category not covered by the Rules.

Domain Bacteria

Bacterial phylum

The bacterial phyla are the major lineages (phyla or divisions) of the domain Bacteria

- "*Abditibacteriota*"
- "*Acidobacteria*"
- "*Actinobacteria*"
- "*Candidatus* Aminicenantes"
- "*Aquificae*"
- "*Armatimonadetes*"
- "*Bacteroidetes*"
- "*Balneolaeota*"
- "*Caldiserica*"
- "*Calditrichaeota*"
- "*Chlamydiae*"
- "*Chlorobi*"
- "*Chloroflexi*"
- "*Chrysiogenetes*"
- "*Candidatus* Cloacimonetes"
- "*Coprothermobacterota*"
- "*Candidatus* Cryosericota"
- "*Cyanobacteria*"
- "*Deferribacteres*"
- "*Deinococcus-Thermus*"
- "*Candidatus* Dependentiae"
- "*Dictyoglomi*"
- "*Elusimicrobia*"
- "*Candidatus* Eremiobacteraeota"
- "*Candidatus* Fermentibacteria"
- "*Fibrobacteres*"
- "*Firmicutes*"
- "*Fusobacteria*"
- "*Fusobacteria*"
- "*Gemmatimonadetes*"
- "*Candidatus* Goldbacteria"
- "*Candidatus* Kapabacteria"
- "*Kiritimatiellaeota*"
- "*Candidatus* Krumholzibacteriota"
- "*Lentisphaerae*"
- "*Candidatus* Margulisbacteria"
- "*Candidatus* Mcinerneyibacteriota"
- "*Candidatus* Melainabacteria"
- "*Candidatus* Microgenomates"
- "*Nitrospinae*"
- "*Nitrospirae*"
- "*Candidatus* Omnitrophica"
- "*Candidatus* Parcubacteria"
- "*Candidatus* Parcunitrobacteria"
- "*Candidatus* Peregrinibacteria"
- "*Planctomycetes*"
- "*Proteobacteria*"
- "*Rhodothermaeota*"
- "*Spirochaetes*"
- "*Candidatus* Sumerlaeota"
- "*Synergistetes*"
- "*Tenericutes*"
- "*Thermodesulfobacteria*"
- "*Thermomicrobia*"
- "*Thermotogae*"
- "*Verrucomicrobia*"

Euzeby,2020

Taxa above the rank of class

Taxonomic categories not covered by the Rules of the Bacteriological Code

The higher taxa proposed by Cavalier-Smith are generally disregarded by the molecular phylogeny community (e.g. Pace, N. R. (2009). "Mapping the tree Life: Progress and Prospect". Microbiology and Molecular Biology Reviews 73 (4): 565-576.

- ***Actinobacteria*** Cavalier-Smith 2002, subdivisio nov.
- ***Aquificae*** Reysenbach 2001, phyl. nov.
- ***Archaebacteria*** Cavalier-Smith 2002, divisio nov.
- ***Bacteria*** Cavalier-Smith 2002, regnum nov.
- ***Chloroflexi*** Garrity and Holt 2001, phyl. nov.
- ***Chrysiogenetes*** Garrity and Holt 2001, phyl. nov.
- ***Crenarchaeota*** Garrity and Holt 2001, phyl. nov.
- ***Crenarchaeota*** Cavalier-Smith 2002, subdivisio nov.
- ***Cyanobacteria*** (*ex* Stanier 1974) Cavalier-Smith 2002, divisio nov., nom. rev.
- ***Deferribacteres*** Garrity and Holt 2001, phyl. nov.
- ***Endobacteria*** Cavalier-Smith 2002, subdivisio nov.
- ***Eobacteria*** Cavalier-Smith 2002, divisio nov.
- ***Eobacteria*** Cavalier-Smith 2002, infraregnum nov.
- ***Euryarchaeota*** Garrity and Holt 2001, phyl. nov.
- ***Euryarchaeota*** Cavalier-Smith 2002, subdivisio nov.
- ***Eurythermea*** Cavalier-Smith 2002, superclassis nov.
- ***Exoflagellata*** Cavalier-Smith 2002, superdivisio nov.
- ***Firmicutes*** corrig. Gibbons and Murray 1978, *divisio*.
- ***Gemmatimonadetes*** Zhang *et al*. 2003, phyl. nov.
- ***Geobacteria*** Cavalier-Smith 2002, subdivisio nov.
- ***Glycobacteria*** Cavalier-Smith 2002, infraregnum nov.
- ***Gracilicutes*** Gibbons and Murray 1978, *divisio*.
- ***Lentisphaerae*** Cho *et al*. 2004, phyl. nov.
- ***Mendosicutes*** Murray 1984, divisio nov.
- ***Negibacteria*** Cavalier-Smith 2002, subregnum nov.
- ***Neobacteria*** Cavalier-Smith 2002, superclassis nov.
- ***Phycobacteria*** Cavalier-Smith 2002, subdivisio nov.
- ***Planctobacteria*** Cavalier-Smith 2002, divisio nov.
- ***Posibacteria*** Cavalier-Smith 2002, divisio nov.
- ***Proteobacteria*** Cavalier-Smith 2002, divisio nov.
- ***Rhodobacteria*** Cavalier-Smith 2002, subdivisio nov.
- ***Sphingobacteria*** Cavalier-Smith 2002, divisio nov.
- ***Spirochaetae*** Cavalier-Smith 2002, divisio nov.
- ***Tenericutes*** Murray 1984, divisio nov.
- ***Thermodesulfobacteria*** Garrity and Holt 2001, phyl. nov.
- ***Thermomicrobia*** Garrity and Holt 2001, phyl. nov.
- ***Thermotogae*** Reysenbach 2001, phyl. nov.
- ***Thiobacteria*** Cavalier-Smith 2002, subdivisio nov.
- ***Unibacteria*** Cavalier-Smith 2002, subregnum nov.

Domain Bacteria

Bacterial phyla

Bacteroidetes

- Phylum: Bacteroidetes
- Class: Bacteroidetes
- Order: Bacteroidales
- Family: Bacteroidaceae
- Genus: Bacteroides
- Species: fragilis
- The phylum Bacteroidetes is composed of three large classes of Gram-negative, nonsporeforming, anaerobic or aerobic, and rod-shaped bacteria that are widely distributed in the environment, including in soil, sediments, and sea water, as well as in the guts and on the skin of animals.

Bacterial nomenclature

All useful information on bacterial or general nomenclature

1. Bergey's Manual of Determinative Bacteriology, now named Bergey's Manual of Systematic Bacteriology (Boone *et al*.,2001) is the standard reference on bacterial classification.
2. CCUG, Sweden
3. Species 2000 CheckList
4. List of Prokaryotic names with Standing in Nomenclature (LPSN) creator by J.P. Euzéby and now maintained by Aidan C. Parte. This website was integrated into the DSMZ's Prokaryotic Nomenclature Up-to-date (PNU) database with new email address: https://lpsn.dsmz.de.

Bacterial nomenclature

All useful information on bacterial or general nomenclature

Bergey's Manual of Determinative Bacteriology Provides identification schemes for identifying bacteria and archaea.	Morphology, differential staining, biochemical tests
Bergey's Manual of Systematic Bacteriology Provides phylogenetic information on bacteria and archaea.	Based on rRNA sequencing
Approved Lists of Bacterial Names Lists species of known prokaryotes. Appear in the ISPP website	Based on published articles

Bacterial nomenclature
Bergey's Manuals

- The "bible" of bacterial identification is a book called Bergey's Manual of Determinative Bacteriology, based on morphology, differential staining, biochemical tests.
- He was the first doctor to isolate the bacterium *Actinomyces* from a human being, in 1907.
- Bergey is long dead, but the American Society for Microbiology continues to revise and publish the book.

David Hendricks Bergey

Born	December 27, 1860, Pennsylvania
Died	September 5, 1937 (aged 76) Philadelphia, Pennsylvania
Alma mater	University of Pennsylvania
Fields	Bacteriology
Institutions	University of Pennsylvania

Alma mater is an allegorical Latin phrase for a university, school, or college that one formerly attended.

Bacterial nomenclature

Bergey's Manuals

- David Hendricks Bergey together with the Society of American Bacteriologist developed a single scheme to cover all the described bacteria.

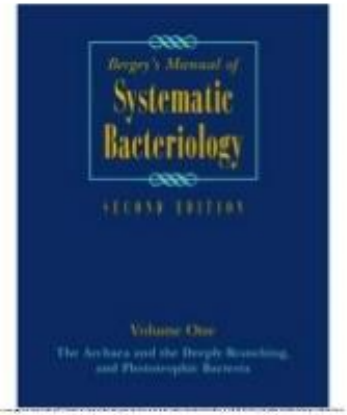
David Hendricks Bergey

Editions

Bergey's Manual of Determinative Bacteriology

***Bergey's Manual of Systematic Bacteriology*, 1st ed.**

***Bergey's Manual of Systematic Bacteriology*, 2nd ed.**



Bergey's Manual of Determinative Bacteriology
2nd-8th Editions (1923-1974)

- Classifies bacteria by cell wall composition, morphology, biochemical tests, differential staining, etc.
- The classification in Bergey's Manual is accepted by most microbiologists as the best consensus for prokaryotic taxonomy.
- Chester's Manual of Determinative Bacteriology,1901 (long out of date).
- Bergey's Manual of Determinative Bacteriology (ed.1,1923);
- Bergey's Manual of Determinative Bacteriology (ed. 2,1925);
- Bergey's Manual of Determinative Bacteriology (ed. 3,1930);
- Bergey's Manual of Determinative Bacteriology (ed. 4,1936);
- Bergey's Manual of Determinative Bacteriology (ed. 5,1939);
- Bergey's Manual of Determinative Bacteriology (ed. 6,1948);
- Bergey's Manual of Determinative Bacteriology (ed. 7,1957);
- Bergey's Manual of Determinative Bacteriology (ed. 8,1974).

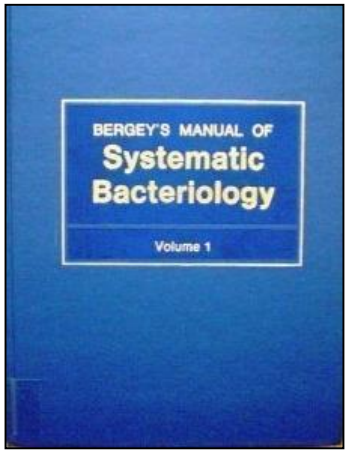
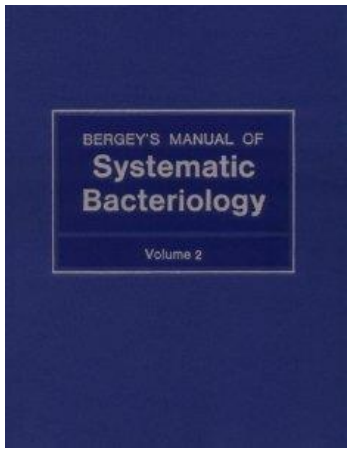
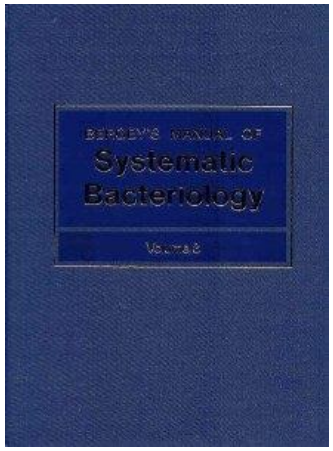
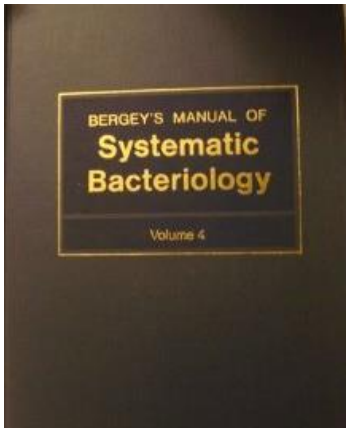
Systems of classification were based entirely on morphologic characters.

Bergey's Manual of Systematic Bacteriology (Vol.1-4)
1st edition(1984-1989)

- Classifies bacteria via evolutionary or genetic relationships.
- John G. Holt, Editor-in-Chief
- Williams & Wilkins, Baltimore, MD
- **Published in 4 volumes:**
- **Volume 1** (1984)
  - Gram-negative Bacteria of general, medical, or industrial importance
- **Volume 2** (1986)
  - Gram-positive Bacteria other than Actinomycetes
- **Volume 3** (1989)
  - Archaeobacteria, Cyanobacteria, and remaining Gram-negative Bacteria
- **Volume 4** (1989)
  - Actinomycetes

Bergey's Manual of Systematic Bacteriology (Vol.1-4)

1st edition (1984-1989)

			
Volume 1 (1984) 964 pages	**Volume 2 (Aug.1986)** 672 pages	**Volume 3 (March 1989)** 744 pages	**Volume 4 (Dec.1989)** 440 pages

In these volumes many higher taxa are defined in terms of phenotype. Mix Phylogenetic/Phenetic-5 Kingdoms.

Bergey's Manual of Systematic Bacteriology(Vol.1-5)

2nd Edition(2001-2012)

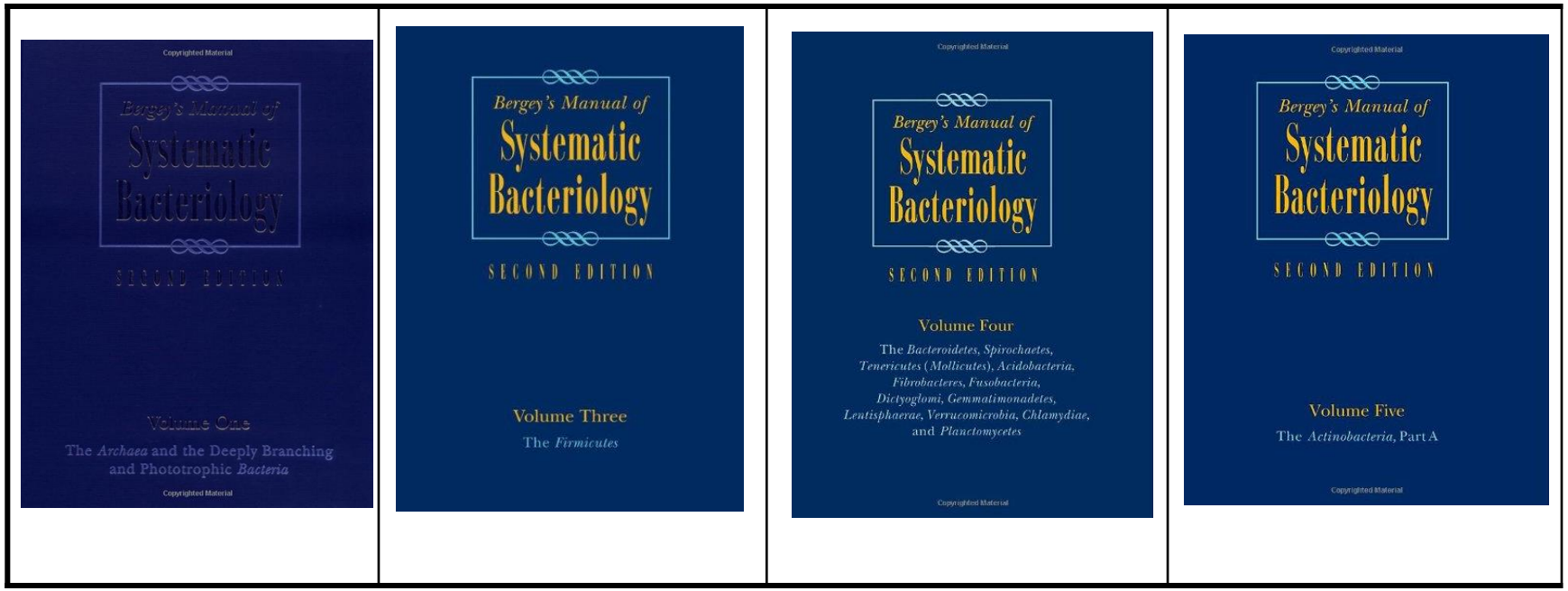
- The second edition is being published in 5 volumes.
- The current volumes differ drastically from previous volumes in that many higher taxa are not defined in terms of phenotype, but solely on 16S phylogeny.
- Based on Phylogenetic-3 Domains.
- Yet retains much of the layout of the first edition.

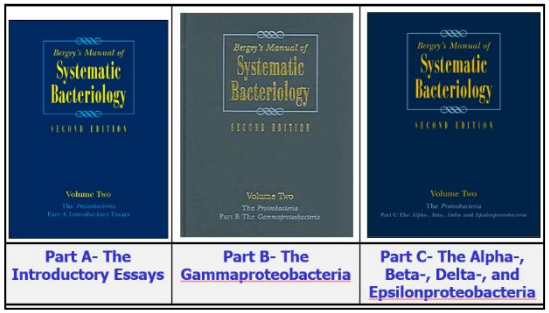
Bergey's Manual of Systematic Bacteriology (Vol.1-5)
2nd Edition (2001-2012)

- **The second edition is being published in 5 volumes.**
- **Volume 1** (2001)
  - The *Archaea* and the deeply branching and phototrophic *Bacteria*
  - **Editor-in-Chief:** George M. Garrity
  - **Editors:** David R. Boone and Richard W. Castenholz. 721 pp.
- **Volume 2** (2005) **divided into three books (Parts A, B & C)**
  The *Proteobacteria*
  - **Editor-in-Chief:** George M. Garrity
  - **Editors:** Don J. Brenner, Noel R. Krieg and James T. Staley. 2816 pp.
- **Volume 3** (2009)
  - The *Firmicutes*
  - **Editors:** Paul De Vos, George Garrity, Dorothy Jones, Noel R. Krieg, Wolfgang Ludwig, Fred A. Rainey, Karl-Heinz Schleifer and William B. Whitman. 1450 pp.
- **Volume 4** (2010)
  - The *Bacteroidetes*, *Planctomycetes*, *Chlamydiae*, *Spirochaetes*, *Fibrobacteres*, *Fusobacteria*, *Acidobacteria*, *Verrucomicrobia*, *Dictyoglomi* and *Gemmatimonadetes*
  - **Editors:** Noel R. Krieg, James T. Staley, Brian Hedlund, Bruce J. Paster, Naomi Ward, Wolfgang Ludwig and William B. Whitman. 974 pp.
- **Volume 5** (2012)
  - The *Actinobacteria*
  - **Editors:** Michael Goodfellow, Peter Kämpfer, Hans-Jürgen Busse, Martha Trujillo, Ken-ichiro Suzuki, Wolfgang Ludwig and William B. Whitman. 1750 pp.

Bergey's Manual of Systematic Bacteriology (Vol.1-5)

2nd Edition (2001-2012)





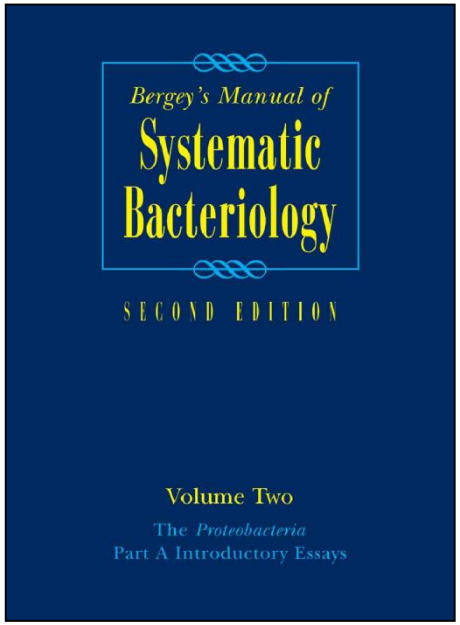
Bergey's Manual of Systematic Bacteriology 2nd Edition, Volume 2 (Parts A, B & C)

The Proteobacteria

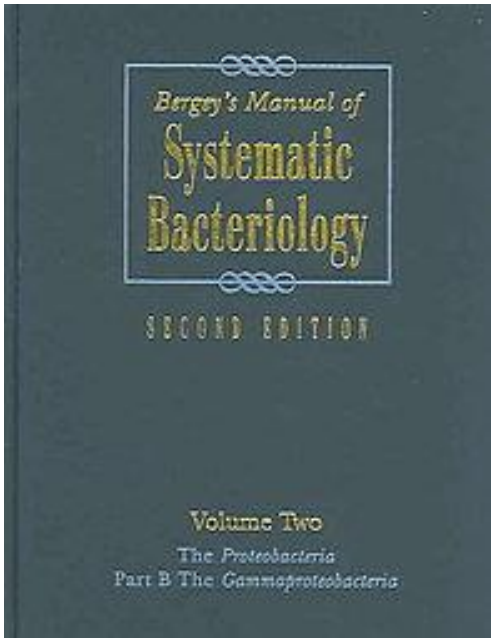
- **Volume 2 "The Proteobacteria."** (2004). Don J. Brenner, Noel R. Krieg, James T. Staley (Volume Editors), and George M. Garrity (Editor-in-Chief) with contributions from 339 colleagues.
- Hardcover: 2816 pages
- The volume provides descriptions of more than 2000 species in 538 genera that are assigned to the phylum Proteobacteria.
- This volume is subdivided into three parts:

A. **Part A- The Introductory Essays** (332 pgs, 76 figures, 37 tables).
B. **Part B- The Gammaproteobacteria** (1136 pages, 222 figures, and 300 tables).
C. **Part C- The Alpha-, Beta-, Delta-, and Epsilonproteobacteria** (1256 pages, 512 figures, and 371 tables).

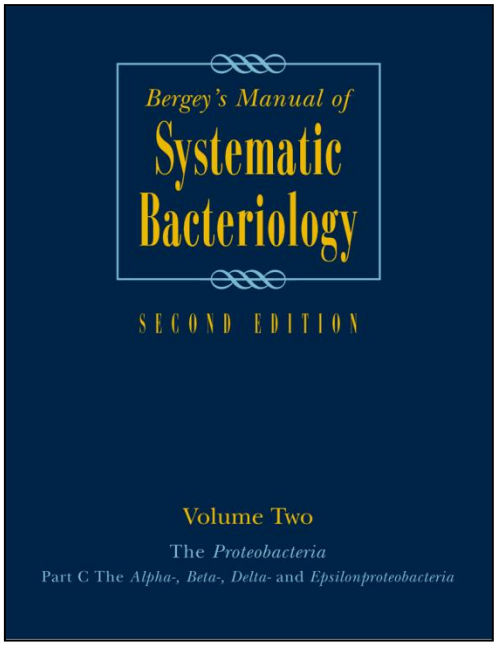
Bergey's Manual of Systematic Bacteriology
Volume 2 (Parts A, B & C)
The Proteobacteria



Part A- The Introductory Essays



Part B- The Gammaproteobacteria



Part C- The Alpha-, Beta-, Delta-, and Epsilonproteobacteria

Bergey's Manual Trust Home Page

Online encyclopedia of Systematics of Archaea and Bacteria (BMSAB)

Bergey's Manual Trust Home Page
Online encyclopedia of Systematics of Archaea and Bacteria (BMSAB)

- Bergey's Manual Trust and John Wiley & Sons, Inc., are pleased to announce a new partnership to produce an online encyclopedia entitled Bergey's Manual of Systematics of Archaea and Bacteria (BMSAB).
- This electronic work will provide:
- Up-to-date descriptions of the taxonomy, systematics, ecology, physiology and other biological properties of all named prokaryotic taxa.
- Release is anticipated in the fall of 2014.
- BMSAB will be available via Wiley Online Library.
- Individual volumes of BMSAB will also be made available in e-book format.
- Visit Bergey's Manual Trust at http://www.bergeys.org/for more information on the Trust and our other publications.

Bacterial nomenclature

German Collection of Microorganisms and Cell Cultures GmbH

DSMZ

DSMZ

Search Catalogues

Ordering/Prices

Patent- and Safe Deposit

Identification and Characterization

Research/Projects

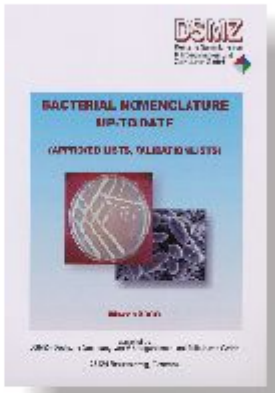
Publications

Download

Links

Bacterial Nomenclature Up-to-date

DSMZ
Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH



A service provided by the DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany

Genera: A ... L

Genera: M ... Z

Search option.

A complete list of the names can be downloaded from our FTP server: bactname.pdf (PDF file), bactname.exe (self-extracting file for MS-DOS systems) or names.txt (uncompressed ASCII text file). To import the list of names into a database or a spreadsheet programme, you can download bactname.zip (compressed Excel file).

Other useful information on bacterial or general nomenclature can be found at

- J. P. Euzéby, Ecole Nationale Vétérinaire de Toulouse, France
- CCUG,Sweden
- Species 2000 CheckList

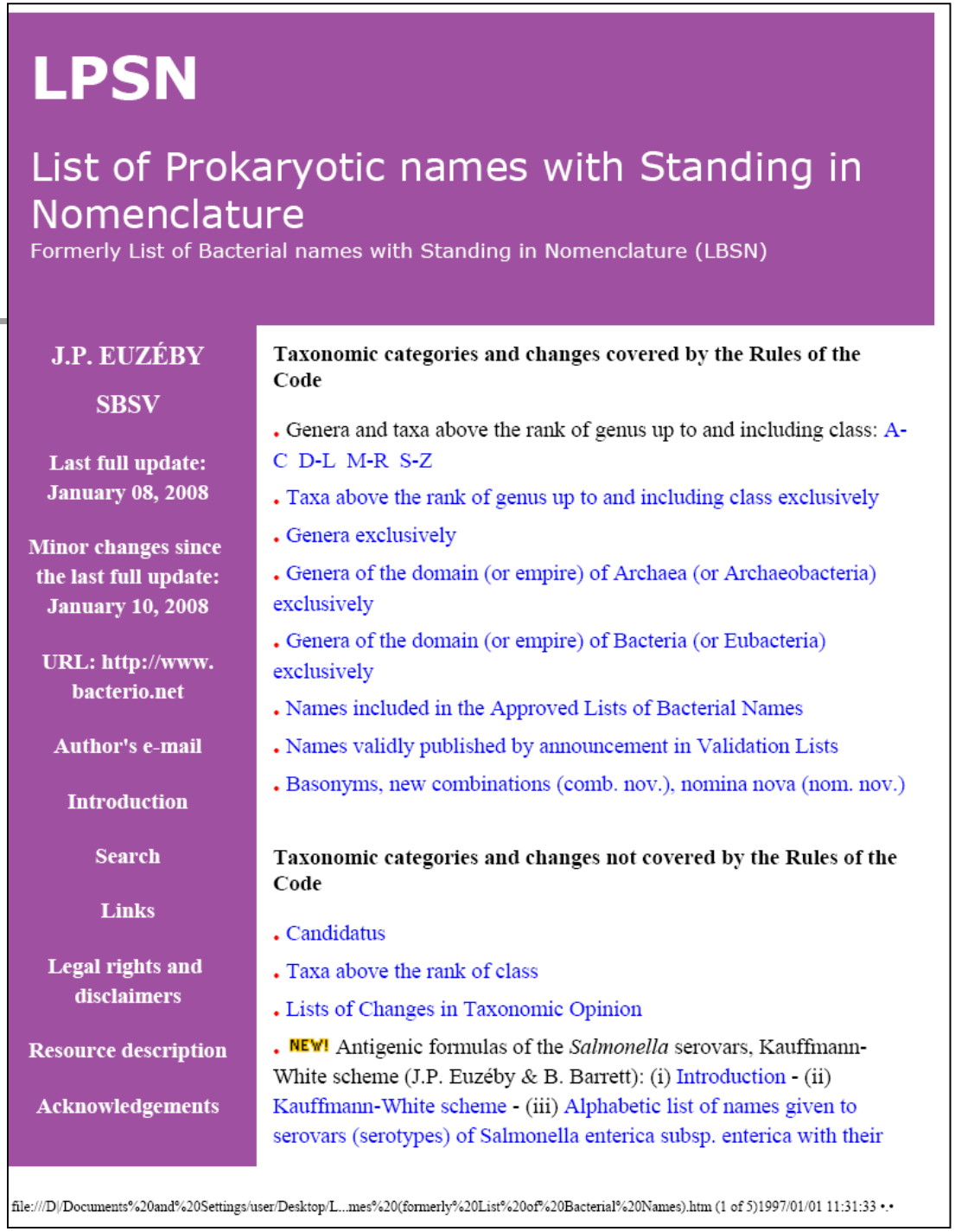
Euzéby website for Bacterial Standing Names

Up-to-date list of validly published names

(a key web source of validated species lists)

now LPSN-DSMZ's website (https://lpsn.dsmz.de).

Euzeby,2010



LPSN

List of Prokaryotic names with Standing in Nomenclature

Formerly List of Bacterial names with Standing in Nomenclature (LBSN)

J.P. EUZÉBY

SBSV

Last full update: January 08, 2008

Minor changes since the last full update: January 10, 2008

URL: http://www.bacterio.net

Author's e-mail

Introduction

Search

Links

Legal rights and disclaimers

Resource description

Acknowledgements

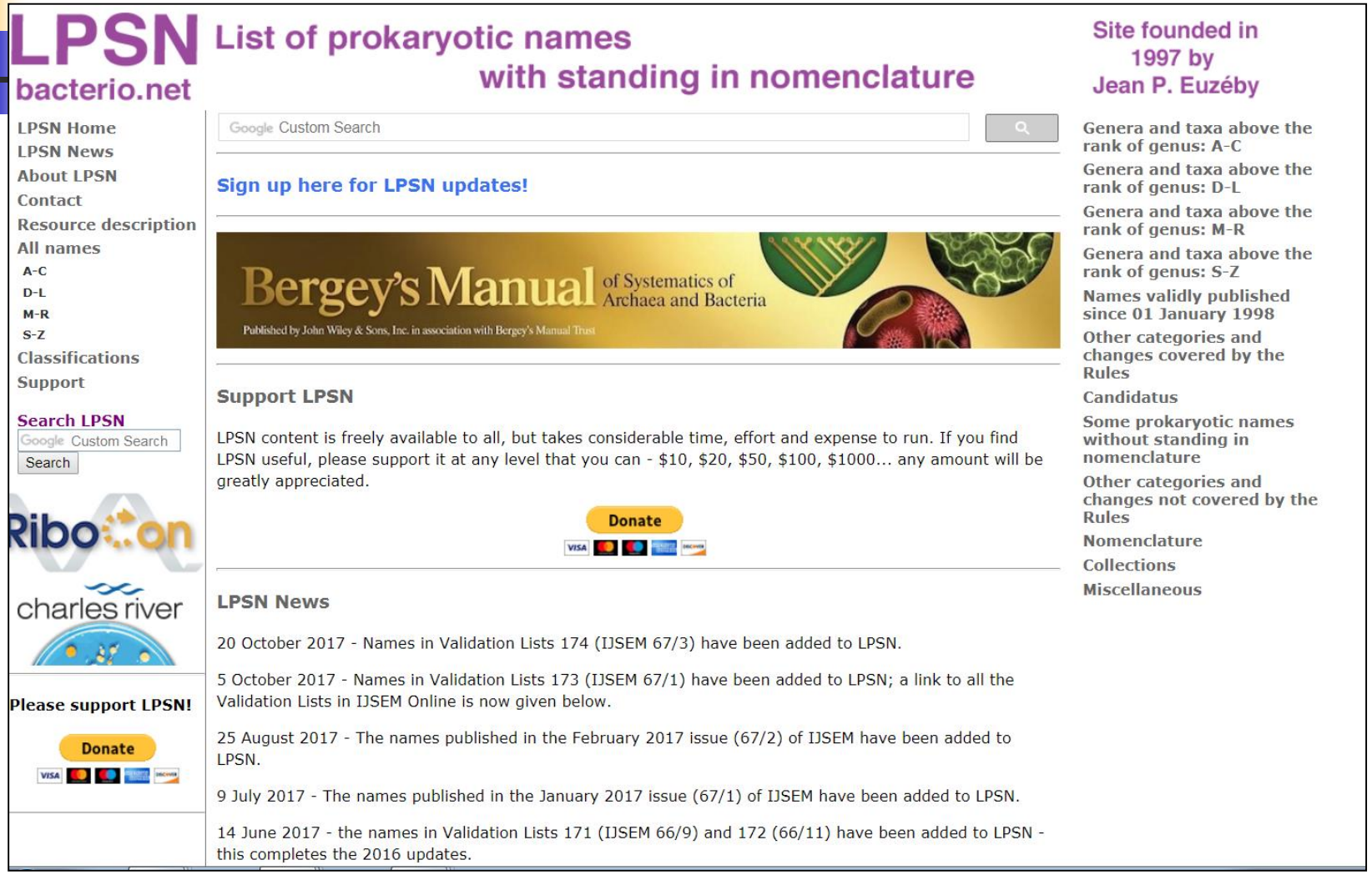
Taxonomic categories and changes covered by the Rules of the Code

- Genera and taxa above the rank of genus up to and including class: A-C D-L M-R S-Z
- Taxa above the rank of genus up to and including class exclusively
- Genera exclusively
- Genera of the domain (or empire) of Archaea (or Archaeobacteria) exclusively
- Genera of the domain (or empire) of Bacteria (or Eubacteria) exclusively
- Names included in the Approved Lists of Bacterial Names
- Names validly published by announcement in Validation Lists
- Basonyms, new combinations (comb. nov.), nomina nova (nom. nov.)

Taxonomic categories and changes not covered by the Rules of the Code

- Candidatus
- Taxa above the rank of class
- Lists of Changes in Taxonomic Opinion
- NEW! Antigenic formulas of the *Salmonella* serovars, Kauffmann-White scheme (J.P. Euzéby & B. Barrett): (i) Introduction - (ii) Kauffmann-White scheme - (iii) Alphabetic list of names given to serovars (serotypes) of Salmonella enterica subsp. enterica with their

LPSN: The List of Prokaryotic Names with Standing in Nomenclature (http://www.bacterio.net)



Euzeby,2017

List of Prokaryotic Names with Standing in Nomenclature (LPSN)

Compiled, edited and curated by Prof. J.P. Euzéby

- LPSN is a database of bacterial names and prokaryotes names.
- It includes a comprehensive classification and taxonomy.
- Formerly "List of Bacterial names with Standing in Nomenclature (LBSN)".
- http://www.bacterio.cict.fr/

List of Prokaryotic names With Standing in Nomenclature

Note: From February 2020, LPSN was integrated into the DSMZ's Prokaryotic Nomenclature Up-to-date (PNU) database as an all-new service. LPSN-DSMZ's website (https://lpsn.dsmz.de).

Table 1.
Bacterial Plant Pathogen
Websites of Interest

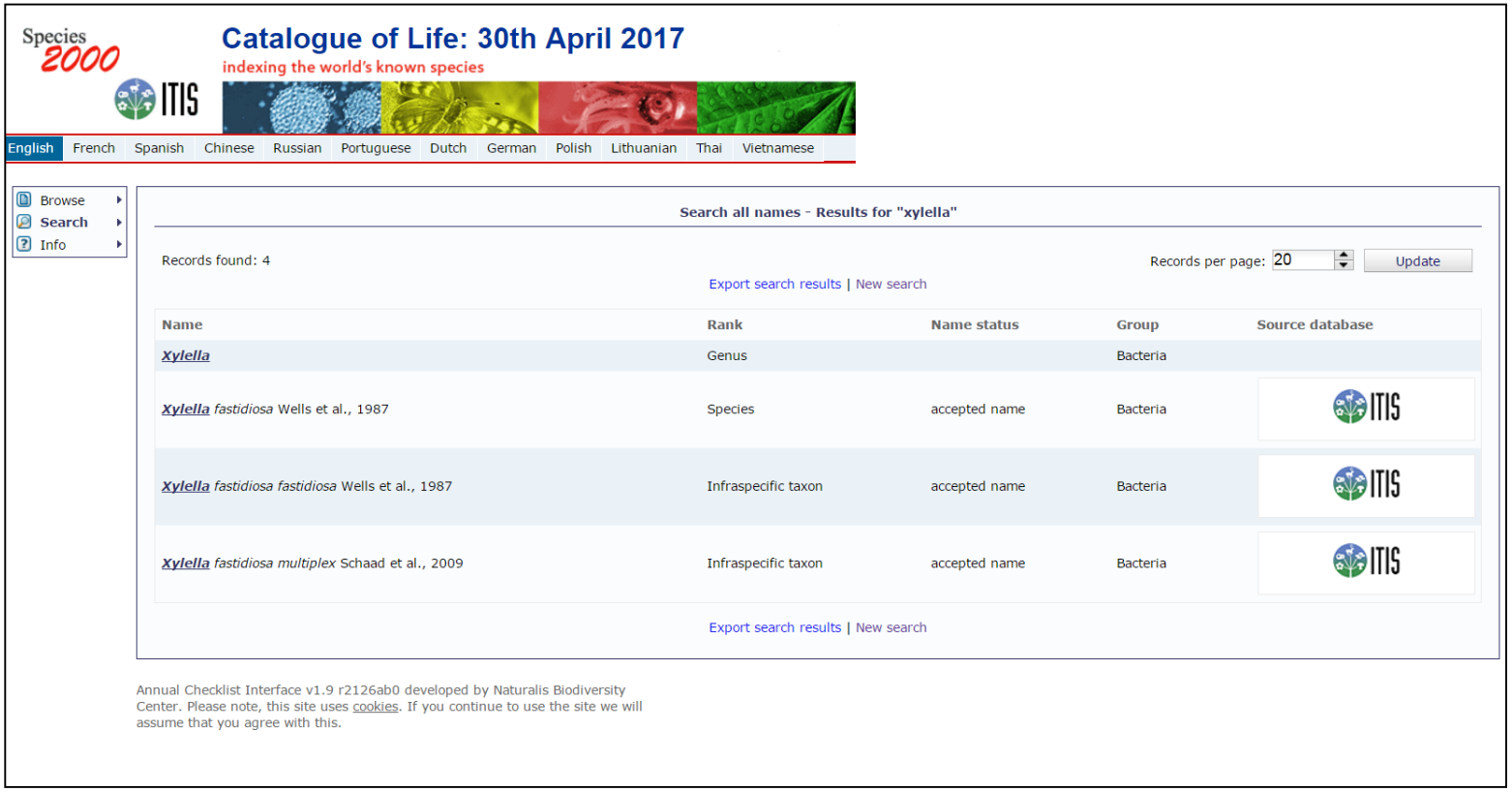
Bacterial Nomenclature	Bacterial Nomenclature Up-to-date	http://www.dsmz.de/bactnom/bactname.htm
	List of Bacterial Names with Standing in Nomenclature	http://www.bacterio.cict.fr/
	Bergey's Manual of Systematic Bacteriology 2nd Edition – Taxonomic Outline of the Procaryotes	http://dx.doi.org/10.1007/bergeysoutline200210

Vidaver and Lambrecht,2004

Catalogue of Life

Species 2000 CheckList

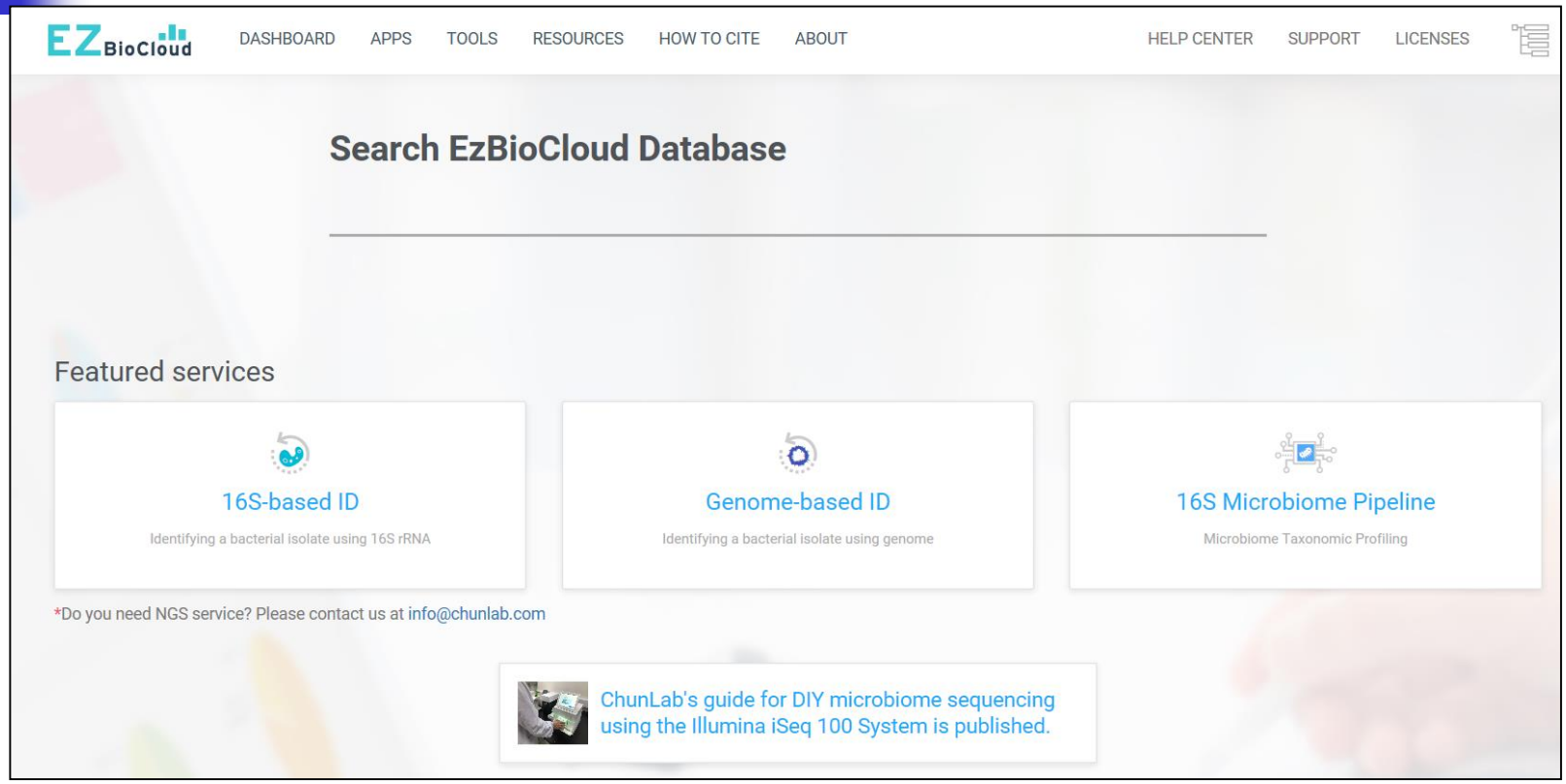
http://www.catalogueoflife.org/col/



Roskov Y, Kunze T, Orrell T, Abucay L, Paglinawan L, Culham A, et al., editors. Species 2000 & ITIS Catalogue of Life, 2014 Annual Checklist [DVD]. 2014; Naturalis, Leiden, the Netherlands: Species 2000.

Up-to-date list of validly published names

EzBioCloud Database



EzBioCloud,2019

What name should be used?

- Although the Code and the Standards govern the correct application and publication of names, there is no authority governing which classification and adherent nomenclature should be used.
- Sometimes, a lack of consensus about which classification (former or new) is most appropriate can lead to two or more legitimate names being used in the literature for the same organism.

Bull *et al.*,2008

What name should be used? *Agrobacterium* spp.

- The reclassification of *Agrobacterium* spp. into *Rhizobium* is a case in point (Young *et al.*,2001b).
- The authors proposed that strains of *Agrobacterium* spp. and *Rhizobium* spp. should be included in a single genus for which the name *Rhizobium* had priority.
- There is little use, as yet, made of the new nomenclature.
- This is probably because *Agrobacterium tumefaciens* is a familiar name in many disciplines, in addition to a lack of agreement with the newly proposed classification (Farrand *et al.*,2003).
- In case of *A. vitis*, it was proposed as *Rhizobium vitis* (Young *et al.*,2001). More recently, it was transferred to the genus *Allorizobium* (Mousavi *et al.*,2014 and 2015).

Agrobacterium spp.
with a validly published and correct name

	Name	Host plant disease
1	***Agrobacterium arsenijevicii*** **Kuzmanović et al. 2019**	**crown gall tumors on raspberry and cherry plum**
2	***Agrobacterium cavarae*** **Flores-Felix et al. 2020**	**galls in grapevines**
3	***Agrobacterium fabacearum*** **Delamuta et al. 2020**	**nodules of plants of the family Fabaceae**
4	***Agrobacterium larrymoorei*** **Bouzar and Jones 2001**	**aerial tumours of *Ficus benjamina***
6	***Agrobacterium radiobacter*** **(ex. *Agrobacterium tumefaciens*) (Beijerinck and van Delden 1902) Conn 1942 (Approved Lists 1980)**	
8	***Agrobacterium rosae*** **Kuzmanović et al. 2019**	**galls on different agricultural crops**
9	***Agrobacterium rubi*** **(Hildebrand 1940) Starr and Weiss 1943 (Approved Lists 1980)**	**galls on blackberry**

In case of *A. vitis*, it was transferred to the genus *Allorizobium* as *Allorhizobium vitis* (=*A. vitis*) (Mousavi *et al.*,2014 and 2015).

Euzeby,2020

What name should be used?
Erwinia spp.

- In contrast, there is reasonably good acceptance of most if not all the new genera derived from *Erwinia*.
- Unfortunately there is no easy rule to help decide which valid name to use.
- Individual authors must critically examine the taxonomic literature to decide which classification is most appropriate and suitable for their needs.

Bull *et al.*,2008

What name should be used?

Permits of some office to use the synonyms of bacterial names

- In many instances the old/common names are still used by growers, but scientists have split the bacterial genera into numerous species and totally renamed others.
- This makes it difficult to understand exactly what we are trying to identify and regulate; and may also pose a problem when code enforcement is required.
- Thus, a new list of bacterial plant pathogens, synonyms, common names was generated.
- I believe this should help to clarify the confusion of our bacterial nomenclature that has evolved over the last several decades.

Dan Opgenorth,2005

What name should be used?

Nomenclature of Plant Pathogenic Bacteria

Original Name	Synonyms
Corynebacterium fascians	*Rhodococcus fascians*
Corynebacterium michiganense subsp. *insidiosum*	*Clavibacter insidiosus*
Corynebacterium michiganense	*Clavibacter michiganensis*
Corynebacterium michiganense subsp. *sepedonicum*	*Clavibacter sepedonicus*
Erwinia carotovora	*Pectobacterium carotovorum*
Erwinia nigrifluens	*Brenneria nigrifluens*
Pseudomonas caryophylli	*Burkholderia caryophylli*
Pseudomonas solanacearum	*Ralstonia solanacearum*

Culture Collections of Prokaryotes
Reference cultures

- When isolates are designated as reference strain (especially when they function as the type strain for a certain species or pathovar they should be deposited in one of the official culture collections such as:
- ATCC: American Type Culture Collection, Rockville MD, USA.
- CNBP: Collection National de Bactéries Phytopathogènes Angers, France.
- ICMP: International Collection of Micro-organisms from Plants, DSIR Auckland, New Zealand.
- NCPPB: National Culture Collection Plant Pathogenic Bacteria CSL, York, UK.
- PD: Culture Collection Plant Protection Service, Wageningen, The Netherlands.

Janse,2006

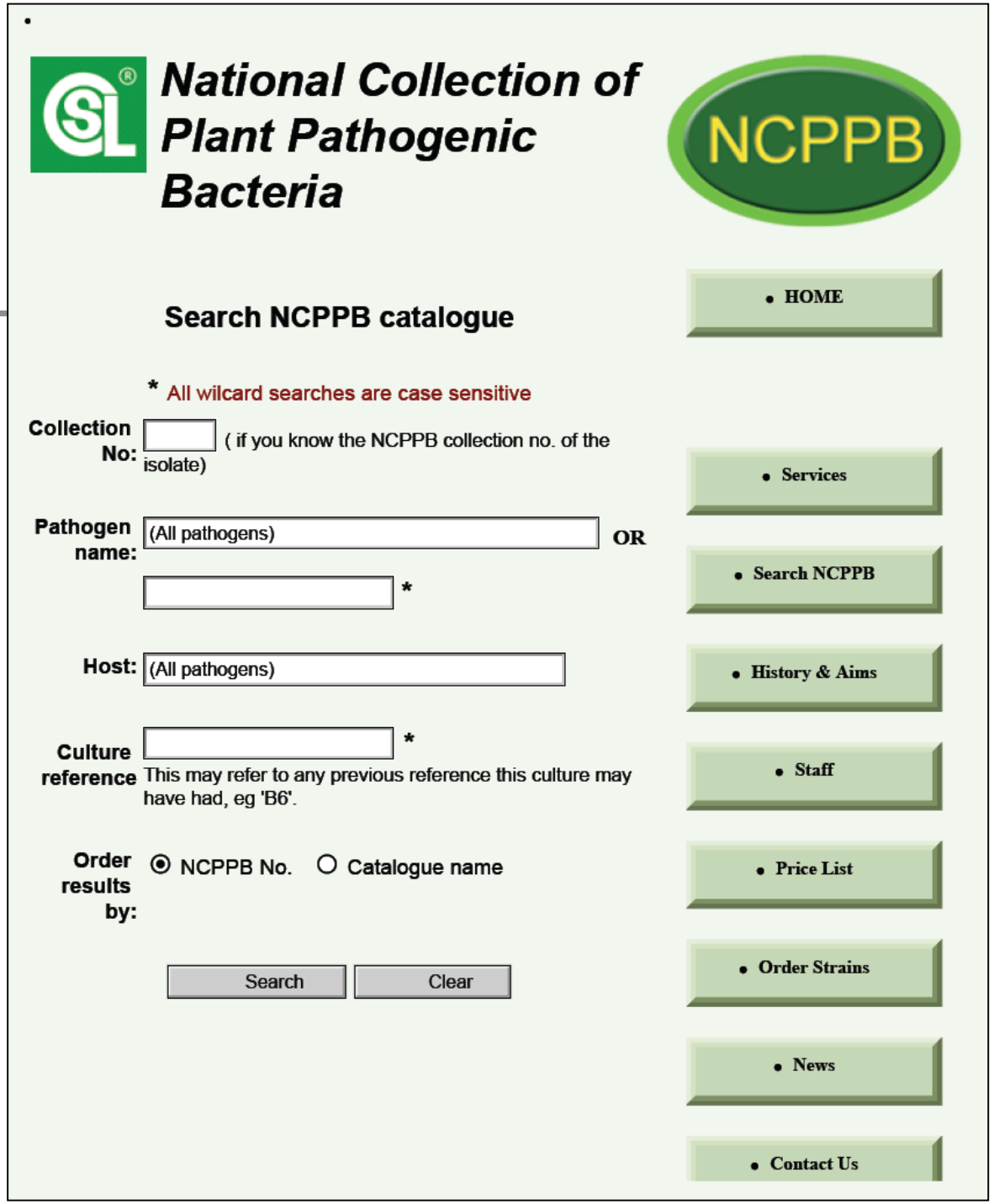
Culture Collections of Prokaryotes
Reference cultures

- ACM: Australian Collection of Microorganisms.
- AMRC: FAO-WHO International Reference Centre for Animal Mycoplasmas, Institute for Medical Microbiology, University of Aarhus, Aarhus, Denmark.
- BKM: All-Union Collection of Microorganisms, Institute of Microbiology, USSR Academy of Sciences, Moscow, USSR.
- BKMW: Culture Collection, Institute of Microbiology, USSR Academy of Sciences, Moscow, USSR.
- CIP: Collection of the Institute Pasteur, Paris, France.
- DSM: Deutsche Sammlung von Mikroorganismen, Schnittspahnstrasse (in English: German Collection of Microorganisms and Cell Cultures), Darmstadt, Federal Republic of Germany.
- KCTC: Korean Collection for Type Cultures.
- NCCB: Netherlands Culture Collection of Bacteria.
- PTCC: Persian Type Culture Collection, Iran.
- ACCC: Agricultural Culture Collection of China.
- ITCC: Indian Type Culture Collection.

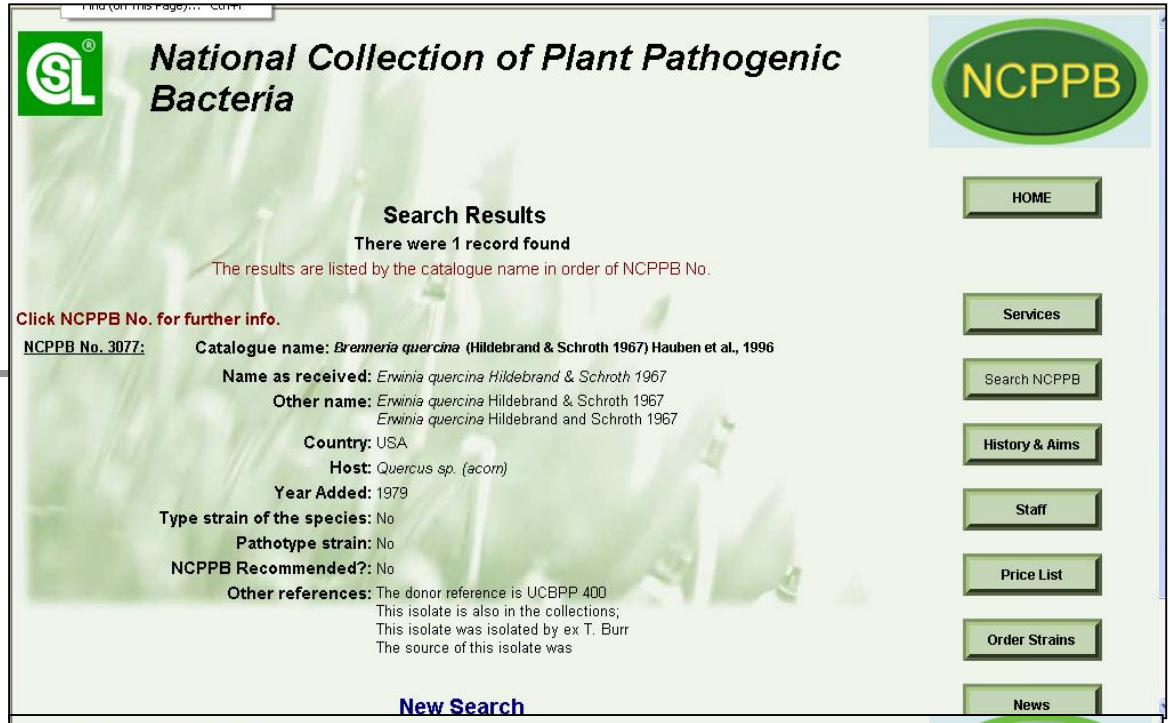
Culture Collections of Prokaryotes
Reference cultures

- IPV: Instituto de Patologia Vegetale, Milan, Italy.
- LIA: Museum of Cultures, Leningrad Research Institute of Antibiotics, 23 Ogorodnikov Prospect, Leningrad L-20, USSR.
- NCMB: National Collection of Marine Bacteria, Torry Research Station, Aberdeen, Scotland, UK.
- NIHJ: National Institute of Health, Tokyo, Japan.
- UQM: Culture Collection, Department of Microbiology, University of Queensland, Brisbane, Australia.
- WINDSOR: Culture Collection, University of Windsor, Windsor, Ontario, Canada.

Searching and ordering strains from NCPPB Culture Collection, UK.

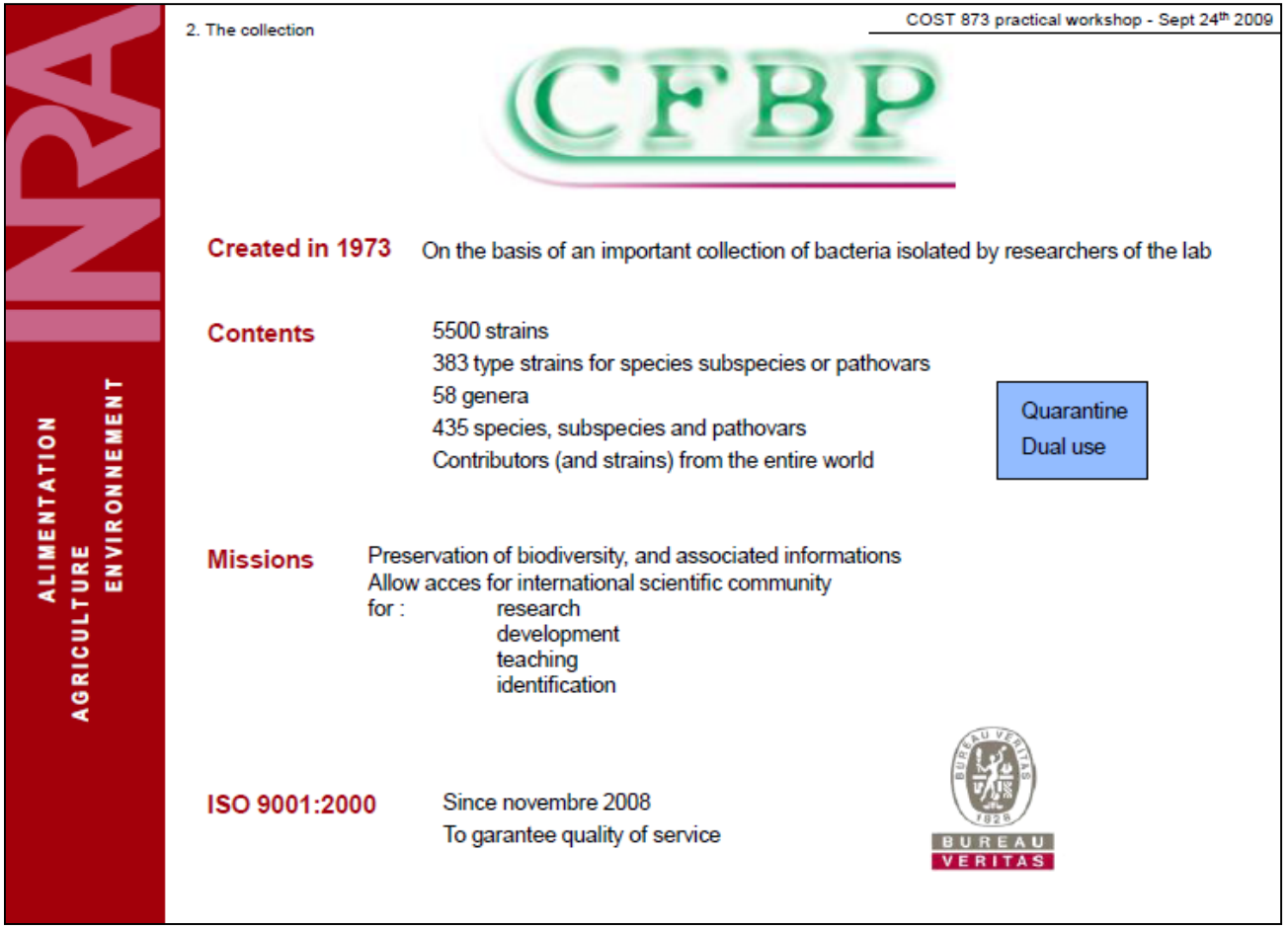


Searching and ordering strains from NCPPB Culture Collection, UK.



CFBP

French Collection for Plant-Pathogenic Bacteria

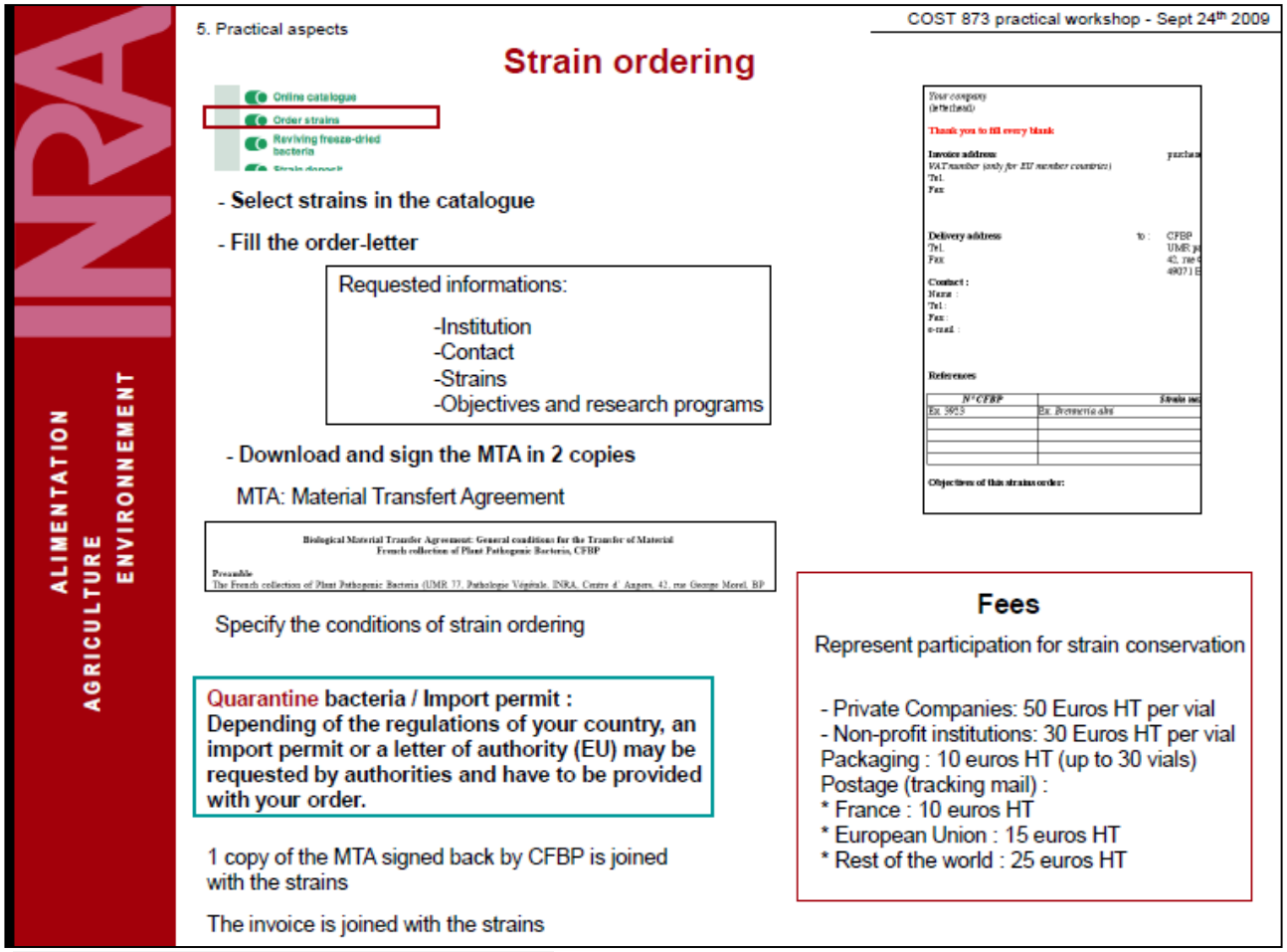


Portier,2009

CFBP

French Collection for Plant-Pathogenic Bacteria

Strain ordering



HT=Hors taxes (not including)

Portier,2009

DSMZ

Microbial identification services

Prices for services and techniques

Services and techniques for Bacteria and Archaea

- General identification of bacteria »
- Phenotypic characterization »
- Production of biomass for the different techniques »
- Analysis of the cellular fatty acid composition »
- Analysis of respiratory quinones »
- Analysis of polar lipids »
- Analysis of mycolic acids »
- Analysis of DAP (2,6-diaminopimelic acid) »
- Analysis of peptidoglycan structure »
- Analysis of whole cell sugars »
- DNA base composition by HPLC »
- DNA - DNA hybridization »
- Ribotyping »
- MALDI-TOF »
- Partial 16S rDNA sequence analysis »
- Complete 16S rDNA sequence analysis »
- Full phylogenetic study by complete 16S rDNA sequence analysis »
- Antibiotic susceptibility testing »

Bacteria and Archaea	Price
General identification of bacteria	450 €
Pre-identification (cost related to complexity)	Inquire
Fee for purification of a culture	from 100 €
Phenotypic characterization	Inquire
Production of biomass for the different techniques	300 - 600 €
Analysis of the cellular fatty acid composition	85 €
Analysis of respiratory quinones	275 €
Analysis of polar lipids	275 €
Analysis of mycolic acids	275 €
Analysis of DAP (2,6-diaminopimelic acid)	125 €
Analysis of peptidoglycan structure	600 €
Analysis of whole cell sugars	160 €
DNA base composition by HPLC	240 €
Preparation of DNA for DNA - DNA hybridization	250 €
DNA - DNA hybridization	125 €
Ribotyping	320 €
MALDI-TOF	125 €
Partial 16S rDNA sequence analysis	220 €
Complete 16S rDNA sequence analysis	425 €
Full phylogenetic study by complete 16S rDNA sequence analysis	800 €
Antibiotic susceptibility testing	Inquire

Material transfer
DNA from dead bacterial cells

- To send DNA from dead bacterial cells no permits are required.
- To prepare heat-killed cells, simply spread the bacteria from a plate onto some type of absorbant paper (like filter paper or tissue paper) and place the paper at 55°C in a dry incubator for at least 10 min.
- This will kill the bacteria, but will not damage the DNA.
- Fold up the filter paper and place each sample in its own small envelope or plastic bag. They can all be mailed together in a larger package.
- You can then send the filter paper as a "non-hazarous DNA sample" through the regular mail.

Virtual Museum of Bacteria

The Foundation of Bacteriology and the Society for Applied Microbiology

- This site (http://bacteriamuseum.org/) brings together many links on bacteria, bacteriology, and related topics available on the web.
- It also provides crystal-clear information about many aspects of bacteria.

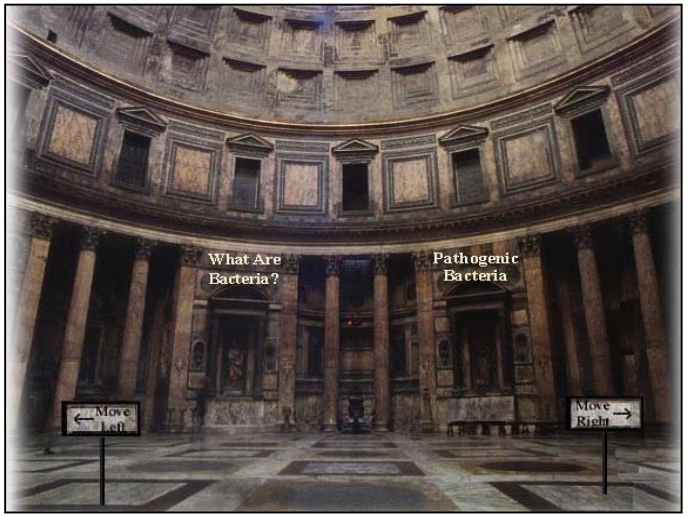
Wassenaar,2012

Virtual Museum of Bacteria

The Foundation of Bacteriology and the Society for Applied Microbiology

- We provide many links to pictures, video's, and explanatory text.
- By browsing through our links you are presented with a synopsis of the diverse world of bacteria,

1. you can learn that not all bacteria are harmful,
2. how they are used in industry, that they belong to the oldest living creatures on Earth, and
3. much more.



Museum of Bacteria

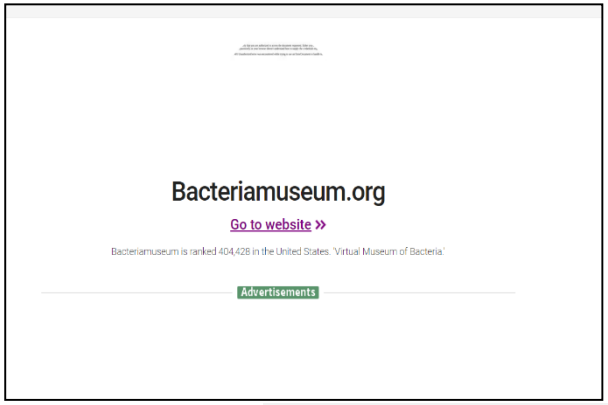
Information on bacteria

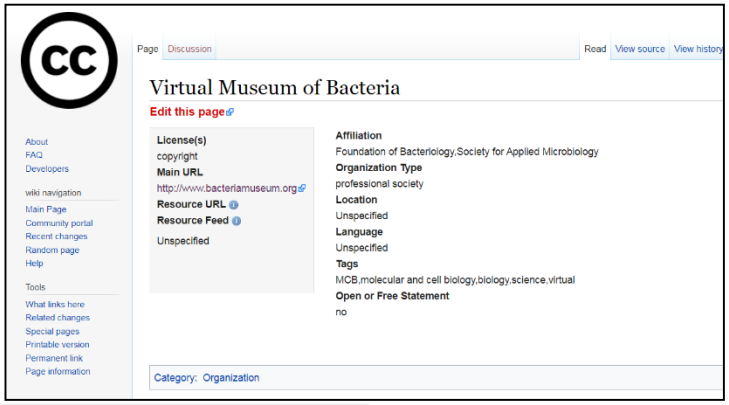
Wassenaar,2007

Virtual Museum of Bacteria

The Foundation of Bacteriology and the Society for Applied Microbiology

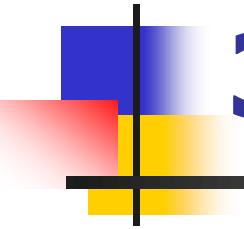
- A thoughtful collection of species-based links and facts about pathogenic bacteria, bacteriophages, origin of life, food/water safety, antibiotics and vaccination.





Asks for the username and password

Website address: http://www.bacteriamuseum.org



3. Classification

Classifications group similar things together

Classification systems

Prokaryote (bacteria and archaea) Classification

- **Two general ways to classify based on:**
- Overall similarity – phenetic
- Evolutionary relationships – phylogenetic

Therefore:

- Taxonomy is traditionally phenotypic.
- Phylogeny is mainly genetic.

The goal of classification

- The goal of classification is to determine genetic similarity.
- As a consequence of the sequencing of organism genomes, the science of classification is very much in flux (classifications are changing often) as phenotype-based classification is augmented by genotype-based classification.
- All is not happy even with genotype-based classification, however, since horizontal gene transfer greatly complicates estimations of genetic similarity, with different parts of genomes displaying different levels of genetic similarity between organisms.

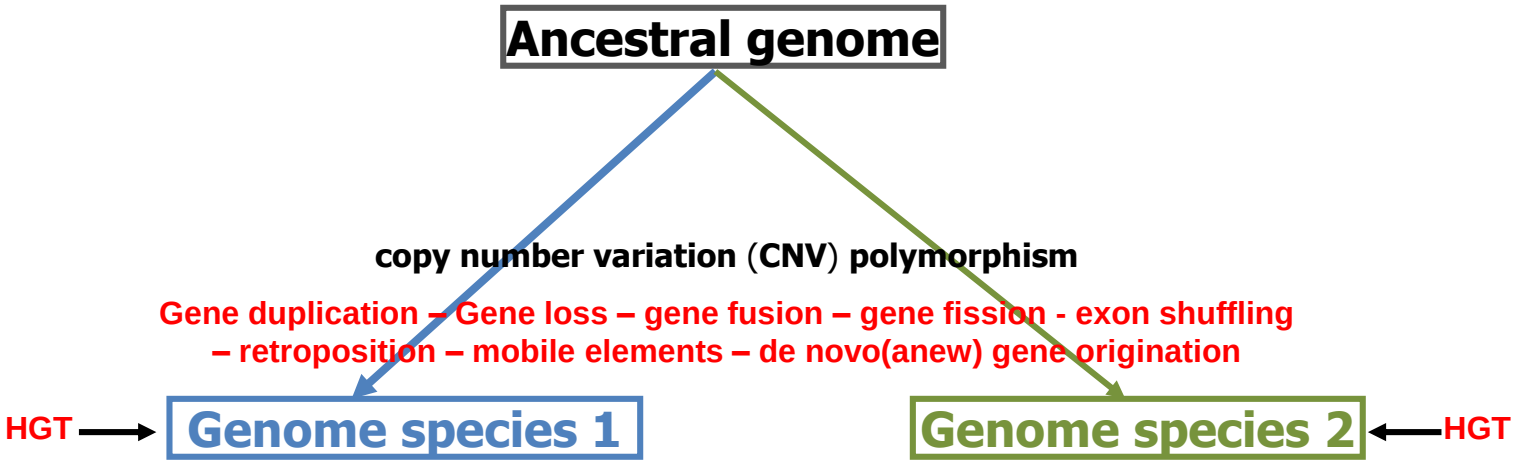
Classification approaches

Subjective	Objective
Phenetic classification	**Phylogenetic classification**
Distinguish among pathogens	Testable hypotheses
Few characters tested	Polyphasic
Monothetic (the use of a single criterion) - special purpose classifications	Polythetic (using several or all possible criteria) classifications
Monophyletic	Polyphyletic
Criteria change depending on species	Criteria universal for all bacteria

Monophyletic: A group consisting of a common ancestor +all descendants of that common ancestor.

Classification approaches

- **Phylogeny** is the study of evolutionary relationships.
- **Phylogenetic analysis** is the means of inferring evolutionary relationships.



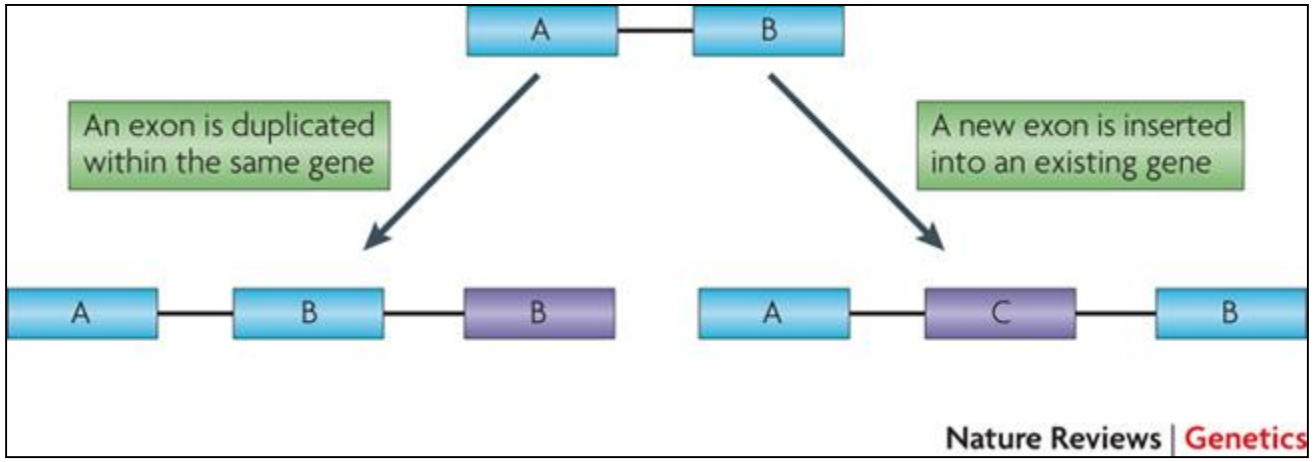
brigitte.boeckmann@isb-sib.ch

Classification approaches

Ancestral genome

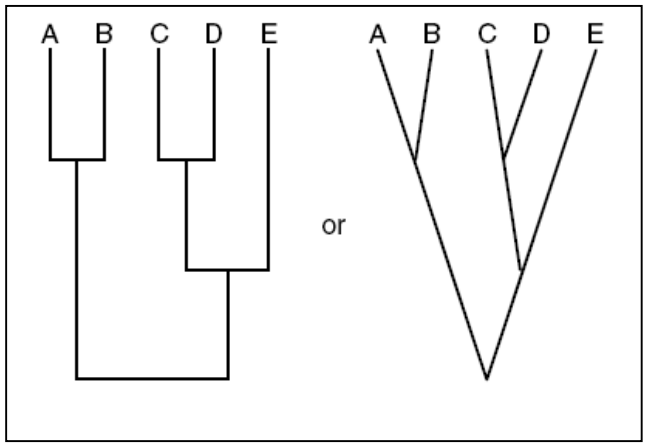
Exon shuffling

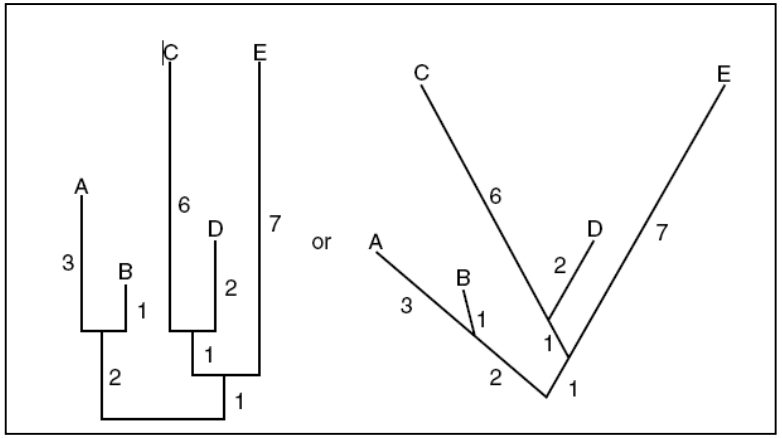
- Exon shuffling (rearrengment) is a molecular mechanism for the formation of new genes.
- It is a process through which two or more exons from different genes can be brought together ectopically (out of place), or the same exon can be duplicated, to create a new exon-intron structure.



Classification approaches
Phylogenetic trees

- Cladogram: branching diagram without branch length estimates.
- Phylogram or phylogenetic tree: branching diagram with branch length estimates.



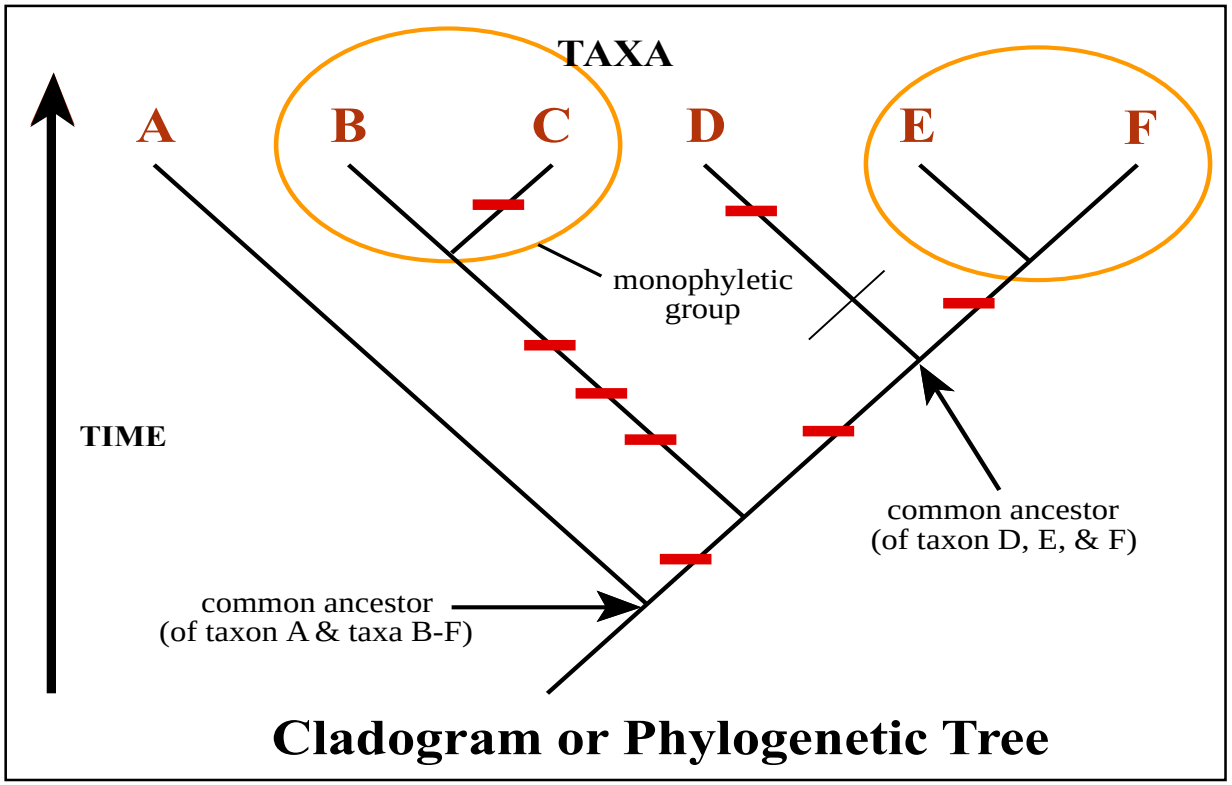


- The branch length in phylogram represents the number of character changes.
- Molecular clock: The molecular clock hypothesis states that DNA and protein sequences evolve at a rate that is relatively constant over time and among different organisms.

Classification approaches

Cladogram or phylogenetic trees

Monophyletic group



Both **phylogenetic trees** and **cladograms** help show the relationships between different organisms, but only **phylogenetic trees** have branches that represent evolutionary time and amount of change.

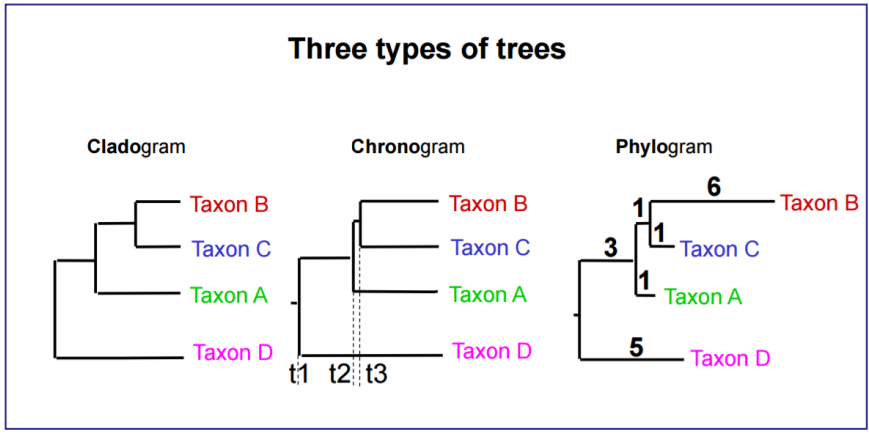
Monophyletic: A group consisting of a common ancestor+all descendants of that common ancestor.

Classification approaches

Three types of trees

Cladogram, phylogram and chronogram

1. Chronogram (Time calibrated phylogeny): represents evolutionary time through its branch lengths.
2. Cladogram: branching diagram without branch length estimates.
3. Phylogram or phylogenetic tree: branching diagram with branch length estimates.



Characteristics used in taxonomy

1. Phenotypic characteristics: Morphology, physiology, Biochemical; Numerical taxonomy.
2. The first genomic parameters: G+C mol% content and whole-genome hybridizations (DDH).
3. Chemotaxonomy: Numerous techniques revealing relevant chemical components.
4. Phylogenetic indicators: 16S rRNA,16S-23S rDNA intergenic spaces (ITS),…

Characteristics used in taxonomy

Characters evaluated

- **Phenotypic**
  - Gram reaction
  - Carbon utilization
  - Fatty acid production
  - Ice nucleation
  - Flagella
  - Phage sensitivity
  - Antibody specificity
  - Pathogenicity
- **Nucleic Acid Based**
  - **ANI - genome sequence**
  - **MLSA**
  - Rep-PCR
  - RAPDS
  - AFLP
  - RFLP
  - ITS-PCR
  - DDH
  - ARDRA
  - 16S rDNA sequencing

Average Nucleotide Identity (ANI) values between your isolate and type strains. MLSA: Multilocus sequence analysis.

Bull,2013

Classification systems
Classification types

- Historically, there have been two main classification schemes used to identify living organisms:

1. Artificial, and
2. Natural classification.

Classification systems
Phylogenetic classification

- T present there are three types of classifications.
- They are:

1. artificial,
2. natural, and
3. phylogenetic.

Woese, C.R., Kandler, O. and M.L. Wheels. Towards a natural system of organisms: proposal for the domains of Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.* 87, 4576-4579. (1990).

Classification systems

1. Artificial classification

1. Only a few morphological characters are considered
2. Members of different groups are usually not similar in hereditary pattern
3. Stable classification
4. Provides only limited information
5. Cannot add new work
6. Difficult for identifying a particular species or genera
7. Not related phylogenetically.

https://www.majordifferences.com/

Phenetic Classification

Artificial classification

Two kingdoms: Plant and Animal

- Before scientists had a clear understanding of the nature of microbes the biological world was classified into two kingdoms:

1. Plant
2. Animal

- Bacteria were placed into plant kingdom(Plantae).
- Clearly, this scheme was inadequate.

Classification systems
2. Natural classification

- It is based on the natural affinities among the organisms and considers not only the external features but also the internal one (phenotypes and genotypes).
- Almost all the characters are considered
- Members of different groups are mostly similar in hereditary pattern
- May change with advancement in knowledge
- Provides plenty of useful information
- Recent advancement in the field can be added
- Closely related to phylogenetically.

Natural system of classification is based on a large number of characters. In this system, the genus and accompanying higher taxa consist of all the species that have evolved from one common ancestral species.

Classification systems
Phylogenetic classification

- A third type of classifications is phylogenetic classification.
- In phylogenetic system of classification reflects possible evolutionary relationships based on ancestry and descent.

1. Based on many phenotyptic and genotypic characters with emphasis evolutionary relationships.
2. Characters chosen have evolutionary significance from primitive to evolved conditions.
3. Gives some idea on origin and evolution of different taxa.

- It generates trees called cladograms, which are groups of organisms that include an ancestor species and its descendants.

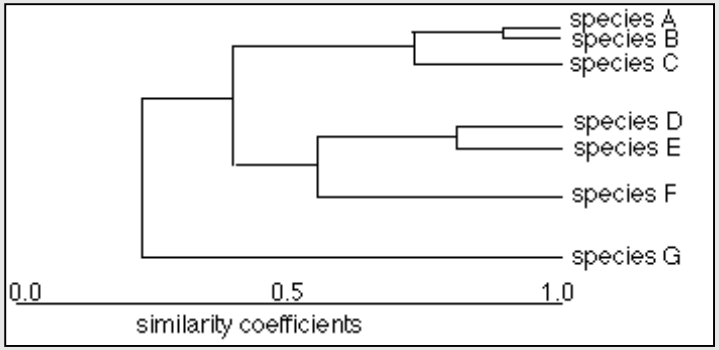
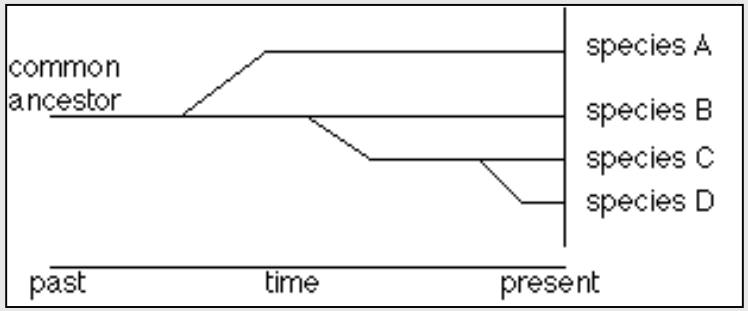
Classification systems
Phenetic system vs. Phylogenetic system

- **Two very different ways to construct a taxonomy:**

Phenetic system	Phylogenetic system
1. Groups organisms based on mutual similarity of phenotypic characteristics. 2. May or may not correctly match evolutionary grouping. ▪ Example: ▪ Group (motile) organisms in one group, non-motile organisms in another group. ▪ This is useful, but does it reflect underlying evolutionary ancestry?	1. Groups organisms based on shared evolutionary heritage. ▪ Example: ▪ Mycoplasma (no wall) and *Bacillus* (walled Gram+ rods) are not obviously similar, would not be grouped together phenetically. ▪ But evolutionarily they are similar, more so than either to Gram-ve organisms.

Classification systems
Phenetic system vs. Phylogenetic system

- **Two very different ways to construct a taxonomy:**

Phenetic system:	Phylogenetic system:
	
	Cladistic: Each branch of the tree descends from common ancestor.

Branches show the path of transmission of genetic information from one generation to the next. Branch lengths indicate genetic change i.e. the longer the branch, the more genetic change (or divergence) has occurred. Typically we measure the extent of genetic change by estimating the average number of nucleotide or protein substitutions per site.

Prokaryotes
The History

1735 Plant and Animal Kingdoms.

1857 Bacteria and fungi put in the Plant Kingdom.

1866 Kingdom Protista proposed for bacteria, protozoa, algae, and fungi.

1937 Prokaryote introduced for cells "without a nucleus".

1961 Prokaryote defined as cells in which nucleoplasm is not surrounded by a nuclear membrane.

1959 Kingdom Fungi.

1968 Kingdom Prokaryotae proposed (Murray,1968).

1978 Two types of prokaryotic cells found (Woese,1977).

Prokaryotes

Most known prokaryotes are bacteria

- The name bacteria was once synonymous with "prokaryotes" but it now applies to just one of the two distinct prokaryotic domains.
- However, most known prokaryotes are bacteria.
- Every nutritional and metabolic mode is represented among the thousands of species of bacteria.
- The major bacterial taxa are now accorded kingdom status by most prokaryotic systematists.

Phenetic Classification

Natural classification

Five-Kingdom System of Biological Classification proposed in 1969 by Robert Whitaker.

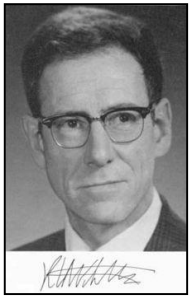
An American plant ecologist

December, 27 - October 20, 1980 (aged 59).

Phenetic Classification

Five-kingdom classification system

Natural classification



- **Proposed in 1969 by Robert Whittaker.**
- Whittaker's system of classification is based on:

1. cell type (prokaryotic or eukaryotic/true nucleus);
2. cell wall (present or absent);
3. mode of nutrition (absorptive or ingestive);
4. Nutritional class (photo/chemo/heterotrophic);
5. body organization (uni- or multi-cellular).

The heterotrophic bacteria can be parasitic or saprophytic. The autotrophic bacteria can be chemosynthetic or photosynthetic.

Phenetic Classification

Five-kingdom classification system

Natural classification

- Whittaker's classification scheme recognizes **five kingdoms**:

1. Monera,
2. Protista,
3. Fungi,
4. Plantae, and
5. Animalia.

Phenetic Classification

The five-kingdom system, proposed in 1969 by Whittaker

Monera (bacteria and prokaryotes)

- Monera are the most primitive of organisms.
- Bacteria are classified in the Monera kingdom.
- Later, the Bacteria were classified under Kingdom Prokaryotae (aka Monera).
- And prokaryotes were considered as oldest living structures on the earth.

Aka= Also known as

Phenetic Classification

Five-kingdom classification system

Natural classification

1. Kingdom Procaryotae (Monera): Oldest known cells. Lived over 3.5 billion years ago. Lack a nucleus and membrane bound organelles.

- The other four kingdoms are eucaryotes:
- Have a true nucleus and membrane bound organelles.

2. Kingdom Protista: Mostly unicellular, lack tissue organization. Most have flagella during life.
3. Kingdom Fungi: May be unicellular (yeasts) or multicellular (molds). Many are saprotrophs.
4. Kingdom Plantae: Multicellular, photosynthetic.
5. Kingdom Animalia: Multicellular, heterotrophs that ingest food through a mouth or oral cavity.

Phenetic Classification

Five-kingdom classification system

Table 8.1 Properties of the five kingdoms

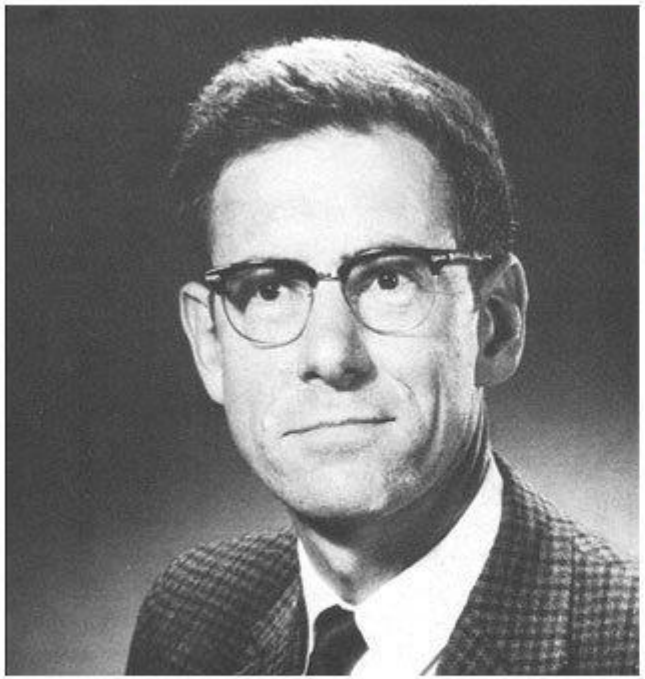
Kingdom	Monera	Protista	Fungi	Plantae	Animalia
Cell type	Prokaryotic	Eukaryotic	Eukaryotic	Eukaryotic	Eukaryotic
Cell organization	unicellular	unicellular	Multicellular and unicellular	Multicellular	Multicellular
Cell Wall	Present in most	Present in some absent in others	Present	Present	Absent
Nutritional Class	Phototrophic, heterotrophicor chemoautotrophic	Heterotrophic and phototrophic	Heterotrophic	Phototrophic	Heterotrophic
Mode of nutrition	Absorptive	Absorptive or ingestive	Absorptive	Mostly Absorptive	Mostly ingestive

All bacteria and possibly blue-green algae are "prokaryotic" organisms, lacking a nuclear membrane.

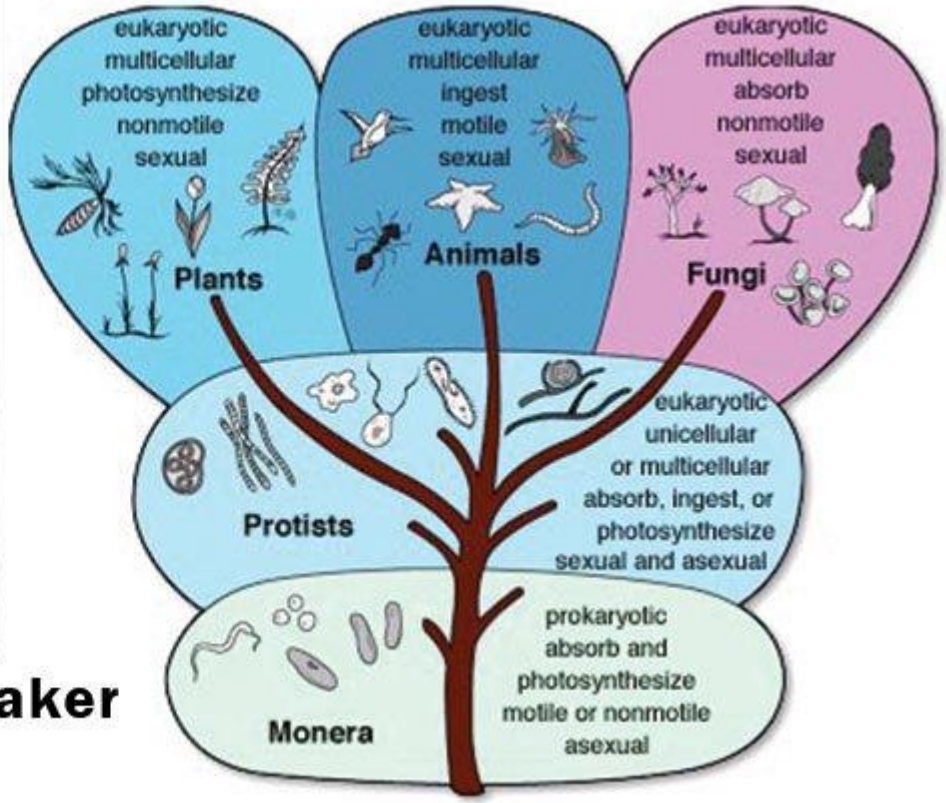
Phenetic Classification

Five-kingdom classification system

Natural classification



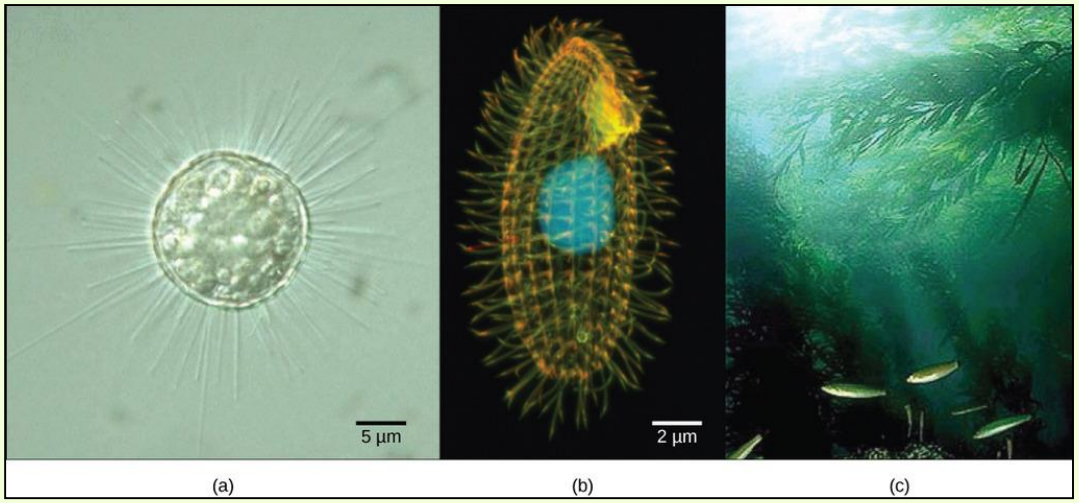
Robert Harding Whittaker
(1920 – 1980)



Phenetic Classification

The five-kingdom system, proposed in 1969 by Whittaker

- Protist varieties:
- Protists range from the microscopic, single-celled (a) *Acanthocystis turfacea* and the (b) ciliate *Tetrahymena thermophila*, both visualized here using light microscopy, to the enormous, multicellular (c) kelps (Chromalveolata) that extend for hundreds of feet in underwater "forests.



Boundless.com

Phylogenetic Classification

His results were the first to prove that all life on earth was related

Three domain of life proposed in 1977 by Carl Woese and introduced by Woese *et al.*, in 1990.

Carl Woese (pronounced woes) was one of the most significant biologists of the 20th century (July 15, 1928 – died aged 84. December 30, 2012).

The process of revisions in the Carl Woese classification scheme
A glance

Carl Woese: from scientific dissident to textbook orthodoxy

- In 1977, Carl Woese proposed that *Archaea* are different from bacteria and constitute a new super-kingdom *Archaebacteria*.
- By 1990 Woese adopted the term 'domain' for the three new branches of life and shortened the name *Archaebacteria* to *Archaea*.
- 20 years later became mainstream science (early 1990s) archaea became widely accepted by microbiologists as a third domain of life.
- And 30 years later (2007) became textbook orthodoxy
- In 2003 Woese was awarded the Crafoord Prize.

Korthof,2007

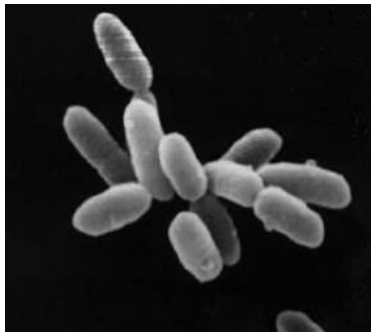
Phylogenetic Classification

The Three Domain System

The famous American microbiologist and physicist

- Because of this vast difference in genetic makeup, Carl Woese,1977 proposed that life be divided into three domains:

1. Eukaryota,
2. Eubacteria, and
3. Archaebacteria.



Cluster of halobacterium (archaea)

- He later decided that the term Archaebacteria was a misnomer (a name wrongly or unsuitably applied to a person or an object and shortened it to Archaea).
- The three domains are shown that each group is very different from the others.

Phylogenetic Classification

Three-domain classification system

Woese *et al.*,1977

- Carl Woese (pronounced woes) was one of the most significant biologists of the 20th century (July 15, 1928 – died aged 84. December 30, 2012).

1. Woese was famous for defining the Archaea (a new domain or kingdom of life) in 1977.
2. He was also the originator of the RNA world hypothesis in 1977, although not by that name.

- The scientific community was understandably shocked in the late 1970s by the discovery of an entirely new group of organisms - the Archaea.

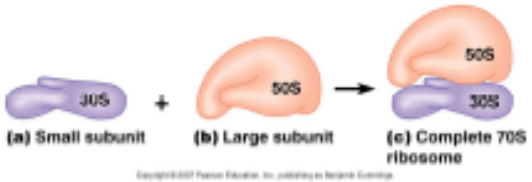
Phylogenetic Classification

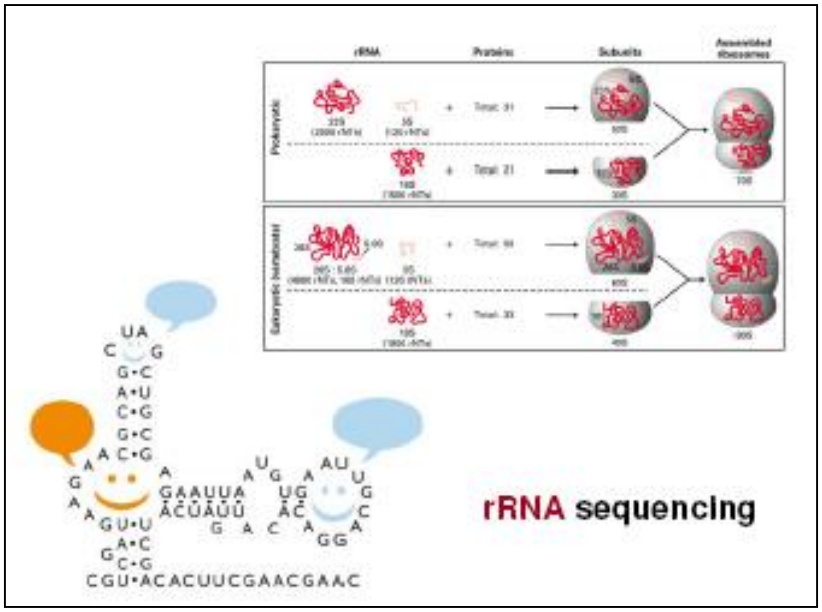
The Three Domain System

16S/18S rRNAs

Carl Woese, 1978

- Since the early 1970's, Woese and colleagues have used rRNA sequencing to determine genetic differences among cells



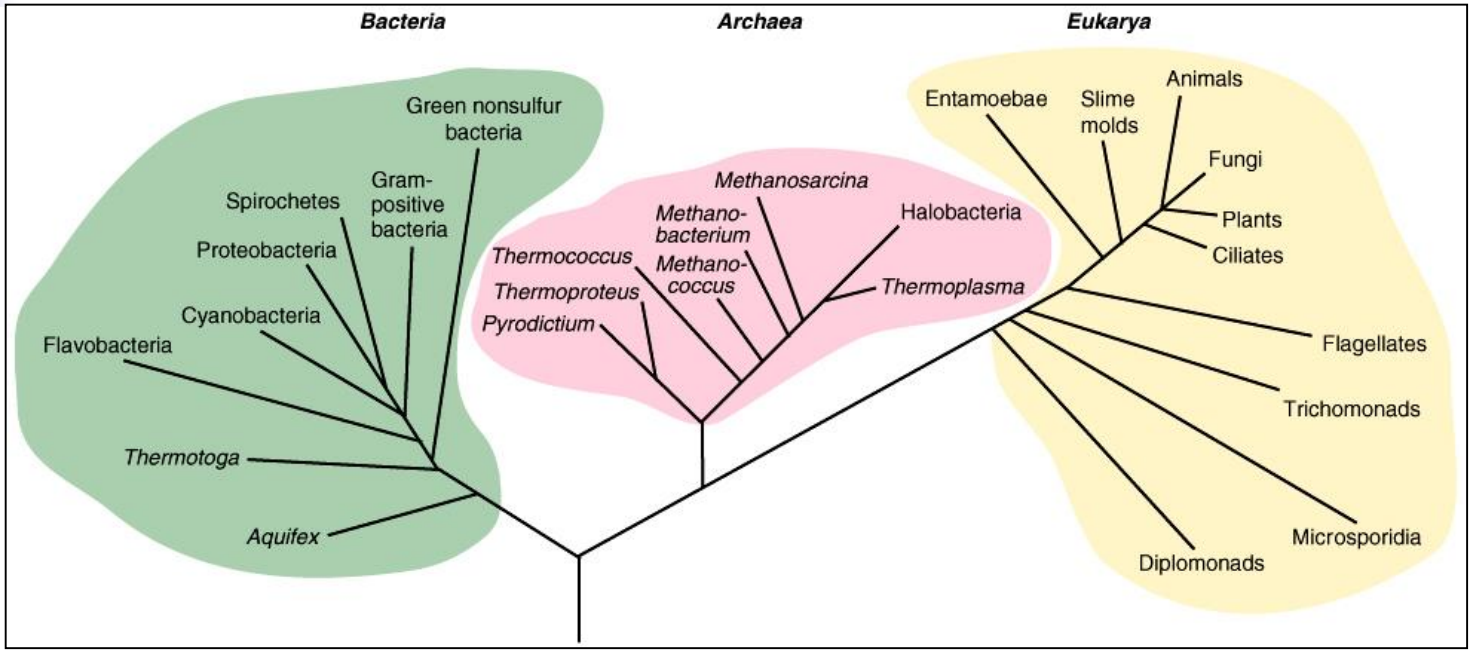


However, 40 years after his legendary classification, it becomes more and more apparent that the three-domain system is not a system carved in stone but in need of revision. In 2011, Woese told Lab Times that "serious considerations are long overdue" (LT 6-2011, p. 36-9).

Phylogenetic Classification

Three-domain classification system

Carl Woese-Late 1970's & early 1980's

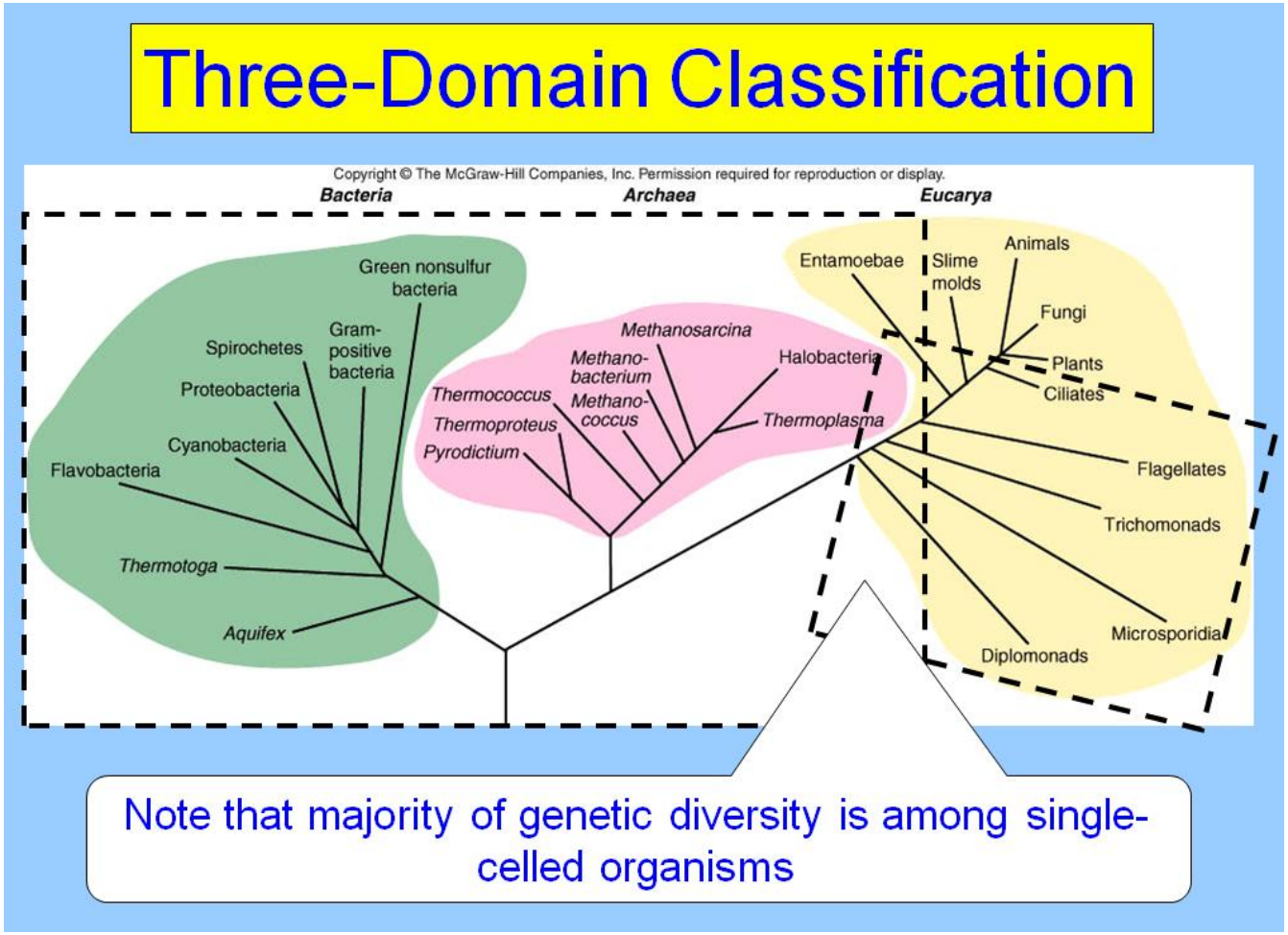


A second group (eukaryotes) is defined by the 18S rRNAs of the eukaryotic cytoplasm-animal, plant, fungal, and slime mold (Woese and Fox,1997). In 1987, Carl Woese divided the Eubacteria into 11 divisions (phyla) based on 16S ribosomal RNA (SSU) sequences, which with several additions are still used today.

Powers,2011;..

Phylogenetic Classification

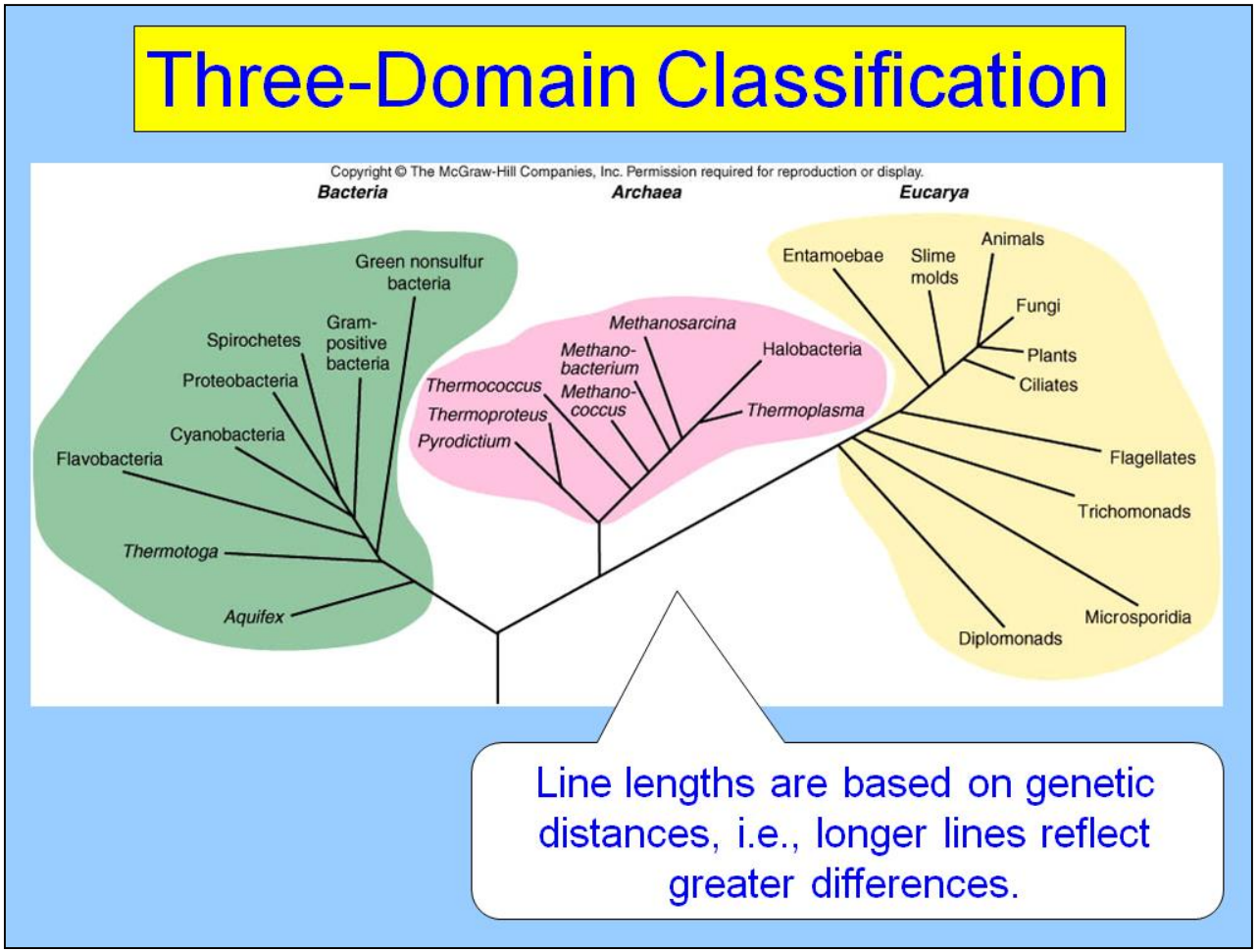
Three-domain classification system



Abedon,2011

Phylogenetic Classification

The three domain system



Abedon,2011

Phylogenetic Classification

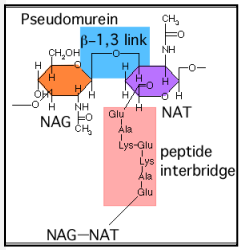
The three domain system

16S/18S rRNAs

- The three domain system, proposed by Woese,1977.
- The three domains are:

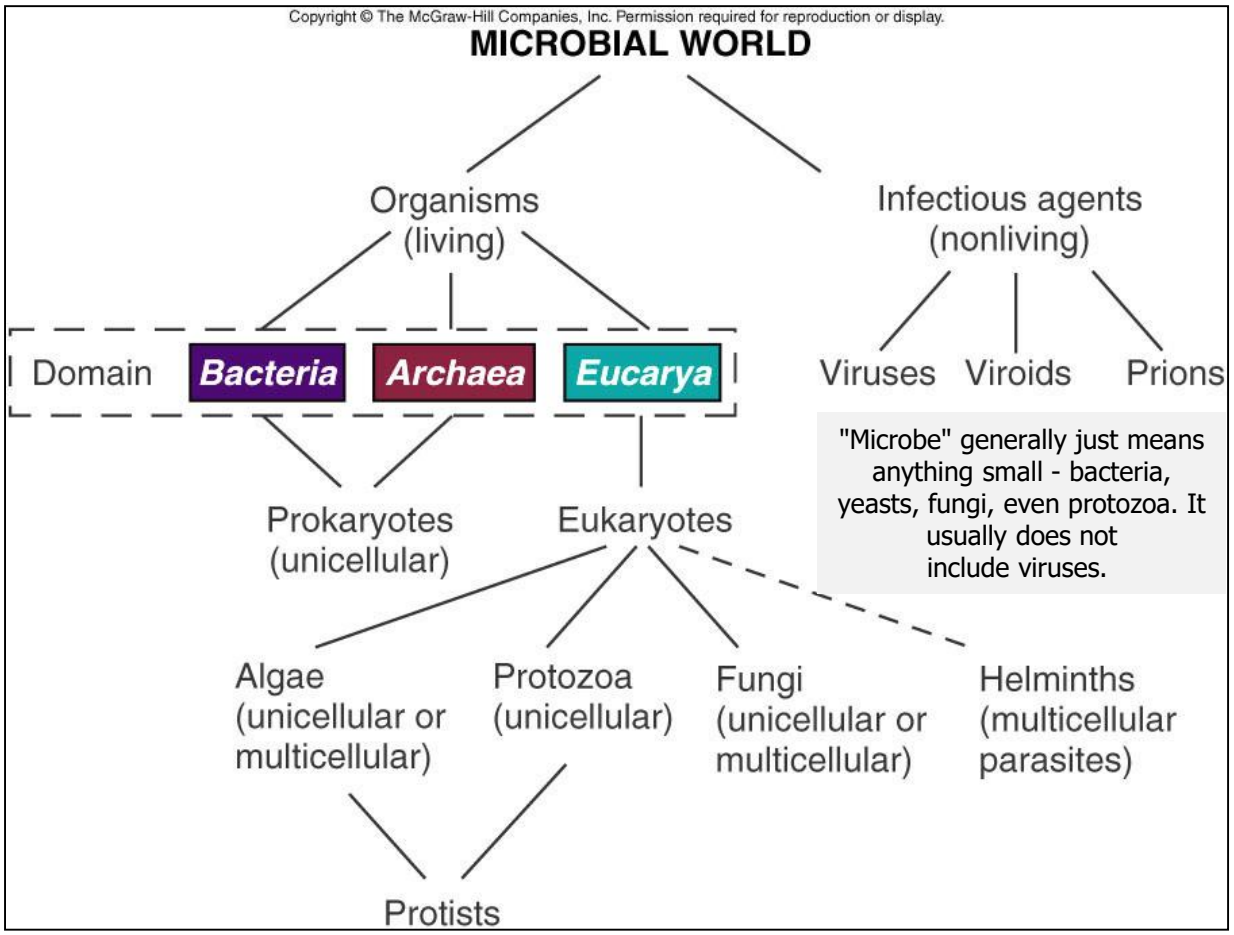
1. **Archaea** (archaebacteria): Includes all prokaryotes lacking peptidoglycan in their cell walls and living in extreme environments).
2. **Bacteria** (eubacteria): Having peptidoglycan and includes prokaryotes (pathogenic and non-pathogenic).
3. **Eukarya** (eukaryotes): Includes the 4 kingdoms of Animals, plants, fungi, protists.

Archaea cell walls are made up of pseudomurien, means they lack peptidoglycan but resembles the same chemistry and function.



Phylogenetic Classification

The three domain system



Seelke,2010

Phylogenetic Classification

The three domain system

- rRNA was chosen for phylogenetic classification because all cells have it.
- It functions the same in all cells.
- Its nitrogen base sequences is moderately conserved generation to generation.
- And, this enabled Woese to separate organisms into 3 cell types, forming a phylogenetic classification scheme.
- They proposed elevating the 3 cell types to domains (above kingdoms).

Phylogenetic Classification

The three domain system

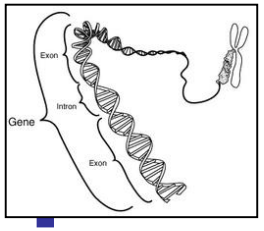
- The Three Domain System is an evolutionary model of classification based on differences in the sequences of nucleotides in the:
- This classification system divides the life based on the differences in the:

1. 16S ribosomal RNA (rRNA) structure,
2. cell's membrane lipid structure, and
3. its sensitivity to antibiotics.

- This system proposes that a common ancestor cell ("Cenancestor") gave rise to three different cell types, each representing a domain.

In 2006, Cavalier-Smith proposed that the last common ancestor to all terrestrial organisms was a non-flagellate negibacterium with two membranes.

Comparison of the main characteristics of three domains



Characteristic	Archaea	Eucarya	Bacteria
Ribosomes	70S	80S	70S
Initiator tRNA	Methionine	Methionine	Formylmethionine
Introns in tRNA	Yes	Yes	No
Membrane lipids	Ether-linked	Ester-linked	Ester-linked
Sensitivity to:			
Choramphenicol	NO	NO	Yes
Kanamycin	NO	NO	Yes
Streptomycin	NO	NO	Yes
Rifampicin	NO	NO	Yes

-Ether (ROR') linked lipids as opposed to ester linked (RCOOR') lipids.
-Introns are non-coding sequences.

Mechanism of protein synthesis
$tRNA^{fMet}$ as initiator tRNA in bacteria

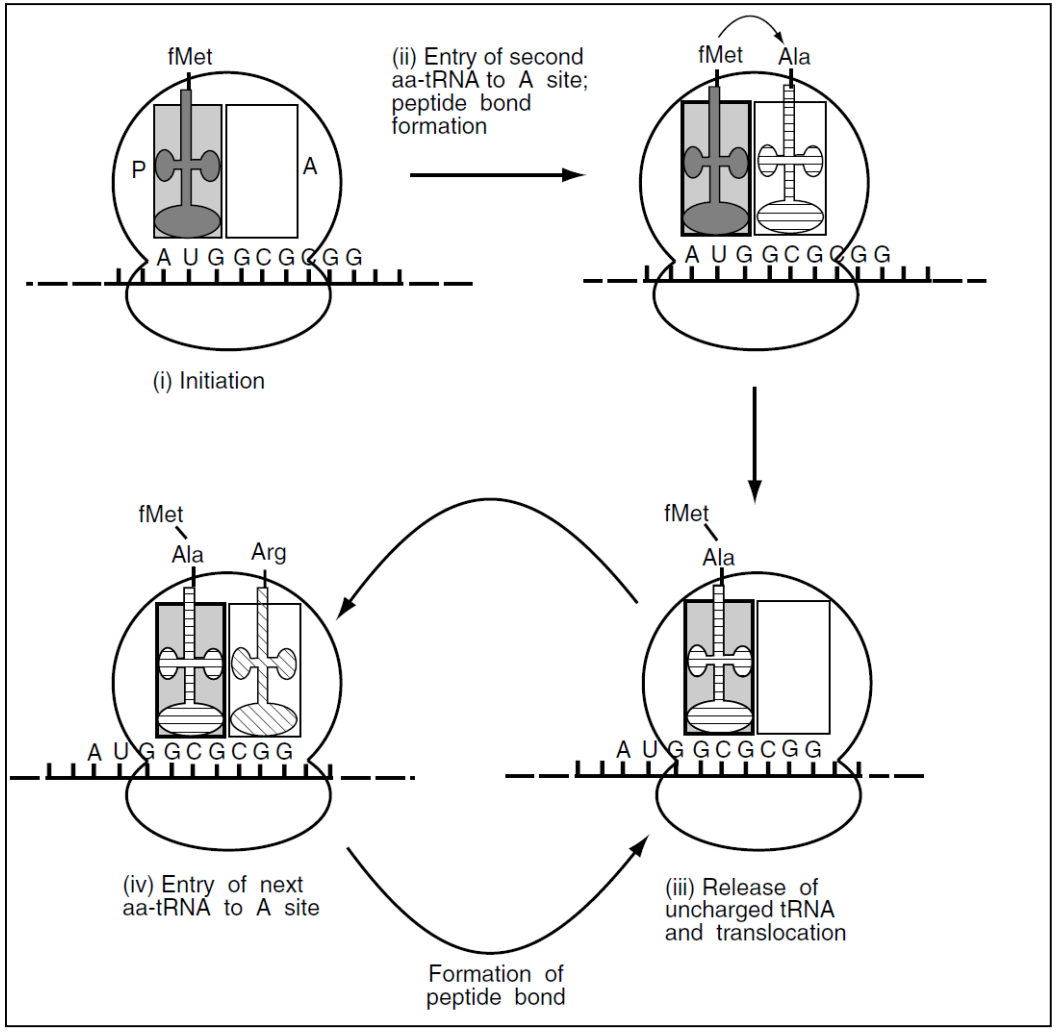
- In bacteria, the initiation codon is recognized by a specific tRNA molecule, $tRNA^{fMet}$.
- After this tRNA molecule is charged with methionine, the amino acid is modified, to N-formylmethionine.
- Aminoacylated tRNA molecules normally bind to a site on the ribosome known as the A (Acceptor) site, while their anticodon region pairs with the mRNA.
- Only after peptide bond formation is the tRNA able to move to a second site on the ribosome, the P (Peptide) site.
- The fMet-$tRNA^{fMet}$ (i.e. the $tRNA^{fMet}$ charged with formylmethionine) is unique in being able to enter the P site directly.

Mechanism of protein synthesis

$tRNA^{fMet}$ as initiator tRNA in bacteria

After this tRNA molecule is charged with methionine, the amino acid is modified, to N-formylmethionine.

- **A (Acceptor) site,**
- **P (Peptide) site.**



Dale and Park,2004

Phylogenetic Classification

The Two Empire System
Cavalier-Smith megaclassification of prokaryotes (life).

Thomas Cavalier-Smith(October 21, 1942-19 March 2021 (aged 78) was a Professor of Evolutionary Biology in the Department of Zoology, at the University of Oxford.
He is well known for his series of classification system of all organisms.

The process of revisions in the Cavalier-Smith's classification scheme

A glance

- His classification has been a major foundation in modern taxonomy, particularly with revisions and reorganizations of kingdoms and phyla.
- In 1981, he proposed that by completely revising Robert Whittaker's Five Kingdom system, there could be nine kingdoms (instead of 5).
- In 1993, he revised his system particularly in the light of the general acceptance of Archaebacteria as separate group from Bacteria.
- By 1998 and 2004 Cavalier-Smith had reduced the total number of Kingdom from eight to six.
- **Rooting the tree of life**
- In 2006, Cavalier-Smith proposed that the last universal common ancestor to all life was a non-flagellate negibacterium with two membranes.

Summary of the sequence from the two-kingdom system up to Cavalier-Smith's six-kingdom system

- In 1866, Ernst Haeckel proposed the addition of a third kingdom- Protista.
- In 1925, Edouard Chatton recognized the distinction between prokaryotes and eukaryotes and proposed a two empire system.
- In 1938, Herbert Copeland incorporated prokaryotic cells into a fourth kingdom- Monera.
- In 1969, Robert Whittaker proposed a five kingdom system, which included a kingdom for Fungi.
- In 1977, Carl Woese identified differences in prokaryotes and in 1990 proposed the currently recognized three domain system.

Summary of the sequence from the two-kingdom system up to Cavalier-Smith's six-kingdom system

Linnaeus 1735	Haeckel 1866	Chatoon 1925	Copeland 1938	Whittaker 1969	Woese et al. 1977	Woese et al. 1990	Cavalier-Smith 2004
2 kingdoms	3 kingdoms	2 empires	4 kingdoms	5 kingdoms	6 kingdoms	6 kingdoms	6 kingdom
	Protista	Prokaryota	Monera	Monera	Eubacteria	Bacteria	Bacteria
Vegetabila	Plantae	Euokaryota	Protista	Protista	Archaebacteria	Archaea	Protozoa
Animalia	Animalia		Plantae	Fungi	Protista	Protista	Chromista
			Animalia	Plantae	Fungi	Fungi	Fungi
				Animalia	Plantae	Plantae	Plantae
					Animalia	Animalia	Animalia

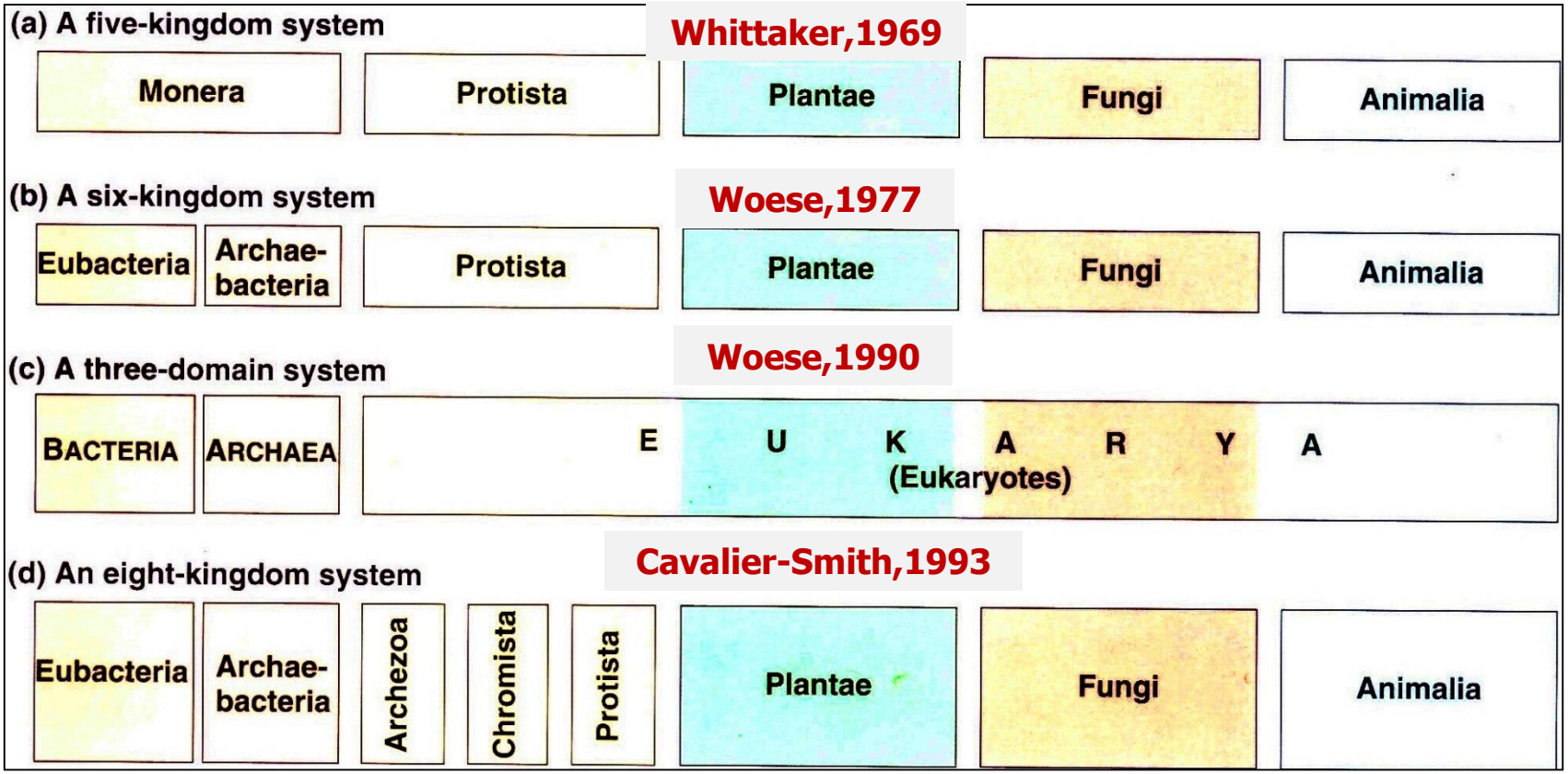
The word 'procariotique' was independently coined in 1925 by Edouard Chatton for cyanobacteria and all other bacteria including archaebacteria proposed by Woese *et al.*,1997.

Wikipedia,2011;..

Summary of the sequence from the two-kingdom system up to Cavalier-Smith's six-kingdom system

Whittaker 1969	Woese et al. 1977	Woese et al. 1990	Cavalier-Smith 1993	Cavalier-Smith, 1998 2004
Kingdoms	**Domains Eubacteria, Archaebacteria, Eukaryota**	**Domains Bacteria, Archaea, Eukarya**	**Empire Prokaryota and Eukaryota**	**Empire Prokaryota and Eukaryota**
5 kingdoms	6 kingdoms	6 kingdoms	8 kingdoms	6 kingdom
Monera	Eubacteria	Bacteria	Eubacteria	Bacteria
Protista	Archaebacteria (ancient bacteria)	Archaea (he discovered third domain of life)	Archaebacteria	Protozoa
Fungi	Protista	Protista	Archezoa	Chromista (he creates a new kingdom)
Plantae	Fungi	Fungi	Protozoa	Fungi
Animalia	Plantae	Plantae	Chromista	Plantae
	Animalia	Animalia	Plantae	Animalia
			Fungi	
			Animalia	

Summary of the sequence from the five-kingdom (whittaker, 1969) system up to Cavalier-Smith's eight-kingdom system (1993)



Archezoa such as Archamoebae have been developed from mitochondriate ancestors. The kingdom Archezoa has been abandoned.

The Two Empire System

Cavalier-Smith megaclassification of prokaryotes (life)

Two Empires: Prokaryota and Eukaryota and six kingdoms

- Some authors have opposed the three domain system proposed by Woese *et al.*,1987 and 1990 due to various reasons.
- One prominent scientist which opposes the three domain system is Thomas Cavalier-Smith.
- Cavalier-Smith,2002 put it, with "mega-classification" of prokaryotes, focusing on taxonomy of higher ranks.

Jones and Bartlett Publishers,2007;Zou *et al.*,2012

The Two Empire System

Cavalier-Smith megaclassification of prokaryotes (life)

Two Empires: Prokaryota and Eukaryota

- An empire is a taxonomic group used in Cavalier-Smith's six-kingdom classification system (1998).
- Cavalier-Smith emphasizes the fundamental difference in cell structure between prokaryotes and eukaryotes by creating two supergroupings called empires:

1. Prokaryota,
2. Eukaryota.

He has also published prodigiously on issues such as the origin of various cellular organelles (including the nucleus, mitochondria), genome size evolution, and endosymbiosis.

Jones and Bartlett Publishers,2007;Wikipedia,2012

The Two Empire System

Cavalier-Smith megaclassification of prokaryotes(life)

Two Empires: Prokaryota and Eukaryota and six kingdoms

1. The kingdom Bacteria is in empire Prokaryota;
2. The other kingdoms are in empire Eukaryota.
3. He also creates a new kingdom called Chromista.

- The latter kingdom (chromista) contains photosynthetic organisms (e.g., diatoms, brown algae, water moulds, etc.), whose chloroplasts differ from those found in plants.

Jones and Bartlett Publishers,2007;Wikipedia,2014

The Two Empire System

Cavalier-Smith megaclassifcation of life

Two Empires: Prokaryota and Eukaryota and six kingdoms

- Prokaryotes constitute a single kingdom, Bacteria.
- Bacteria is divided into two new subkingdoms:

1. Subkingdoms Negibacteria, with a cell envelope of two distinct genetic membranes.

- **Negibacteria as a root of the universal tree**

2. Subkingdom Unibacteria, comprising the new phyla Archaebacteria and Posibacteria.

- Other new bacterial taxa are established in a revised higher-level classification that recognizes only:

1. 8 phyla, and
2. 9 classes.

The Two Empire System

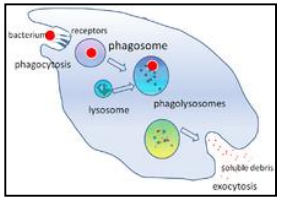
Cavalier-Smith megaclassification of life

Two Empires: Prokaryota and Eukaryota and six kingdoms

- Thomas Cavalier-Smith,2002 takes into account many phenotypic characteristics, and is not sequence-based.
- Cavalier-Smith basically concludes that:

1. Double-membraned Gram-negative bacteria (he calls them Negibacteria) lie near the root of the bacterial tree (3700 Mya), and
2. That the Archaea and Eucarya are relatively recent (850 Mya) emergents from a line that also gave rise to the modern Gram-positive bacteria and actinobacteria.
3. In 2006, Cavalier-Smith proposed that the last common ancestor to all terrestrial organisms was a non-flagellate negibacterium with two membranes.

The Two Empire System

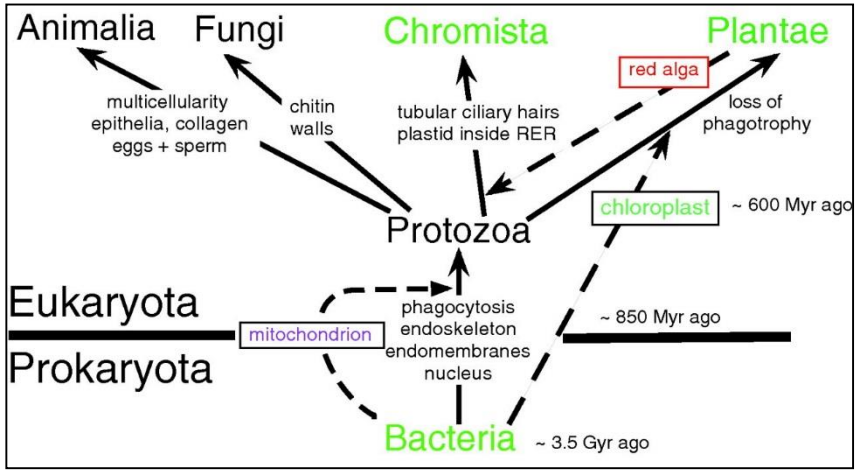


Cavalier-Smith megaclassification of life

Two Empires: Prokaryota and Eukaryota and six kingdoms

- **The six-kingdom, two-empire classification of life.**

1. Three major lineage mergers (symbiogeneses involving cell enslavement after phagocytic engulfment) are shown as dashed lines;
2. Four additional mergers that transferred chloroplasts from green plants or chromists into different protist lineages to make novel kinds of algae (Cavalier-Smith 2007c).



Phagocytes are cells that protect the body by ingesting harmful foreign particles, bacteria, and dead or dying cells.

The abbreviations Gya (giga-annum) or bya are for billion years ago.

Cavalier Smith,2010

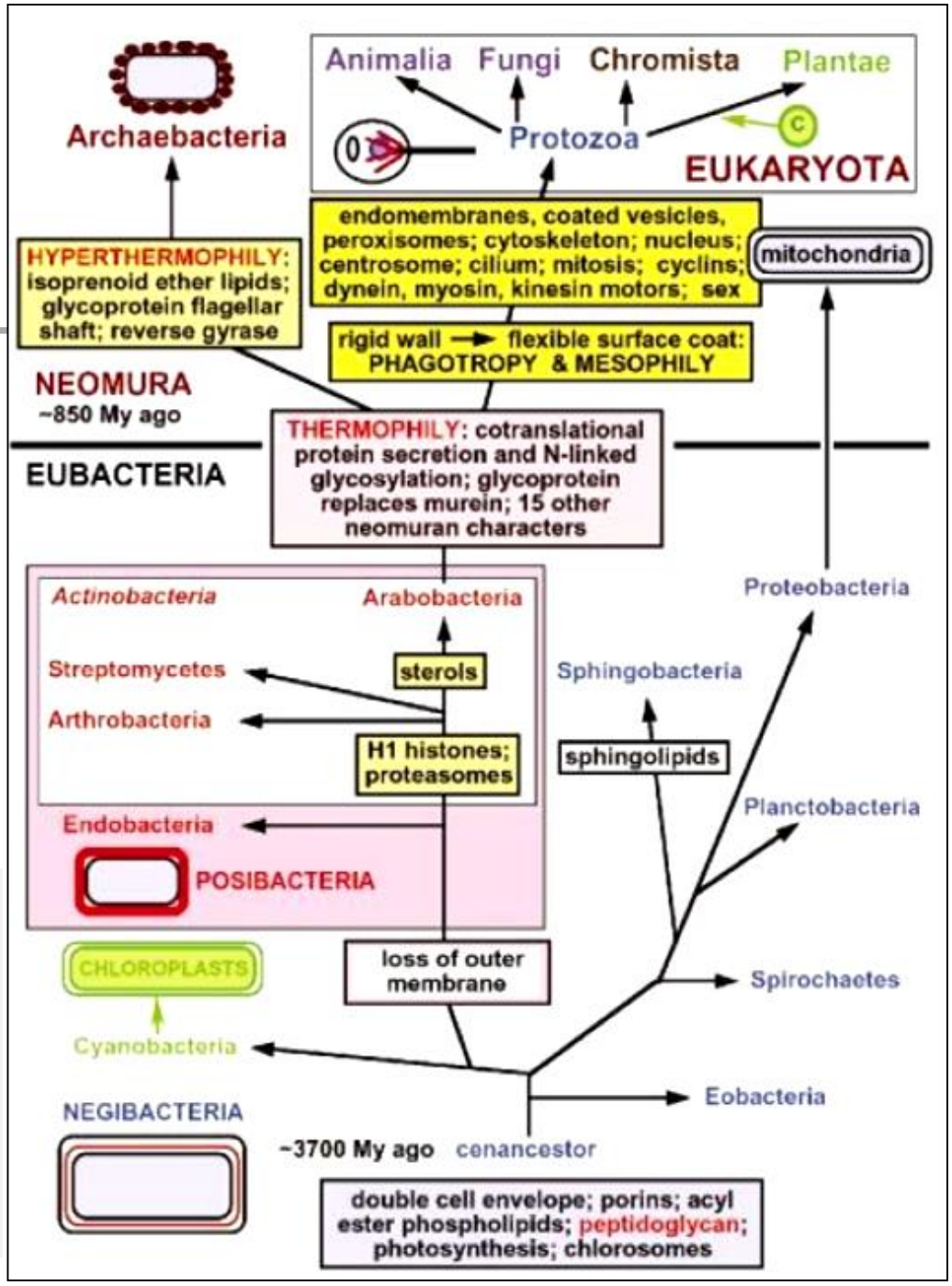
Terminology

- **Arbobacteria** (insect-vectored bacteria) are pathogens Transmittable by Arthropods (insects). e.g. phytoplasmas, "Candidatus Liberibacter" and Rickettsia sp.,..
- **Arthrobacteria** are coryneform bacteria.
- **Sterols**, the third lipid class.
- **Protozoa** (also **protozoan**, plural protozoans) is an informal term for a group of single-celled eukaryotes
- **Rhizaria** have been broadly divided into core Cercozoa, Endomyxa,.. It is a diverse collection primarily of free-living **protozoan** organism.
- **Glidobacteria,** the new infrakingdom including Chlorobacteria, Hadobacteria, Cyanobacteria and excluding Actinobacteria and Firmicutes.
- **Adobacteria** such as Deinococci.
- **Planctobacteria,** the origin of eukaryotic and archaebacterial cells. Cavalier-Smith considered as phylum but later it was considered as a superphylum.
- The **PVC group** is a superphylum of bacteria named after its three important members, Planctomycetes, Verrucomicrobia, and Chlamydiae.

The bacterial origins of eukaryotes as a two-stage process

Neomura (new wall) is a clade composed of Archaea and Eukaryota.
The Archaea and the Eukaryotes stem from Gram positive bacteria (Posibacteria), which in turn derive from gram negative bacteria (Negibacteria).

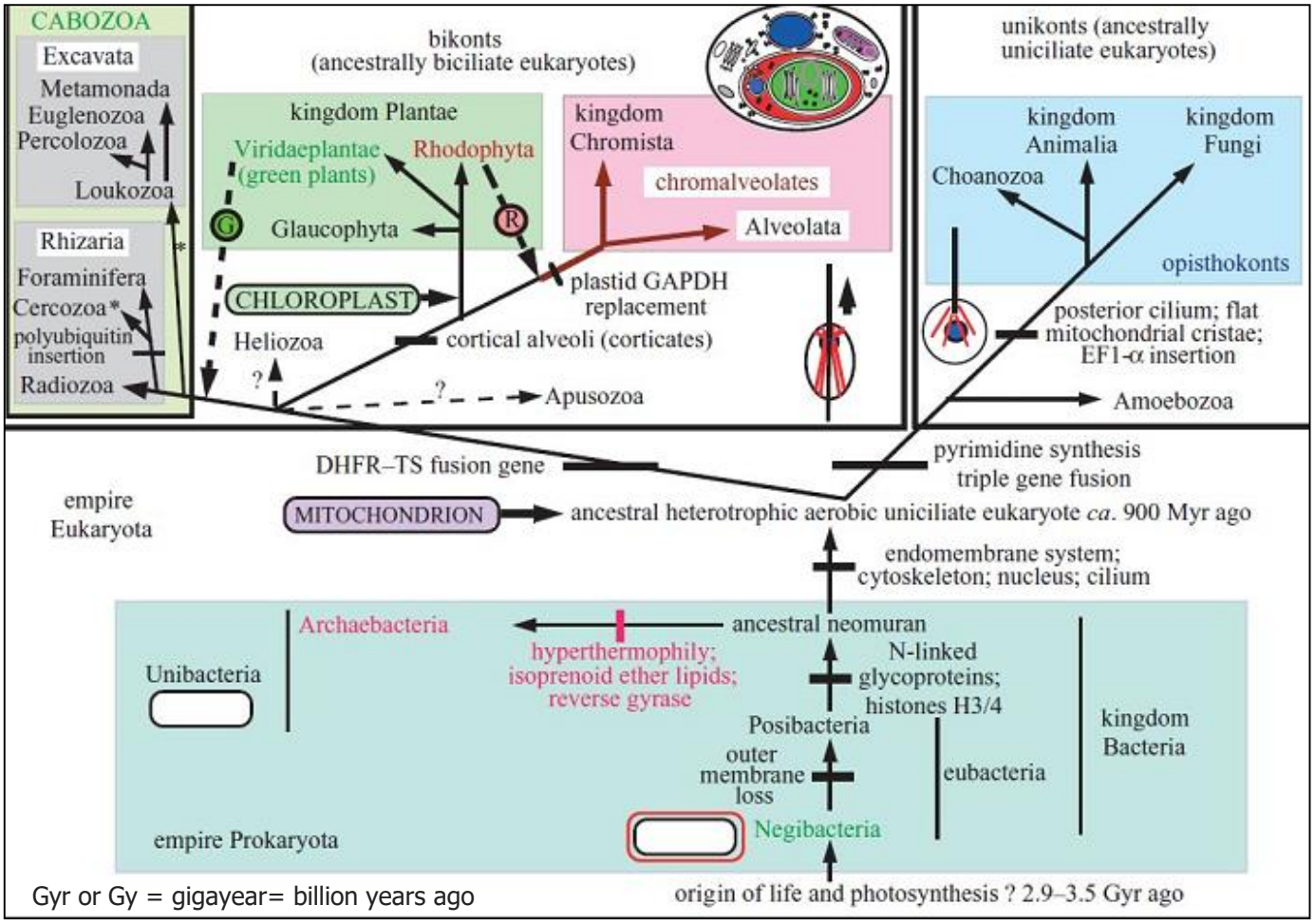
Neomura means "new walls.
The adjectival form is Neomuran, and
A single individual from the group is called a Neomuran.
The cholorosome contains bacteriochlorophyll c, d, or e.
Reverse gyrase is an enzyme that introduces positive supercoils into covalently closed DNA.
Sterols, the third lipid class.
Acyl lipids



The Two Empire System

Cavalier-Smith megaclassification of life

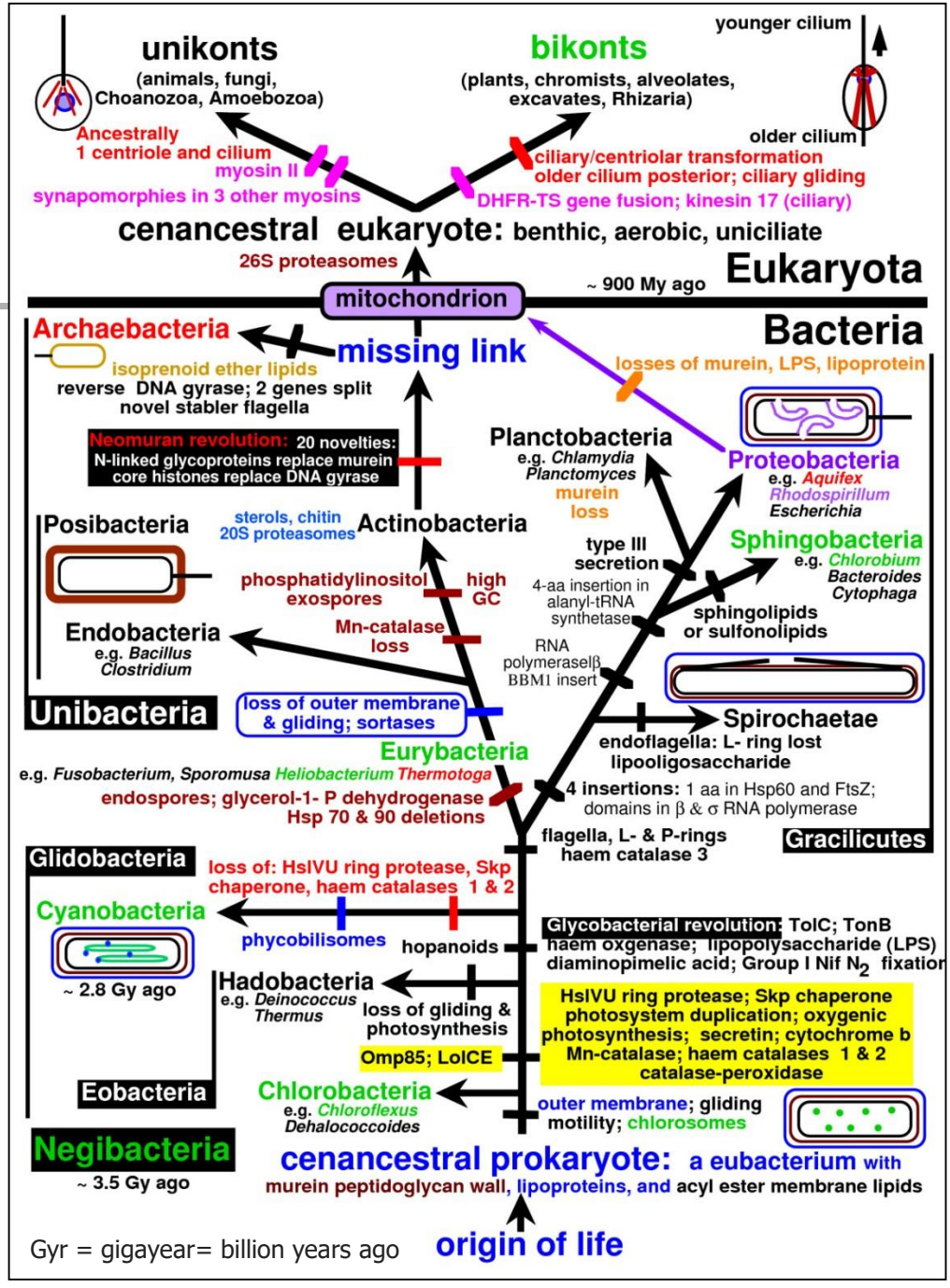
Two Empires: Prokaryota and Eukaryota



Cavalier Smith,2004

The rooted tree of life

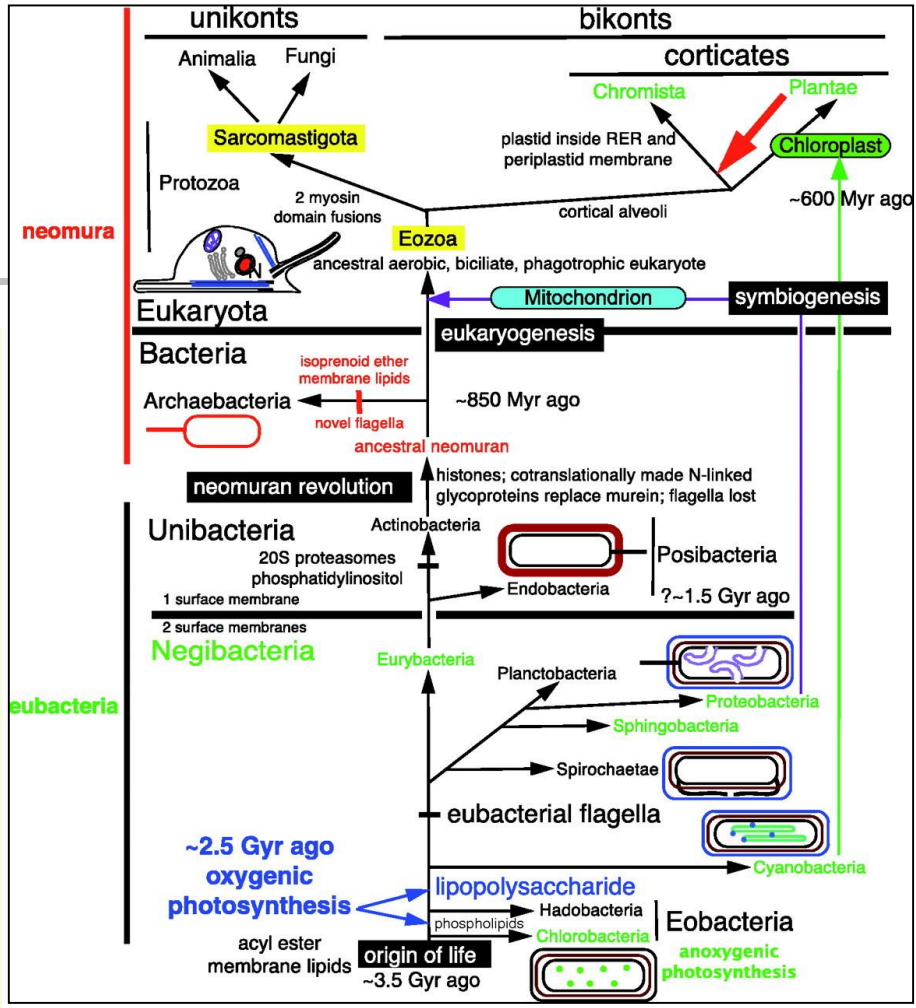
- The new negibacterial infrakingdom Gracilicutes segregates four phyla from the other negibacteria.
- Eubacteria must be significantly older than eukaryotes.
- The age of ~900 My for eukaryotes is based on a recent Bayesian analysis of 143 proteins multiply calibrated from the fossil record.



Cavalier-Smith,2006

The bacterial origins of eukaryotes as a two-stage process

- **Intracellular coevolutionary theory:**
- The last common ancestor of eukaryotes was a sexual phagotrophic protozoan with mitochondria.
- The eukaryotic cytoskeleton and endomembrane system originated through cooperatively enabling the evolution of phagotrophy.
- Eukaryotes plus their archaebacterial sisters form the clade Neomura.



Bikont is a eukaryotic cell with two flagella; thought to be the ancestor of all plants while **unikont** is a eukaryotic cell with a single flagellum; thought to be the ancestor of all animals.

Cavalier-Smith,2010

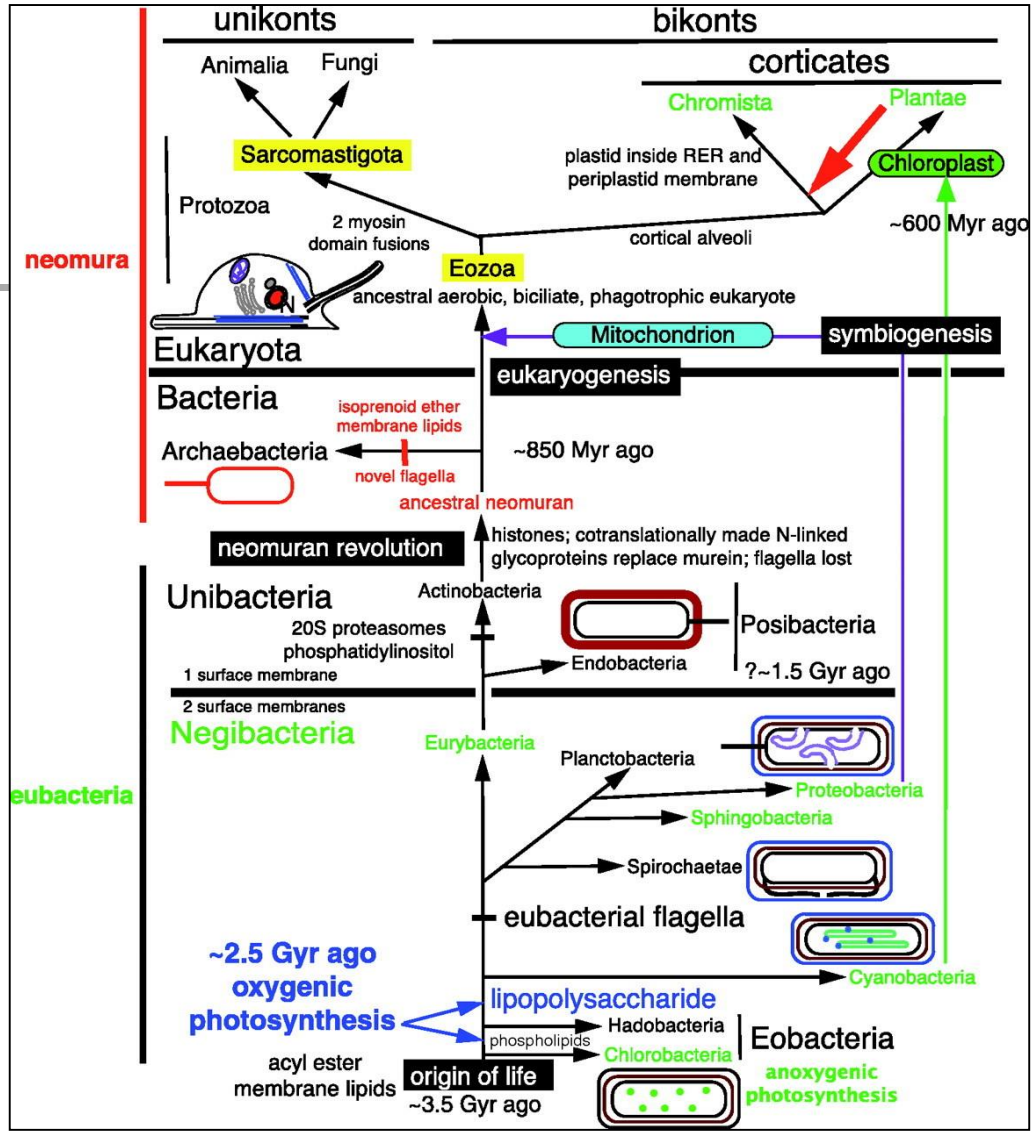
The bacterial origins of eukaryotes as a two-stage process

Neomura (new wall) is a clade composed of Archaea and Eukaryota.

The Archaea and the Eukaryotes stem from Gram positive bacteria (Posibacteria), which in turn derive from gram negative bacteria (Negibacteria).

Neomura means "new walls.
The adjectival form is Neomuran, and
A single individual from the group is called a Neomuran.
The cholorosome contains bacteriochlorophyll c, d, or e.
Reverse gyrase is an enzyme that introduces positive supercoils into covalently closed DNA.

Cavalier-Smith,2010



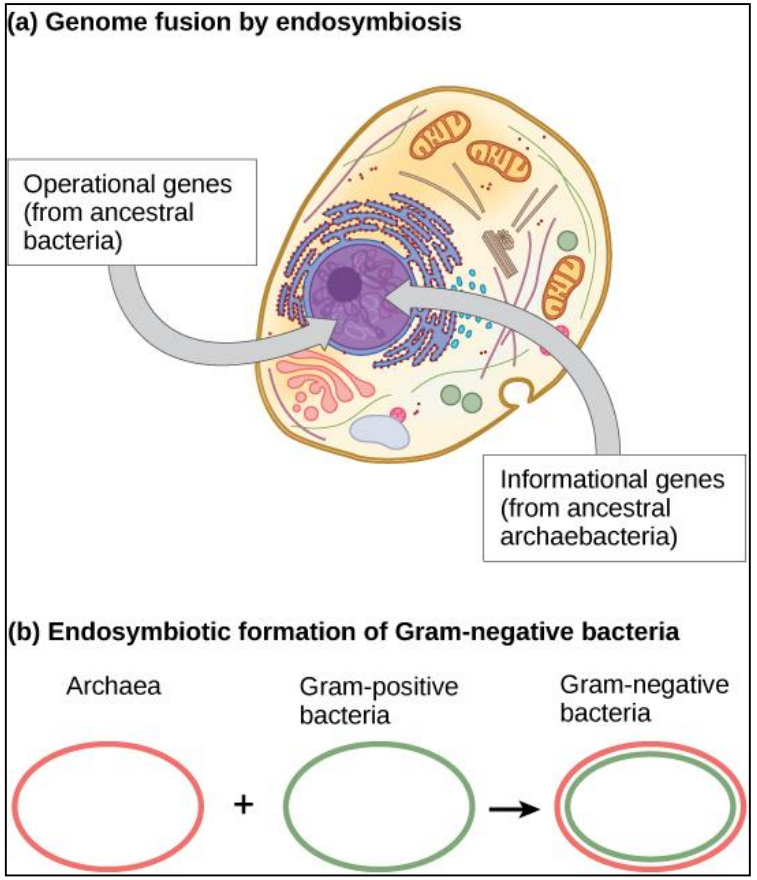
Bikont is a eukaryotic cell with two flagella; thought to be the ancestor of all plants while **unikont** is a eukaryotic cell with a single flagellum; thought to be the ancestor of all animals.

Endosymbiosis in eukaryotes

- The theory that mitochondria and chloroplasts are endosymbiotic in origin is now widely accepted.
- More controversial is the proposal that

a) the eukaryotic nucleus resulted from the fusion of archaeal and bacterial genomes; and that

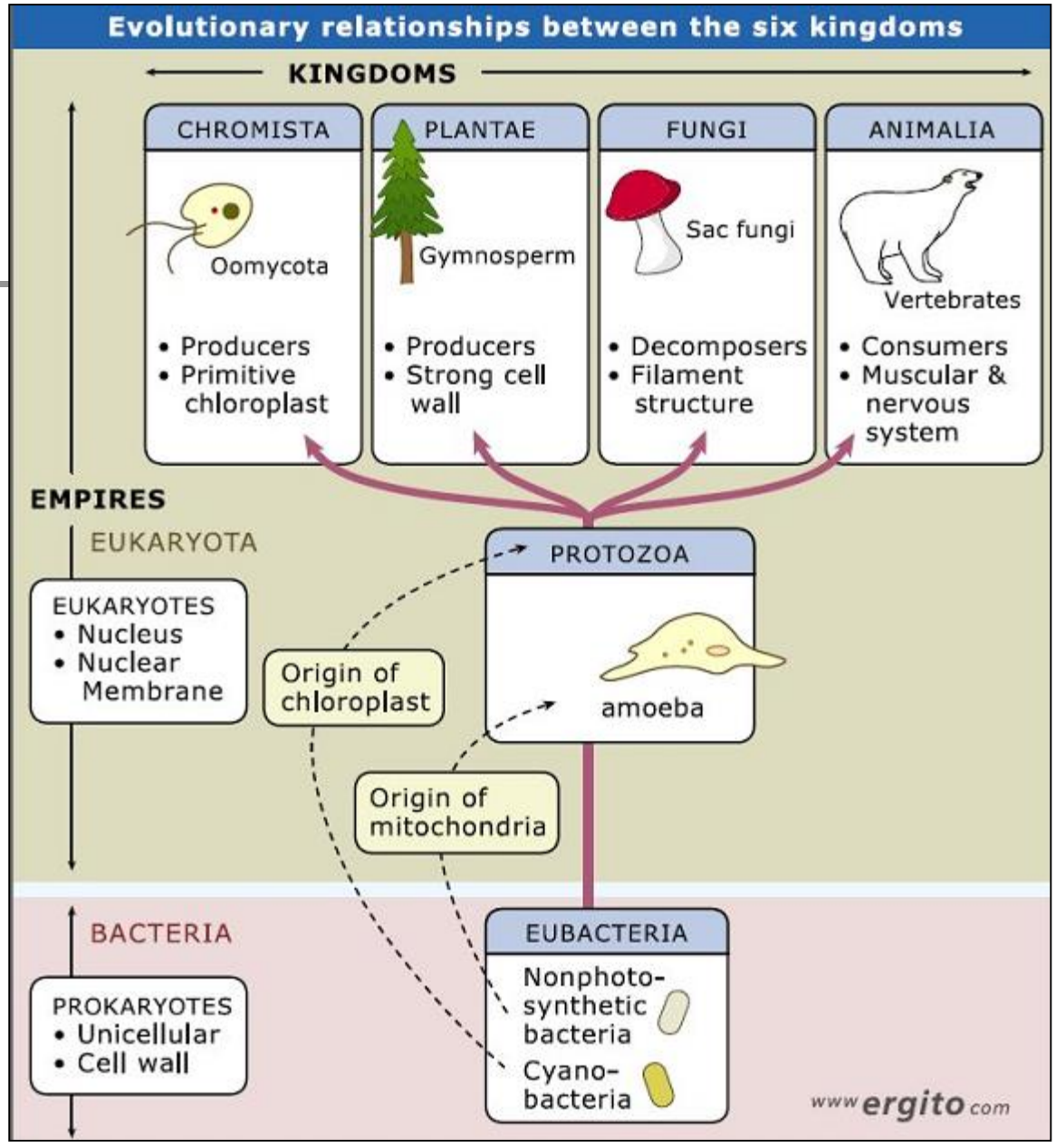
b) Gram-negative bacteria, which have two membranes, resulted from the fusion of Archaea and Gram-positive bacteria, each of which has a single membrane.



The Two Empire System:
Prokaryota and Eukaryota

Possible evolutionary relationships are indicated by branching patterns.

Cyanobacteria are relatives of the bacteria but not eukaryotes. Because they are photosynthetic and aquatic, cyanobacteria are often called "blue-green algae". Chloroplasts share a common ancestor with the cyanobacteria.

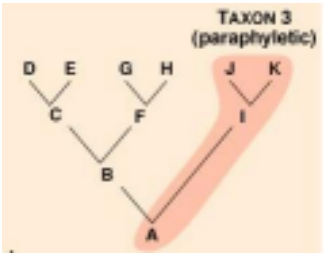


Jones and Bartlett Publishers,2007

The Two Empire System

Cavalier-Smith megaclassifcation

Logical arguments



- Though fairly well known, many of his claims have been controversial and have not gained widespread acceptance in the scientific community to date.
- Most recently, he has published a paper citing the paraphyly of his bacterial kingdom, the origin of Neomura from Actinobacteria and taxonomy of prokaryotes.

The arrangement of the members of a paraphyletic group is called a paraphyly. Paraphyletic group, a group of organisms which contains a common ancestor and some, but not all, of the descendants.

Wikipedia,2020

The Two Empire System

Cavalier-Smith megaclassifcation

Logical arguments

- Thomas Cavalier-Smith schemes based on several logical arguments, which are highly controversial and generally disregarded by molecular biology community and are often not even mention in reviews due to the subjective nature of the assumptions made for logical arguments.
- In total, his views have been influential but controversial, and not always widely accepted.

Wikipedia,2012;..

The Fourth natural classification scheme Ruggiero *et al.*,2015

Two superkingdoms: Prokaryota and Eukaryota and seven kingdoms

- Cavalier-Smith and his collaborators revised the classification in 2015, and published it in PLOS ONE journal.
- In this scheme they reintroduced the division of prokaryotes into two kingdoms:

1. Bacteria (=Eubacteria), and
2. Archaea (=Archebacteria).

- This is based on the consensus in the:

1. Taxonomic Outline of Bacteria and Archaea (TOBA), and
2. The Catalogue of Life.

Ruggiero, M.A., D.P. Gordon, T.M. Orrell, N. Bailly, T. Bourgoin, R.C. Brusca, T. Cavalier-Smith, M.D. Guiry and P.M. Kirk. 2015. Correction: A Higher Level Classification o f All Living Organisms. PLoS ONE 10 (6): e0130114. doi:10.1371/journal. pone.0130114.

The Fourth natural classification scheme Ruggiero *et al.*,2015

Two superkingdoms: Prokaryota and Eukaryota and seven kingdoms

- We are proposing a **two-superkingdom** (Prokaryota and Eukaryota), seven-kingdom classification that is a practical extension of Cavalier-Smith's six-kingdom schema(1998).
- Our schema includes:
- **The prokaryotic kingdoms:**

1. Archaea (Archaebacteria), and
2. Bacteria (Eubacteria), and

- **The eukaryotic kingdoms:**

1. Protozoa,
2. Chromista,
3. Fungi,
4. Plantae, and
5. Animalia.

Michael Ruggiero

Ruggiero, M.A., D.P. Gordon, T.M. Orrell, N. Bailly, T. Bourgoin, R.C. Brusca, T. Cavalier-Smith, M.D. Guiry and P.M. Kirk. 2015. Correction: A Higher Level Classification o f All Living Organisms. PLoS ONE 10(6): e0130114. doi:10.1371/journal. pone.0130114.

Ruggiero *et al.*,2015

Two superkingdoms: Prokaryota and Eukaryota and seven kingdoms

Linnaeus 1735	Haeckel 1866	Chatoon 1925	Copeland 1938	Whittaker 1969	Woese *et al.*, 1977	Woese *et al.*, 1990	Cavalier-Smith,1993	Cavalier-Smith,1998 &2004	Ruggiero *et al.*, 2015
2 kingdoms	3 kingdoms	2 empires	4 kingdoms	5 kingdoms	6 kingdoms	3 domains	8 kingdoms	6 kingdoms	7 kingdoms
(not treated)	Protista	Prokaryota	Mychota	Monera	Bacteria	Eubacteria	Eubacteria	Bacteria	Bacteria
					Archaebacteria	Archaea	Archaebacteria		Archaea
		Euokaryota	Protoctista	Protista	Protista	Eukarya	Archezoa	Protozoa	Protozoa
							Protozoa		
							Chromista	Chromista	Chromista
Vegetabila	Plantae		Plantae	Plantae	Plantae		Plantae	Plantae	Plantae
				Fungi	Fungi		Fungi	Fungi	Fungi
Animalia	Animalia		Animalia	Animalia	Animalia		Animalia	Animalia	Animalia

Wikipedia,2017

The major revisions of Cavalier-Smith's classification schemes

Eight to six and lastly seven kingdom systems

Cavalier-Smith, 1993	Cavalier-Smith,1998 & 2004	Ruggiero *et al.*, 2015
8 kingdoms	6 kingdoms	7 kingdoms
Eubacteria	Bacteria	Bacteria
Archaebacteria		Archaea
Archezoa	Protozoa	Protozoa
Protozoa		
Chromista	Chromista	Chromista
Plantae	Plantae	Plantae
Fungi	Fungi	Fungi
Animalia	Animalia	Animalia

Cavalier-Smith and his collaborators revised the classification in 2015 (Ruggiero *et al.*,2015).

Wikipedia,2020

Ruggiero *et al.*,2015

Two superkingdoms: Prokaryota and Eukaryota and seven kingdoms

Logical arguments

- The proposed classification of prokaryotes and eukaryotes which was published by Ruggiero *et al.*,2015 may be:

1. reasonable in their classification of the prokaryotes,
2. the eukaryote section does not pass standards of modern biology.

Adl *et al.*,2018

Ruggiero *et al.*,2015

Two superkingdoms: Prokaryota and Eukaryota and seven kingdoms

Logical arguments

- It is definitely not the only possible classification, and not the consensus one.
- Based on the same phylogenetic scheme, several equally "correct" classifications can be drafted.
- In general, I see no particular benefit in this kind of detailed classifications: many groups can be treated differently, in different ranks and circumscriptions.
- There are many alternative classification schemes for particular large groups. As to plants, I have many questions regarding the selected approach and accepted taxa and "taxa".

Ruggiero *et al.*,2015

Two superkingdoms: Prokaryota and Eukaryota and seven kingdoms

Logical arguments

- I do not understand why one need all the time to redo taxonomy? It only confuses researchers and gives nothing new to biology.
- In recent years, the system of organisms altered many times.
- Now it is almost impossible to determine not only the species and the genus of the specimens, but more higher rank up to kingdom.
- Anyway Nature is richer than any classification.
- It is better don't change the classification and use of the old one, more or less established.

Some recent natural classification scheme
The universal tree of life: an update
Based on ribosomal protein trees

- **Universal multiprotein ribosomal protein (RP) trees- Offer markedly higher resolution for prokaryote phylogeny than rDNA:**

1. Forterre,2015
2. Kovacs *et al*.,2018
3. Cavalier-Smith and Chao,2020

Some recent natural classification scheme Based on ribosomal protein trees

Forterre,2015

- Rooting of the domain Bacteria and internal branching in this domain have been adapted from the ribosomal protein trees of Koonin and colleagues (Yutin *et al.*,2012).
- These authors have suggested several superphyla beside the previously recognized PVC superphylum, which includes:
- Planctomycetes,
- Verrucomicrobia, and
- Chlamydiae.

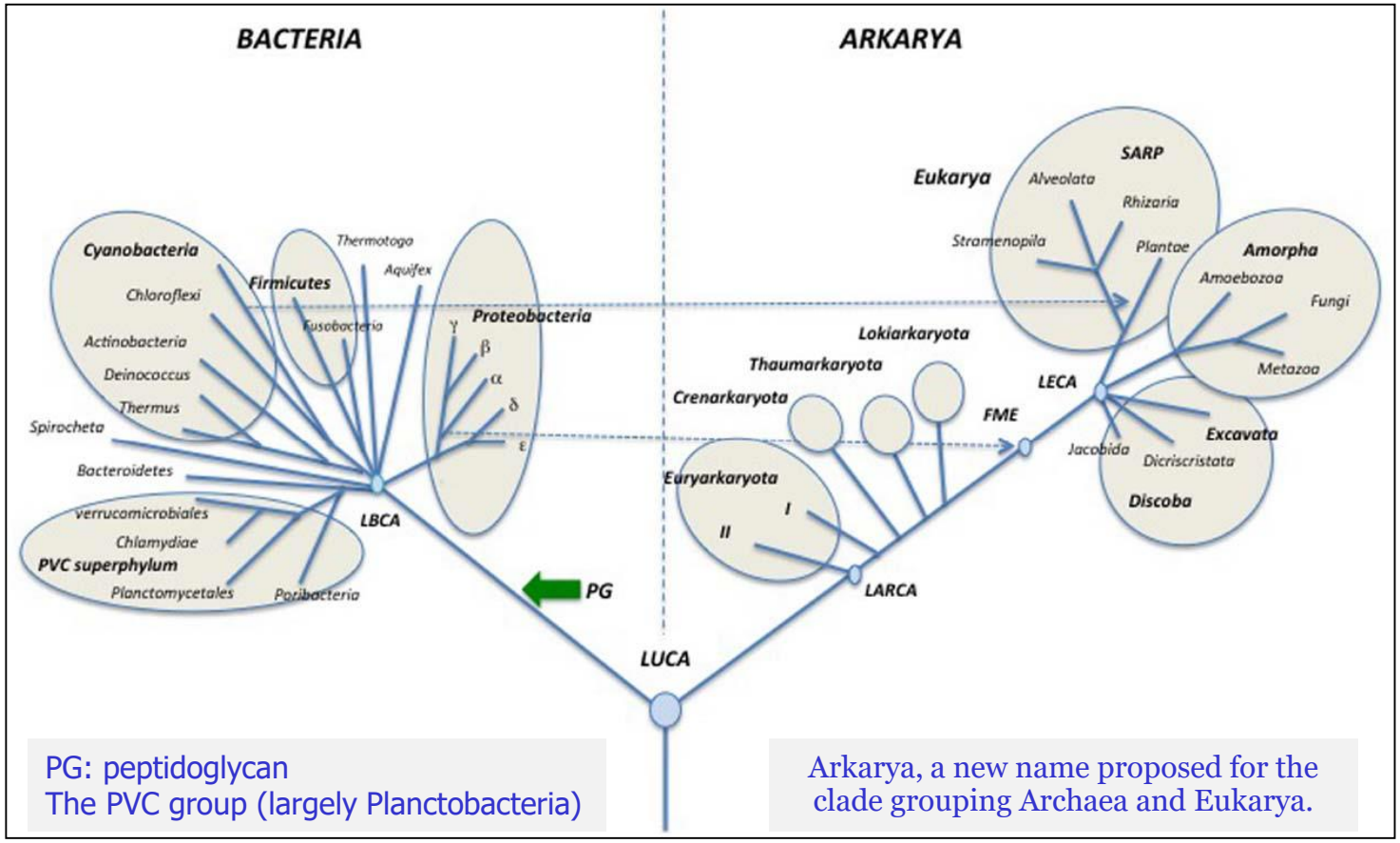
Cavalier-Smith calls the PVC group as Planctobacteria and considers it a phylum and not superphylum.

Forterre,2015,Wikepdia,2020

The universal tree of life: an update

Universal multiprotein ribosomal protein (RP) trees

Offer markedly higher resolution for prokaryote phylogeny than rDNA



PG: peptidoglycan
The PVC group (largely Planctobacteria)

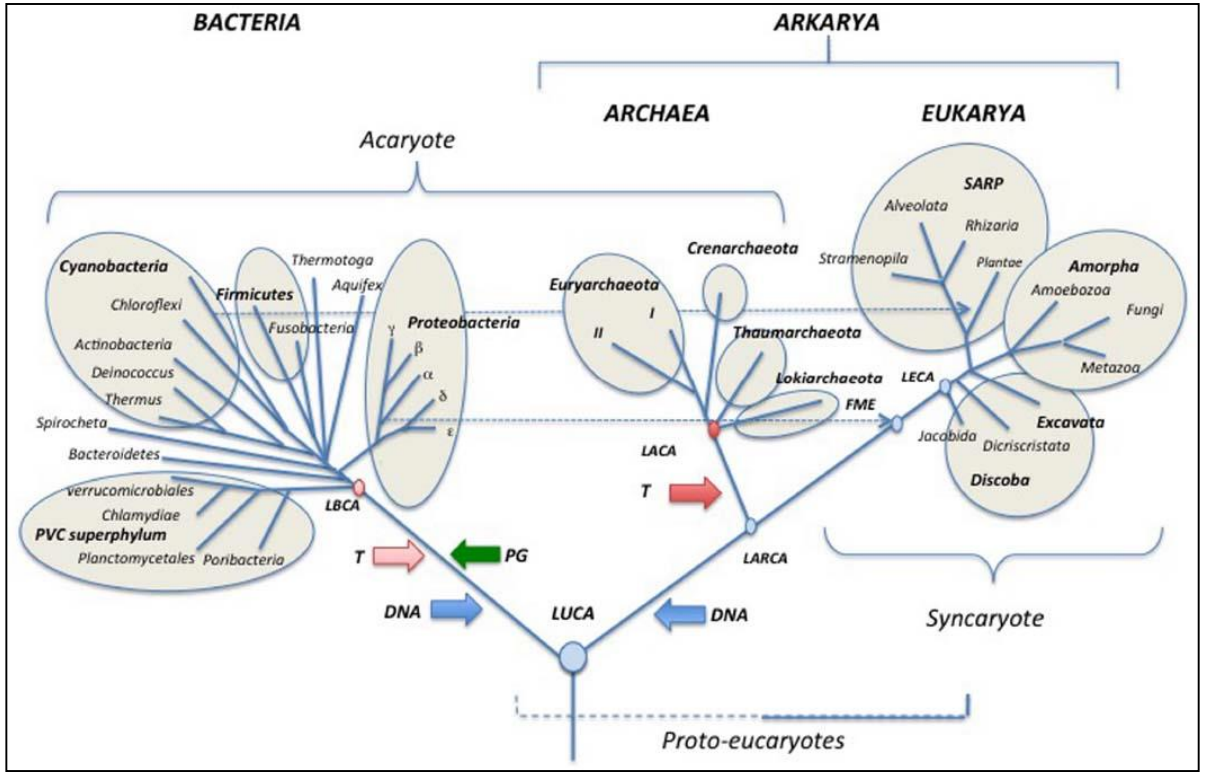
Arkarya, a new name proposed for the clade grouping Archaea and Eukarya.

Forterre,2015

The universal tree of life: an update

Universal multiprotein ribosomal protein (RP) trees

Offer markedly higher resolution for prokaryote phylogeny than rDNA



Schematic universal tree updated from (Woese *et al.*,1990)

DNA (blue arrows) introduction of DNA; T (pink and red arrows) thermoreduction. LBCA: Last Bacterial Common Ancestor, pink circle: thermophilic LBCA; LACA: Last Archaeal Common Ancestor, red circle, hyperthermophilic LACA. LARCA: Last Arkarya Common Ancestor; FME: First Mitochondriate Eukarya; LECA: Last Eukaryotic Common Ancestor; blue circles, mesophilic ancestors. SARP: Stramenopila, Alveolata, Rhizobia.

Forterre,2015

Some recent natural classification scheme Based on ribosomal protein trees

Kovacs *et al*.,2018

- Rooting of the domain Bacteria and internal branching in this domain have been adapted from the ribosomal protein trees of Koonin and colleagues (Yutin *et al*.,2012).
- These authors have suggested several superphyla beside the previously recognized PVC superphylum, which includes:
- Planctomycetes,
- Verrucomicrobia, and
- Chlamydiae.

Cavalier-Smith calls the PVC group as Planctobacteria and considers it a phylum and not superphylum.

Forterre,2015,Wikepdia,2020

The universal tree of life: an update

Universal multiprotein ribosomal protein (RP) trees

Offer markedly higher resolution for prokaryote phylogeny than rDNA

- The ribosome is the most commonly observed intracellular multiprotein complex in bacteria.
- Functions, origins, and evolution of the translation system are best understood in the context of unambiguous and phylogenetically based taxonomy and nomenclature.
- Here, we map ribosomal proteins onto the tree of life and provide a nomenclature for ribosomal proteins that is consistent with phylogenetic relationships.

Kovacs *et al.*,2018

The universal tree of life: an update

Universal multiprotein ribosomal protein (RP) trees

Offer markedly higher resolution for prokaryote phylogeny than rDNA

- We have increased the accuracy of homology relationships among ribosomal proteins, providing a more informative picture of their lineages.
- We demonstrate that bL33 (bacteria) and eL42 (archaea/eukarya) are homologs with common ancestry and acute similarities in sequence and structure.
- bL33 and eL42 are composed of zinc ribbon protein folds, one of the most common zinc finger fold-groups of, and most frequently observed in translation-related domains.

Kovacs *et al.*,2018

The universal tree of life: an update

Universal multiprotein ribosomal protein (RP) trees

Offer markedly higher resolution for prokaryote phylogeny than rDNA

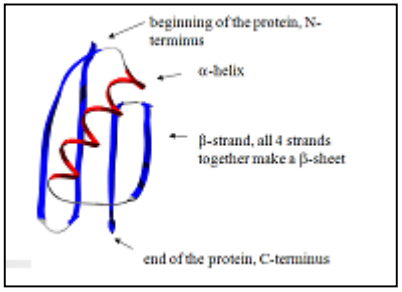
- Bacterial-specific ribosomal protein bL33 and archaeal/eukaryotic-specific ribosomal protein eL42 are now both assigned the name of uL33, indicating a universal ribosomal protein.
- We provide a phylogenetic naming scheme for all ribosomal proteins that is based on phylogenetic relationships to be used as a tool for studying the systemics, evolution, and origins of the ribosome.

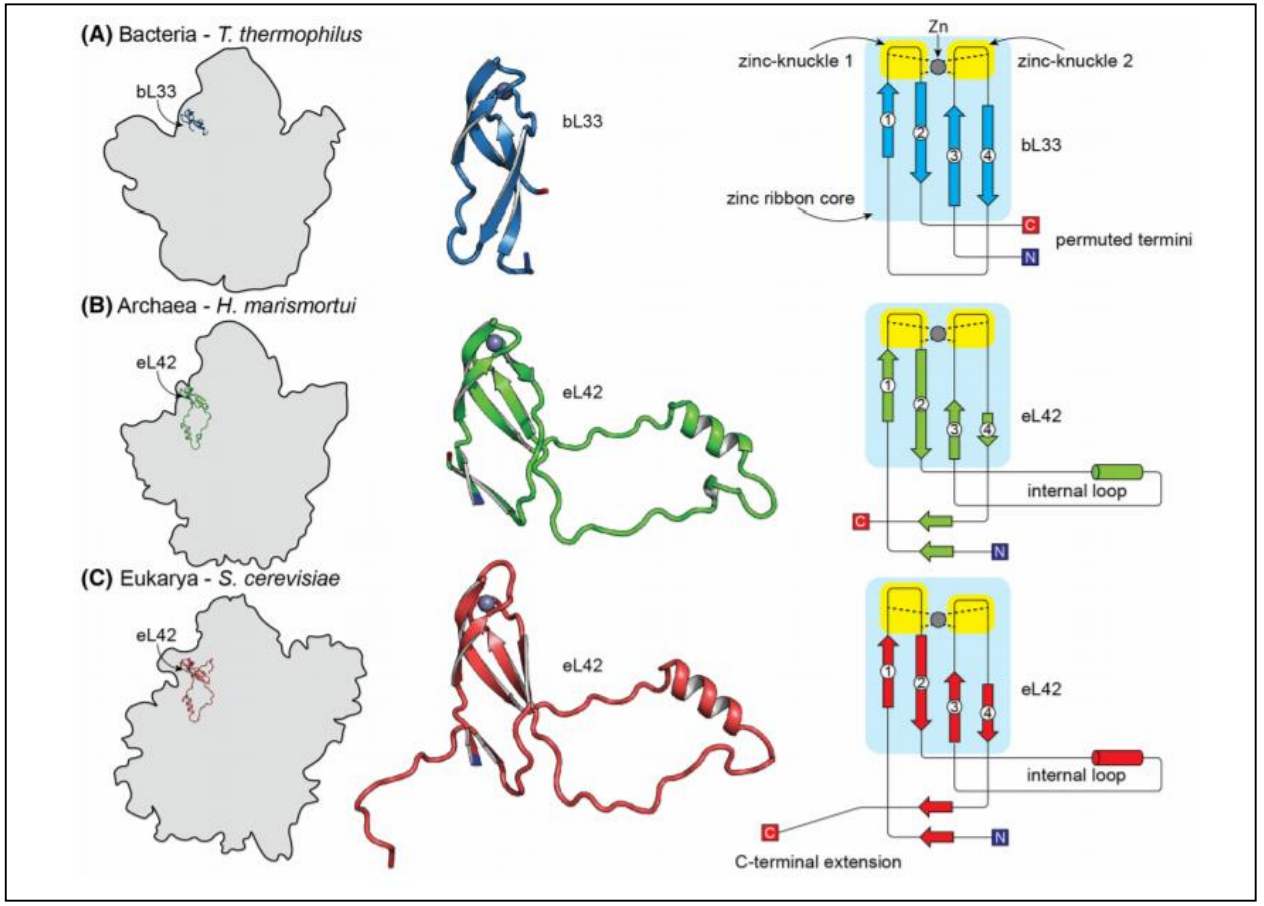
Kovacs *et al.*,2018

The universal tree of life: an update

Universal multiprotein ribosomal protein (RP) trees

Offer markedly higher resolution for prokaryote phylogeny than rDNA





Kovacs *et al.*,2018

Some recent natural classification scheme Based on ribosomal protein trees
Cavalier-Smith & Chao,2020

- Multiprotein ribosomal protein (RP) trees now offer markedly higher resolution for prokaryote phylogeny than rDNA.
- Universal multiprotein ribosomal protein (RP) trees, more accurate than rRNA trees, are taxonomically undersampled.
- Recent discoveries and the results of our RP trees now mean that we recognise 14 distinct eubacteria phyla.

Cavalier-Smith,T. & E. E-Yung Chao.2020. Multidomain ribosomal protein trees and the planctobacterial origin of neomura (eukaryotes, archaebacteria). Protoplasma. 257, 621–753. Published: 3 January 2020.

Date od death 19 March 2021 (aged 78).

Neomuran revolution

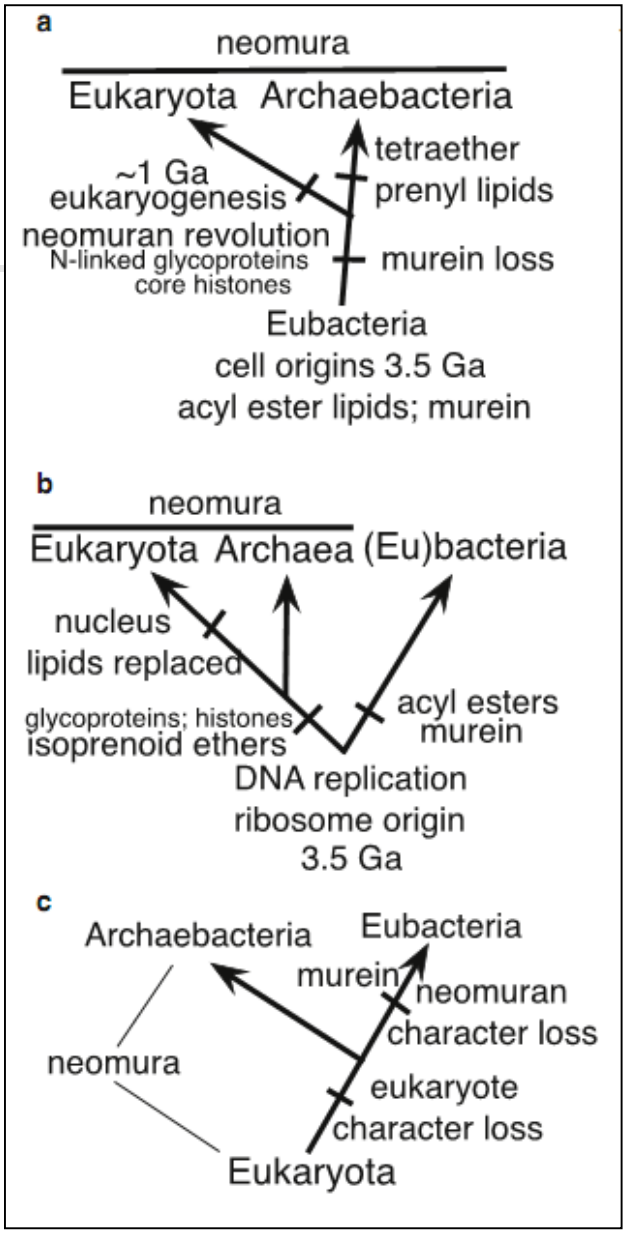
Murien loss

Planctobacterial origin of neomura

- The archaebacteri a/eukaryote clade was called neomura, meaning new walls, to contrast it with eubacteria that typically have walls of murein peptidoglycan.
- From the outset, it was controversial whether:

1. archaebacteria are ancestral to eukaryotes or
2. are their sisters still not unambiguously decided.

Gyr or Gy = gigayear= billion years ago
Eubacteria(3.5 Ga old) are > 3 × older than neomura (eukaryotes, archaebacteria with 1 Ga old). The Prenyllipids found in higher plants.
prenylation (modification by isoprenoid lipids).

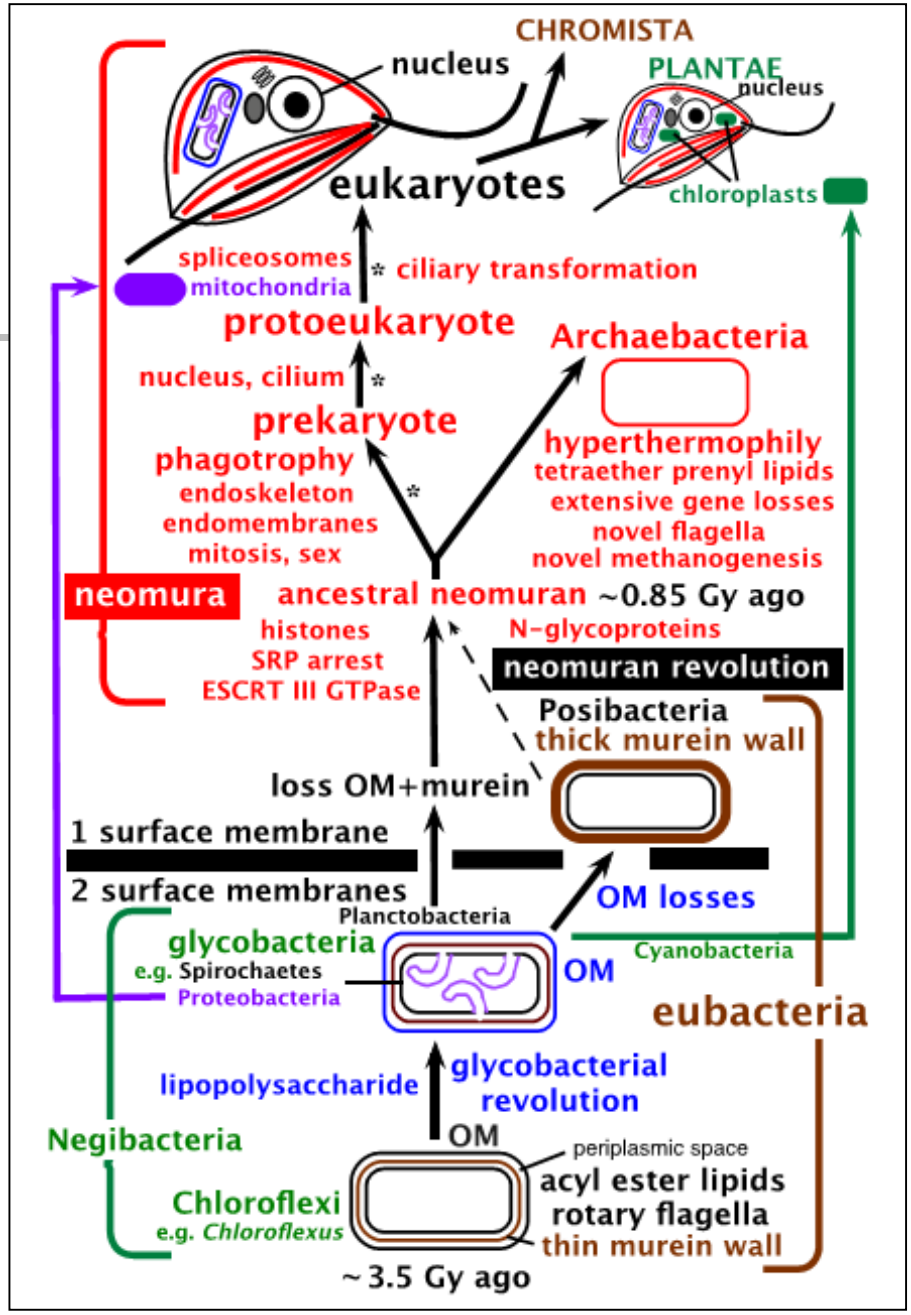


Cavalier-Smith & Chao,2020

Negibacterial root eubacteria

- The major kinds of cell and likely evolutionary relation ships.
- Chloroflexi (e.g., Green Non-sulfur Bacteria) as negibacteria, have an acyl ester phospholipid bilayer OM, being the earliest diverging negibacteria prior to LPS origin remains as strong as ever.
- Chloroflexi, the most divergent prokaryotes, which originated photosynthesis (new model proposed).

Gyr or Gy = gigayear= billion years ago



Cavalier-Smith & Chao,2020

Universal multiprotein ribosomal protein (RP) trees

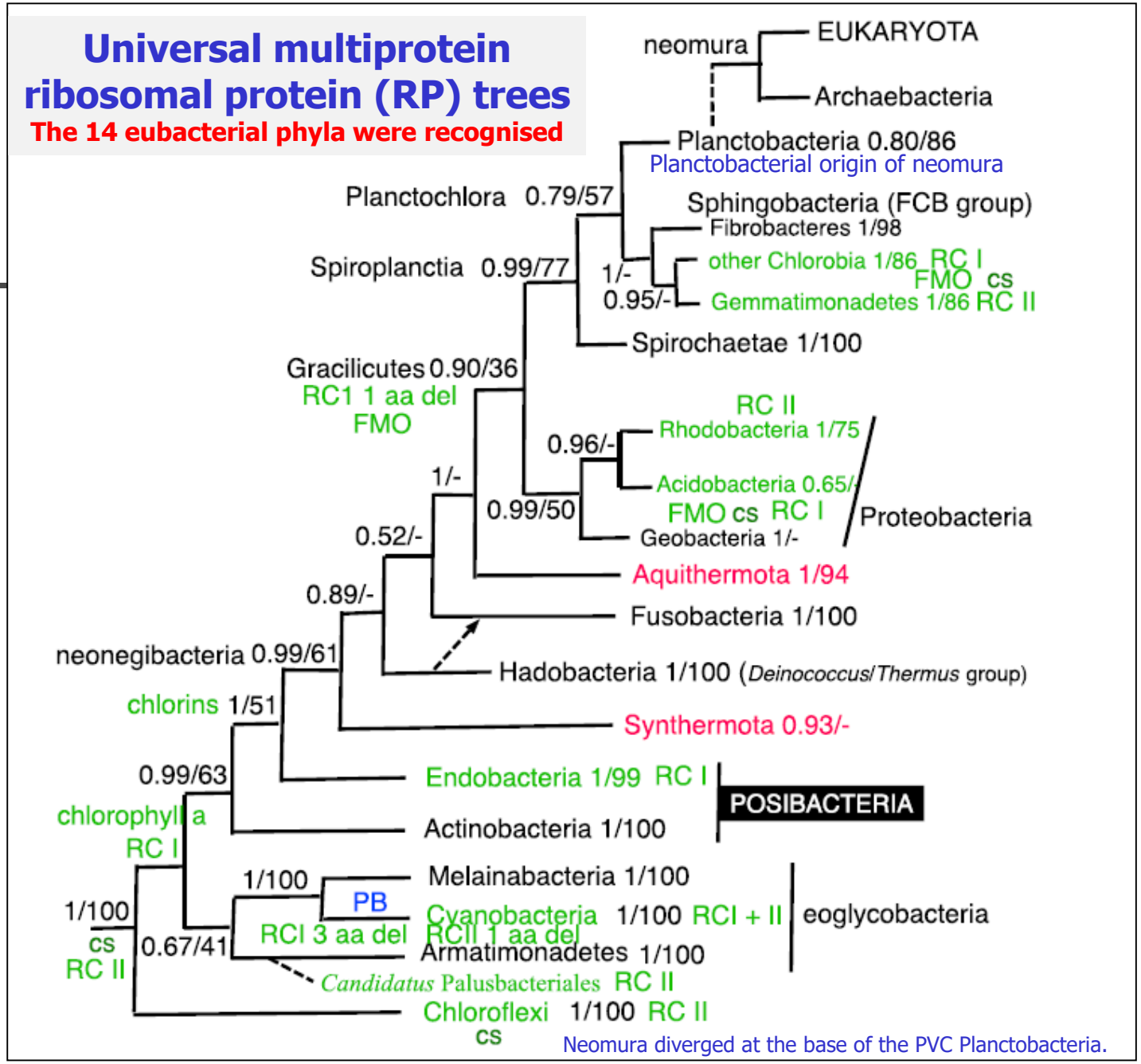
Offer markedly higher resolution for prokaryote phylogeny than rDNA

- Site-heterogeneous 26-RP trees for 143 eubacteria including all formally named phyla show 14 robust clades that merit recognition as phyla (**under half the number previously recognised**), several comprising more than one solely-rDNA-based 'phylum', whose constituent clades are better treated as subphyla or classes to simplify eubacterial classification.
- Relative branching order of the 14 robust phyla is much more strongly supported than in site-homogeneous analyses, and significantly doubtful only for the relative order of Hadobacteria and Fusobacteria, which might really be a single clade rather than successive branches.

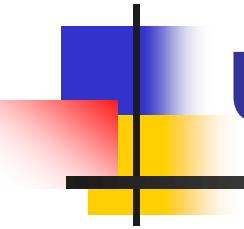
Cavalier-Smith & Chao,2020;…

Universal multiprotein ribosomal protein (RP) trees

The 14 eubacterial phyla were recognised



Cavalier-Smith & Chao,2020



Updating Prokaryotic Taxonomy

Establishing universal categories for all living organisms, including prokaryotes

(Towards a universal species concept and natural classification system)

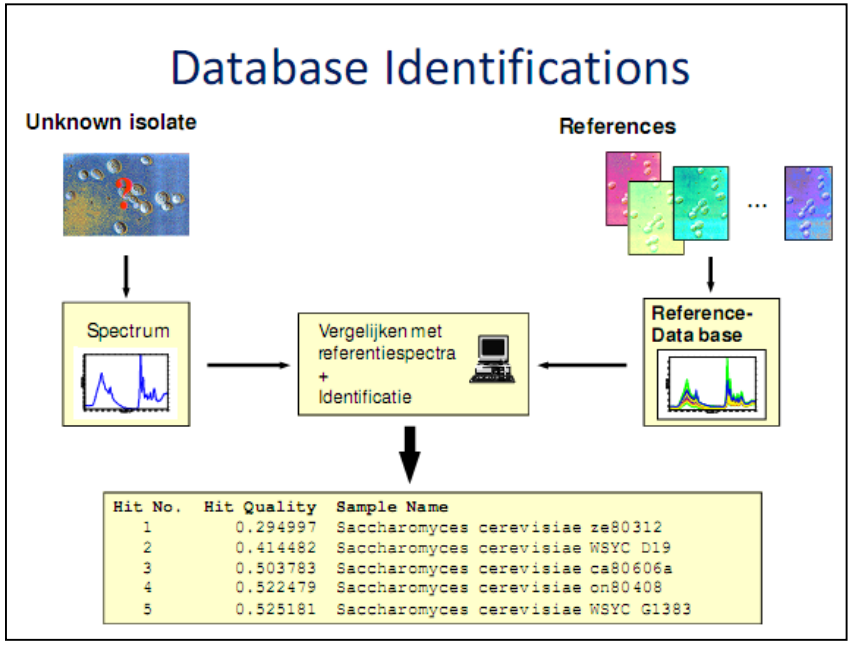
Overview

1. General introduction in bacterial taxonomy
2. Identification of bacteria
3. Comparison of microbial identification methods:
   - traditional methods
   - molecular methods
   - chemotaxonomic methods
4. Database identifications
5. Summary and general considerations.

Elke,2010

Database identifications

- Quality and completeness → crucial for accurate identification.
- Species not included → never be given as identification result.
- Include a set of representative strains, not only the type strain.
- Expand with new genera and species (formally described).
- Define well evaluated acceptation criteria.
- Commercial databases (CE/IVD) available.



IVD (In Vitro Diagnostics) devices (e.g. antibody, antigen, coated ELISA plates). IVD includes 'kit' as being an IVD in itself. The CE marking indicates a product's compliance with EU legislation.

Elke,2010

Important breakthroughs in prokaryote taxonomy

In the last 30 years

1. Computerized developments of numerical taxonomy;
2. A little later, but almost simultaneously, came the introduction of the first genomic parameters, G+C mol% content and whole-genome hybridizations, and these were followed by:
3. The appearance of numerous techniques revealing relevant chemical components (i.e., chemotaxonomy).
4. The most recent relevant innovation introduced into prokaryotic taxonomy is the use of gene or protein sequences, and especially that of the 16S rRNA gene, as molecular clocks.

Rosselló-Mora,2005

Important breakthroughs in prokaryote taxonomy

After a lapse of more than 30 years

- We have definitively entered the genomic era.
- The number of completely sequenced genomes is geometrically increasing with time simultaneously with the decrease in cost of such techniques.
- Thus, the race has begun to find those parameters that can be retrieved and that are useful for scientific taxonomic purposes using the current information data set.
- There is hope that genomics will produce tools to definitively understand "natural relationships" between microorganisms.

Rosselló-Mora,2005

Important breakthroughs in prokaryote taxonomy

Genome signature-Difference in sequence

- Genome signature is the relative abundance profile of di-, tri-, and tetranucleotides present in the genome.
- This method, developed by Karlin and colleagues and known as Karlin signatures.
- Genes that are shared between organisms can differ considerably in sequence.
- The percentage sequence divergence in orthologous genes is described by the ANI parameter (ANI: average nucleotide identity).

Vandamme,2013

Important breakthroughs in prokaryote taxonomy

After a lapse of more than 30 years

- Two new genome-based parameters applicable to prokaryotic taxonomy are:

1. Average amino acid identity (AAI): Pairwise genome comparisons and averaging the sequence identities of shared orthologous genes (amino acid).
2. Average nucleotide identity (ANI): Pairwise genome comparisons and averaging the sequence identities of shared orthologous genes (nucleotide).

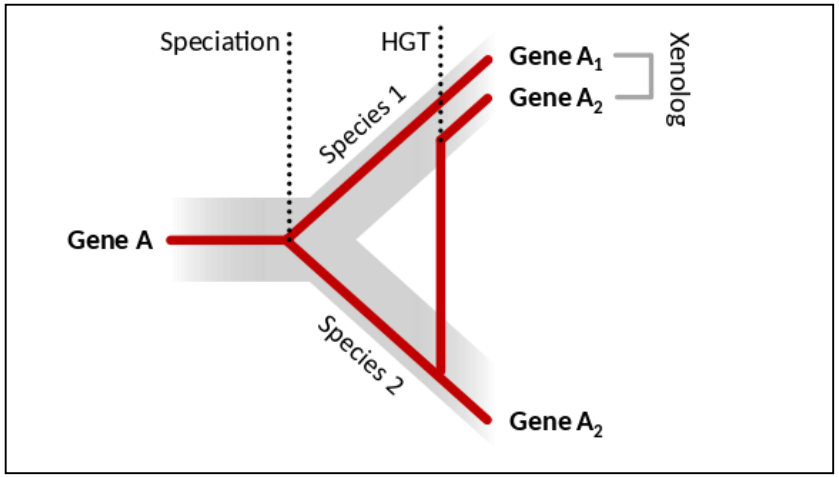
- Both achieve, to a great extent, the goal that whole-genome DNA hybridizations (DDH) pursued and that has been especially determinant for the circumscription of prokaryotic species.

Rosselló-Mora,2005

Orthologous genes: Genes in different species that are homologous (similar) because they are derived from a common ancestral gene.

Homologs vs Orthologs

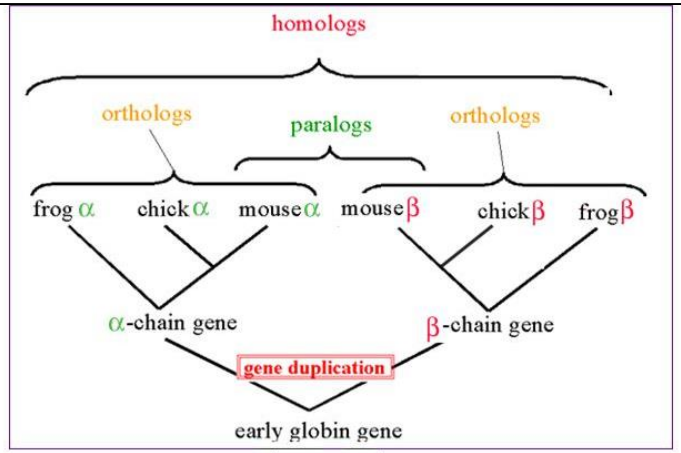
- A homologous gene (similar) is a gene inherited in two species by a common ancestor.
- Orthologous, the copy of the same genes are generated by speciation, not by gene duplication.
- Xenolog (plural xenologs) (genetics): A type of ortholog where the homologous sequences are found in different species because of horizontal gene transfer.



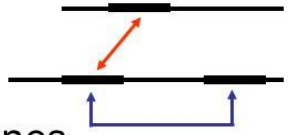
Homologs

Orthologs vs paralogs

- Orthologous genes: Genes in different genomes (species) with a common origin.
- Paralogous genes: Genes in the same genomes (species) with a common origin.



Orthologous genes are homologous (corresponding) genes in different species

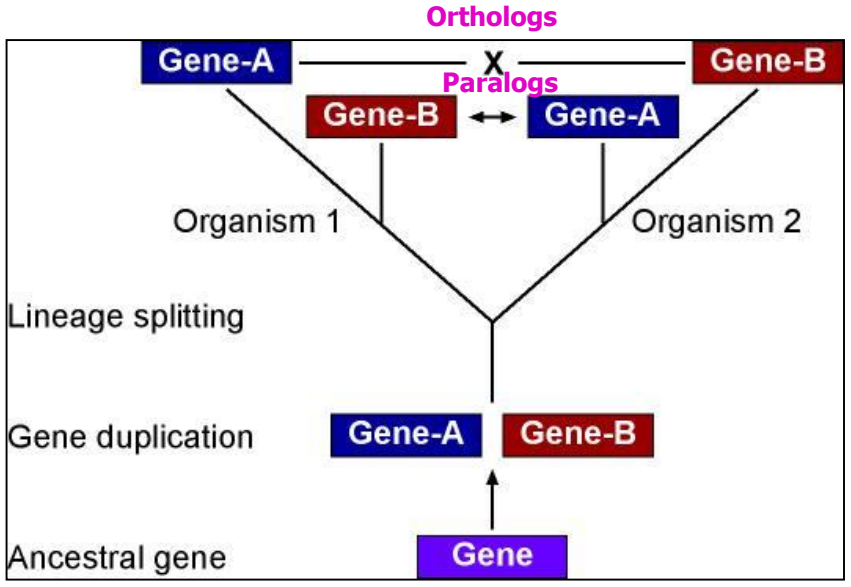


Paralogous genes are homologous genes within the same species (genome)

Orthologs and paralogs

Ortho-paralog tree

1. Orthologous genes: Genes in different species (organisms 1 and 2) that are homologous (similar) because they are derived from a common ancestral gene (X).
2. Paralogs are genes related by gene duplication (Gene A and Gene B from ancestral gene (↔).



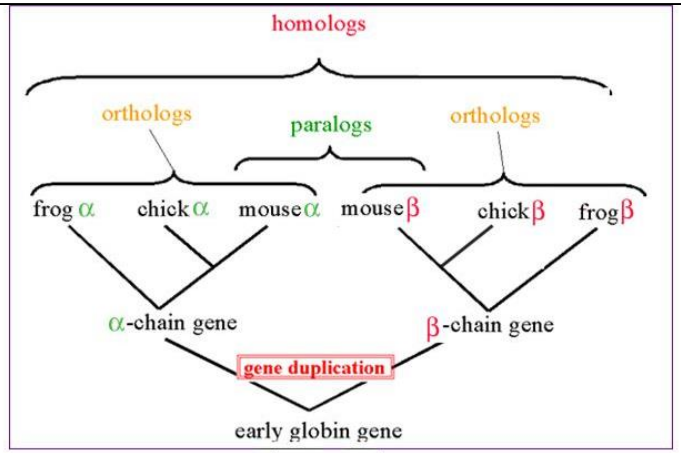
See also species definition section

GPS; www.icp.ucl.ac.be/~opperd/private/orthol.html

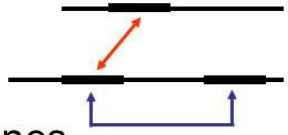
Homologs

Orthologs vs paralogs

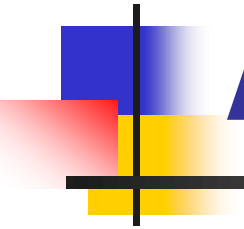
- Orthologous genes: Genes in different genomes (species) with a common origin.
- Paralogous genes: Genes in the same genomes (species) with a common origin.



Orthologous genes are homologous (corresponding) genes in different species



Paralogous genes are homologous genes within the same species (genome)



A. Phenetic Classification

Conventional Bacterial Taxonomy

(Numerical taxonomy)

Numerical Taxonomy

Phenetic Taxonomy

- Numerical taxonomy is a common approach to phenetic taxonomy.
- Numerical identification is a methods which is used a large number of tests (~100) and the results are scored as positive or negative.
- Several control species are included in the analysis.
- All characteristics are given equal weight and a computer based analysis is carried out to identify the bacteria.

Categories of characters used in classification

Category	Examples
Cultural	Colonial morphology, pigmentation
Morphological	Cell shape, staining reactions, motility
Physiological	Growth temperatures, anaerobic growth
Biochemical	Acid from carbohydrates, nitrate reduction
Nutritional	Organic acids as sole carbon and energy sources, vitamin requirements
Chemotaxonomic	Amino acids in interpeptide bridges of cell wall, types of lipids in membranes
Inhibitory tests	Sensitivities to antibiotics, dye tolerance
Serological	Agglutination by antisera to reference strains
Genomic	mol%GC in DNA, DNA–DNA reassociation

Logan,1994

Categories of characters used in classification

Classical characteristics

- Use a variety of characteristics: e.g., Gram stain, cell shape, motility, size, aerobic/anaerobic capacity, nutritional capabilities, cell wall chemistry, immunological characteristics such as:
- **Morphology and structure**: Shape, motility, Gram reaction, etc.
- **Microscopy**: Gram stain, acid-fast stain, cell morphology, cell arrangements
- **Colony morphology**: Size, color, border shape, etc.
- **Growth on**: Differential, Selective media agar, etc.
- **Cultural characteristics**: Photroph, chemotroph, aerobe/anaerobe, etc.
- **Biochemical activities**: Catalase, urease, oxidase, C/N utilizations
- **Pathogenicity**: Specific host infection, tissue colonization, etc.
- **Serological reactions**: Serotype or serovar, etc.
- **Phage typing**: Bacterial host specific, lyzotype
- **Gross protein analysis**: Electrophoretic patterns of the cell's proteins
- **Amino acid sequencing**: Amino acid sequence (s) of a given protein (s)
- **Chemotaxonomy**: Cell wall, Lipid & Protein compositions, Pyrolysis.

Pyrolysis is a process of chemically decomposing organic materials at elevated temperatures in the absence of oxygen. E.g. cellulose pyrolysis mainly produces three substances: (1) furans, (2) pyrans, and (3) linear small molecules.

Chemotaxonomy methods

- Classify organisms based on differences and similarities in chemical markers (cell wall constituents, lipids, whole cell proteins)
- Chemotaxonomic fingerprints of bacteria:

1. FAME GC (Fatty Acid Methyl Esters),
2. SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis) of whole cell proteins,
3. FTIRS (Fourier-transform infrared spectroscopy),
4. MALDI-TOF MS (Matrix assisted laser desorption ionization-time of flight mass spectrometry),
5. ESI MS(ionization mass spectrometry),…

Elke,2010

Cellular components and methods in chemotaxonomy

Site or level	Components	Analytical methods*
Extracellular	Products of metabolism	TLC, GLC, HPLC and isotachophoresis of extracts and derivatives
	Polysaccharides	GLC of derivatives
Whole-organism	Whole-organism	FT-IR spectrometry
	Pyrolysate	Pyrolysis GLC or MS
	Fatty acids	GLC of methyl esters
	Proteins	PAGE, two-dimensional electrophoresis and isoelectric focusing
	Polyamines	HPLC of dansyl derivatives
	Sugars	GLC of methyl glycosides
Gram-negative outer membrane	Lipopolysaccharide	Purification by UC then GLC of methyl esters and glycosides
	Polar lipids	TLC of solvent extracts
Mycobacterial outer membrane	Mycolic acids	TLC of methyl esters
	Free lipids	TLC of solvent extracts
Cell wall	Peptidoglycan:	
	1 Diaminopimelate isomers	PC and TLC of whole organism or wall hydrolysates
	2 Other amino acids	PC and TLC of wall hydrolysates, GLC of derivatives, AAAA
	3 Glycan type	Glycolate test
	Polysaccharides	As for whole-organism sugars
	Teichoic acids	Chemical analysis and serology
Plasma membrane	Isoprenoid quinones	TLC, RPTLC, GLC, HPLC, RPHPLC and DPMS of solvent extracts
	Polar lipids	
	Cytochromes	Spectrophotometry
Proteins	Enzymes	PAGE, growth independent enzyme assays (rapid enzyme tests) and functional characterization
	Amino acids	Sequencing

* Abbreviations: AAAA, automatic amino acid analysis; DPMS, direct probe mass spectrometry; FT-IR, Fourier-transform infrared; GLC, gas–liquid chromatography; HPLC, high-performance liquid chromatography; MS, mass spectrometry; PAGE, polyacrylamide gel electrophoresis; PC, paper chromatography; RPHPLC, reverse-phase HPLC; RPTLC, reverse-phase TLC; TLC, thin-layer chromatography; UC, ultracentrifugation.

Logan,1994

Numerical taxonomy

- In the 1950s-1960s it became evident that the analysis of large numbers of characteristics provided a more stable classification and a superior means to classify and identify bacteria.
- First generations of computers were used to analyze large data sets of biochemical and phenotypic characteristics.

Vandamme,2013

Numerical Taxonomy
Phenetic Taxonomy

- To create phenetic classification systems;
- multistep process;
- code information about properties of organisms

  e.g., 1 = has trait; 0 = doesn't have trait
- use computer to compare organisms on ≥50 characters;
- determine association coefficient (e.g. Simple matching coefficient)
- construct similarity matrix
- identify phenons and construct dendograms.

Numerical Taxonomy
Similarity coefficient

- Three genetic similarity coefficients were estimated and compared for their usefulness:

1. Simple matching (S_{SM}),
2. Jaccard's (S_J), and
3. Dice's (S_D).

For each coefficient, the similarity matrix was used to construct dendrograms with the help of the unweighted pair grouping by mathematical averaging (UPGMA) methods, using the SAHN and TREE programmes in NTSYS.

Numerical Taxonomy

Three main similarity coefficients

- **Typical coefficients are:**

1. **Simple matching coefficient (SSM)**- Based on the number of characteristics that match between two organisms; ignore mismatches.
2. **Jaccard coefficient (SJ)**- Ignores characteristics both organisms lack.
3. **Sorensen-Dice coefficients (S_D)**- In cluster analysis, the Jaccard and Sorensen-Dice coefficients revealed extremely close results, because both of them exclude negative co-occurrences.

Similarity analyses were done with the NTSYS-PC program.

Numerical Taxonomy

Three main similarity coefficients

Coefficients	Similarity expression(%)	Source
Simple matching (SM)	$\frac{a + d}{a + b + c + d}$	Sokal and Michener, 1958
Jaccard (J)	$\frac{a}{a + b + c}$	Jaccard, 1901
Sorensen-Dice (SD)	$\frac{2a}{2a + b + c}$	Sorensen, 1948; Dice, 1945

Dalirsefat *et al.*,2009

Similarity coefficients
Simple matching coefficient (S_{SM})

- Simple matching coefficient (SM) as a measure of genetic diversity for each pair of isolates:

$$SM = (a + d)/(a + b + c + d)$$

Where

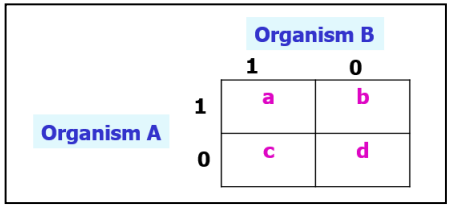
- *d* is the total number of bands absent in isolates 1 and 2.

Lebot *et al.*,2003

Simple matching coefficient (S_{SM})

The calculation of association coefficient for two organism

- In this example, organism A and B are compared in terms of the characters they do and do not share. The terms in the association coefficient equations are defined as follows:



a= number of characters coded as present (1) for both organisms;

b and c= numbers of characters differing as (1,0 or 0,1) between two organisms;

d= number of characters absent (0) in both organisms
Total number of characters compared=a+b+c+d.

$$\text{The simple matching coefficient } (S_{SM}) = \frac{a+d}{a+b+c+d}$$

Where, a character present and d character absent.

Numerical Taxonomy

Simple matching coefficient (S_{SM})

Usually expressed in percent

- Relies on similarity coefficients.
- Calculates the percentage of characteristics that two organisms or groups have in common.
- If use 10 characteristics, then match organisms:

Example:
A and B share 8 characters out of 10: similarity coefficient S_{ab} is 8/10=0.8 (80%). Can use many such values to establish similarity matrix.
Dendrograms help display this information clearly.

Numerical Taxonomy

Simple matching coefficient (S_{SM})

Usually expressed in percent

- The number of common characters is considered as a quantitative measure for the taxonomic relationship (not phylogenetic relatedness).
- Similarities are quantitatively expressed by the matching coefficient of Sokal and Michener (Skerman,1967), usually expressed in percent:

$$S_{SM} = \frac{\text{sum of positive and negative matches}}{\text{total number of tests}} \times 100$$

A and B share 8 characters out of 10: similarity coefficient S_{ab} is 8/10=0.8 (80%).

The Prokaryotes (chapter 1.4),2006

Similarity coefficients

Jaccard similarity coefficient (SJ or J)

- The Jaccard similarity index (sometimes called the Jaccard similarity coefficient) compares members for two sets to see:

1. which members are shared, and
2. which are distinct.

- It's a measure of similarity for the two sets of data, with a range from 0% to 100%.
- The higher the percentage, the more similar the two populations.

$$\text{The Jaccard coefficient } (S_J) = \frac{a}{a+b+c}$$

Statistics How To,2016

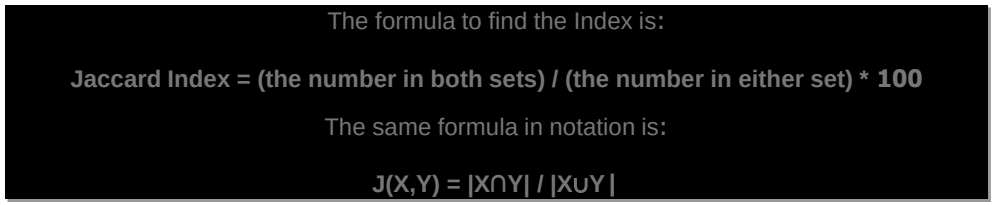
Similarity coefficients

Jaccard similarity coefficient (SJ or J)

- How to Calculate the Jaccard Index:
- The formula to find the Index is:

- Jaccard Index = (the number in both sets) / (the number in either set) x100
- The same formula in notation is:

$$J(X,Y) = |X \cap Y| \,/\, |X \cup Y|$$



∩(union)=The symbol used for the intersection of two sets. Intersection is a group of items that belong to two different sets.

Statistics How To,2016

Similarity coefficients
Sorensen-Dice coefficients(S_D)

- Similarity matrices were generated using DICE and simple matching coefficients and the resulting dendrograms were compared:

$$DICE = 2a/(2a + b + c)$$

where

- a is the number of shared bands,
- b is the number of bands present only in isolate 1, and
- c is the number of bands present only in isolate 2.

Lebot *et al.*,2003

Numerical Taxonomy
Jaccard and Sorensen-Dice coefficients

- The Jaccard and Sorensen-Dice coefficients revealed extremely close results, because both of them exclude negative co-occurrences.

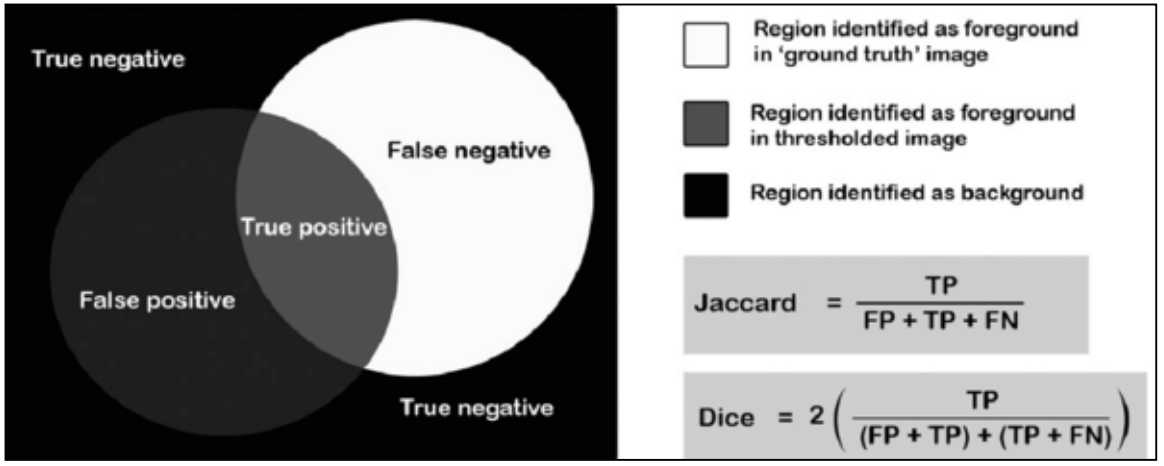
Coefficients	J	SD	SM
J	1.00		
SD	1.00	1.00	
SM	0.75	0.76	1.00

- The sorensen-Dice or Jaccard coefficients might be preferable to the simple matching coefficient when using RAPD analysis to compare groups of distantly related taxa.

Dalirsefat *et al.*,2009

Numerical Taxonomy
Jaccard and Sorensen-Dice coefficients

- An explanation of the Dice and Jaccard Indices of similarity.
- FN, false negative; FP, false positive; TN, true negative; TP, true positive.



Levin *et al.*,2018

Numerical taxonomy

Characterization and numerical taxonomy of novel oxalate-oxidizing bacteria

Clusters	A	B	C	D	E	F	G
Number of strains (*n*)	2	3	2	3	DSM1105^T	5	NS14
Growth on/at :							
MacConkey agar	50	67	+	-	+	-	+
Cetrimide agar	+	+	+	33	+	60	-
42 °C	+	67	-	67	+	-	-
3% NaCl	+	+	+	33	-	20	+
4% NaCl	50	-	+	-	-	-	+
2% Oxalate	50	67	+	-	-	-	+
Antibiotic resistance							
Chroramphenicol	-	33	50	-	-	80	-
Erythromycine	-	33	-	33	-	80	-
Streptomycine	50	-	-	67	-	80	-
Ampicilline	+	33	+	67	-	+	-
Bacitracine	+	+	+	+	+	+	-
Enzyme activity							
Nitrate reduction	+	-	-	67	+	-	+
Denitrificiation	-	-	-	-	+	-	-
Urease	+	33	-	33	+	+	-
DNase	+	+	-	33	+	20	-
ONPG	50	+	-	-	-	-	+
N-acetly-glucosamine	+	-	-	-	-	-	+
Lipase (C14)	-	67	50	-	-	-	-
Valine arylamidase	+	+	+	33	-	80	+
Cystine arylamidase	+	+	50	67	-	80	-
Trypsine	-	+	-	+	-	+	-
Acid phosphatase	+	+	50	67	+	+	-
α-galactosidase	50	-	-	-	-	-	-
β-glucosidase	50	67	-	-	-	-	+
β-galactosidase	-	67	-	-	-	-	+
α-chymotrypsin	-	67	-	-	-	20	-
Alkaline phosphatase	+	+	+	+	+	+	-
α-glucosidase	-	-	-	-	-	-	+
N-acetly-β-glucosaminidase	-	-	-	-	-	-	+

The values given in the table are the percentage of positive strains; [+]: all strains positive, [–]: all strains negative, y: yellow, p: pink.

Sahin *et al.*,2002

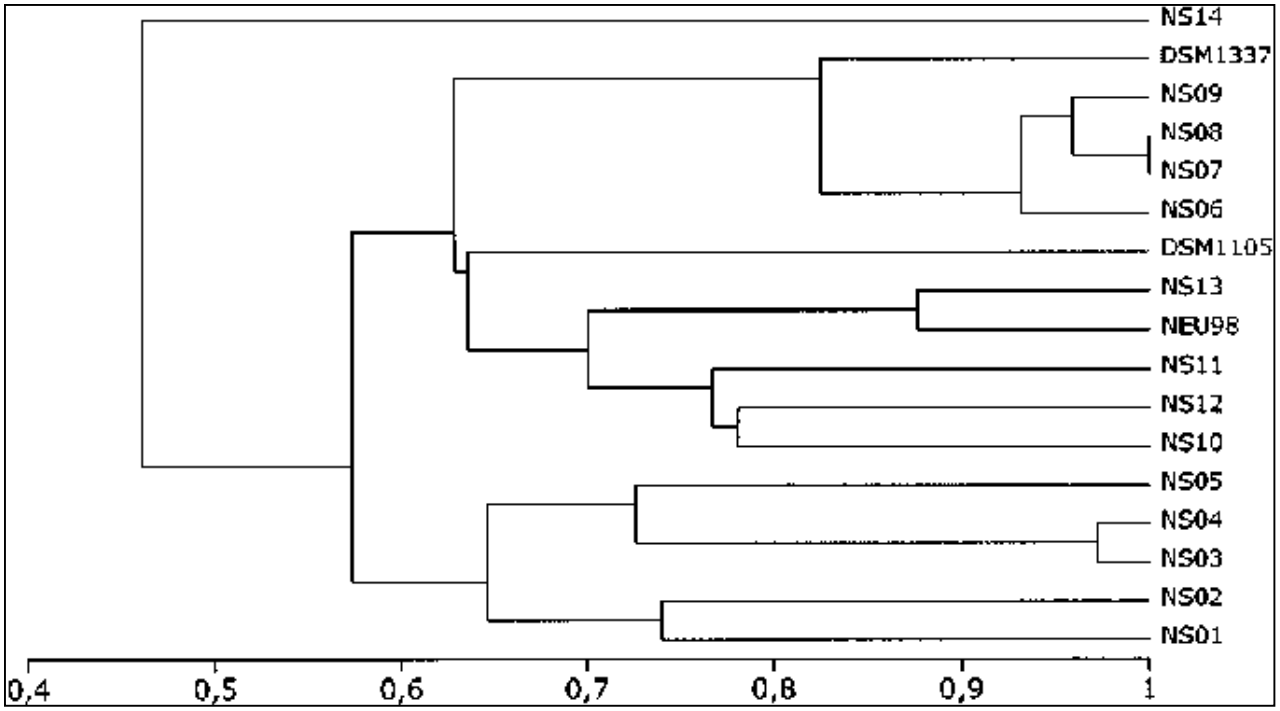
Numerical taxonomy
Continued

Clusters	A	B	C	D	E	F	G
Number of strains (*n*)	2	3	2	3	DSM1105^T	5	NS14
Pigmentation	-	-	y	-	-	p	y
Motility	+	-	+	+	+	+	-
Assimilation of :							
Glucose	+	67	-	-	-	60	+
Arabinose	+	67	-	-	-	-	+
Mannose	+	33	-	-	-	-	+
Maltose	+	-	50	-	-	-	+
Fructose	+	+	-	33	+	80	+
Xylose	+	+	-	33	-	80	+
Lactose	-	33	-	+	-	-	-
Galactose	+	+	-	+	-	-	-
Sucrose	50	-	-	33	-	20	-
Glycine	-	33	-	33	-	-	-
Gluconate	+	33	50	-	+	-	+
Caprate	50	-	-	-	+	-	-
Adipate	+	-	50	-	+	-	-

Sahin *et al.*,2002

Numerical taxonomy

Dendrogram showing phenotypic similarities among the isolates and reference strains

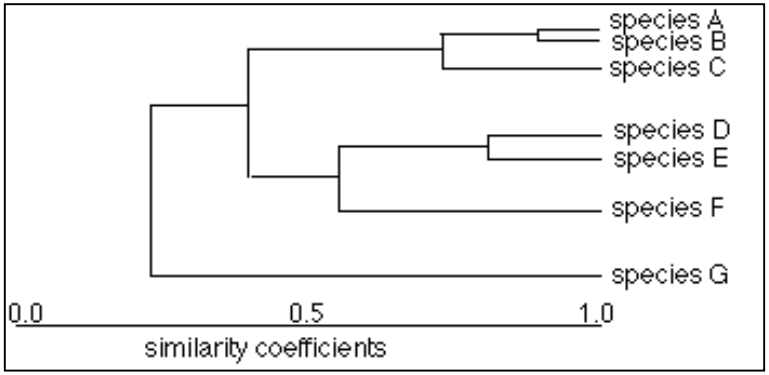


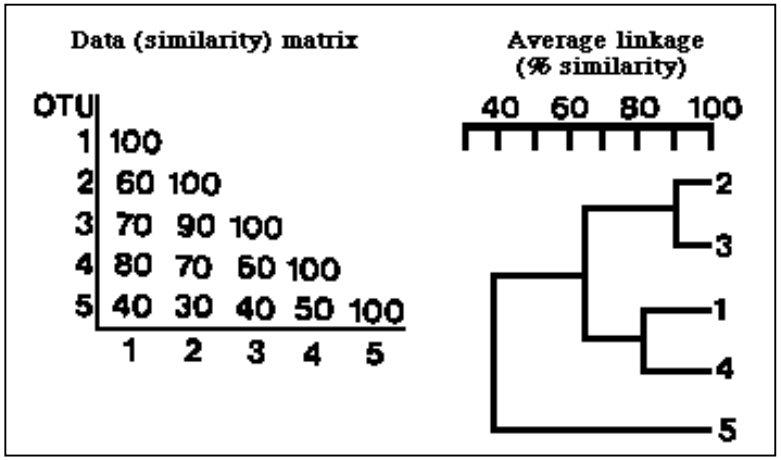
Scale shows the similarities (SSM).

Sahin *et al.*,2002

Numerical Taxonomy
Similarity coefficients

- Numerical taxonomy, the use of computers.
- The similarity matrix and conversion to dendrogram (phenogram).
- Greater similarity coefficient, closer relatedness is inferred.
- Greater than ~70% and inference is that two compared bacteria are of the same species.





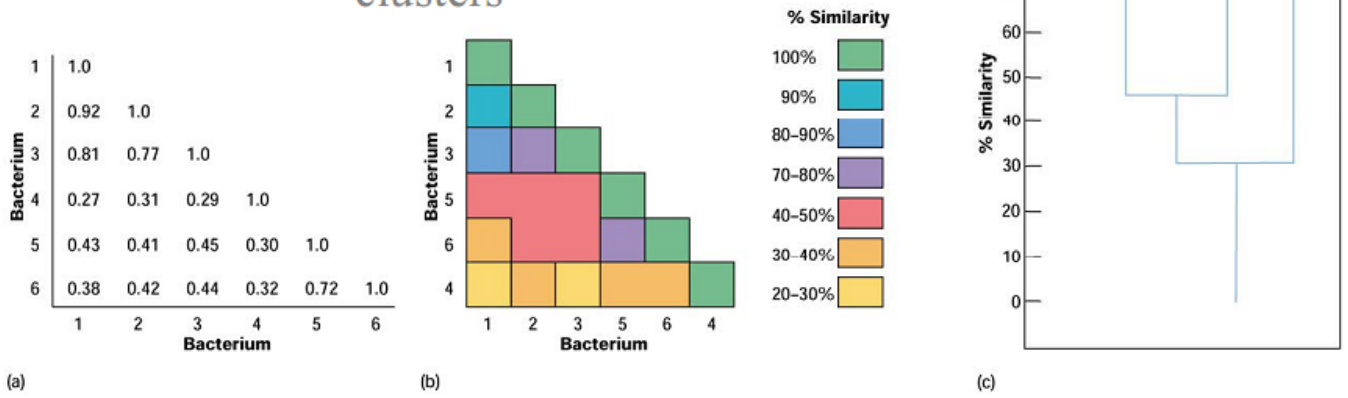
Operational taxonomic unit (OUT)

Numerical taxonomy

Similarity coefficients

- dendogram – treelike diagram used to display results
- phenon – group of organisms with great similarity
  – phenons with ≥80% similarity = bacterial species

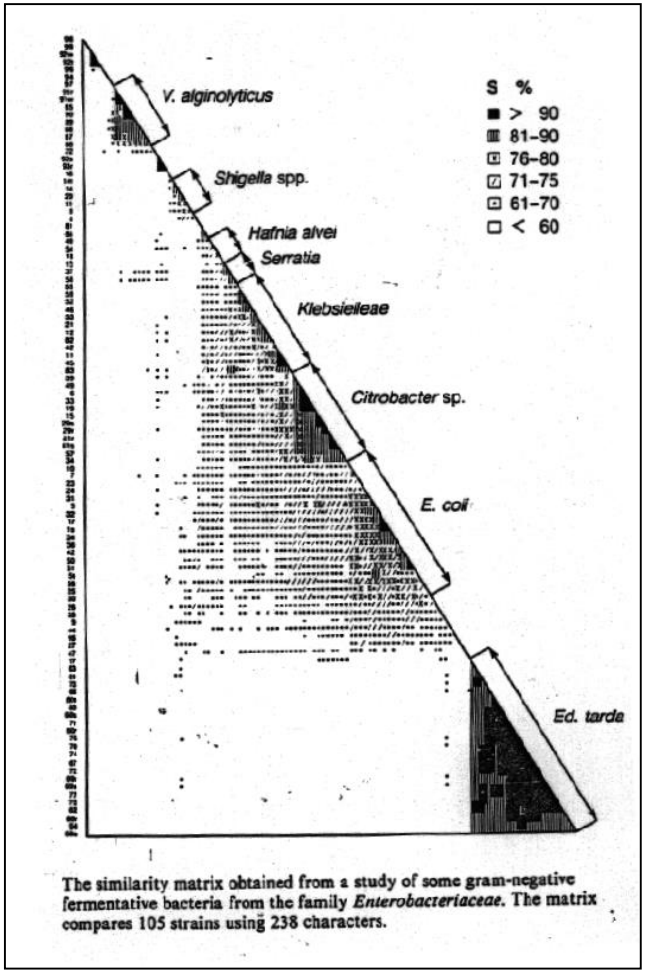
similarity matrix → rearranged and joined to show clusters → dendogram



Numerical taxonomy

Similarity coefficients

- Pattern matrix of the 124 test cultures based on 158 unit characters.
- The coefficients for matches between very many organisms by computer into a matrix (similarity matrix) **-->**

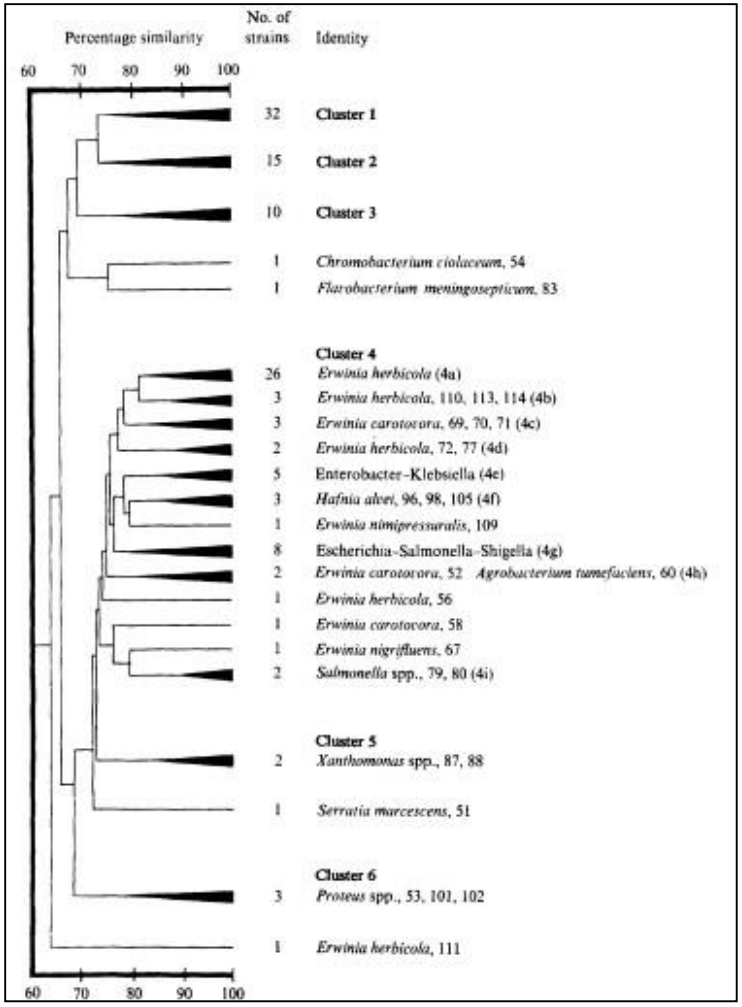


The similarity matrix obtained from a study of some gram-negative fermentative bacteria from the family *Enterobacteriaceae*. The matrix compares 105 strains using 238 characters.

Numerical taxonomy

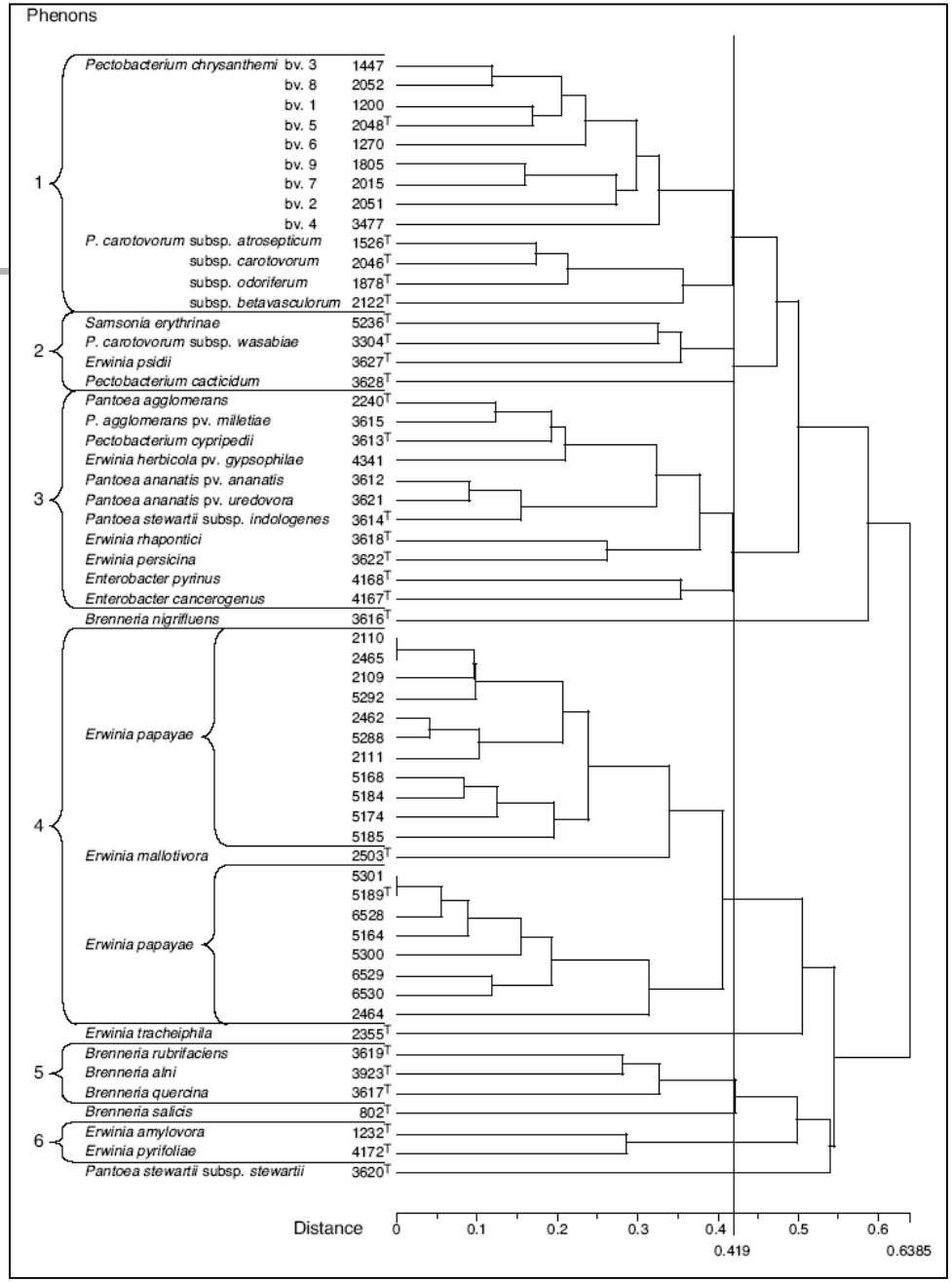
Numerical taxonomy of some yellow-pigmented bacteria isolated from plants

- Simplified dendogram showing the relationship between clusters and subgroups based on the coefficient and average linkage clustering.



Goodfellow *et al.*,1976

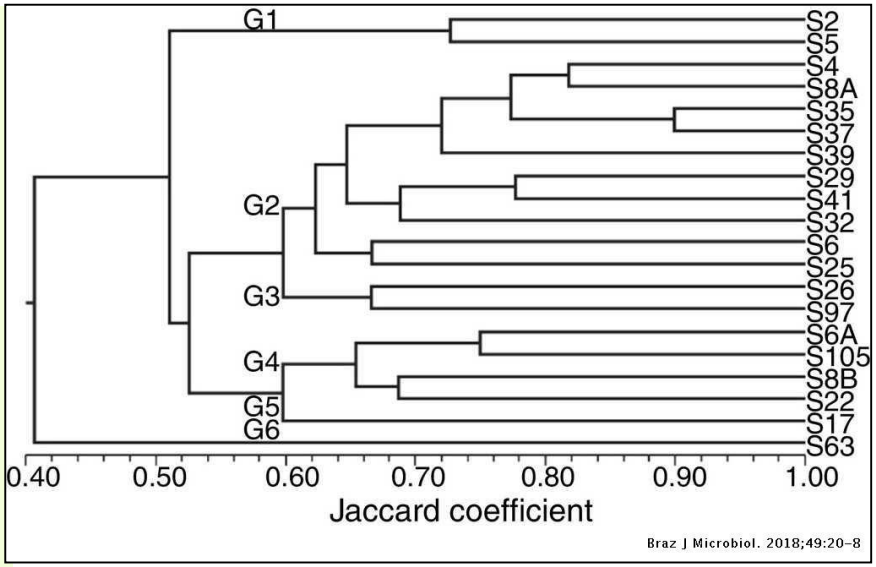
Dendrogram of phenotypic characteristics strains of phytopathogenic *Enterobacteriaceae*.



Numerical taxonomy

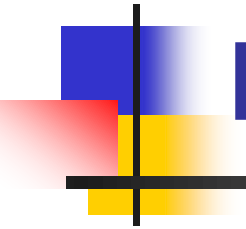
Similarity matrix obtained from the use of Jaccard coefficient

- The similarity analysis based on: morphological characteristics, enzymatic activity, carbon source use and antibiosis assay revealed six clusters formation at 60% of similarity, indicating great diversity among the 20 isolates regarding these traits.



Consensus dendrogram obtained by combining the morphological, enzymatic activity, carbon source use and antibiosis data among 20 endophytic isolates obtained from upland rice plants. Dendrogram was generated by the algorithm UPGMA and similarity matrix obtained from the use of Jaccard coefficient.

Braga *et al.*,2018



B. Polyphasic Taxonomy

A General Approach to Bacterial Systematics

Classical plus Molecular Taxonomy

Polyphasic taxonomy

- Taxonomic scheme which employs a wide range of stable and accepted phenotypic and genotypic information.
- The use of a range of radically different techniques on the same set or organisms allow:

1. To produce a classification;
2. To describe numerous new species.

Polyphasic taxonomy

- There is no single molecule that represents all organismal relationships adequately.
- Different molecules carry different types of information.
- A wealth of other methods was developed which were, just like the original biochemical tests, used to classify and identify bacteria.
- All of these methods carried some information that could be used as indirect measure of whole genome similarity between isolates.

Vandamme,2013

Polyphasic Taxonomy

- Taxonomy (systematics or biosystematics) is traditionally divided into three parts:

i. Classification: The orderly arrangement of organisms into taxonomic groups on the basis of similarity,

ii. Nomenclature: The labeling of the units defined in (i),

iii. Identification: The process of determining whether an organism belongs to one of the units defined in (i) and labeled in (ii).

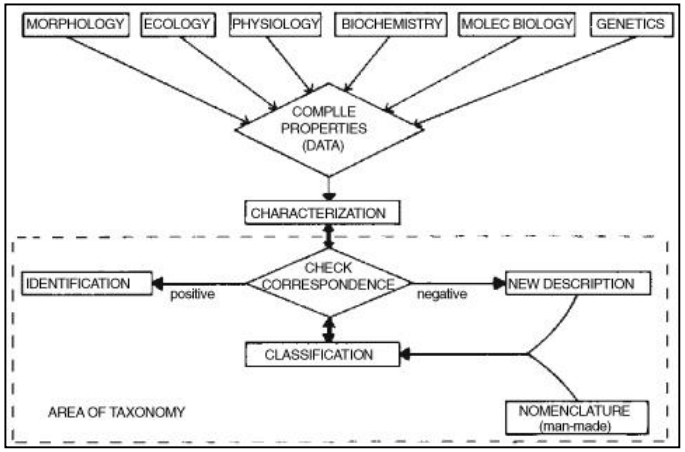
- Two additional parts are needed to completely define modern biosystematics:

iv. Phylogeny

v. Population genetics.

Polyphasic Taxonomy
Systematics

- The term systematics is often used in synonymy with taxonomy.
- This term (systematics) is better defined as: The scientific study of organisms with the ultimate object of characterizing and arranging them in an organized manner.
- Actually encompasses other disciplines including:
- Morphology,
- Ecology,
- Epidemiology,
- Biochemistry,
- Molecular biology, and
- Physiology.



Chester R. Cooper, Jr.,2004;..

Polyphasic Taxonomy

Ranks important to plant pathologists

Bacterial species/subspecies

Definitions

New Host-New Species
Operational

Species Concepts

Biological
Evolutionary
Ecological
Combinations

Vinatzer and Bull, 2010

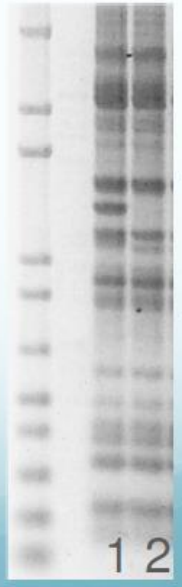
Pathovar

Pseudomonas cannabina pv. *alisalensis* on broccoli raab

- An infrasubspecific term referring to a group of phytopathogenic bacteria differentiated on the basis disease expression
- A novel pathovar is an organism with a novel host range or novel disease syndrome

Ecotypes

Lineage 1 isolated from many different crucifers



Lineage 2 isolated from Radish and Arugula

Carolee T. Bull

Typing methods

Typing methods

Identification at the level of individual strains:

- identification/elimination of source
- characteristics affecting transmission
- characterization of hypervirulent clones
- etc.

Phenotypic methods:	Metabolic activities, protein analysis, susceptibility to phages, etc.
Genotypic methods:	plasmid analysis, **PFGE**, ribotyping, **PCR**-ribotyping, **MLST**, etc.

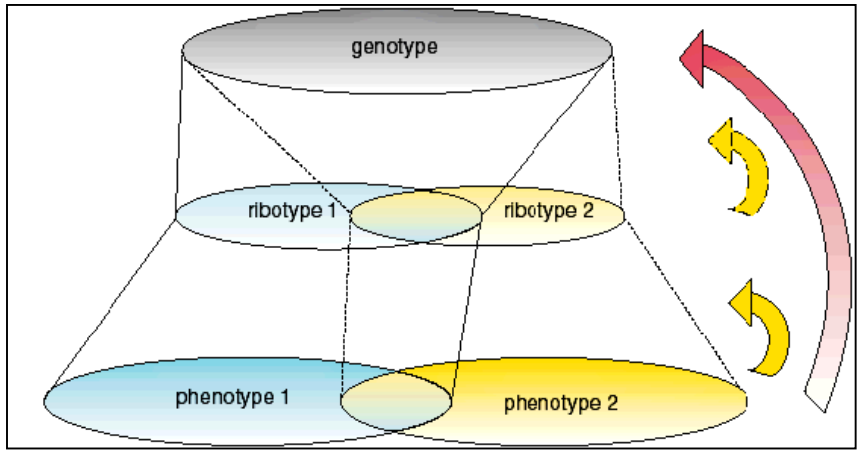
Pulsed-field gel electrophoresis (PFGE)

Altwegg,2005

Typing methods
Genotype, Ribotype and Phenotype

- A cell is a system made of:

1. Genotype,
2. Ribotype, and
3. Phenotype.



The ribotype is the collection of molecules and structures based on RNA, i.e., the mRNA, tRNA and the ribosomes. These molecules perform the critical function of translating the genotype into the phenotype.

Modern theories of the cell origin and evolution

Genotype, Ribotype and Phenotype

- Modern evolutionary theory is firmly based on the duality of the genotype and the phenotype.
- However Barbieri,1985 has described a new view in which life is based on a trinity of genotype-phenotype and ribotype.
- At the molecular level:

1. The genotype is the DNA,
2. The phenotype is the proteins, and
3. The ribotype is the collection of molecules such as mRNA, tRNA and the ribosomes.

Ray,1990

Modern theories of the cell origin and evolution

Genotype, Ribotype and Phenotype

- Each cell is not a duality of genotype and phenotype but a trinity made of genotype, phenotype and ribotype.
- The ribotype was defined as the ribonucleoprotein system of the cell, and it was underlined that it represents a new cell category (the ribotype theory).

1. Phenotype is the seat of metabolism;
2. Genotype the seat of heredity;
3. Ribotype is the seat of genetic coding.

- It was proposed that life on earth originated with the ancestors of today's ribotypes.

Barbieri,1981

A ribosome is composed of rRNA and ribosomal proteins known as a ribonucleoprotein.

Ribotyping
rRNA-based phylogenetic analyses

- All bacteria have ribosomal genes, but the exact sequence is unique to each species, serving as a genetic fingerprint.
- Therefore sequencing the particular 16S gene and comparing it to a database would yield identification of the particular species.
- Ribotyping is a molecular technique for bacterial identification that uses information from rRNA-based phylogenetic analyses.

rDNA PCR of the 16S-23S intergenic spacer region, ITS ('ribotyping'). The number of fragments generated by ribotyping is a reflection of the multiplicity of rRNA operons present in a bacterial species.

Ribotyping
Procedure

- Extract the genomic DNA from your bacterial isolate.
- Cut it with a specific restriction enzyme.
- Run the DNA in an agarose gel by electrophoresis.
- DNA fragments are separated by size as they move through the gel.
- Transfer the DNA pieces to a nylon membrane.
- Incubate the nylon membrane with a specific enzyme-linked DNA probe (a DNA fragment that hybridizes to the genes coding for 16S and 23s rRNA).

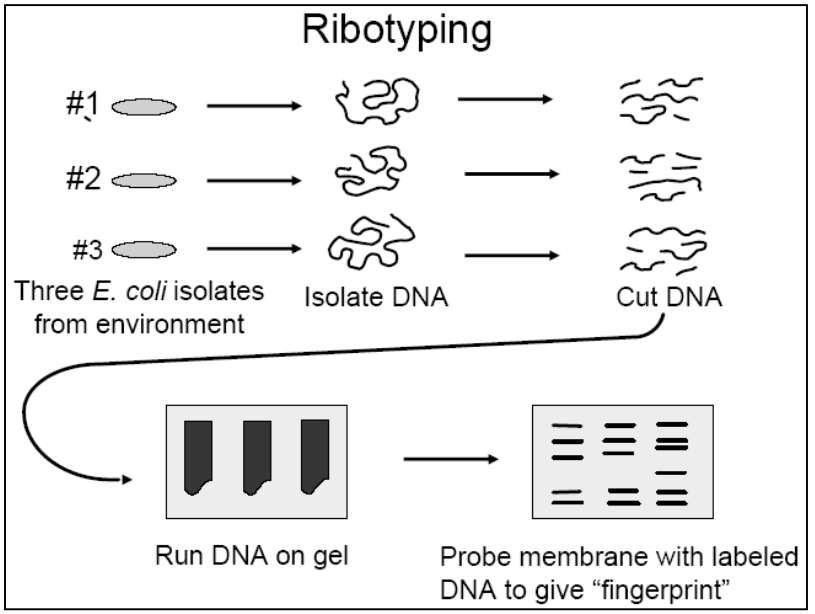
Ribotyping
Procedure

- Wash the nylon membrane and add the enzyme substrate to produce a chromogenic or fluorescent signal and visualize the pieces of DNA of interest.

1. In the case of chromomeric signal the reaction will be seen in the nylon membrane.
2. In the case of fluorescent reaction the nylon membrane is coupled with a photographic film and the results are seen on film.

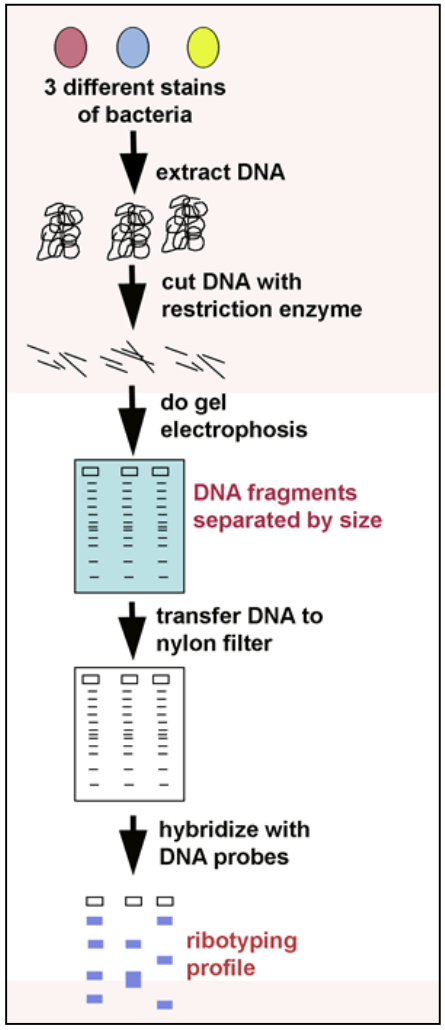
Ribotyping
Procedure

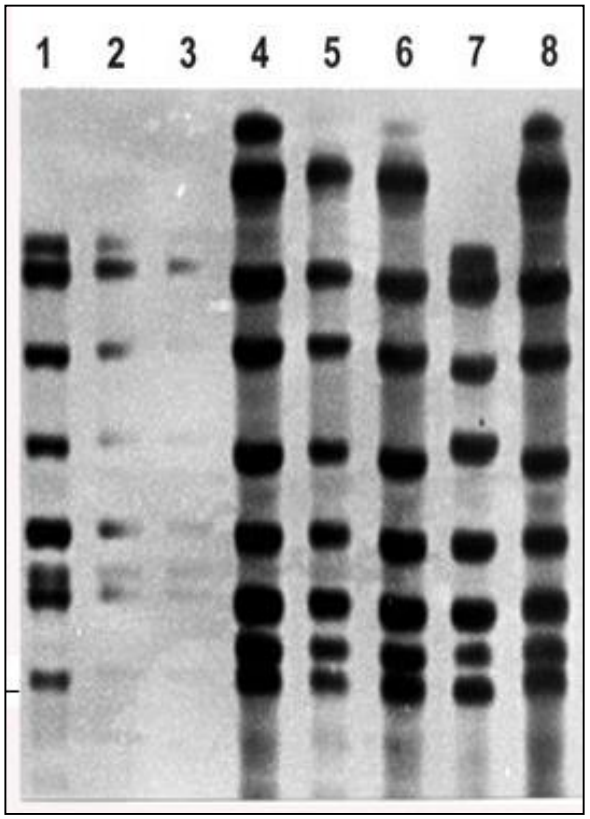
- Genomic DNA encoding the 16S and 16-23S ribosomal RNA molecules is digested by specific restriction enzymes.
- Each restriction enzyme cuts DNA at a specific nucleotide sequence, resulting of fragments of different lengths.
- Different strains of bacteria have the specific "cut-sites".
- Each strain generates a unique pattern of DNA pieces.
- The DNA fragments are transferred onto nylon membranes and hybridized with a labelled 16S or 23S rRNA probe.



Enzyme	Cutting site
*Bam*HI	GGATCC
*Hae*III	GGCC
*Pst*I	CTGCAG
*Hinf*I	GANTC
*Msp*I	CCGG

Ribotyping
Procedure





Hybridized with a labelled 16S or 23S rRNA probe.

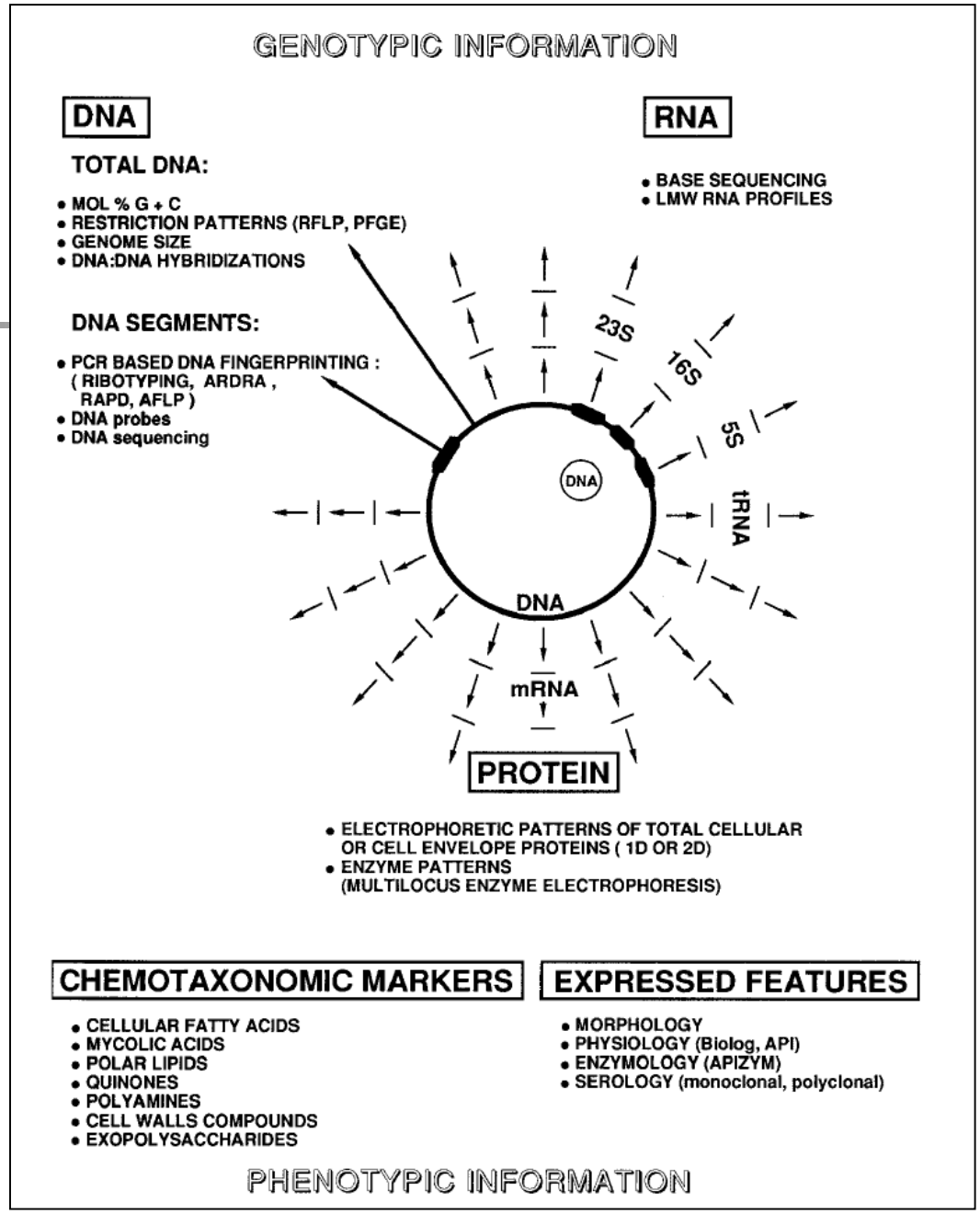
https://courses.cit.cornell.edu

Characteristics used in polyphasic taxonomy

- **Phenotypic Methods**
  - Classical phenotypic analyses
  - Numerical analysis
  - Automated systems
- **Genotypic Methods**
  - Determination of the DNA base ratio (moles percent GC)
  - DNA-DNA hybridization (DDH) studies
  - rRNA homology studies (Ribotyping: molecular fingerprinting)
  - DNA-based typing methods
- **Typing methods**
  - Cell wall composition
  - Cellular fatty acids
  - Isoprenoid quinones (lipid molecules)
  - Whole-cell protein analysis
  - Polyamines
- **Siderotyping is a new determinative tool for accurate identification of pseudomonads at the species level (Meyer *et al.*,2002).**

Peter Vandamme, Professor at Universiteit Gent, Belgium.

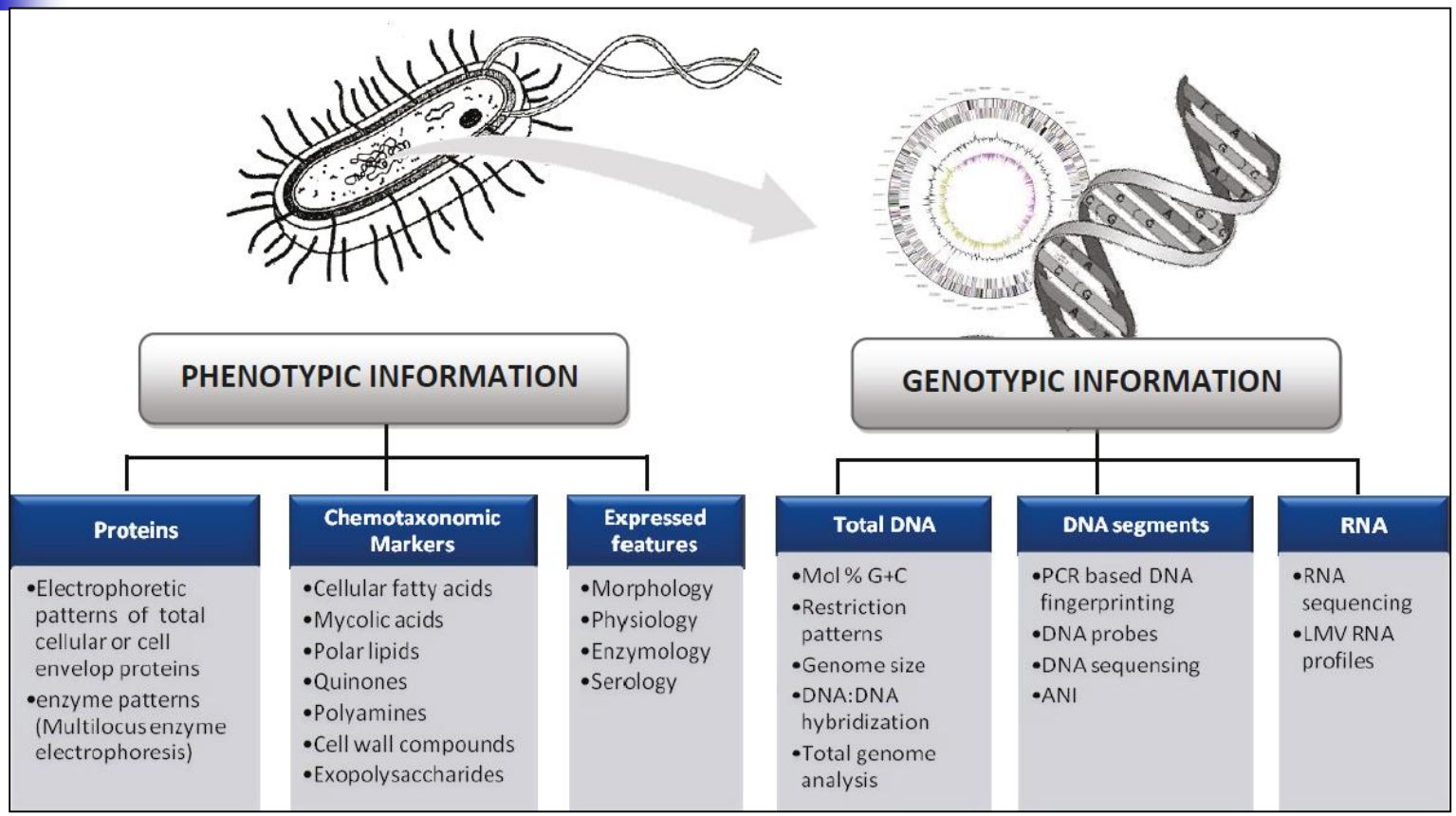
The seat and the nature of most of the genotypic and phenotypic methods used for monophasic studies or polyphasic comparisons of bacteria.



Vandamme *et al.*,1996

Schematic overview of various cellular components and techniques used in polyphasic bacterial taxonomy

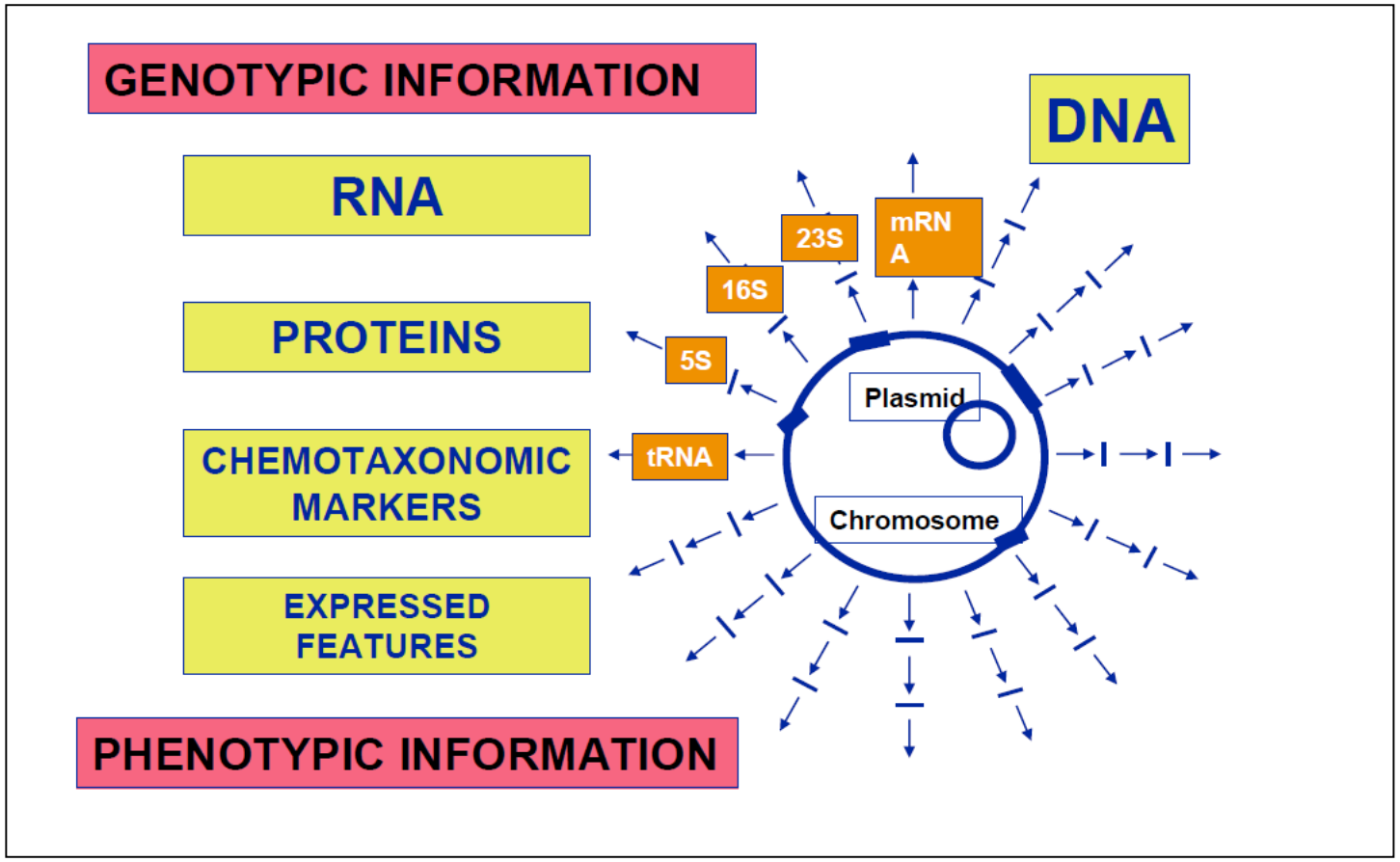
Adapted from Vandamme *et al*.,1996



Average Nucleotide Identity (ANI). A nucleotide is made up of three parts: a phosphate group, a 5-carbon sugar, and a nitrogenous base.

Amoozegar and Mehrshad,2013

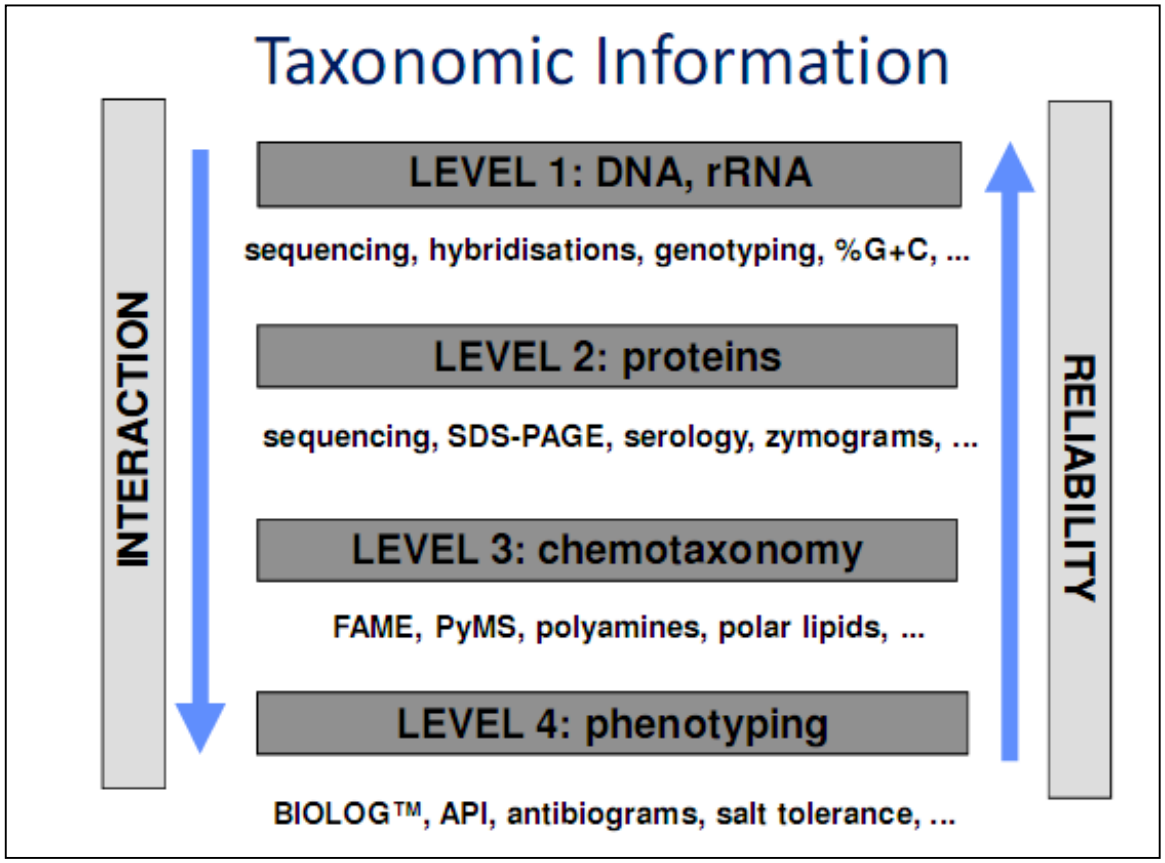
Characteristics used in polyphasic taxonomy



Coenye,2003

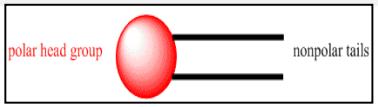
Characteristics used in polyphasic taxonomy

Interactions and reliability



Polar lipids are amphiphilic lipids with a hydrophilic head and a hydrophobic tail.

Lucas,2010

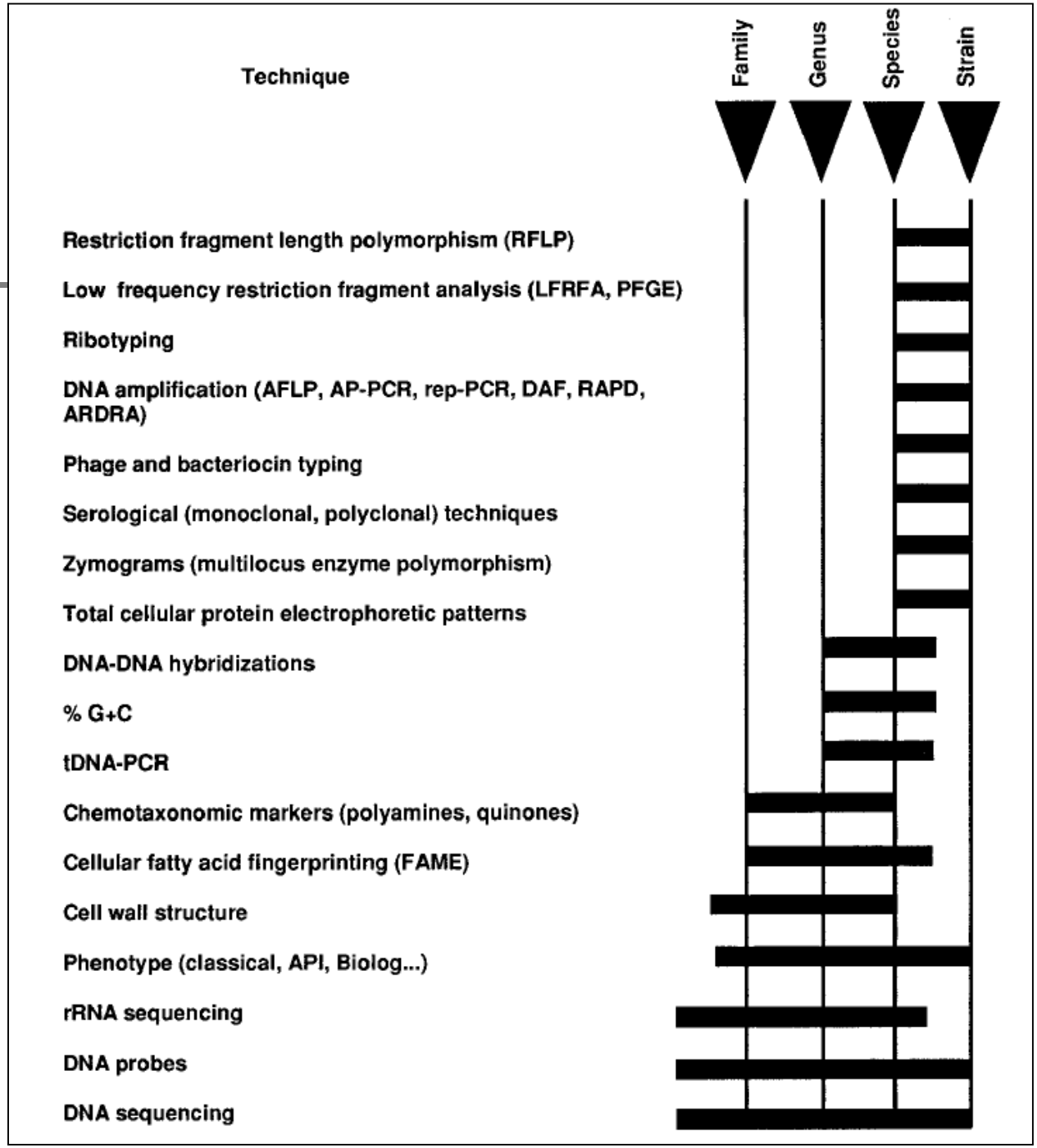


Polyphasic taxonomy:

Taxonomic resolution of some of the currently used techniques.

Zymograms: In electrophoretic strip (as of starch gel) or a representation of it exhibiting the pattern of separated enzymes and especially isoenzymes after electrophoresis.

Vandamme *et al.*,1996



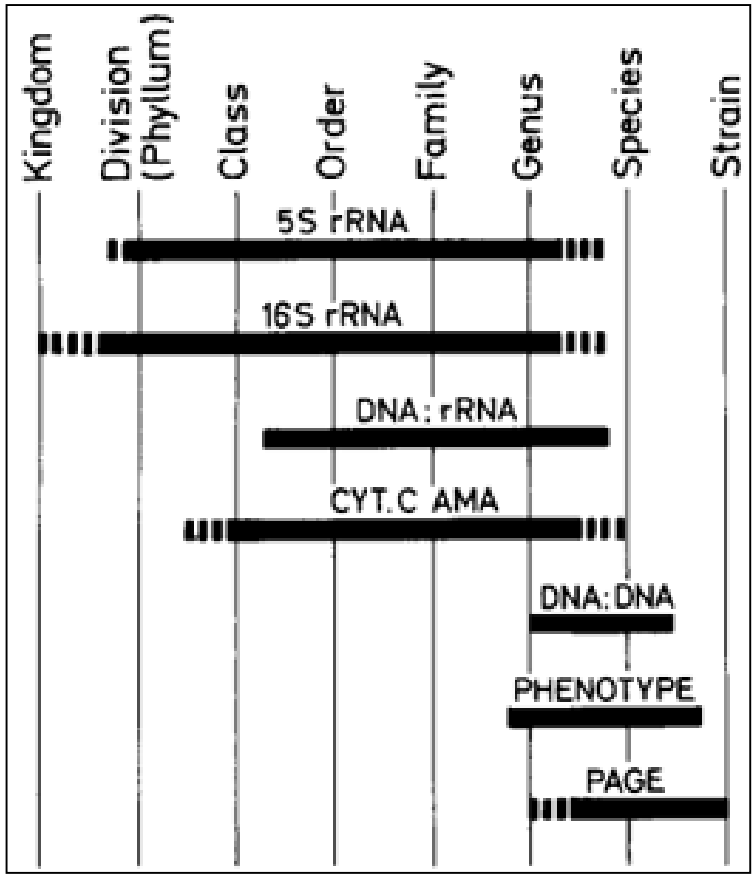
Taxonomic validity range of several methods

1. 5S rRNA sequencing; 16S rRNA cataloging is valid up to about a class or a division.
2. Complete sequencing extends the validity range to about a kingdom.
3. DNA-rRNA hybridization; cytochrome *c* amino acid (CYT.C) sequencing; DNA-DNA hybridization are useful on:

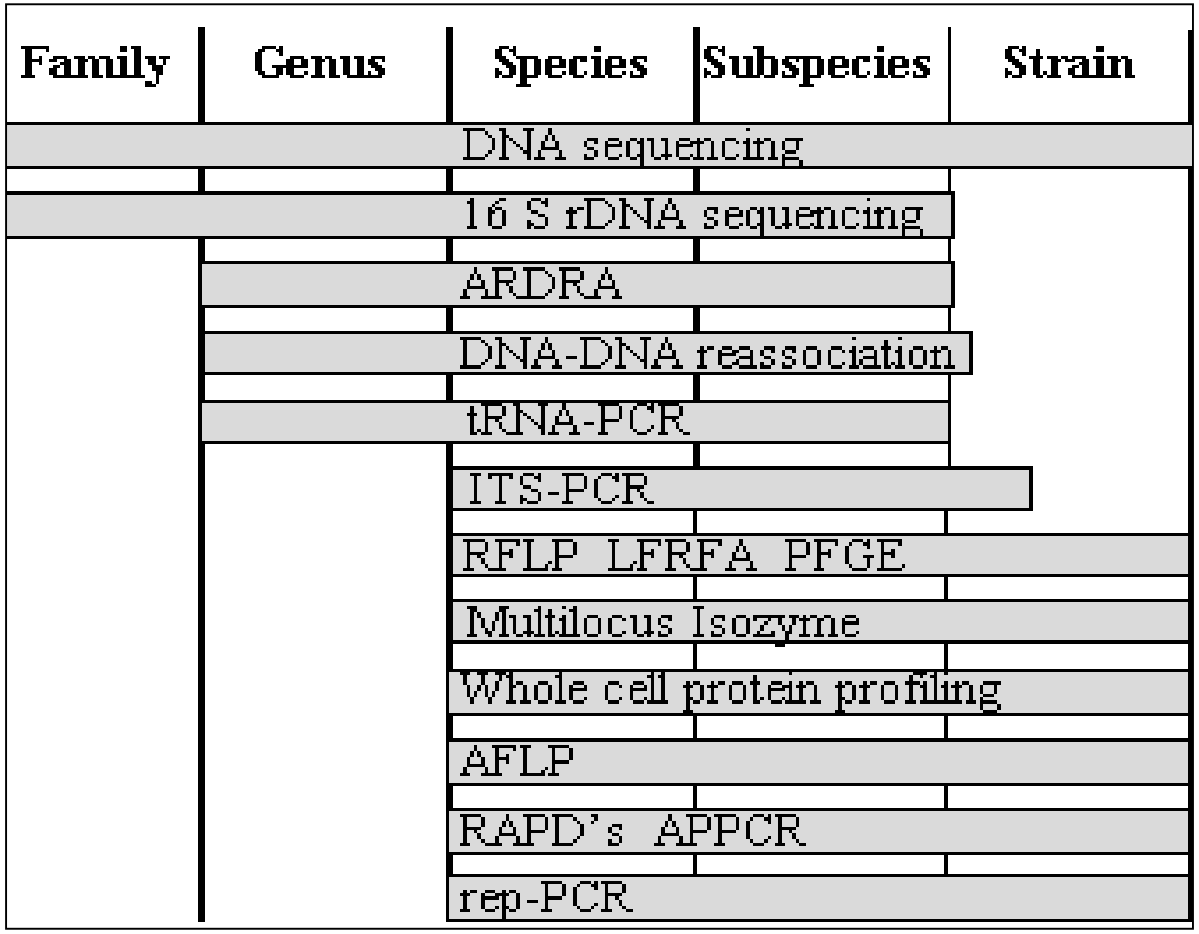
I. Phylogenetic or taxonomic conclusions,

II. Numerical analysis of phenotypic features;

III. Numerical analysis of protein gel electrophoresis (PAGE).

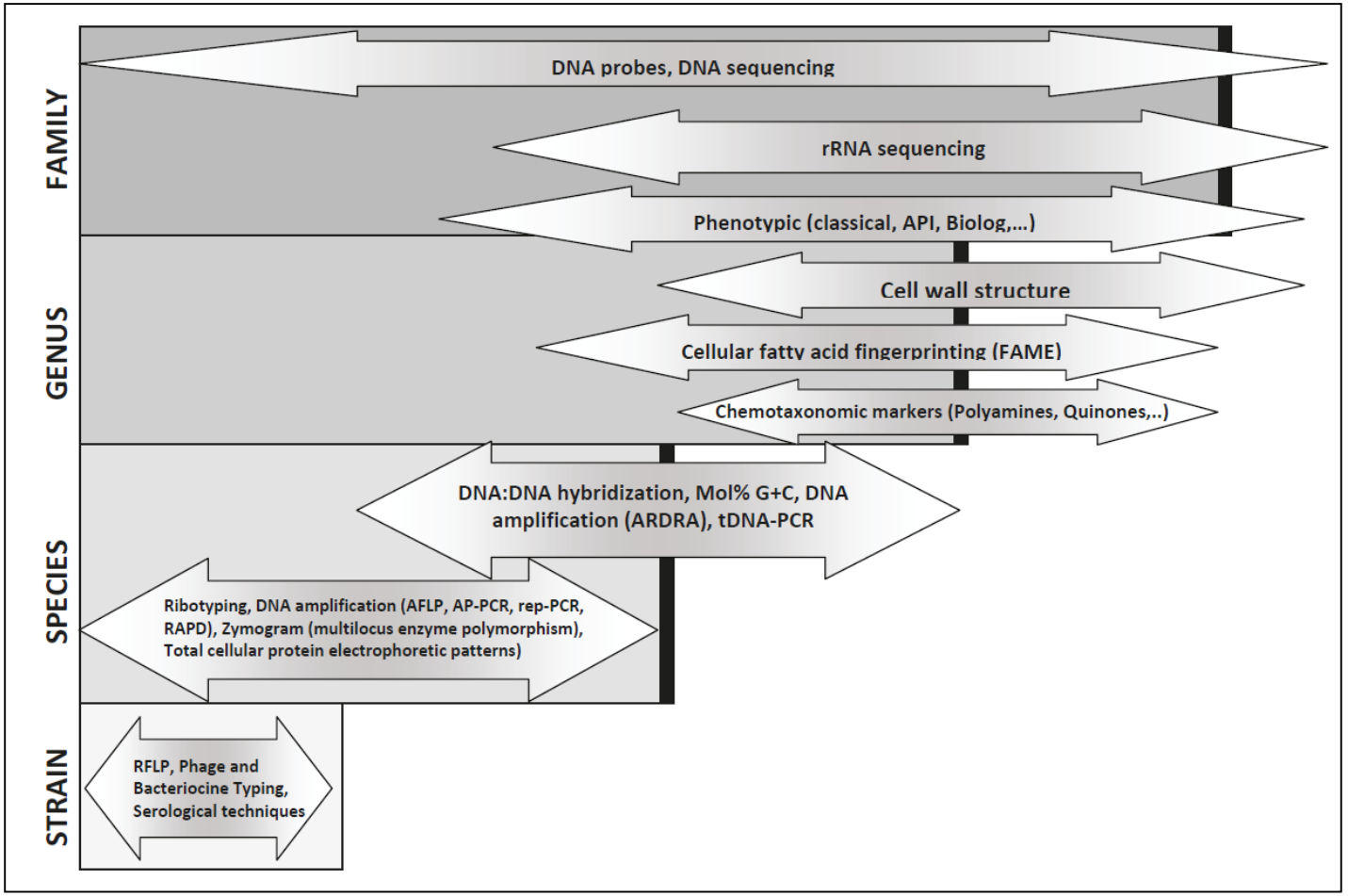


Relative resolution of various fingerprinting and DNA techniques



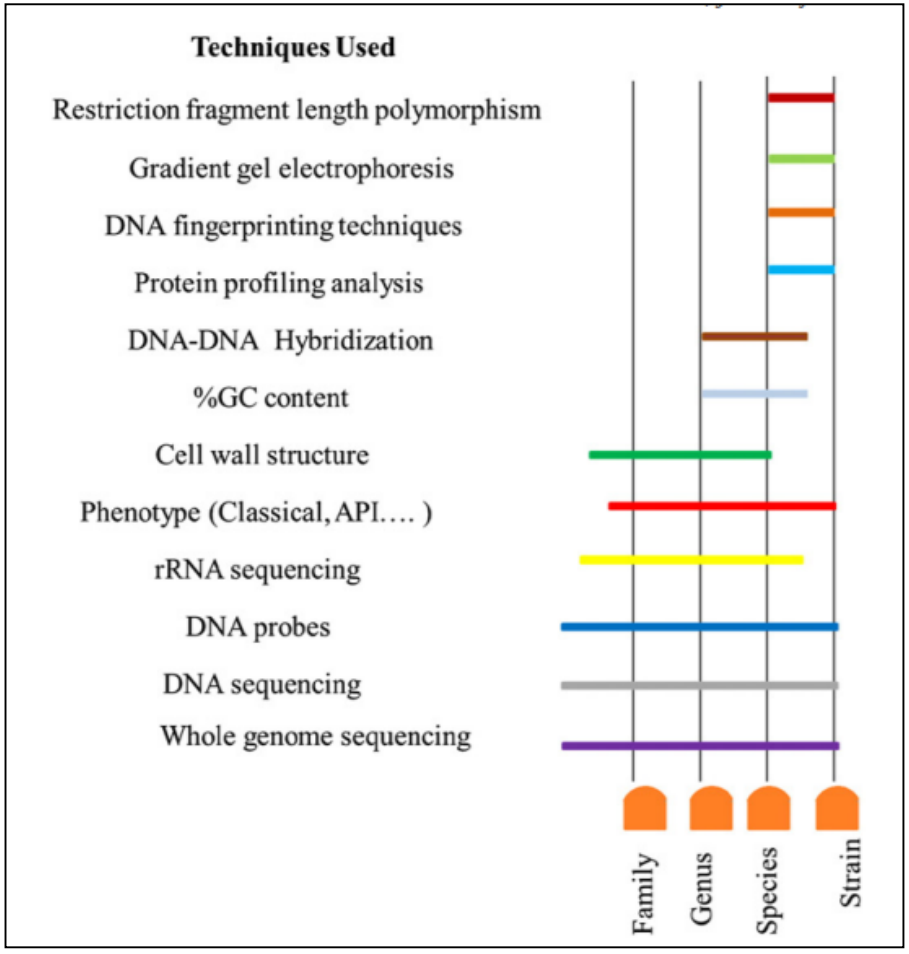
Rademake and Bruijn

Taxonomic resolution of some of the currently used techniques in bacterial taxonomy
Adapted from Vandamme *et al.*,1996



Amoozegar and Mehrshad,2013

Taxonomic resolutions of the currently used techniques



Das *et al.*,2014

JBT JOURNAL OF BIOMOLECULAR TECHNIQUES
Formerly known as ABRF News

Typing and target characteristics, electrophoresis platforms and references of several typing techniques.

	Characterization of:	Target used for characterization	Electrophoresis platform
AFLP	Whole genome or plasmid	Restriction site and additional nucleotide	Polyacrylamide[a] or agarose
rep-PCR Genomic oligonucleotide fingerprinting	Repetitive element oligonucleotide sequence	Repetitive element oligonucleotide	Agarose or polyacrylamide[a]
ARDRA	Ribosomal gene	Oligonucleotide sequence and restriction site	(MetaPhor) Agarose
ITS-PCR-FLP	Inter-transfer ribosomal gene spacer sequence	Oligonucleotide sequence and fragment length	Agarose or polyacrylamide[a]
IGS-PCR-FLP	Inter-ribosomal gene spacer sequence	Oligonucleotide sequence and fragment length	Agarose or polyacrylamide[a]
IGS-PCR-RFLP	Inter-ribosomal gene spacer sequence	Oligonucleotide sequence and restriction site	Agarose
T-RFLP	(Ribosomal) gene	Oligonucleotide sequence and restriction site	polyacrylamide[a]
SSCP	(Ribosomal) gene	Specific oligonucleotide and total fragment sequence	polyacrylamide[a]
T/DGGE	(Ribosomal) gene	Specific oligonucleotide and total fragment sequence	Polyacrylamide
RAPD/AP-PCR/DAF	Whole genome	Repetitive random oligonucleotide sequence	Agarose
PFGE	Whole genome	Low frequency restriction site	Agarose
MLST Plasmid profiling	Several genes Plasmid	DNA sequence Plasmid sizes	DNA sequence Agarose

[a]Fluorescent labels and capillary electrophoresis can be applied.

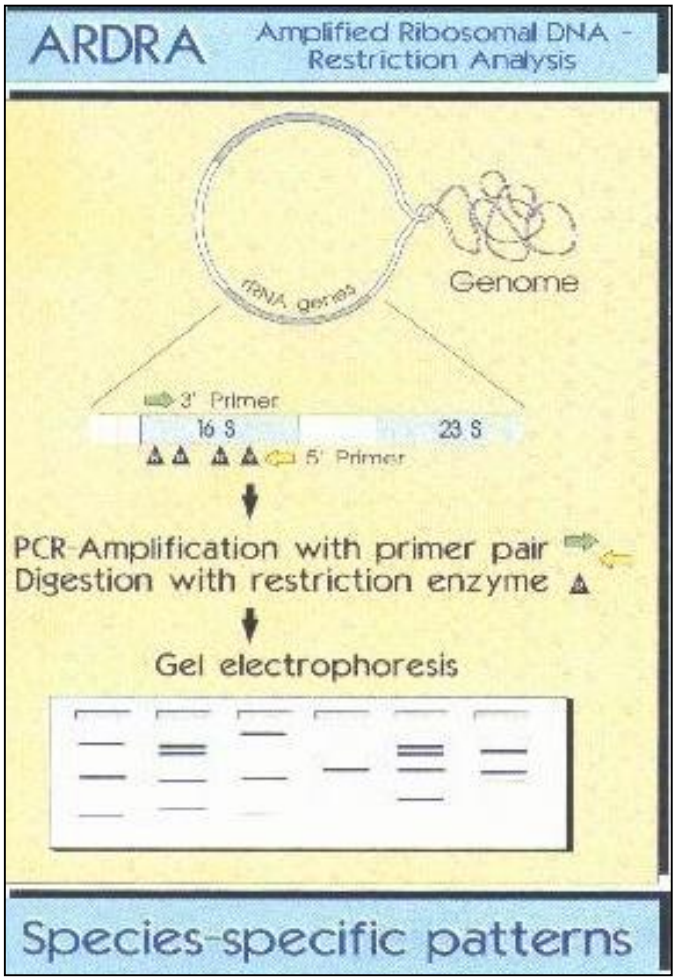
Osborn and Smith,2005

Microbial diversity

Screen Clones for differences

ARDRA is an RFLP on the 16SrDNA amplified by PCR

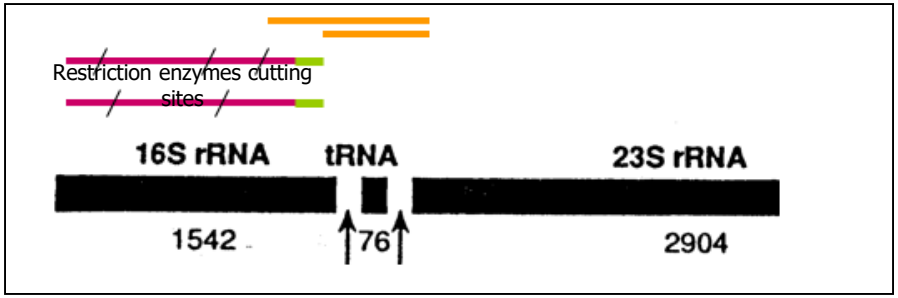
- ARDRA has been used:
1. post-cloning
2. pre-cloning
3. isolates or communities.
- Procedures:
- Extract DNA from sample;
- PCR Amplification of 16SrDNA;
- Clone all amplicons;
- Screen clones for differences (e.g. ARDRA).



Microbial diversity

Quick Assessment of Diversity
"One band = One bug" (?)
(little phylogeny information)

- **RISA (ribosomal intergenic spacer analysis):**
  - Often get overlapping bands (on band = > 1 bug)
  - Phylogenetic information limited by 16SrDNA overlap
- **ARDRA (amplified ribosomal DNA restriction analysis):**
  - ARDRA is an RFLP on the 16SrDNA amplified by PCR.
  - Good for identification of isolates; esp. with multiple restriction enzymes.
  - Too many bands makes it hard to interpret mixed populations.
- **T-RFLP (terminal restriction fragment polymorphism):**
  - Steps like ARDRA, but terminal 3' end of gene is fluorescent
  - Multiple restriction enz. Give best results; maybe used to query RDP.



Microbial diversity

Genotyping of bacterial isolates by ARDRA

- Amplified ribosomal DNA restriction analysis (ARDRA) was performed on the PCR-amplified 16S rDNA products from each of the isolates using three specific restriction *Hha*I, *Afa*I, and *Msp*I.
- Five microliters of each PCR product was digested for 2 h at 37°C with 1.5 U of each restriction endonuclease.
- Aliquots (5 µL) of each digested product were analyzed by gel electrophoresis in an 8% nondenaturing acrylamide gel (acrylamide: N,N'-Methylenebisacrylamide, 29:1) and by silver nitrate staining, as described previously.
- Fragment sizes were estimated using a low range, 50 bp DNA ladder and a final grouping of isolates was performed by a visual comparison of the restriction patterns.

Huang *et al.*,2012

ARDRA

Amplified Ribosomal DNA Restriction Enzyme Analysis

P. tolaasii

- The high degree of polymorphism cause problems in the identification of the *P. tolaasii*.
- Therefore, DNA samples were isolated from *P. tolaasii* strains and PCR reactions were performed with special primers.
- So called ARDRA was performed with *Cfo*I, *Hae*III, *Hinf*I and *Sac*I restriction enzymes, respectively.

1. The *Cfo*I *enzyme* provided 4 groups, but
2. The Hinf*I* and *Sac*I digestion resulted 8 different patterns.

Sajben et al

ARDRA

Amplified Ribosomal DNA Restriction enzyme Analysis

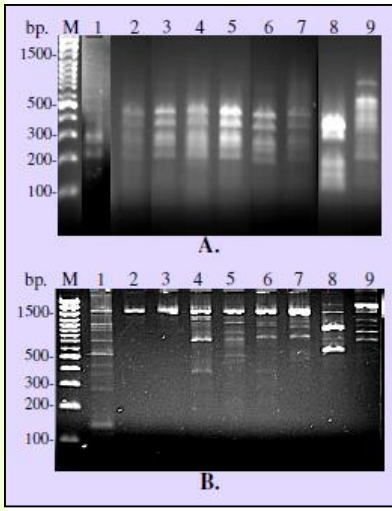
P. tolaasii

- The ARDRA pattern of the strains of *P. tolaasii*.

A. Digestion with *Cfo*I enzyme.

B. Digestion with *Sac*I enzyme.

- The strains in the lines as the following: 1: 6, 2: 7, 3: 8, 4: 29, 5: 17,6: 15, 7: 37, 8: 25, 9: 56.
- M: GeneRuler 100bp DNA Ladder Plus



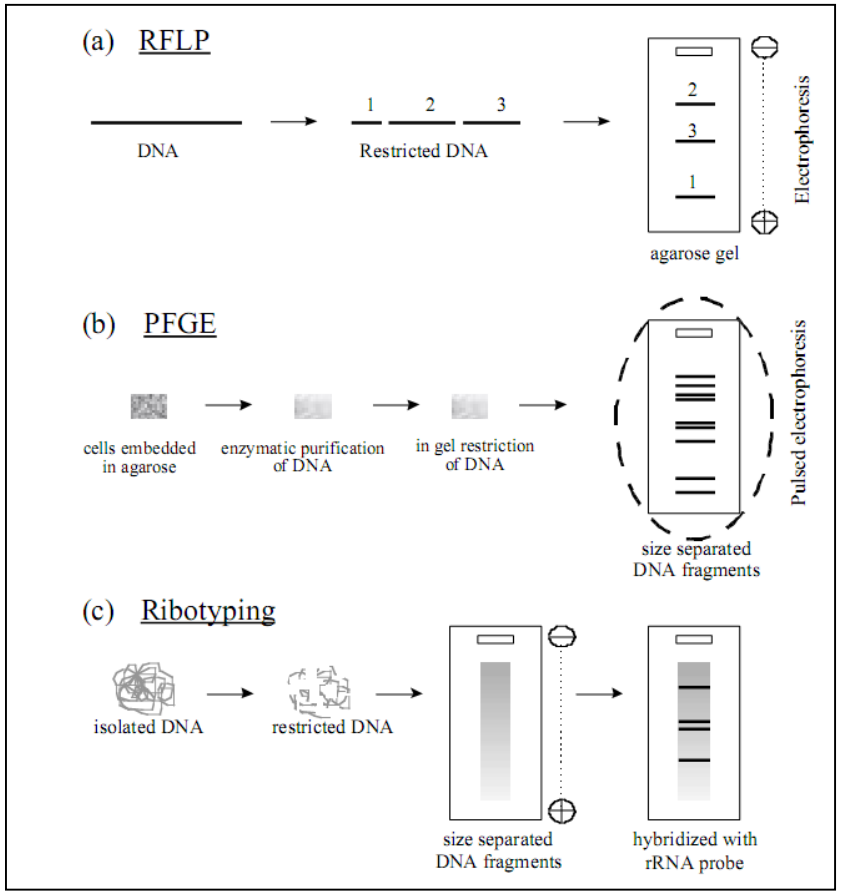
Pseudomonas strains	*Cfo*I	*Hae*III	*Sac*I	*Hin*fI
6	1	1	1	1
7	2	2	2	2
8	2	2	2	2
29	2	3	3	3
17	2	3	4	4
15	2	4	5	5
37	2	4	6	6
25	3	5	7	7
56	4	6	8	8

Table 2. Classification of the strains on the basis of ARDRA results

Sajben et al

Illustration of commonly used fingerprinting techniques
RFLP, PFGE and ribotyping

- **Ribotyping**
- DNA is isolated from a cultured isolate, restricted and size separated in an agarose gel.
- The gel is then hybridized with labeled rRNA probe, which binds to fragments containing copies of the rRNA operon.
- Following probe detection, fragments with bound probe are visualised, forming a characteristic RFLP.



O'Sullivan,2000

Typing methods
Chemotaxonomy

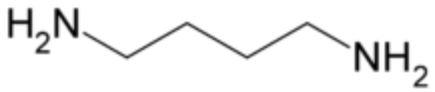
- Many of the cellular compounds which belong to the bacterial phenotype have been used in typing systems to characterize strains at the infraspecific level. e.g.

1. Polyamines (a compound containing more than one amino group. Many polyamines function as essential growth factors in microorganisms);
2. Xanthomonadins
3. Cell wall composition
4. Cellular fatty acids.

Vandamme *et al.*,1996;..

Typing methods

Polyamines (PAs)



putrescine

Considered as a plant hormone but the mechanisms of action of polyamines differ greatly from those of plant hormones

- **Polyamine** is an organic compound having two or more primary amino groups -NH_2.
- These substances play important roles in both eukaryotic and prokaryotic cells.
- The two most common polyamines found in bacteria are putrescine and spermidine.
- Polyamines present in bacterial nucleoids neutralize negative charges on phosphates (El-Safey,2011).
- If cellular polyamine synthesis is inhibited, cell growth is stopped or severely retarded.
- They appear to be essential in growth and cell division.

Typing methods

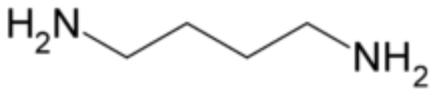
Polyamines (PAs)

Considered as a plant hormone but the mechanisms of action of polyamines differ greatly from those of plant hormones

- Polyamines are polycations that interact with negatively charged molecules such as DNA, RNA and proteins.
- PAs have specific and diverse roles in multiple cellular processes, including apoptosis (programmed cell death), cell division and differentiation, differentiation, DNA and protein synthesis, gene expression, signal transduction.
- What foods are high in polyamines?
- Mushrooms, peas, hazelnuts, pistachios, spinach, broccoli, cauliflower and green beans also contain significant amounts of both polyamines. The lowest levels are found in the fruit category.

Typing methods

Polyamines (PAs)



putrescine

Considered as a plant hormone but the mechanisms of action of polyamines differ greatly from those of plant hormones

- Apoptosis meaning "dropping off" and refers to the falling of leaves from trees in autumn.
- Apoptosis plays an important role in the treatment of cancer as it is a popular target of many treatment strategies.
- However, many troubling questions arise with the use of new drugs or treatment strategies that are designed to enhance apoptosis and critical tests must be passed before they can be used safely in human subjects.

Typing methods

Chemotaxonomy respiratory quinones

- The quinolones are a family of synthetic broad-spectrum antibacterial drugs.
- Respiratory quinolones are a class of antimicrobials with a high activity against most respiratory pathogens.

Typing methods

Isoprenoid quinones

Lipid molecules

- Isoprenoid quinones are one of the most important groups of compounds occurring in cytoplasmic membranes of living organisms.
- These Lipid molecules play important roles in electron transport, oxidative phosphorylation, and possibly, active transport.

There is difference between isoprenoid quinone composition (lipid molecules of cytoplasmic membranes) and whole-cell fatty acid profiles. Fatty acids are the major constituents of lipids and lipopolysaccharides and have been used extensively for taxonomic purposes.

Typing methods

Isoprenoid quinones

Lipid molecules

- Lipid molecules present in all species of respiratory and photosynthetic microorganisms and exhibit marked structural variations depending upon the microbial taxon.
- Taking advantage of this, quinones have been used:

1. As chemotaxonomic markers in microbial systematics, but also
2. As good measures of microbial populations in the environment in terms of quantity, quality, and activity.

Typing methods

Isoprenoid quinones

Lipid molecules

- Two major structural groups of bacterial isoprenoid quinones can be recognized:

1. Naphthoquinones
2. benzoquinones.

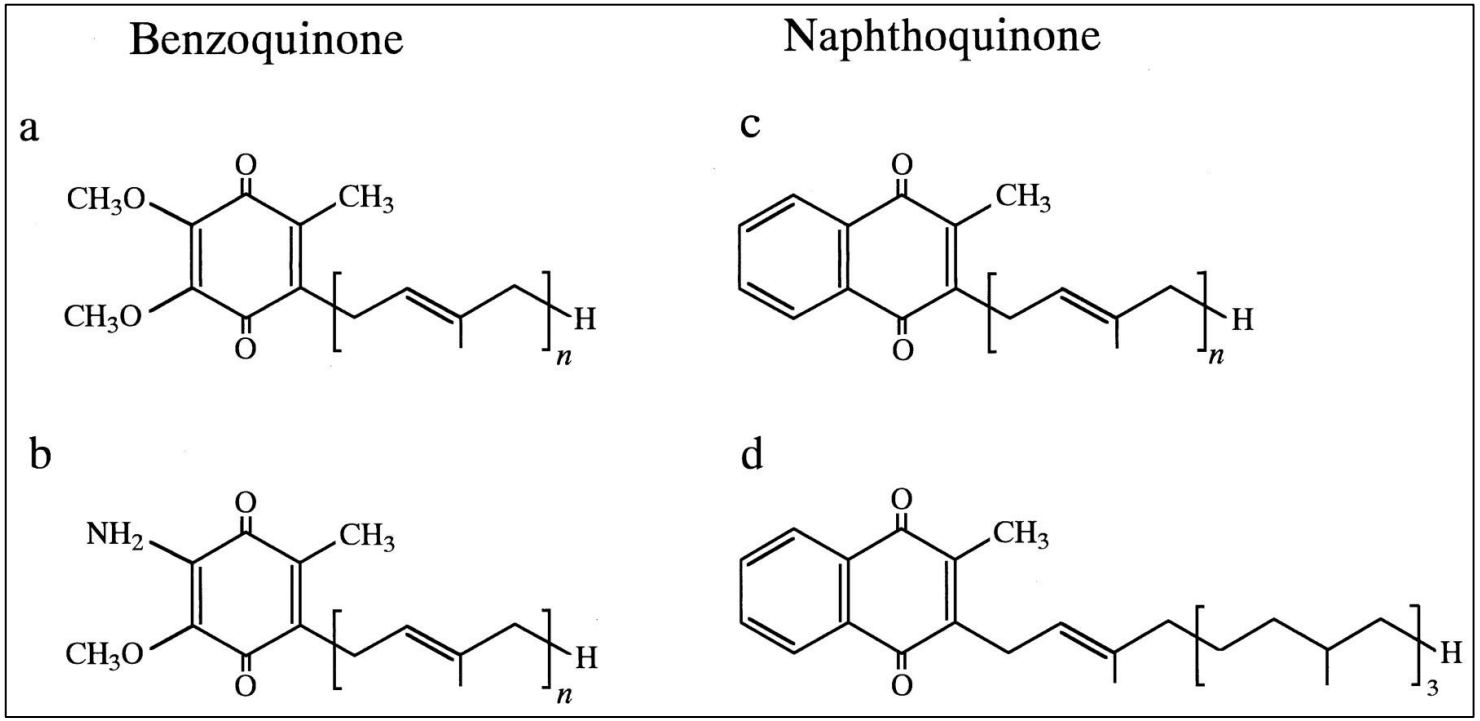
- Naphthoquinones can be divided further into two main types on the basis of structural considerations. These are the:

1. phylloquinones
2. Menaquinones.

Collins and Jones,1981

Typing methods

Four quinone structural classes found in activated sludge



The large variability of the side chains (differences in length, saturation, and hydrogenation) can be used to characterize bacteria at different taxonomic levels.

Vandamme,2013

Typing methods

Isoprenoid quinones

Determination of isoprenoid quinones

- Cells were hydrolyzed by adding 0.2 ml of 50% aqueous KOH and 3 ml of 1% pyrogallol in methanol and boiling (100°C) for 10 min.
- The cooled hydrolysates were then extracted with ether-hexane (1:1) as previously described.
- Quinones were finally dissolved in methanol and examined by reverse-phase high performance liquid chromatography (HPLC).

Dess *et al.*,1986

Different thresholds for a species delineation

Comparison of DDH and 16S rRNA gene sequence similarity

Gold standards for bacterial species demarcation

- DNA-DNA hybridization (DDH) values have been used by bacterial taxonomists since the 1960s to determine relatedness between strains and are still the most important criterion in the delineation of bacterial species.
- Commonly, a species is defined as a set of strains with approximately 70% or greater DNA-DNA relatedness or 97% 16S rRNA identity.

Gupta and Sharma,2015

Different thresholds for a species delineation

Comparison of DDH and 16S rRNA gene sequence similarity

Gold standards for bacterial species demarcation

- DNA-DNA hybridization (DDH) is technically challenging, labor intensive and a time consuming method.
- The DDH values lower than 50% cannot be used for estimating the genetic relatedness between distantly related species.
- In case of 16S rRNA gene sequence analysis, unlike DDH, the information is archival and can be used for various analyses.

Gupta and Sharma,2015

Different thresholds for a species delineation

Comparison of DDH and 16S rRNA gene sequence similarity

Gold standards for bacterial species demarcation

- Though the 16S rRNA gene sequence is capable of classifying a genome to the family or genus level, it is not very efficient in the differentiation of species.
- For example, organisms with greater than 97% sequence identity may still belong to different species.

Gupta and Sharma,2015

Different thresholds for a species delineation

Comparison of DDH and 16S rRNA gene sequence similarity

Gold standards for bacterial species demarcation

- Multi-locus sequence analysis (MLSA) of housekeeping genes is another molecular method which has recently become popular for investigating taxonomic relationships.
- MLSA of selected housekeeping genes accurately predicts the relationships between closely related genomes without the need for genome-wide comparison.
- However, the main limitation in MLSA is the dependence on the choice of housekeeping genes which varies between different taxa.

Gupta and Sharma,2015

DNA–DNA hybridization (DDH)

Nucleic acid base composition

G+C content

- The G/C content, or ratio, is the percentage of nitrogenous bases on a DNA molecule which are either guanine or cytosine.
- In PCR procedures, the GC content of primers are used to determine their annealing temperature to template DNA.
- A high GC content level indicates a higher melting temperature as it takes more energy to break three hydrogen bonds versus two.

Kotrla,2014

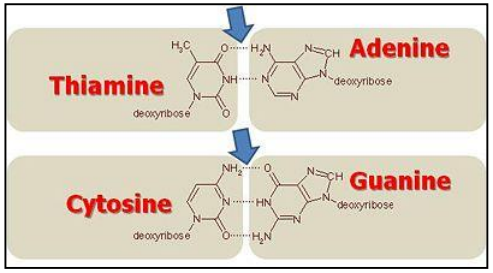
DNA–DNA hybridization (DDH)

Nucleic acid base composition

G+C content

- Here is a sample calculation.

$$\frac{G+C}{A+T+G+C} \times 100$$



- You have a fragment of DNA with the following structure: AAGGTATCGGTATCGAAATT
- Count up the bases and plug into the formula:

7 As, 6 Ts, 5 Gs and 2 Cs

5+ 2/7+6+5+2 x 100=

7/20 x 100=

35% G/C Content.

Count up the bases and plug into the formula: 7 As, 6 Ts, 5 Gs and 2 Cs

5+2/7+6+5+2 * 100 =
7/20 * 100 =
35% G/C Content

Kotrla,2014

DNA–DNA hybridization (DDH)

Nucleic acid base composition

G+C content

$$\text{Mol\%}\ \ (G+C) = \frac{G+C}{G+C+A+T} \times 100\%$$

- Estimated by determining the melting temperature (T*m*) of the DNA.
- Higher G+C gives a higher melting temperature.
- Closely related organisms should have similar G+C ratio.
- e.g.
- *Erwinia* GC% 50-58
- *Agrobacterium* GC% 57-63
- *Pseudomonas* GC% 58-70
- *Bacillus* GC% 30-70
- *Clostridim* GC% 22-55

Hendrix,2004

DNA-DNA hybridization

DNA-DNA reassociation

1. DNA of unknown is labeled (growth in 3 H thymidine; nick translation, nonradioactive methods),
2. Denatured (heat),
3. Reacted with denatured test DNA bound to a nitrocellulose filter, washed, amount of radioactivity retained on filter reflects % homology.

- The more closely related the strains are, the greater the homology.

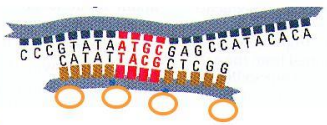
El-Safey,2011

DNA-DNA hybridization

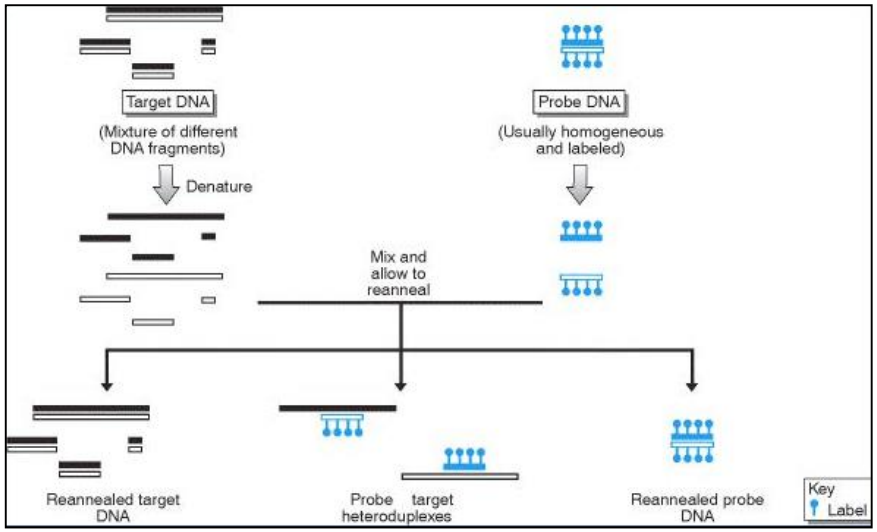
DNA-DNA reassociation

Hybridisation

- **Hybridisation:** process in which a labeled single stranded piece of DNA (RNA) finds (and basepairs with) the complementary strand

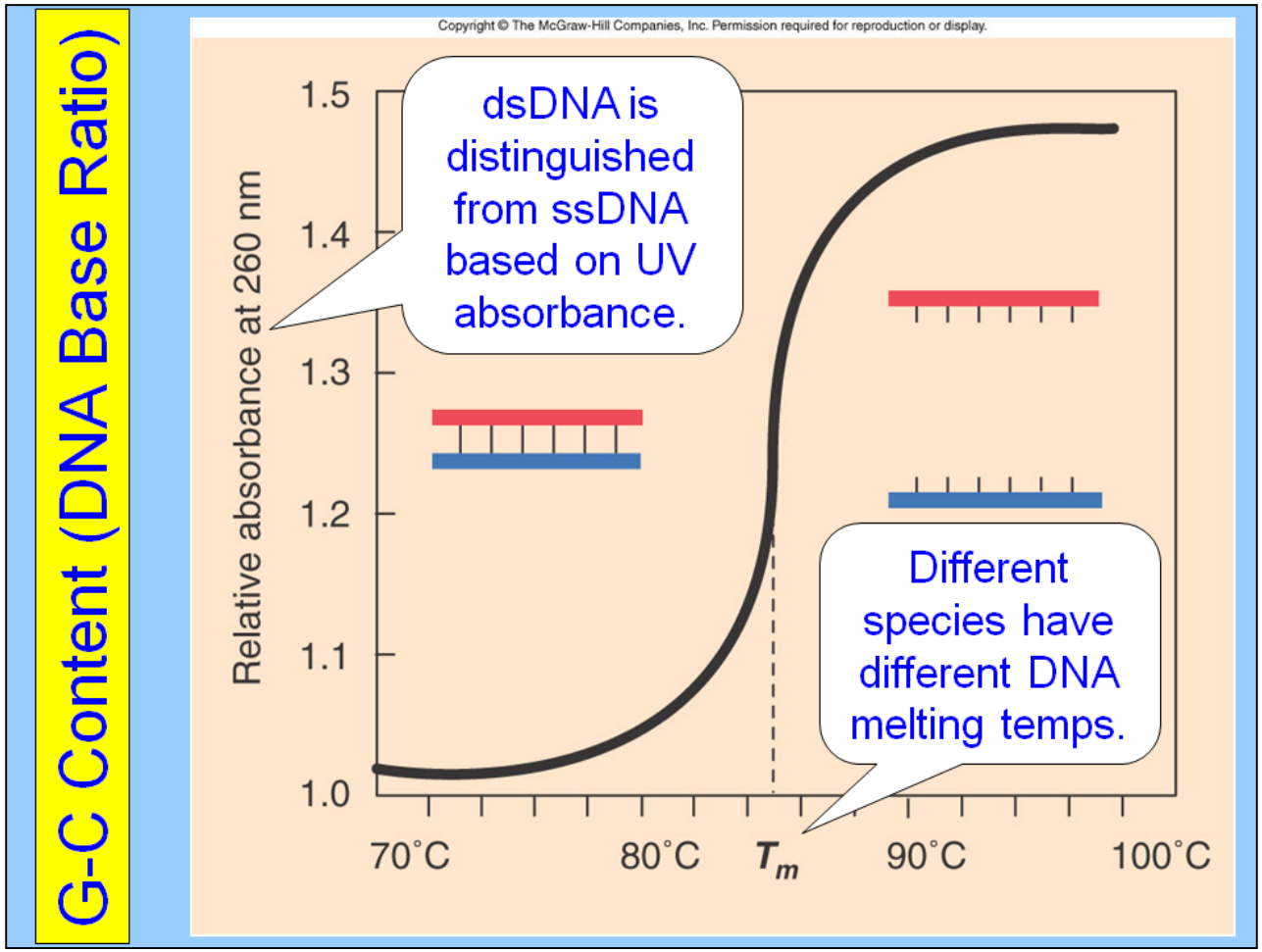


- **Probe:** piece of single-stranded DNA (or RNA) that is labeled (○) to detect the corresponding complementary sequence in hybridisation



Nucleic acid base composition

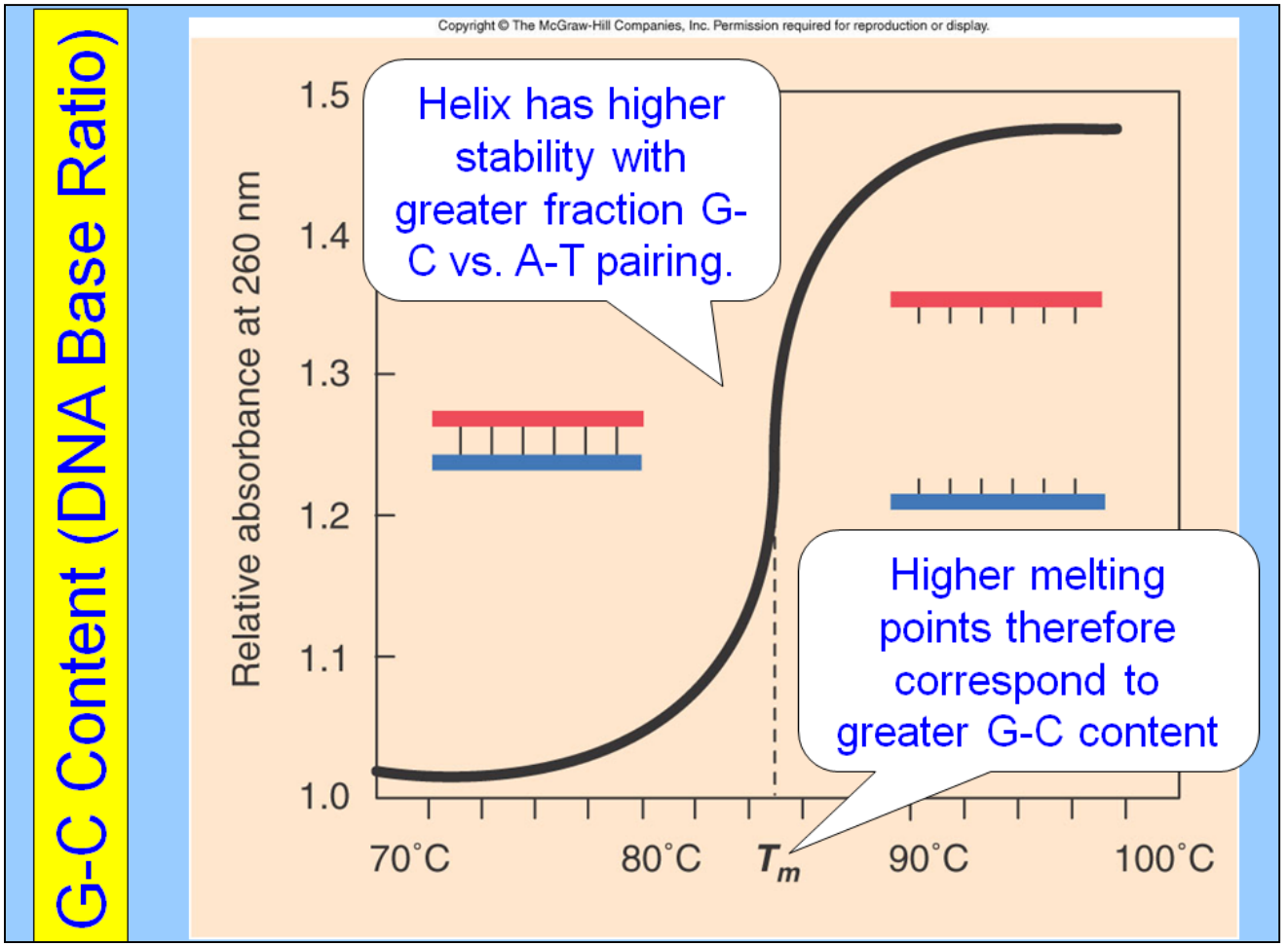
G+C content



Abedon,2011

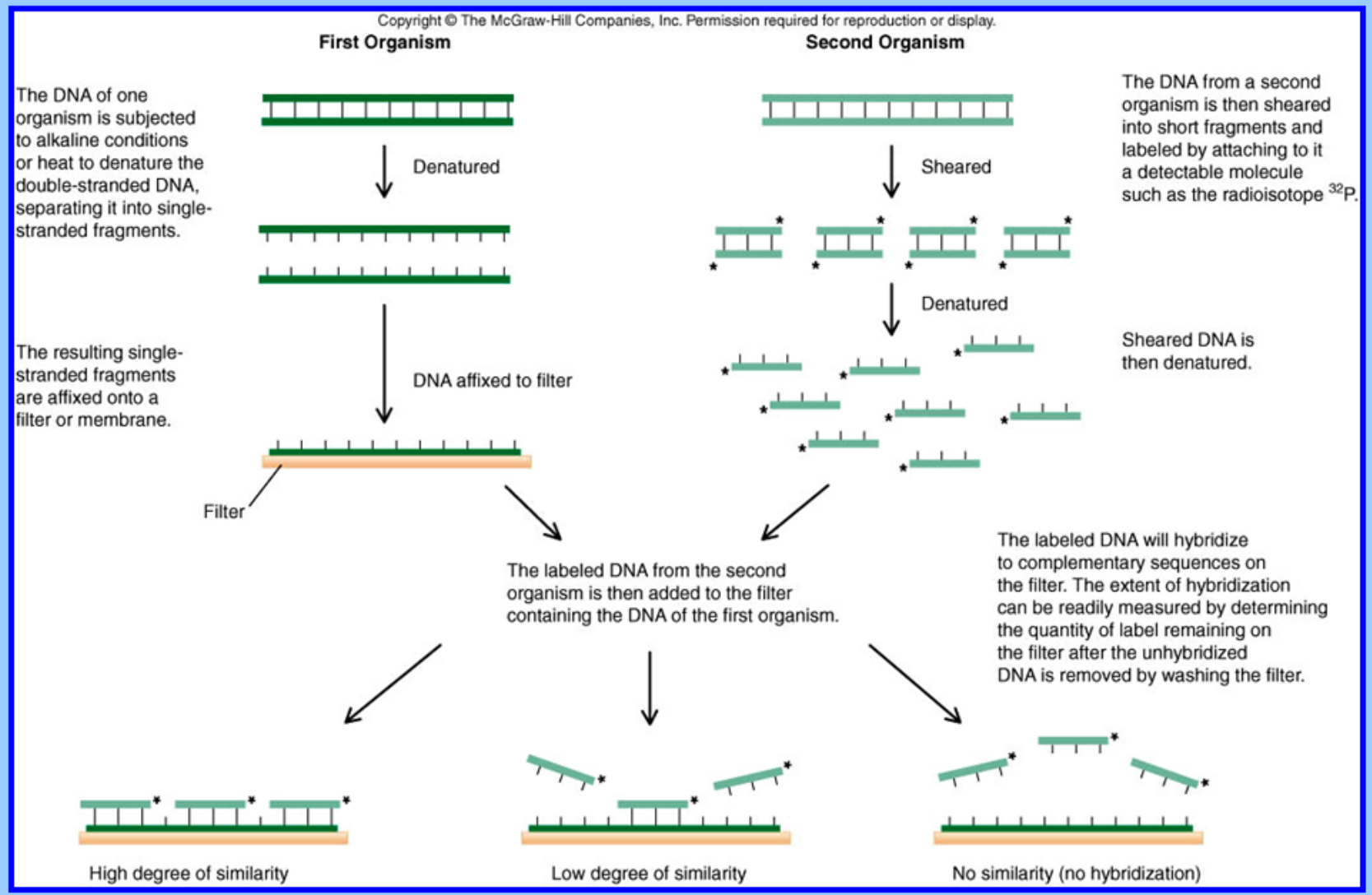
Nucleic acid base composition

G+C content



Abedon,2011

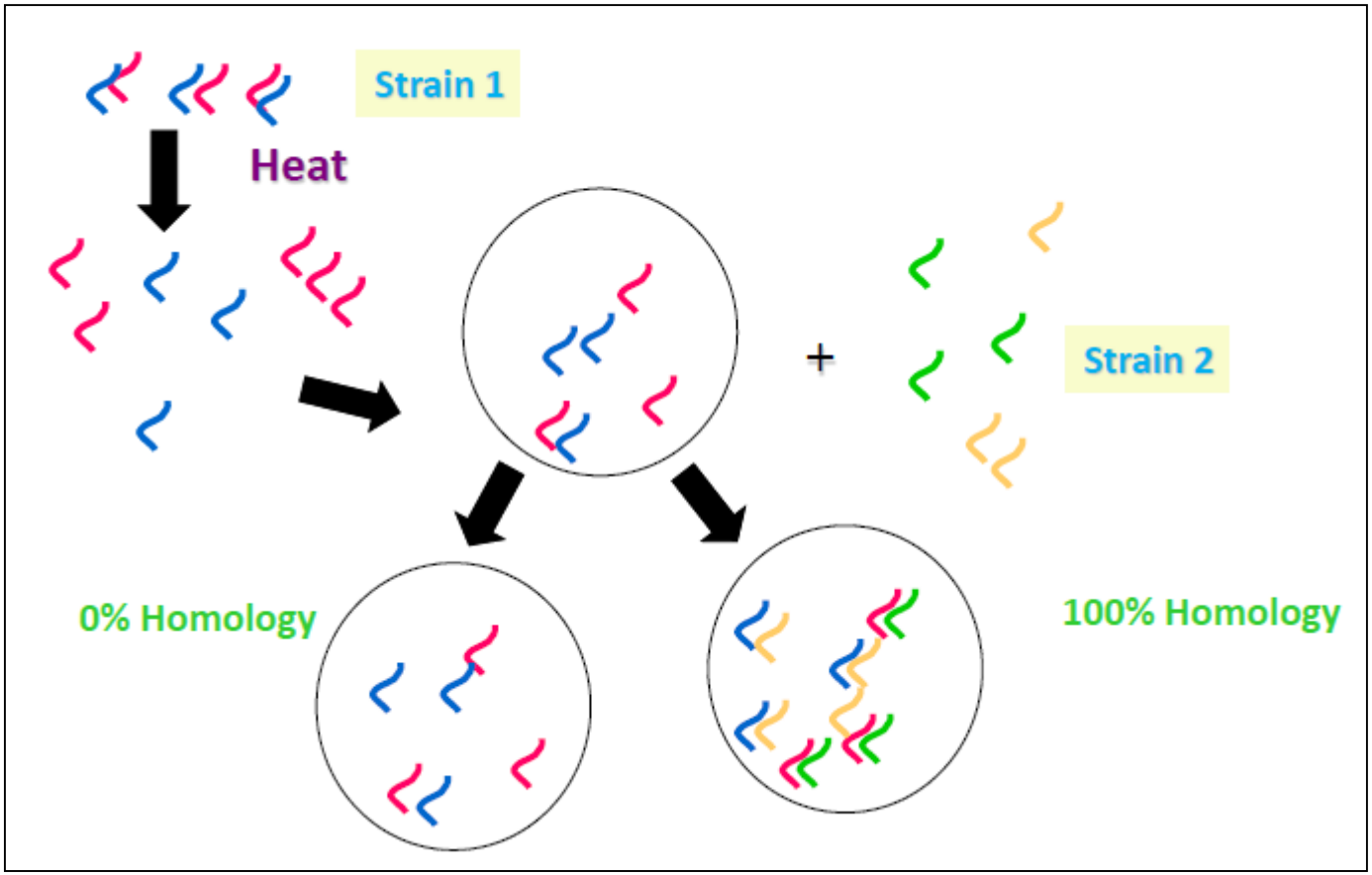
DNA-DNA Hybridization



Abedon,2011

DNA-DNA hybridization

Typing via RFLP analysis



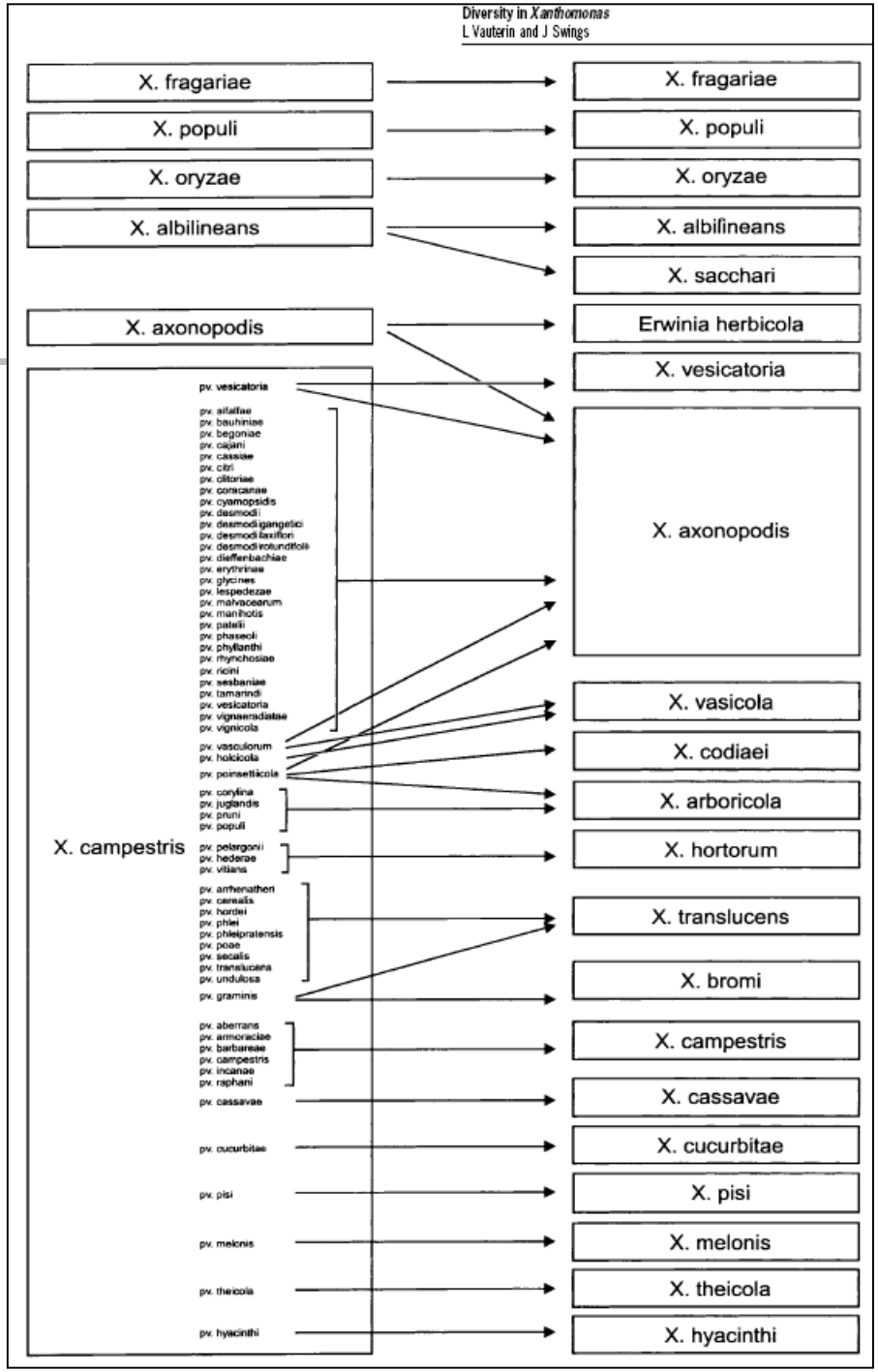
Fox,2008

Classification

Based on DNA homology

Schematic representation of the rearrangements proposed within the genus *Xanthomonas*, resulting from a global taxonomic study of more than 1000 strains and DNA hybridization experiments between 183 selected strains.

Vauterin and Swings,1997



Classification of *Acidovorax*
Based on DNA-DNA hybridization

- Native DNA of two *Acidovorax valerianellae* causal agent of lamb's lettuce strains, CFBP 4730T and CFBP 4723, was labelled with tritiated nucleotides (^{3}H nucleotides) by nick-translation.
- The S1 nuclease/trichloroacetic acid method was used as indicated by Gardan *et al.* 2000.
- The reassociation temperature was 70° C.
- Levels of DNA relatedness among *Acidovorax valerianellae* and related strains hybridization was determined at 70° C.
- ND, Not determined.

Source of unlabelled DNA	Relative binding with labelled DNA from:	
	CFBP 4730^T	CFBP 4723
A. valerianellae sp. nov.		
CFBP 4730^T	100	91
CFBP 4720	100	100
CFBP 4721	84	100
CFBP 4723	100	100
CFBP 4725	100	98
CFBP 4726	95	99
CFBP 4728	89	88
CFBP 4731	100	100
CFBP 4732	100	93
CFBP 4733	92	89
CFBP 4734	100	100
A. anthurii CFBP 3232^T	24	ND
A. avenae subsp. *avenae*		
CFBP 2425^T	19	ND
CFBP 1201	23	ND
A. avenae subsp. *cattleyae* CFBP 2423^T	35	ND
A. avenae subsp. *citrulli* CFBP 4459^T	29	ND
A. konjaci CFBP 4460^T	15	ND

Gardan *et al.*,2003

DNA-RNA homology
More sensitive method

1. DNA homology does not go much beyond the genus or family level.
2. RNA (esp. rRNA) does.

- Same principles as DNA homology but hybridization is between DNA and rRNA.
- Three types of RNA: mRNA, tRNA, rRNA.
- More sensitive - can pick homologies not detected by DNA-DNA hybridization.
- Because

1. rRNA is highly conserved - critical for protein synthesis.
2. rRNA constitutes small % of total genetic material - therefore, examining specific conserved genes, not whole chromosome that is less conserved.

El-Safey,2011

Specific genes

Phylogenies based on specific genes

- There are several important requirements if we wish to use a gene phylogeny to infer organismal relationships:

1. The gene must be present in all organisms of interest. Examples include genes whose products function in replication, transcription, or translation.
2. The gene can not be subject to transfer between species (lateral transfer).
3. The gene must display an appropriate level of sequence conservation for the divergences of interest.
4. The gene must be sufficiently large to contain a record of the historical information.

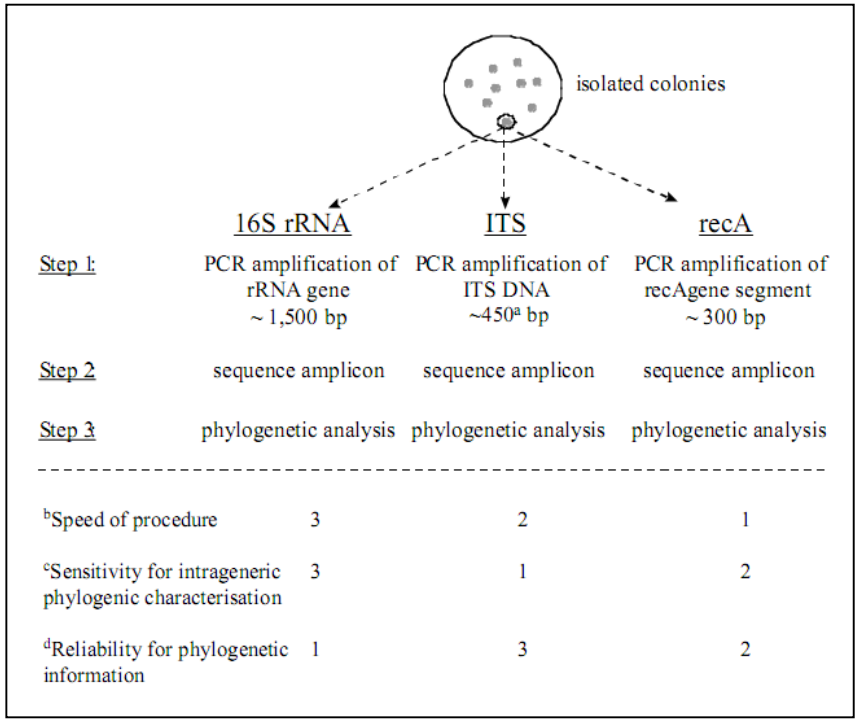
See also Species Concept in Prokaroytes

Specific genes

Phylogenies based on specific genes

- Comparison of the three sequence-based typing and phylogenetic characterization approaches.
- The three sequences which have been used are:

1. 16S rRNA;
2. Sequence between the 16S and 23S rRNA genes, termed the internal transcribed spacer (ITS); and
3. An internal portion of the recombinase *recA* gene.



An amplicon is a piece of DNA or RNA that is the source and/or product of amplification or replication events.

O'Sullivan,2000

Specific genes

Housekeeping (metabolic/essential)genes

Phylogenetic markers

- Molecular techniques in a comparative analysis of housekeeping genes such as *oprI*, *rpoD*, *gyrA*, *gyrB*, etc. but also macromolecules like 16S rRNA.
- A housekeeping gene is a gene that codes for proteins needed all the time.
- Most investigators suggest that the presence of the almost universally conserved essential or housekeeping genes defines a chromosome.
- These could include:

1. 16S rRNA genes,
2. Heat-shock proteins (HSPs) such as *groE*, *dnaK*, *gyrB* and *rpoD*.

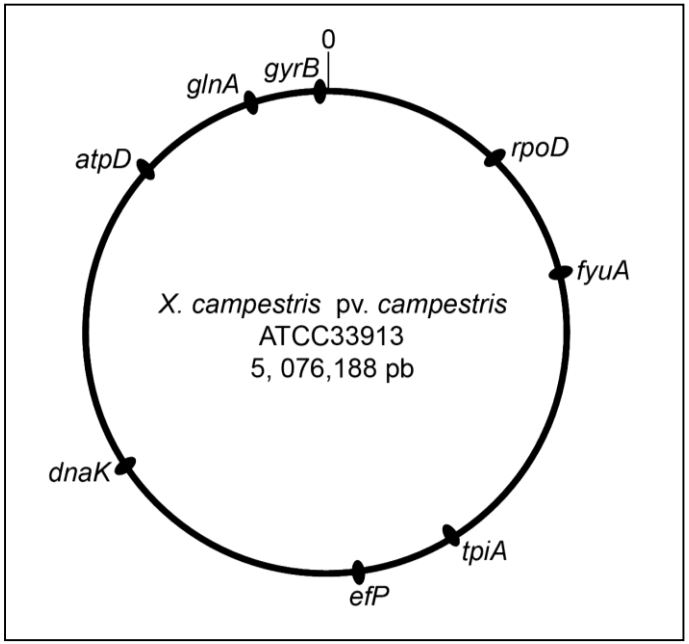
Heat shock proteins (HSP) are expressed in response to various biological stresses, including heat, high pressures, and toxic compounds. It is also one of the most abundant cellular proteins found under non-stress conditions.

Specific genes

Housekeeping (metabolic/essential) genes

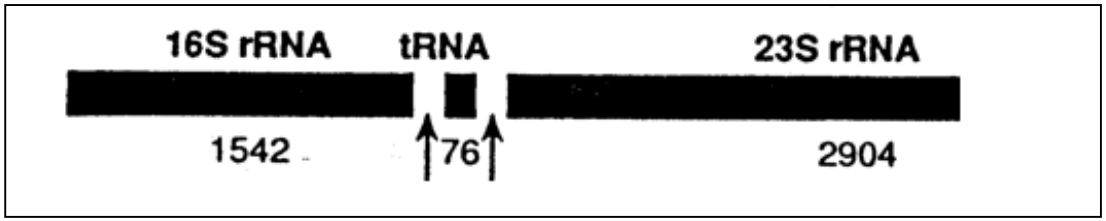
X. c. pv. *campestris*

- Schematic of the relative location of the seven housekeeping genes and the *fyuA* gene used in this study on a schematic chromosomal map of *X. c.* pv. *campestris* strain ATCC 33913.
- **Heat-shock proteins:**
- *gyrB* (DNA gyrase subunit B);
- *rpoD* (RNA polymerase sigma-70 factor);
- *dnaK* (heat shock protein 70, molecular chaperone DnaK);
- **Other proteins:**
- *atpD* (ATP synthase beta chain);
- *efP* (elongation factor P);
- *glnA* (glutamine synthetase I);
- *tpiA* (triosephosphate isomerase) and
- *fyuA* (TonB-dependent receptor).



Specific genes
Signature sequences
16S rDNA sequences

- Particular groups of microbes will possess unique nucleotides at specific locations in 16S rRNA molecules.



Chester R. Cooper, Jr.,2004;..

Specific genes

16S rDNA sequences (signature sequences)

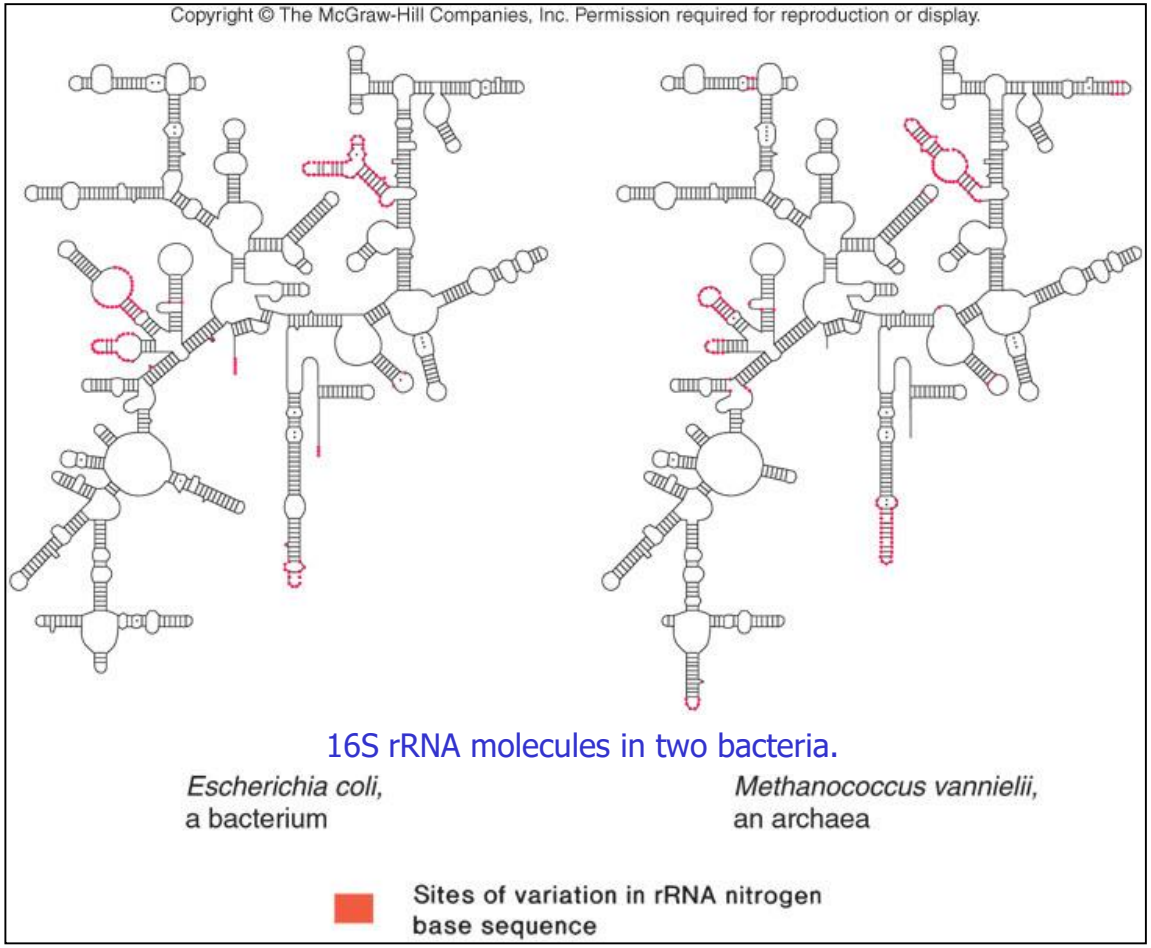
The first-line molecular tool

- Knowledge of bacterial diversity obtained after more than 100 years of pure culture study is incomplete, and very few of the total number of microbial species are in culture (Torsvik *et al.*,1995).
- However, phenotypic methods are restricted to only those bacteria that can be isolated and cultured.
- So far, a small portion of all prokaryotes has been cultivated, and the majority of soil bacteria observed microscopically can not be cultivated.
- Since the mid-1980s, the use of small-subunit ribosomal ribonucleic acid (SSU rRNA) based techniques has facilitated a culture independent approach of investigating microorganisms as they occur in nature.

Specific genes

The conserved macromolecule 16S rDNA

Red color: site of variation in rRNA nitrogen base sequence



Powers,2010

rRNA sequence

16S rRNA sequence analysis

Molecular clock

- Differences in the nucleotide sequence are used to classify prokaryotes.
- The DNA sequence that encodes rRNA is highly conserved among bacteria of common ancestry.
- Sequence DNA encoding rRNA of different isolates, determine % identity, have a molecular chronometers (molecular clock)–an accurate determination of phylogenetic distance.
- Phylogenetic trees are now established on basis of rRNA sequence.

El-Safey,2011

Specific genes

16S rDNA sequences (signature sequences)

Less than 97% identical usually hybridize less than 70%

- Prokaryote species are most precisely defined on the basis of their 16S rRNA sequences.
- Such a convention is supported in that DNA from two bacteria whose 16S rRNA sequences are less than 97% identical usually hybridize less than 70%.
- When to claim discovery of a new genus using nucleic acid data is on somewhat shakier ground than that for species.
- Groups of sufficiently similar species make up a genus and a collection of sufficiently similar genera make up a family.

Specific genes

16S rDNA sequences (signature sequences)

- The rRNA sequences were shown to be a very useful molecular marker for phylogenetic analyses.
- Among the three rRNA molecules, 16S rRNA has been the most widely studied.
- Thus most information is available for this molecule.
- This information is provided in the description of most of the newly classified bacterial species.
- It is also becoming increasingly popular to propose new bacterial species using data generated from 16S rRNA sequencing studies.

Ribosomal RNAs

A powerful evolutionary and investigative biomarkers

- Ribosomal RNA and the corresponding genes (*rrn*) are now widely used as powerful evolutionary and investigative biomarkers for the following reasons (Olsen *et al.*,1986):

i. Ribosomal RNAs are essential to protein synthesis, and therefore are ubiquitous to all organisms, and structurally and functionally conserved;

ii. Ribosomal RNAs are readily isolated and identified;

iii. They contain variable and highly conserved regions in both primary and secondary structure; and

iv. They appear to change in sequence very slowly, and they do not exhibit horizontal gene transfer found with many other prokaryotic genes; therefore relationships between rRNAs reflect evolutionary relationships.

16S ribosomal DNA

A set of 16S rDNA PCR primers for exploring bacterial diversity

- Most of these new methods are based on sequences of the 16S rRNA gene, a gene encoding a molecule of RNA used in bacterial and archaeal ribosomes.
- The 16S rRNA gene is approximately 1500 bases in length and contains regions that are:

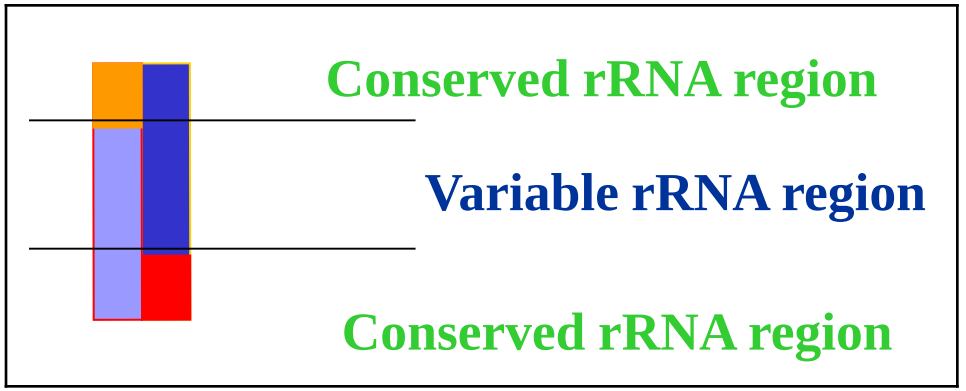
1. Highly 'conserved' (i.e., have the same sequence in all bacteria and archaea), and
2. Highly 'variable' (i.e., have sequences that are unique at the genus or species level).

- Thus the conserved regions of the gene can be used to bind primers for PCR and sequencing, and the variable regions to determine the identity of the organism.

Moran,2010

Discovering new species
Based on ribosomal DNA genes

- Recover DNA from different environments;
- Amplify the ribosomal DNA genes (rDNAs) using relatively conserved primers.
- Sequence the variable region inside, and compare sequences with web based Ribosomal Database Project – RDP.



✓Amplify all bacterial DNA;
✓Assess biodiversity of variable region.

Web based Ribosomal Database Project - RDP

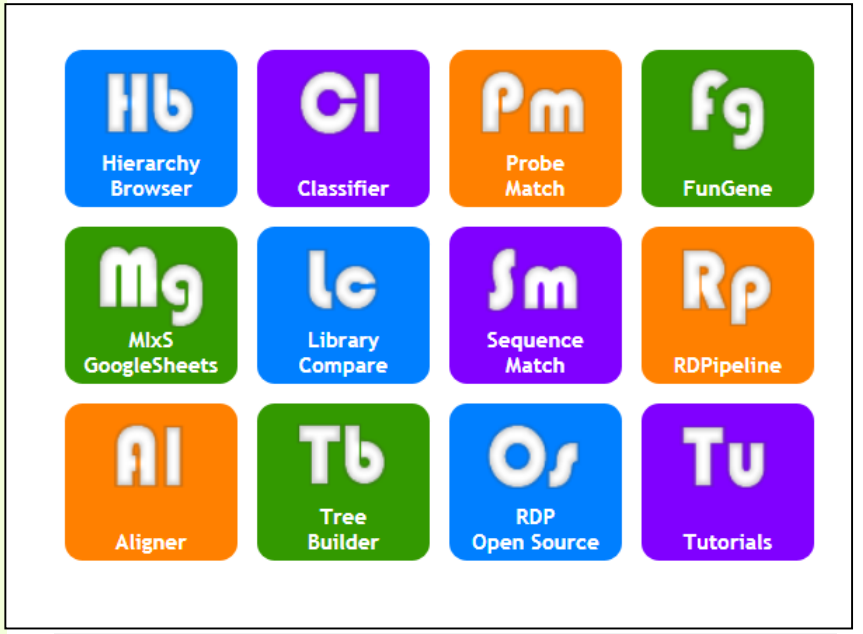
Based on ribosomal DNA genes

- The Ribosomal Database Project (RDP) provides ribosome related data and services to the scientific community, including:
- Online data analysis, and
- Aligned and annotated (explain) bacterial and archaeal small-subunit 16S rRNA sequences.
- To date, 3,356,809 16S rRNAs :: 125,525 Fungal 28S rRNAs have been made available for comparison.
- As these databases rapidly expand, they constantly improve the process of matching new sequences to known microorganisms.

Web based Ribosomal Database Project - RDP

Based on ribosomal DNA genes

- RDP provides quality-controlled, aligned and annotated Bacterial and Archaeal 16S rRNA sequences, and Fungal 28S rRNA sequences, and a suite of analysis tools to the scientific community.
- New to RDP release 11:
- Most of the RDP tools are now available as open source packages for users to incorporate in their local workflow.



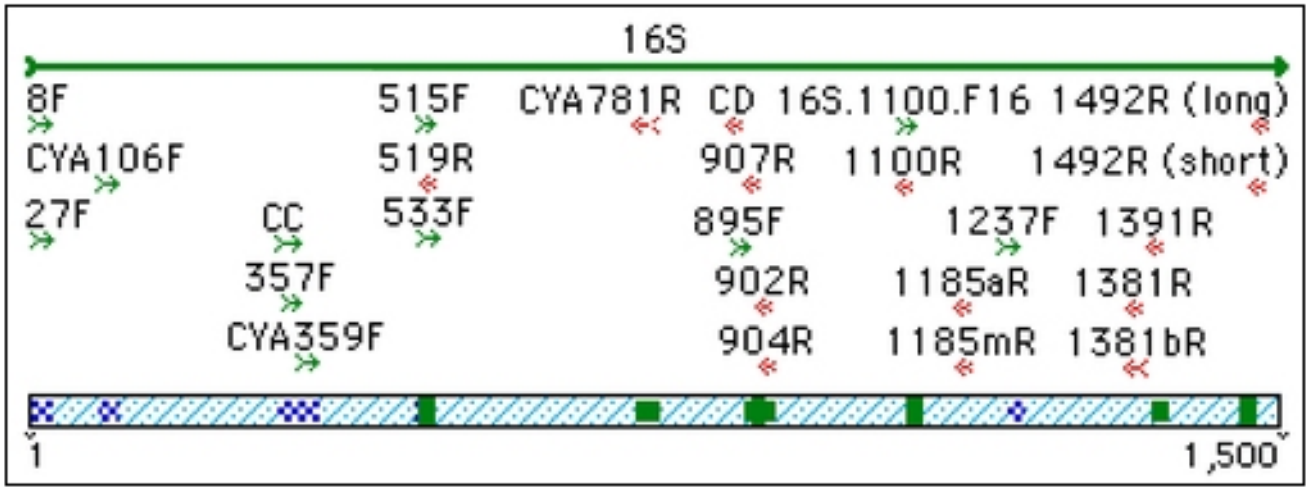
To date, 3,356,809 16S rRNAs :: 125,525 Fungal 28S rRNAs have been made available for comparison.

RDP Taxanomy 18 :: August 14, 2020

Primer map for the 16S SSU rRNA gene

A set of 16S rRNA gene primers

- The 16S rRNA gene is approximately 1500 bases in length.
- The primer map for the 16S SSU rRNA gene includes both universal (8F, 27F,515F, 907R, 1492R...) and specific primers (902R,904R,...).



Lutzoni Lab; Bélisle,2010

rRNA sequence

23S rRNA gene sequence analysis

Another molecular clock

- The 23S rRNA gene like 16S rRNA most widely used molecular chronometers for:

1. Inferring microbial phylogeny, and
2. Have been instrumental in developing a comprehensive view of microbial phylogeny and systematics.

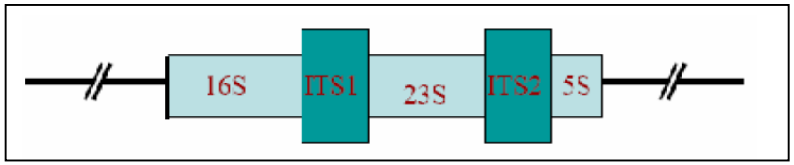
Specific genes
16S-23S rDNA intergenic spacer (ITS)

- Another molecular phylogenetic marker, the 16S-23S rDNA intergenic spacer (ITS), for distinguishing between these species.
- PCR-RFLP of ITS sequences from several bacterial species e.g. *Xanthomonas* species has established a consistent degree of variation among them.
- ITS sequences are located between the 16S and 23S ribosomal subunits and, in most Gram-negative bacteria, contain a $tRNA^{Ala}$ and a $tRNA^{Ile}$ in their sequence.
- ITS sequences have been under less intensive selection pressure and are considered to be 10 times as variable as 16S rDNA.

$tRNA^{Ile}$:A transfer RNA which is specific for carrying isoleucine to sites on the ribosomes in preparation for protein synthesis.

Specific genes
Based on 16S-23S rDNA ITS

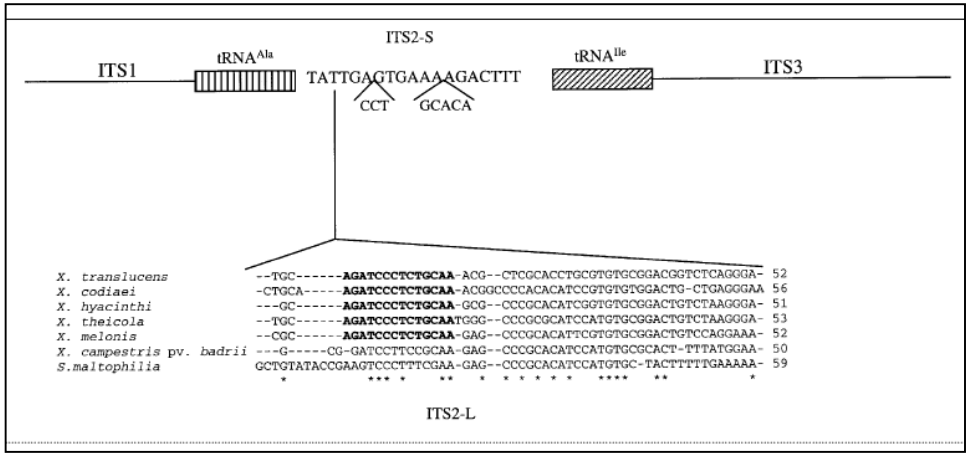
- Several other criteria have also been applied for differentiation at the species, pathovar or strain levels.
- E.g.
- All strains within the genus *Xanthomonas* that showed 70% or more DNA–DNA homology exhibited 100% 16S rDNA sequence similarity.
- Better discrimination was obtained by analysis of the 16S-23S rDNA intergenic spacer (ITS) sequence, which showed approximately ninefold higher diversity than 16S rDNA.



Genomic structure of rRNA gene complex in prokaryotes. Internal Transcribe Spacers (ITS) are highly variable regions within the rRNA gene complex and are typically used for identification of closely related bacterial species.

Overall structure of the 16S-23S rDNA ITS *Xanthomonas*

- These tRNAs were highly conserved and divided the ITS sequence into three regions (ITS1, ITS2 and ITS3).
- Conserved nucleotide blocks are indicated in bold; consensus nucleotides are indicated by asterisks.
- Major differences in size were found in the ITS2 region.
- Most *Xanthomonas* species had a shorter sequence nt (named ITS2-S), but five species exhibited a longer sequence, of 78-85 bases (ITS2-L) caused either by:
  - The addition of a long stretch of 51-56 nt and two smaller (CCT and GCACA) sequences, or by
  - The loss of these sequences from the other species.

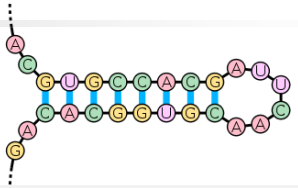


Gonçalves and Rosato,2002

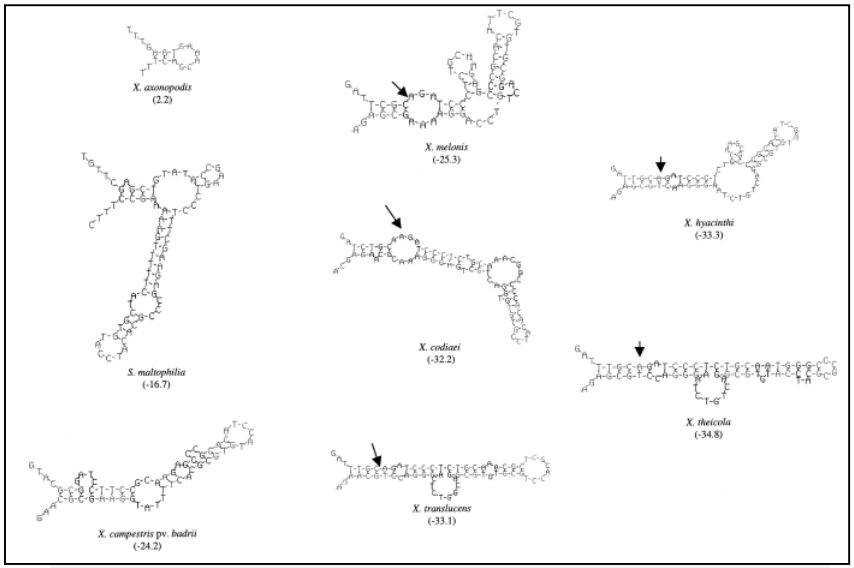
Putative secondary structures of ITS2 Xanthomonads

Hairpin structure

- Although ITS2-S was a shorter sequence, formation of a small secondary hairpin structure (stem-loop) could be predicted.
- Stem-loop intramolecular base pairing is a pattern that can occur in single-stranded DNA or, more commonly, in RNA (Wikpedia,2012).



An example of an RNA stem-loop(hairpin or hairpin loop).

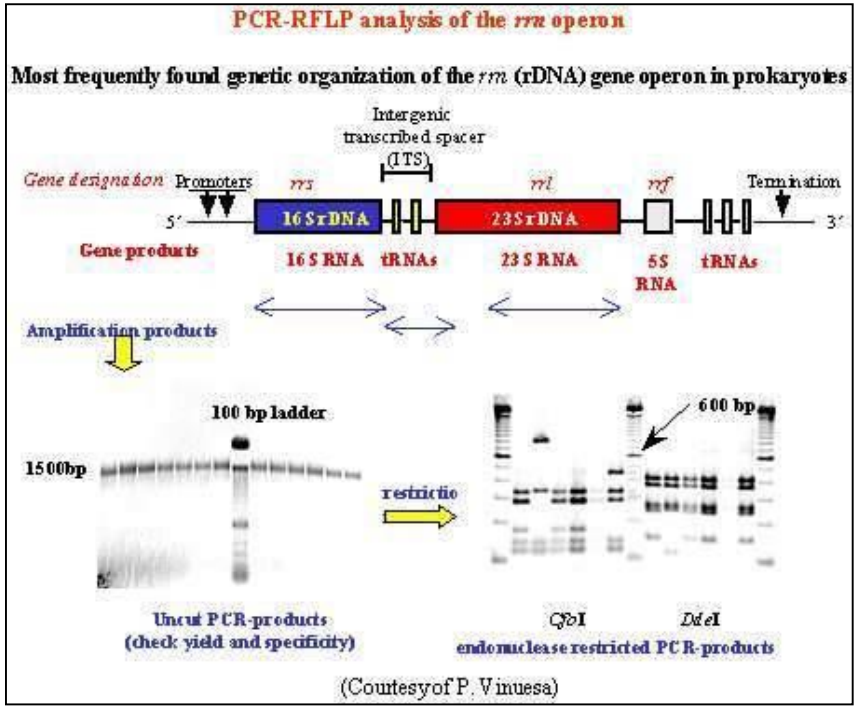


Arrows refer to 5´end of the conserved 14-nt block). Values in parentheses indicate free energy (kcal mol^{-1}).

Gonçalves and Rosato,2002

Specific genes
16S-23S rDNA intergenic spacer (ITS)

- RFLP analysis is discriminative at low taxonomic level, often strain level and is now often combined with PCR, e.g. digestion after PCR of the only some loci e.g. 16S rRNA gene (Coenye *et al.*,2001).
- PCR-RFLP is the faster and easer way method.



rrs" genes encode 16S rRNAs,"*rrl*" genes encode 23S rRNAs, and "*rrf*" genes encode 5S rRNAs. The rrf, rrs and rrl rRNA genes encode for the structural rRNA molecules required for ribosome assembly and function (5S, 16S and 23S rRNAs, respectively).

Specific genes

Based on Length and GC content of ITS sequences from *Xanthomonas* species

Strain	Spacer length (bp)				G+C content (mol%)
	ITS1	ITS2	ITS3	Total	
X. albilineans ICMP 196	106	14	223	497	53·50
X. arboricola pv. *juglandis* ICMP 35^T	109	19	226	507	51·08
X. axonopodis pv. *axonopodis* LMG 538^T	113	18	232	519	51·63
X. axonopodis pv. *passiflorae* ICMP 3151^P	113	18	229	507	53·06
X. bromi LMG 947^T	112	19	228	510	51·96
X. campestris pv. *campestris* LMG 568^T	112	19	225	509	50·69
X. cassavae LMG 673^T	113	19	228	513	51·66
X. codiaei LMG 8678^T	113	79	228	578	54·30
X. cucurbitae LMG 690^T	110	19	219	501	52·49
X. fragariae LMG 708^T	114	19	226	512	51·40
X. hortorum pv. *hederae* LMG 733^T	113	19	226	510	51·17
X. hyacinthi LMG 739^T	103	75	166	493	55·57
X. melonis LMG 8670^T	112	75	230	569	53·08
X. pisi LMG 847^T	110	19	227	510	50·39
X. sacchari LMG 471^T	106	15	218	492	53·05
X. theicola LMG 8684^T	105	77	208	541	57·11
X. translucens pv. *translucens* LMG 876^T	103	76	208	540	55·37
X. vasicola pv. *holcicola* LMG 736^T	113	19	227	510	50·98

Gonçalves and Rosato,2002

Celebrating the 20th anniversary of the first *Xanthomonas* genome sequences

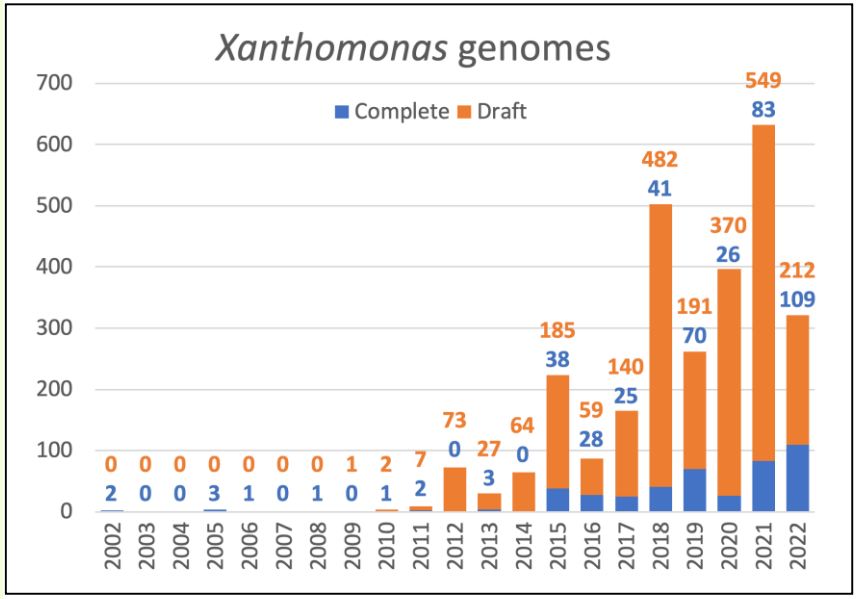
- Celebrating the 20th anniversary of the first Xanthomonas genome sequences– how genomics revolutionized taxonomy, provided:

1. insight into the emergence of pathogenic bacteria,
2. enabled new fundamental discoveries, and
3. helped developing novel control measures – a perspective from the French network on Xanthomonads.

Koebnik and Cesbron *et al.*,2024

Celebrating the 20th anniversary of the first *Xanthomonas* genome sequences

- NCBI Xanthomonas genome statistics (as of 13 July 2023). *Xanthomonas* genome assembly
- metadata were extracted from NCBI GenBank at https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=338.
- GenBank assembly levels 'Contig', 'Scaffold' and 'Chromosome' were considered together as Draft level.
- The complete list of genomes and relevant metadata are available in Supplementary Table S1.



Koebnik and Cesbron *et al.*,2024

Celebrating the 20th anniversary of the first *Xanthomonas* genome sequences

- Phylogenetic tree of the 32 valid species of *Xanthomonas* provided after TYGS analysis (Meier-Kolthoff *et al.*,2022).
- Tree inferred with FastME 2.1.6.1 (Lefort *et al.*, 2015) from GBDP distances calculated from genome sequences retrieved from Genbank.
- The branch lengths are scaled in terms of GBDP distance formula d5.
- The numbers on branches are GBDP pseudo-bootstrap support values > 70% from 100 replications, with an average branch support of 97.2% (Farris, 1972).
- The Newick file was edited in iTOL (https://itol.embl.de/) and rooted on the outgroup *Stenotrophomonas maltophilia*.
- The complete list of genomes and GenBank Assembly accession numbers are available in Supplementary Table S2.

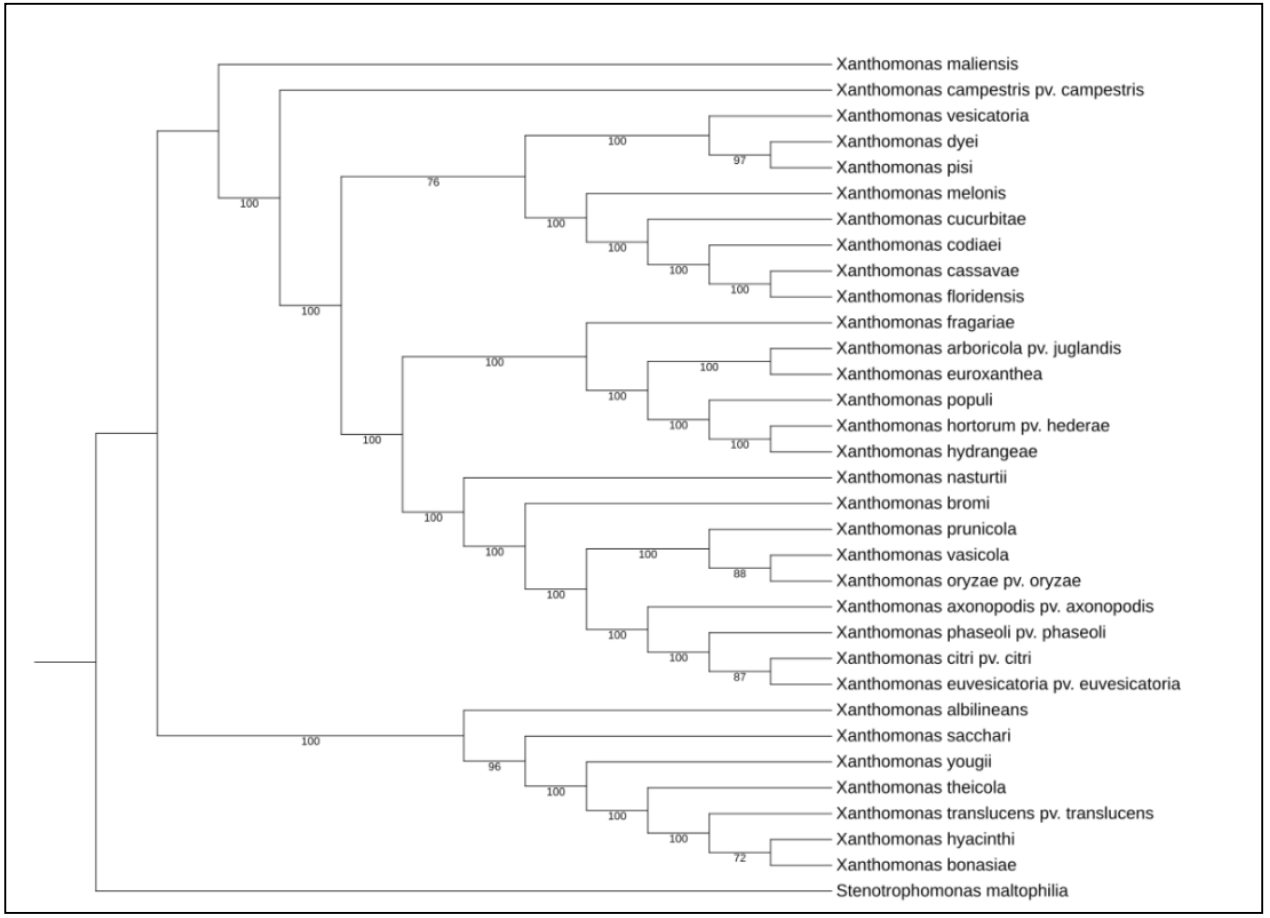
Koebnik and Cesbron *et al.*,2024

Celebrating the 20th anniversary of the first *Xanthomonas* genome sequences

- Phylogenetic tree of the 32 valid species of *Xanthomonas* provided after TYGS analysis (Meier-Kolthoff *et al.*,2022).
- Tree inferred with FastME 2.1.6.1 (Lefort *et al.*, 2015) from GBDP distances calculated from genome sequences retrieved from Genbank.
- The branch lengths are scaled in terms of GBDP distance formula d5.
- The numbers on branches are GBDP pseudo-bootstrap support values > 70% from 100 replications, with an average branch support of 97.2% (Farris, 1972).
- The Newick file was edited in iTOL (https://itol.embl.de/) and rooted on the outgroup *Stenotrophomonas maltophilia*.
- The complete list of genomes and GenBank Assembly accession numbers are available in Supplementary Table S2.

Koebnik and Cesbron *et al.*,2024

Celebrating the 20th anniversary of the first *Xanthomonas* genome sequences



Koebnik and Cesbron *et al.*,2024

Polyphasic taxonomy approach
Protein sequences

- DNA sequences can potentially possess more information than protein sequence.
- Protein sequences, on the other hand, are the functional units (e.g. enzymes, transcription factors, receptors) upon which selection can act.

Polyphasic taxonomy approach
Protein sequences

- Although rRNAs have been extremely useful in unravelling (solving) the phylogenetic relationships of organisms, certain problems cannot be solved.
- A valuable alternative would be the use of protein-coding sequences for the construction of phylogenetic trees and species delineation.
- Fortunately during the last few years, investigators have analyzed an increasing number of protein sequences from Archaea, Bacteria and Eukarya.
- e.g. heatshock (HSP70) proteins, chaperonins, etc.

Fred Opperdoes,1997;..

Polyphasic approach

Ribosomal RNA or protein sequences for phylogenetic inference?

Ribosomal RNA trees

- Comparative analyses of rRNA sequences, initiated in the 1970s, suggest that the living world is divided into three domains: Eucarya, Archaea, and Bacteria.

Protein trees

- Recently protein-coding genes were used for phylogenetic analyses.
- The entire sequence of the 16S rRNA (1,500 bp) or *rpoB* gene (4,000 to 4,500 bp) containing <1% undetermined positions would be required for the description of a new taxon.

Polyphasic taxonomy approach
Based on *gyrB* ,*rpoD*, *Irp*, *recA* genes

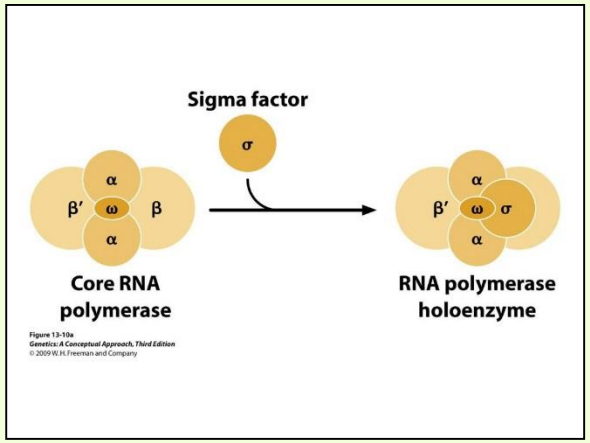
- Based on new molecular markers such as *gyrB*, *rpoD* (RNA polymerase sigma-70 factor), *Irp*, *recA* sequences and application of PCR-RFLP technique, bacterial classification, identification and detection systems were obtained for:
- *Pseudomonas*,
- Actinomyces,
- *Xylella fastidiosa*,
- etc.
- These markers can be very useful in microbial ecology and other fields of bacteriology and improve the reliability of the phylogenetic trees.

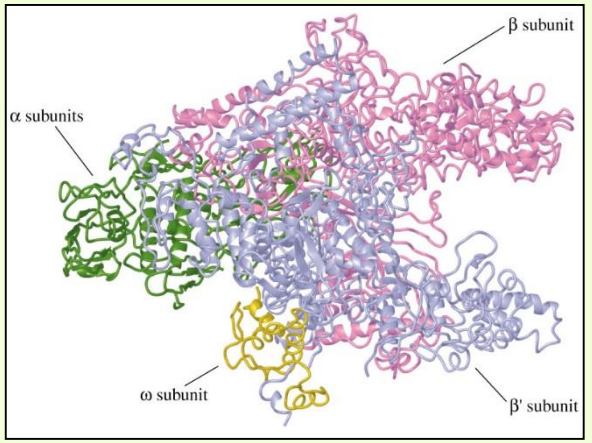
Polyphasic approach

Based on *RpoB*, *gyrB* sequencing and ribotyping

1. *RpoB*, *RpoD* and *RpoS*
2. *GyrA* and *B* sequencing, and
3. Ribotyping.

- Are very useful for species differentiation within *Pseudomonas* and other genera.





Universal and specific primers used to amplify gyrB, nifD, fusA, rpoB and recA gene fragments

Locations refer to the *Escherichia coli* numbering system

Primer pair*	Primer sequences (5′–3′)	Location
***fusA* primers**		
EF-GF/EF-GR[a]	ACIATGGATTGGATGGADCA/TTCIACTTTCATCATIGGTTC	360–1420
EF-G30F/EF-G510R[b]	GTGCAGAMTCTICTTGAT/TAATGCCCAGCTTTTCCTGGTC	30–510
***gyrB* primers**		
GYR48F/GYR1010R[b]	GACGATGATCCGGTGGTAGC/CGATGATACCATCTTCGAGAC	48–1010
GYR130F/GYR1130R[b]	ATCTCCTGGGAGGAGAGCATC/CTGGAGCTGGAGATTCGCCG	130–1130
GYRBF/GYRBR[a]	CATCATATGGTNTGGGAGATWGTNGATAAT/CATTTTGTCGAAICKIGCTTTTTCIACGTT	100–1180
GYR838F/GYR1541R[b]	GGGAAGGTGTACCGCCAGT/GAGGGTGGAACCCACTTGG	838–1541
***nifD* primers**		
NIFDF/NIFDR[a]	GATTTTCADGADAADGATATT/CCAIGGIATICCDTATTTTC	405–963
NIFDF2/NIFDR2[b]	CAGGAAAAIGATATTGTCTACGG/TAGTCGTCCTIGTGGGCGAACTCGTA	180–987
***recA* primers**		
RECAF/RECAR[a]	TGTGCITTTATWGATGCIGAGCATGC/CCCATGTCICCTTCKATTTCIGCTTT	21–618
REC258F/REC858R[c]	AAGACGGACGACCTGCTCG/TGAGGAAAAGGCGGGAATTT	258–858
RECA11F/RECA479R[b]	CATCTGGCAAGACTACCCTG/GGATGTCCATGCGCACGAAG	11–479
RECA11F/RECA560R[b]	CATCTGGCAAGACTACCCTGG/TGCGACTTTGTTATTAACGACCTT	11–560
RECA50F/RECA563R[b]	AGCCTTGGAAATTGCCGAG/TCCGCCGTAGGAGAACCAG	50–563
REC300F/REC630R[b]	TGTTTGTTGATGGCCGAGCA/GGTTGCCGAACATGACACCGAT	300–630
REC300F/REC910R[b]	TGTTTGTTGATGGCCGAGCA/CCCGCCGTAGGAGAAACCAGGC	300–910
***rpoB* primers**		
RPOBF/RPOBR[a]	CAGTTHATGGATCAG/GTTHTGNCDTTGCATGTT	117–642
RPOBF2/RPOBR2[b]	ACTCGTGAGCGTGCTGGTTTT/ATCAAAGCCCGGTGAGCAT	9–636
RPO216F/RPO666R[b]	CAGGAACGGGATGAGCGAT/CTCCGAGGTTACCCACAAGC	216–666
RPO175F/RPO800R[b]	ATGTTCGACCCCATGAGTGC/GAGCGGATGAGTCTCCAGGA	175–800
RPO310F/RPO940R[b]	GAGCGSATGAGTCTSCAGGA/CITTGCATGTTSGAGCCCAT	310–940

*Amplification parameters: *a*, 92–94 °C for 5 min, 20 cycles of 92–94 °C for 1 min, 41–55 °C for 1·5–3 min (−0·5 °C per cycle) and 70–71 °C for 1·5–3 min and then 20 cycles of 92–94 °C for 1 min, 38–57 °C for 1–1·5 min and 71–72 °C for 1·5 min, followed by 71 °C for 10 min; *b*, 94 °C for 5 min, 35 cycles of 94 °C for 45–60 s, 47–60 °C for 1–1·5 min and 72 °C for 1–1·5 min; followed by 72 °C for 10 min; *c*, 94 °C for 5 min, 10 cycles of 92 °C for 45 s, 55 °C for 1·5 min (−0·2 °C per cycle) and 72 °C for 1·5 min, 20 cycles of 94 °C for 45 s, 58 °C for 1 min and 72 °C for 1 min, followed by 72 °C for 10 min.

Holmes *et al.*,2004

Polyphasic approach

Based on *gyrB* and *rpoD* genes

Pseudomonas

- Analysis based on the combination of *gyrB* and *rpoD* sequences yielded better resolution than 16S rRNA and resulted into four major clusters viz.:
- Cluster I (*P. fluorescenes*)
- Cluster 11 (*P. putida*, *P. plecoglossicida*),
- Cluster I11 (*P. pseudoalcaligenes*), and
- Cluster IV (*P. aeruginosa*).
- This may be due to the high evolutionary rate of protein coding sequences, *gyrB* and *rpoD* (Yainamoto and Harayaina,1988; Yamamoto *et al.*,2000).

Viz: It is used as a synonym for "namely", "that is to say", "to wit", "which is", or "as follows".

Polyphasic taxonomy approach

1. Based on *rpo* genes

- RNA polymerases "transcribe" the genetic information on DNA into RNA strands.
- For a long time, *E. coli* was thought to contain just a single RNA polymerase.
- RNA polymerase is a complex enzyme consisting of at least six non-identical subunits such as α, β, β' and σ.
- The genetic loci coding for the subunits, namely *rpo*A (*α*), *rpo*B (ß), *rpo*C (β') and *rpo*D (σ).

Housekeeping sigma factors

E. coli sigma factors

- σ70 (RpoD)- The "housekeeping" sigma factor, transcribes most genes in growing cells.
- σ38 (RpoS)- The starvation/stationary phase sigma factor
- σ28 (RpoF)- The flagellar sigma factor
- σ32 (RpoH)- The heat shock sigma factor
- σ24 (RpoE)- The extracytoplasmic stress sigma factor
- σ54 (RpoN)- The nitrogen-limitation sigma factor
- σ19 (FecI)- The ferric citrate sigma factor
- All sigma factors are distinguished by their characteristic molecular weights.
- For example, σ^{70} refers to the sigma factor of RNA polymerase in *E. coli* with a molecular weight of 70 kDa.

Sigma factor (σ factor) is a prokaryotic initiation enzyme factor that must be part of RNA polymerase for attachment to promoter sites on DNA. RpoB is beta subunit of RNA polymerase and function in transcription and nonspecific DNA binding, sigma binding.

Sigma factors
Functions

DNA --replication--> DNA
DNA --transciption--> RNA ----------> Protein
RNA --Reverse transciption--> DNA

- Sigma is a specificity factor.
- It directs RNA polymerase to the promoter and ensures that transcription is initiated only where it is supposed to be initiated.
- The sigma subunit does two things:

1. It reduces the affinity of the enzyme for non-specific DNA.
2. It greatly increases the affinity of the enzyme for promoters.

Transcription and gene expression

Transcription in bacteria

A sigma factor (σ factor) is a prokaryotic initiation enzyme factor that must be part of RNA polymerase for attachment to promoter sites on DNA.

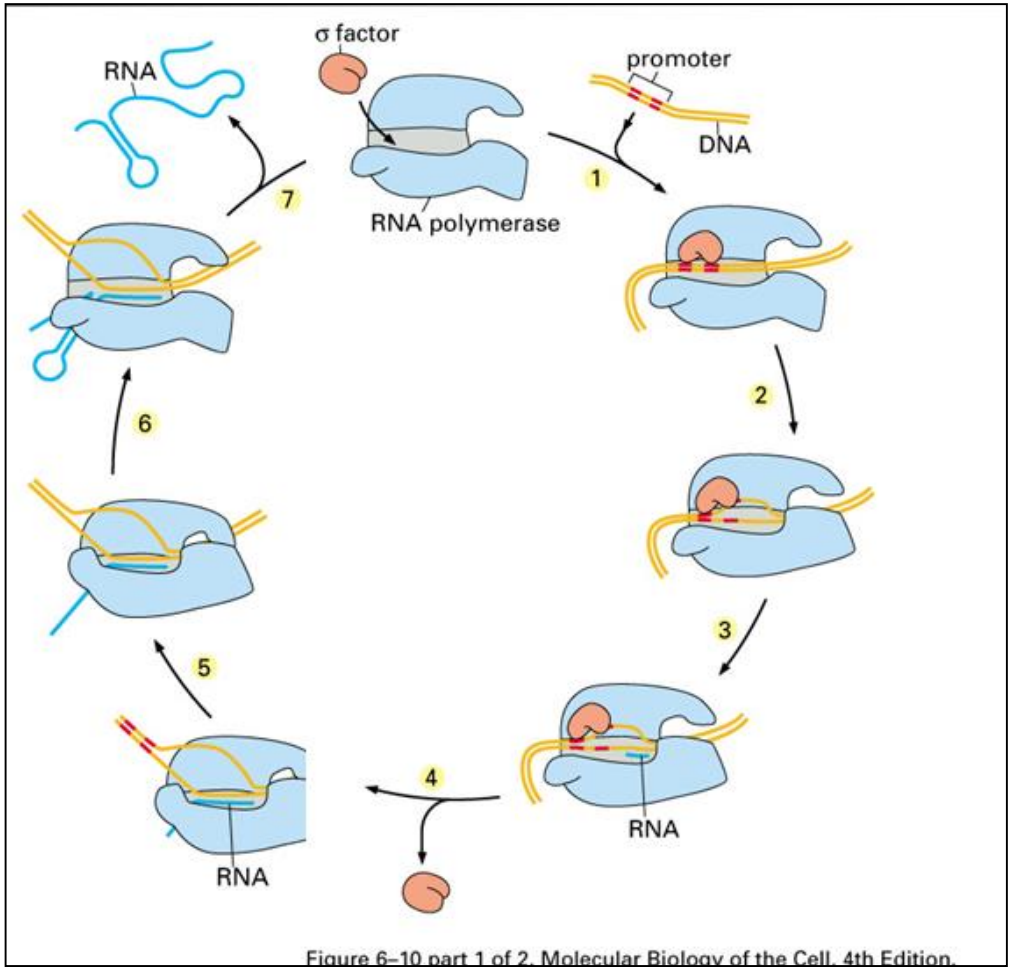


Figure 6–10 part 1 of 2, Molecular Biology of the Cell, 4th Edition.

rpoS gene

RpoS

The starvation/stationary phase sigma factor

- Bacterial adaptation to changing conditions and to the host environment requires coordinated changes in gene expression that permit more efficient use of metabolites and increased survival.
- An important form of gene control is through the use of alternative sigma factors that direct RNA polymerase to recognize a distinct group of genes.
- One such sigma factor is RpoS, which is widely present in many Gram-negative bacteria.
- RpoS is important for adaptation under nutrient-limited conditions. e.g. iron limit(siderophore).

rpoS gene
Function

- The *rpoS* gene is essential for the:

1. Expression of a variety of stationary-phase-induced genes, as well as for the
2. Expression of stationary-phase-specific stress resistance phenotypes, such as increased resistance to high temperature, high concentrations of H_2O_2, and high osmotic challenge.

rpoS gene
Function and similarity with *ropD*

- The *rpoS* has been known as *nur*, *katF*, *appR*, *csi-2*, and *abrD*, a result of independent studies of different phenotypes.
- This gene encodes the sigma factor σ^s.
- σ^s is a protein of 41.5 kDa that controls a regulon of at least 30 genes which are expressed during starvation and at the transition into stationary phase.
- The *rpoS* sequence exhibits extensive homology to *rpoD*, which encodes σ^{70}, the sigma subunit of RNA polymerase in *E. coli*.

Krogfelt, *et al.*,2000

Polyphasic approach
Based on *rpoS* gene

The design of primers for the *rpoS* gene

The sequence of *Erwinia carotovora rpoS* gene was described

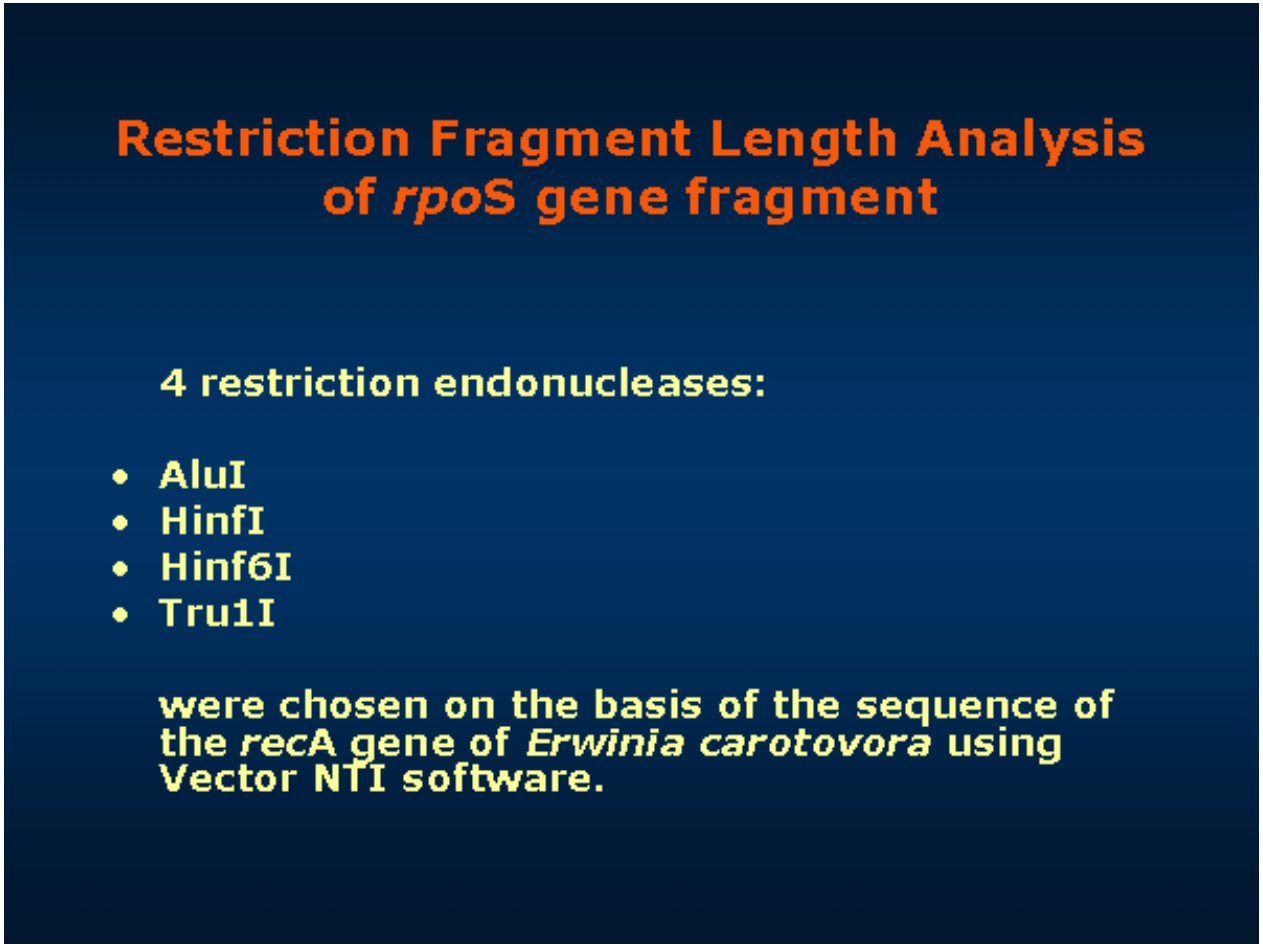
The sequence of designed primers (*Waleron et al.*, in press):

- rpoS1 5′- ATGAGCCAAAGTACGCTGAA -3′
- rpoS2 5′- ACCTGAATCTGACGAACACG -3′

was checked in the complete GeneBank and EMBL databases using the BLAST-n program

Lojkowska *et al.*,2004

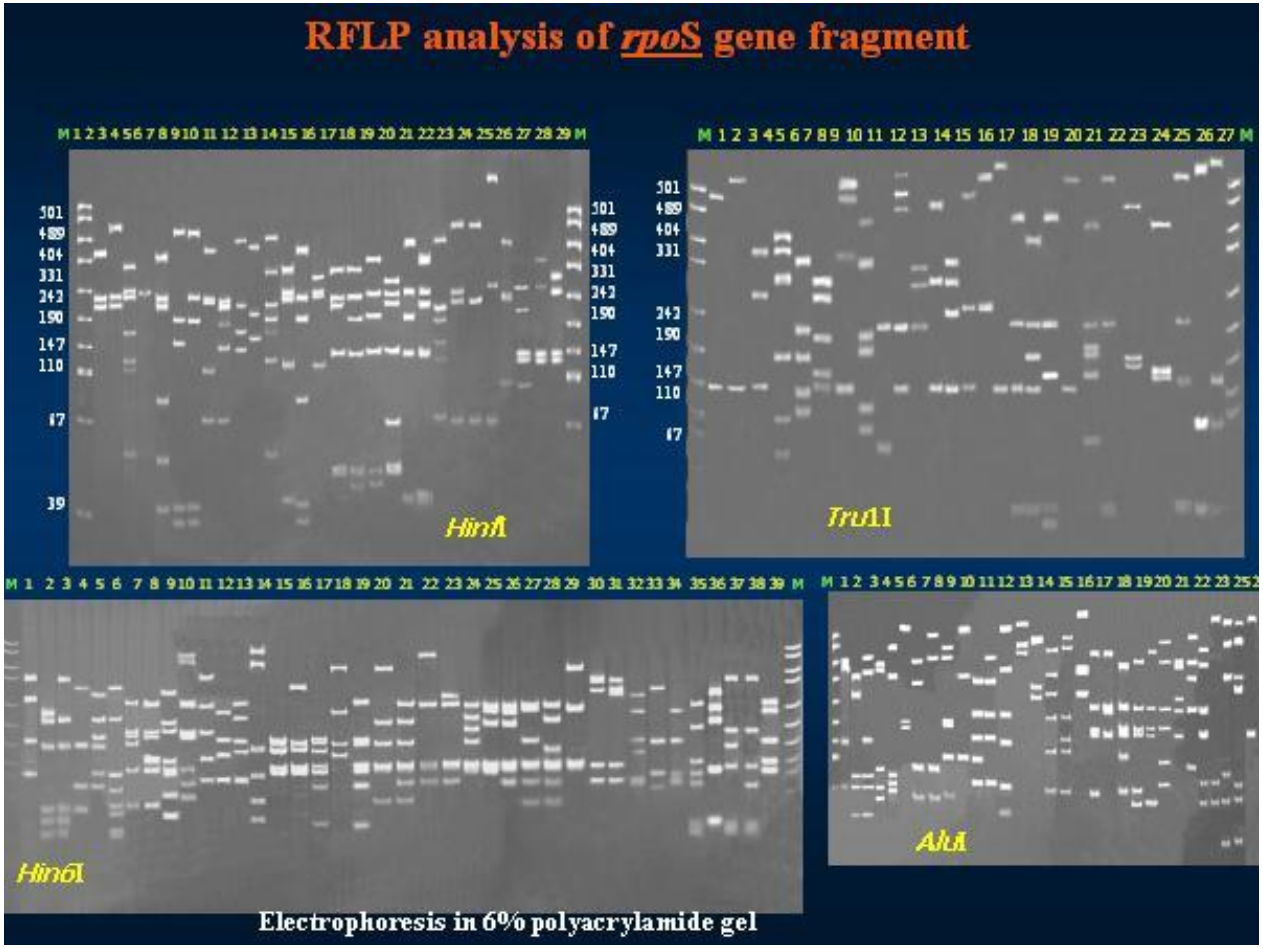
Polyphasic approach
Based on *rpoS* gene



Lojkowska *et al.*,2004

Polyphasic approach

Based on *rpoS* gene



Lojkowska *et al.*,2004

Polyphasic approach
Based on *rpoS* gene

Number of profiles described for the different species from the genus *Erwinia based on* RFLP analysis of *rpoS* gene fragment

Species	Number of strains	Amount of profile	Numbers of profils
Eca	24	3	21-23
Ecc	69	20	29-48
Ecb	12	1	24
Eco	12	2	25-26
Ecw	5	1	27
E. amylovora	39	1	1
E. ananas	5	2	15, 16, 17
E. cacticida	10	2	6, 28
E. camegiena	1	1	6
E. chrysanthemi	23	15	49-63
E. cypripedi	5	1	9
E. herbicola	6	2	18, 19
E. mallotivora	1	1	5
E. milletiae	1	1	18
E. nigrifluens	3	1	13
E. persicinum	3	2	7, 8
E. psidii	1	1	4
E. pyrifoliae	6	2	2, 3
E. quercina	1	1	14
E. rhapontici	6	1	10
E. rubifaciens	3	1	12
E. salicis	2	1	11
E. stewartii	2	1	20

Lojkowska *et al.*,2004

Polyphasic approach

Based on *RpoD* gene

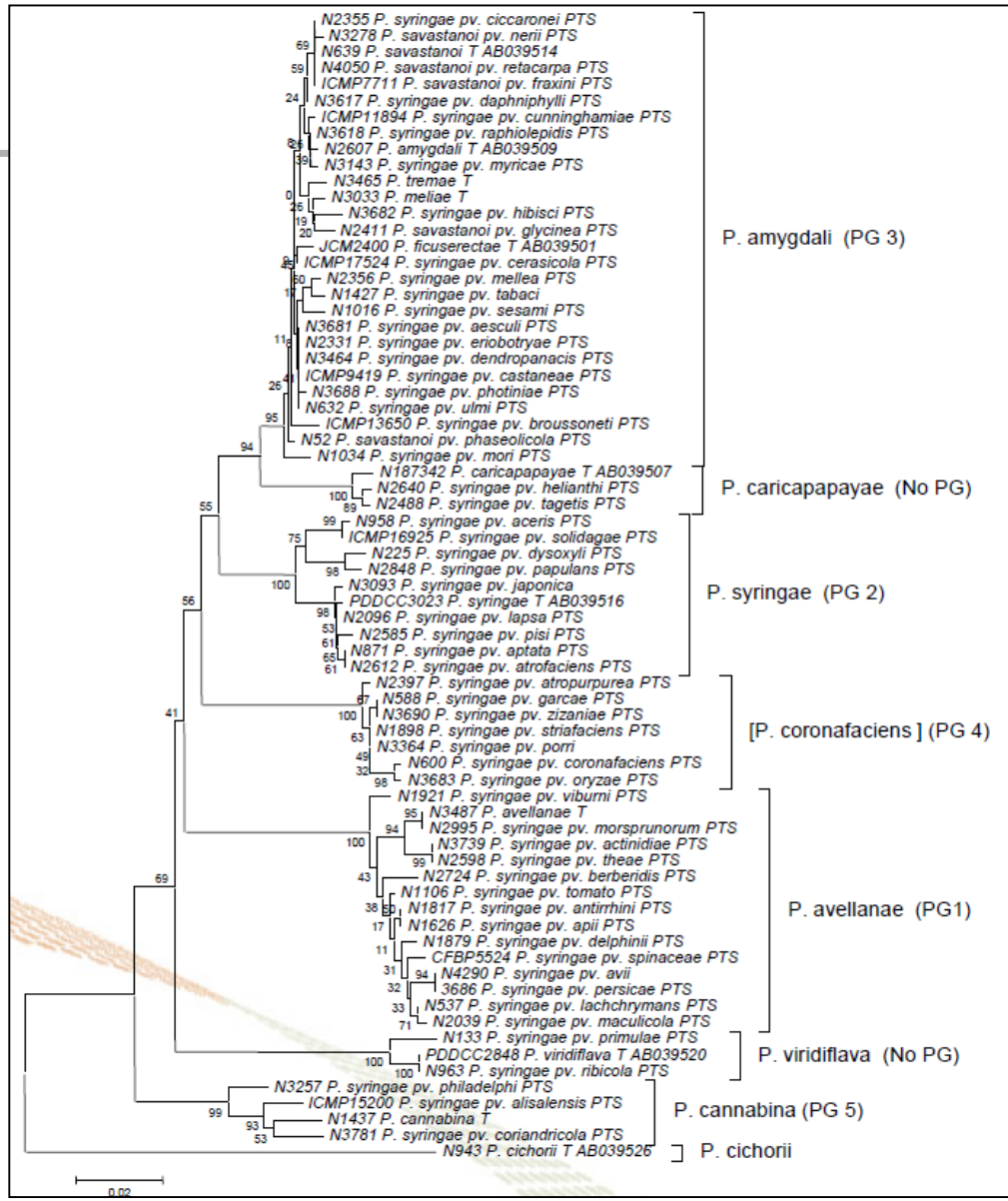
RpoD

Pseudomonas

species

identification.

σ70 (RpoD)- The "housekeeping" sigma factor, transcribes most genes in growing cells.

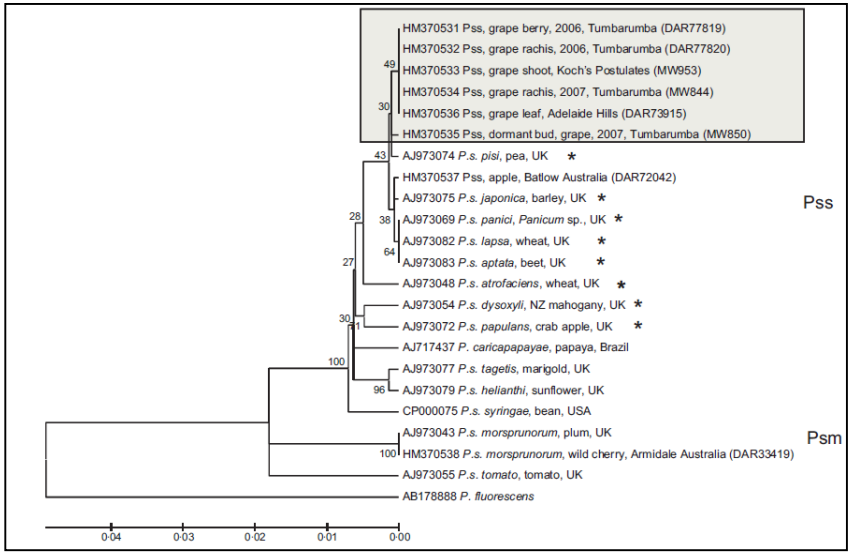


Neil Parkinson

PCR detection of *Pss*

Amplification of both genes of *16S rRNA* and *rpoB* (β-subunit of RNA polymerase)

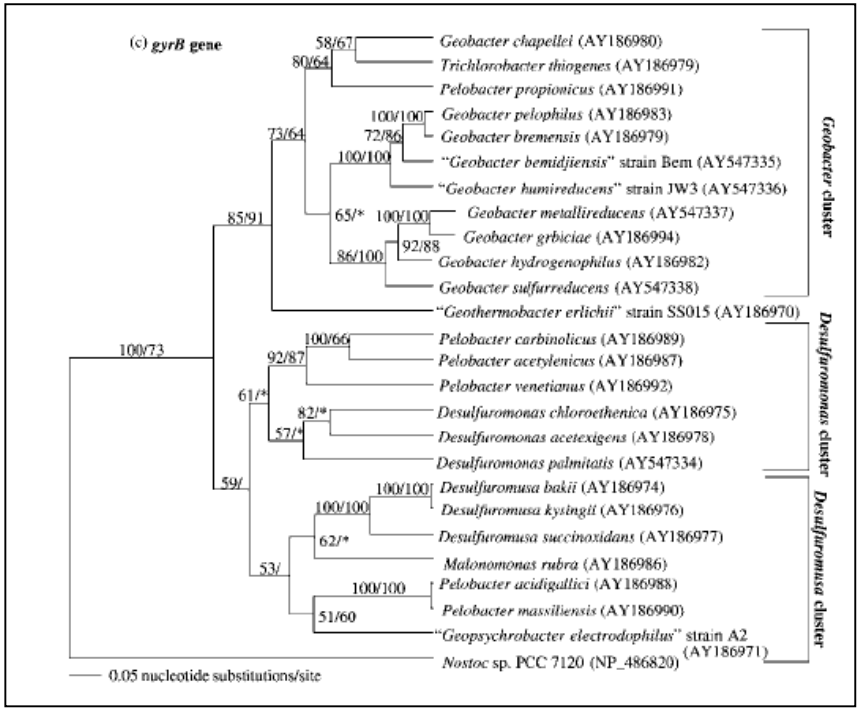
- UPGMA tree derived from *rpoB* gene sequences from representative collected Australian grape strains (boxed); Australian apple and Prunus strains; and published sequences from GenBank.
- *Pseudomonas syringae* pv. *syringae* (Pss) and *P.s.* pv. *morsprunorum* (Psm) groups are clearly separated.
- Sequence for *Pseudomonas fluorescens* (AB178888) was included as an outgroup.

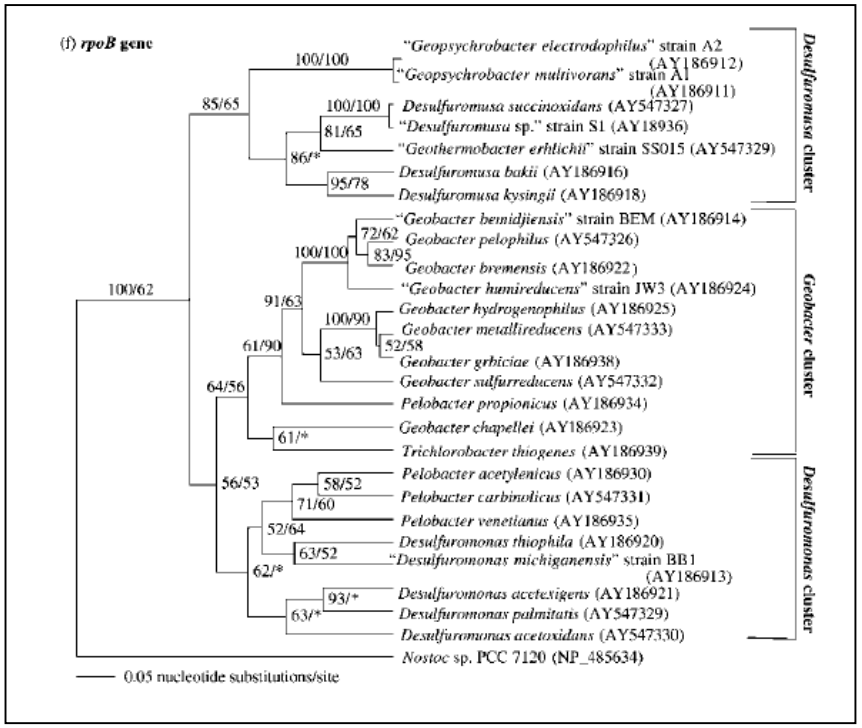


***rpoB* gene, a reliable biomarker to distinguish some of *P. syringae* pathovars.**

Whitelaw-Weckert *et al.*,2011

Phylogenetic analysis of sequences
Comparison of *gyrB* and *rpoB* genes





Holmes *et al.*,2004

Polyphasic approach

2. Based on *recA* gene

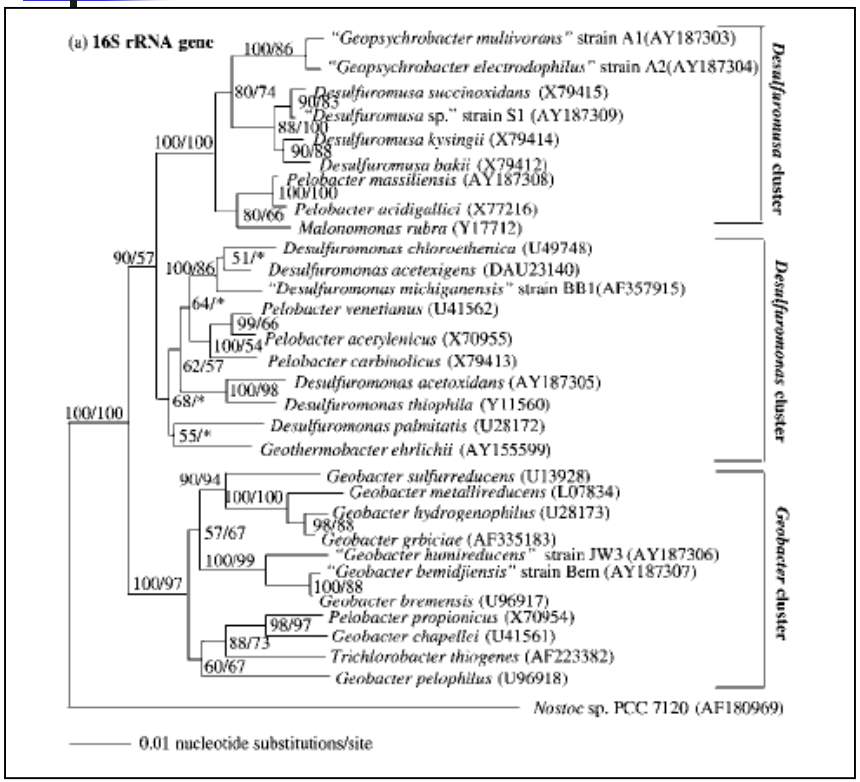
Recombination protein

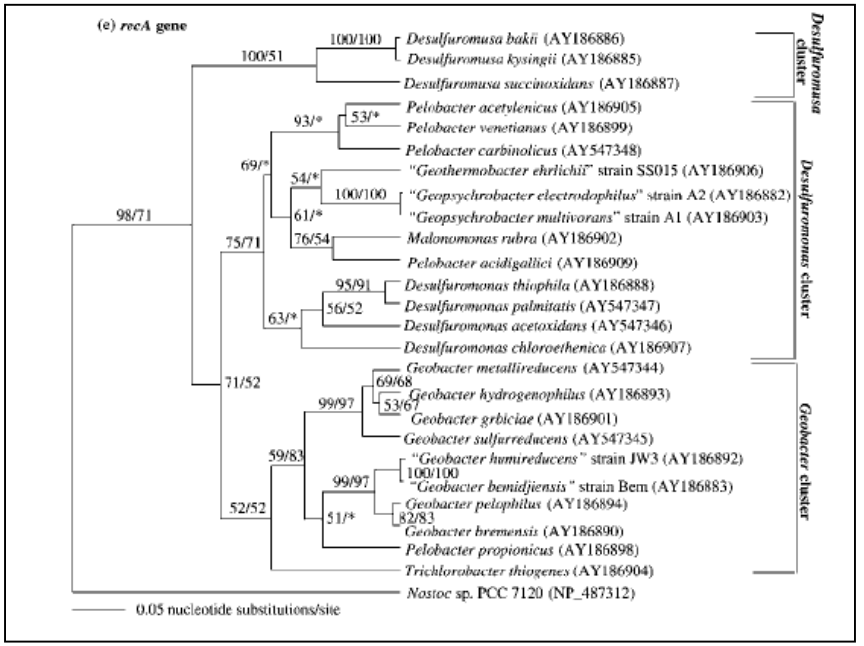
- RecA is a ubiquitous protein, present in nearly all bacteria.
- RecA in *E. coli* is a 38 KDa protein.
- RecA has a structural and functional homolog in every species.
- RecA has multiple activities, all related to DNA repair.
- The bacterial RecA protein is a recombinase, functioning in recombinational DNA repair in bacteria.
- During DNA pairing and strand exchange in genetic recombination RecA protein binds strongly to ssDNA to form a nucleoprotein filament.

Polyphasic approach
Based on *recA* gene

- This reaction initiates the exchange of strands between two recombining DNA double helices and promotes recombinational DNA repair and maintenance.
- RecA is also regulated by the action of other proteins such as RecF, RecO, RecR, DinI, RecX, RdgC, PsiB, and UvrD proteins.
- All of these proteins function in a network that determines where and how RecA functions.

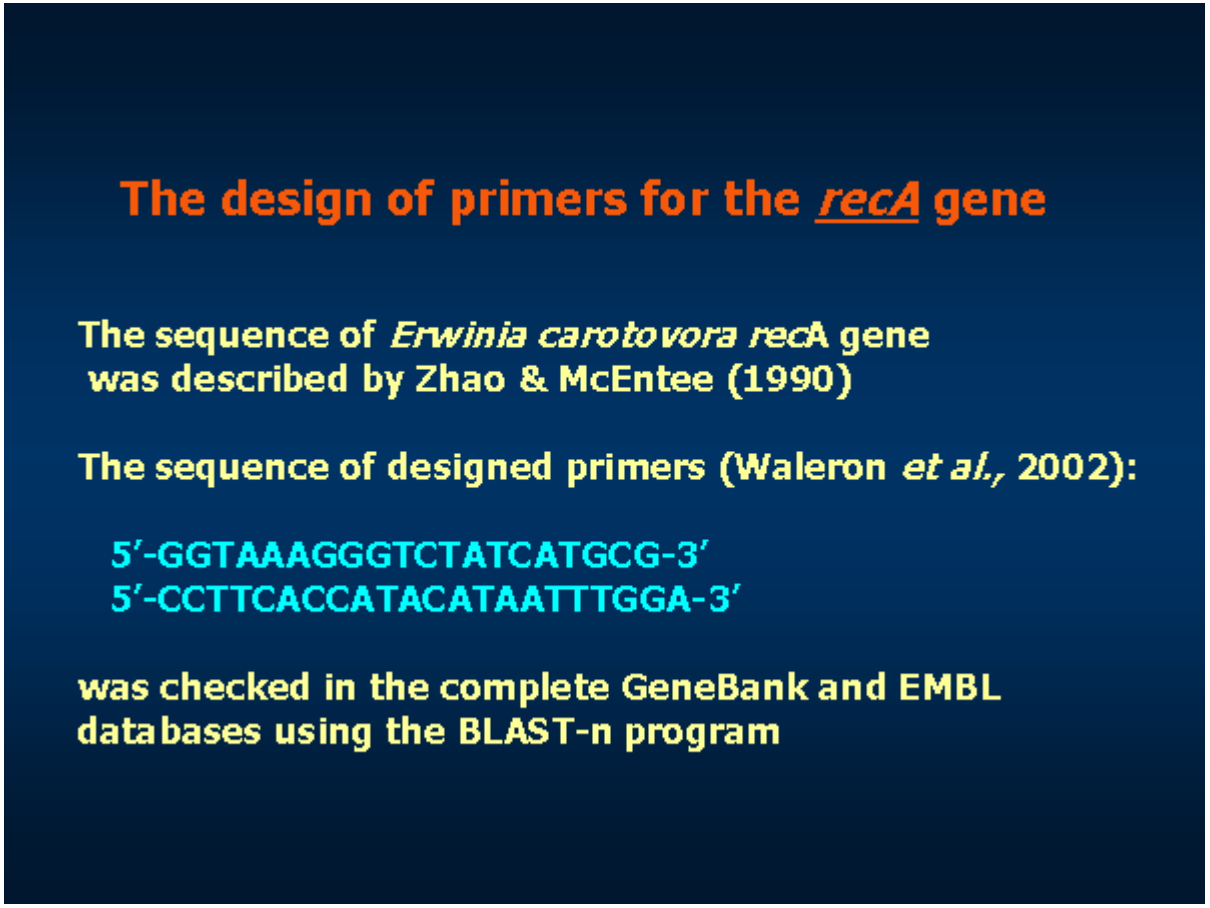
Phylogenetic analysis of sequences
Comparison of 16S rRNA and *recA* genes





Holmes *et al.*,2004

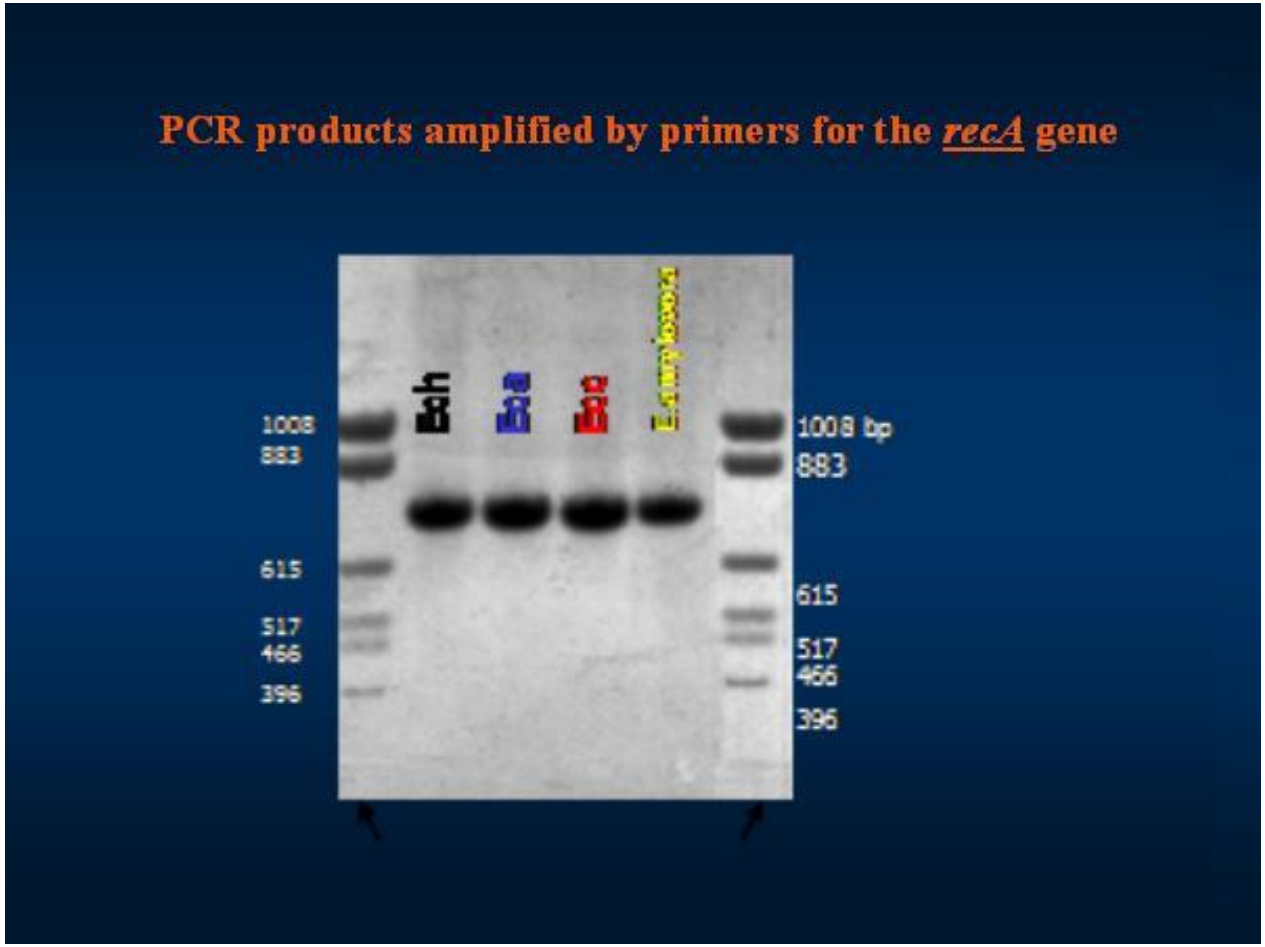
Polyphasic approach
Based on *recA* gene



Lojkowska *et al.*,2004. in EPPO Conference on Quality of Diagnosis and New Diagnostic Methods for Plant Pests.

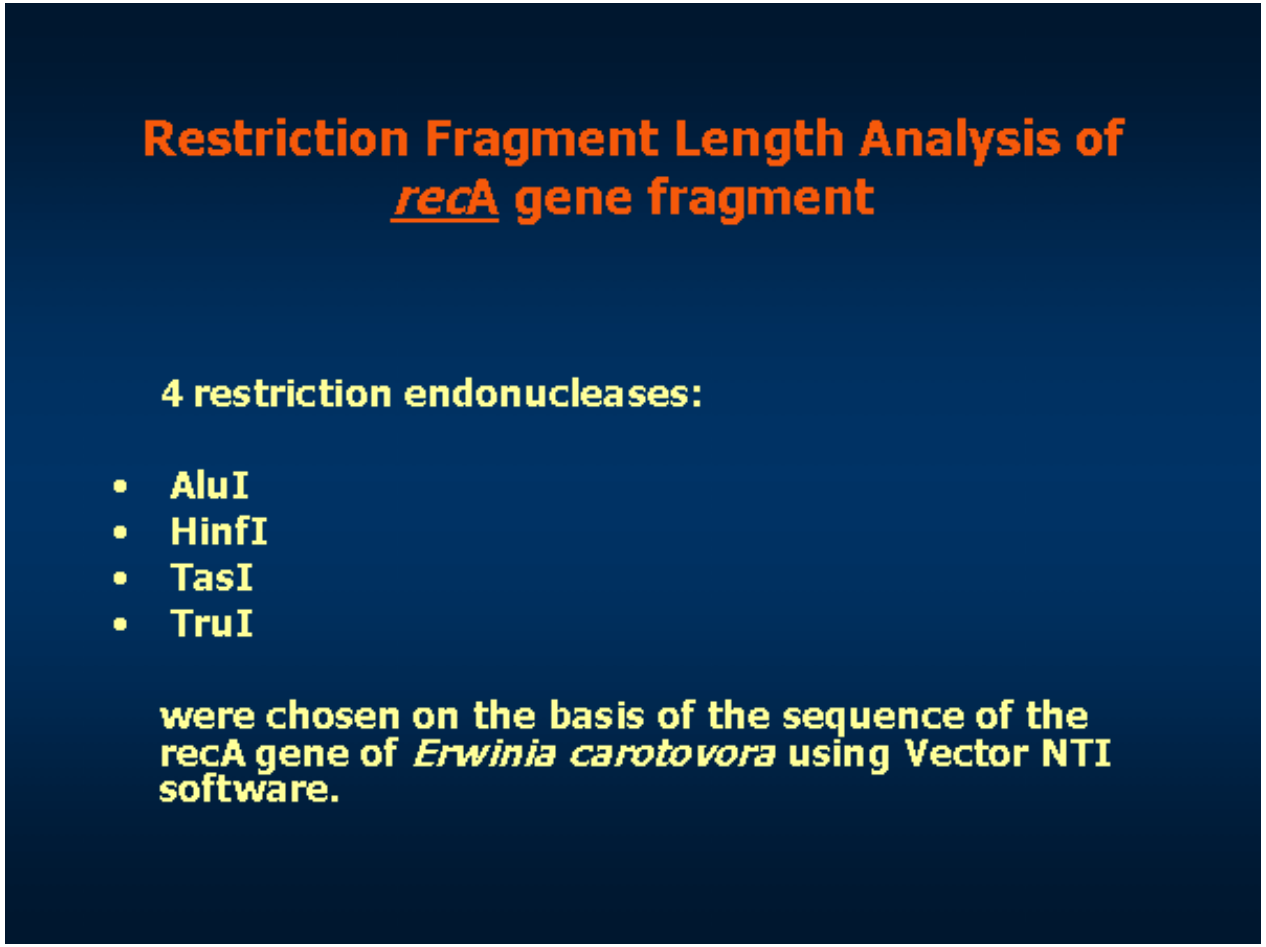
Polyphasic approach

Based on *recA* gene



Lojkowska *et al.*,2004

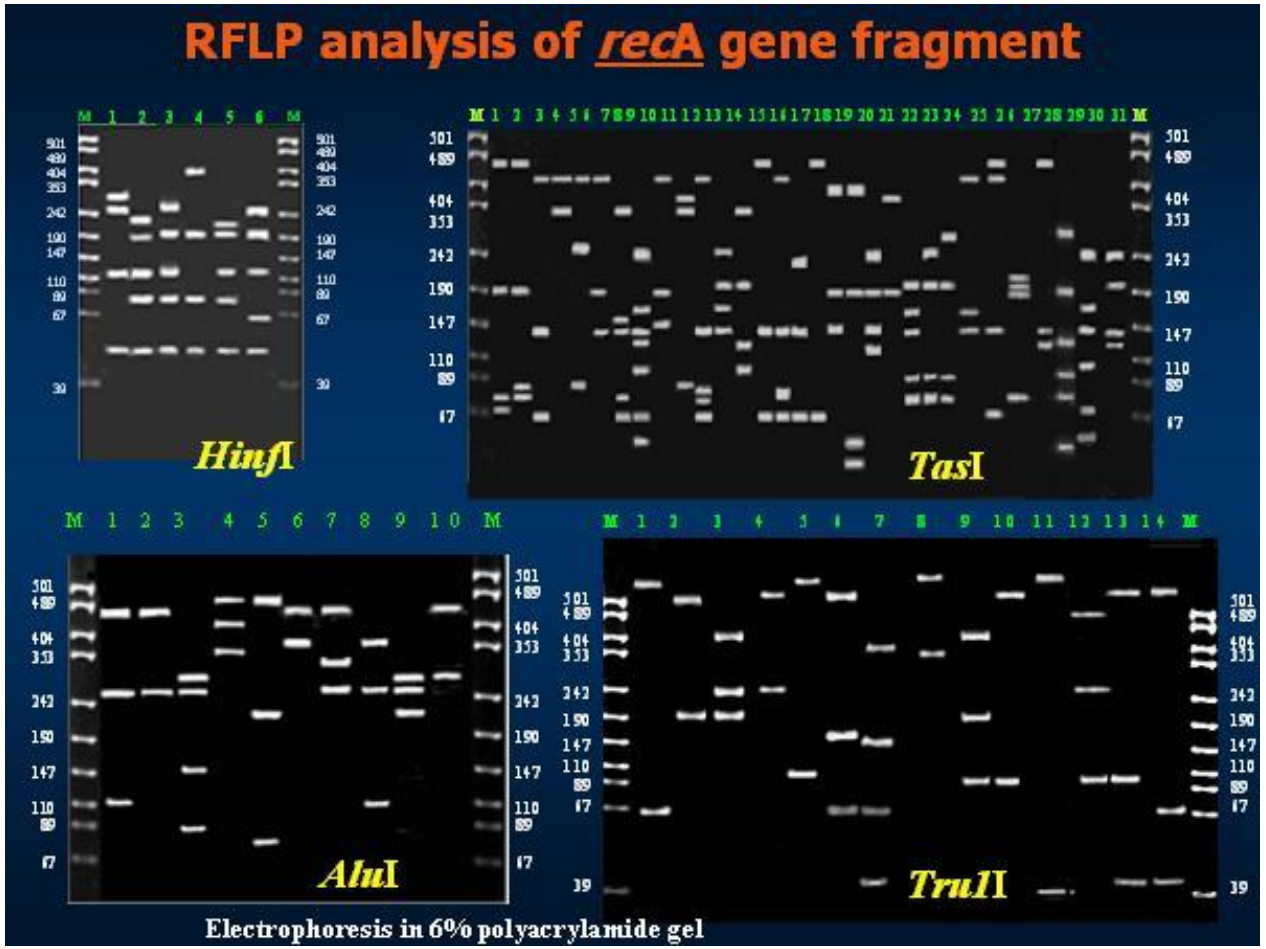
Polyphasic approach
Based on *recA* gene



Lojkowska *et al.*,2004

Polyphasic approach

Based on *recA* gene



Lojkowska *et al.*,2004

Polyphasic approach
Based on *recA* gene

Number of profiles described for the species from the genus *Erwinia based on* RFLP analysis of *recA* gene fragment

Species	Number of strains	Amount of profils	Number of profis
Eca	106	2	1, 2
Ecc	158	18	3-20
Ecb	12	1	21
Eco	12	1	22
Ecw	5	1	23
E. amylovora	39	1	32
E. ananas	5	2	25, 26
E. cacticida	16	3	29, 30, 31
E. carnegiena	1	1	31
E. chrysanthemi	93	15	33-46
E. cypripedi	5	1	24
E. herbicola	6	2	27, 28
E. mallotivora	1	1	55
E. milletiae	1	1	28
E. nigrifluens	3	1	51
E. persicinum	3	1	49
E. psidii	1	1	54
E. pyrifoliae	6	1	58
E. quercina	1	1	56
E. rhapontici	6	1	53
E. rubifaciens	3	1	50
E. salicis	2	1	52
E. stewartii	2	1	48
E. tracheiphila	1	1	57

Lojkowska *et al.*,2004

Polyphasic approach

3. Based on DNA gyrase

Unwinds the coiled DNA to relieve tension

- DNA topoisomerase converts supercoiled DNA to a relaxed form.
- DNA topoisomerases are divided into two classes:

1. Type I enzymes (topoisomerases I, III and V) break single-strand DNA.
2. Type II enzymes (topoisomerases II, IV and VI) break double-strand DNA.

- DNA gyrase, a type II bacterial topoisomerase composed of:

1. Two A subunits (GyrA)encoded by the *gyrA*, and
2. Two B subunits (GyrB), encoded by the *gyrB* genes.

Polyphasic approach

Species discrimination using *gyrB*

gyrB an essential role in DNA replication

- Most species were clearly resolved.
- Multiple strains clustered within the species grouping.
- Only *X. albilineans* has not amplified -yet
- Some new species were not resolved. These probably represent misclassifications.

Parkinson and co-workers

Polyphasic approach

Based on *gyrA* gene

Negatively supercoils closed circular double-stranded DNA

- Thus, *gyrA* gene codes for two of the four subunits of the bacterial DNA gyrase enzyme i.e. GyrA.
- The gyrase enzyme introduces negative supercoils into DNA in an ATP-dependent reaction.
- *gyrA* is characterized by areas of high conservation and areas of variability, making it suitable for bacterial identification.
- The *gyrA* gene is used as a specific chromosomal marker in identification of many bacteria.

Polyphasic approach
Based on *gyrA* gene

The design of primers for the *gyrA* gene

The sequence of *Erwinia carotovora gyr*A gene was described by Rosanas *et al.* (1995)

The sequence of primers designed (Waleron *et al.*, in press):

5′-TGGTGACGCGTCGTACCATT-3′
5′-ATACCCACAGCATGACCGCC-3′

was checked in the complete GeneBank and EMBL databases using the BLAST-n program

Lojkowska *et al.*,2004

Polyphasic approach
Based on *gyrA* gene

The design of primers for the *gyrA* gene

The sequence of *Erwinia carotovora gyr*A gene was described by Rosanas *et al.* (1995)

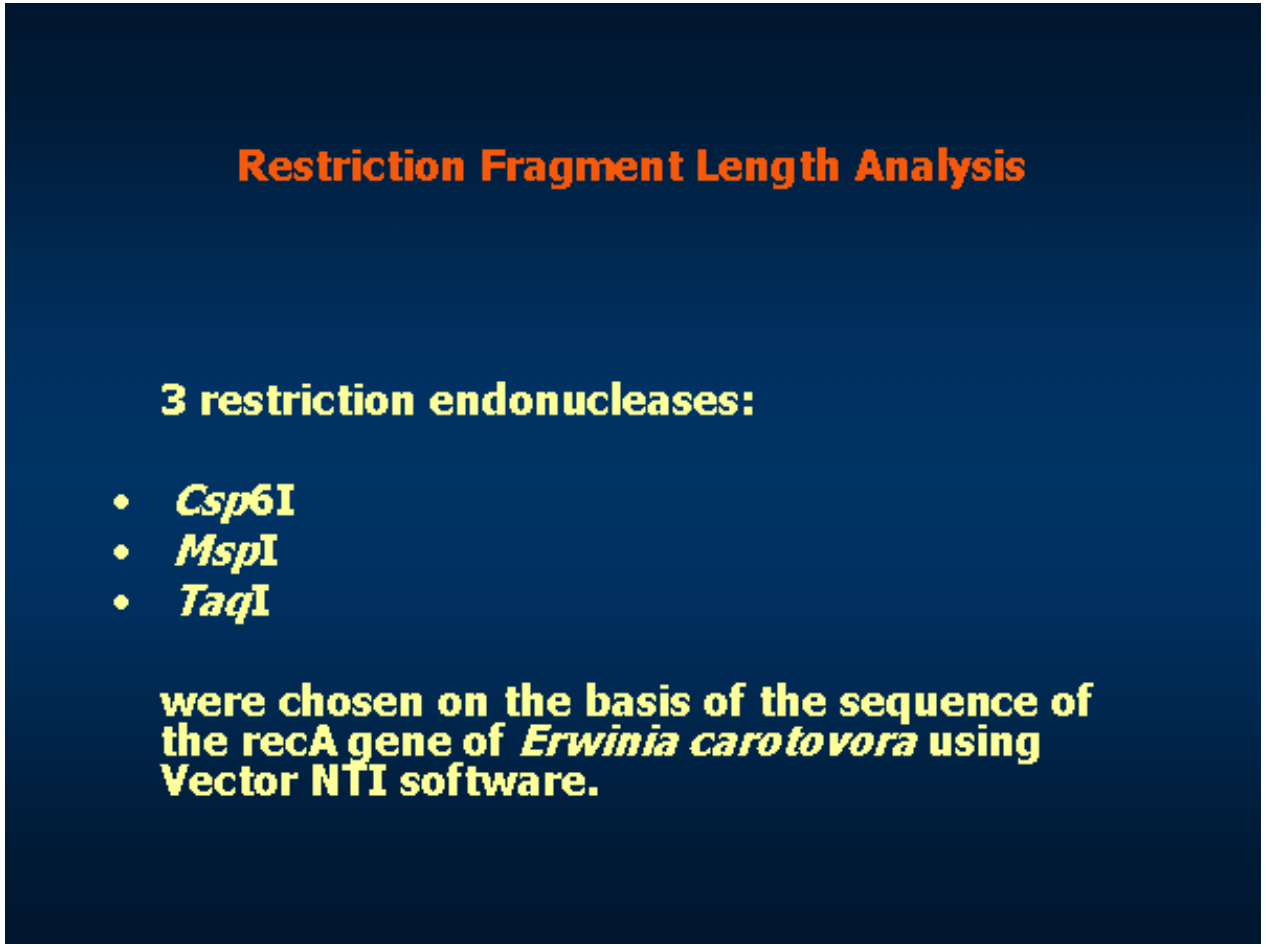
The sequence of primers designed (Waleron *et al.*, in press):

5′-TGGTGACGCGTCGTACCATT-3′
5′-ATACCCACAGCATGACCGCC-3′

was checked in the complete GeneBank and EMBL databases using the BLAST-n program

Lojkowska *et al.*,2004

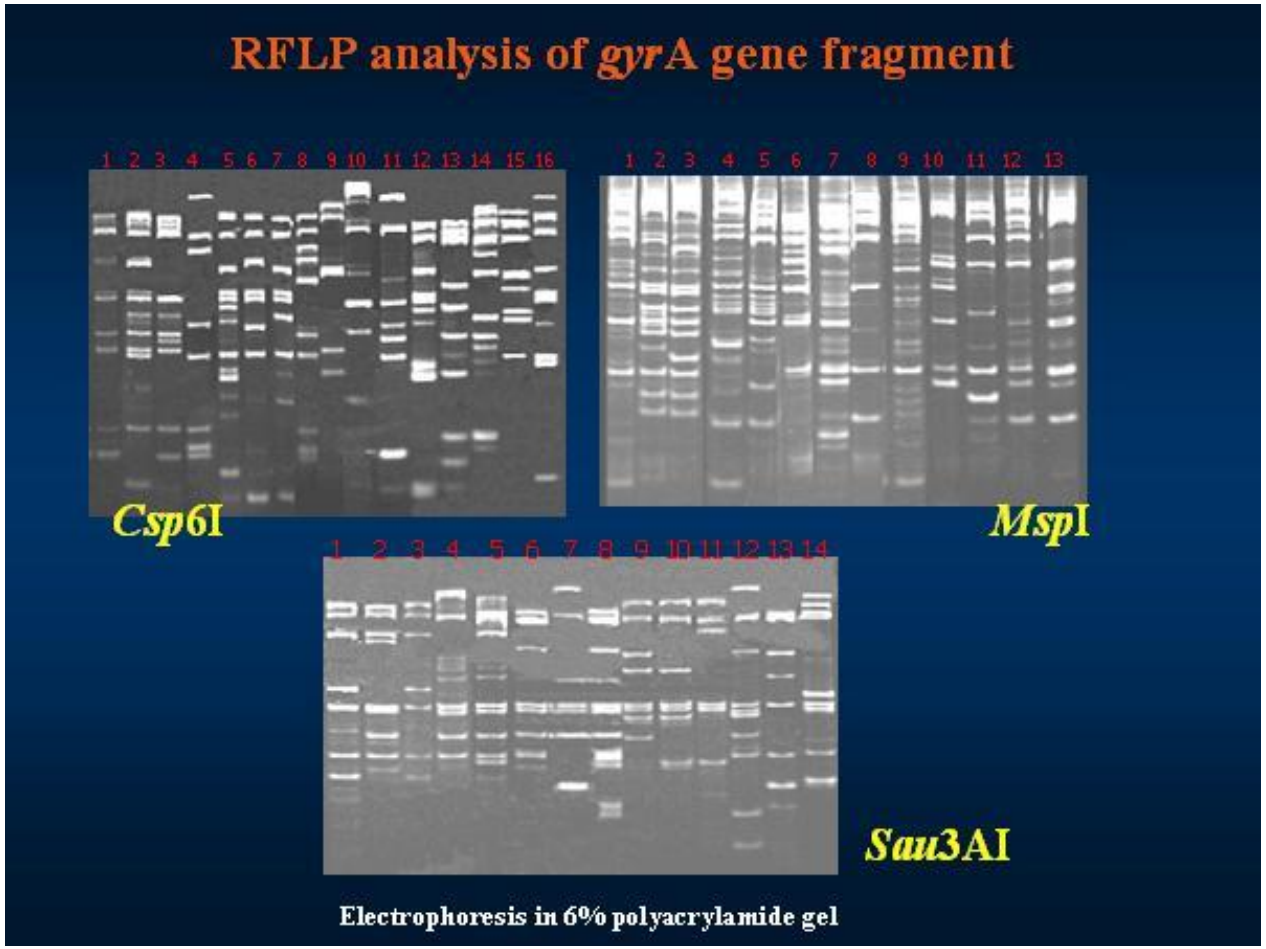
Polyphasic approach
Based on *gyrA* gene



Lojkowska *et al.*,2004

Polyphasic approach

Based on *gyrA* gene



Lojkowska *et al.*,2004

Polyphasic approach
Based on *gyrA* gene

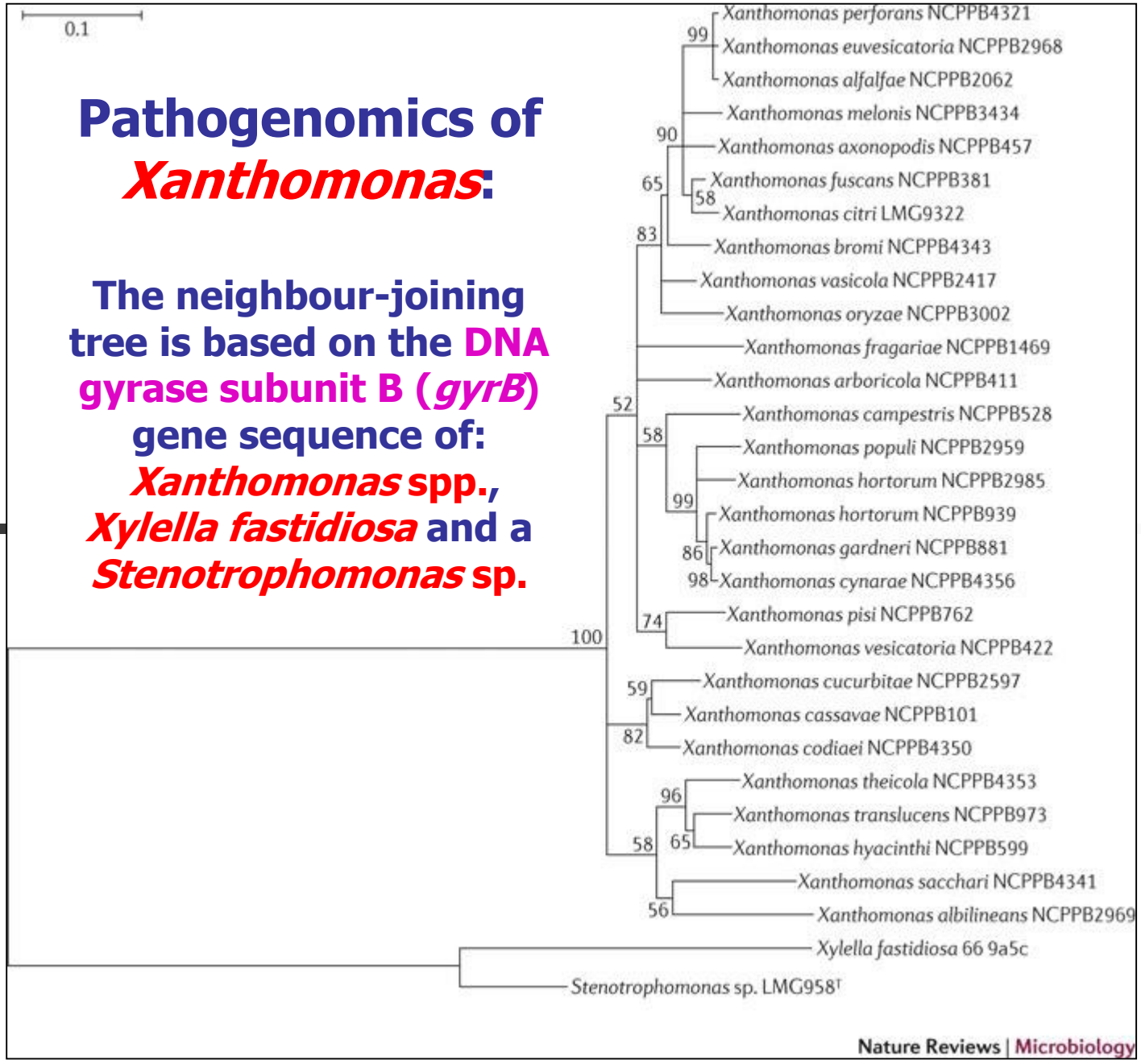
Number of profiles described for the different species from the genus *Erwinia* based on RFLP analysis of *gyrA* gene fragment

Subspecies	Number strains	Amount profils	Number profils
E. carotovora subsp. atroseptica	43	7	1-7
E. carotovora subsp. carotovora	94	24	8-31
E. carotovora subsp. batavasculorum	12	1	33
E. carotovora subsp. odorifera	12	1	31
E. carotovora subsp. wasabiae	8	1	32
E. rhapontici	3	1	34
E. cacticida	12	3	35-37

Lojkowska *et al.*,2004

Pathogenomics of *Xanthomonas*:

The neighbour-joining tree is based on the DNA gyrase subunit B (*gyrB*) gene sequence of: *Xanthomonas* spp., *Xylella fastidiosa* and a *Stenotrophomonas* sp.



Ryan *et al.*,2011

Polyphasic approach

4. Based on *lrp* gene

Leucine-responsive regulatory protein (*lrp*) gene

- Because of universal presence of *lrp* gene in genus *Xanthomonas*, as well as in other Gram-negative bacteria, and the fact that horizontal transmission is improbable, analysis of the *lrp* gene may be a widely adaptable tool for polyphasic taxonomic studies of bacteria.
- This method should be particularly valuable as a complement to current rDNA- and rep-PCR-based approaches, due to its capability for high resolution at the specific and infraspecific levels.

Polyphasic approach

Based on *lrp* gene

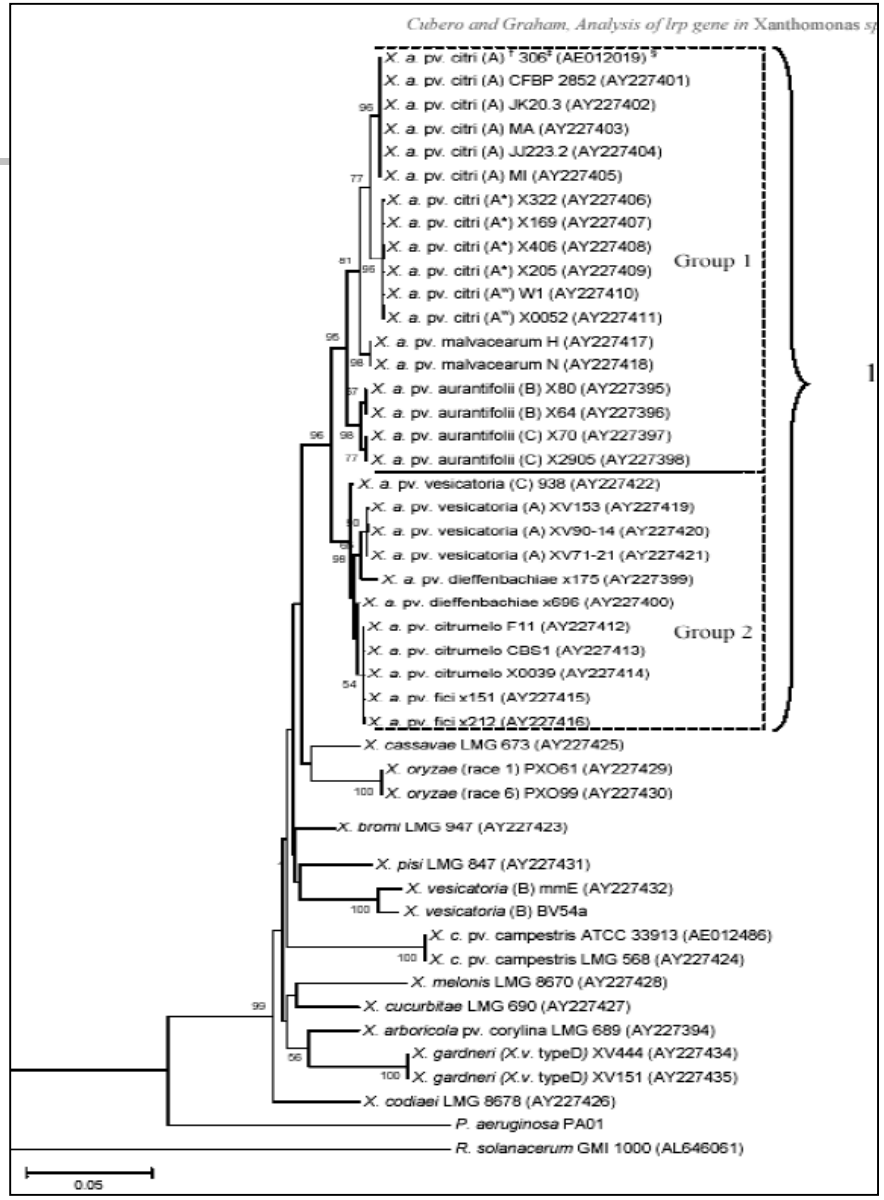
A new tool for polyphasic analysis of the genus *Xanthomonas*

- Sequence analysis of the leucine-responsive regulatory protein (*lrp*) gene (A transcriptional regulator) revealed a high level of similarity within the genus *Xanthomonas*.
- Higher variability was observed at the nucleotide level than at the amino acid level among the different species and pathovars.
- Variability in the *lrp* gene was higher than that in 16S rDNA alone or 16S rDNA combined with the rDNA ITS region and slightly lower than that in the ITS region alone.
- Thus, the moderate variability in the *lrp* gene sequence may be used to investigate phylogenetic relationships and to clarify the classification of xanthomonads.

Cubero & Graham,2003

Phylogeny based on Leucine responsive regulatory protein (*LRP*) analysis

- A dendrogram based on pairwise comparison of all *lrp* sequences showed that *Xanthomonas* species grouped together, with *P. aeruginosa* and *R. solanacearum* as outgroups.
- The largest cluster within *Xanthomonas* was cluster 1 which encompassed all strains that are presently proposed as pathovars of *X. axonopodis* (Vauterin *et al.*,2000).



Cubero & Graham,2003

New methods for description and classification of bacterial species

Mulitlocus enzyme electrophoresis (MLEE)

- Mulitlocus enzyme electrophoresis (MLEE):
- MLST is different from MLEE in assignation of alleles(one of a pair of genes at a single locus) by nucleotide sequencing rather than the electrophoretic mobility of their gene products.
- Classifies bacteria on the basis of the isoforms of a combination of approx. 15 metabolic enzymes.
- Drawback:

1. Determines phenotypes and not genotypes.
2. Phenotype of the enzyme can easily be altered in response to environmental conditions and badly affect the reproducibility of MLEE results.
3. low throughput (intensive laboratory work);
4. not widely used.

Medini *et al.*,2008

Polyphasic approach
Multiplex PCR (mPCR)

- A multiplex PCR is a phylogentic method has been designed to aid in identification and differentiation of the bacteria like *Agrobacterium*, *Erwinia*, *R.solanacearum*, etc.
- e.g. One-step multiplex PCR is used for simultaneously identification of the both *ams*-genes and pEA29 of *E. amylovora*.

Multiplex PCR: a novel method of simultaneous amplification of multiple DNA fragments (usually all target genes). The multiplex PCR assay is useful for low-cost screening of large numbers of isolates with rapid analytical capacity and could be utilized in most laboratories.
Multilocus sequence typing (MLST): a reliable method for simultaneous amplification of multiple DNA fragments (usually housekeeping genes). However, the method is costly, time-consuming and difficult to use for screening large numbers of isolates.

The species complex

Diversity of tumorigenic agrobacteria isolated from poplar crown galls from Orléans

- At present, tumorigenic bacteria belong to the:

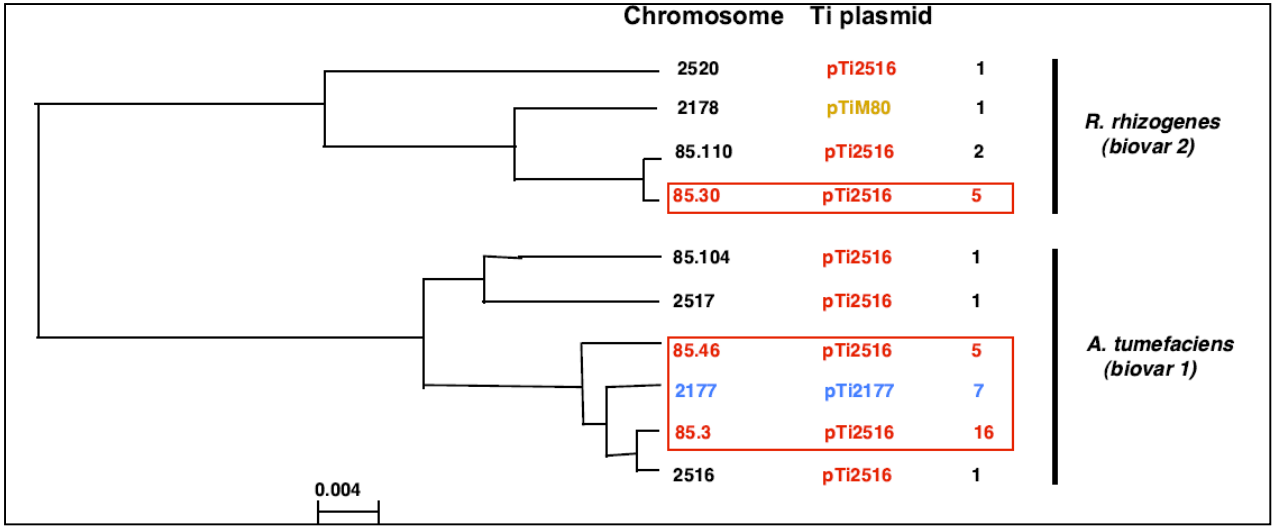
1. Species complex *A. tumefaciens* (= *A. radiobacter* = biovar 1) which consist of 10 genomovars (G1 to G7, G9, G13 and G14), and
2. The six species *A. vitis* (= biovar 3), *A. rubi*, *A. larrymoorei*, *A. rhizogenes* (= biovar 2), the recently described *A. fabrum (Lassalle et al.*,2011) likely involved in plant-bacteria commensal interaction and *Rhizobium skierniewicense* (Pulawska *et al.*,2012).

The species complex

Diversity of tumorigenic agrobacteria isolated from poplar crown galls from Orléans

- There is generally more than a single strain involved in crown gall outbreaks at both:

1. Chromosome, and
2. Ti plasmid levels.



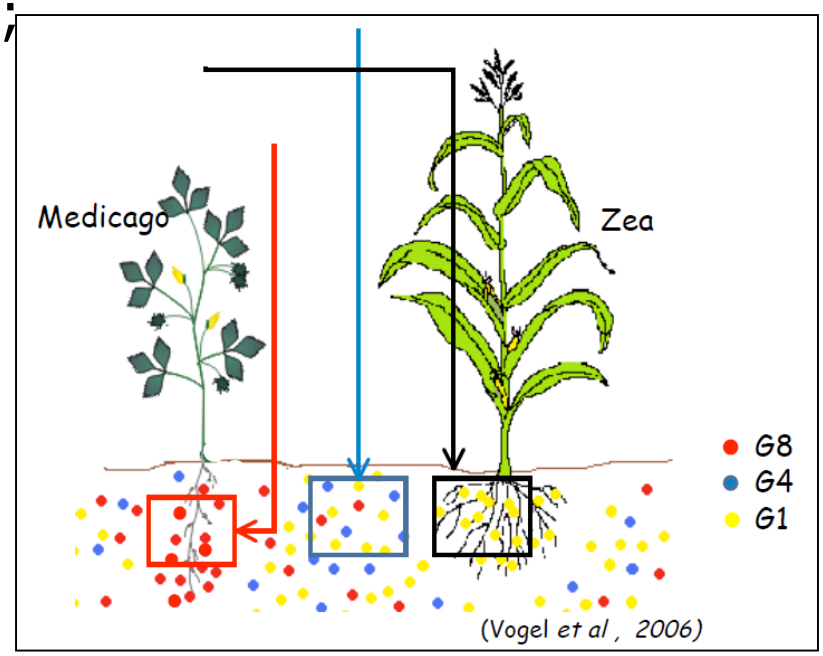
The species complex

Genomic species diversity in *Agrobacterium* spp.

A. tumefaciens (=*A. radiobacter* =biovar 1) consist of 10 genomovars

- The species complex *A. tumefaciens* (=*A. radiobacter* =bv.1) consist of 10 genomovars (G1 to G7, G9, G13 and G14).
- *Agrobacterium* genomovars are differentially selected by plants.

1. G1, G8 et G4 in the same soil;
2. G1 in maize rhizospheres;
3. G8 in medic rhizospheres.



The question is are genomovars ecovars ?

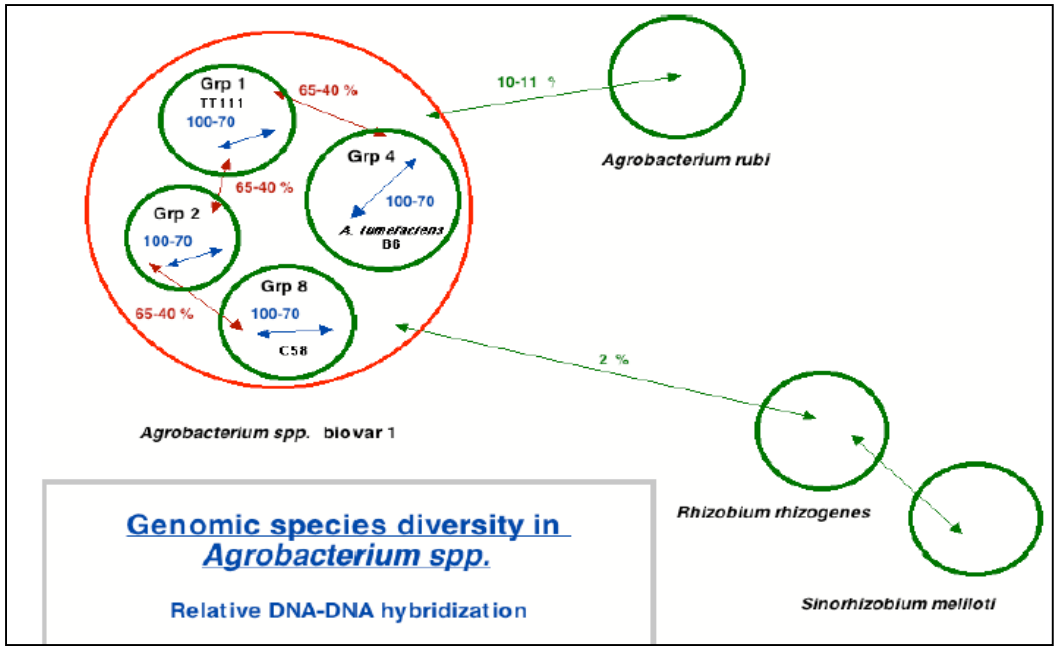
Genomovars are phenotypically similar, genotypically distinct groups of strains.

The species complex

RBS(ribosome-binding site) reveals several genomovars in the biovar 1/*A. tumefaciens* complex

- How to rapidly identify *Agrobacterium* species in the outside environment to study their ecological specificity without performing tedious RBS studies?

A ribosomal binding site (RBS) is a sequence on mRNA that is bound by the ribosome when initiating protein translation.

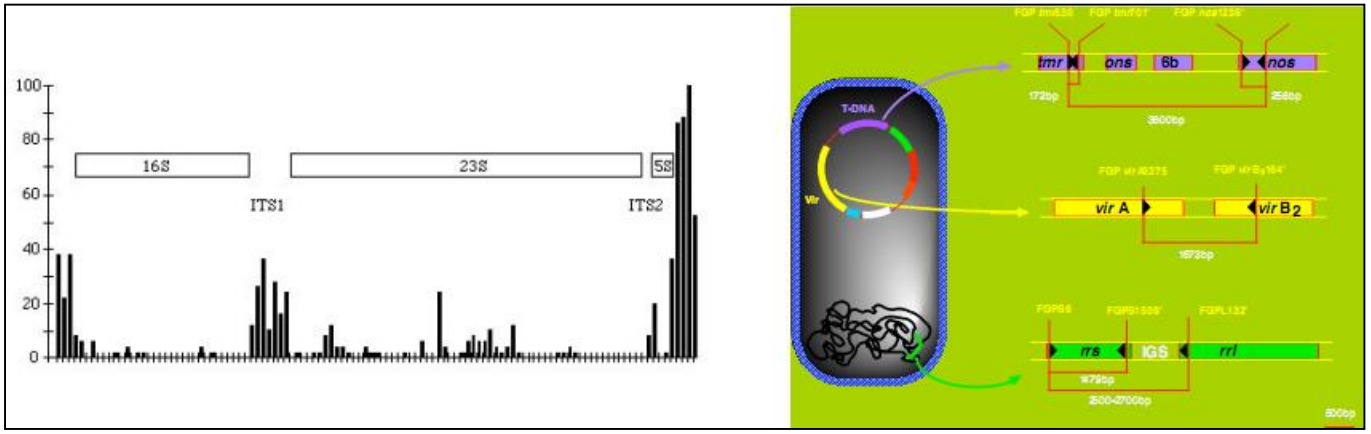


PCR detection and identification of tumorigenic agrobacteria

PCR primers designed based on chromosome (16S+ITS), or Ti plasmids

- 16S+16S-23S-ITS regions useful at the infraspecific level.
- PCR primers designed to type:

1. Chromosome (16S+ITS), or
2. Ti plasmids (vir or T-DNA conserved regions).



PCR detection and identification of tumorigenic agrobacteria
A multiplex PCR

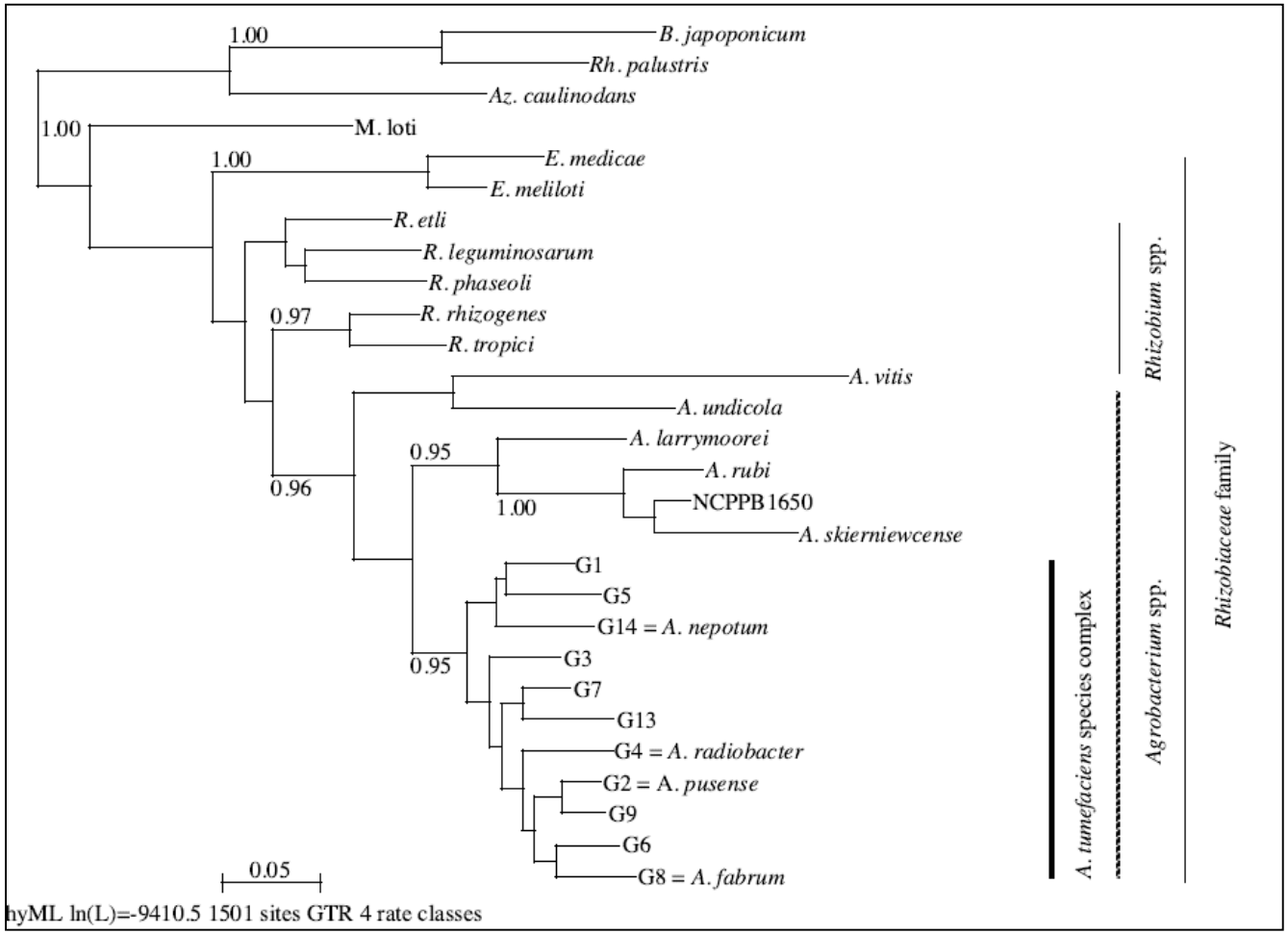
- The genetic diversity was evaluated for these isolates by comparing DNA samples using multiplex PCR with three specific primer pairs.
- The isolates were segregated into three main groups:

1. The first group that is isolated carry octopine type Ti plasmids;
2. The second group that is isolated carry vitopine Ti plasmids and,
3. The third group that is isolated carry both octopine and vitopine type Ti plasmids.

Tolba and Zaki,2011

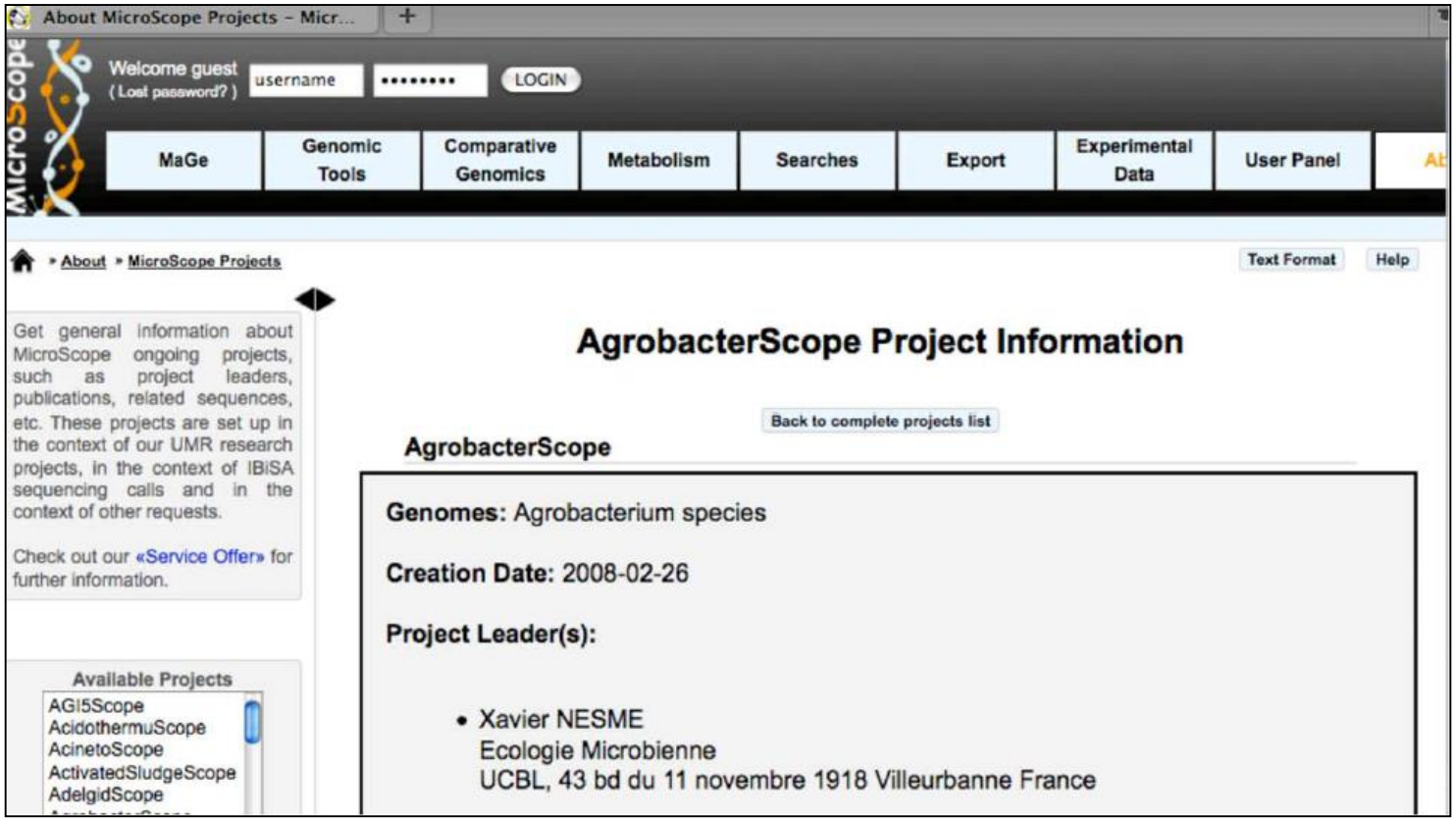
Revised nomenclature of *Agrobacterium* spp.

Last version obtained in the present 2011 COST meeting



Agrobacterscope Project Information

16 new *A. tumefaciens* genomes sequenced: at least one per genomovar



New methods for description and classification of bacterial species
Multilocus sequence typing (MLST)

- Multilocus sequence typing (MLST):
- Based on the partial sequences of 7 housekeeping genes of approx. 450 bp each.

1. high throughput (output);
2. allows direct comparison between different laboratories;
3. > Database (MLST Public Repository).

- Drawback:

1. For some species there is too little sequence variation in the housekeeping genes for a sufficient discrimination.

- Drawback for 16S rRNA and MLST:

1. Only limited genome coverage.

Medini *et al.*,2008

New methods for description and classification of bacterial species

Multilocus sequence typing (MLST)

MLST Public Repository

- MLST is based on the analysis of allelic profiles generated by comparing sequences to an online repository(store).
- **PulseNet vs. MLST websites:**
- PulseNet (the molecular subtyping network for foodborne bacterial disease) is only accessible by their member laboratories due to privacy and confidentiality issues. Whereas
- MLST websites host publicly accessible databases where any laboratory can submit data.

TABLE 1
Online molecular typing databases

Method	Database	URL
MLST	MLST.net	http://www.mlst.net
	Pubmlst.org	http://www.pubmlst.org
	Institut Pasteur MLST	http://www.pasteur.fr/mlst/
	European Working Group for Legionella Infections Sequence-based typing database	http://www.hpa-bioinformatics.org.uk/legionella/legionella_sbt/php/sbt_homepage.php
	Environmental Research Institute, University College Cork	http://mlst.ucc.ie/
MLVA	MLVAbank	http://minisatellites.u-psud.fr/MLVAnet/
	Groupe d'Etudes en Biologie Prospective	http://www.mlva.eu
	MLVAplus	http://www.mlvaplus.net/
	Institute Pasteur MLVA: MLVA-NET	http://www.pasteur.fr/mlva
	MLVA.net	http://www.mlva.net
ccrB typing	Staphylococci ccrB sequence typing	http://www.ccrbtyping.net/
dru typing	dru typing database	http://www.dru-typing.org
spa typing	Ridom Spa Server	http://spaserver.ridom.de/
CRISPR typing	CRISPRdb	http://crispr.u-psud.fr/crispr/

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeat; MLST: multilocus sequence typing; MLVA: multilocus variable-number tandem repeat analysis.

Carriço *et al.*,2013

Polyphasic approach

Based on multilocus sequence typing (MLST) analysis

- Multilocus sequence typing (MLST) is another phylogentic method in molecular biology for the typing of multiple loci.
- The procedure characterizes isolates of bacterial species using the DNA sequences of internal fragments of multiple (usually seven) housekeeping genes.
- Approximately 450-500 bp internal fragments of each gene are used, as these can be accurately sequenced on both strands using an automated DNA sequencer.
- For each housekeeping gene, the different sequences present within a bacterial species are assigned as distinct alleles and, for each isolate, the alleles at each of the loci define the allelic profile or sequence type (ST).

Allele: One of a pair of genes at a single locus.

Principle of MLST

Multilocus sequence typing (MLST) analysis

- Molecular phylogeny deduced from a single locus may be unreliable due to the stochastic nature of base substitutions or to rare horizontal gene transfer events (Yamarnoto *et al.*,2000).
- MLST directly measures the DNA sequence variations in a set of housekeeping genes and characterizes strains by their unique allelic profiles.
- The principle of MLST is simple:
- The technique involves PCR amplification followed by DNA sequencing.
- Nucleotide differences between strains can be checked at a variable number of genes (generally seven) depending on the degree of discrimination desired.

Wikipedia,2011

Principle of MLST

Multilocus sequence typing (MLST) analysis

- The workflow of MLST involves:

1. Data collection,
2. Data analysis, and
3. Multilocus sequence analysis.

- In data collection, definitive identification of variation is obtained by nucleotide sequence determination of gene fragments.
- In data analysis:
- All unique sequences are assigned allele numbers, and
- Combined into an allelic profile, and
- Assigned a sequence type (ST).
- In the final section of MLST the relatedness of isolates are made by comparing allelic profiles.

Wikipedia,2011

Polyphasic approach

Genomic approaches to estimate diversity

Based on multilocus sequence typing (MLST) analysis

- Sequence 400-600bp of at least 6 housekeeping genes.
- Identical alleles given same profile designation ("Sequence Type").
- Advantages:

I. Higher variability in housekeeping genes compared to 16S rDNA. Increased length sequence analyzed (differentiation below sp. level)

II. Sequencing at least 6 genes reduces risk of interference from horizontal gene transfer.

De La Fuente,2009

MLST: www.mlst.net

MLST
Multi Locus Sequence Typing

Policy Document

Sunday 7th

DATA ANALYSIS

DATABASES

- *B.burgdorferi*
- *B.cereus*
- *B.pseudomallei*
- *C.albicans*
- *C.glabrata*
- *C.krusei*
- *C.tropicalis*
- *C.jejuni*
- *C.neoformans var grubii*
- *E.coli*
- *E.faecalis*
- *E.faecium*
- *H.influenzae*
- *H.pylori*
- *Leptospira spp.*
- *M.catarrhalis*
- *N.meningitidis*
- *S.agalactiae*
- *S.aureus*
- *S.enterica*
- *S.epidermidis*
- *S.pneumoniae*
- *S.pyogenes*
- *S.suis*
- *V.vulnificus*

DATABASES

Please select the database you would like to access / query from the left hand menu

Navigation

- **PubMLST**
  - MLST Home
  - Search / site map
- **Software**
  - Bio-Linux
  - Web tools
  - Software
- **Bacteria**
  - *A. baumannii*
  - *Arcobacter*
  - *B. cereus*
  - *Bordetella*
  - *B. hyodysenteriae*
  - *Brachyspira*
  - *B. cepacia*
  - *C. fetus*
  - *C. helveticus*
  - *C. insulaenigrae*
  - *C. jejuni & C. coli*
  - *C. lari*
  - *C. upsaliensis*
  - *Chlamydiales*
  - *H. pylori*
  - *M. haemolytica*
  - *Neisseria*
  - *P. aeruginosa*
  - *P. gingivalis*
  - *S. agalactiae*
  - *S. maltophilia*
  - *S. oralis*
  - *S. uberis*
  - *S. zooepidemicus*
  - *Streptomyces*
  - *V. parahaemolyticus*
  - *V. vulnificus*
  - *Wolbachia*
- **Eukaryotes**
- **Other dbases**
- **Mirrors**
- **Developers**

MLST of *Burkholderia cepacia* complex

Genes

The *Burkholderia cepacia* complex MLST scheme uses fragments of the following seven house-keeping genes:

ATP synthase beta chain (*atpD*)
Glutamate synthase large subunit (*gltB*)
DNA Gyrase subunit B (*gyrB*)
Recombinase A (*recA*)
GTP binding protein (*lepA*)
Acetoacetyl-CoA reductase (*phaC*)
Tryptophan synthase subunit B (*trpB*)

PCR amplification

Reaction conditions for all the primers were as follows: initial denaturation at 96°C for 1 min; 30 cycles of denaturation at 96°C for 1 min, primer annealing at 58°C for 1 min, extension at 72°C for 2 min; followed by a final extension step of 72°C for 5 min. Each 50 µl amplification reaction mixture comprised: 10 µl Q solution (Qiagen), 4.0 µl chromosomal DNA (5-20 ng/µl), 2.0 µl forward primer (10 pmol/µl), 2.0 µl reverse primer (10 pmol/µl), 5.0 µl 10x PCR buffer (Qiagen, contains 15 mM MgCl2), 1.0 ?l dNTP solution (Qiagen, 10 mM each dNTP), 0.25 µl Taq polymerase (Qiagen, 5units/µl) and 25.75 µl PCR-grade water. All Qiagen solutions from PCR CORE Kit (Cat No. 201225).

The amplification product was then purified using MinElute UF plates (Qiagen) following the manufacturers protocol before being used in a sequencing reaction. Sequencing was carried out on each DNA strand with BigDye Terminator Ready Reaction Mix v3.1 (PE Biosystems, Foster City, US) under standard sequencing conditions according to the manufacturer's protocol. Unincorporated dye terminators were removed by precipitation with 95% alcohol. The reaction products were separated and detected on an ABI PRISM genetic analyser 3100 (PE Biosystems) using a standard sequencing module with a Performance Optimised Polymer and 5 cm array.

The primers used in MLST analysis of *B. cepacia* complex are listed below.

Gene	Direction	Primer Sequence
atpD	F	GATCGTACAGTGCATCGG
	R	ATCGTGCCGACCATGTAG
gltB	F	CGCTCGAAGATCAAGCAG
	R	GGGAACACCTTCACGAAC
gyrB	F	CGACAACTCGATCGACGA
	R	GACAGCAGCTTGTCGTAG
recA	F	GATAGCAAGAAGGGCTCC
	R	CTCTTCTTCGTCCATCGCCTC
lepA	F	CGACGGCAAGGTCTACAA
	R	AGCATGTCGACCTTCACG
phaC	F	CTCAGCGAATTGCGTACG
	R	CCGTTCAGCGAGAAGTCG
trpB	F	GATCTACCTGAAGCGCGA
	R	GTGTGCATGTCCTTGTCG

Citing the database

The preferred format for citing this website in publications is:

This publication made use of the *Burkholderia cepacia* complex Multi Locus Sequence Typing website (http://pubmlst.org/bcc/) developed by Keith Jolley and sited at the University of Oxford (Jolley et al. 2004, *BMC Bioinformatics*, **5**:86). The development of this site has been funded by the Wellcome Trust.

Status

Profile database
Profiles: 563
Last updated: 2009-05-28

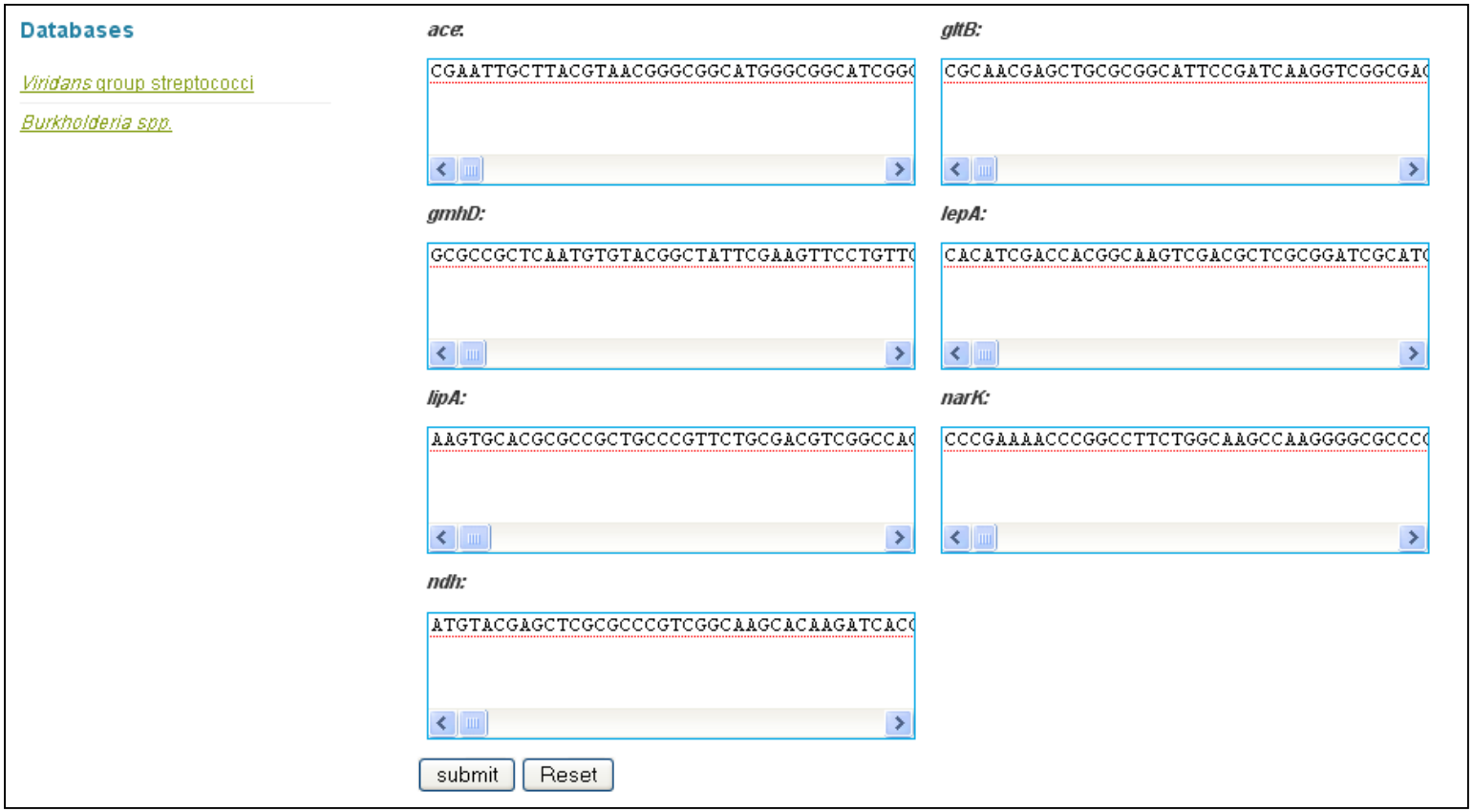
Isolate database
Isolates: 923
Last updated: 2009-05-28

MLST: www.mlst.net
Burkholderia spp.

- Please paste your sequences as one continuous string (no spaces) in the following boxes and click submit.
- The boxes contain sequences which can be used to demonstrate the features of eMLSA.net.
- Click 'Reset' to enter your own gene sequences.

MLST: www.mlst.net

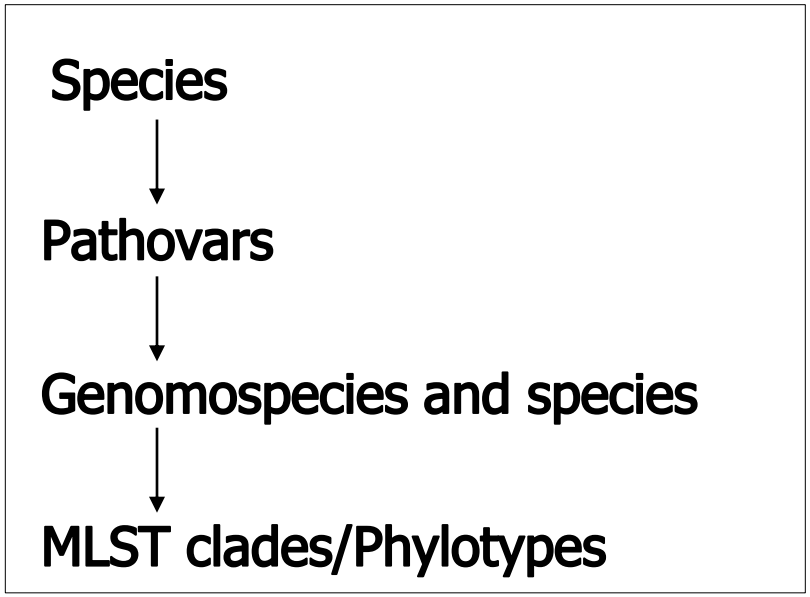
Burkholderia spp.



Classification of pseudomonads

Based on multilocus sequence typing (MLST) analysis

- MLST is a highly discriminating, rapid, and portable DNA-based strain typing method in which regions from several housekeeping loci are sequenced from each strain.
- A preliminary multilocus sequence typing (MLST) analysis of 367 strains within the genus *Pseudomonas* identified major clades corresponding to *P. syringae*, *P. viridiflava*, *P. fluorescens*, *P. putida*, *P. aeruginosa*, and *P. stutzeri*.



Guttman *et al.*,2008;David Stead

Classification of pseudomonads

Based on multilocus sequence typing (MLST) analysis

- We report here on the multilocus sequence typing (MLST/MLSA) of the genus *Pseudomonas*.
- Four loci totaling 2046 bp of sequence were obtained from 367 Pseudomonads.
- The four housekeeping genes sequenced were:

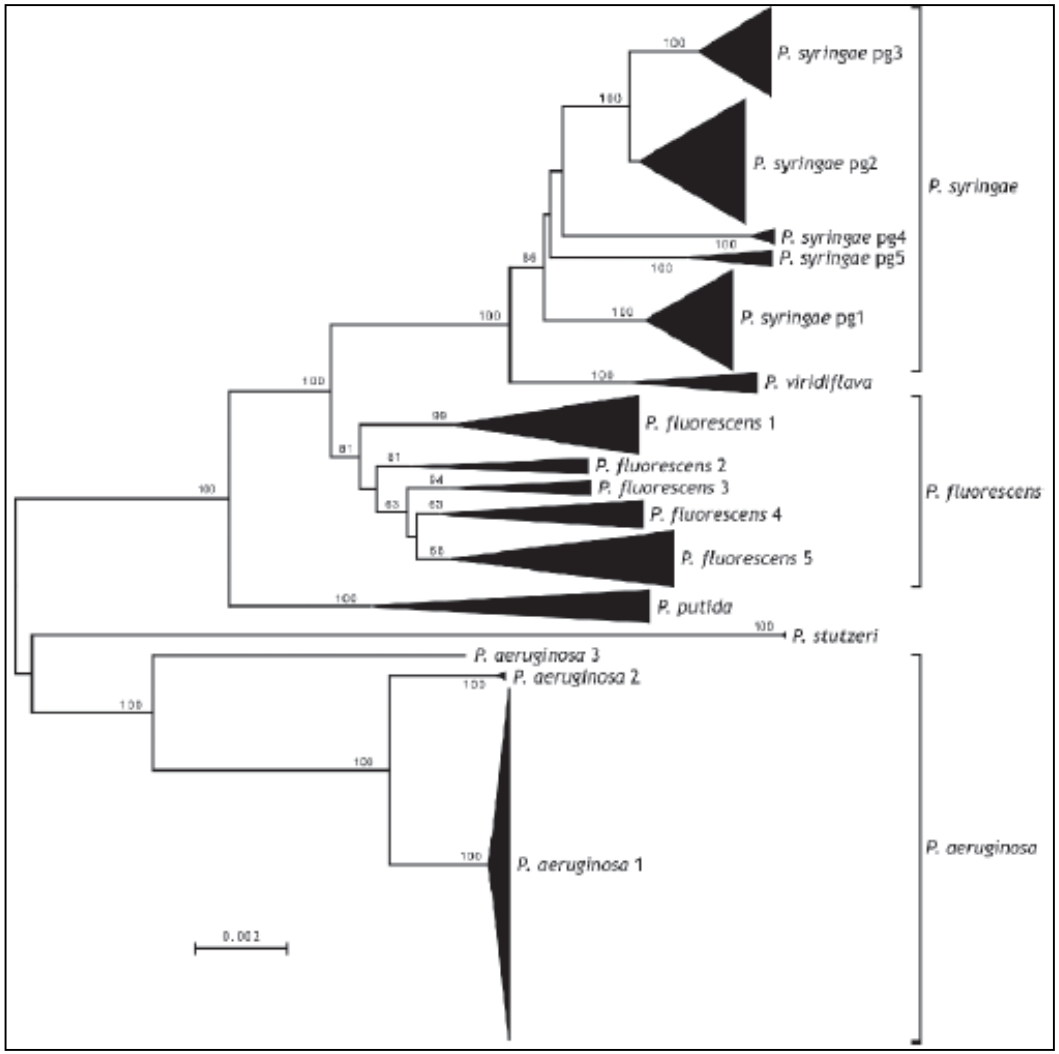
1. *rpoD*, encoding sigma factor 70,
2. *gyrB*, encoding DNA gyrase B,
3. *gltA* (also known as *cts*), encoding citrate synthase, and
4. *gapA*, encoding glyceraldehyde-3-phosphate dehydrogenase.

Guttman *et al.*,2008

Polyphasic approach *Pseudomonas*

Based on Multilocus sequence typing (MLST) analysis

MLSA of *Pseudomonas* species.



Guttman *et al.*,2008

Classification of pseudomonads

Based on multilocus sequence typing (MLST) analysis

P. syringae

- Applying MLST technique on sequencing 6-8 selected housekeeping genes and identification of polymorphic nucleotide sites has been introduced for the characterization of *P. syringae* isolates.
- Combination of the alleles at the different loci results in unique diploid sequence types that can be used to discriminate strains.

Polyphasic approach

New taxonomy of the *R. solanacearum* species

Based on three genes including *egl* (endoglucanase gene sequence), a new marker used for *R. solanacearum* diversity

- The most recent study compared the:

1. 16S rDNA sequences,
2. *egl* (endoglucanase sequences), and
3. *hrpB* sequences from the most representative set of strains studied to date and consistently found four major divisions (phylotypes) that correlated with geographic origin.

Phylotypes: A group that is part of the same phylogenetic cluster based on the analysis of three loci.

Denny,2006

Polyphasic approach

New taxonomy of the *R. solanacearum* species

Based on pan-genome and comparative genomic analysis approach drawing on previously published *R. solanacearum* genome sequences

- Based on the pan-genome (the set of genes present in the collective genomes of the *R. solanacearum* species complex) and computation of genomic distances between sequenced *R. solanacearum* strains, we again suggested that the *R. solanacearum* species complex should be restructured into three different species:

1. One containing phylotypes I and III,
2. A second containing phylotype II, and
3. A third containing *R. solanacearum* strains from phylotype IV.

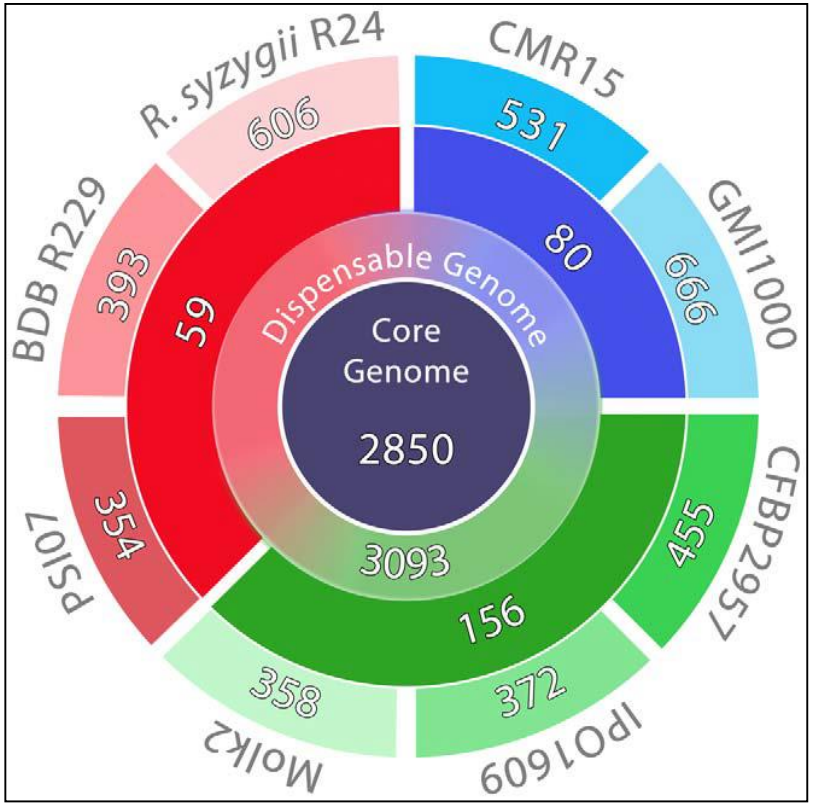
Remenant *et al.*,2010 and 2011

Polyphasic approach

New taxonomy of the *R. solanacearum* species

Based on pan-genome and comparative genomic analysis approach drawing on previously published *R. solanacearum* genome sequences

- Number of genes in the species complex pan genome.
- From inside to outside:
- 1, Core Genome,
- 2, Dispensable Genome,
- 3, Specific Genome at the phylotype level.
- Blue: phylotype I and III;
- Green: phylotype II;
- Red: phylotype IV, and
- 4, Specific Genome at the strain level.



Remenant *et al.*,2010 and 2011

Polyphasic approach

Pathogen diversity

Phylotype and sequevar determination in *R. solanacearum* species complex

- The *Ralstonia solanacearum* species complex consists of four phylogenetically distinct major lineages, named phylotypes.
- Each phylotype contains strains primarily isolated from specific geographic areas:

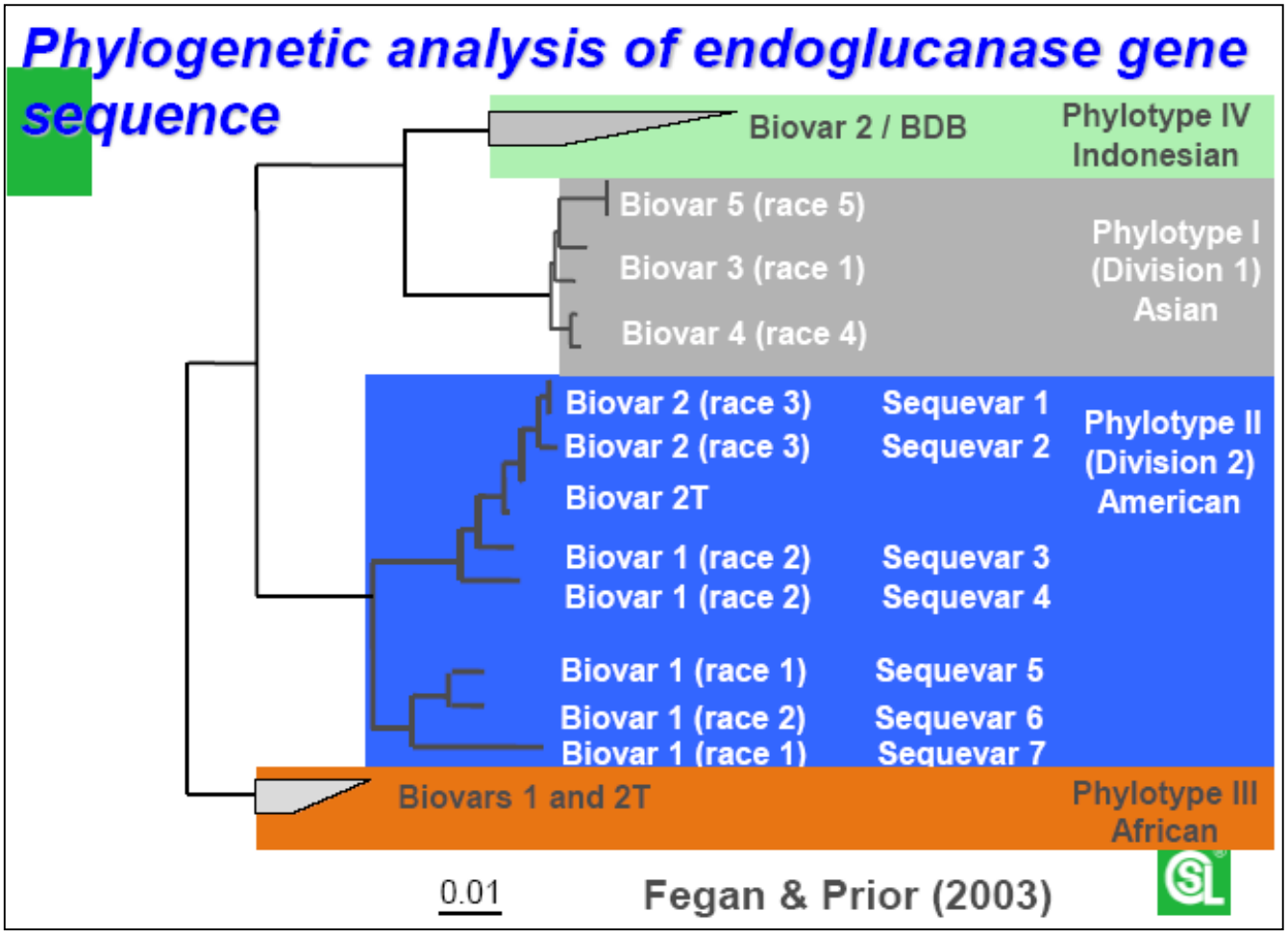
1. Phylotype I strains are from Asia;
2. Phylotype II are from the Americas;
3. Phylotype III are from Africa; and
4. Phylotype IV are from Indonesia, Japan, Australia, and the Philippines.

Patrice,2001;Denny,2006

Pathogen diversity

Sequevar determination

Determination uses endoglucanase gene sequence



David Stead

Polyphasic approach

Based on Multilocus sequence typing (MLST) analysis

MLSA of the *R. solanacearum* complex

- 92 *Ralstonia* strains were studied:

1. 90 *R. solanacearum* of the 4 phylotypes and various ecotypes;
2. Type strains of *R. insidiosa* and *R. pickettii*.

- 8 single-copy and evenly distributed genes:

1. Housekeeping, chromosome (*leuS*, *rplB*, *gdhA*, *gyrB*, *adk*, *mutS*); and
2. Virulence-related, megaplasmid genes (*Egl*, *fliC*).

- Note that *fliC* (Flagellin encoded gene), *egl* (encodes endoglucanase) and *hrpB* (*hrpB* encodes an araC *(1-ß-D-*arabinofuranosylcytosine) type ranscriptional regulatory protein) are located on the megaplasmid.

Wicker *et al.*,2012

Multilocus sequence typing (MLST) analysis

R. solanacearum

- Virulence-related genes which are located on the megaplasmid, are implicated directly as (*egl*) or indirectly as (*hrpB* and *fliC*) in disease-causing processes.

1. Flagellin- encoded by the *fliC* gene, is the essential subunit of the flagellar filament that is needed for invasive virulence.
2. *hrpB-* encodes an araC (*1-β-D*-arabinofuranosylcytosine) type transcriptional regulatory protein that governs multiple virulence pathways.
3. *egl-* encodes an endoglucanase that likely acts at the front line of host invasion by partially degrading host cell walls.

Wicker *et al.*,2012

Multilocus sequence typing (MLST) analysis

Phylotype and sequevar determination

- Primer sequences used in PCR:

Gene	Primer designation[a]	Primer sequence (5 ' ⟶ 3')	Annealing temperature
hrpB	RShrpBF	TGCCATGCTGGGAAACATCT	64
	RShrpBR	GGGGGCTTCGTTGAACTGC	
egl	EglF	AAATCCAGATATCGAATTGCCAA	57
	EglR	GCGTGCCGTACCAGTTCTG	
fliC	Rsol_fliCF	GAACGCCAACGGTGCGAACT	63
	Rsol_fliCR	GGCGGCCTTCAGGGAGGTC	

- This new classification scheme broadly reflects:

1. The ancestral relationships, and
2. Geographical origins of the strains.

Patrice,2001;Denny,2006;Mikhail *et al.*,2012

Pathogen diversity
Phylotype and sequevar determination

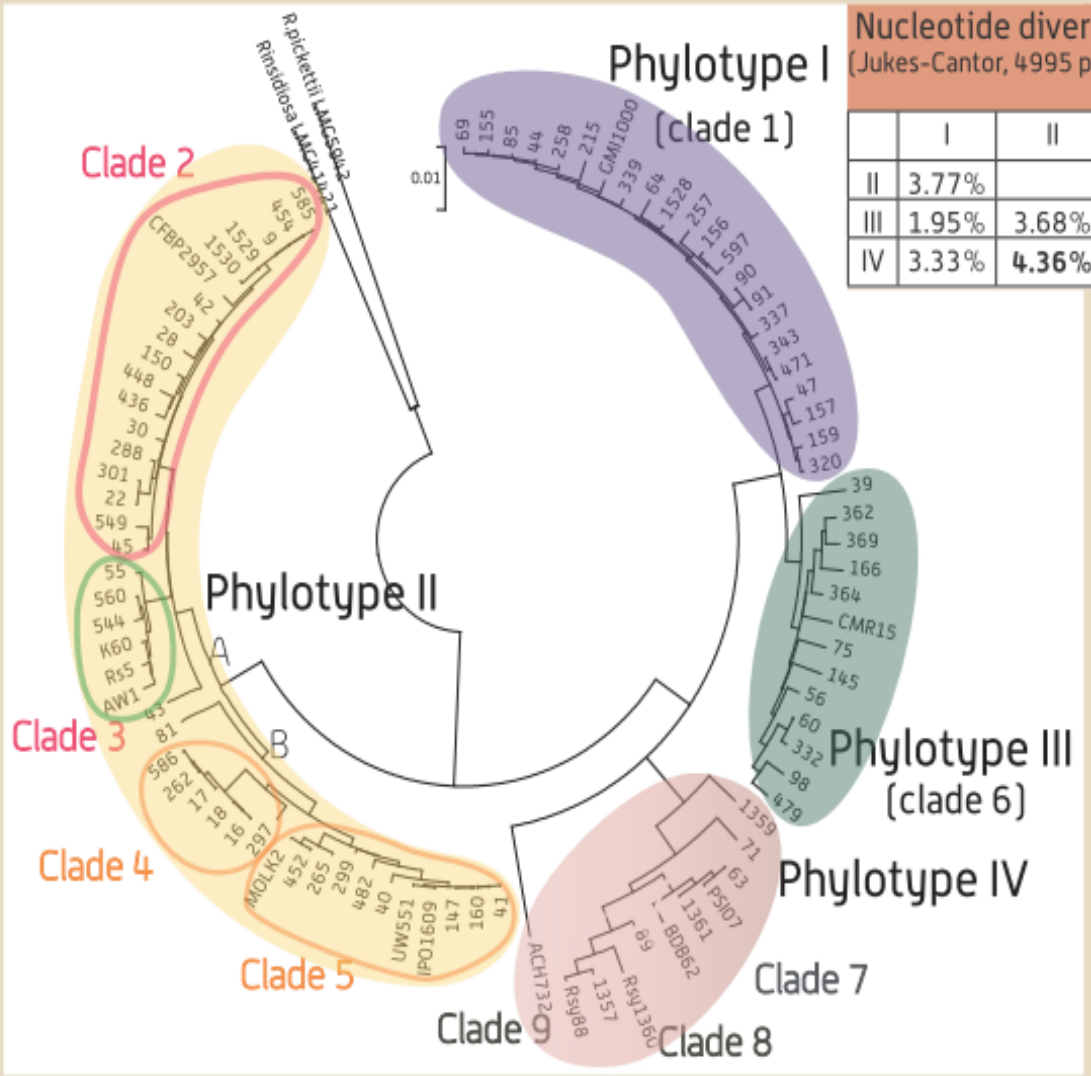
- The three virulence-related genes located on the megaplasmid in virulent and avirulent isolates of *R. solanacearum*.

Gene	Fragment length (bp)	Virulent isolates						Avirulent isolates		
		S1	S2	S3	S4	S5	S6	S7	S8	S9
hrpB	810	+	+	+	+	+	+	+	+	+
egl	686	+	+	+	+	+	+	+	+	+
fliC	318	+	+	+	+	+	+	+	+	+

Mikhail *et al.*,2012

Multi-Locus Sequence Analysis(MLSA)

R. solanacearum



Clonal Complex	Strains	Phylotype/ecotype
1	RUN0009, 0030, 0203,0 454, 0585, UW588, UW595	IIA, clade 2 (seqvar6, 35, 52) BW and Moko strains from Guatemala, Venezuela, Carribean, Cameroon
2	RUN0016, 0017, 0018, 0262, 0568	IIB, clade 4 (sequevar 4&4NPB) Moko strains from Peru, emerging strains from Martinique
3	UW551, IPO1609, RUN0041, RUN0160, RUN147	IIB, clade 5 (sequevar 1) Brown rot strains from Kenya, Cameroon, Nigeria, Réunion, Netherlands)
4	K60, AW1, Rs5, RUN0055	IIA, clade 3 (sequevar 7) BW strains from Kenya, USA
5	RUN0022, 0301	IIA, clade 2 (seqvar24) Moko strains from Brazil
6	RUN0090 &91	I Mulberry strains, China
7	RUN0265, RUN0452	IIB, clade 5 (sequevar 3) Moko strain/Honduras, peanut/ Indonesia
8	RUN0320, 0343	I BW,China & Madagascar

Maximum Likelihood tree estimated under the TIM+I+G model using the concatenated sequences of six housekeeping genes (chromosome) from 92 *Ralstonia strains*. Robustness of the nodes was evaluated by bootstraping (1000 replications).
Clades were defined within phylotypes according to their phylogenetic relevance (bootstrap values over 80%)

Wicker *et al.*,2012

Multi-Locus Sequence Analysis
MLSA of *R. solanacearum* species

- The 4 phylotypes are subdivided in different clades, but DNA divergences observed between phylotypes are below species delineation values.
- Phylotypes display different population structures:

1. Recombining (phylotype I),
2. Slightly clonal (phylotype III),
3. Highly clonal (phylotype II).

Phylotype		Clade	EGL (Fegan&Prior ,2005, unpublished 2008)	Features
I (Asian)		1	Sequevars 12-34	bacterial wilt
II (American)	*A*	2	Sequevar 7	US biovar 1, bacterial wilt
	A	3	Sequevars 6, 24, 35,36,39..	Moko disease, bacterial wilt
	B	4	Sequevar 4	Moko disease, bacterial wilt (emerging strains)
	B	5	Sequevar 1-2, 3	Brown rot Moko disease
III (African)		6	Sequevars 19-23,29, 42-44,49	Brown rot and bacterial wilt
IV		7	Sequevar 9	*R.syzygii*
		8	Sequevar 10	BDB & BW
		9 ?	Sequevar 11	ACH432

Wicker *et al.*,2012

New methods for description and classification of bacterial species

The Bacterial Pan-Genome

- There are up to 7,411 complete sequenced bacteria genomes up to date.
- The dramatic increase in genome sequencing during the last years has changed our ideas about bacteria diversity, from single gene to whole community DNA surveys; we have learned that the nature largest gene repository resides in bacteria.
- The largest amount of life's gene functions diversity resides in bacteria.

Alcaraz,2014

New methods for description and classification of bacterial species
Bacterial pan-genome

- Pan genome: Can be divided in core genome and accessory genome.
- The total of core genome and accessory genome are known as pan-genome.

1. Core genome: genes shared by all strains of the same species.
2. Dispensable genome: consisting of partially shared and strain-specific genes, i.e. genes present in some but not all of the same species.

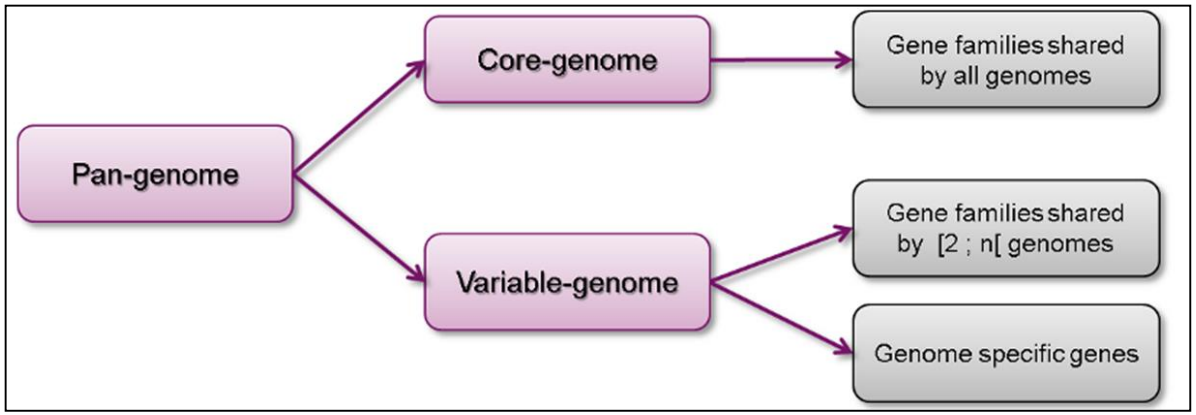
In molecular biology a pan-genome (or supra-genome) describes the full complement of genes in a species.

Campbell and co-workers; Alcaraz,2014

Bacterial pan-genome

It is the union of all the gene families and specific genes of all the strains

- Pan-genome includes:
- The core-genome containing gene families shared by all the organisms (intersection of gene families).
- The variable-genome containing genes families shared by two or more organisms and strain specific genes.



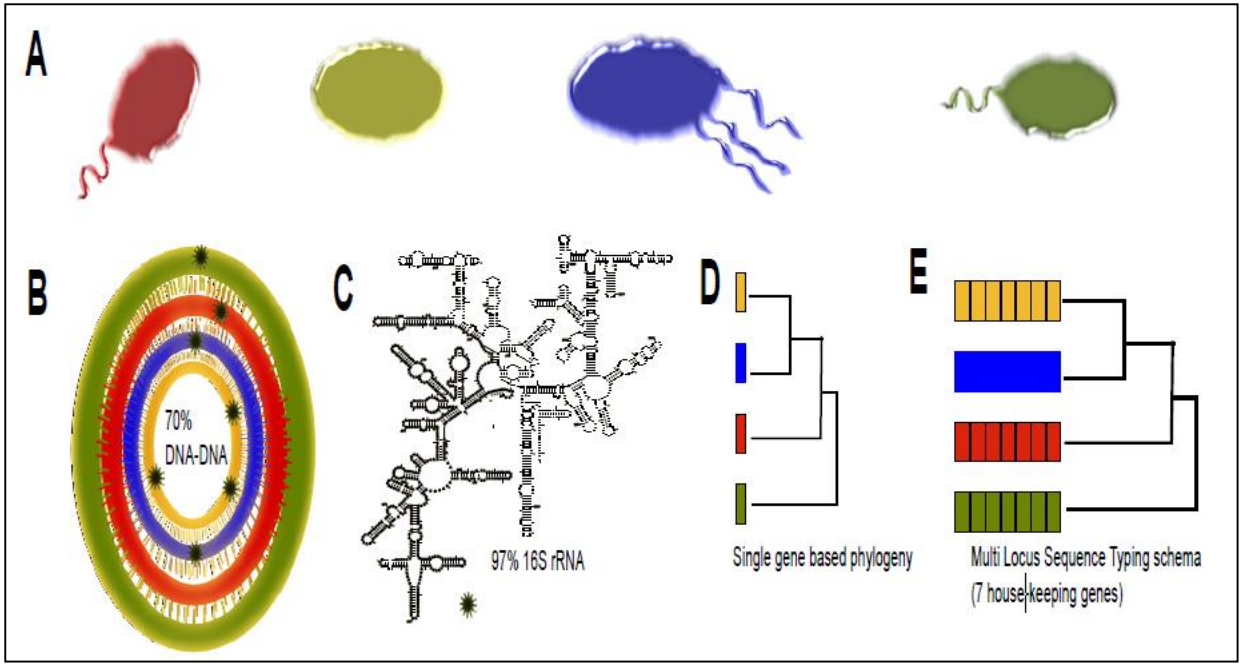
Core genome vs. Pan-Genome

1. Core genome phylogenetic: are the next level of MLST schema, comparing whole shared genome information coded in the genomes of interest strains.
2. Pan-genomics: comparing shared (core genome) and partially shared (accessory genome) genome information and can be used as a workhorse to describe both taxonomical and functional diversity within bacteria.

- The core genome set is logically much smaller than the pan-genome.

Alcaraz,2014;..

The current systematics of Bacteria
From morphological traits(A) to Multi Locus Sequence Typing (MLST)



DNA-DNA hybridization of different strains (B), the use of universally conserved 16S rRNA sequence comparison has a cut-o value of 97% identity (C), sequencing of the 16S a single gene phylogenetic analysis (D) and Multi Locus Sequence Typing (MLST) schema uses information of multiple (~7) coding gene sequences (E).

Alcaraz,2014

The Bacterial Pan-Genome

Update on the *Burkholderia* pan genome

- Among the more than 35,000 gene families that are represented in the *Burkholderia* pan-genome, a set of 717 core gene families that are conserved across all the *Burkholderia* genomes.
- To display taxonomic groupings and evolutionary relationships, three types of phylogenetic trees were constructed, based on different sets of genetic information:

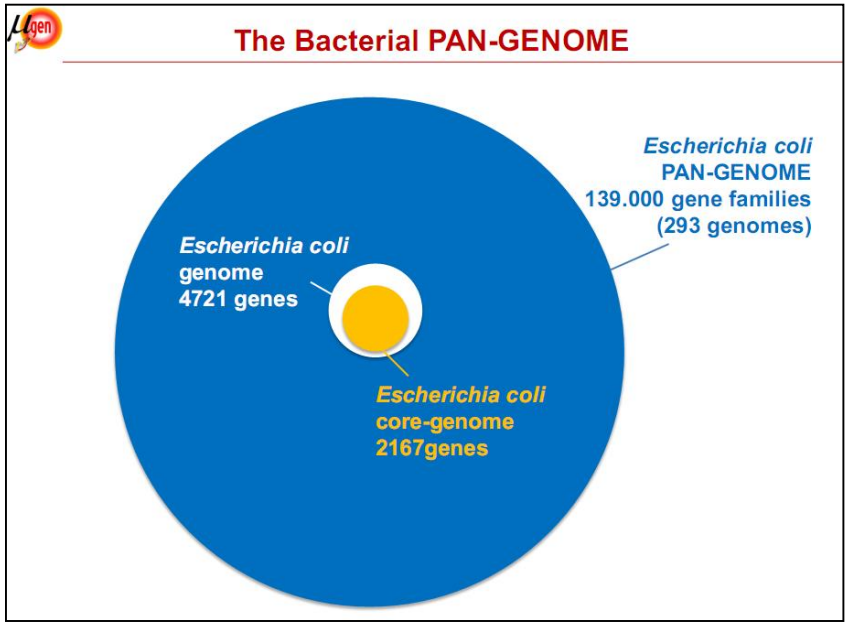
1. 16S rRNA genes,
2. multiple locus sequence tags (MLST), and
3. pan-genome sequences.

Leekitcharoenphon co-workers

The Bacterial Pan-Genome

Relevance of the pan-genome model

1. Evolution of Prokaryote,
2. Ecology,
3. Pathogenicity,
4. Biotechnology.



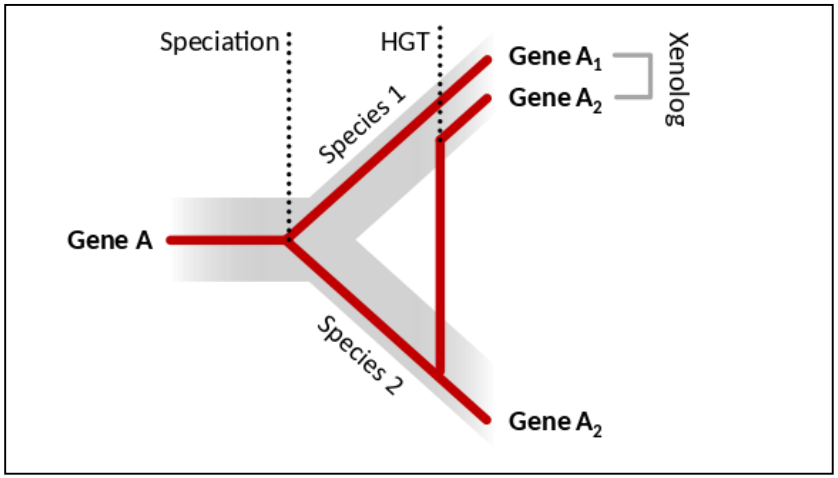
The predicted pan-genome in *E. coli* comprises 15,741 gene families, and only 993 (6%) of the families are represented in every genome, comprising the core genome (Lukjancenko *et al.*,2010).

Microgen,2013

The Bacterial Pan-Genome

Gene families in prokaryotes

- In prokaryotes, gene families could be the result of incomplete xenologous gene replacement by which a gene from another genome gets incorporated into a gene family with which it shares some sequence similarity.
- Xenolog (plural xenologs) (genetics): A type of ortholog where the homologous sequences are found in different species because of horizontal gene transfer.

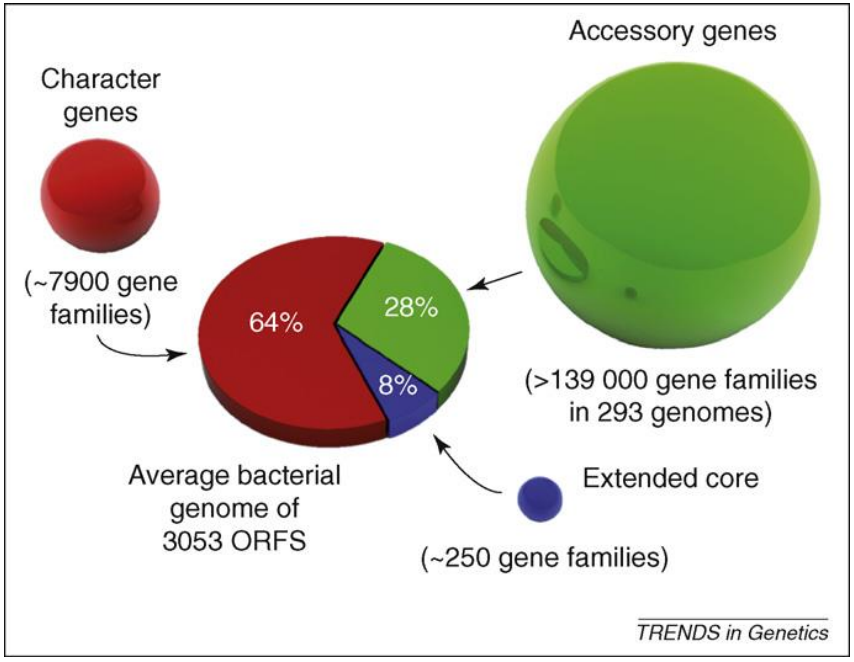


A homologous gene (similar) is a gene inherited in two species by a common ancestor. Orthologous, the copy of the same genes are generated by speciation, not by gene duplication.

The Bacterial Pan Genome

The bacterial pan genome is of infinite size

1. Analyzed 573 sequenced genomes.
2. Chose 15,000 random ORF(open reading frames or protein coding sequences).
3. Distinguished 3 groups of ORF.



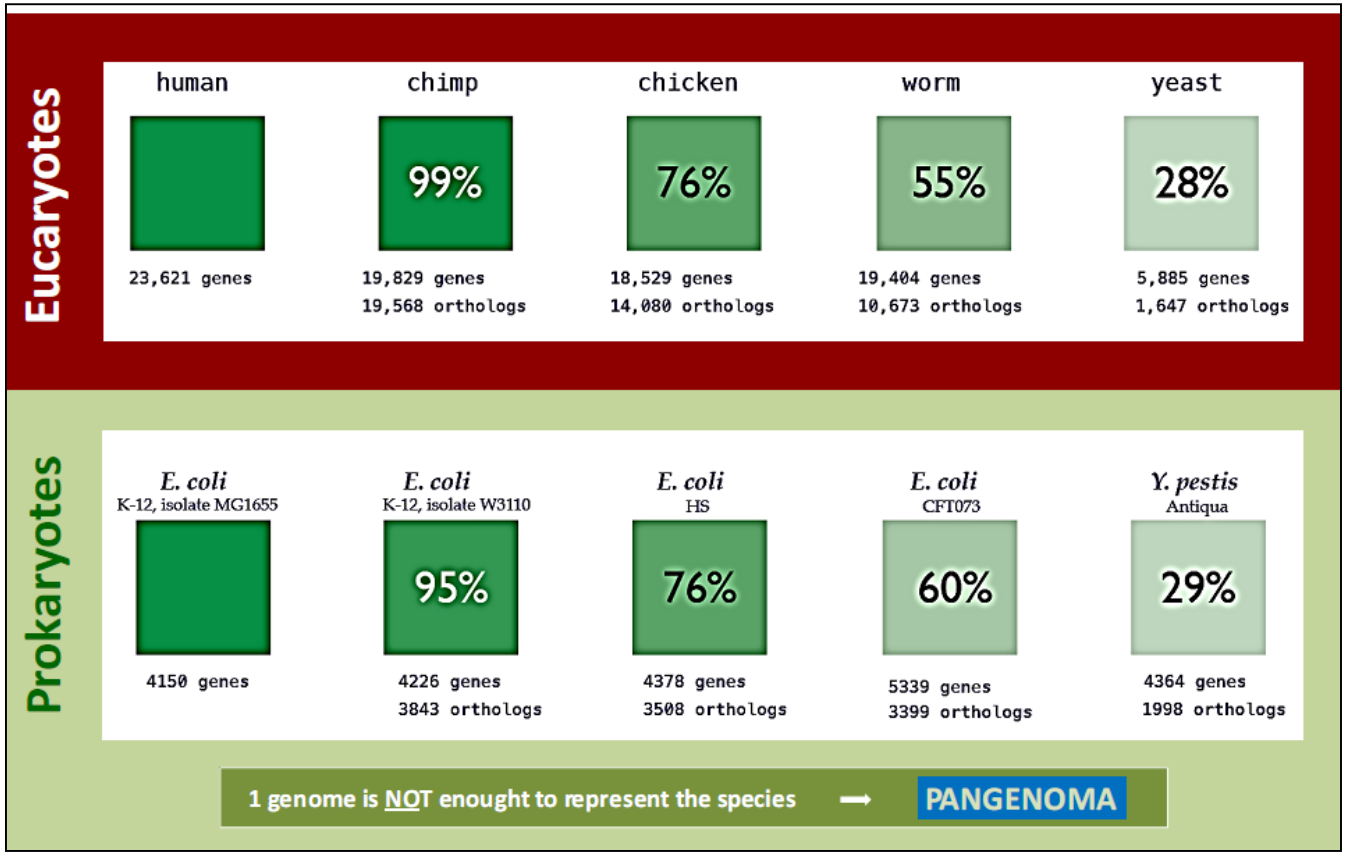
A nucleic acid sequence (DNA or RNA sequence) with a reading frame that contains nonstop (termniation) codons; it can therefore potentially be translated into a polypeptide. Each sequence contains all the information to produce a particular protein.

Lapierre and Gogarten,2009;Campbell and co-workers;..

The Bacterial Pan-Genome

Intra-species variation

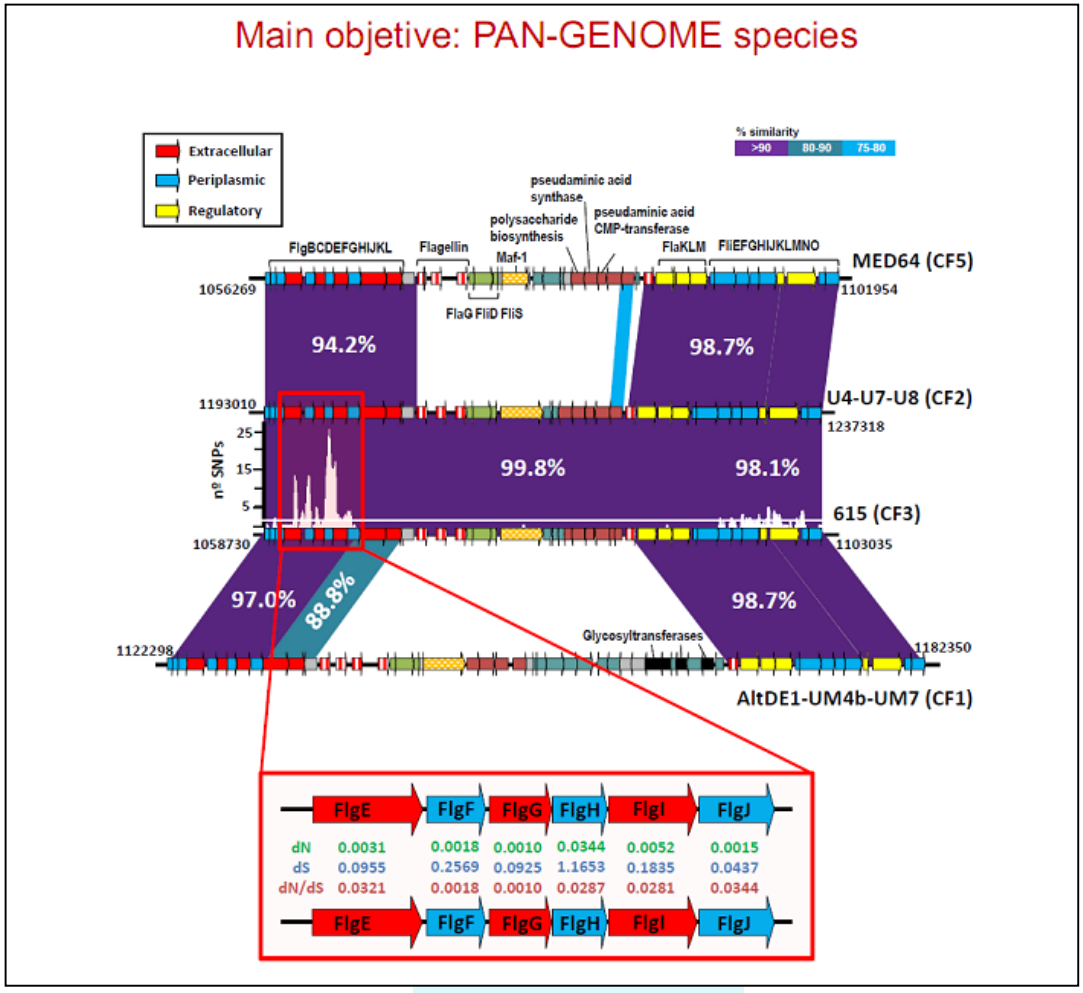
1 genome is NOT enough to represent the species



Orthologous genes: Genes in different species that are homologous (similar) because they are derived from a common ancestral gene.

Microgen,2013

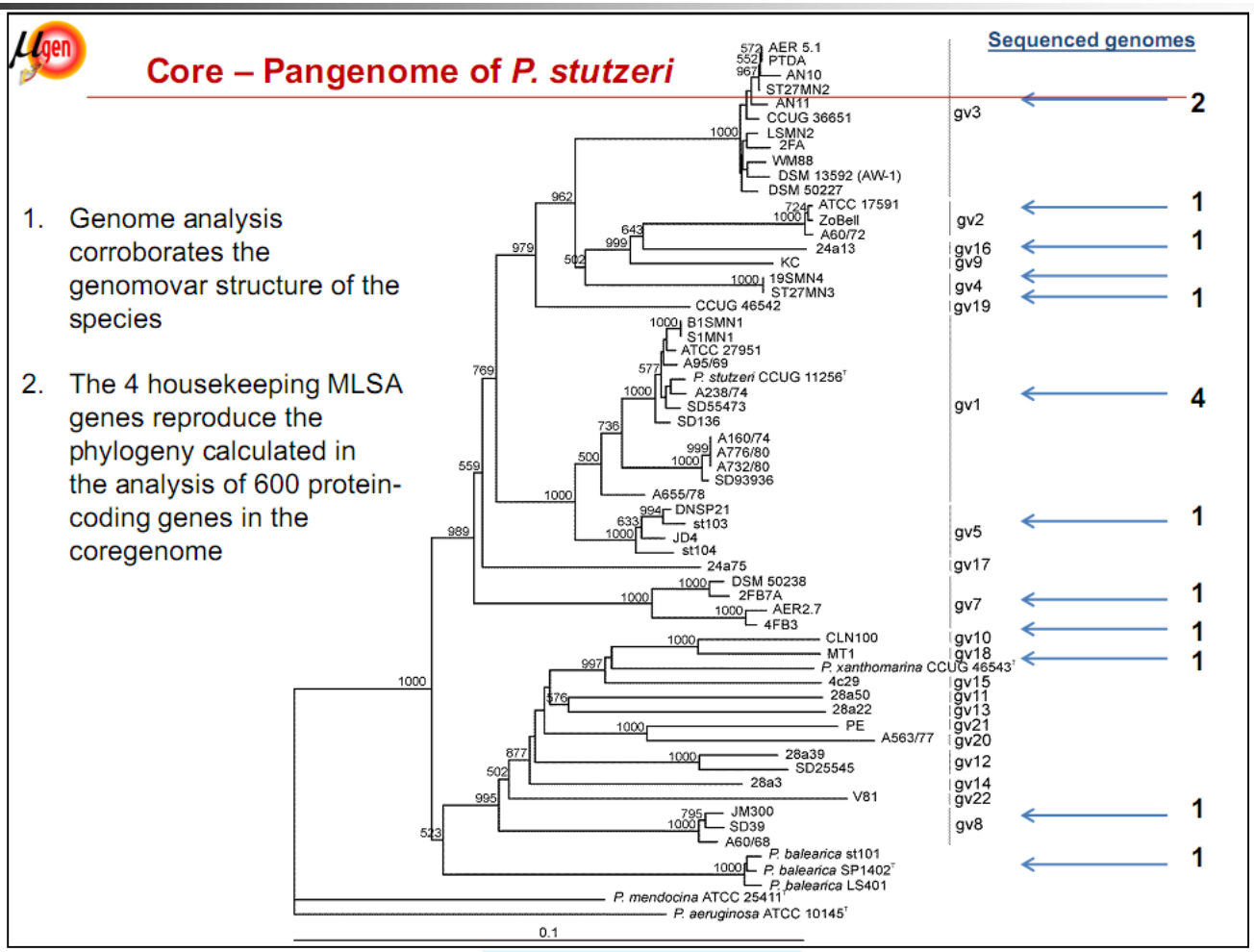
Pan-genome and Core genome
Pan-genome species



Microgen,2013

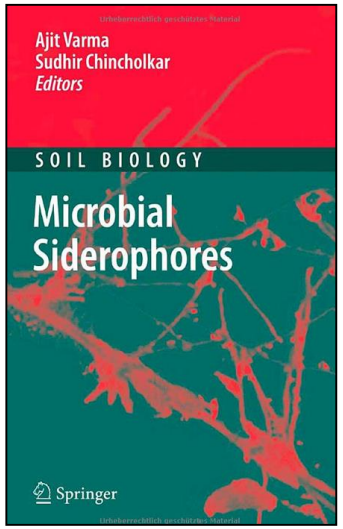
Pan-genome

Pan-genome of *Pseudomonas stutzeri*



Microgen,2013

Siderotyping



Siderotyping is a powerful technique for discriminating species within the genus *Pseudomonas*

Bacterial typing method
Siderotyping

- Siderotyping (short for siderophore-typing) has been recently proposed as a rapid and efficient bacterial typing method for identification at the species level of
- Fluorescent, and
- non-fluorescent *Pseudomonas*.

Crude Siderophore crystals

Ali and Vidhale,2013;..

Siderotyping
Phenotypic characterization

- A precise taxonomic allocation which parallels the polyphasic taxonomy data can be simply achieved by siderotyping.
- Siderotyping fulfill the demand for fast and reliable species identification.
- It may be useful in identification and systematics of numerous siderophore-producing microbial groups other than fluorescent *Pseudomonas* species including:
- *P. corrugata*,
- *Burkhoderia graminis*, etc.
- Siderotyping also proved to be an interesting way to phenotypically discriminate some new species.
- e.g. in characterization of *Pseudomonas tolaasii* and *"Pseudomonas reactans"*.

The element iron
Fe Need

- Life without iron is impossible.
- Up to 10^5 Fe-ions are typically required in key metabolic processes of a single bacterial cell.
- Though iron is the fourth most abundant element in the Earth's crust, it is present under aerobic conditions at nearly neutral pH in the form of extremely insoluble minerals like hematite, goethite, and pyrite or as polymeric oxidehydrates, carbonates, and silicates.
- Thus, iron (III) at physiological pH is high insolubile.

F. Matzanke;..

The element iron

Fe Need

- In fact iron exists in nature either as:

1. Ferrous (Fe++), or
2. Ferric (Fe+++) ions.

- In the presence of oxygen, iron (II) is rapidly oxidized to iron(III) which precipitates as a polymeric oxyhydroxide.
- Fe(III) is the predominant form of iron in aerobic and microaerobic environments and is highly insoluble.

Fe(IV)	Fe(III)	Fe(II)	Fe(I)
hardest	most common in aerobic environments	most common in anaerobic environments	**softest**

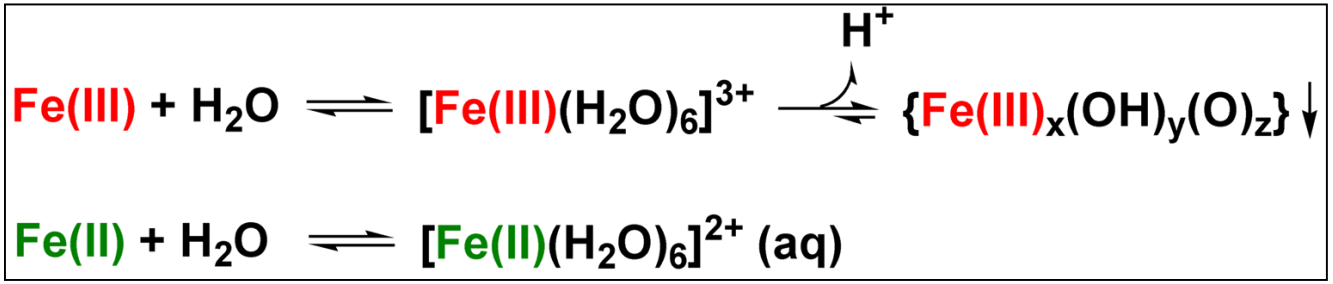
DuBois,2011;AgriInfo.in©2009

Siderophores

Greek: sideros + phore = iron carrier

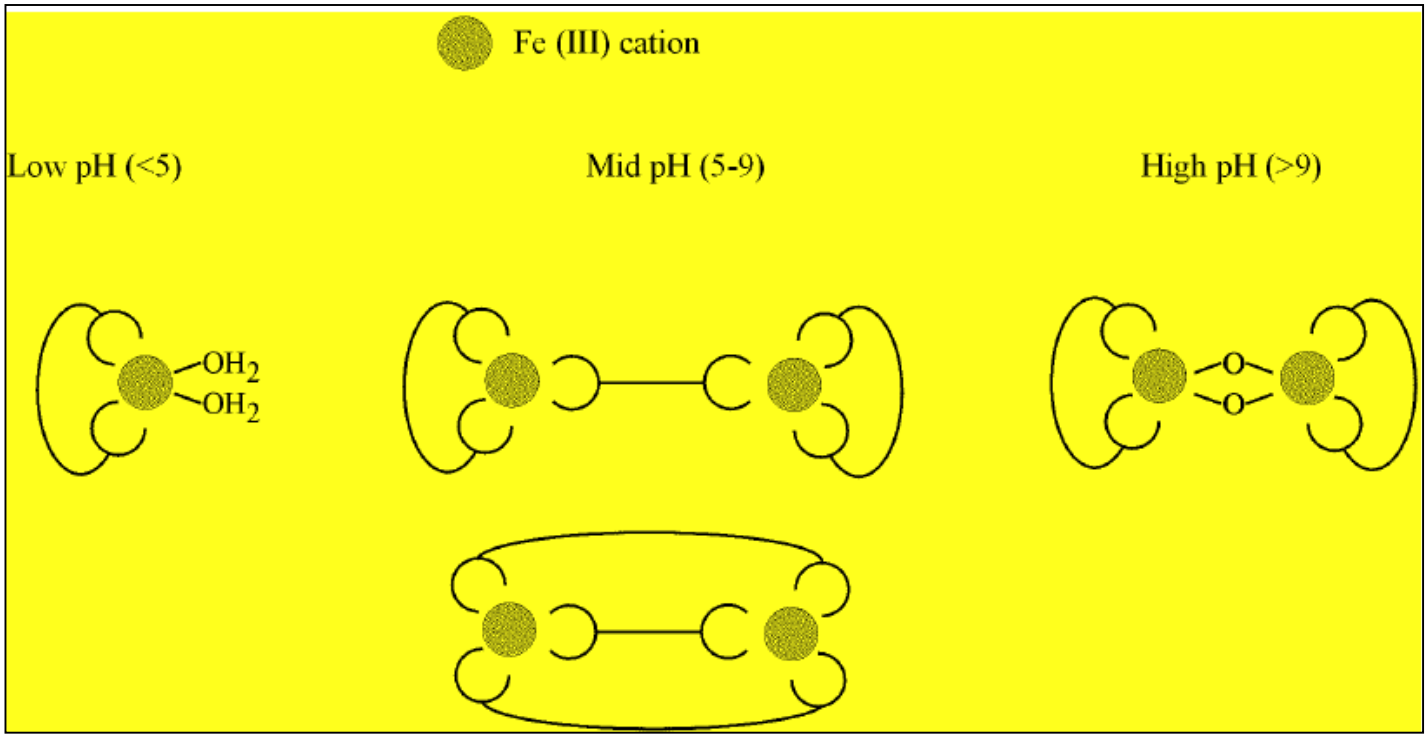
Iron needs of microorganisms

- Under aerobic conditions this essential element is present in the form of insoluble oxide hydrates [Fe_2O_3 1 nH_2O].
- Most life requires [Fe(III)] ~ mM or 10^{-6} M.
- The concentration of Fe(III) is far too low ($\leq 10^{-17}$ M) to sustain the microorganism (10^{-6} M).



DuBois,2011;..

The element iron
At different pH levels



F. Matzanke

The element iron

Availability and Biological need

- Plants as well as microorganisms require traces of iron, manganese copper, zinc, molybdenum, calcium boron, cobalt, etc.
- Iron is always abundant in terrestrial habitats, and it is often in an unavailable form for utilization by plants and leads to the serious deficiency in plants.

Siderophores

Greek: sideros + phore = iron carrier

Iron needs of microorganisms

- Iron is required by aerobic bacteria and other living organisms for a variety of biochemical reactions in the cell. e.g.

1. Energy generation by electron transport,
2. DNA synthesis, and
3. Defense against toxic reactive oxygen species, that the element is indispensable to their survival.

- Aerobic microorganisms need sufficient iron for passive diffusion of ions into the cells and for their normal growth.

The element iron
Iron needs of microorganisms

- Microorganisms can take up iron using two different pathways:

1. The first, which takes place at iron concentrations above 1×10^{-4} M is a low affinity system based on the diffusion of iron ions through biological membranes.
2. The second is a high affinity system that has developed under iron deficiency conditions and consists of an iron chelator – a siderophore.

Jankiewicz & Kuzawińska,2009

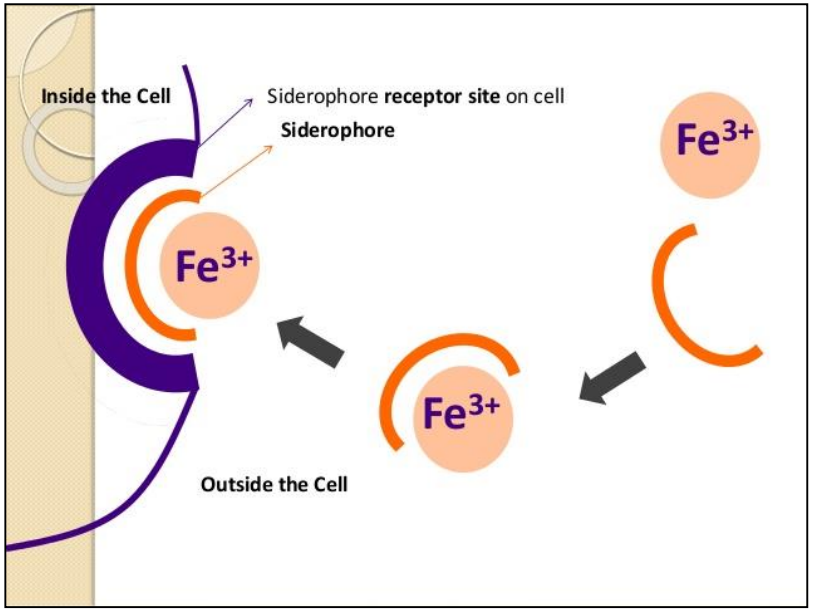
The element iron
Availability and Biological need

- Most of the aerobic microorganisms live in an environment where iron exists in the oxidized, insoluble ferric hydroxide form.
- They produce iron-binding compounds (siderophores) in order to take up ferric iron.
- Siderophores are low-molecular-weight compounds (< 1500 Da) possessing a high affinity for Fe (III).
- Siderophores can form water soluble complexes with Fe^{3+}.
- After complexation iron (III) is available in a water-soluble form, as ferri-siderophore (ferri-pyroverdine), which can be taken up by microorganisms.

Siderophores

Iron acquisition by microorganisms

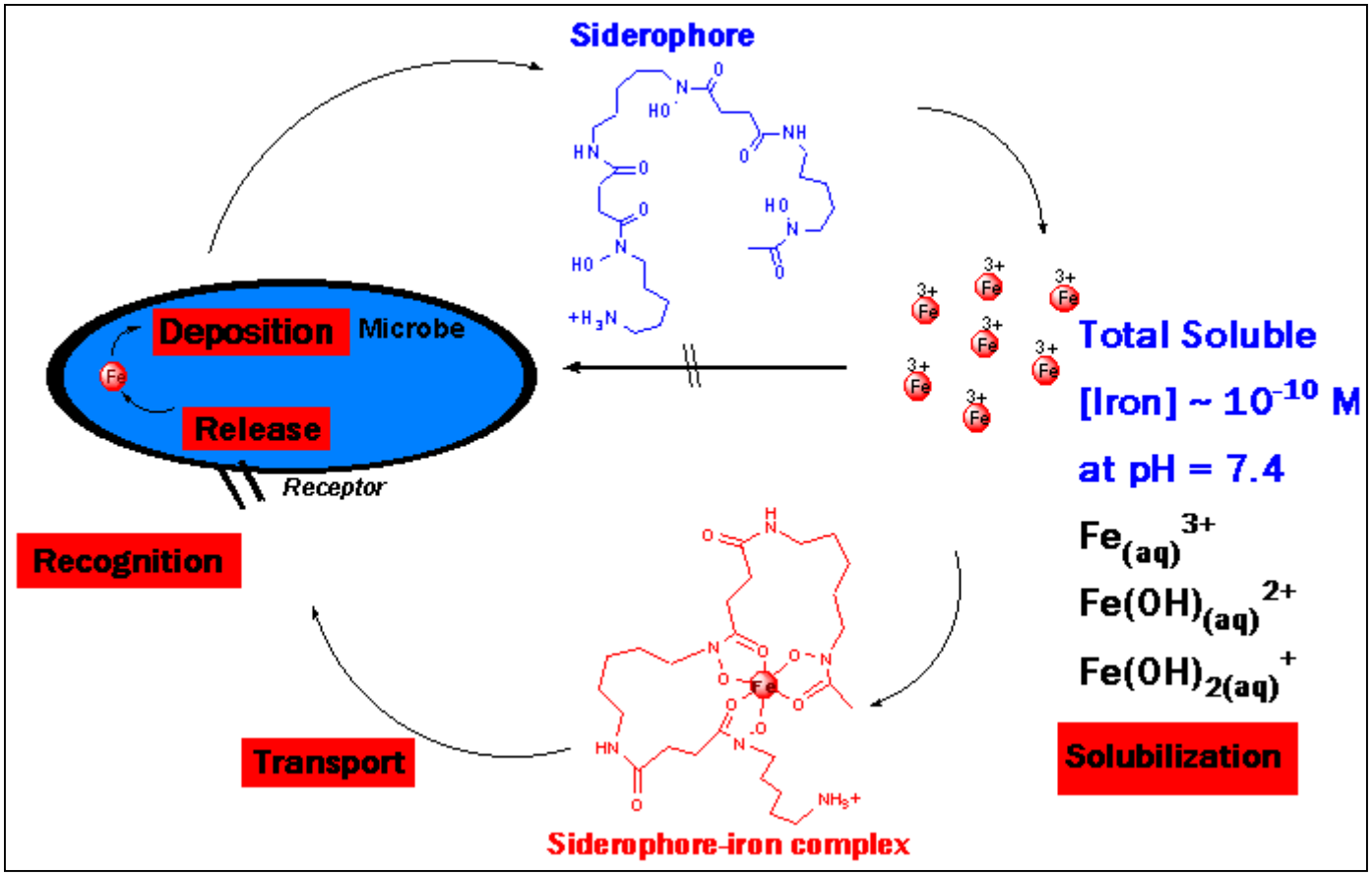
- To survive in such an environment, organisms were found to secrete iron-binding ligands called siderophores having high affinity to sequester iron from the micro-environment.
- Most microorganisms respond to low-iron stress by producing extracellular, low molecular-weight (500-1000 Daltons), iron transport compounds (or siderophones) which bind iron selectively and with great binding power.



Manju,2014;..

Siderophores

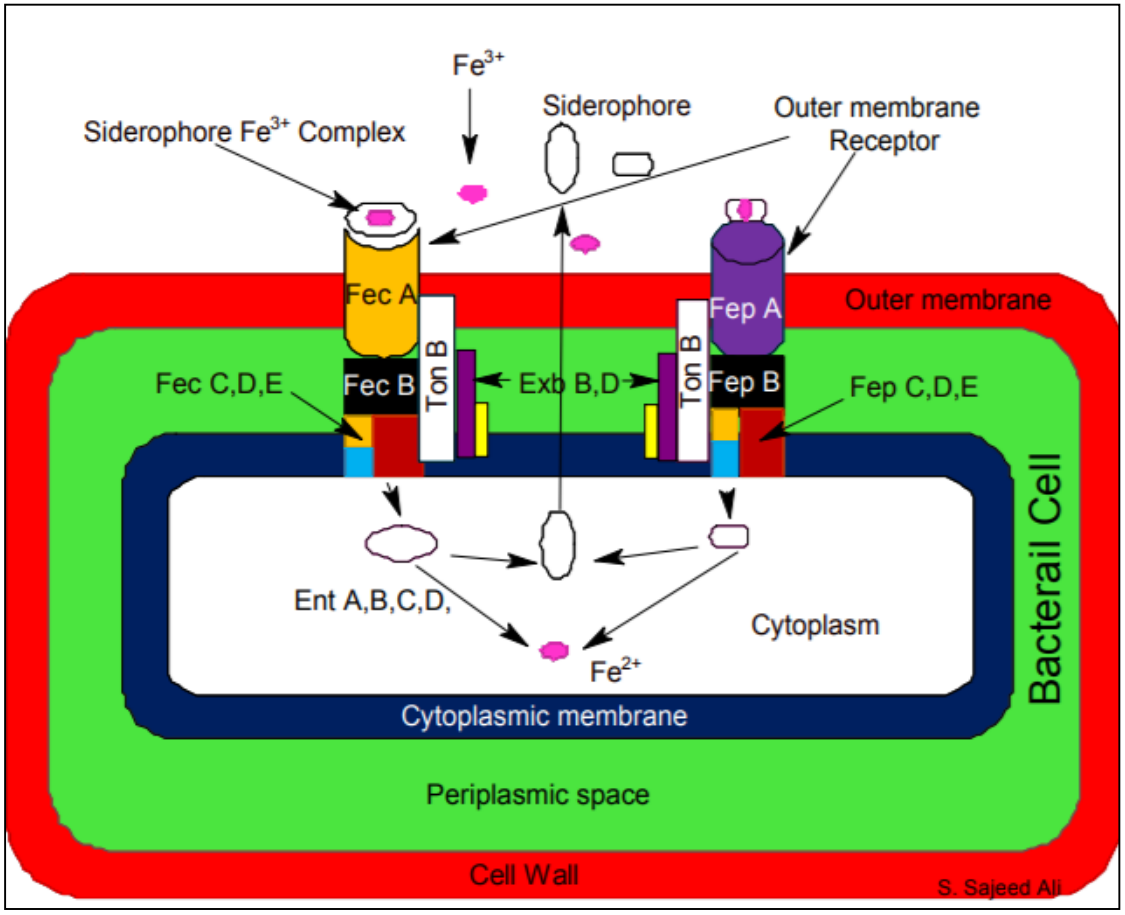
Iron acquisition by microorganisms



Siderophores

Iron acquisition by microorganisms

Mechanism of siderophore mediates iron transport in bacteria

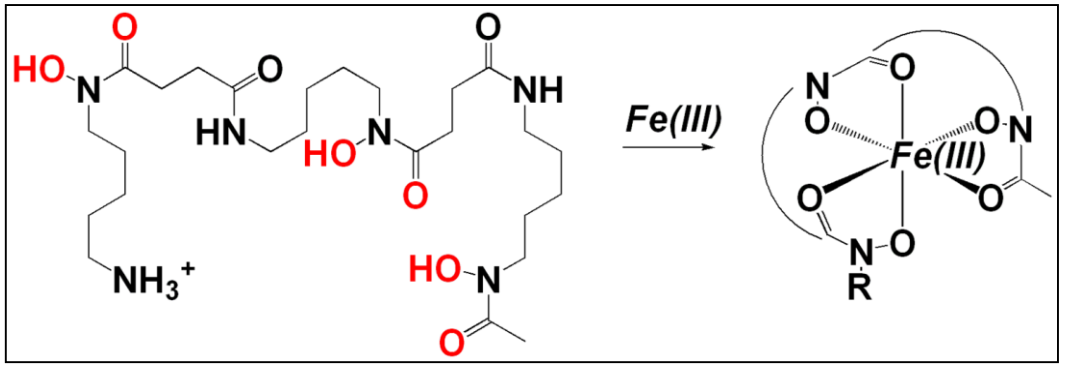


Ali and Vidhale,2013

Siderophores

Iron acquisition by microorganisms

- Siderophores are multidentate chelating ligands with high affinity for Fe (III) and often for other hard cations.
- Many bacteria, fungi, plants produce them.
- Recognize their own and sometimes siderophores from other species.



DuBois,2011

Siderophores

Iron acquisition by *Pseudomonas*

Pyoverdines

- Pyoverdine, the yellow-green, water-soluble, fluorescent pigment of the fluorescent *Pseudomonas* species, is a powerful iron (III) scavenger and an efficient iron (III) transporter.
- The majority of fluorescent pseudomonads produce complex peptidic siderophores called pyoverdines (Pvds) or pseudobactins, which are very efficient iron scavengers.
- Pyoverdine is the main siderophore produced by fluorescent *Pseudomonas* strains.
- PVDs bind Fe(III) efficiently.

Siderophores
Pyoverdines

- A tremendous variety of pyoverdines has been also observed in *Pseudomonas*, each species producing a different pyoverdine.
- Pyoverdines are siderophores containing the pigments (fluorescent).
- This variety can be used as an interesting tool to study the diversity and taxonomy of fluorescent pseudomonads.

Siderophores

Iron acquisition by *Pseudomonas*

Other siderophores different from pyoverdines

- Several other structurally divergent siderophores have also been described for this bacterial genus.
- These also having iron chelating properties. e.g.

1. Pyocyanin produced by *P. aeruginosa*,
2. Pyochelin produced by *P. aeruginosa*,
3. Pseudomonin produced by two *P. fluorescens* strains, and
4. 2,6-pyridine dithiocarboxylic acid (PDTC), by *Pseudomonas stutzeri*.

- These are usually produced in smaller amounts than pyoverdine and their affinity for iron are usually much lower.

Siderophores
The structure of pyoverdines

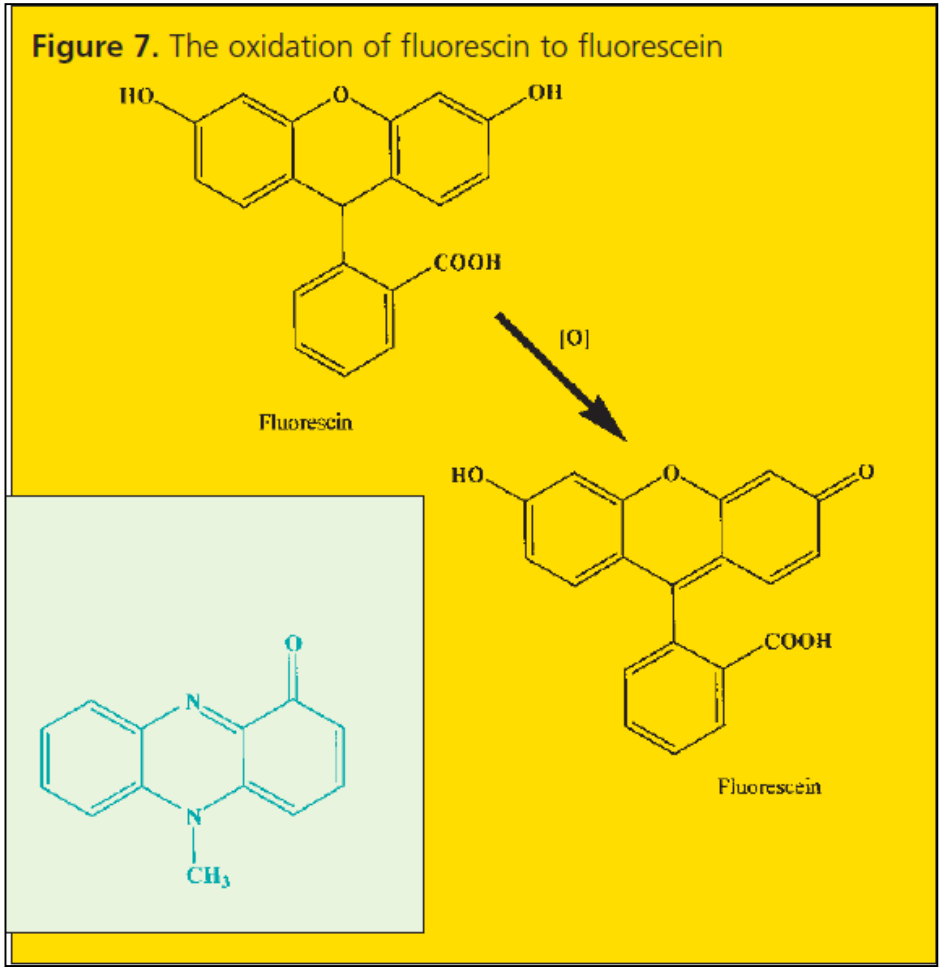
- Fluorescent *Pseudomonas* species produce Fe(III)-chelating siderophores (fluorescien) called pyoverdins (PVDs), pyofluorescein or simply fluorescein that fluorescence under UV light.
- The pyoverdines are a large family of complex siderophores which bind metal ions, especially iron, and make structure of beautifully complex molecules.

Moss,2002

Siderophores

The structure of pyoverdines

Fluorescin: A colorless, amorphous substance which is produced by the reduction of fluorescein.



Moss,2002

Siderophores

Numbers of PVDS in different bacteria

Pseudomonas

- More than 500 different siderophores have been identified from microorganisms.
- Some bacteria produce more than one type of siderophore. e.g.
- *P. fluorescens* with 12 PVDs, and
- *P. putida* with 12 PVDs.
- It is interesting to know some *Pseudomonas* spp. which are not pigmented (nonfluorescent), still synthesize siderophores. e.g.
- *P. stutzeri*, *P. pseudoalcaligenes*, *Pseudomonas corrugata*.

Siderophores

Siderophore transport in microorganisms

- Siderophore-mediated iron uptake in microorganisms is both

1. A receptor- and
2. An energy-dependent process.

- Moreover, many receptor-siderophore interactions are very specific.

Siderophores

Specific single receptor vs many nonspecific receptors

- Until recently, it was generally assumed that receptors had a strict specificity for the cognate PVD or pseudobactin.
- Recently, however, it was found that the receptor for type II *P. aeruginosa* PVD can also recognize two other PVDs from *P. fluorescens* PL7 and PL8 (Meyer *et al.*,1999;2002b).
- Evidence also exists for the presence of several extra receptors for the uptake of heterologous PVDs not produced by the bacterium.
- Some of *Pseudomonas putida* or *P. fluorescens* strains have a remarkable capacity to use a broad spectrum of heterologous PVDs, whereas their PVDs cannot be used by competitor *Pseudomonas*.

Microorganisms can use either siderophores produced by themselves or siderophores produced by other microorganisms. The cross utilization of pyoverdines by different strains depends on the structure of the peptide part of the siderophore.

Siderophores

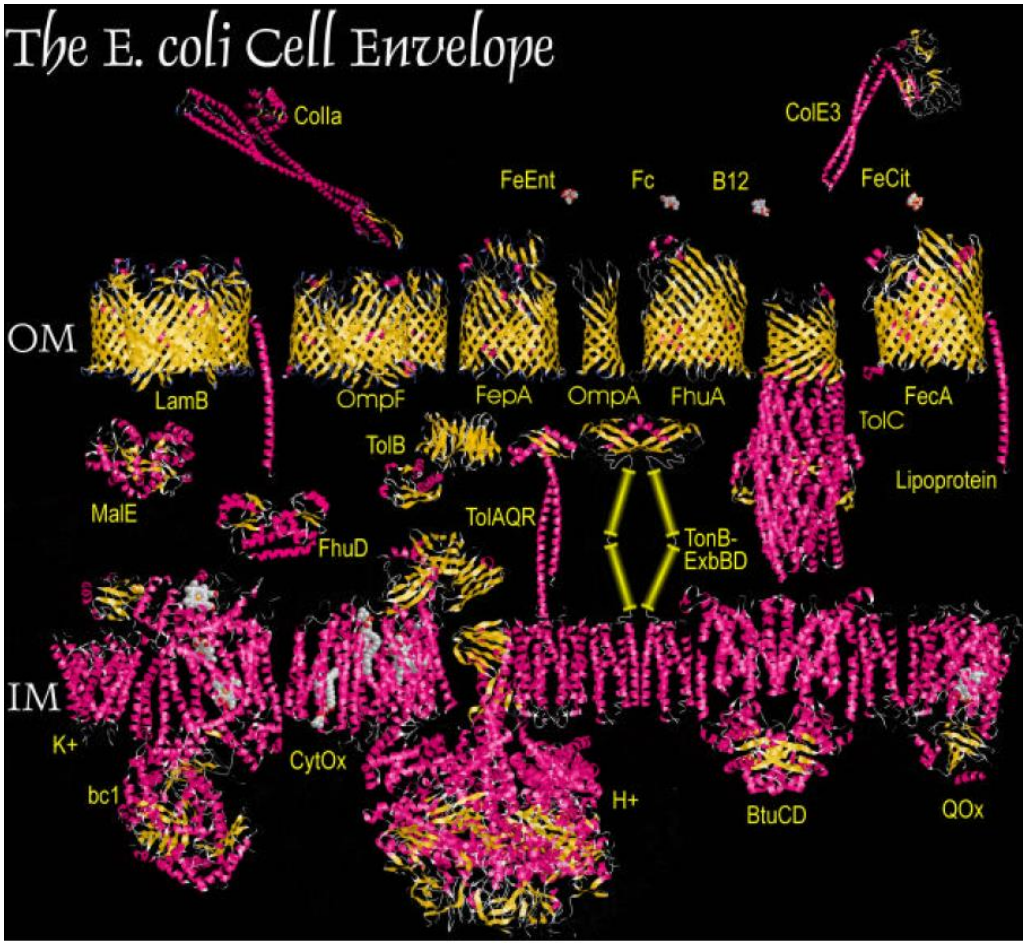
Iron uptake by soil and water bacteria

- The majority of bacteria and fungi use siderophores to solubilize and transport ferric iron (Fe^{3+}) in the environment.
- Siderophore binds Fe^{3+} ions under iron deficiency conditions.
- The complexes Fe-siderophore are identified by specific receptors (TonB proteins) on OM.
- Pseudomonads is their ability to use a large number of heterologous siderophores via different TonB-dependent receptors.

TonB protein of *E. coli* is a very close structure with *P. putida* and *P. aeruginosa*.

Proteins of the Gram-negative bacterial cell envelope

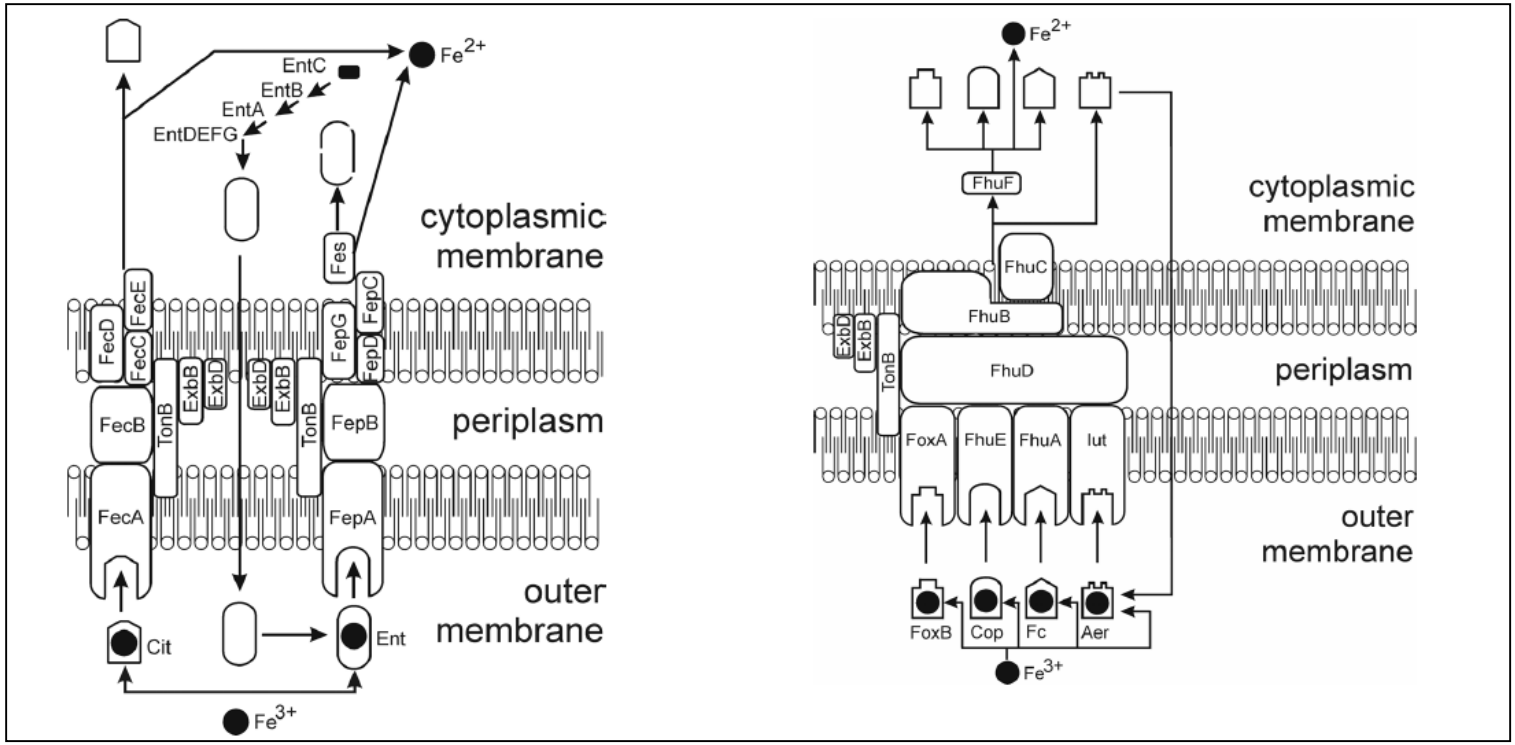
TonB proteins



Phillip E. Klebba

E. coli cell envelope

Various proteins are shown, sets of which represent specific siderophore-transport systems



Outer membrane receptors (OMR) shown here are FepA (enterobactin), IutA (aerobactin), Fec A (FeIII dicitrate), FhuA (ferrichrome), and FhuE (coprogen, FeIII rhodotorulate, and ferrioxamine B, FoxA (ferrioxamine B).

F. Matzanke

Siderophores
Iron uptake by *E. coli*

- Iron is also not freely available in humans and other mammals.
- Most iron is found (stored) intracellularly in

1. Heme proteins, and
2. Transferrin and Ferritin, the iron storage compounds.

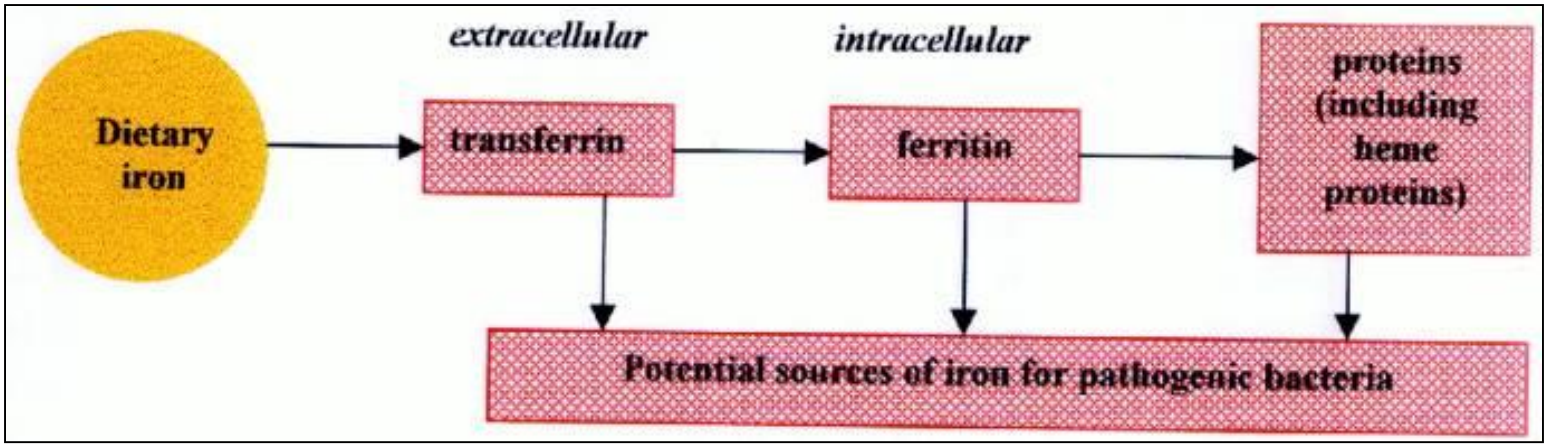
- Iron outside cells is tightly bound to proteins.
- Bacteria that grow in humans or other animals and cause infections must have a mechanism to remove iron from these proteins and use it for their own energy and growth needs.

Heme or haem is **a major source of iron,** present in haemoglobin.

McGraw-Hill Concise Encyclopedia of Environmental Science,2008

Siderophores

Iron uptake by *E. coli*



Ann Rev Microbiol. 54: 881-941,2000

Siderophore functions

Iron uptake by *E. coli*

- Outer membrane is the main target for iron uptake.
- Siderophores bind iron(III) before being recognized as ferrisiderophores by specific receptors, which function as gated porin channels in concert with the TonB protein that energizes the receptor protein.
- Iron is delivered by either lactoferrin or transferrin, heme or heme-hemophore, and homologous or heterologous siderophores to specific OMTs located in the outer membrane.
- The OMTs transport iron/iron ligand across the outer membrane and the transport is energized by the pmf of the cytoplasmic membrane via a complex formed of three proteins, TonB-ExbB-ExbD.

Siderophore functions

Iron uptake by *E. coli*

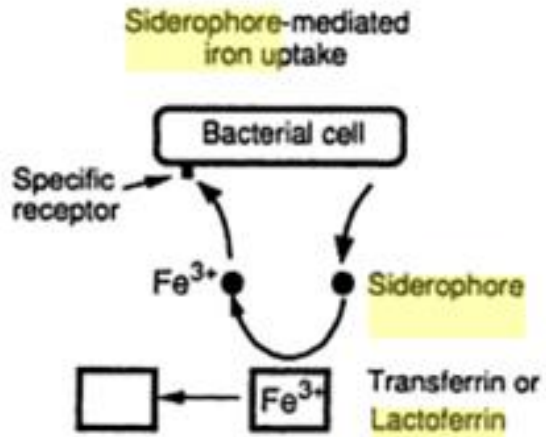
- In the periplasm, Fe^{3+}, heme, or siderophore-Fe^{3+}bind to periplasmic binding proteins (PBPs) that are specific for either Fe^{3+}, hem or ferric-siderophore.
- The PBPs deliver the iron compounds to an ABC transporter for the transport across the cytoplasmic membrane.
- The energy required for transport across the cytoplasmic membrane is provided by the hydrolysis of ATP.
- Inside the cell, the iron ion (III) is reduced, the complex disintegrates and the siderophore can be used again.

Ligand: A molecule other than an enzyme substrate that binds tightly and specifically to a macromolecule, usually a protein forming a macromolecule-ligand complex.

Siderophore functions

Iron uptake by Gram-negative bacteria

Different ways by which bacteria obtain iron from transferrins



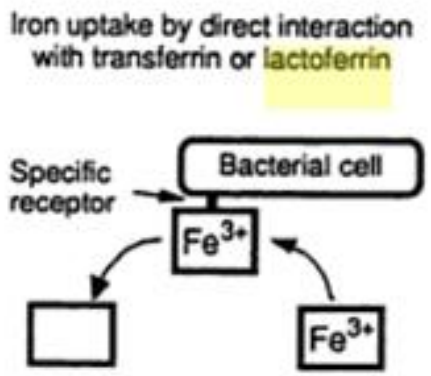


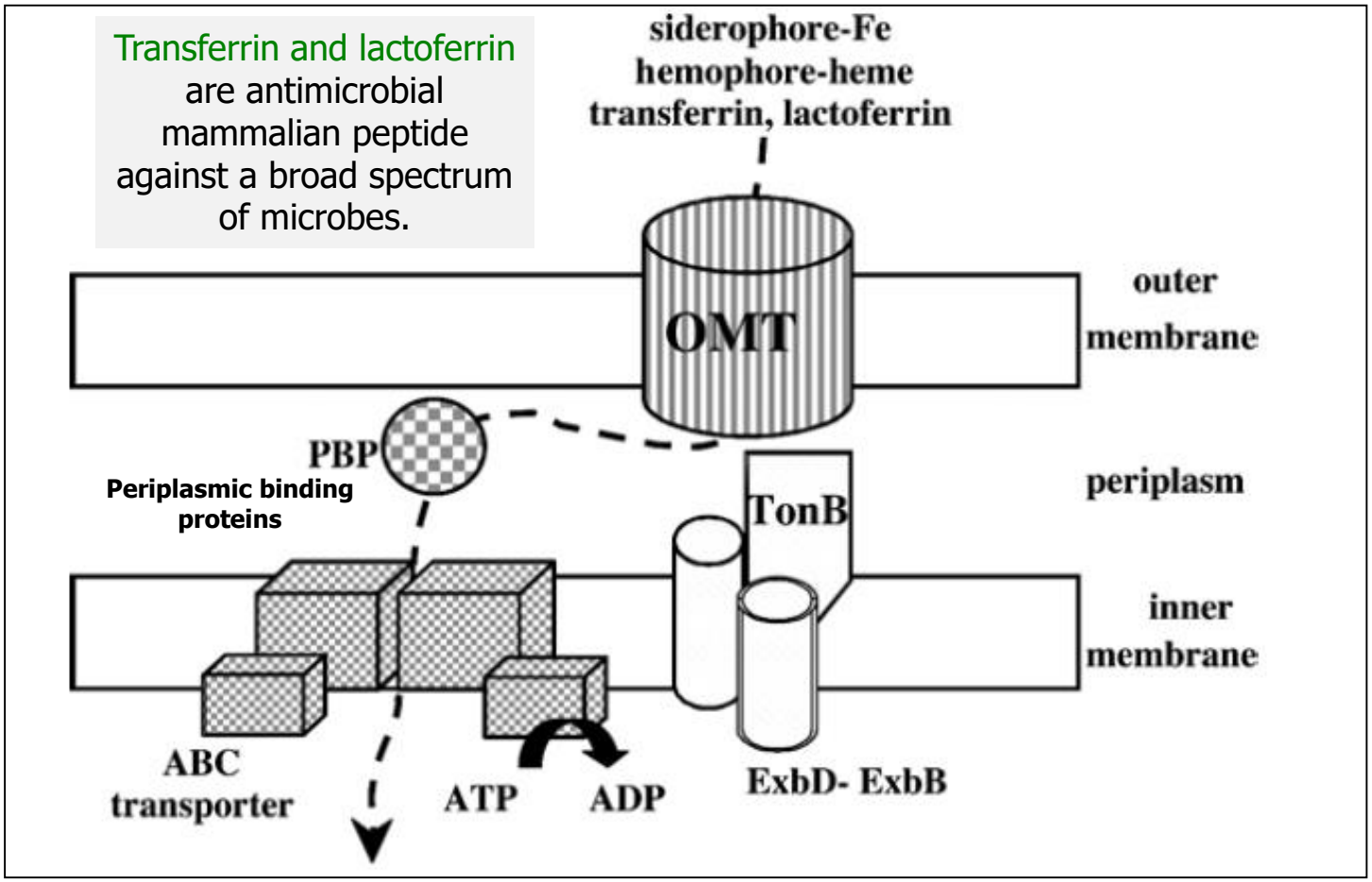
Fig. 1. Schematic representation of different ways by which bacteria obtain iron from transferrins. Modified from Otto et al. (1992).

Otto *et al.*,1992

Siderophore functions

Iron uptake by *E. coli*

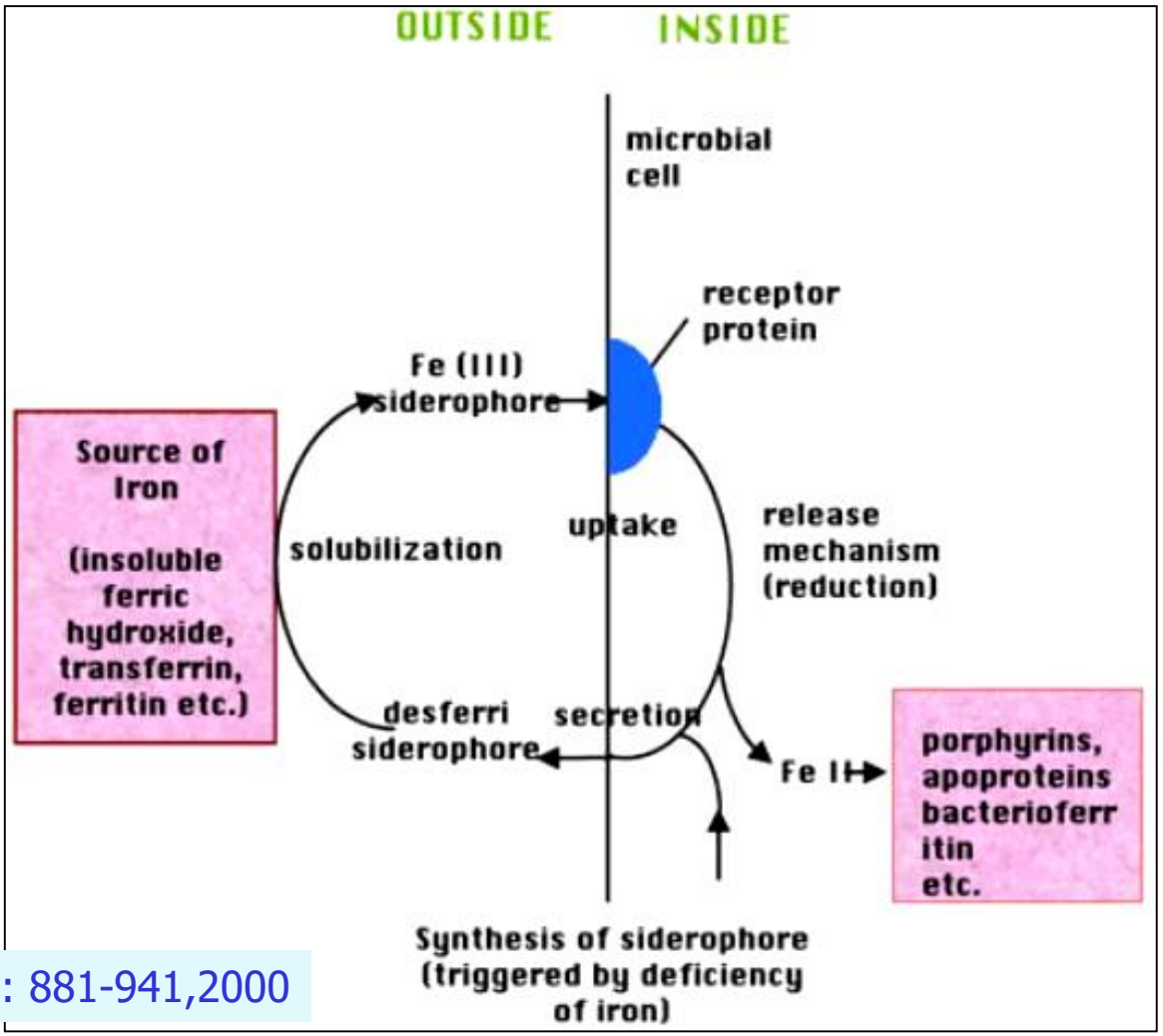
Transferrin and lactoferrin are antimicrobial mammalian peptide against a broad spectrum of microbes.



Schalk,2006

Siderophore functions

Iron uptake by *E. coli*



Ann Rev Microbiol. 54: 881-941,2000

Siderophore functions

Iron uptake by *Dickeya dadantii* and *Erwinia amylovora*

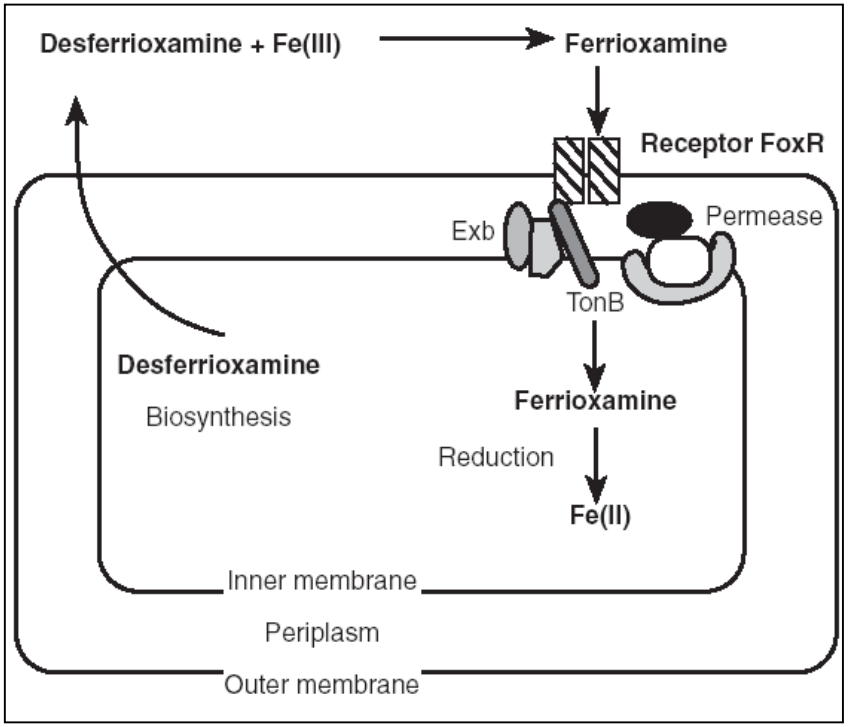
- Besides the production and utilization of siderophores, *P. atrosepticum* and *D. datantii* have the capacity to use other iron sources.
- Indeed, *Dickeya dadantii* (*Erwinia chrysanthemi*) and *Erwinia amylovora* are able to use haem iron, whereas only *P. atrosepticum* can transport the ferric citrate complex and only *D. dadantii* can acquire ferrous iron.
- These different modes of iron capture indicate that these species have to cope with various environmental and ecological conditions during their pathogenic life cycle.

Cornelis and Andrews,2010

Desferrioxamine E

Role in pathogenicity of *E. amylovora*

- Desferrioxamine E is the major siderophore of *E. amylovora*.
- Compounds of the desferrioxamines (DFO family), of which DFO E appeared to be predominantly produced, were identified in various strains of *E. amylovora*.



Vanneste,2000

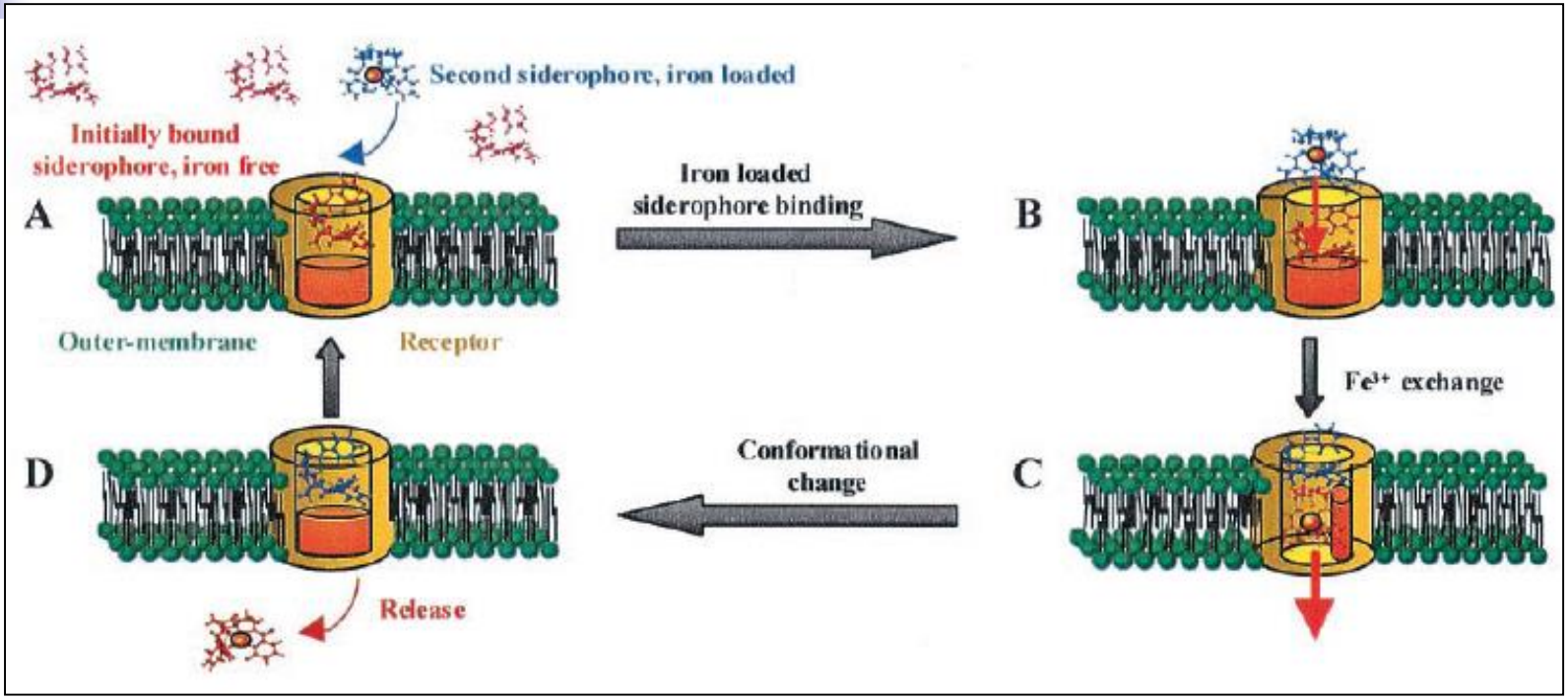
Siderophore functions

The other proposed siderophore iron uptake and delivery mechanism

- The ligand exchange step occurs at the cell surface and involves the exchange of iron from a ferric siderophore to an iron-free siderophore already bound to the receptor.
- The ferric-siderophore complex must cross the outer membrane and the cytoplasmic membrane before delivering iron within the cytoplasm.
- Once in the periplasmic space, the ferric-siderophore binds to its cognate periplasmic binding protein and is then actively transported across the cytoplasmic membrane by an ATP-transporter system.

Siderophore

The other proposed siderophore iron uptake and delivery mechanism



Proposed model of the siderophore shuttle iron exchange mechanism for iron transport in Gram-negative bacteria. The steps involved in binding to the receptor, ligand exchange, and iron translocation.

Stintzi *et al.*,2000

Siderophore synthesis
By non-ribosomal peptide synthetases

- Biosynthesis of siderophores is executed on a cellular level by a set of enzymes specific for the respective siderophore.
- These enzymes are known as non-ribosomal peptide synthetases (NRPSs).
- The corresponding genes are located on the chromosome or on a plasmid.
- NRPSs are large multimodular enzymes, which perform non-ribosomal peptide synthesis.

Cornelis and Matthijs 2002;F. Matzanke

Siderophore production
Influence of sigma factors

- In fluorescent pseudomonads, production of PVDs and their cognate receptors is influenced by PvdS an extracytoplasmic sigma factor (ECF).
- PvdS is needed for the transcription of several PVD biosynthesis genes, including NRPSs and other peptide modification genes.
- The alternative sigma factor, RpoS, has been also shown negatively to affect the production of PVD in *P. aeruginosa*.
- In *P. putida*, in contrast, RpoS was found to have no effect on siderophore production.

Cornelis and Matthijs 2002

Siderophore production
Influence of quorum-sensing

- Another factor that seems to influence PVD production is the quorum-sensing system.
- In *P. aeruginosa*, it was found that mutants in the lasI/lasR quorum-sensing genes produced half the amount of PVD compared with the wild type.
- Whiteley *et al*.,1999 found that the *pvdI* gene, encoding one of the NRPSs (non-ribosomal peptide synthetases) for the biosynthesis of the peptide chain of PVD, is controlled by quorum-sensing autoinducers.

Cornelis and Matthijs,2002

Siderophores

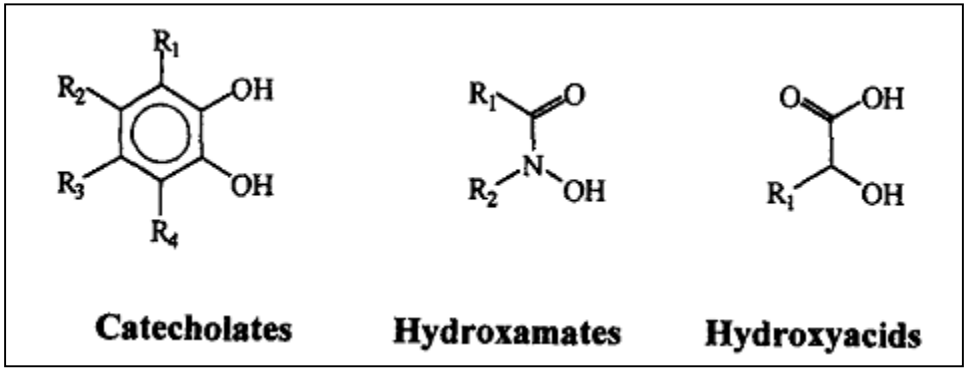
Four types of siderophores in bacteria

- Bacteria produce four types of siderophores:

1. hydroxymate (OH-group).e.g. aerobactin in *E. coli*,
2. catecholate (phenolates). E.g. enterobactin from *E. coli*.

- Salicylate (salicylic acid, a siderophore of *Pseudomonas fluorescens* CHAO), and

1. carboxylate (e.g. derivatives of citric acid). Citric acid can also act as a siderophore.



Siderophore types

Examples of siderophores produced by fluorescent pseudomonads and non-fluorescent pseudomonads

Siderophore	Organism
Fluorescent pseudomonads	
Pyoverdine	*P. fluorescens*
Pyochelin	*P. aeruginosa* PAO1
Pseudobactin	*P. putida* B10
Non-fluorescent pseudomonads	
Desferriferrioxamine E	*P. stutzeri*
Corrugatin	*P. corrugata*

Desferrioxamine E is produced mainly by the *Erwinia amylovora* but also by *Streptomyces coelicolor* and *P. stutzeri*.

Wikipedia,2011;Chie,2014

Siderophore types

Hydroxamate siderophores produced by some other bacteria and fungus

- Hydroxamate group-bearing siderophores are mainly synthesized by fungi and Gram-positive filament-forming bacteria (streptomycetes).
- Actinomycetes produce both hydroxymate and salicylate types of siderophores.

Siderophore	Organism
Ferrichrome	*Ustilago sphaerogena*
Desferrioxamine B (Deferoxamine)	*Streptomyces pilosus* *Streptomyces coelicolor*
Desferrioxamine E	*Streptomyces coelicolor*
Fusarinine C	*Fusarium roseum*
Ornibactin	*Burkholderia cepacia*

Wikipedia,2011;..

Siderophore types

Structure of some microbial siderophores

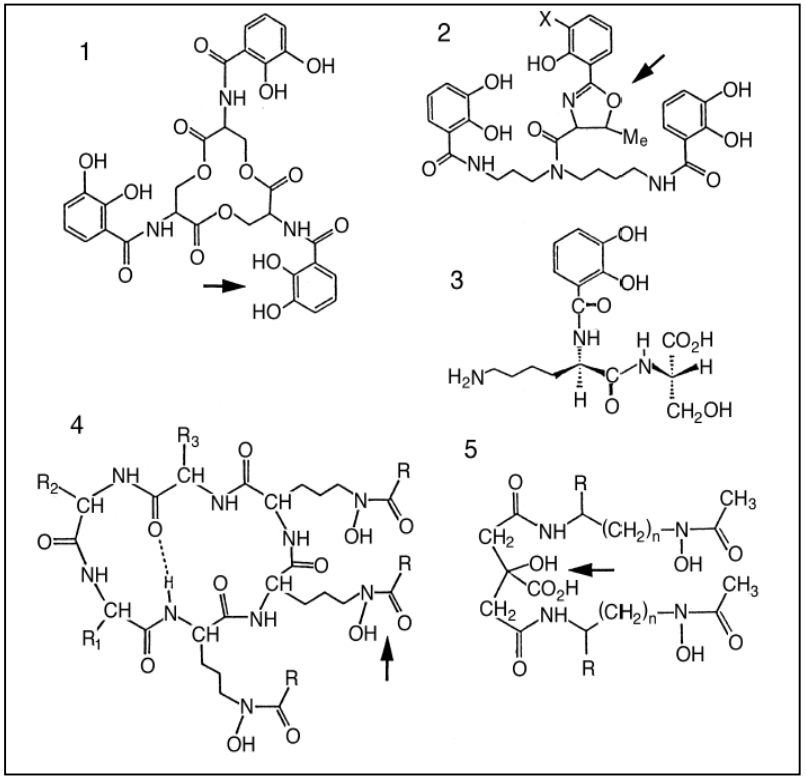
- Siderophores containing catechol ligands:

1. Enterobactin or enterochelin by *E. coli*.
2. Chrysobactin by *Dickeya chrysanthemi*.

- Hydroxamate siderophores:

4. Desferrichrome E by *E. amylovora*.
5. Agrobactin by *A. tumefaciens*.

- Iron (III) ligands are shown by arrows.



Catechol is the conjugate acid of a chelating agent. Basic solutions of catechol react with iron(III) to give the red $[Fe(C_6H_4O_2)_3]^{3-}$.

Vanneste,2000

Pyoverdine structure

Pyoverdine structural parts

Pseudomonas

- A molecule of pyoverdine consists of three parts:

1. A conserved dihydroxyquinoline chromophore;
2. A peptide chain with 6 to 12 amino acids, depending on the producing strain. The peptide chains vary among strains and species.
3. A side chain, usually a dicarboxylic acid or its amide.

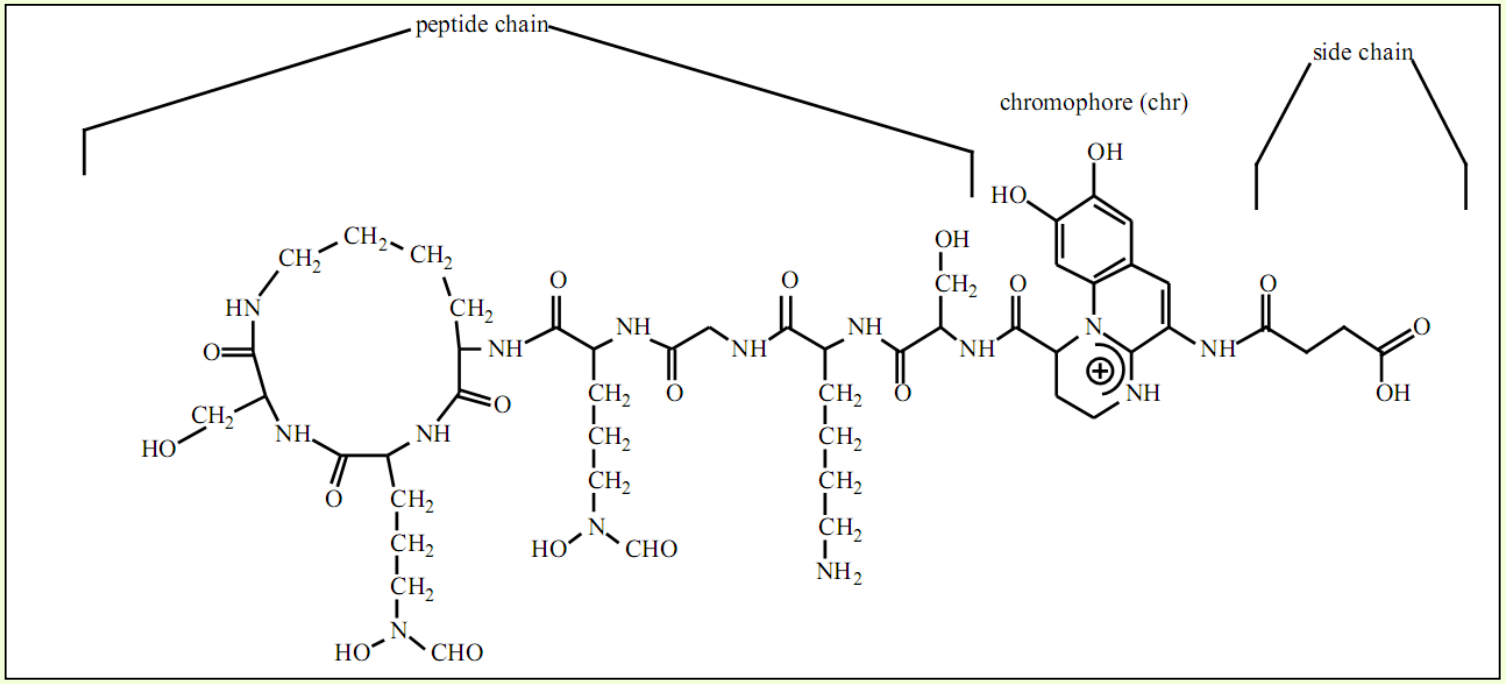
- Many structures of PVDs, differing mainly in their peptide chain(amino acid sequence of the peptide chain).
- These structural differences are used in the siderotyping of bacteria belonging to the genus *Pseudomonas*.

Jankiewicz & Kuzawińska,2009

Pyoverdine structure

Pyoverdine from *P. fluorescens*

- Dihydroxyquinoline-based chromophore that is responsible for iron binding and an N-terminal side chain bound to the chromophore.

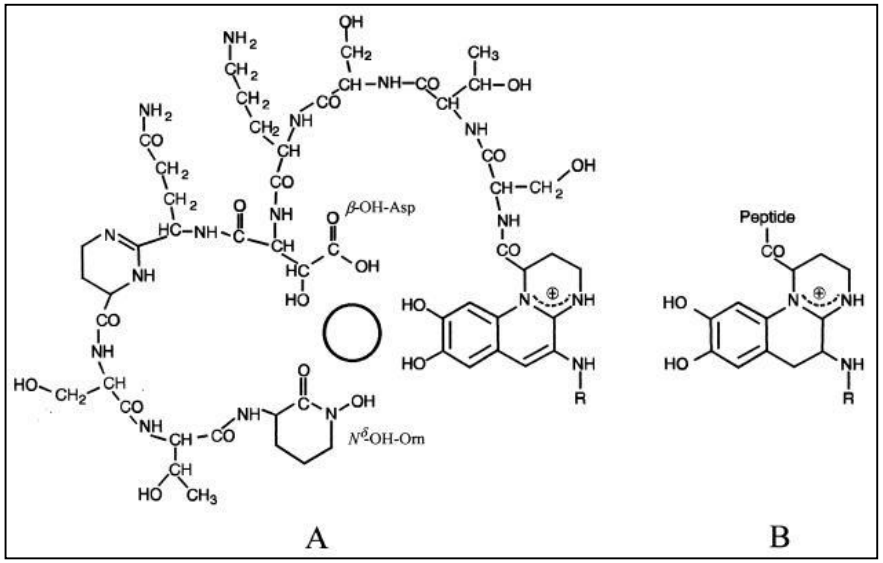


Fuchs *et al.*,2001; Drake and Gulick,2011

Pyoverdine structure

Example of a pyoverdin (A) and the cognate dihydropyoverdin (B) produced by a *P. putida* strain

- The circle in the center of the molecule (A) indicates the place where the iron is chelated.
- Both hydroxyl groups on the chromophore and FHO (N-formyl-N-hydroxy-ornithine) side chain oxygens form interactions with iron.



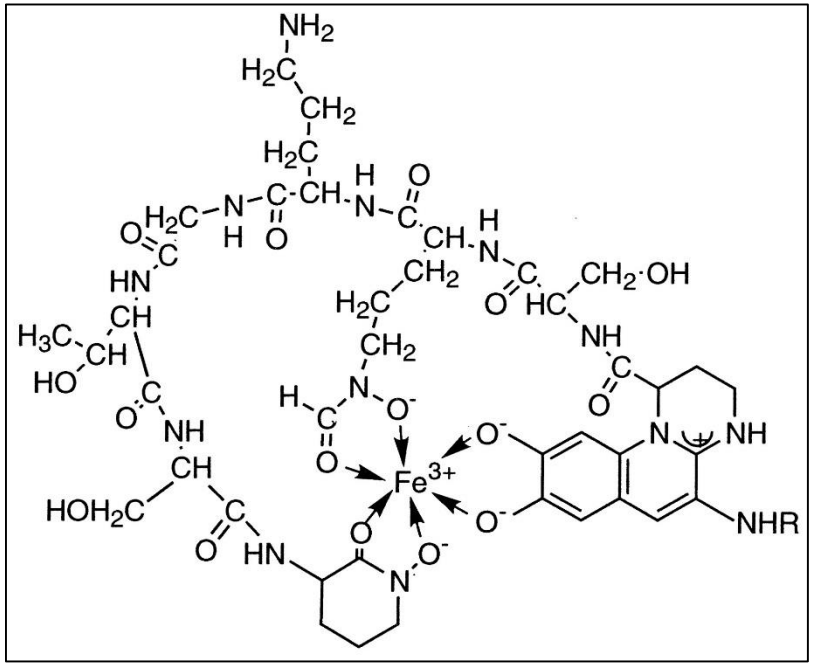
Bultreys *et al.*,2001;Peek *et al.*,2012

Pyoverdine structure

Structure of *P. aeruginosa* pyoverdine

- The acyl chain (R group) could be:

1. Succinic acid,
2. Succinamide, or
3. Glutamic acid.



Meyer *et al.*,2002

Pyoverdine structure

1. Peptide chain

- Comprising 6 to 12 amino acids, depending on the producing strain.
- The sequence of the peptide chain is usually strain-specific but can vary within a species.
- The peptide chain has two functions:

1. It provides two ferric iron binding sites, and
2. Is responsible for identifying the complex of siderophore and iron (ferri-pyoverdine molecule) on the surface of the producing cell by specific receptors located in the outer membrane.

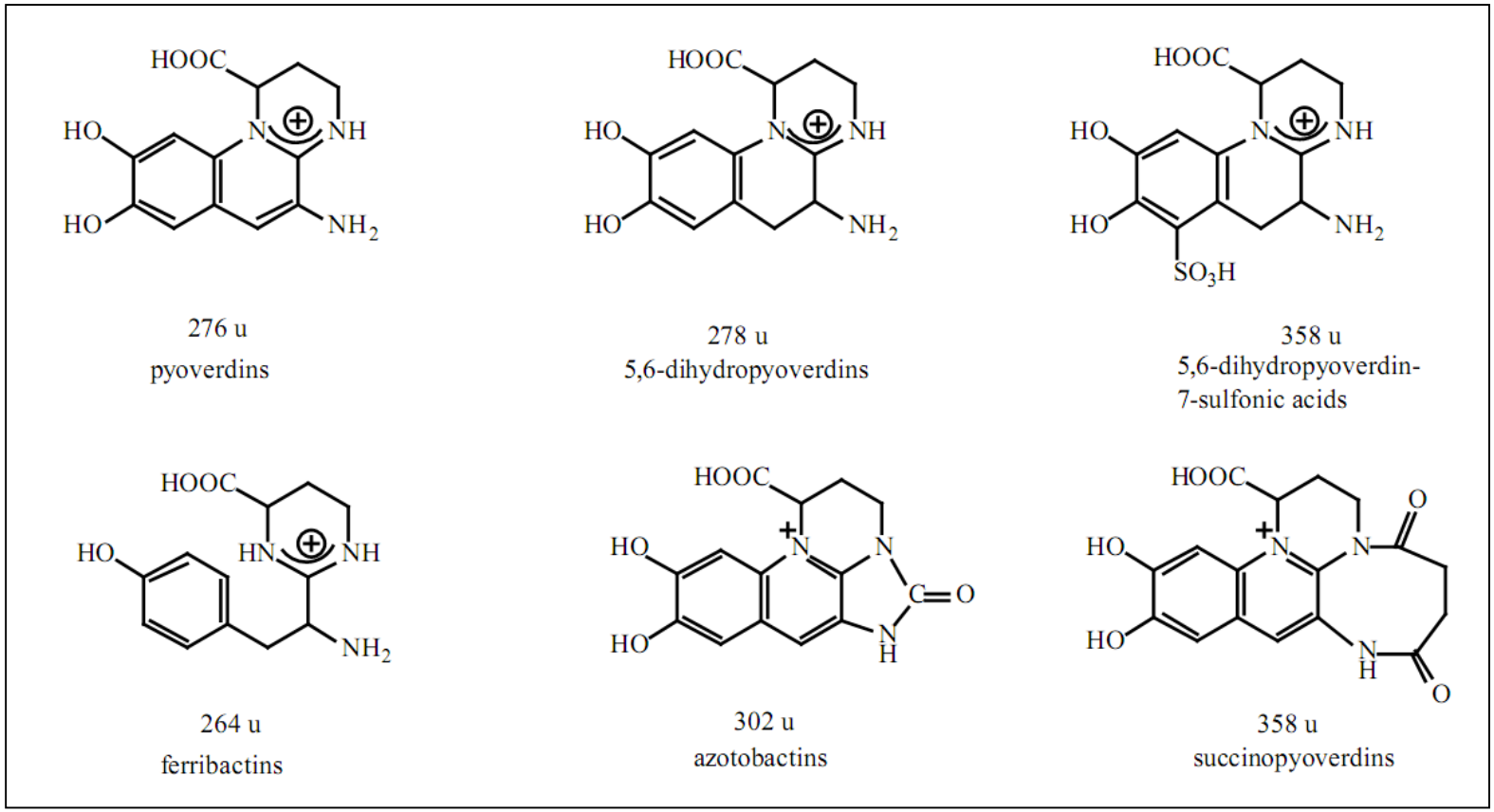
Pyoverdine structure
2. Quinoline chromophore

- Imparts the color and fluorescence to the molecule.
- The chromophore is identical for all pyoverdines and its catechol unit is one of the binding sites for iron(III).
- Heterogeneity of pyoverdines is due to the of the chromophore and side chain part.
- Presence of different chromophors are different in:

1. color,
2. Fluorescence, and
3. UV/Vis-spectra.

Pyoverdine structure

Chromophore types that are produced along with the one of pyoverdine



Fuchs *et al.*,2001

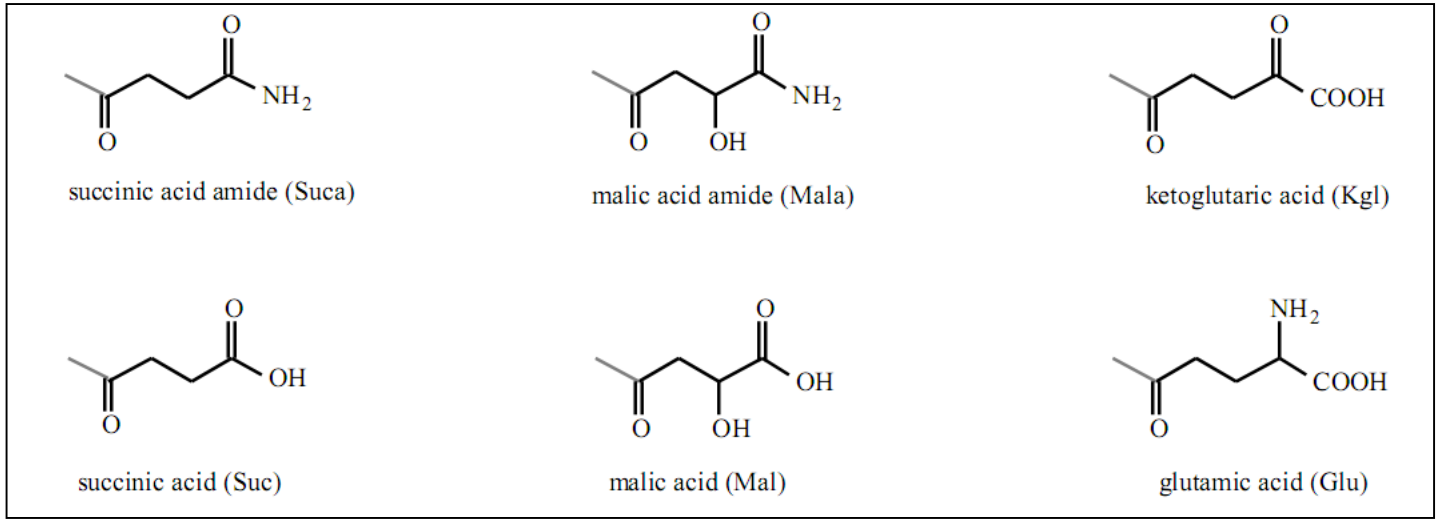
Pyoverdine structure
3. Acyl side chain

- The amino group of the chromophore is bound to a dicarboxylic acid side chain or its amide.
- The pyoverdines of the same bacterial strain differ only in the structure of the acyl side chain.
- The appearance of different side chains leads to different pI-values (useful for characterizing the pyoverdine bands) and allows the separation of the compounds through chromatography and electrophoresis.

Pseudomonas	**Peptide chain identical with**	**Side chain**
P. fluorescens	*P. fluorescens* 13525	ketoglutaric acid (Kgl)
P. fluorescens	*P. fluorescens* 13525	succinic acid amide (Suca)
Pseudomonas sp.	*P. fluorescens* 13525	succinic acid (Suc)

Pyoverdine structure

Acyl side chains of pyoverdine chromophores and related compounds



Fuchs *et al.*,2001

Pyoverdine structure

Peptide chain contains common and uncommon amino acids

- Amino Acids of Pyoverdines (amino acids involved in the Fe^{3+}-complexation with chelating functionalities are printed in bold).
- Siderovars: Regroup strains that produce pyoverdins with the same peptide chain.

Amino acids of pyoverdines	
Proteinogenic amino acids	**Unusual (non-proteinogenic) amino acids**
Ala	2,4-diaminobutyric acid (Dab)
Asn	**β-hydroxy-Asp**
Asp	**β-hydroxy-His**
Arg	***cyclo*-N-hydroxy-Orn**
Gln	**N-formyl-N-hydroxy-Orn**
Glu	**N-acetyl-N-hydroxy-Orn**
Gly	**N-hydroxybutyryl-N-hydroxy-Orn**
Lys	Orn
Ser	
Thr	
Val[3]	

Fuchs *et al.*,2001

Siderotyping methods

- The purified pyoverdins were analysed by the following microbiological and bioanalytical methods:

1. Growth stimulation (microbiological methods (visual tests) in order to evaluate the biological activity of pyoverdine against selected plant pathogens);
2. Iron uptake (uptake specificity; cross uptake);
3. PVD isoelectric focusing (IEF patterns);
4. PCR protocol for detection of siderophore
5. Immunoblotting (serological methods);
6. Amino acid analysis;
7. Mass spectrometry along with chemical degradation and synthesis.

Microbiological methods

Detection of siderophores

1. Visual tests

- The CAS (Chrome Azurol S) solid and liquid media were prepared and stored at 4°C for several months.

1. Petri dishes assay
2. Glass tubes assay

- Since this media is very sensitive to variations in pH or $FeCl_3$concentrations, it may be useful to perform a quality control test prior to using the plates in the lab.

Bultreys and Gheysen,2008;Louden *et al.*,2011

CAS agar preparation
Schwyn and Neilands,1987

- The following is a detailed, step-by-step procedure taken from Schwyn and Neilands (1987). Preparing the CAS agar can be a tedious and difficult process. By providing a detailed protocol for doing this, we hope to reduce the difficulty.
- Clean all glassware with 6M HCl to remove any trace elements, then rinse with ddH_2O.
- **A. Blue Dye:**
- a. Solution 1:
- i. Dissolve 0.06 g of CAS (Fluka Chemicals) in 50 ml of ddH_2O.
- b. Solution 2:
- i. Dissolve 0.0027 g of $FeCl_3$-6 H_2O in 10 ml of 10 mM HCl.
- c. Solution 3:
- i. Dissolve 0.073 g of HDTMA in 40 ml of ddH_2O.
- d. Mix Solution 1 with 9 ml of Solution 2. Then mix with Solution 3. Solution should now be a blue color. Autoclave and store in a plastic container/bottle.
- **B. Mixture solution:**
- a. Minimal Media 9 (MM9) Salt Solution Stock
- i. Dissolve 15 g KH_2PO_4, 25 g NaCl, and 50 g NH_4Cl in 500 ml of ddH_2O.
- b. 20% Glucose Stock
- i. Dissolve 20 g glucose in 100 ml of ddH_2O.
- c. NaOH Stock

CAS agar preparation

Schwyn and Neilands,1987

Continued

- i. Dissolve 25 g of NaOH in 150 ml ddH_2O; pH should be ~12.
- d. Casamino Acid Solution
- i. Dissolve 3 g of Casamino acid in 27 ml of ddH_2O.
- ii. Extract with 3% 8-hydroxyquinoline in chloroform to remove any trace iron.
- iii. Filter sterilize.
- **C. CAS agar Preparation:**
- a. Add 100 ml of MM9 salt solution to 750 ml of ddH_2O.
- b. Dissolve 32.24 g piperazine-N,N′-bis(2- ethanesulfonic acid) PIPES.
- i. PIPES will not dissolve below pH of 5. Bring pH up to 6 and slowly add PIPES while stirring. The pH will drop as PIPES dissolves. While stirring, slowly bring the pH up to 6.8. Do not exceed 6.8 as this will turn the solution green.
- c. Add 15 g Bacto agar.
- d. Autoclave and cool to 50°C.
- e. Add 30 ml of sterile Casamino acid solution and 10 ml of sterile 20% glucose solution to MM9/ PIPES mixture.
- f. Slowly add 100 ml of Blue Dye solution along the glass wall with enough agitation to mix thoroughly.
- g. Aseptically pour plates.

Louden *et al.*,2011

CAS liquid/agar medium
With original and modified compositions

1. CAS-liquid medium (MM9):

- Tris buffer, casamino acids 0.3%, sucrose 0.2%,thiamine·HCl 0.0002%, and succinate 0.2%, w/v).

2. Modified CAS-agar medium:

- 200 ml of blue agar with 20 ml resultant dark blue liquid, 150 ml double distilled-deionized water, 20 ml 10× MM9 salts, 3 g agar, 2.4 g of a 50% (w/w) NaOH solution, 6 ml casamino acids (10%, w/v), 2 ml glucose (10%, w/v), 0.2 ml thiamine.HCl (0.2%, w/v), and 0.6 ml L-tryptophan (1%, w/v), respectively.

Tian *et al.*,2009

Detection of siderophores

Siderophore production on a CAS siderophore testing agar

Petri dishes assay

- The chelator-iron (III) complex tints the agar (Chrome Azurol S(CAS)-agar plate) with a rich blue background.
- The color change from blue to orange halo around the colony of *Pseudomonas putida* indicates the excretion of siderophore and its dimension approximates the amount of siderophore produced.

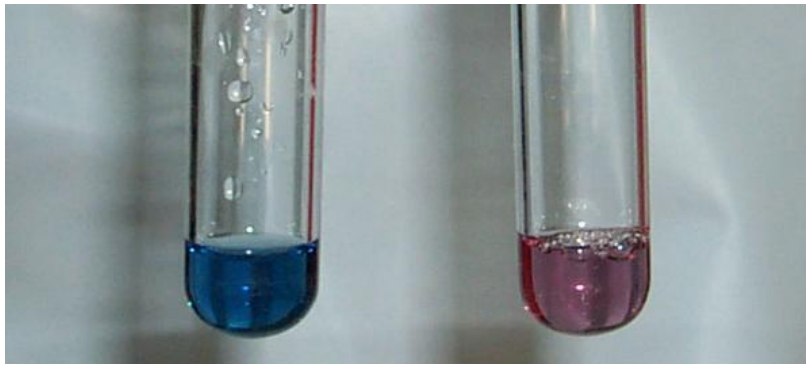
Jankiewicz & Kuzawińska,2009

Detection of siderophores

Siderophore production on a CAS siderophore testing broth

Liquid CAS solution

- The Fe(III) gives the agar a rich blue color and concentration of siderophores excreted by iron starved organisms results in a color change to orange.
- After inoculating it takes only a short time before a color change will occur usually no more that 6 hours.

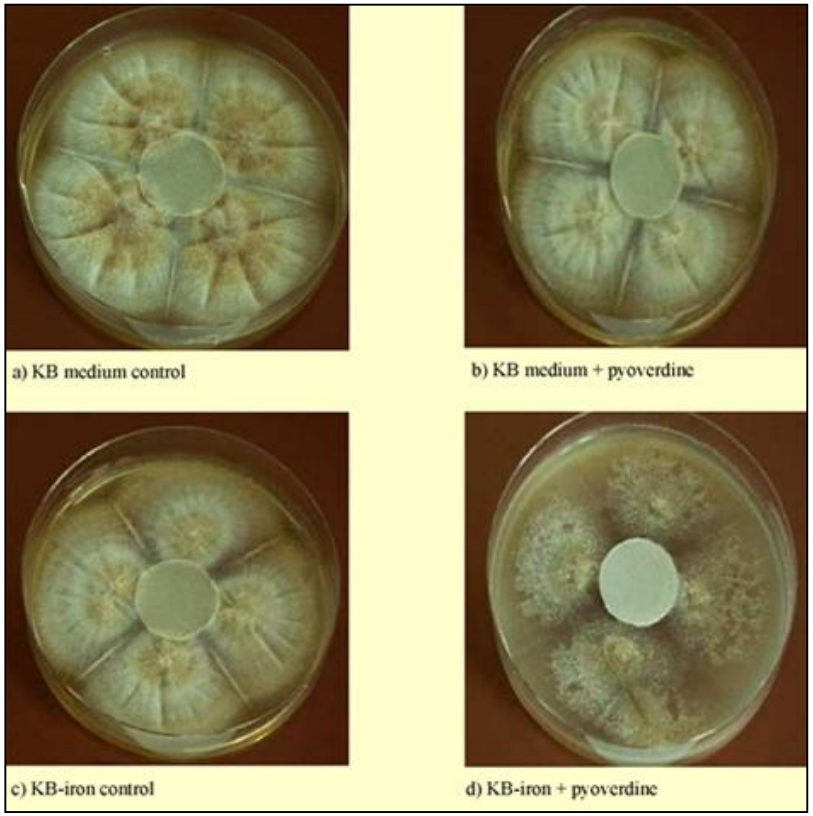


The change in color of the blue dye-chrome azurol sulphonate (CAS) assay solution to orange indicating the presence of siderophore.

Microbiological methods

Inhibition of the mycelium of *Fusarium solani* growth in the presence of the purified pyoverdine *P. putida*

- The experiments were carried out using:
- KB medium (Photo 2a, b) as well as
- KB medium that had been previously depleted of iron (Photo 2c, d).



Jankiewicz & Kuzawińska,2009

1. Isolation of pyoverdins
2. Complementary tests

- The bacteria were grown in SSM mineral medium.
- The final pH of the medium was adjusted to 7.0 with 2 M solution of NaOH.
- All culture media were prepared using deionized water.
- All laboratory glass was kept in a 2 M HCl for 12 h and then rinsed several times with deionized water.
- The bacteria were grown at 28ºC with shaking (120 rpm) for 72 h and then were spun down in Sigma centrifuge at 24000×g.
- The obtained supernatant was taken as a source of pyoverdine.
- Intensity of bacterial growth was monitored based on optical density of the media (OD 550nm).

Jankiewicz & Kuzawińska,2009

2. Purification of pyoverdins

General method

- The method of Meyer and Abdallah,1978.
- The pyoverdins produced in succinate medium were extracted with chloroform/phenol from the culture supernatants followed by separation in Sephadex CM25 column chromatography and eluting with 0.2 M pyridine/acetic acid buffer (pH 6.5).
- The compounds were rechromatographed twice on CM Sephadex.
- Iron-free pyoverdins were obtained by mixing the purified ferric pyoverdins with EDTA solution (1M) at pH 7 for 30 min.
- This was then applied to a Sephadex G25 gel filtration column to separate the pyoverdin from the ferric EDTA complex.
- Pyoverdine in column was eluted with 20 mM acetate buffer (pH 5.0) containing 100 mM NaCl.
- Fractions of 1.6 ml were collected.

Ethylenediaminetetraacetic acid(EDTA) is mainly used to sequester metal ions in aqueous solution.

3. Detection of siderophores

3.1. Spectrophotochemical detection of siderophore production

- The presence of pyoverdine was determined photometrically by measuring absorbance at λ_{405}.
- The content of pyoverdine was calculated taking into account the value of the molar absorbance coefficient 20000 $mol \cdot cm^{-1}$(Meyer J.M., Abdallah M.A.,1978).

Jankiewicz & Kuzawińska,2009

3. Detection of siderophores

3.1. Spectrophotochemical detection of siderophore production

- Each strain was grown for 3 days in one petri dish containing 10 ml of GASN liquid medium and one block of GASN agar medium.
- Then 36 µl of an $FeCl_3$ solution (1 M) was added, and the cultures were shaken for 20 min, centrifuged (20 min, 10,000 × *g*), and filtered.
- The pH was modified between pH 3.0 and 7.0, and the colors of the solutions were noted.
- After being possibly diluted three times, when the siderophores were too concentrated to be directly analyzed, the culture media equilibrated at pH 7.0 were analyzed with UV-VIS spectrophotometer, starting from 700 to 360 nm.
- Liquid GASN medium was used as a control.

Bultreys *et al.*,2001

3. Detection of siderophores

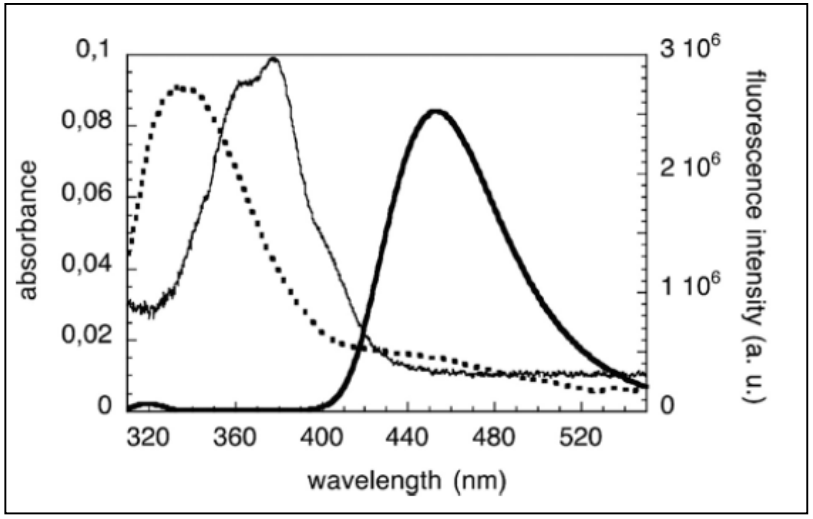
3.1. Spectrophotochemical detection of siderophore production

Spectral properties of pyoverdins

- UV spectrum of iron-free Pvd (thin line), fluorescent spectra of purified FpvA (dotted line), and iron-free Pvd (thick line).

1. For the UV spectrum of iron free Pvd (thin line), Pvd was dissolved in pyridine-acetic acid buffer pH 5.0.
2. For the fluorescent spectra of purified FpvA (dotted line) and iron-free Pvd (thick line) the excitation wavelength was set at 290 and 380 nm, respectively.

- The protein and iron-free Pvd were dissolved in Tris-HCl buffer pH 8.0 and 1% (v/v) octyl-POE.



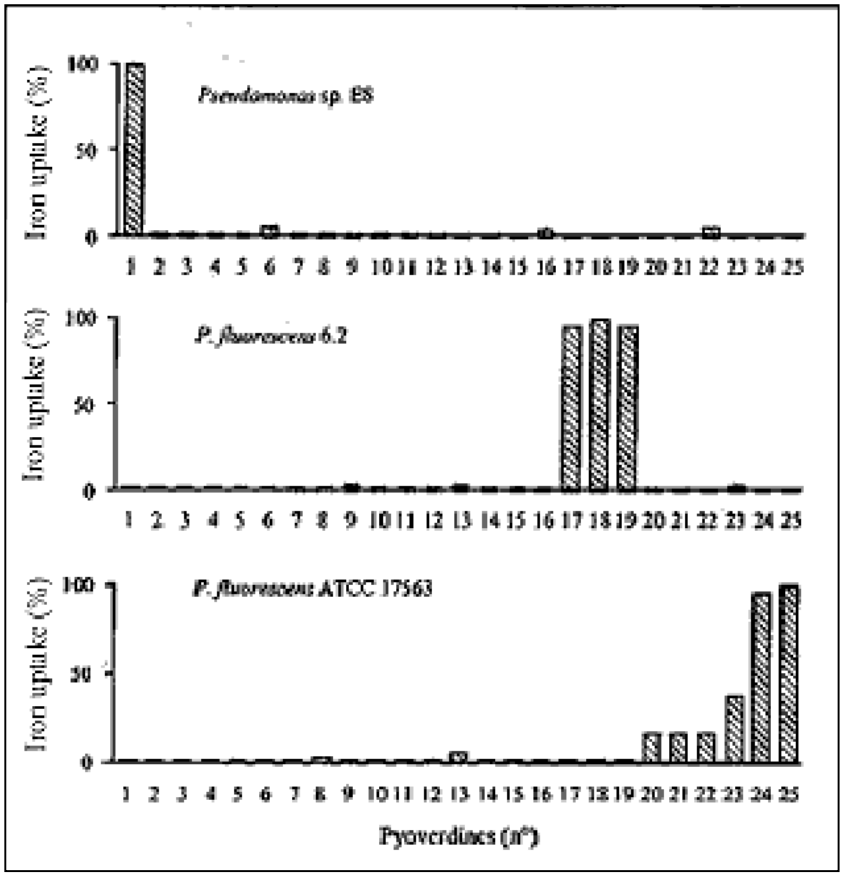
3. Detection of siderophores
3.2. Iron uptake specificities

- Another efficient biological way to discriminate pyoverdines is to define their iron transport capacities.
- Cellular ^{59}Fe radioactivity was measured after 20 min incubation of the iron starved cells with the respective (^{59}Fe) iron-pyoverdine complexes and expressed as the percentage of label reached with the homologous pyoverdine:
- PVD n° 1 for *Pseudomonas* sp. E8
- PVD n° 18 for *P. fluorescens* 6.2
- PVD n° 25 for *P. fluorescens* ATCC (17563).

Fuchs *et al.*,2001

Iron uptake specificities of *Pseudomonas* sp.

- Iron uptake specificities of:
- *Pseudomonas* sp. E8,
- *P. fluorescens* 6.2 and
- *P. fluorescens* ATCC 17563 towards 25 pyoverdines originated from different bacterial strains.



Fuchs *et al.*,2001

3. Detection of siderophores

3.3. Isoelectrophoretic patterns of pyoverdines

Characterization of bacterial PVDs by IEF

- So far, about half of the ca. 50 published structures correspond to pyoverdines were first detected by means of siderotyping techniques.
- A comparison of IEF (isoelectric focusing) patterns and uptake specificity with those of known pyoverdine strongly suggested that the selected compounds should correspond to new structures.

Fuchs *et al.*,2001

3. Detection of siderophores

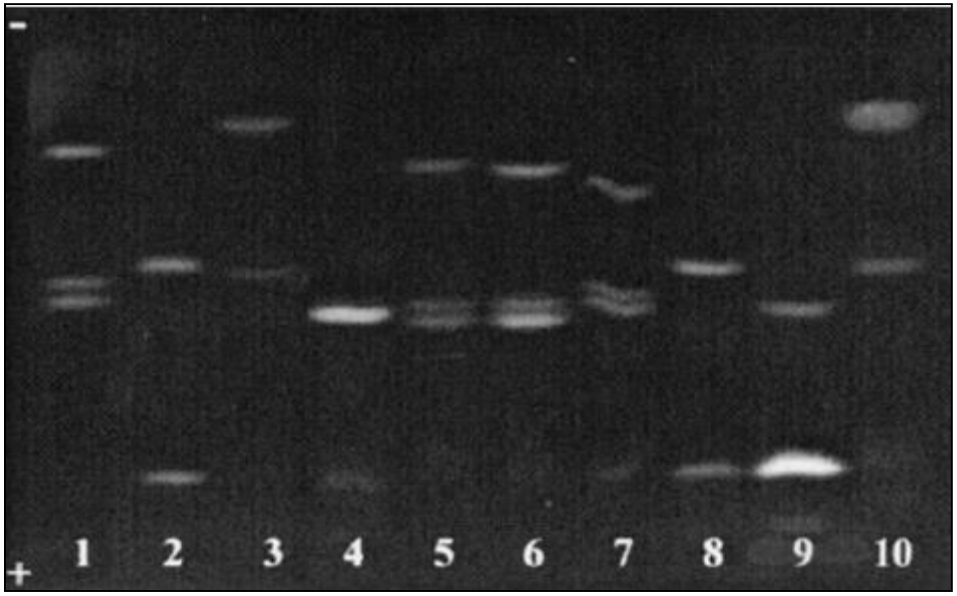
Isoelectrophoretic patterns of pyoverdines

Characterization of bacterial PVDs by IEF

- PVDs produced by different strains of fluorescent pseudomonads can easily be differentiated from each other by isoelectric focusing of partially purified siderophores, as the different peptides result in PVDs with different pIs that can be visualized directly under UV light.
- Electrophoresis of PVDs on ampholine-containing polyacrylamide gel (PVD-IEF) results in the separation of the different molecular forms of PVD present in the supernatant of an iron-starved fluorescent pseudomonad culture.
- Strains belonging to the same group usually developed an identical pyoverdine-IEF pattern.

PVD-IEF patterns

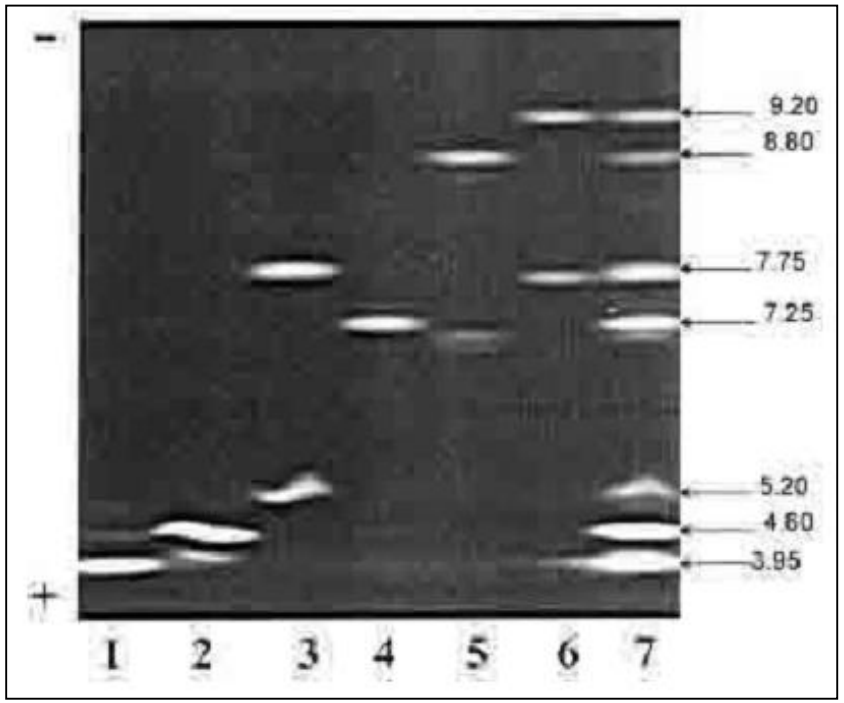
- 10 different PVD-IEF patterns were observed upon analyzing the culture supernatants of the *P. tolaasii* and "*P. reactans*" strains grown under iron-deficient conditions (CAA medium).



PVD-IEF patterns

Pyoverdine isoelectric focusing profile

- lane 1, *P. putida* strain C; lane 2, *P. putida* ATCC 12633; lane 3, *P. fluorescens* strain ii; lane 4, *P. rhodesiae* CFML 92-104; lane 5, *P. aeruginosa* ATCC 15692; lane 6, *P. fluorescens* strain 1.3.
- Lane 7 represents the isoelectrophoretic pattern of an equal mix of the pyoverdines depicted in lanes 1 to 6 which could be used as an internal standard for pyoverdine isoelectrophoretic pH (pI) determination.
- The pI values indicated on the right side were determined by the gel slicing method.



Fuchs *et al.*,2001

PVD-IEF patterns

A precise pI(pyoverdine isoelectrophoretic pH) determination

- The appearance of different side chains leads to different pI-values (pH at the isoelectric point of a compound) and allows the separation of the compounds through chromatography and electrophoresis.
- Each band could be characterized by the pH value measured at the place on the gel where the PVD isoform localizes.

Fuchs *et al.*,2001

A precise pI determination

pI values

Method

- A precise pI determination is useful for characterizing the pyoverdine bands.
- The commercially available pI standard mixture made of proteins is not convenient since it requires a Comaasiie Blue staining with several steps that in the case of pyoverdines leads to their decomposition.
- Therefore, a direct measurement of the pH reached at different portions of the gel is done by slicing the gel in 0.5 cm bands.
- Each band is incubated in 2ml of a 10 mM KCl solution for 30 min, the pH of the solution being measured with a mini electrode-equipped representative of the pH gradient on the gel, which allows the determination of the pH value at the place of the gel pH-meter.

Fuchs *et al.*,2001

pI-values of different pyoverdines (peptide chains) as determined by the silicing method

Strain	Pyoverdine peptidic structure [a]	pI values [b]
P. putida C	Asp-BOHOrn-Dab-Thr-Gly-Ser-Ser-OHAsp-Thr	(5.2), (4.65), **3.95**
P. putida ATCC 12633	Asp-Lys-OHAsp-Ser-Thr-Ala-Glu-Ser-cOHOrn	**4.6**, (4.0)
P. fluorescens ii	Ala-Lys-Gly-Gly-OHAsp-Gln-Ser-Ala-Ala-Ala-Ala-cOHOrn	**7.75, 5.2**
P. rhodesiae CFML92-104	Ser-Lys-FOHOrn-Ser-Ser-Gly-c(Lys-FOHOrn-Ser-Ser)	(8.4), (7.35), **7.25**
P. aeruginosa ATCC 15692	Ser-Arg-Ser-FOHOrn-c(Lys-FOHOrn-Thr-Thr)	**8.8**, (8.7), 7.0, (6.9)
P. fluorescens 1.3	Ala-Lys-Gly-Gly-OHAsp-Gln/Dab-Gly-Ser-cOHOrn	**9.2**, 7.2

- Values in parentheses correspond to minor bands.
- Values in bold characters correspond to the major bands used as internal standards (see previous slide).

Fuchs *et al.*,2001

pI values of PVDs of *P. tolaasii* and "*P. reactans*" strains and grouping of the strains into siderovars.

Bacterial strain	Origin	pI values of PVD isoforms	Siderovar
LMG 2342^T	LMG Collection[a]	9.1, 7.5, 7.3	*P. tolaasii* sv. 1
LMG 6641	LMG Collection	9.1, 7.5, 7.3	
PS8.2	Venlo, The Netherlands	9.1, 7.5, 7.3	
PS8.5	Venlo, The Netherlands	9.1, 7.5, 7.3	
PS8.14	Venlo, The Netherlands	9.1, 7.5, 7.3	
PS8.14V	This study	9.1, 7.5, 7.3	
PS19.2	Mutzig, France	9.1, 7.5, 7.3	
PS19.12	Mutzig, France	9.1, 7.5, 7.3	
PS20.1	Dassel, Germany	9.1, 7.5, 7.3	
PS22.2	Oulu, Finland	9.1, 7.5, 7.3	
PS3a	Kiukainen, Finland	7.7, 5.3	*P. tolaasii* sv. 2
PS3c	Kiukainen, Finland	7.7, 5.3	
PS3d	Kiukainen, Finland	7.7, 5.3	
PS3e	Kiukainen, Finland	PVD^{-b}	
PS7.5	Kiukainen, Finland	7.7, 5.3	
PS7.10	Kiukainen, Finland	7.7, 5.3	
PS9.4	Kiukainen, Finland	7.7, 5.3	
PS12.1	Kiukainen, Finland	7.7, 5.3	
PS12.4	Kiukainen, Finland	7.7, 5.3	
PS12.5	Kiukainen, Finland	7.7, 5.3	
PL24.1	Kurjenkylä, Finland	9.2, 7.6	"*P. reactans*" sv. 1
PL24.14	Kurjenkylä, Finland	9.2, 7.6	
PS8.1	Venlo, The Netherlands	7.2, 5.2	"*P. reactans*" sv. 2
PS8.21	Venlo, The Netherlands	7.2, 5.2	
PS11.4	Venlo, The Netherlands	7.2, 5.2	
PS16.18	Mellerud, Sweden	7.2, 5.2	
PS18.10	Rödeby, Sweden	7.2, 5.2	
PL21.4	Creil, France	7.2, 5.2	
PL5.13	Humppila, Finland	7.2, 5.2	
PS21.4	Yläne, Finland	7.2, 5.2	
PL23.3	Kurjenkylä, Finland	7.2, 5.2	
PL24.10	Kurjenkylä, Finland	7.2, 5.2	
PL26.16	Jämijärvi, Finland	7.2, 5.2	
PL28.3	Jämijärvi, Finland	7.2, 5.2	
PL28.10	Jämijärvi, Finland	7.2, 5.2	
PL30.1	Jämijärvi, Finland	7.2, 5.2	
PS7.2	Kiukainen, Finland	8.9, 7.3, 7.1	"*P. reactans*" sv. 3
PL25.2	Kurjenkylä, Finland	8.9, 7.3, 7.1	
LMG5329	LMG Collection	8.8, 7.3, 7.1	"*P. reactans*" sv. 4
PS8.6	Venlo, The Netherlands	8.8, 7.3, 7.1	
PS8.12	Venlo, The Netherlands	8.8, 7.3, 7.1	
PS11.10	Venlo, The Netherlands	8.8, 7.3, 7.1	
PS13.10	Kiukainen, Finland	8.8, 7.3, 7.1	
PS16.14	Mellerud, Sweden	8.8, 7.3, 7.1	
PS19.6	Mutzig, France	8.8, 7.3, 7.1	
PS20.11	Dassel, Germany	8.8, 7.3, 7.1	
PS3b	Kiukainen, Finland	8.4, 7.4, 7.3	"*P. reactans*" sv. 5
PS6.4	Kiukainen, Finland	8.4, 7.4, 7.3	
PL10.9	Vantaa, Finland	7.6, 5.3	"*P. reactans*" sv. 6
PL4.14	Humppila, Finland	7.3, 5.4	"*P. reactans*" sv. 7
PL4.20	Humppila, Finland	7.3, 5.4	
PL21.3	Yläne, Finland	7.3, 5.4	
PL26.3	Jämijärvi, Finland	7.3, 5.4	
PL29.9	Jämijärvi, Finland	7.3, 5.4	
PL29.12	Jämijärvi, Finland	7.3, 5.4	
PL4.6	Humppila, Finland	9.2, 7.7	"*P. reactans*" sv. 8
PL5.5	Humppila, Finland	9.2, 7.7	
PS6.10	Kiukainen, Finland	9.2, 7.7	
PS12.14	Kiukainen, Finland	9.2, 7.7	
PS14.1	Kiukainen, Finland	9.2, 7.7	
PS14.8	Kiukainen, Finland	9.2, 7.7	

Cross uptake studies
Cross-reactivity

- Pvds of *Pseudomonas* species, usually restricted to the producer strain or to strains producing an identical compound.
- However, cross-reactivity does also occur.
- This may pose uncertainty for relying on siderotyping methods.
- Also promising future is expected from recent developments which indicate that pyoverdines might be considered as potent and easy-to-handle taxonomic markers for the fluorescent species of the genus *Pseudomonas*.

Meyer,2000;..

Cross uptake studies involving 25 strains of fluorescent Pseudomonads and their respective pyoverdines

- Cross uptake studies involving 25 strains of fluorescent Pseudomonads and their respective pyoverdines.
- The squares in dark grey correspond to groups of strains involving within each group pyoverdines with identical IEF patterns.
- Light grey squares correspond to assays with pyoverdines presenting some differences in their IEF patterns.

Bacterial Strains	Iron uptake as mediated by pyoverdine of strain n°																								
	n°	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Pseudomonas sp. E8	1	++																								
P. syringae ATCC 19310	2		++																							
P. fluorescens 9AW	3			++																						
P. putida ATCC 12633	4				++																					
P. fluorescens 51W	5					++																				
P. fluorescens ii	6					+	++																			
P. aeruginosa Pa6	7							++																		
P. fluorescens CCM 2798	8								++																	
P. fluorescens CHAO	9									++																
P. tolaasii	10										++															
P. aeruginosa ATCC 27853	11											++														
P. aptata Ps4a	12												++	++												
Pseudomonas sp. SB8.3	13												++	++												
P. putida strain Gwose	14														++	++										
P. fluorescens ATCC 17400	15														+	++	++									
P. fluorescens 1.3	16														+	++	++									
Pseudomonas sp. 267	17																	++	++	++						
P. fluorescens 6.2	18																	++	++	++						
Pseudomonas sp. B10	19																	++	++	++						
P. chlororaphis ATCC 9446	20																				++	++	++	+		
P. fluorescens 10CW	21																				++	++	++	+		
P. fluorescens ATCC 13525	22																				++	++	++	+		
P. fluorescens ATCC 15692	23																				+	+	+	++	+	+
P. fluorescens 18.1	24																				+	+	+	+	++	++
P. fluorescens ATCC 17563	25																				+	+	+	+	++	++

Fuchs *et al.*,2001

PCR protocol for detection of siderophore
Useful in identification of some pvs. of *Pseudomonas syringae*
Yersininabactin gene, *irp1*

- The siderophore pyoverdin, produced by all fluorescent strains of *P. syringae*, is a character used for diagnosis (Bultreys *et al*.,2003).
- Among stone fruit pathogens, the siderophore yersiniabactin is produced by the pathovars *morsprunorum* race 2, *avii*, and *persicae* and, generally, by genomospecies 3, 7 and 8 of *P. syringae* (Bultreys *et al*.,2006).
- Pyoverdin is not involved directly in pathogenicity of *P. syringae* pv.*syringae* but is a virulence factor for *P. syringae* pv.*tabaci*.
- The *irp1* gene encoding a polyketide synthase/peptide synthetase involved in yersiniabactin synthesis is used in diagnostics.

Bultreys and Kaluzna,2010

Role of siderophore yersiniabactin

- The siderophores pyoverdin and yersiniabactin play a (possibly indirect) role in pathogenesis and in competition with other microorganisms on plant surfaces(ecological role).

Bultreys and Kaluzna,2010

PCR for detection of two toxins and yersiniabactin siderophore

Differentiate pvs. *syringae*, *morsprunorum* and *avii*

Table 4 – Differential PCR's for races of *Pseudomonas syringae* pv. *morsprunorum*, *P. s.* pv. *syringae* and *P. s.* pv. *avii*

Bacterium	PCR *cfl* coronatine	PCR *syr B* or *syrD*	PCR irp1 yersiniabac
P.s. pv. *morsprunorum* race 1	+, few -	-	-
P.s. pv. *morsprunorum* race 2	-	-	+
P.s. pv. *syringae*	-	+, few -	-
P.s. pv. *avii*	-	-	+

- = negative; + = positive

Janse,2010

Yersiniabactin gene *irp1*

Differentiate pvs. *morsprunorum* race 2 and *avii* from *P. syringae* pv. *syringae*

- The development of a PCR protocol for detecting the gene *irp1* involved in the production of the siderophore yersiniabactin.
- Detection of the yersiniabactin gene *irp1* by PCR using primers PSYE2/PSYE2R helps to differentiate pvs. *morsprunorum* race 2 and *avii* from *P. syringae* pv.*syringae*.
- The strains of pathovar *morsprunorum* race 2 and *avii* both give positive results.

Alain Bultreys

Yersiniabactin gene *irp1*

Differentiate pvs. *morsprunorum* race 2 and *avii* from *P. syringae* pv. *syringae*

Group[a]	Strain	Host	Year of isolation	Genes encoding toxin production		
				*syr*B (syringomycin)	*cfl* (coronatine)	*irp*1 (yersiniabactin)
Psm race 1	701A	Sweet cherry	2005	–	+	–
	702	Plum	1994	–	–	–
	704	Sweet cherry	1994	–	+	–
	710	Sweet cherry	1996	–	+	–
	755	Plum	1999	–	–	–
	771	Plum	1999	–	–	–
	782	Sweet cherry	2001	–	–	–
	787	Plum	2001	–	–	–
	788	Plum	2001	–	–	–
	793	Plum	2001	–	–	–
	LMG 2222*	*Prunus avium* cv. 'Napoleon'		–	+	–
Psm race 2	701	Sour cherry	1994	–	–	+
	719	Sour cherry	1997	–	–	+
	732	Sour cherry	1997	–	–	+
	733	Sour cherry	1997	–	–	+
	745	Sour cherry	1999	–	–	+
	764	Sour cherry	1999	–	–	+
	CFBP 3800*	*Prunus cerasus*		–	–	+
Atypical[b]	791	Sour cherry	2001	–	–	–
Pss	702A	Plum	2005	+	–	–
	753	Apricot	1999	+	–	–
	757	Plum	1999	+	–	–
	760	Sour cherry	1999	+	–	–
	762	Apricot	1999	+	–	–
	763	Sour cherry	1999	+	–	–
	2905*	Sour cherry	1978	+	–	–
	LMG 1247*	*Syringa vulgaris*		+	–	–

Kałużna & Sobiczewski,2009

The Species Concept in Prokaryotes
Species definition for bacteria

Different species concepts

Biological species concept may or may not be applied to prokaryotes!!!

What is a species?
General definition

- Species = Latin term meaning "kind" or "appearance".
- Linnaeus (founder of modern taxonomy) described species in terms of their physical form (morphology).
- Morphology is still the most common method used for describing species.
- Modern taxonomists also consider genetic makeup and functional and behavior features when describing species.

Speciation

The formation of new species

Speciation and Taxonomy

- Speciation refers to the formation of species.
- Speciation is the evolutionary process by which new biological species arise.
- The evolution of two or more new species from one pre-existing species, will occur if different groups within a species evolve to become different from each other -- so different that they would be considered different species.

Speciation

Kinds of species concept and modes of speciation

- Kinds of species concept:

1. Biological species concept;
2. Morphological species concept;
3. Recognition species concept;
4. Ecological species concept;
5. Evolutionary species concept.

- Types or modes of speciation:

1. Allopatric
2. Sympatric

Speciation

1. Kinds of species concept

1. Biological species concept(a group of individuals that can breed together. However, they cannot breed with other groups).
2. Morphological species concept (on the basis of phenotypic features).
3. Recognition species concept (unique set of molecular, morphological, or behavioral characteristics).
4. Ecological species concept (defines a species on the basis of where they live and what they do).
5. Evolutionary species concept (defines a species as a sequence of ancestral and descendent populations that are evolving independently of other such groups).

Biological species concept

- Biological species: A population or group of populations whose members have:

1. The potential to interbreed with one another in nature, and
2. To produce viable, fertile offspring, but cannot
3. Produce viable, fertile offspring with members of other species.

Biological species concept

Biological species concept does not apply to prokaryotes!!!

- The biological species concept cannot be applied to:
- Organisms that are completely asexual in their reproduction:

1. Some protists and fungi,
2. Some commercial plants (bananas), and
3. Many bacteria are exclusively asexual.

Biological species concept

Biological species concept may be applied to prokaryotes!!!

- Regarding biological species concept, most commonly applied to higher organisms, could be applied to bacteria, implying that there is a gene pool shared by members of species.

Biological species concept

Biological species concept may be applied to prokaryotes

Gene transfer (gene flow)

- All bacteria share a common gene pool and all belong to the same species.
- Mixing of 'species' gene pools would give the equivalent of the Biological Species Concept.
- 10-15% of *E. coli* genome derived from lateral gene transfer.

Biological species concept

Biological species concept may be applied to prokaryotes

Gene transfer (gene flow)

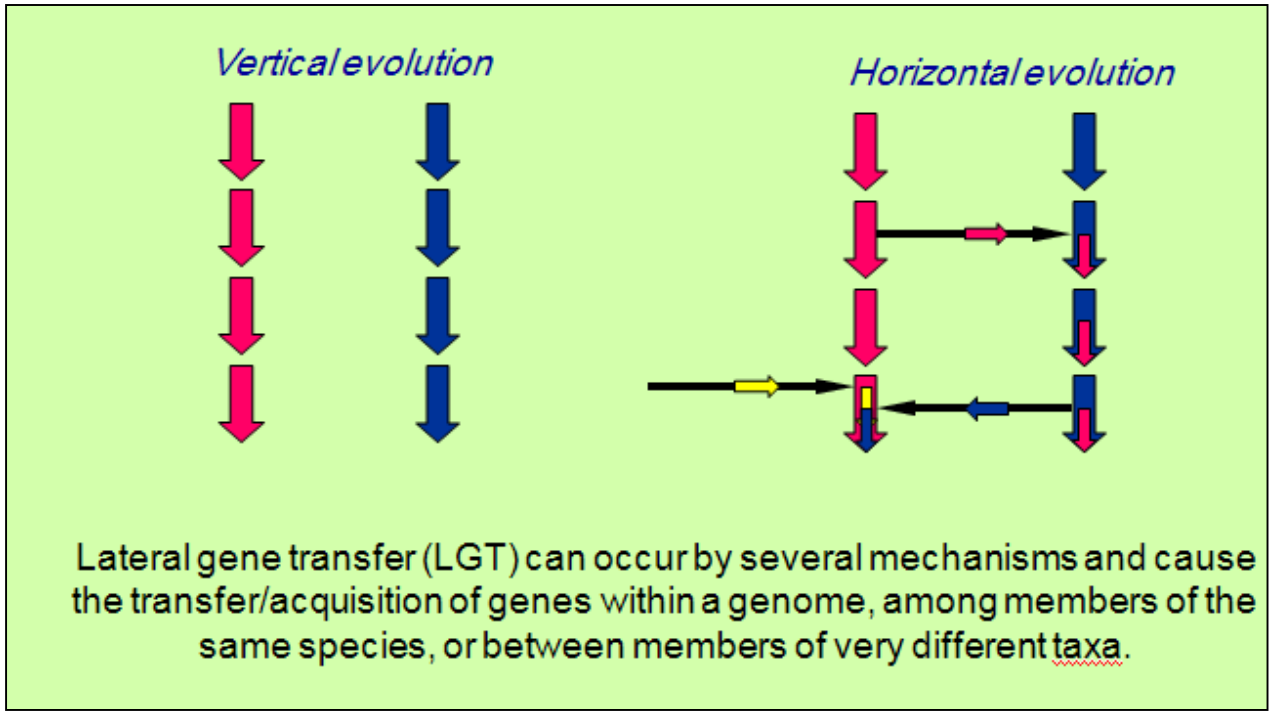
- Basically, bacteria do not respect conventional species concepts.
- By this, I am referring to their habit of sharing small segments of their DNA with neighboring bacteria, regardless of "species".
- Because bacterial DNA swapping is different to what we see when parents provide copies of their DNA to their offspring (vertical gene transfer), this process has another name: horizontal gene transfer.

GrrlScientist,2011

Gene transfer

Vertical gene transfer vs Lateral gene transfer

- Lateral (or horizontal) gene transfer denotes any transfer, exchange or acquisition of genetic material that differs from the normal mode of transmission from parents to offspring (vertical transmission).



Speciation

2. Types or modes of speciation

- Two organisms, or two populations, or two varieties, species or taxa, can live in relation to each other in two ways, geographically speaking.

1. Allopatric speciation:

- Speciation occurs with physical separation.
- They can live in different regions (*allo*-meaning "other", plus *patria* gives allopatry).

2. Sympatric speciation:

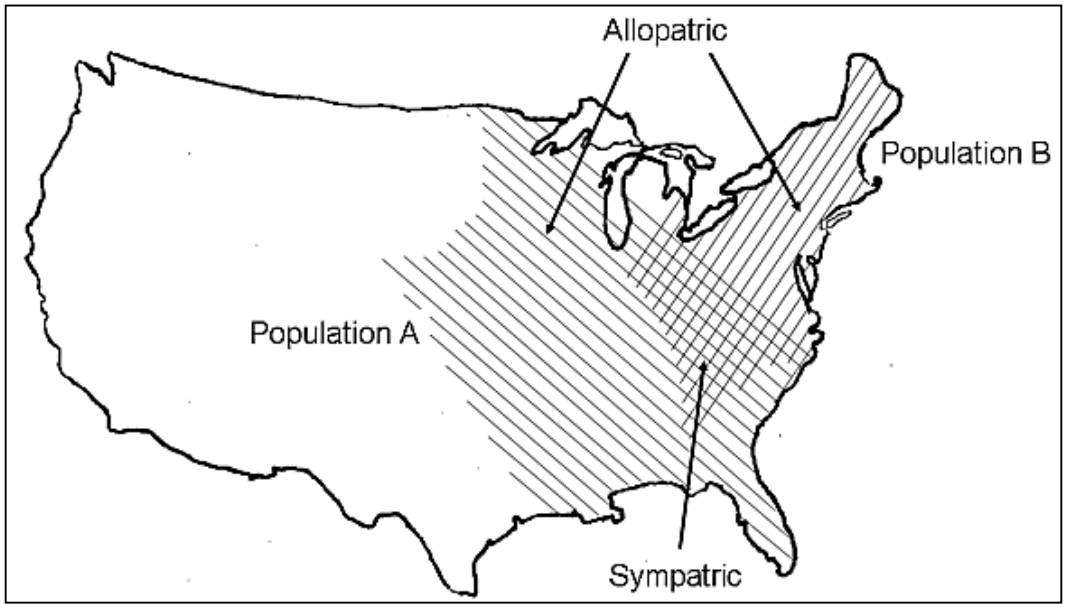
- Speciation occurs without physical separation.
- They can live in the same region (*sym-* meaning "together", plus *patria*, gives *sympatry*).

Scoville,2014;..

Speciation

Comparison of allopatric and sympatric speciation

1. Allopatric: refers to populations that are geographically isolated (they do not overlap).
2. Sympatric: refers to populations whose ranges overlap.

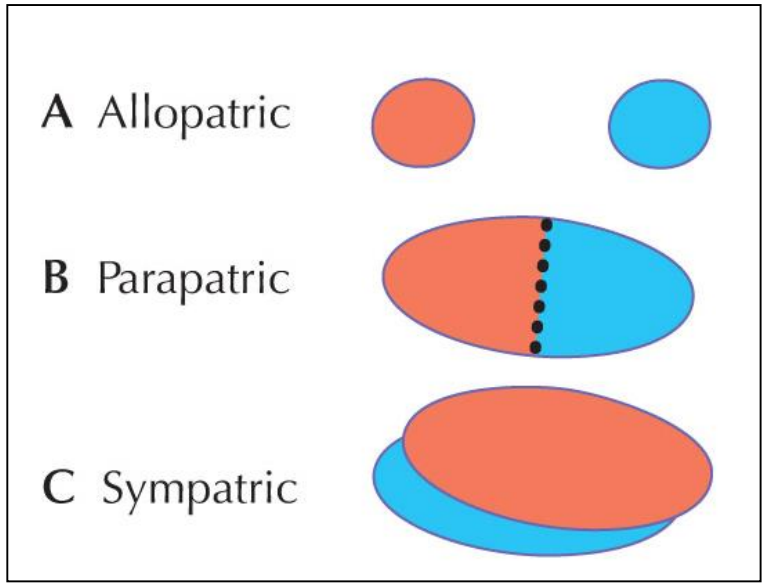


Speciation

Types or modes of speciation

Allopatric, sympatric, platypatric

1. **Allopatric speciation:**
   - Two populations are separated by some geographically barrier. **No HGT will occurs**. They therefore diverge in complete independence.
2. **Parapatric speciation:**
   - There is only partial separation of the zones of two diverging populations afforded by geography(dotted line).
3. **Sympatric speciation:**
   - There is no spatial separation.They can live in the same region.



Evolution,2007;..

Sympatric speciation

Bacteria

Obligate pathogens or symbionts

- Sympatric speciation is very different from the other forms because new species emerge from populations living in highly overlapping or even identical areas.
- It may be more common in bacteria than in multicellular organisms.
- Because:

1. bacteria can transfer genes to each other, as well as
2. transfer genes to offspring when they divide.

https://biologydictionary.ne

Sympatric speciation

Bacteria

Obligate pathogens or symbionts

- Bacteria that evolve in close association with a host, e.g. obligate pathogens or symbionts, and exhibit limited dispersal hence mixing outside the host, are likely to exhibit well-defined genetic clusters as an effect of their coevolution with their host (sympatric speciation).
- This scenario (sympatric speciation) may be generally applicable to bacteria that have a very narrow ecological niche and their dispersal between inhabitable niches is restricted by survival and/or distance factors.

Konstantinidis *et al.*,2006

Four modes of speciation

Allopatric, sympatric, peripatric and parapatric

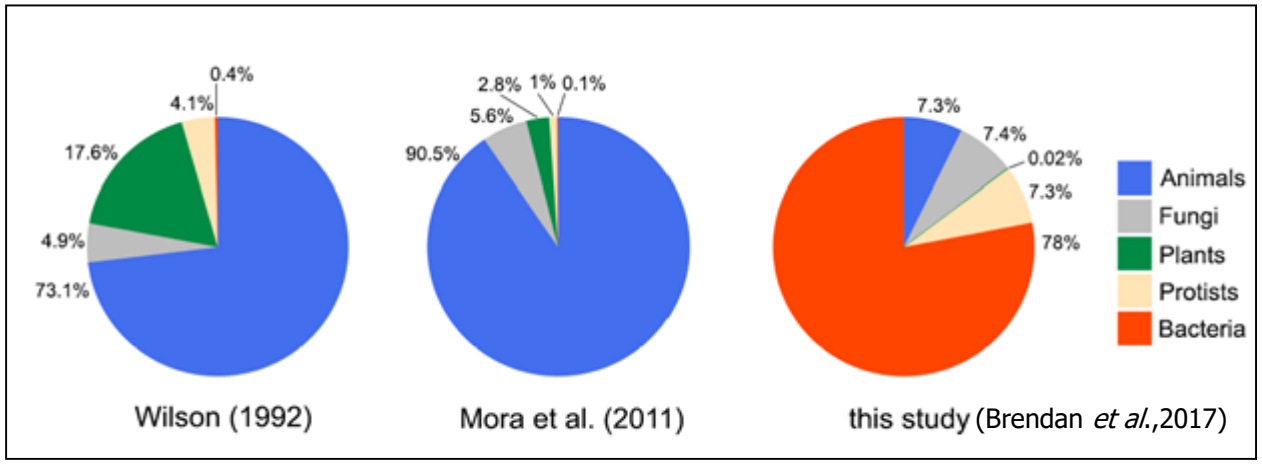
Biodiversity

- All forms of natural speciation have taken place over the course of evolution; however it still remains a subject of debate as to the relative importance of each mechanism in driving biodiversity.

Scoville,2014;Wikipedia,2014

Species discovery
Exploring biodiversity

1. To date, about 1.5 million species have been formally described in the scientific literature, most of them insects.
2. 350-400,000 species of land plant have been described.



Bacteria comprise less than 1% of all described species.

Species discovery
Exploring biodiversity
Counting in a bacterial world

- An informal survey of 24 microbiologists at a conference in Spain last year tried to reach a consensus.

1. Two people guessed that the Earth was home to between 10,000 and 100,000 species of bacteria;
2. Another five said between 100,000 and 1,000,000;
3. While nine microbiologists put the put the upper limit at ten million species;
4. Eight guessed there were even more..

The Short Sharp Science blog,2018

Species discovery

Exploring biodiversity

Counting in a bacterial world

- A 2011 paper in the journal PLOS Biology put the total number of species at 8.7 million, but that study's methodology calculated the existence of only 10,000 bacterial species, a contradiction of the 2004 review that put the minimum known bacterial species above 35,000.
- It is suspected that many bacteria cannot be cultured i.e. the niche conditions cannot be reproduced. One of the reasons is that to date less than 5000 prokaryotic species have been described.
- The result is that only a relatively small number of bacterial species are described (7029 in year 2005).
- So most bacterial diversity remains to be discovered……. How?

Pappas,2016;..

Species discovery

Exploring biodiversity

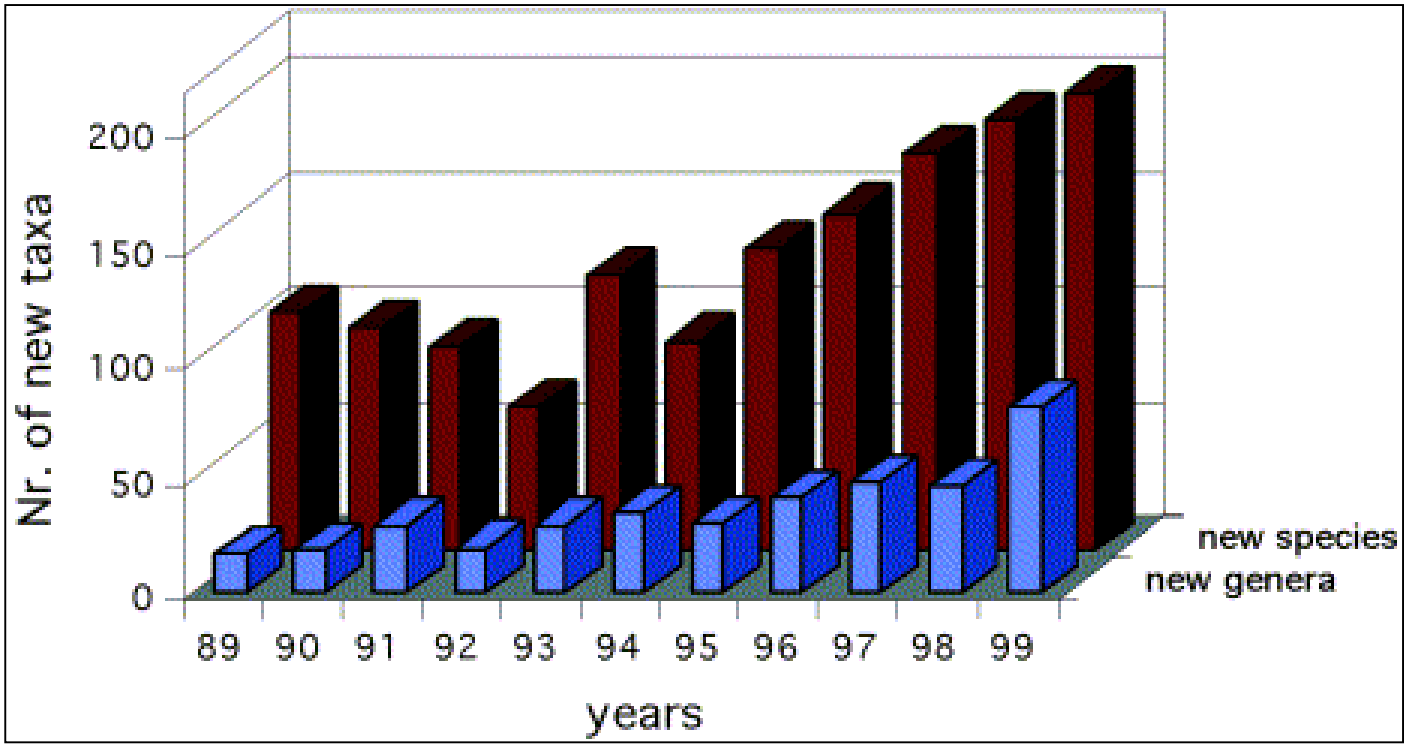
Counting in a bacterial world

- Question: Why we should care whether there are 10,000 or a million or even a billion different kinds of bacteria?
- Answer: If we're interested in discovering new drugs from bacteria - and pharmaceutical companies certainly are - it's important to know what's out there.

The Short Sharp Science blog,2018;…

Descriptions of new taxa through the years 1989 and 1999

The current number is 10,929



During the last years, the number of new isolates as well as the amount of information useful for systematics has increased significantly. New combinations of already existing taxa have not been included.

Species concept

Microbial diversity

Types of diversity

- **Morphological diversity:**
- Bacilli, cocci, and spirals are 3 common shapes, but we've also seen filamentous forms, pleiomorphic forms. Although most bacteria are tiny, there are many varieties of size, ranging from submicroscopic up to a few bacteria that can be seen with the naked eye.
- **Structural diversity:**
- There are major differences between gram-positive and gram-negative bacteria, and even more profound structural differences between bacteria and archaea.
- Other differences include presence or absence of walls, external appendages (capsules,..), endospores, etc.

Species concept

Microbial diversity

Types of diversity

- **Metabolic diversity:**
- Heterotrophs vs autotrophs.
- e.g. bacterial nutrition (how they obtain energy and carbon), and differences between catabolic styles such as fermentation vs respiration (aerobic and anaerobic).
- **Genetic diversity:**
- Small ribosomal subunit sequencing has profoundly altered our perception of the extent of genetic diversity.
- Now that genomes are being sequenced for many microbes, the full extent of this diversity is being understood as never before.
- The great bulk of life's diversity is not in the eukaryotes, but in the bacteria and archaea.

Prokaryote species concept

Problems with prokaryote species concept

1. No useful fossil record.
2. No universal concept exist
3. Original classification were produced simultaneously by different microbiologists.
4. Problems with microbial diversity
5. Problems with gene transfer

Species concept

1. Problem with exist of useful fossils

- Paleo-bacteriology cannot identify species – no useful fossil record.

Species concept

2. Problems with microbial diversity

- Microbial genomes can be remarkably plastic, being capable of substantial change within very short periods of time.
- Genetic heterogeneity in individual microorganisms can arise from a number of random, semi-random, or programmed events.
- Tempo(speed) of evolution likely to be affected by:

1. Integration of entire genes/pathways, as well as,
2. Variation by mutation.

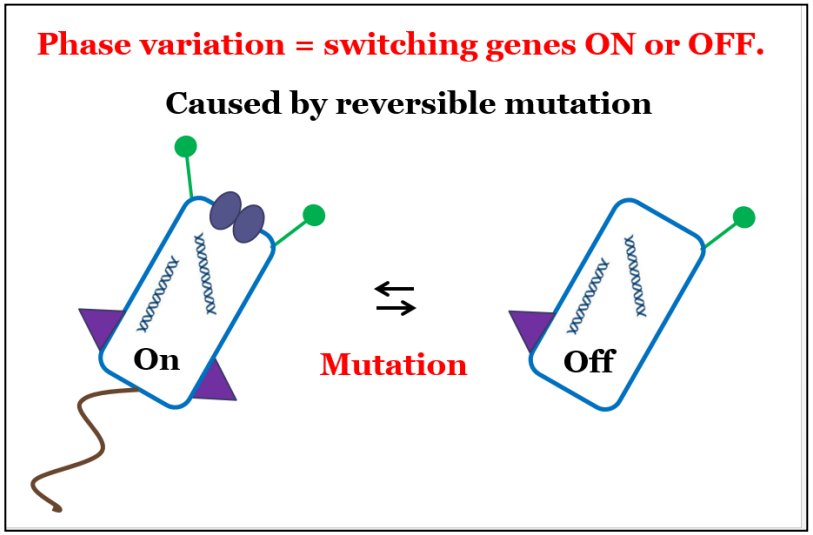
Species concept
Problems with microbial diversity

- Modes and mechanisms of genetic variability include:

1. Spontaneous point mutations (single base is substituted).
2. Random transcription events.
3. Phage-related phenomena (e.g. transduction and lysogeny).
4. Chromosomal duplications and gene amplification;
5. The presence, absence, and copy number of mobile genetic elements such as plasmids and transposons.
6. Flagellar or capsular phase variation.
7. Even intracellular genetic heterogeneity, such as that arising from transcription of multiple rRNA operons within a single cell.

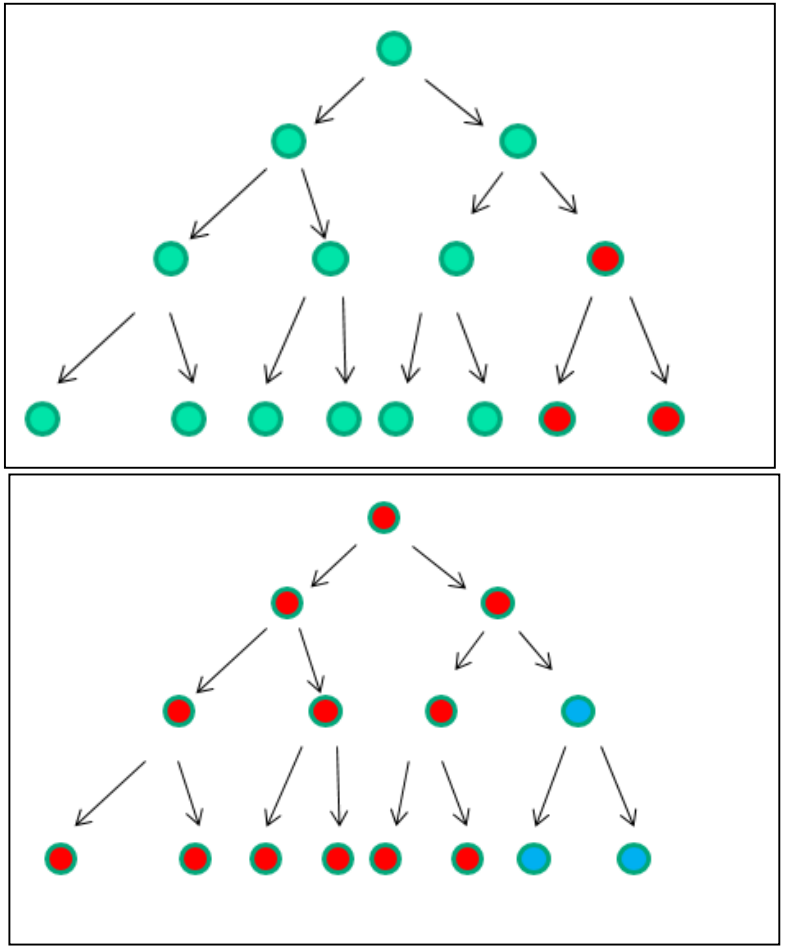
Phase variation
Mutation

- **Phase variation:**
- Phase variation is the switching on and off of genes in response to mutation.
- These mutations are reversible, meaning genes can be switched on and off interchangeably.

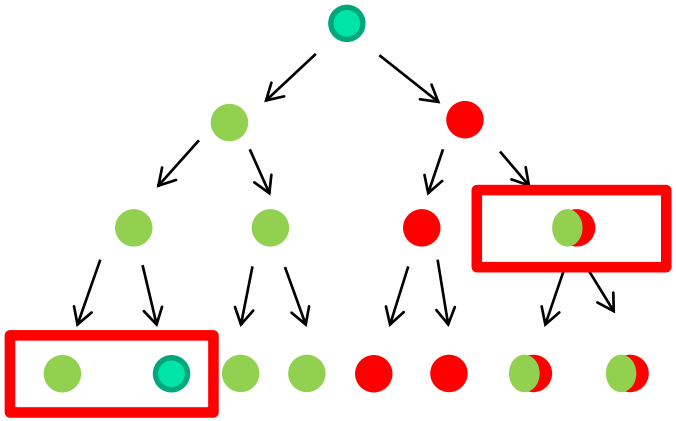


What does phase variation look like?

- Mutation is random.
- Like rolling a dice.
- One cell will divide into two daughter cells.
- With every division, there is a chance they will mutate (mutation rate).
- This mutation will be passed down to the daughter cells.
- It is a reversible process.



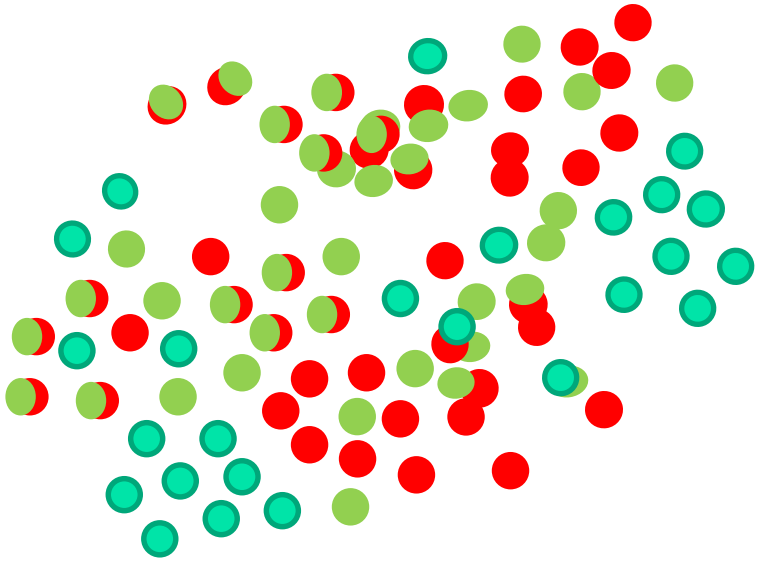
What does phase variation look like?



- Many bacteria have more than one phase variable gene
- Both genes can mutate at the same time

- Mutation isn't this quick in real bacteria.

What is the effect of phase variation on populations of cells?



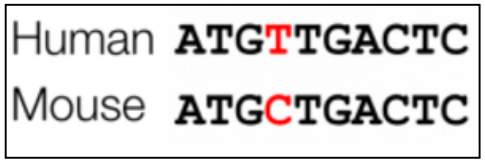
All of the different cells have the same DNA but have different antigens expressed.

Species concept

How do we estimate genetic change?

Branch length

- Spontaneous point mutations (single base is substituted).
- Estimating the extent of genetic change is not a trivial task.
- A naïve method is to align pairs of sequences, count up the number of differences and divide by the sequence length.



Simple sequence alignment.

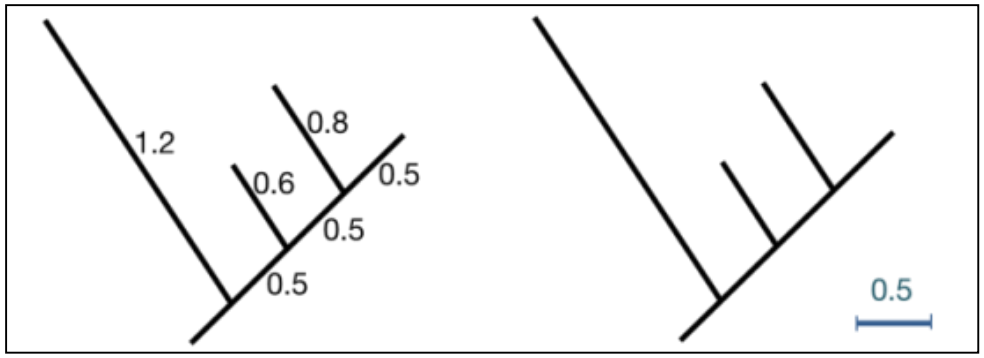
- In above figure, we can see that there is one site that is different between the two sequences, and we could say that based upon this tiny sample there are 1/10 = 0.1 substitutions per site.

Species concept

How do we estimate genetic change?

Branch length/scale bar

- Informative branch lengths are typically drawn to scale and indicate the number of substitutions per site.
- Branch lengths are occasionally shown on the phylogeny (left), but it is far more common to see branch lengths represented by a scale bar (right).



A scale bar usually indicates distances (p-distance). Measuring evolutionary distance between two sequences. Scale bar represents base or amino acid substitution per site.

Species concept

How do we estimate genetic change?

Branch length

- **Beware of very long branches!**
- To get a value of one substitution per site using the simple method above would require the pair of sequences to be completely different to each other at all 10/10 sites.
- It is unlikely you would align such sequences since two random nucleotide sequences are likely to be 25% identical.
- So if you see figures in the literature with branches longer than ~3 substitutions per site then you might want to worry about the confidence we have in those estimates!

Species concept

How do we estimate genetic change?

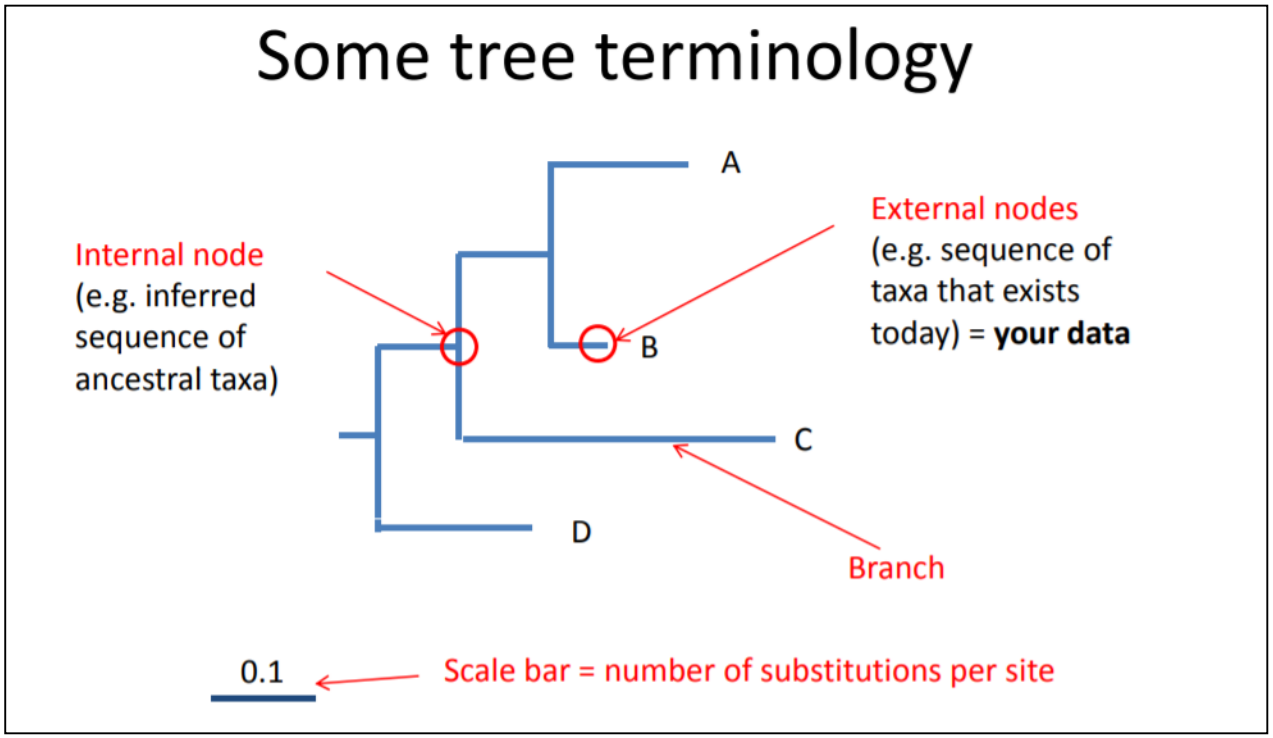
Scale bar represent p-distance

- Pairwise distance (p-distance) and p is the percent difference.
- p-distance($P=n_d/n$) = number of different aa/nt (amino acids or nucleotides) between two aligned sequences of length n.
- The "scale bar" is a reference, basically a ruler, allowing someone viewing the tree to measure the lengths of the branches in the tree, and to compare different trees.
- Typically, the scale bar line represents an evolutionary distance of 0.10 or 0.05.
- The scale bar represents p distance.

Species concept

How do we estimate genetic change?

Scale bar represent p-distance



In this figure scale bar= no. of substitution per site (one site that is different between the two sequences).

0.1= one nucleotide differences in 100 nucleotides per site.

Tadmor,2010;..

Species concept

How do we estimate genetic change?

Scale bar represent p-distance

- p-distance ($P=n_d/n$) is a simple measure of evolutionary distance is a count of the number of sites between any two sequences that are different.
- This is known as the observed distance or p-distance.
- P- distance doesn't account for multiple hits: A→ C→T.
- E.g.
- If two mutations occurs at an individual site: an A changing to a C and then to a T.
- p-distance will only account for one mutation (A→ C) and not two mutations (A→ C→ T).

Tadmor,2010

Species concept

How do we estimate genetic change?

Scale bar represent p-distance

- P- distance doesn't account for back mutation or hits: A→C→A.
- P- distance doesn't account for parallel mutations or hits: A→C; A→C.
- Underestimates d.
- Saturates at p=0.75 (not a good estimate of d when p is high).

Calculating distances in MEGA4:
p distance

p distance = 100 - percent identity/100

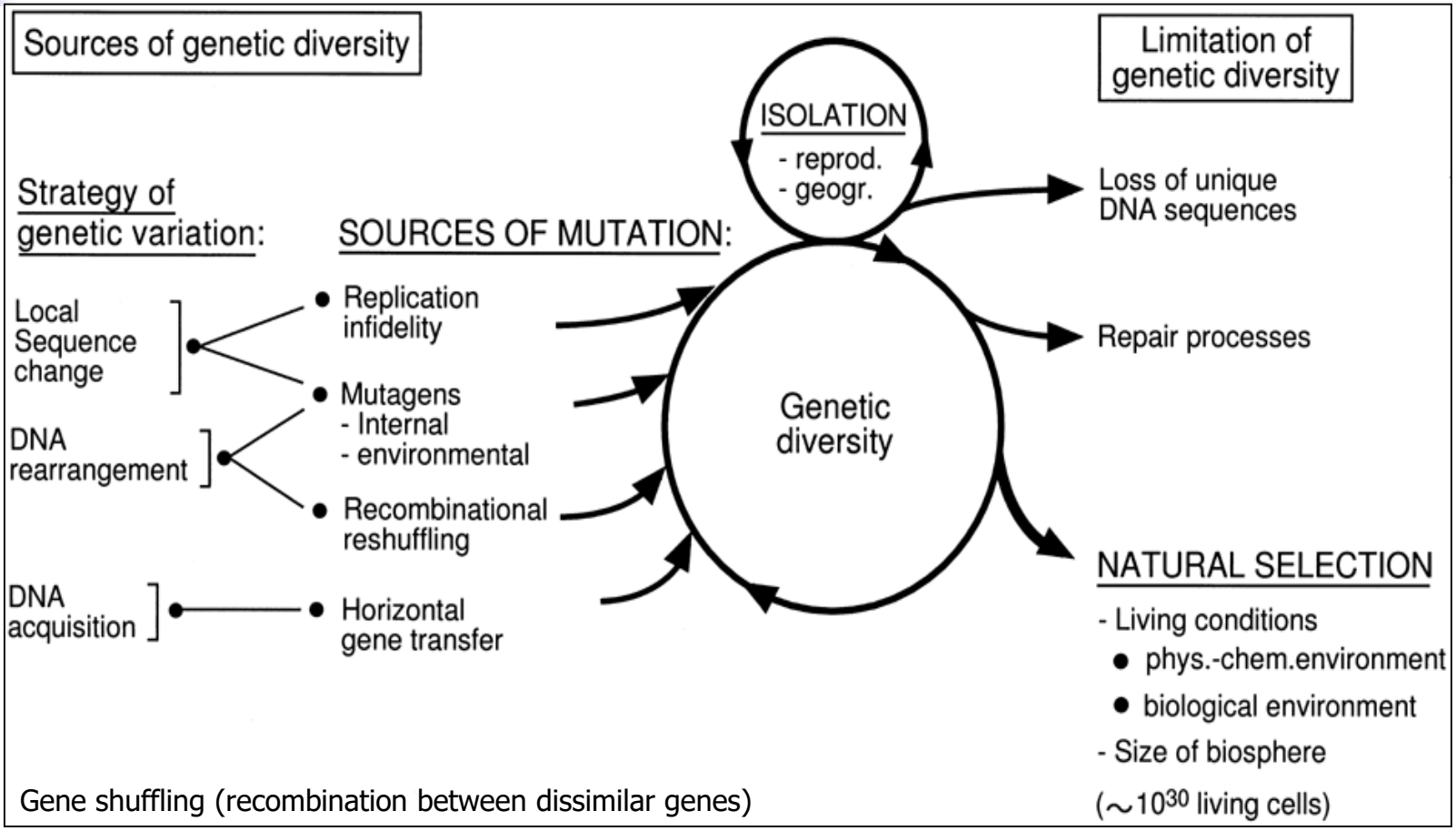
	1	2	3	4	5	6
1. Homo sapiens (human) ...						
2. Papio anubis (olive baboon) ...	0.0513					
3. Gallus gallus (chicken) ...	0.2587	0.2681				
4. Xenopus laevis (frog) ...	0.4056	0.3869	0.3660			
5. Danio rerio (zebra fish) ...	0.3846	0.4009	0.3776	0.4336		
6. Salmo salar (atlantic salmon) ...	0.3963	0.3963	0.3986	0.4522	0.2960	

Note that p distances < 0.75

Tadmor,2010

Origin of diversity: genetic drift

Synoptic view of elements of the molecular evolution of prokaryotic microorganisms

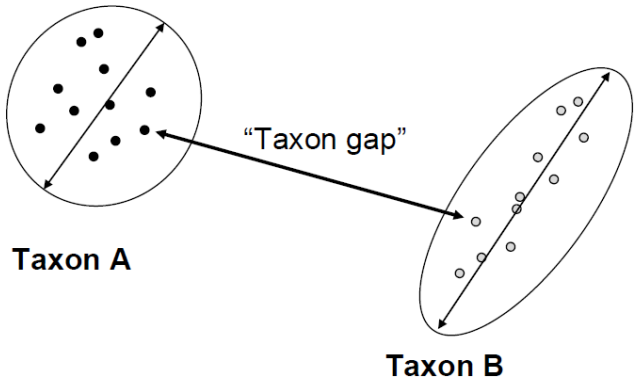


Biosphere: an area where organisms live.

Arber,2000;Vandamme,2013

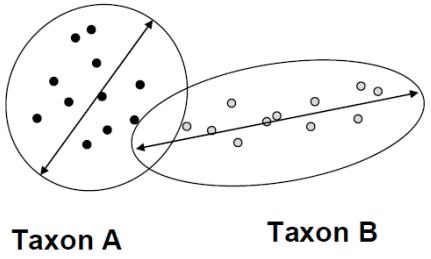
Taxonomic resolution
Intra vs. interspecies diversity

Taxonomic resolution :
intra vs. interspecies diversity



Interspecies distance > intraspecies diversity
separability > heterogeneity

Taxonomic resolution :
intra vs. interspecies diversity



interspecies distance < intraspecies diversity
separability < heterogeneity
no "taxon gap"

Tom Coenye

Species concept

3. Problems with gene transfer (gene flow)

- The greater evolutionary potential was seen in bacterial pathogens, due to their:

1. Rapid generation time, and
2. Ability to exchange genetic material horizontally (LGT/HGT).

Species concept
Prokaryotic genome

- The core or backbone of genes usually coding for components of the informational part of the cellular machinery (i.e., DNA replication, DNA transcription, and RNA translation).
- Sometimes the core is separated into:

1. a hard core (never transferred), and
2. a soft core (rarely transferred).

O'Malley and Boucher,2005

Species concept
Prokaryotic genome

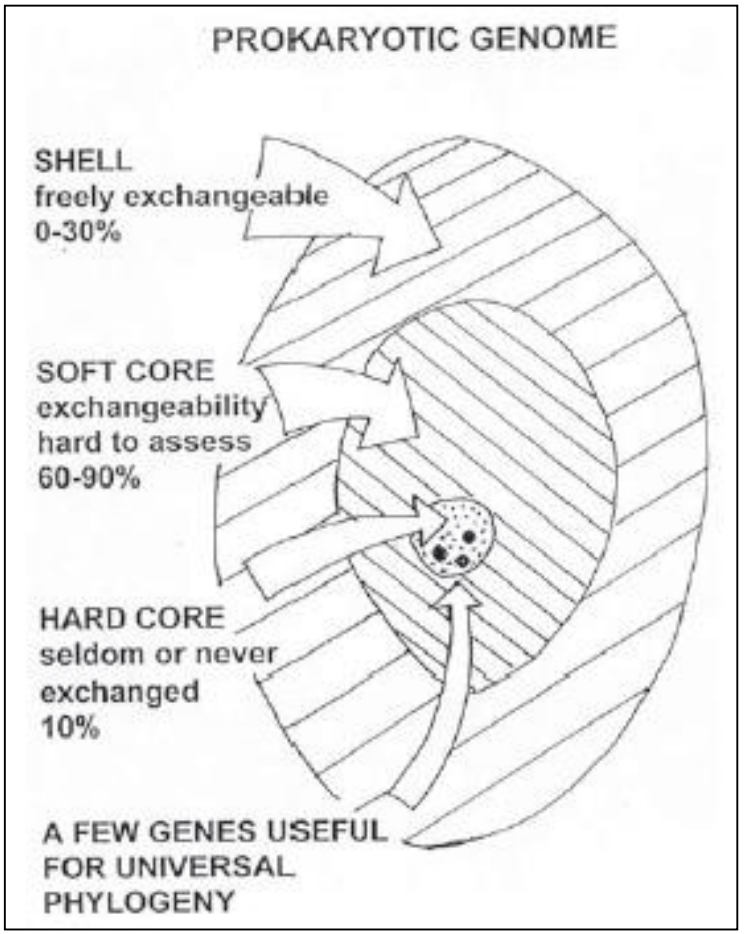
- Thus, it seems that genes could pertain to three categories:

1. The hard core that is composed of genes that are never transferred (or at least not transferred at the considered scale),
2. The soft core with genes rarely transferred (maybe about one atypical node out of 100), and
3. The shell genes where all genes susceptible to HGTs.

Philippe and Douady,2005

Species concept
Prokaryotic genome

1. **Shell:** Freely exchangeable 0-30%. A typical prokaryote genome has a shell (up to 30% of its genes) made up of genes that are quickly gained or lost and are patchily distributed among strains of its "species."
2. **Soft core:** Exchangeability hard to assess 60-90%. Most of a genome's genes will be in the "soft" core, comprising genes that are patchily distributed among phyla.
3. **Hard core:** It also has a "hard" core of genes shared with (common to) all prokaryotes, seldom or never exchanged 10%. Only a few genes useful for phylogeny.



Species concept

Prokaryotic genome

Bacterial pan-genome

- Pan-genome is defined as the set of all unique gene families found in one or more strains of a prokaryotic species.
- Due to the extensive within-species diversity in the microbial world, the pan-genome is often many times larger than a single genome. Studies of pan-genomes have become popular due to the easy access to whole-genome sequence data for prokaryotes.
- A pan-genome study reveals species diversity and gene families that may be of special interest, e.g. because of their role in bacterial survival or their ability to discriminate strains.

Snipen and Liland,2015

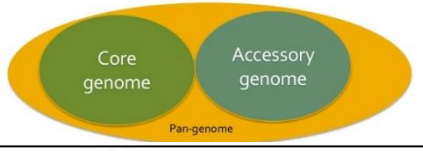
Species concept

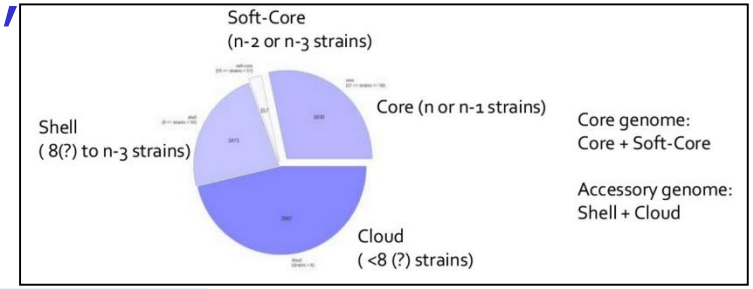
Prokaryotic genome

Bacterial pan-genome

- It has been observed that genes could be divided into classes depending on their degree of conservation within the pan-genome.
- This is the basis for the use of mixture models to predict pan-genome size.
- In the bacterial pan-genome was divided into three major categories:

1. Core (hard-core+soft-core),
2. Shell, and
3. Cloud.





Snipen and Ussery,2010;..

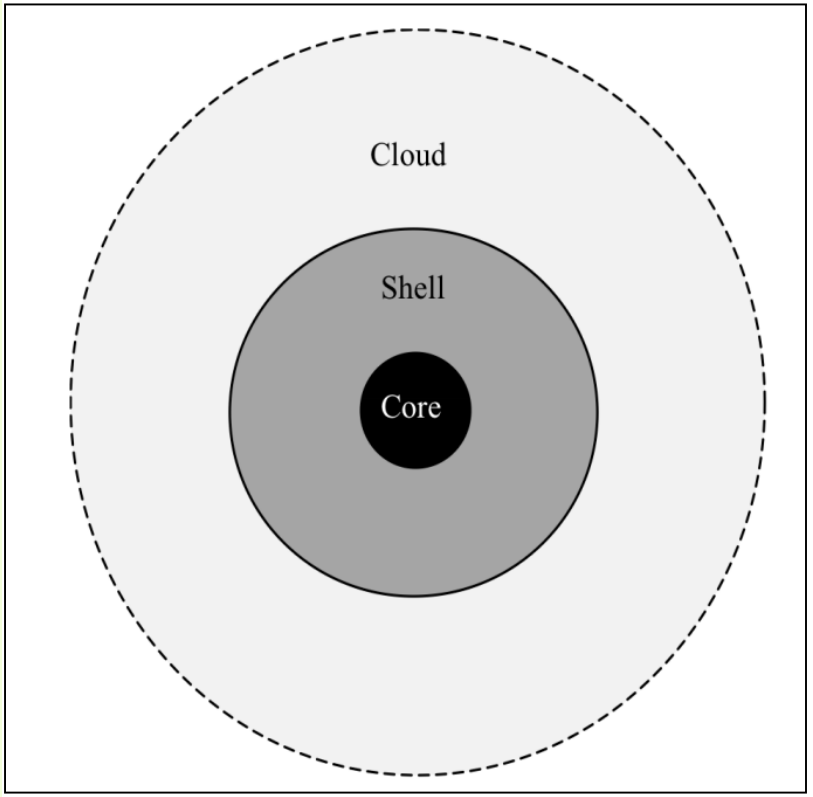
Species concept

Prokaryotic genome

Bacterial pan-genome

- The bacterial pan-genome can be divided into:

1. the core (genes always occurring in any genome inside the pan-genome)
2. the shell (genes frequently occurring), and
3. cloud (rarely occurring genes).



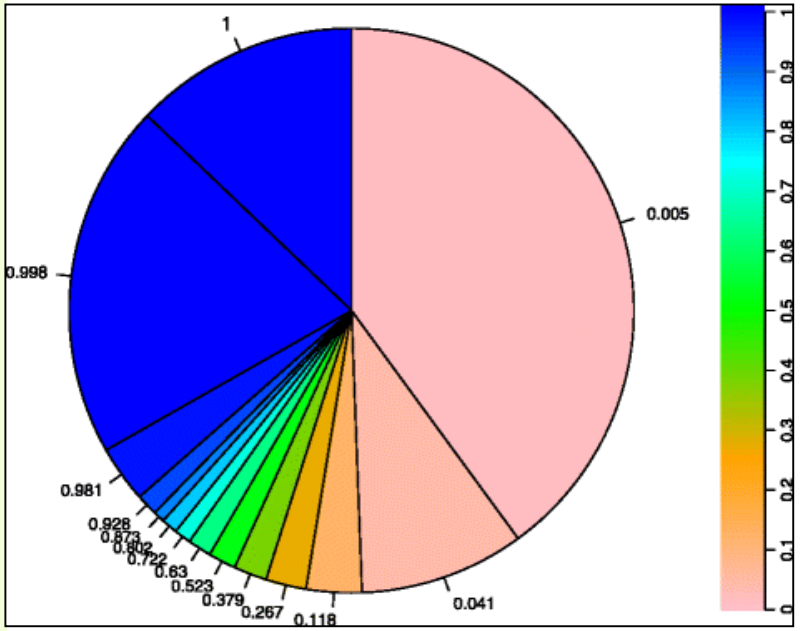
Snipen and Ussery,2010

Species concept

Prokaryotic genome

The pie symbolizes the entire pan-genome, and the 14 sectors correspond to various groups of gene clusters, where the color of the sector shows the detection probability, also indicated by the numbers

- The blue sectors are highly conserved (core) genes.
- These are found in (almost) all genomes.
- The greenish sectors are accessory (shell) genes found in (large) subsets of genomes.
- The orange/pink sectors are the rarely occurring (cloud) genes.
- Most of these have not yet been described.
- Only one or a few such genes are found in each genome, but if we keep on sequencing they will make up a large proportion of the pan-genome.



Core genes (always or almost always present), shell genes (often present, but lacking in subsets of genomes) to cloud genes (observed in only a few genomes).

Snipen and Liland,2015

Species concept

Prokaryotic genome

Bacterial pan-genome

1. Core: The core genes, i.e. the gene families present in all genomes.
2. Shell: The shell represents the genes found in the majority of the genomes, and
3. Cloud: the corresponding Cloud consists of genes only observed in a minority of the genomes.

- The size of the cloud and shell can be significantly larger than the core genome, reflecting the diversity (or lack thereof) of various types of bacteria in different ecological niches.

Snipen and Ussery,2010

Species concept

Prokaryotic genome

Bacterial pan-genome

- For example, the shell and cloud would be expected to be larger for *Actinobacteria* and other organisms that produce secondary metabolites.
- On the other hand, pathogenic, parasitic and commensal species that are not routinely found in the environment could have smaller clouds.

Snipen and Ussery,2010

Prokaryotic genome

Unique bacterial genes

- The functions of most genes, including those on extra-chromosomal elements, aren't known.
- It is estimated that each bacterium has about 40% of its genome devoted to unique genes.

Species definition in bacteria
Based on core genome

- The core genes present in the majority of the members of that species and auxiliary genes each present in some of the members.
- Therefore, the species definition in bacteria should be based on analysis of sequence variation in housekeeping genes.
- Because the core genome is less prone to horizontal gene transfer.

Prokaryotic genome

Flexible genome

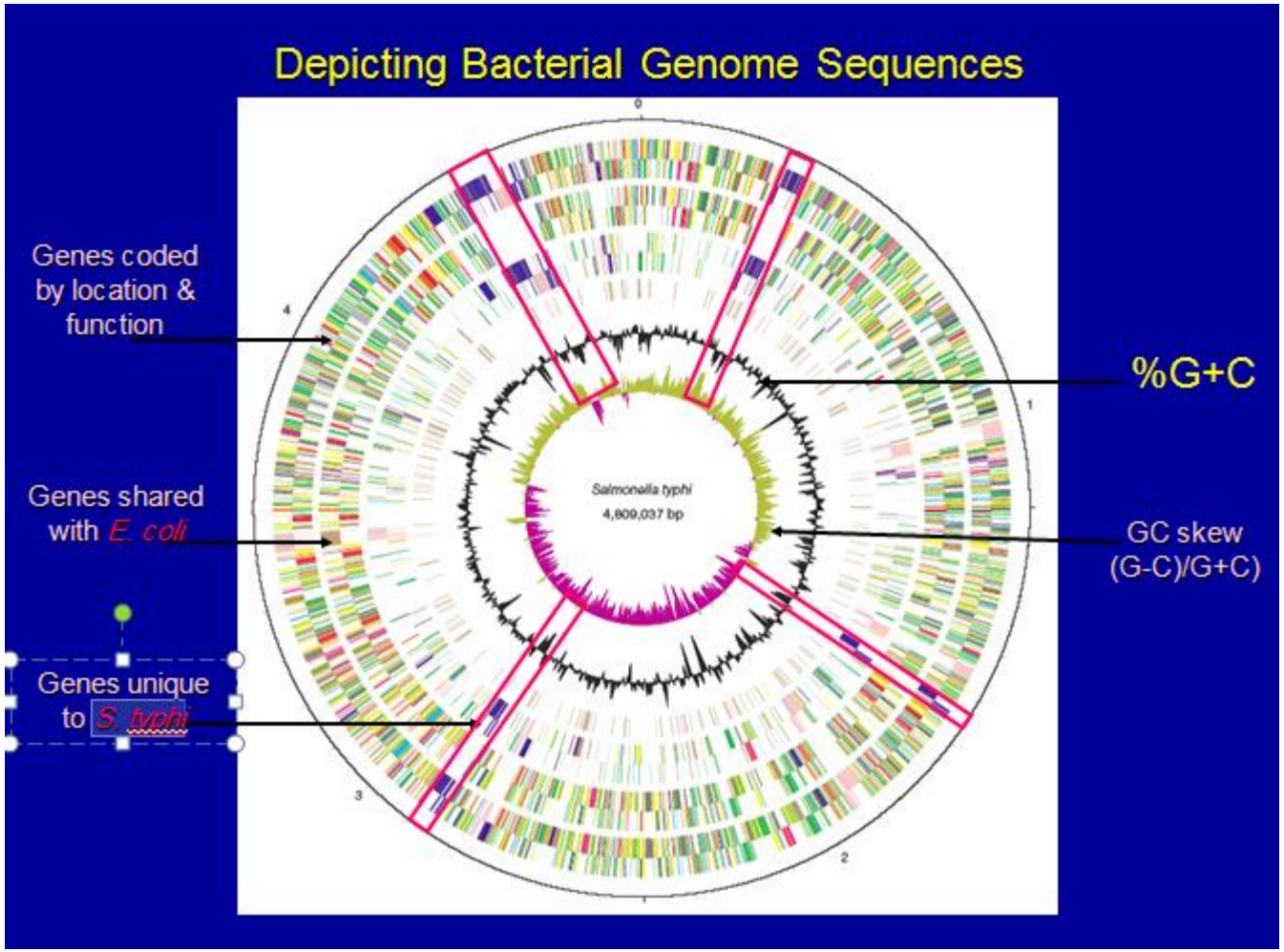
- In contrast to the core genome, the flexible genome or accessory genome largely evolves through horizontal genetic exchange.
- Flexible genome consists of genes that vary among strains within a species.

Prokaryotic genome

The core genome vs. flexible genome

- The core genome:
- Conserved among the members of a species.
- The flexible genome:
- A flexible gene pool conferring strain-, pathotype- or ecotype-specific characteristics which allow adaptation to special conditions such as colonization of specific niches or pathogenicity.

Depicting bacterial genome sequences
Shared and unique genes



Species concept

Evolution of bacterial plant pathogens

Horizontal gene transfer (HGT)

- Emergence of new plant pathogens or increase in virulence of the existing pathogens has always been a matter of great concern to plant pathologists.
- In contrast to eukaryotes, bacteria have acquired a significant proportion of their genetic diversity through the acquisition of DNA sequences from distantly related organisms.
- **With recent estimates suggesting that on average 81% of prokaryotic genes have been involved in HGT at some point.**

Dagan *et al.*,2008; Thind,2013

Species concept

Horizontal gene transfer (HGT)

Prokaryotes vs. eukaryotes

- Although observed rates of acquisition of horizontally transferred genes in eukaryotes are generally lower than in prokaryotes, it appears that, far from being a rare occurrence, HGT has contributed to the evolution of many, perhaps all, animals and that the process is ongoing in most lineages.
- Between tens and hundreds of foreign genes are expressed in all the animals we surveyed, including humans.
- The majority of these genes are concerned with metabolism, suggesting that HGT contributes to biochemical diversification during animal evolution.

Dagan *et al.*,2008

Species concept

Evolution of bacterial plant pathogens

Horizontal gene transfer (HGT)

- HGT is very important in the sense that it may cause almost instant emergence of new pathogens, in comparisons to the hundreds or even thousands of years needed for emergence of new pathogens by other mechanisms.

Thind,2013

Species concept

Evolution of bacterial plant pathogens

Horizontal gene transfer (HGT)

- HGT is an important driving force for:

1. the evolution of bacterial genome;
2. rapid adaptation to changing environments;
3. pathoadaptive mutations, which increase virulence by means of minor genetic alterations such as point mutation, and gene rearrangement or deletion, also contribute to the evolution of virulence.

Thind,2013

Species concept

Evolution of bacterial plant pathogens

Example: *Pantoea agglomerans*

- This bacterium is an epiphytic bacterium associated with many plants, was transformed into a host-specific gall-forming pathogen on gypsophila and beet.
- Barash and Manulis-Sasson (2009), in their recent review article, have given an excellent account of transformation of *Pa. agglomerans* into an hrp-dependent and host-specific pathogen.
- The transformation was achieved by the evolution of a pathogenicity plasmid that harbours most, if not all of the genes, required for gall formation and host specificity.

Thind,2013

Species concept

Evolution of bacterial plant pathogens

Example: *Streptomyces* spp.

- The emergence of new plant pathogenic species of *Streptomyces* suggests that the pathogenicity island of *Streptomyces* spp. is moblizable, at least in some strains.
- The appearance of *St. turgidiscabies* in Japan is the latest example of a newly emerged species (Loria *et al.*,2006).

Pathogenicity islands (PAIs) and antimicrobial resistance islands (REIs) are key to the evolution of pathogens.

1. PAIs promote disease development, and
2. REIs give a fitness advantage to the host over multiple antimicrobial agents.

Thind,2013; PAIDB v2.0

Prokaryotic genome

The core genome vs. flexible genome

Core genome	Flexible genome
Essential for cell viability	Non-essential
Genes are stable	Variable
No selection pressure for horizontal acquisition	> Selection pressure
Can't undergo large indels	Large indels possible
Strong phylogenetic signal	Can be weak

Indel: An insertion or deletion occurring in a protein or nucleic acid sequence.

Methods for generation of mutant populations

Indels

- The most reliable method to ascertain gene function is:

1. to disrupt the gene, and
2. determine the phenotype change in the resulting mutant individual.

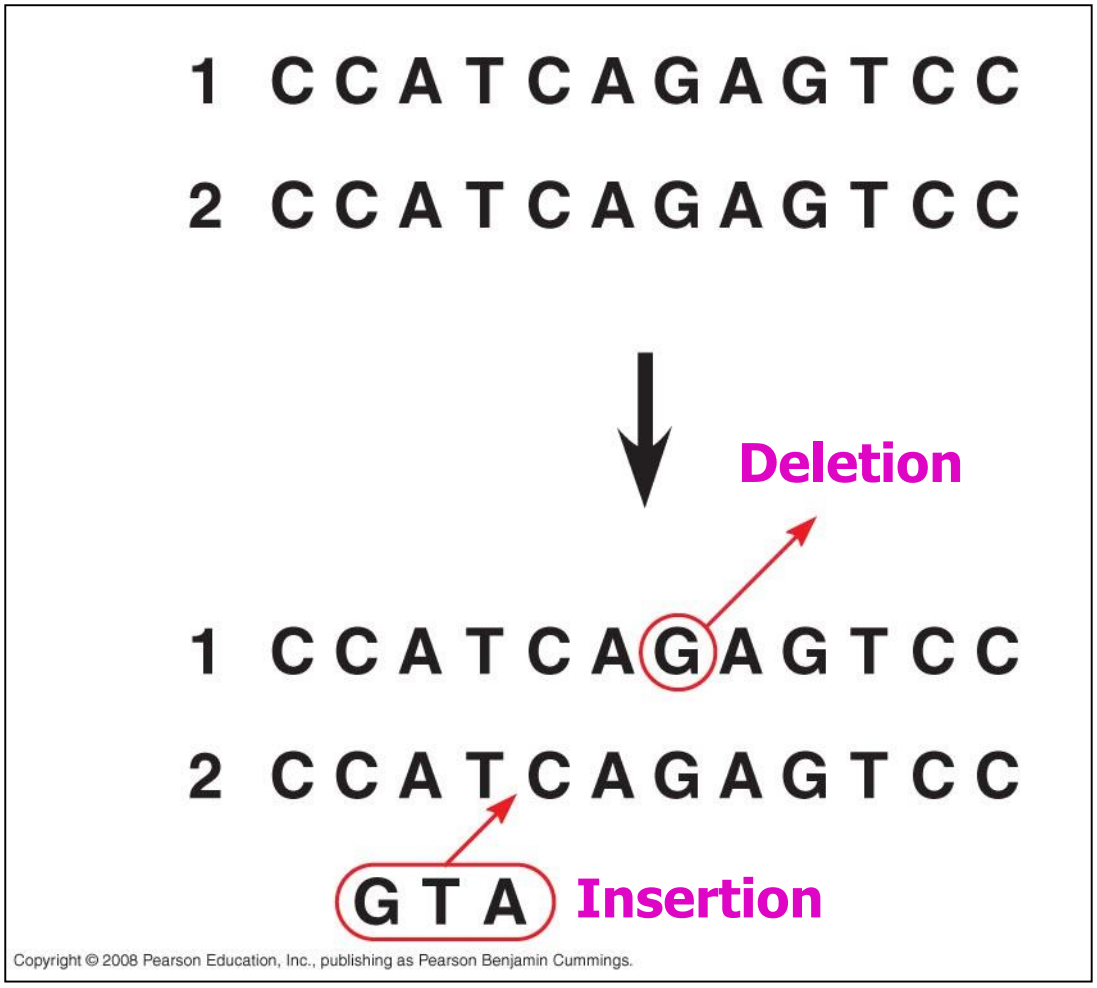
- Two most popular methods to generate mutants:

1. Insertional mutagenesis;
2. Deletion mutagenesis.

Prokaryotic genome

DNA sequence data

Most important type of data

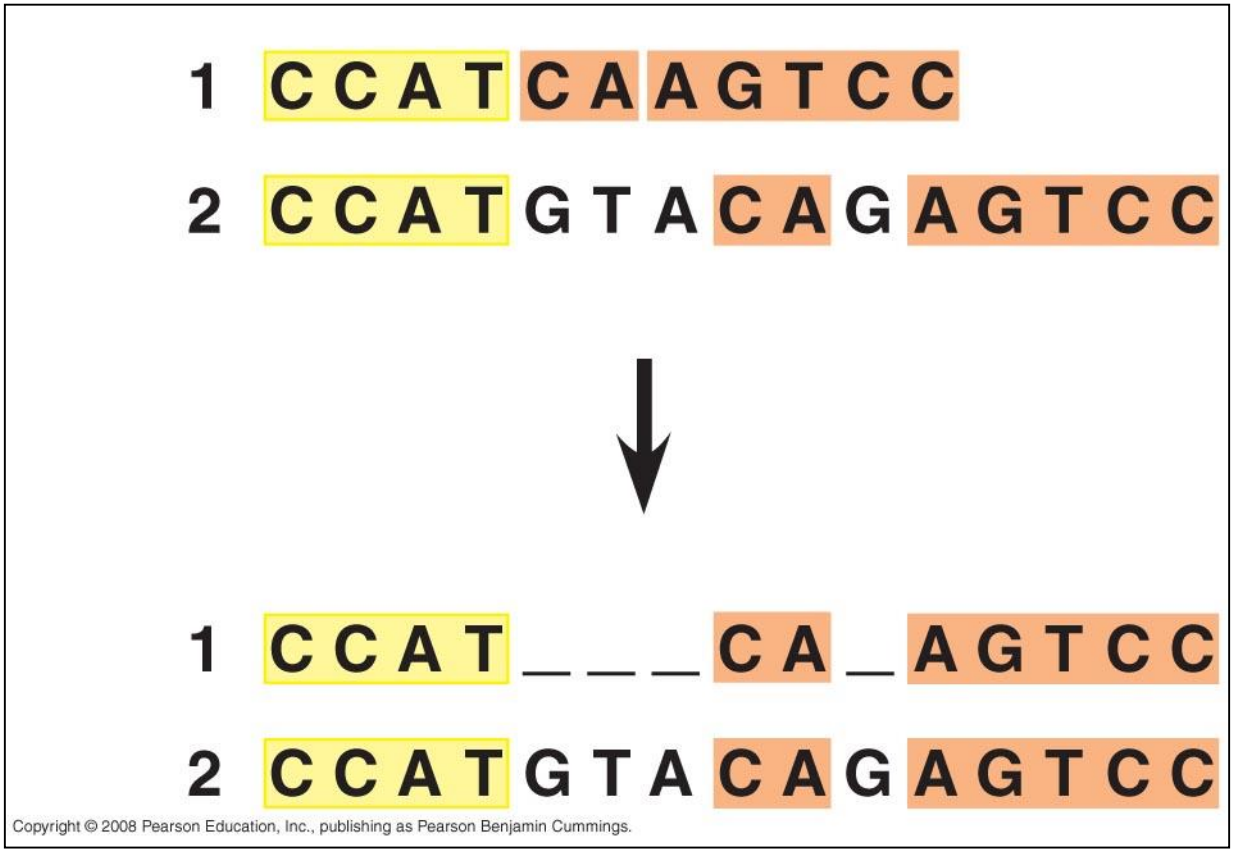


Indel: An insertion or deletion occurring in a protein or nucleic acid sequence.

Prokaryotic genome

DNA sequence data

Alignment



Each nucleotide position = Character
Character states = specific nucleotide

Prokaryotic genome

Gene flow: gene loss and acquisition

Genome evolution

- The process by which the content and organization of genetic information of a species changes over time is known as genome evolution.
- Mechanisms of horizontal gene flux include mobile genetic elements such as:
- Conjugative plasmids,
- Bacteriophages,
- Transposons,
- Insertion elements, and
- Genomic islands(GIs),
- Recombination of foreign DNA into host DNA.

Schmidt and Hensel,2004

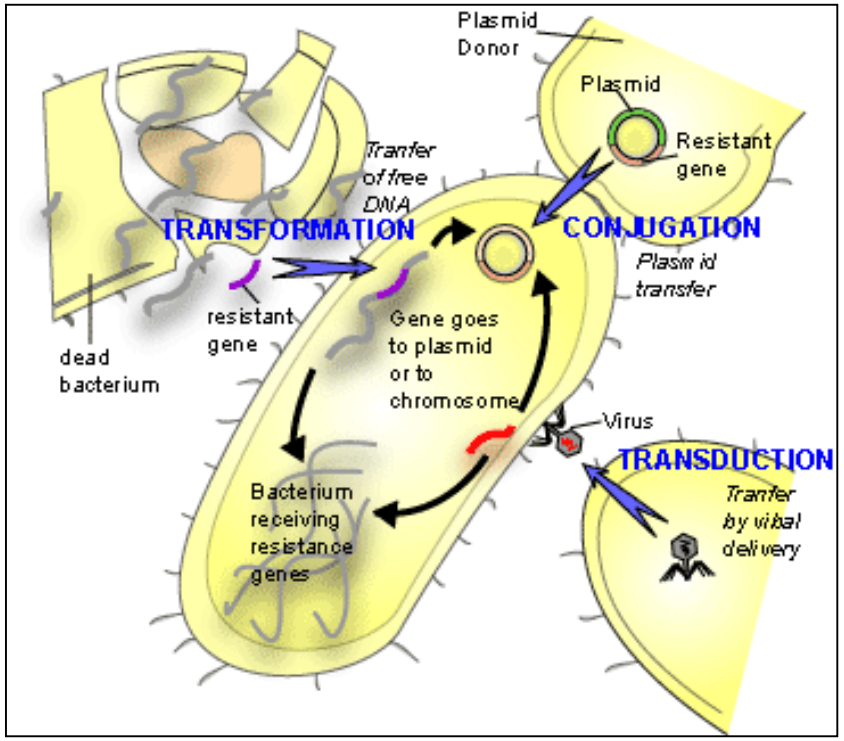
Gene flow

Horizontal gene transfer

Horizontal transfer, lateral gene transfer

- Lateral or horizontal gene transfer (HGT) is a process whereby genetic material contained in small packets of DNA can be transferred between individual bacteria.
- There are three possible mechanisms of HGT.
- These are:

1. Transduction,
2. Transformation,
3. Conjugation.



Gene flow

Horizontal gene transfer

Horizontal transfer, lateral gene transfer

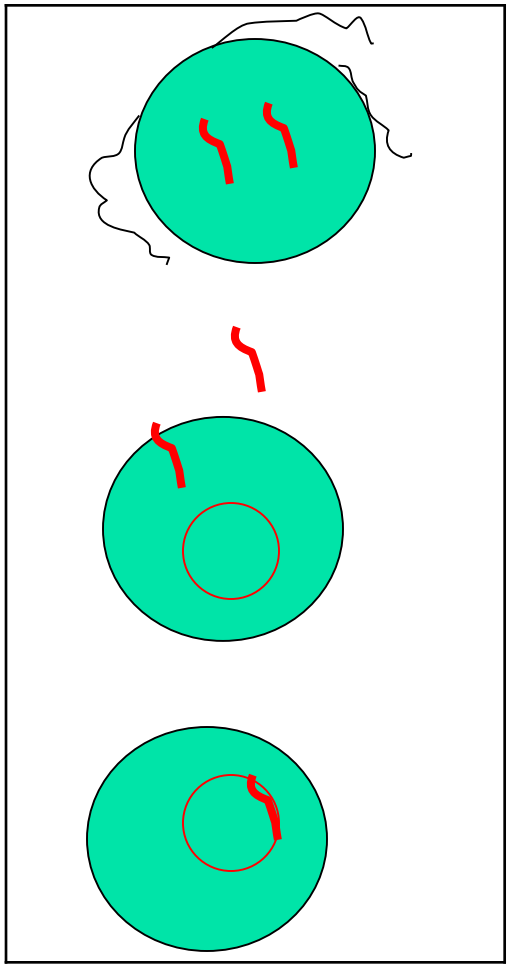
- Various mechanisms of gene transfer:

1. Transformation
2. Conjugation
3. Transduction

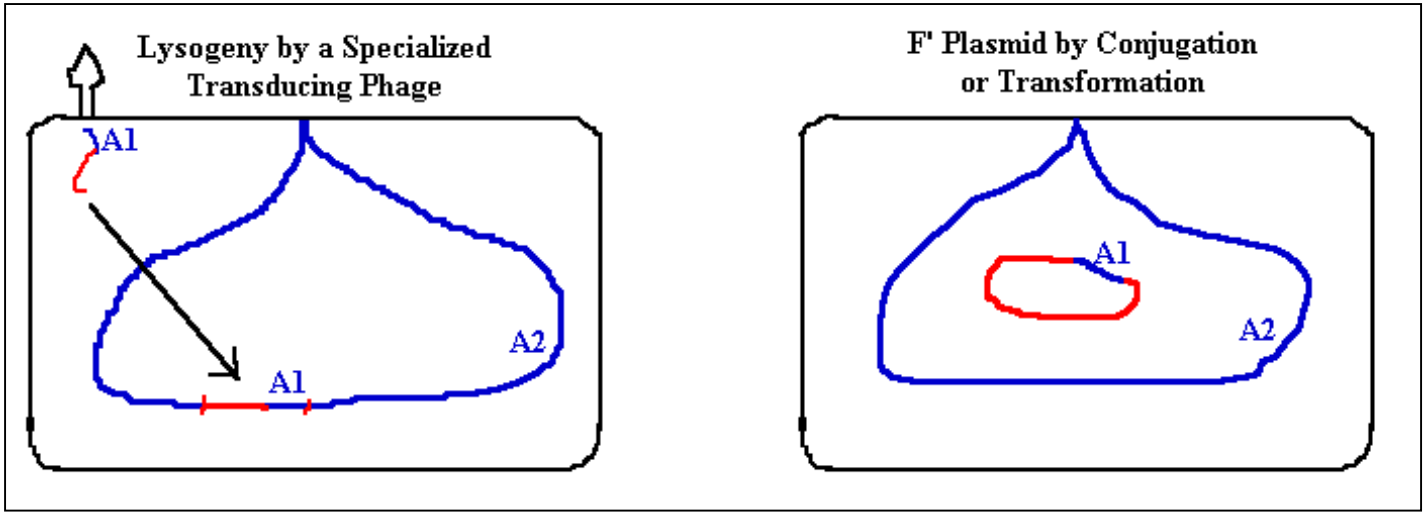
- Donor remains unchanged but recipient genetically modified.
- Therefore, Donor/recipient can be taxonomically completely unrelated.

Transformation

- Free DNA may become available from dying or disintegrating bacteria.
- This can enter a living bacterial cell and join into its chromosome.
- The process by which free DNA becomes a part of the chromosome is called recombination.



Comparison of
Transformation & Transduction



Two processes to create a partial diploid in bacteria are **specialized transduction** and **F' conjugation**.

Conjugation

- Conjugation is considered a major pathway for horizontal (or lateral) gene transfer among bacteria.
- Conjugation requires cell-to-cell contact and operates by DNA replication resulting in unidirectional transfer of genetic material from a donor to a recipient cell.
- It is mediated mainly by conjugative plasmids, although conjugative transposons are also capable of triggering the process of conjugation.

Dionisio *et al.*,2002

Conjugation

Conjugative plasmids

- Plasmids that can be transferred via conjugation is called conjugative plasmids.
- The transfer of chromosomal genes of bacteria may or may not occur.
- Conjugative plasmids can mediate gene transfer between bacterial taxa in diverse environments such as soil and rhizosphere.
- Since bacteria share the same habitat with several other bacterial species, conjugative plasmid can infect different coexist bacterial species.
- Many carrying conjugative plasmids can transfer virulence and antibiotic-resistance genes between pathogenic and nonpathogenic bacteria.

Conjugative plasmids
Mediators of DNA transfer

- However, the ability conjugative plasmids to pass between different bacterial strains or species has been considered to be of lower efficiency than their ability to pass between similar bacteria, due to diverse barriers such as restriction systems.

Dionisio *et al.*,2002

Ti and *Sym* plasmids
Agrobacterium and *Rhizobium*

- *Agrobacterium* is closely related to *Rhizobium*.
- Most of the genes involved in crown gall disease are not borne on the chromosome of *A. tumefaciens* but on a large plasmid, termed the T_i (tumour-inducing) plasmid.
- In the same way, most of the genes that enable *Rhizobium* strains to produce nitrogen-fixing nodules are contained on a large plasmid termed the *Sym* (symbiotic) plasmid.
- Thus, the characteristic biology of these two bacteria is a function mainly of their plasmids, not of the bacterial chromosome.

Ti and *Sym* plasmids
Agrobacterium and *Rhizobium*

- The central role of plasmids in these bacteria can be shown easily by "curing" of strains.
- If the bacterium is grown near its maximum temperature (about 30°C in the case of *Agrobacterium* or *Rhizobium*) then the plasmid is lost and pathogenicity (of *Agrobacterium*) or nodule-forming ability (of *Rhizobium*) also is lost.
- However, loss of the plasmid does not affect bacterial growth in culture.
- The plasmid-free strains are entirely functional bacteria.

Ti and *Sym* plasmids
Agrobacterium and *Rhizobium*

- In laboratory conditions it is also possible to cure *Agrobacterium* or *Rhizobium* and then introduce the plasmid of the other organism.
- Introduction of the T_i plasmid into *Rhizobium* causes this to form galls; introduction of the *Sym* plasmid into *Agrobacterium* causes it to form nodule-like structures, although they are not fully functional.

Challenging questions about the nature of bacteria

- Studies such as these raise many interesting and challenging questions about the nature of bacteria.
- For example

1. What does the name of a bacterial species or genus really mean, if the organism can change so drastically by loss or gain of a non-essential plasmid?
2. How much gene exchange occurs by means of plasmids and other mobile genetic elements within natural populations?

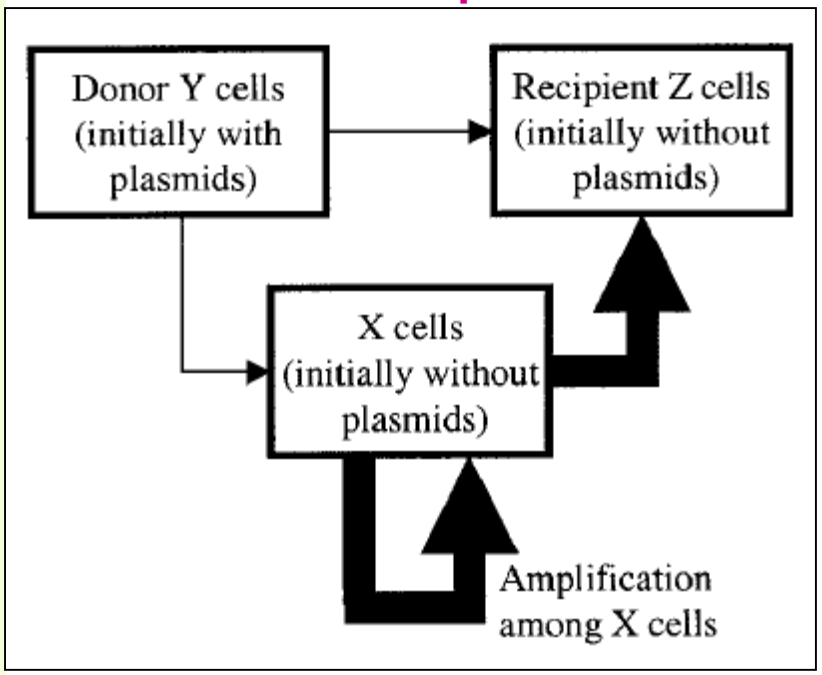
Conjugative plasmids

Mediators of DNA transfer

- Consider three bacterial populations, X, Y, and Z, living in the same habitat.
- Suppose that Y cells bear a conjugative plasmid.
- The arrows represent the conjugation events:
- Larger arrows represent higher efficiency of the plasmid transfer.

1. The plasmid from Y is going to infect both X and Z plasmid-free cells.
2. If the conjugation rate among X cells is high, the plasmid number among them will amplify.
3. Following this, plasmids from X cells will be massively transferred to plasmid-free Z cells.

The model for the amplification effect



Dionisio *et al.*,2002

Prokaryotic genome

Horizontal gene transfer

Pathogenicity islands

- The acquisition of genes by lateral transfer is a major factor in the adaptation of clones and in bacterial evolution over all time-frames from generation of clonal diversity.
- Many pathogenic properties are encoded on plasmids, phages or pathogenicity islands, and the transfer of such properties is widely thought to be associated with the origin of pathogenic clones.

For more details please see the Bacterial Pathogenesis-Part II.

Prokaryotic genome

Horizontal gene transfer

Pathogenicity islands

- Pathogenicity islands (PAIs) is a subgroup of genomic islands.

1. **Genomic islands** are present in the majority of genomes of pathogenic as well as nonpathogenic bacteria and may encode many different functions, which depend largely on the environmental context in which the bacterium grows.
2. **Pathogenicity islands** comprise large genomic regions [10-200 kilobases (kb) in size] that are present on the genomes of pathogenic strains but absent from the genomes of nonpathogenic members of the same or related species.

- PAIs carry clusters of virulence genes whose products contribute to the pathogenicity of the bacterium.

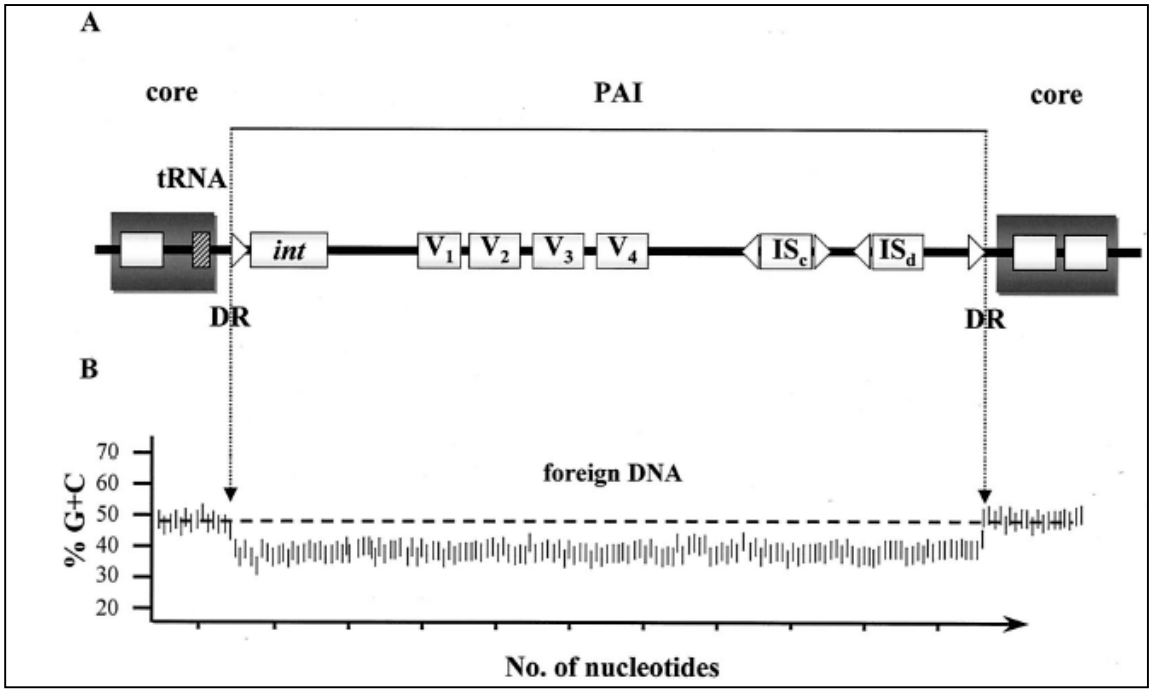
Prokaryotic genomes
Horizontal gene transfer
Pathogenicity islands

- PAIs are often associated with tRNA loci, which may represent target sites for the chromosomal integration of these elements.
- Pathogenicity islands have particular genetic elements such as insertion sequences or tRNA genes, which are sites for recombination into the DNA.
- Cryptic mobility genes may also be present, indicating the provenance as transduction.
- These particular genetic elements were previously able to spread among bacterial populations by horizontal gene transfer, a process known to contribute to microbial evolution.

Hacker *et al.*,2008;…

General structure of PAI

- The PAI boundaries are frequently determined by DRs (direct repeats with perfect sequence repetition (triangle), which are used for insertion and deletion processes.



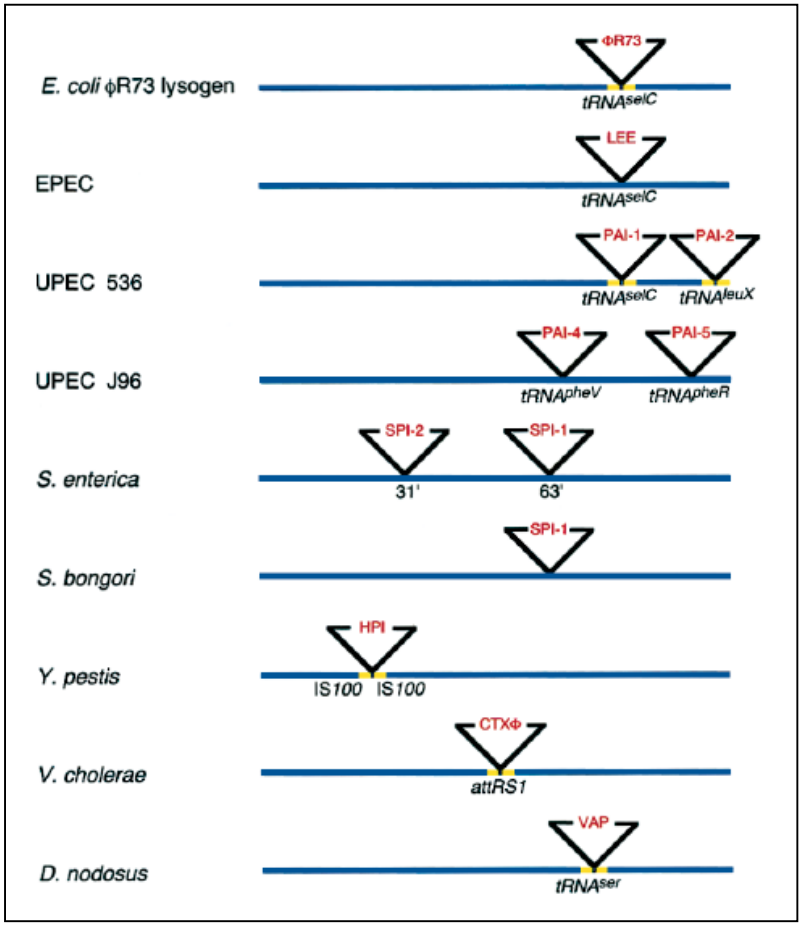
Relatively low GC contents (below 40%) are assembled in a "pathogenicity island"

Schmidt and Hensel,2004

Pathogenicity islands

Location of selected pathogenicity islands and phages of Gram-Negative Bacteria

- Chromosomes (blue lines) and pathogenicity islands (black triangles) are not depictured.
- The presence of repeated sequences at the site of insertion, which is shown below the island, is indicated by short yellow lines.



Groisman and Ochman,1996

Pathogenicity islands

Speciation of pathogens

- PAI apparently have been acquired during the speciation of pathogens from their nonpathogenic or environmental ancestors.
- All knowledge about PAI, their structure, their mobility, and the pathogenicity factors they encode will be helpful:

1. In gaining a better understanding of bacterial evolution;
2. Interactions of pathogens with eukaryotic host cells;
3. Providing delivery systems for vaccination,
4. Tools for cell biology and for the development of new strategies for therapy of bacterial infections.

Schmidt and Hensel,2004

Bacterial genomes

Genomic islands

Pathogenicity islands

- Pathogenicity islands were first described in human pathogens of the species *Esehericchia coil*, but have recently been found in the genomes of various pathogens of humans, animals, and plants.
- These regions generally have a different G+C content than their host genome (below 40%), suggesting acquisition via horizontal gene transfer (HGT).
- It is possible to recognize genes that arose by lateral gene transfer by simply examining genome sequences.

Kjemtrup *et al.*,2000; Buonaurio,2008

Bacterial genomes

Genomic islands

Pathogenicity islands

- PAIs comprise large DNA regions (up to 200 kb of DNA) and often carry more than one virulence gene, the G+C contents of which often differ from those of the remaining bacterial genome.
- The base composition is expressed as percentage of guanine and cytosine (G+C) bases, and the average G+C content of bacterial DNA can range from 25 to 75%.
- Most pathogenic bacterial species have G+C contents between 40 and 60%.

Schmidt and Hensel,2004; Hacker *et al.*,2008

General structure of PAI

Properties of PAIs

PATHOGENICITY ISLANDS

. Segments of the chromosome harboring large clusters of virulence genes

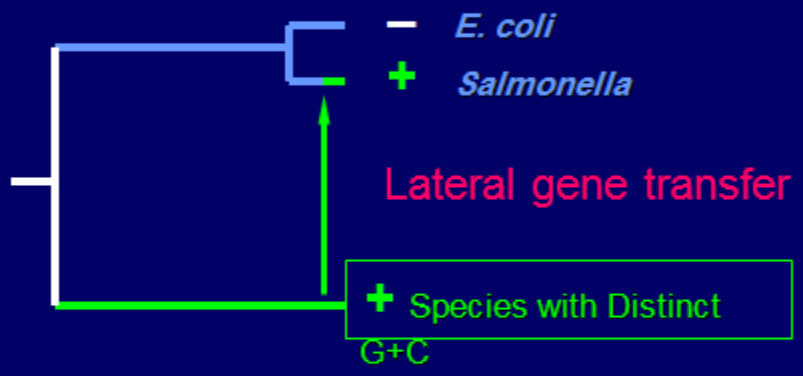
. Present in pathogenic strains but absent or sporadically distributed in related non-pathogenic species

. Typically have a G+C content different from that of the rest of the chromosome

. Often associated with tRNA genes and/or mobile genetic elements at their boundaries

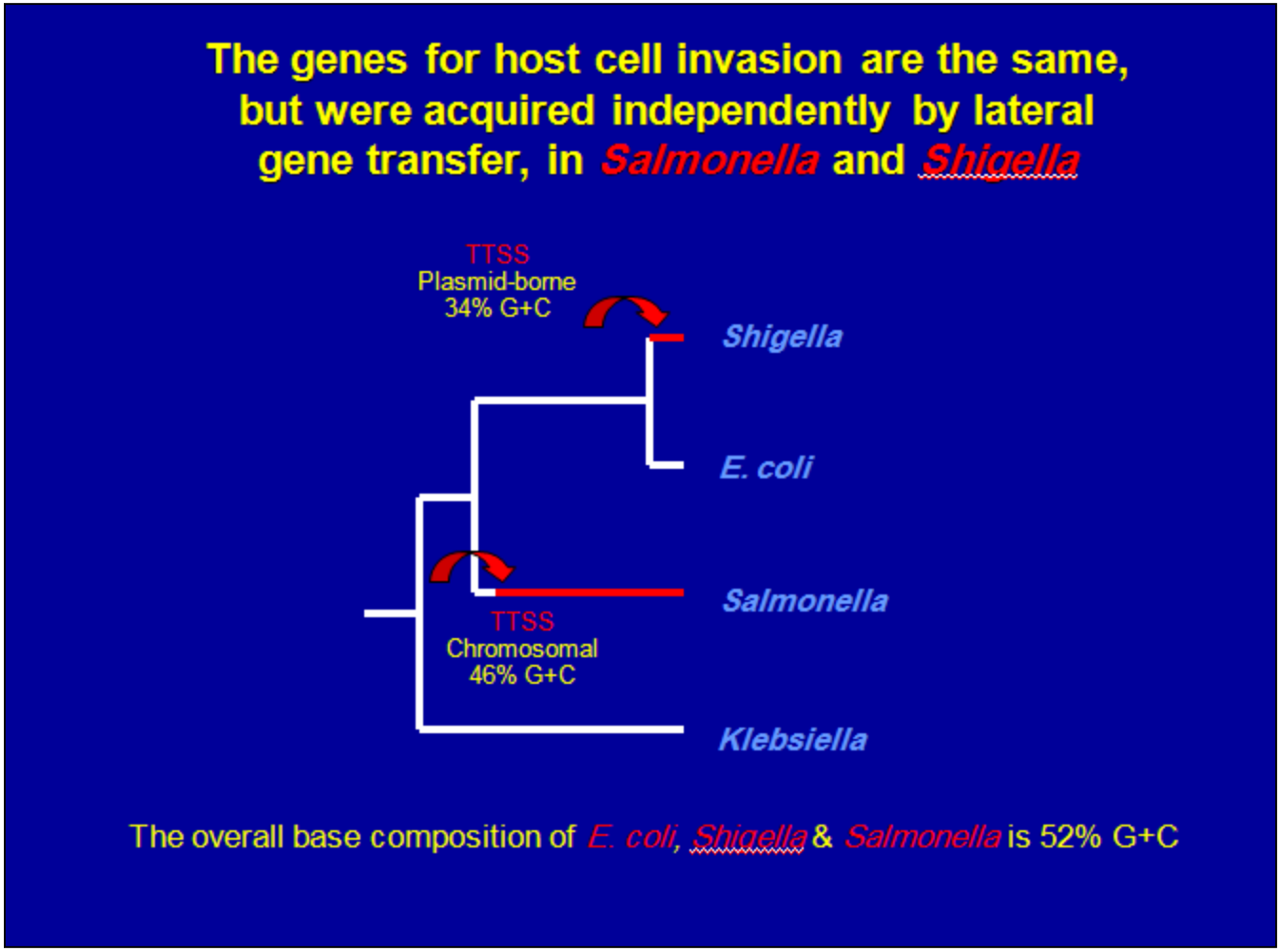
Pathogenicity islands
Horizontal gene transfer

Why do pathogenicity islands have atypical G+C contents?



Lateral gene transfer is the source of "atypical" &"species-specific" genes

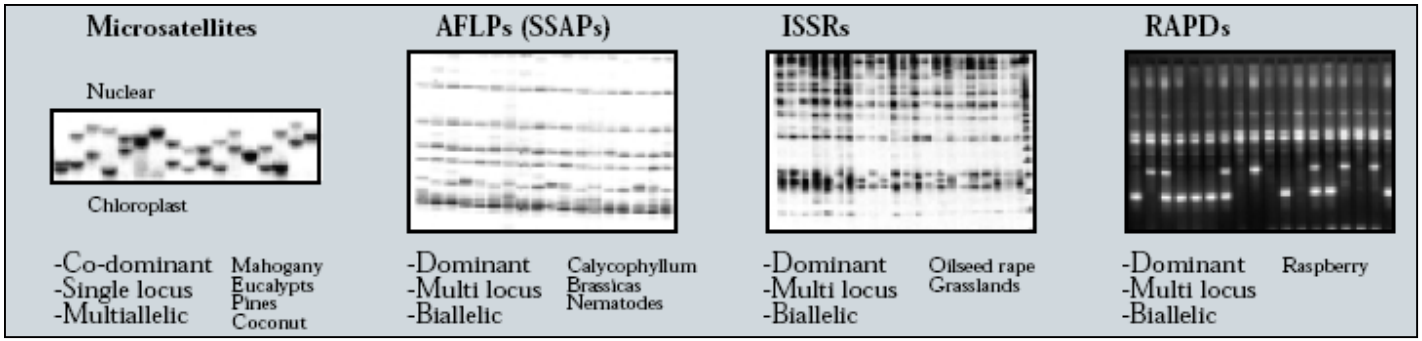
Pathogenicity islands
Horizontal gene transfer



Gene flow

Some PCR-based genetic technologies to assess gene flow and the spatial organization of genetic variation in plants/bacteria

- Some genetic technologies to assess gene flow and the spatial organization of genetic variation in plants.
- To use these techniques, a reasonable amount of tissue (mg) must be available.
- Therefore to generate this amount, the bacteria must be able to be grown in laboratory conditions.

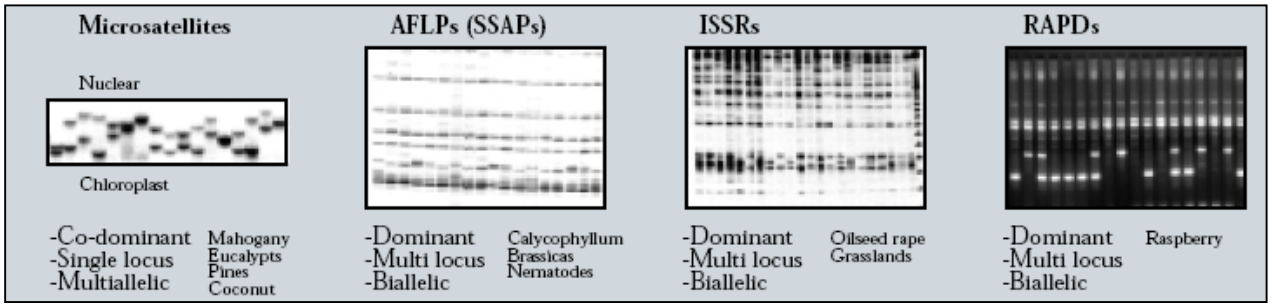


Gene flow

High resolution molecular fingerprinting techniques

Microsatellites (SSR) and ISSR

- **Microsatellites, also known as Simple Sequence Repeats (SSRs)**: A repeating sequence in DNA. Repeating sequences of 1-6 base pairs of DNA.
- Microsatellites are typically neutral and co-dominant. They are used as molecular markers in genetics.
- These markers often present high levels of inter- and intra-specific polymorphism, particularly when the number of repetitions is 10 or greater.
- **ISSR (inter-simple sequence repeat)**: A general term for a genome region between microsatellite loci. Sequences amplified by ISSR-PCR can also be used for DNA fingerprinting.



Gene flow

High resolution molecular fingerprinting techniques

Microsatellites

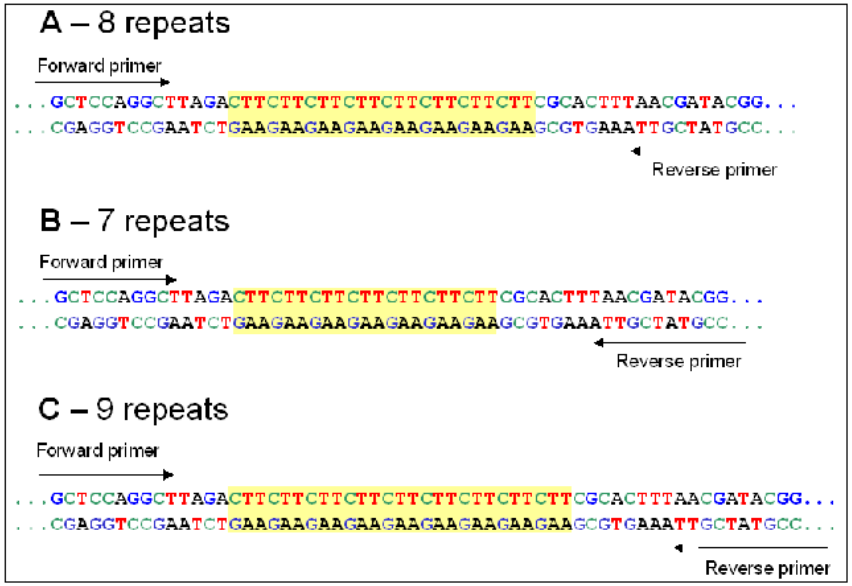
- The sequence surrounding the repeat region is usually conserved, allowing PCR primers to be designed so that the repeat region and a short flanking sequence can be amplified.
- Individuals may differ in the number of repeats present, meaning that the length of the PCR product varies.
- Products can be scored using various forms of high resolution electrophoresis (polyacrylamide or semi automated sequencers).
- Used as molecular markers, they are co-dominant, often highly polymorphic and relatively easy to score.
- The high polymorphism of microsatellites is attributed to relatively high rates of error during DNA replication (slippage) and during recombination (unequal crossover).

Gene flow

High resolution molecular fingerprinting techniques

Microsatellites

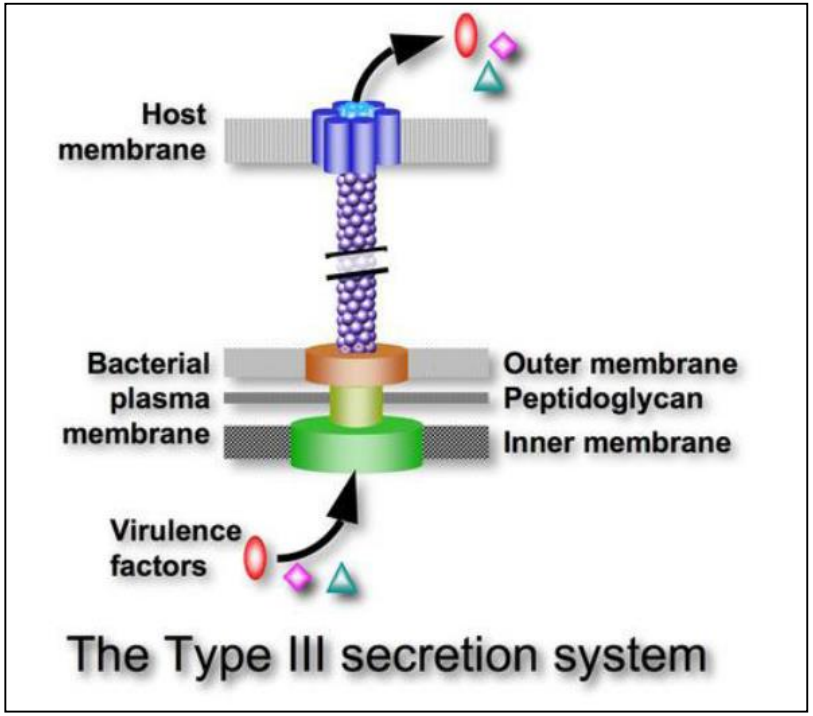
- Synonyms :
- SSR (Simple Sequence Repeat); or
- STR (Short Tandem Repeat).
- Microsatellites are sections of DNA composed of repeats of short motifs (e.g. CA, GTG, TGCT etc) arranged in tandem.



King,2006

Type III secretion system
Structure

- Injection apparatus, together with effector proteins,for parasitising euckaryotic cells.
- Required for by proteobacteria to produce infection of plants and animals.
- Required for colonization and plant defence activation in growth promoting strains.



The Type III secretion system

Neil Parkinson

Type III secretion system
Highly integrated component proteins

- Type III secretion (TTS) systems have been discovered in many Gram-negative bacteria that are pathogenic for animals, plants and even insects.
- These secretion pathways contain highly integrated component proteins, conserved across many bacterial species, and which consist of about 20 or more genes.
- The type III pathway has been found to secrete a number of different proteins in these bacteria, in particular virulence and pathogenicity factors.

users.ugent.be/~tcoenye/epa.pdf;...

Type III secretion system

Distribution of type III secretion systems among saprophytic/nonpathogenic/symbiotic bacteria

- Type III protein secretion systems (TTSSs) have recently been reported in non-pathogenic bacteria.
- In the symbiotic bacteria such as:
- *Rhizobium* sp. *parasponia*
- *Bradyrhizobium japonicum*.
- In five saprophytic species such as *P. fluorescens* and *P. putida*.

Mazurier *et al.*,2004

Type III secretion system

Distribution of type III secretion systems among pathogenic bacteria

- TTSS is employed for suppression of plant defence systems.
- Comparing the genes and their organization, TTS systems can be grouped in five 'families' (see figure).
- Interestingly, some bacteria harbour not only one system, but two from different families (*Salmonella enterica*, *Yersinia enterocolitica* from biogroup 1B, *Y. pestis* and *Burkholderia*).

Arnold and Jackson,2011;Gophna *et al.*,2003

Type III secretion system
T3SE evolution

- The genome sequences have revealed a considerable variability in T3SE numbers per strain and a large number of T3SE families.
- A comparison of *P. syringae* pv.*tomato* strainT1 against strain DC3000 found that the genomes were highly similar except for theT3SE repertoires, which had diverged significantly.
- DC3000 is a pathogen of both tomato and *Arabidopsis thaliana* whereas T1 is pathogenic on tomato but not on *A. thaliana*.
- So the differences in their respective T3SE repertoires may help to explain the different host range of these strains.

Arnold and Jackson,2011

Type III secretion system
Phylogenetic analysis

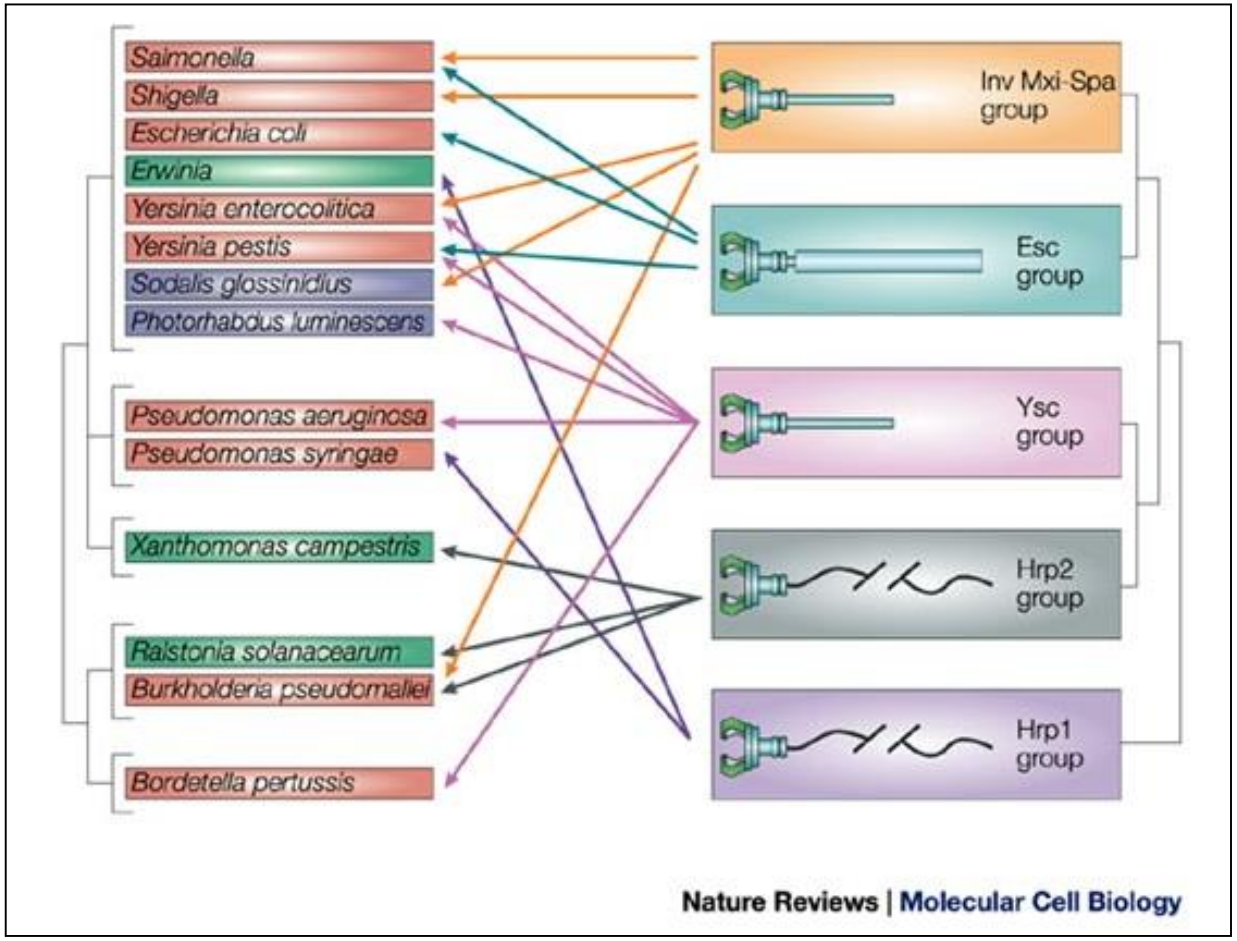
- A recent phylogenetic analysis showed that the evolutionary tree of TTS systems differs completely from the phylogeny of their bacterial host, and hence that they have been distributed in bacterial populations by horizontal transfer.
- They are generally found:

1. on plasmids (*Yersinia*, *Shigella*), or
2. on discrete 'pathogenicity islands' in the bacterial chromosome.

Cornelis,2002

A phylogenetic tree of a TTS system (right) compared with the bacterial phylogenetic tree based on 16S RNA

Plant pathogens are shown in green



Cornelis,2002

Type III secretion system
TTSS protein sequences

- It has often been suggested that TTSS genes evolved from genes encoding flagellar proteins.
- Sequence similarity confirm this high degree similarity between TTSS proteins and flagellar proteins.
- Few intriguing(interesting) questions present themselves:
- Did TTSS originate from flagella?
- Did TTSS genes evolve first in plant-pathogens as an adaptation of the flagellar basal body.

Gophna *et al.*,2003

Phylogenetic relationships of type III secretion proteins with flagellar proteins

- Protein sequences used in this study were variable including SctN/FliI, SctV/FlhA, SctR/Flip and SctS/FliQ homologs.
- Our TTSS phylogenetic trees reveal no clear division between bacteria of mammals and plants, thus, lending no support for the assumption that the TTSS emerged first in plant bacteria.

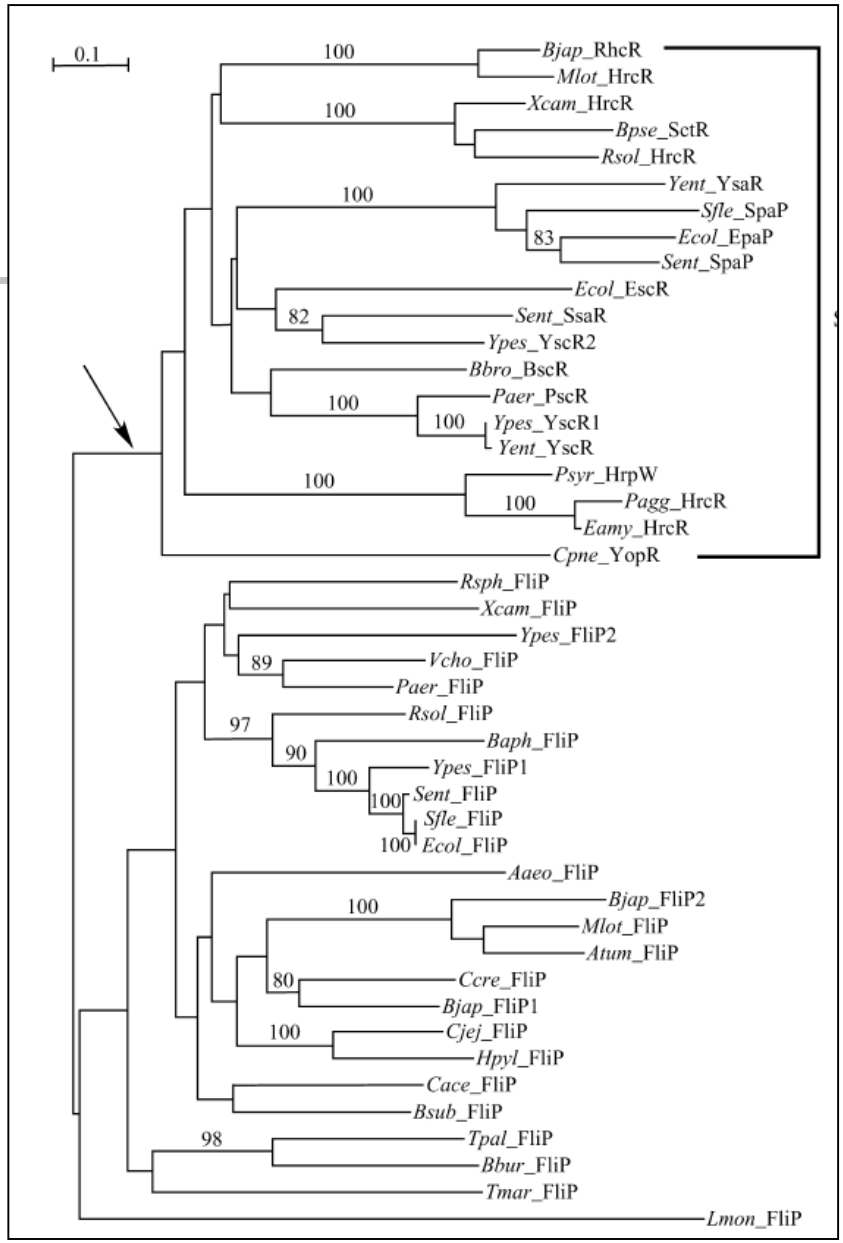
Gophna *et al.*,2003

Phylogenetic relationships of type III secretion proteins with flagellar proteins

- In *Pseudomonas*, two different species have TTSS clusters that are highly dissimilar, each adapted to its own host.
- The mammalian pathogen *P. aeuroginosa* and the plant pathogen *P. syringae* have SctN orthologs that are phylogenetically related to those of Enterobateriaceae.

Gophna *et al.*,2003

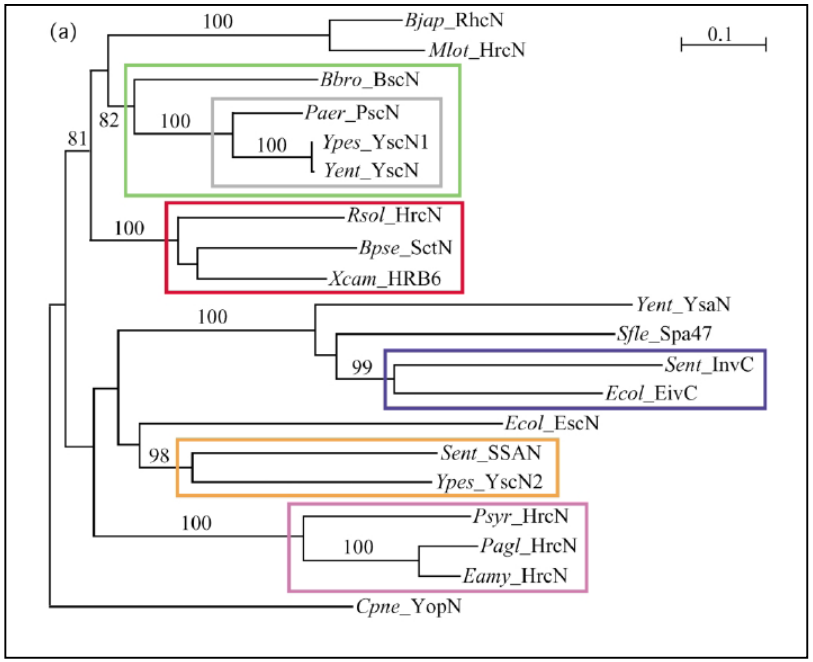
- Unrooted neighbor-joining phylogenetic trees of proteins (SctR/FliP) from the flagellar export and type III secretion systems.
- The arrows indicates that it is possible to position the roots.



Gophna *et al.*,2003

Neighbor-joining trees for SctN protein sequences

- Unrooted neighbor-joining phylogenetic trees of proteins (SctR/FliP) from the flagellar export and type III secretion systems.
- The arrows indicates that it is possible to position the roots.

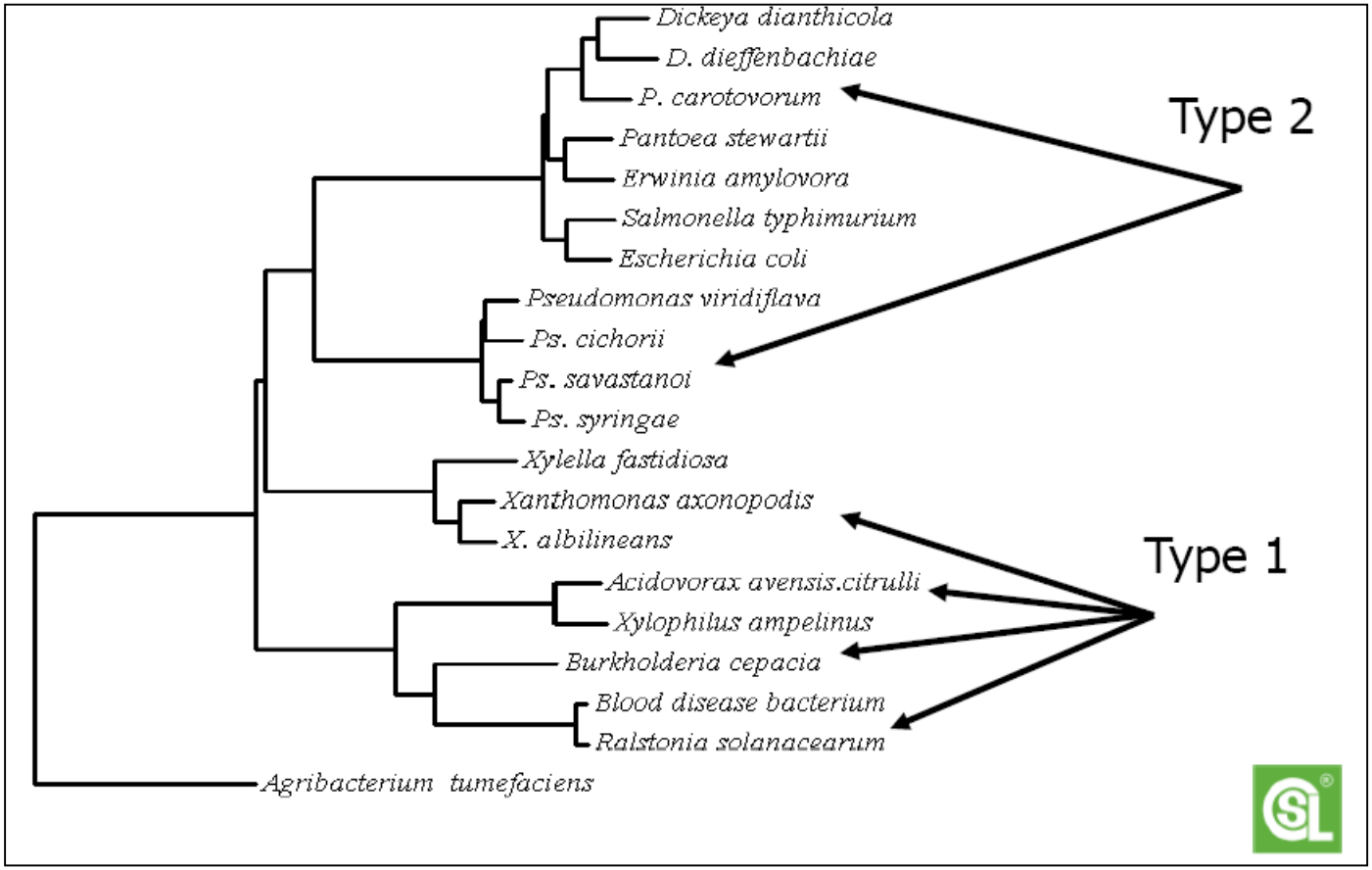


Gophna *et al.*,2003

Type three secretion system

TTSS phylogenetic trees

Plant pathogenic proteobacteria



Neil Parkinson

Polyphasic Taxonomy
Species definition

Current bacterial species definition based on:

1. **Phenotype,**
2. **Ribotype, and**
3. **Genotype information.**

The Species Concept in Prokaryotes
What is a bacterial species?

- The taxonomic hierarchy stands on the basic category "species."
- This unit is considered to be the unique real entity of the whole classification schema, whereas all other higher categories are considered to be abstract.
- The search for a universal species concept has led to heated debates, and establishing universal categories for all living organisms, including prokaryotes, has created general dissatisfaction.

Rosselló-Mora,2005

Definition of species
Definition of "species" in microbiology

- Classic definition: A collection of microbial strains that share many properties and differ significantly from other groups of strains.
- Species are:

1. Identified by comparison with known "type strains";
2. Well-characterized pure cultures;
3. References for the identification of unknowns.

- There are several collections of type strains, including the American Type Culture Collection (ATCC).

Hendrix,2004

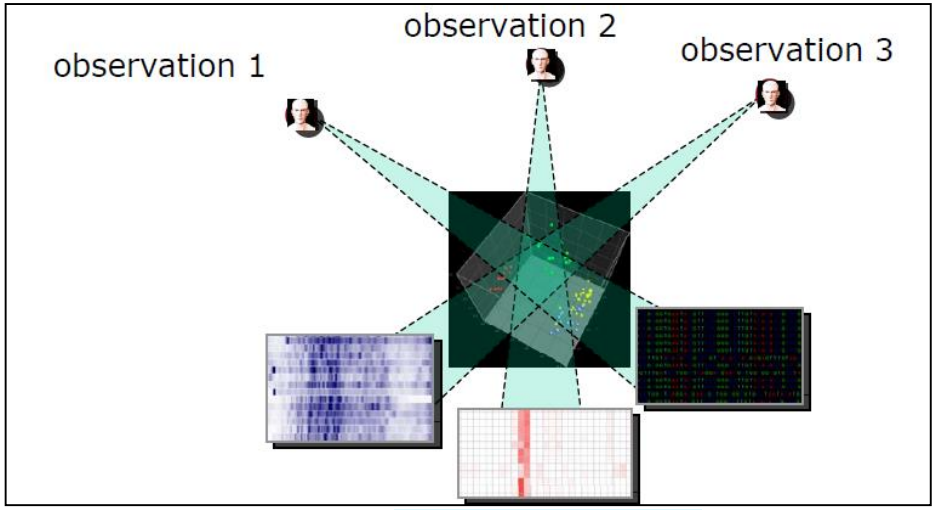
Definition of species
Definition of "species" in microbiology

- Strain:
- A population of microbes descended from a single individual or pure culture.
- Different strains represent genetic variability within a species:
  - Biovars: Strains that differ in biochemical or physiological differences.
  - Morphovars: Strains that vary in morphology.
  - Serovars: Stains that vary in their antigenic properties.

Hendrix,2004

Species Concept in Prokaryotes
Polyphasic species definition

- Species are defined by pragmatic, arbitrary, and sometimes artificial methods based on 16S rRNA gene sequences, DNA-DNA hybridisation, morphology, physiology and chemotaxonomy (...)”
- The system is functional in many ways



P. Vandamme

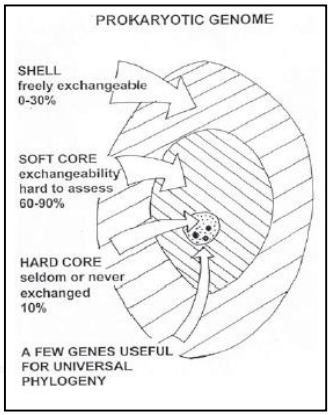
Species Concept in Prokaryotes
Polyphasic species definition

- The bacterial species appears to be an assemblage of isolates originating from a common ancestor population in which genetic drift resulted in clones with different degrees of recombination and characterized by:

1. a certain degree of phenotypic consistency;
2. a significant degree of DNA-DNA hybridization;
3. over 97% of 16S rRNA gene sequence similarity.

Vandamme,2013

Polyphasic Taxonomy
Species concept



- Whole genome sequences can become part of polyphasic taxonomy and the standard description of bacterial species.
- Genomes seem to be composed of:

1. a core set of genes that is conserved among strains of the same species, and
2. accessory genes that are strain specific.

- Content and size of core vary with species.

Vandamme,2013

Species Concept in Prokaryotes

Polyphasic approach

The only solution for species definition

- Today, prokaryote taxonomists agree that a reliable classification and acceptable species concept can only be achieved by the exploration of the internal diversity of taxa by a wide range of techniques in what is generally known as the 'polyphasic approach' including:

1. Phenotype,
2. Ribotype, and
3. Genotype informations.

Species Concept in Prokaryotes

Methods and parameters used in prokaryotic species designation

1. Phenotypic Based
2. Classical
3. Chemotaxonomy
4. Microbial Identification Systems
5. DNA Based
6. rRNA Analysis
7. DNA Based typing methods

Species Concept in Prokaryotes
Taxonomy and species definitions

- The observation of diversity: phenotypic and genotypic coherence allows to define bacterial species.
- Taxonomy and species definitions vary with technology: old and new practices.

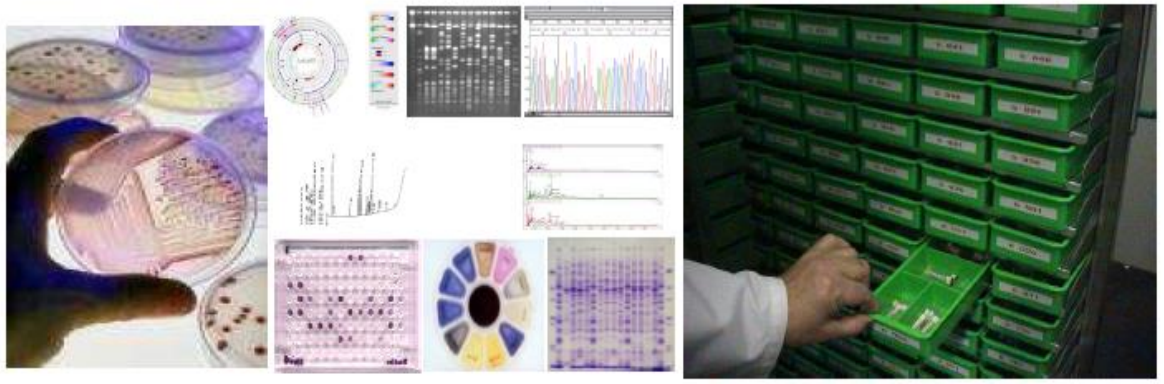
1. Phenotypic and numerical taxonomy
2. DNA
3. Phylogeny
4. Polyphasic taxonomy
5. Whole genome sequences.

Vandamme,2013

Species Concept in Prokaryotes

Taxonomy and species definitions

http://www.lm.ugent.be/



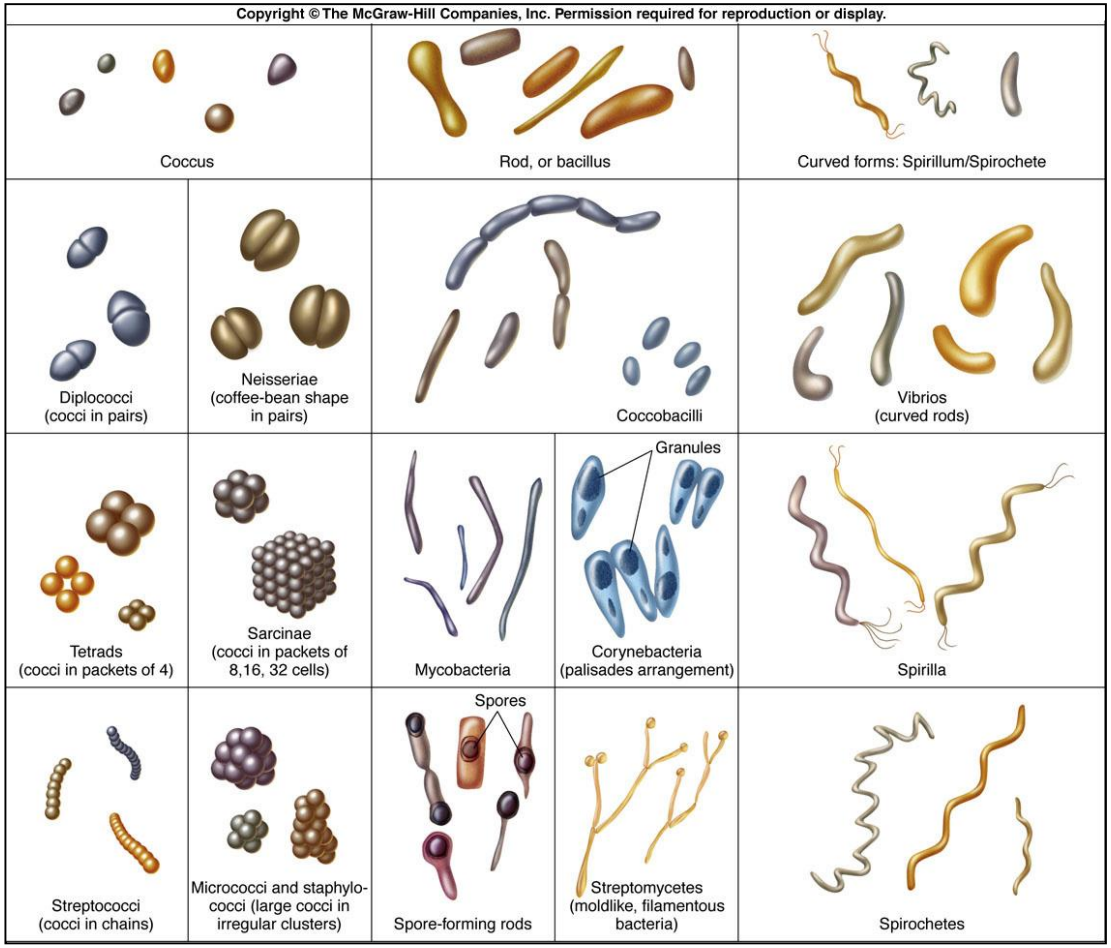
Vandamme,2013

Species Concept in Prokaryotes
1. Phenetic methods

1. Morphology, physiology and growth conditions.
2. Tests directly or indirectly different phenotypic properties.
3. Difficult to perform cladistic analysis based on phenotypic data (phylogenetic reconstructions).
4. Useful for establishing relationships within species.
5. Lack resolution above this taxon level.

Species Concept in Prokaryotes
Phenetic methods

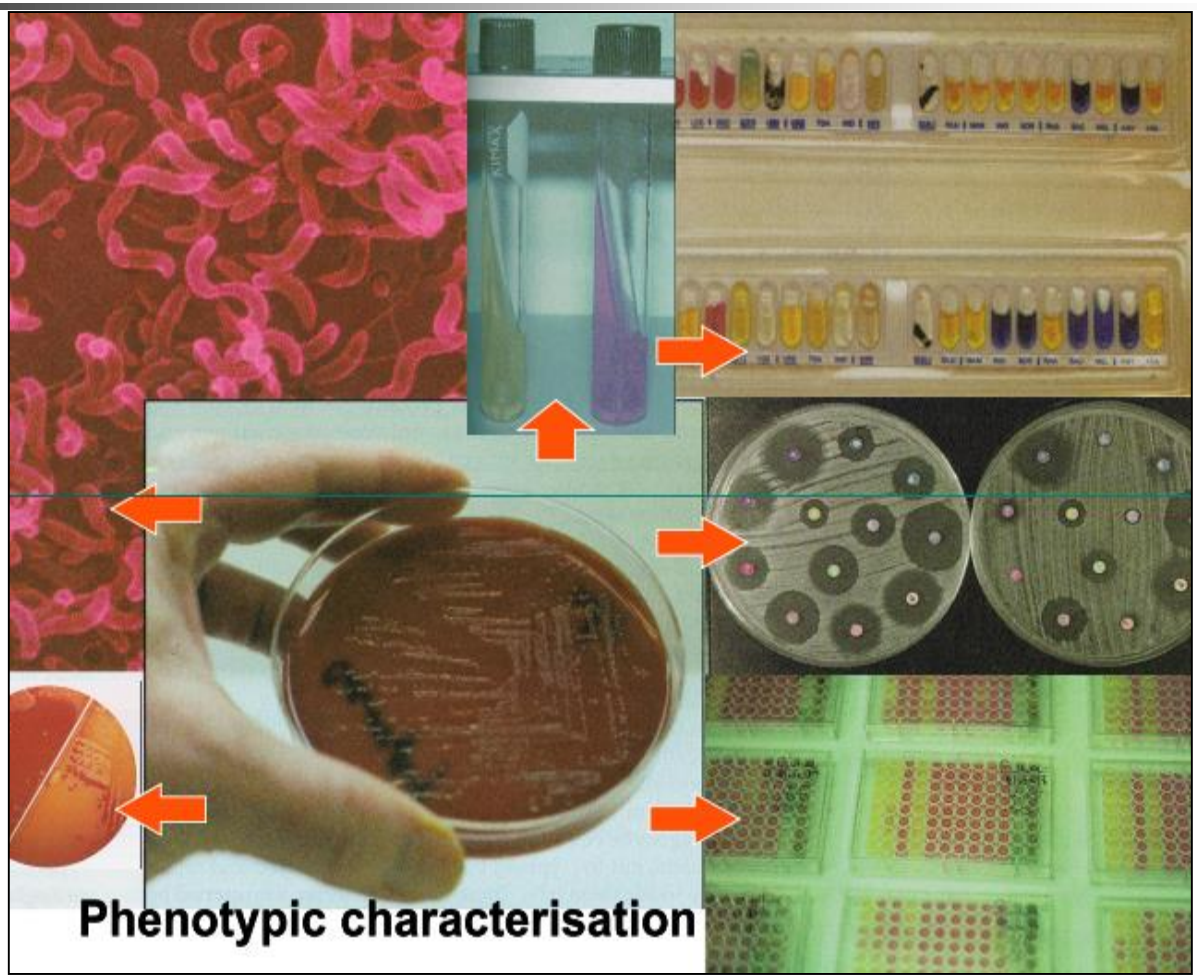
4 Main Shapes of Bacteria:
- **Cocci**
- **Bacilli**
- **Spirilla**
- **Spirochetes**



Powers,2011

The bacterial species concept, definition & taxonomy

Conventional bacterial taxonomy placed heavy emphasis on analyses of phenotypic properties of the organism



Vandamme,2013

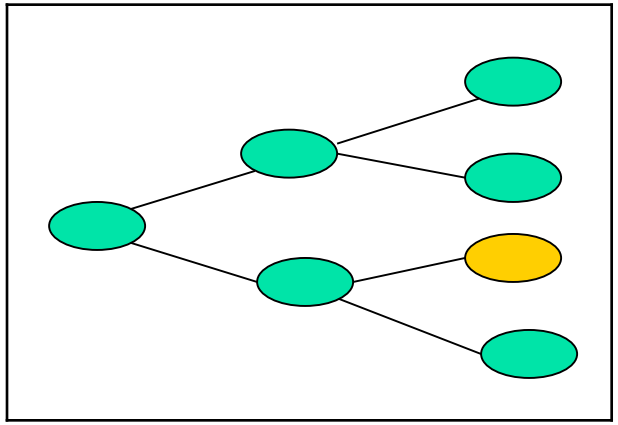
Species Concept in Prokaryotes
Limits of phenetic methods

- Unlike eukaryotes, bacteria reproduce by binary fission and bacterial populations consist of clones (population of cells derived from a single cell).
- Mutation generates variation and diversity among these clones.
- However, genetic recombination by lateral transfer of genes also occurs, and the frequency varies greatly.

Lan and Reeves,2001

Species Concept in Prokaryotes
Limits of phenetic methods

- Based on asexual division (clones) rather than sexual reproduction within the group.
- Mutations during division introduce new phenetic characters.
- Impossible to deduce bacterial lineages by examination of present-day phenetic characters – lateral gene flow would confuse the phylogeny.



Prokaryotic species: **A population of cells with similar characteristics.**
Clone: **Population of cells derived from a single cell;**
Strain: **Genetically different cells within a clone;**
Culture: **grown in the lab.**

Species Concept in Prokaryotes
Limits of phenetic methods

- The clone is not a defined taxonomic level, making it quite possible to name clones within clones, reflecting the reality of bacterial populations.
- **Longevity of clones:**

1. Clones are identified by very high levels of similarity of sequence in housekeeping genes.
2. How long a clone will retain this characteristic depends on the level of clonality of the species.
3. There are a wide range in levels of clonality among bacterial species.

Clonality: A measure of the ability of a group of the same cells to form clones.

Lan and Reeves,2001

The Species Concept in Prokaryotes
Minimum standards for species definition
2.1. Mole % G+C and DNA/DNA Homology

- Minimum standards for species definition:

1. Mole % G+C;
2. DNA homology.

- But the most impotant molecular clock is 16S rRNA gene.
- In recent years, species discrimination was based on:

1. 70% DNA-DNA Homology, and
2. 2% 16S nucleotide (base) sequence difference.

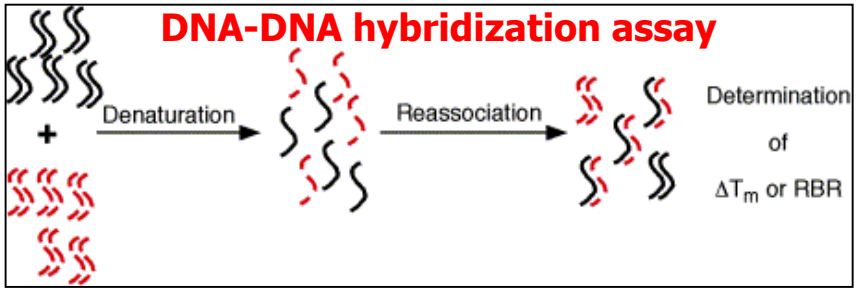
At present all they believe that definitions of bacterial species and subspecies ranks can be obtained in the light of detailed information from whole genome typing approaches.

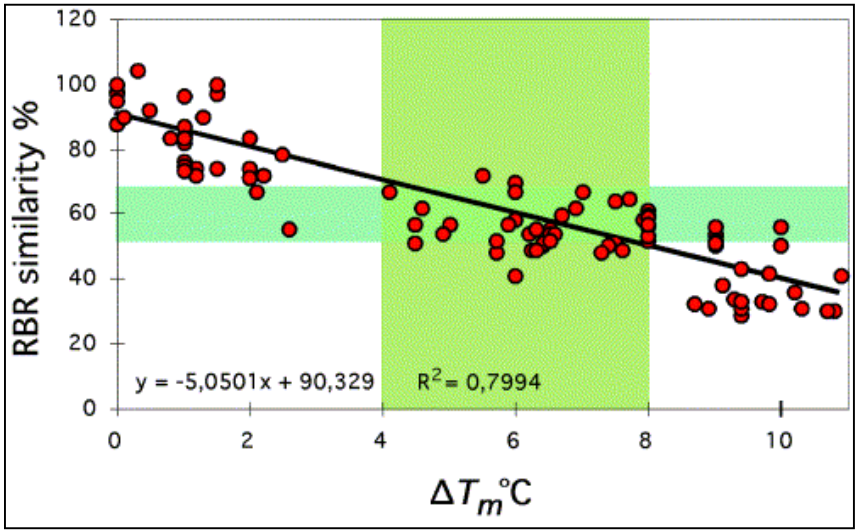
Species Concept in Prokaryotes

DNA-DNA hybridization

Old gold standard for bacterial species demarcation

- Correlation between RBR (relative binding ratio) and Tm values.
- Commonly accepted values for species boundaries are indicated in green.

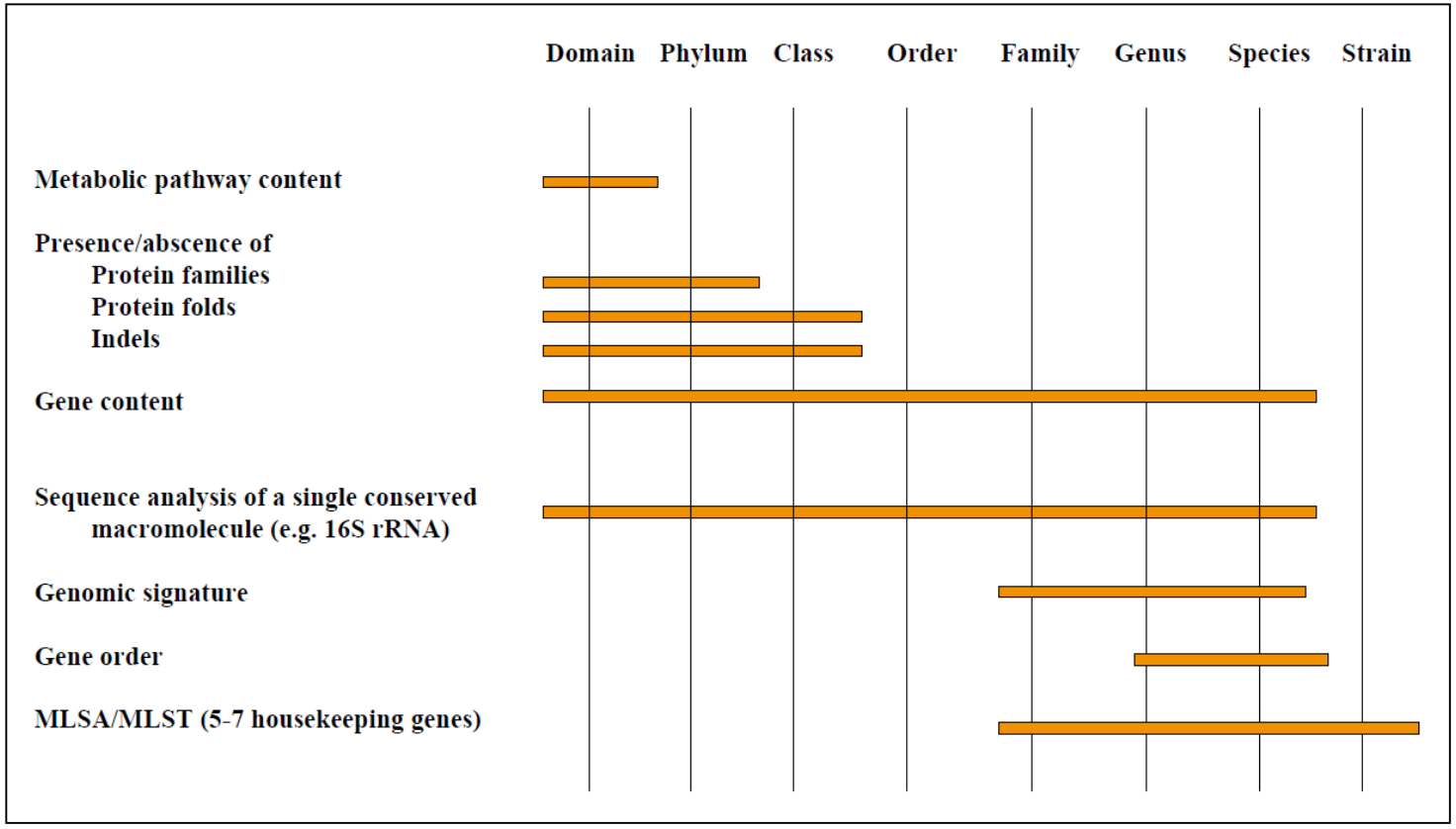




Species Concept in Prokaryotes

Gene-sequence-based criteria for species definition

Sequence analysis of conserved macromolecules



Gene order: The sequential location of genes on a chromosome. Mutations occur when the number or order of bases in a gene is disrupted. The genomic signature refers to the characteristic frequency of oligonucleotides in a genome or sequence. It has been observed that the genomic signature of phylogenetically related genomes is similar.

Species Concept in Prokaryotes

2.2. 16S rDNA sequences (signature sequences)

Molecular clocks (chronometers)

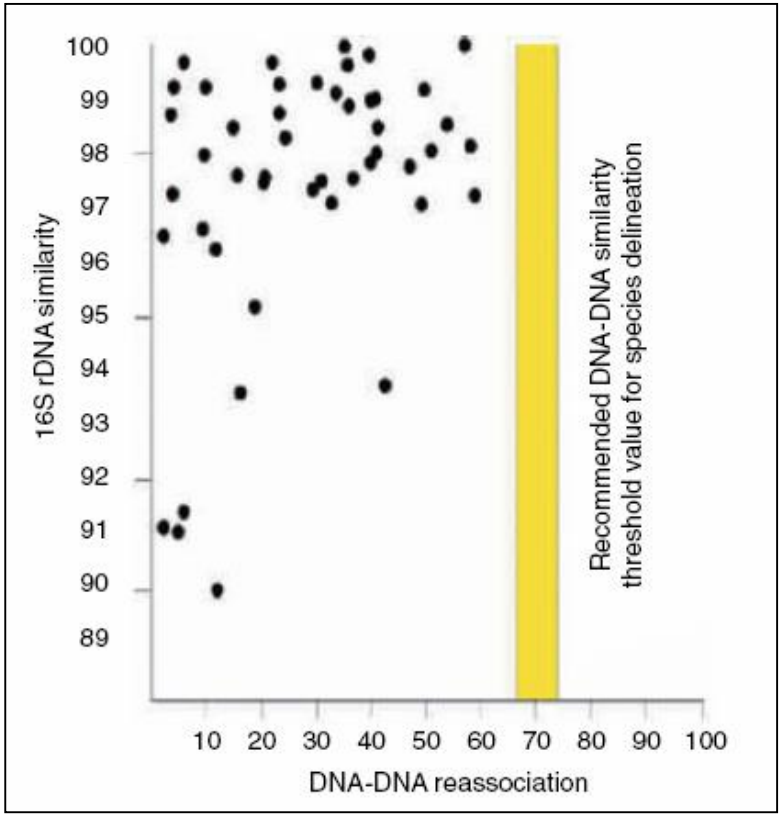
- The most widely used molecular clocks ('single locus appraoches' are small subunit ribosomal RNA (SSU rRNA) genes found in all domains of life (not the case with other chronometers)
- 16S rRNA in prokaryotes and 18S rRNA in eukaryotes.
  - Functionally constant;
  - Sufficiently conserved,
  - Sufficient length;
  - Without (?) lateral gene transfer or recombination.

Vandamme,2013

Species Concept in Prokaryotes

The correlation blot of 16S rDNA and DNA-DNA similarity values

- Correlation blot between 16S rDNA and DNA-DNA reassociation similarities, indicating for several phylogenetically highly related species (as determined by 16S rDNA values of > 98%) that DNA-DNA similarities are clearly below the threshold value of 70% recommended for species delineation.



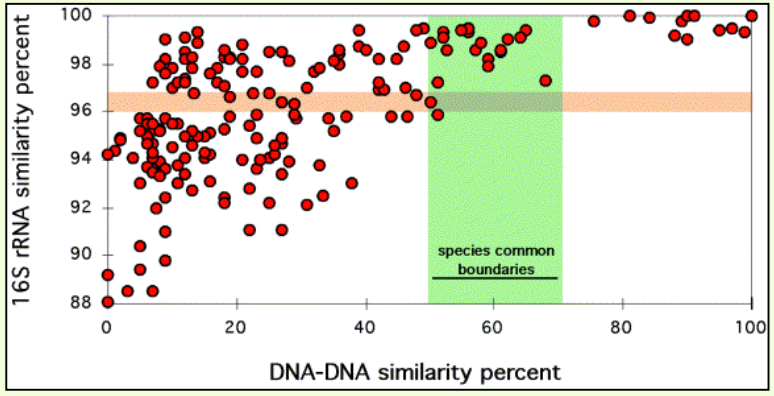
It is now generally accepted that DDH is only required when 16S rRNA gene sequence similarity between two strains is over 97%, even though higher thresholds of 98.7-99.0% have also been used.

DNA-DNA hybridizations and 16S rRNA gene similarities

- Comparison of DNA-cDNA and 16S rRNA similarities.
- Stackebrandt & Goebel (1994):

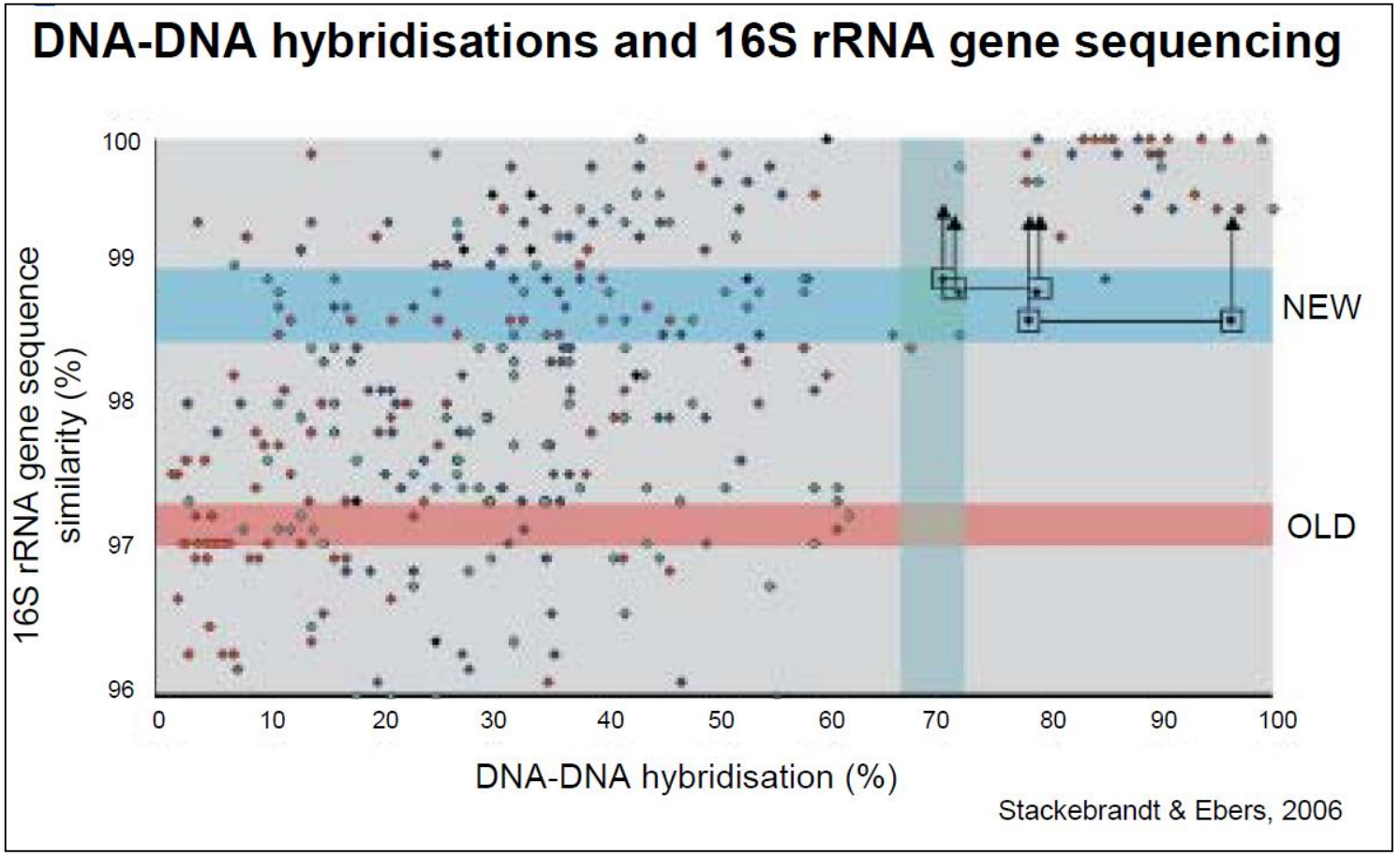
1. Above 97.5% 16S rRNA gene sequence similarity, DDH values can be high or low.
2. Organisms with less than 97.0% 16S rRNA gene sequence similarity will not show meaningful DDH values.

- Stackebrandt & Ebers (2006): revision of threshold to 98.7-99%.



Rossello-Mora & Amann,2001;T. Coenye

DNA-DNA hybridizations and 16S rRNA gene similarities



Stackebrandt & Ebers,2006;T. Coenye

Species Concept in Prokaryotes

16S rDNA sequences (signature sequences)

The sequences of the 16S rRNA

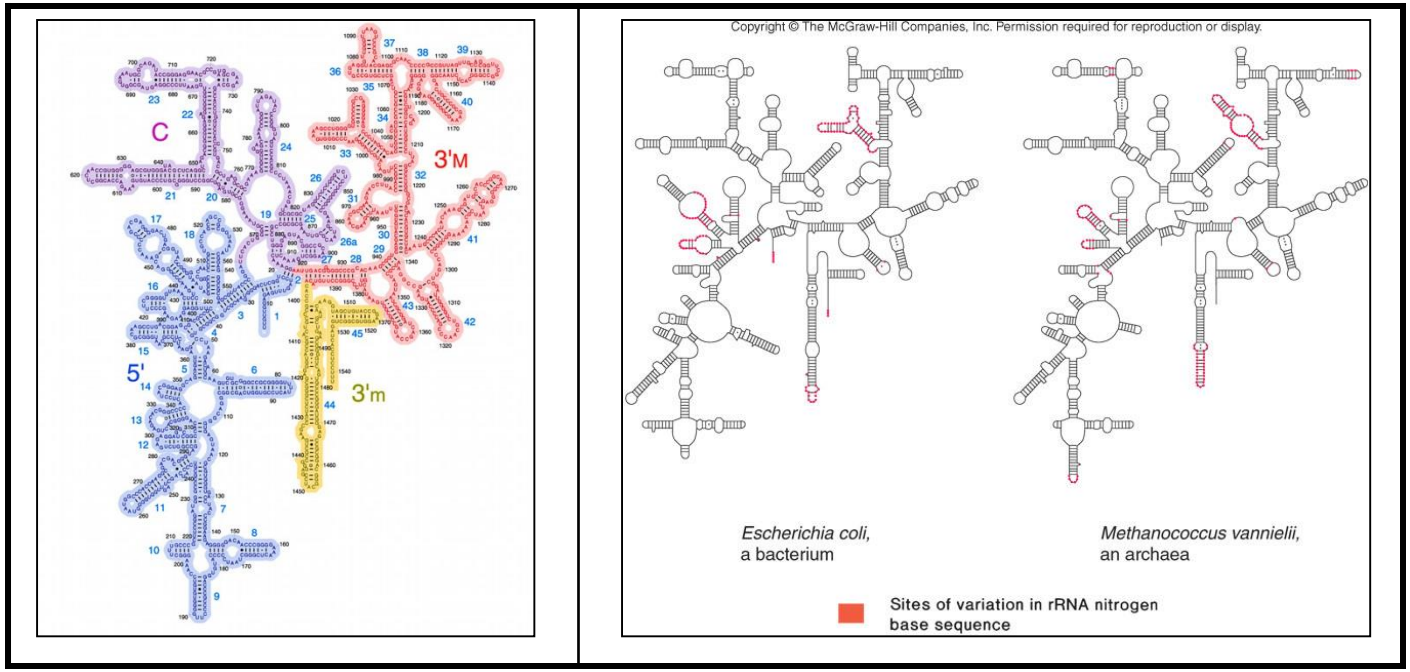
- One of the most attractive potential uses of 16S rRNA gene sequence informatics is to provide genus and species identification for isolates that do not fit any recognized biochemical profiles, for strains generating only a "low likelihood" or "acceptable" identification according to commercial systems, or for taxa that are rarely associated with human infectious diseases.

Janda and Abbott,2007

Species concept

Limits of 16S rRNA sequences in elucidating the branching order of individual phyla

- Despite this success, trees based on 16S rRNA sequences lack resolution when it comes to elucidating the branching order of individual phyla.



16S rRNA molecular structure

Species Concept in Prokaryotes

16S rDNA sequences (signature sequences)

The sequences of the 16S rRNA

- Although 16S rRNA gene sequencing is:

1. highly useful in regards to bacterial classification,
2. it has low phylogenetic power at the species level, and
3. poor discriminatory power for some genera, and DNA relatedness studies are necessary to provide absolute resolution to these taxonomic problems.

- The genus *Bacillus* is a good example of this.
- The type strains of *B. globisporus* and *B. psychrophilus* share >99.5% sequence similarity with regard to their 16S rRNA genes, and yet at the DNA level exhibit only 23 to 50% relatedness in reciprocal hybridization reactions.

Janda and Abbott,2007

Species Concept in Prokaryotes

16S rDNA sequences (signature sequences)

The sequences of the 16S rRNA

- These groups include (not exclusively),
- the family *Enterobacteriaceae* (in particular, *Enterobacter* and *Pantoea*),
- rapid-growing mycobacteria, the *Acinetobacter baumannii-A. calcoaceticus* complex,
- *Achromobacter*,
- *Stenotrophomonas*, and
- *Actinomyces*.
- Some of these problems are related to bacterial nomenclature and taxonomy while others are related to different issues cited in next slides.

Janda and Abbott,2007

Species Concept in Prokaryotes

16S rDNA sequences (signature sequences)

The sequences of the 16S rRNA

- Unfortunately, no universal definition for species identification via 16S rRNA gene sequencing exists, and authors vary widely in their use of acceptable criteria for establishing a "species" match.
- In none of these studies does the definition of a species "match" ever exceed 99% similarity (1% divergence).
- Based on the data listed above, even this threshold value may not be sufficient in all instances to guarantee an accurate identification.

Janda and Abbott,2007

Species Concept in Prokaryotes

16S rDNA sequences (signature sequences)

The sequences of the 16S rRNA

- In the case of *Aeromonas veronii* the genome can contain up to six copies of the 16S rRNA gene that differ by up to 1.5% among themselves.
- This implies intragenomic heterogeneity of the 16S rRNA gene among aeromonads and would preclude the use of this technology alone for species identification.
- The collective data described above strongly suggest that any microbial identifications using 16S rRNA distance scores of 1% are unsatisfactory for a diagnostic or public health reference laboratory.

Janda and Abbott,2007

Limits of 16S rRNA sequences in identification of bacteria at species level

Selected examples of bacterial genera and species with identification problems using 16S rRNA gene sequencing

Genus	Species
Aeromonas	*A. veronii*
Bacillus	*B. anthracis*, *B. cereus*, *B. globisporus*, *B. psychrophilus*
Bordetella	*B. bronchiseptica*, *B. parapertussis*, *B. pertussis*
Burkholderia	*B. cocovenenans*, *B. gladioli*, *B. pseudomallei*, *B. thailandensis*
Campylobacter	Non-*jejuni-coli* group
Edwardsiella	*E. tarda*, *E. hoshinae*, *E. ictaluri*
Enterobacter	*E. cloacae*
Neisseria	*N. cinerea*, *N. meningitidis*
Pseudomonas	*P. fluorescens*, *P. jessenii*
Streptococcus	*S. mitis*, *S. oralis*, *S. pneumoniae*

Species Concept in Prokaryotes

16S rDNA sequences (signature sequences)

The sequences of the 16S rRNA

1. It is believed that a 16S rRNA gene sequence similarity in the range of 0.5 to 1% would be required for the definition of a species.

- But some others use:

1. Threshold values of 1% to define a genus.
2. Threshold values of 1.5% to define a species.

Note: unlike DNA hybridization (70% reassociation) there are no defined "threshold values" for 16SrRNA (e.g., 98.5% similarity) above which there is universal agreement of what constitutes definitive and conclusive identification to the rank of species (Janda and Abbott,2007).

Species Concept in Prokaryotes

16S rDNA sequences (signature sequences)

The sequences of the 16S rRNA

- The cumulative results from a limited number of studies to date suggest that:

1. 16S rRNA gene sequencing provides genus identification in most cases (90%), but
2. less so with regard to species (65 to 83%), with
3. from 1 to 14% of the isolates remaining unidentified after testing.

Janda and Abbott,2007

Species Concept in Prokaryotes

16S rDNA sequences (signature sequences)

The sequences of the 16S rRNA

- Difficulties encountered in obtaining a genus and species identification include:

1. the recognition of novel taxa,
2. too few sequences deposited in nucleotide databases, species sharing similar and/or
3. identical 16S rRNA sequences, or
4. nomenclature problems arising from multiple genomovars assigned to single species or complexes.

Janda and Abbott,2007

16S rRNA sequence analysis:

Caveats (warning or caution)

16S rRNA sequence identity may not be sufficient to guarantee species identity

- Often insufficient diversity to distinguish closely related species (Fox *et al.*,1992).
- How close is close: **16S rRNA sequence identity may not be sufficient to guarantee species identity.**

SPECIES_NAME	STRAIN_NUMBER	
"Burkholderia cepacia"	"PVFi5A"	100.0
Burkholderia pyrrocinia	ATCC 15958T	99.8
Burkholderia diffusa	LMG 24065T	99.8
Burkholderia contaminans	SAR-1	99.7
Burkholderia metallica	LMG 24068T	99.7
Burkholderia anthina	R-4183	99.7
Burkholderia lata	ATTC 17760T	99.7
Burkholderia ambifaria	AMMDT	99.6
Burkholderia cepacia	ATCC 25416T	99.6
Burkholderia stabilis	LMG 14294T	99.6
Burkholderia arboris	LMG 24066T	99.5
Burkholderia seminalis	R-24196T	99.5
Burkholderia latens	LMG 24064T	99.4
Burkholderia ubonensis	EY 3383T	99.3
Burkholderia vietnamiensis	TVV 75T	99.3
Burkholderia cenocepacia	LMG 16656T	99.3
Burkholderia multivorans	LMG 13010T	99.1
Burkholderia dolosa	LMG 18941	99.1
Burkholderia thailandensis	E264T	98.8
Burkholderia pseudomallei	1026b	98.8
Burkholderia glumae	LMG 2196T	98.7
Burkholderia mallei	ATCC 23344T	98.7
Burkholderia oklahomensis	C6786T	98.6
Burkholderia gladioli pv. gladioli	ATCC 10248	98.5
Burkholderia plantarii	LMG 9035T	98.4

Vandamme,2013

Limits of 16S rRNA sequences in identification of bacteria at species level

16 sequence data has been compared to identification results obtained either in conventional or commercial test formats

NO. of strains	Group studied[a]	16S			Commercial system (s)	Species identification (%)[c]		
		Size (s) (bp)	Database[b]	Criteria (%)[c]		Conv	Comm	16S
72	GNB	1,189,527,418	MicroSeq	CM	Conv, MIDI, Biolog	90	67.7-84.8	89.2
328	Mycobacteria		MicroSeq	99≤	Conv.	42		62.5
83	GNB, GPB		MicroSeq	CM	Vitek 2, Phoenix		77.1	100
231	*Bacteroides*	711 ,899	500	99≤	Conv	74.5		83.1
47	CNS	1,500	527	97<	API StaphID, Phoenix		63.8-85.1	87.2
20	GPA	1,500	GenBank	98≤	Vitek ANA, RapID ANA II,		45-20	65
			MicroSeq		API 20A			
107	GNNFB	796	GenBank, EMBL, DDBJ	99≤	API 20NE, Vitek 2		53.2-54.2	91.6

a CNS, coagulase-negative staphylococci; GNB, gram-negative bacteria; GNNFB, gram-negative nonfermentative bacteria; GPA, gram-positive anaerobes; GPB, gram-positive bacteria. *b* DDBJ, DNA Data Bank Japan; EMBL, European Molecular Biology Laboratory. *c* CM, closest match; Comm, commercial system; Conv, conventional phenotypic tests((Janda and Abbott,2007).

Limits of 16S rRNA sequences in identification of bacteria at species level

Recommended guidelines for use of 16S rRNA gene sequencing for microbial identification

Category	Guidelines
Strain to be sequenced	• Phenetic profile of strain is not known by general grouping to present difficulties for identification by 16S rRNA gene analysis. • For such strains requiring molecular identification, another housekeeping gene is required (e.g., *rpoB*)
16S rRNA gene sequencing	• Minimum: 500 to 525 bp sequenced; ideal: 1,300 to 1,500 bp sequenced • <1% position ambiguities
Criteria for species identification	• Minimum: >99% sequence similarity; ideal: >99.5% sequence similarity; • Sequence match is to type strain or reference strain of species that has undergone DNA-relatedness studies; • For matches with distance scores <0.5% to the next closest species, other properties, including phenotype, should be considered in final species identification.

Janda and Abbott,2007

Species Concept in Prokaryotes

Limits of 16S rRNA sequences with species definition

- On the brink of the new millenium, the significance of 16S rRNA as a major taxonomic criterion is receding.
- Unfortunately, the resolving power of the 16S rRNA is insufficient (not enough polymorphic) to guarantee correct delineation of bacterial species.
- Furthermore, its validity as a marker for phylogenetic inferences is being questioned.
- At the same time, recognition of lateral (horizontal) gene transfer as an important element in prokaryote evolution is gaining momentum (Lake *et al.*,1999).

Gest,1999;…

Species Concept in Prokaryotes

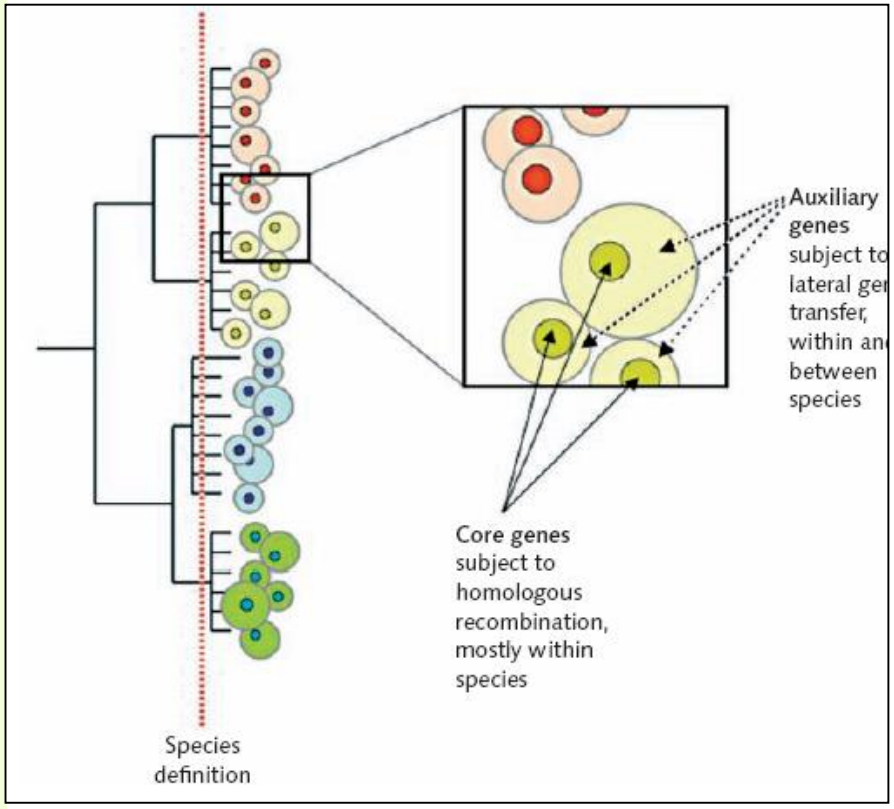
Limits of 16S rRNA sequences with real prokaryote diversity

- Ribosomal DNA (rDNA) sequence analyses most notably those based on 16S rRNA sequences, account for a very small portion of the real prokaryote diversity.
- Thus, if we only take into account the recognized prokaryotic species for diversity calculations, their total number would never be regarded as a significant proportion of the total Earth's biodiversity.

Species concept

Limits of 16S rRNA sequences with real prokaryote diversity and microdiversity

- Trees based on 16S rRNA or other phylogenetic marker genes sequences often show 'microdiversity', exhibiting clusters of sequences more closely related than the value accepted as defining species (vertical hatched red line).
- Even in species so defined, genomes can show substantial (up to 30 %) variation in size and gene content.
- A 'species genome' or 'pangenome' can be imagined to comprise core genes shared by all its strains, and a set of auxiliary genes found only in some strains.
- These two classes of genes may have different evolutionary modes and tempo.



Ribosomal RNA trees

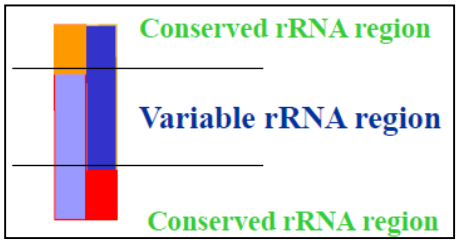
The pangenome is the collection of all genes coming from isolates of the same species.

Species Concept in Prokaryotes
2.3. Based on 16S-23S rDNA ITS

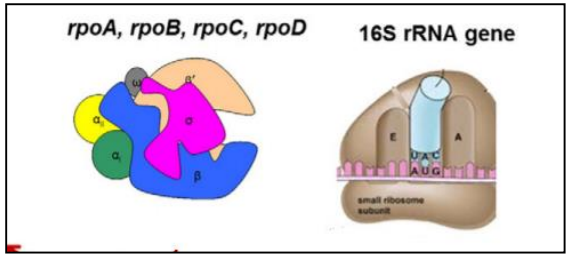
- All strains within the genus *Xanthomonas* that showed 70% or more DNA–DNA homology exhibited 100% 16S rDNA sequence similarity.
- Better discrimination was obtained by analysis of the 16S-23S rDNA intergenic spacer (ITS) sequence, which showed approximately ninefold higher diversity than 16S rDNA.

Species Concept in Prokaryotes
2.3. Based on 16S-23S rDNA ITS

- The rDNA genetic loci in eubacteria include, in 5' to 3' order: 16S, 23S, and 5S rRNA genes, which are separated by intergenic transcribed spacer (ITS) regions (16S-23S ribosomal intergenic transcribed spacer (ITS) region.
- The 16S-23S internally transcribed spacer region(ITS) is more variable in sequence, thus providing significantly greater taxonomic resolution than is found for the 16S small subunit (SSU or *rrs*) and 23S large subunit (LSU or *rrl*) genes.



Target sequences
Ribosomal RNA operon (rrn)

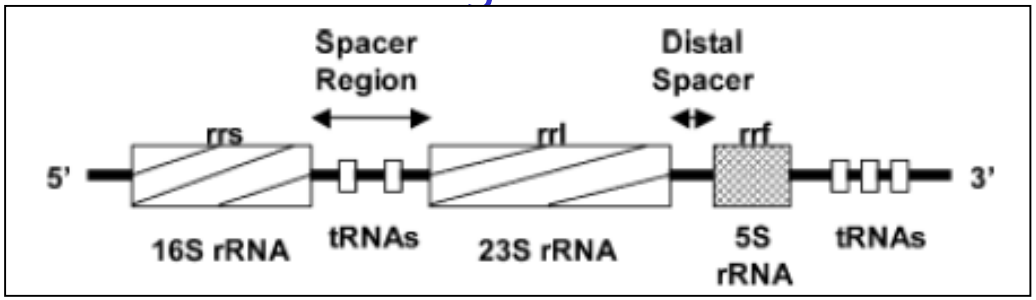


rrs, small ribosomal genes encode 16SrRNA, *rrl*, large genes encode 23S rRNAs and *rrf*, ribosome release factor genes encode 5S rRNAs.

- The rrn locus consisted of:

1. A 16S rRNA gene (rrs), followed by
2. An intergenic transcribed spacer (ITS) containing two genes of $tRNA^{Ile}$ and $tRNA^{Ala}$, a 23S rRNA gene (rrl), an ITS devoid of tRNA genes and a 5S rRNA gene (rrf).

- The internally transcribed spacer region (ITS) between the 16S and 23S rRNA genes appears to be more variable than 16S and 23S rRNA genes.

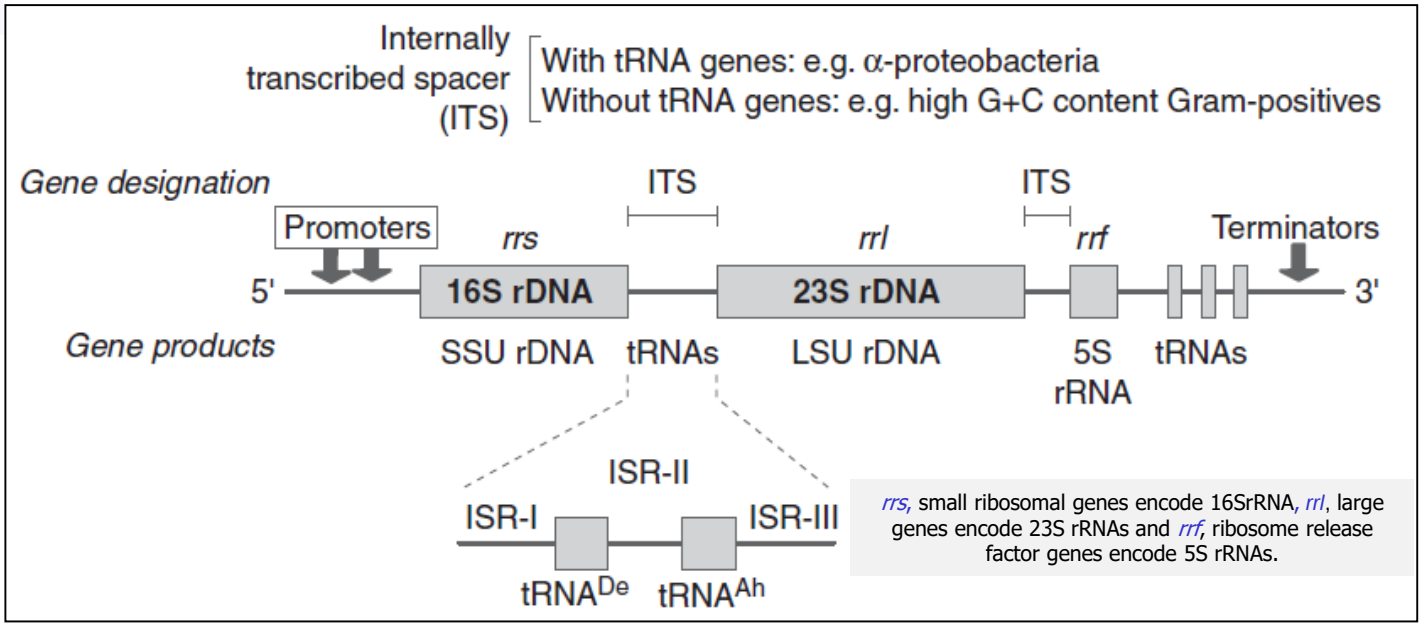


Schematic diagram of a typical ribosomal rRNA operon.

Species Concept in Prokaryotes

General structure of the prokaryotic ribosomal rRNA operon (*rrn*)

Based on 16S-23S rDNA ITS



General structure of the prokaryotic ribosomal rRNA operon (*rrn*). The detection of heterogeneity of the 16S-23S ribosomal intergenic/internally transcribed spacer (ITS) region has become rather common over the past years for identification and typing purposes of bacteria.

The 16S-23S IGS region may or may not contain tandem $tRNA^{Ile}$ (isoleucyltRNA synthetase) and $tRNA^{Ala}$ genes. Several groups of prokaryotes do not contain such genes linked to the *rrn* operon, and others contain only one of the two tRNA genes ($tRNA^{Ile}$ or $tRNA^{Ala}$) or alternatively the single $tRNA^{Glu}$ gene. Several species have been reported to contain split *rrn* operons, in which *rrs*, *rrl* and *rrf* are not all linked to each other. The 23S/5S ribosomal RNA genes (*rrl*/*rrf*) are separate from the 16S ribosomal RNA gene (*rrs*).

Osborn and Smith,2005;..

Species Concept in Prokaryotes

2.4. MLST loci-concatenated sequence analysis for species definition

- Concatenation: The operation of joining two character strings end-to-end.
- Recently the sequencing of several housekeeping genes has been proposed as an alternative to DNA–DNA Hybridization.
- Example of genes:
- 16S rDNA, *gltA*, *groEL*, *rpoB*, *ftsZ*, *rib*C and internal transcribed spacer (ITS).

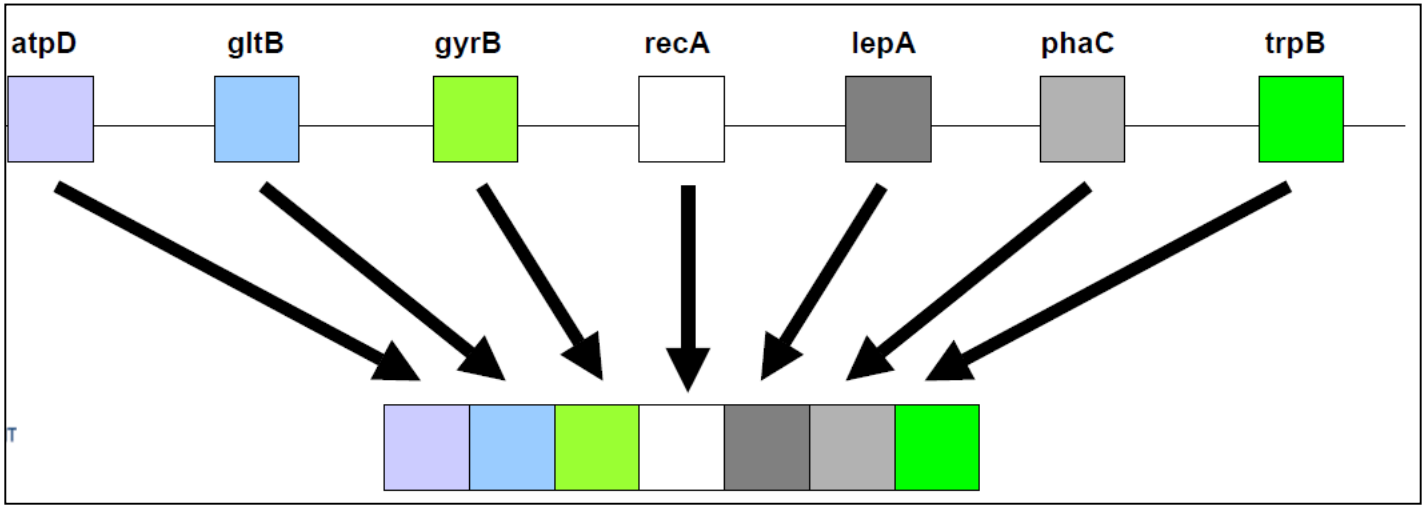
ribC (riboflavin synthase), *groEL* (Hsp60 chaperone), *gltA* (citrate synthase), *nlpD* (glycoprotein), *ftsZ* (cell division protein), *rpoB* (RNA polymerase beta subunit). The *rpoB* gene encodes the β subunit of bacterial RNA polymerase. *rpoB* is also found in plant chloroplasts where it forms the beta subunit of the plastid-encoded RNA polymerase (PEP).

La Scola *et al.*,2003

Species Concept in Prokaryotes

MLST loci- concatenated sequence analysis for species definition

- The sequences of the seven loci are put end to end to form one large sequence which can be used in base pair comparisons.



Species Concept in Prokaryotes

Gene-sequence-based criteria for species definition

The primers used for PCR amplification of MLST housekeeping genes

- The primer pairs used for the PCR amplification of internal fragments of these genes are:

Gene	Product size (bp)	Forward primer 5'-3' sequence	Reverse primer 5'-3' sequence
16S rDNA	472	AGAGTTTGATCCTGGYTCAG	CTTTACGCCCARTAAWTCCG
batR	487	GACCGCAATATTTTGACATC	GCATCCATCAAAGCATCACGACTT
ribC	283	AGCGAGGATCAAAACAAC	GCTCTTCAACACAATTAACG
groEL	369	GTTGATGATGCCTTGAAC	TGGTGTGTCTTTCTTTGG
gltA	338	GGGGACCAGCTCATGGTGG	AATGCAAAAAGAACAGTAAACA
nlpD	494	GGCGCTGGTATGATACAA	GACATCTGTGCGGAAGAA
ftsZ	483	GCCTTCTCATCCTCAACTTC	CTTTGTTTTAAACGCTGCC
rpoB	471	CTGGACGTACATCCTACA	AACAGCAGCTCCTGAATC

Iredell *et al.*,2003

The Species Concept in Prokaryotes

Different thresholds for a species delineation

2.5. Two new genome-based parameters

- Two new genome-based parameters applicable to prokaryotic taxonomy are:

1. Average amino acid identity (AAI): Pairwise genome comparisons and averaging the sequence identities of shared orthologous genes (amino acid).
2. Average nucleotide identity (ANI): Pairwise genome comparisons and averaging the sequence identities of shared orthologous genes (nucleotide).

- Both achieve, to a great extent, the goal that whole-genome DNA hybridizations (DDH) pursued and that has been especially determinant for the circumscription of prokaryotic species.

Rosselló-Mora,2005

Orthologous genes: Genes in different species that are homologous (similar) because they are derived from a common ancestral gene.

The Species Concept in Prokaryotes

New threshold for a species delineation

Comparison of ANI and DNA-DNA hybridization similarity

- Phylogenetic signal present in core genes (ANI values):
- 95% ANI corresponds with 70% DNA-DNA hybridization.
- ANI does not necessarily correlate with gene content.
  - ANI values reflect phylogeny;
  - Gene content reflects ecology.
- Bacteria with considerable differences in gene content are classified in the same species in spite of considerable genomic differences.
- However, due to the labour-intensive and error-prone nature of DDH experiments, average nucleotide identity (ANI) was replaced with DDH.

Vandamme,2013;Kim *et al.*,2014

Whole genome sequencing (WGS)
Next Generation Sequencing (NGS)
Genome-wide Average Nucleotide Identity (gANI) metric

- Whole-genome comparison-based classification methods such as Average Nucleotide Identity (ANI) have been demonstrated to correlate well with the previous 'gold-standard' DNA–DNA hybridization (DDH) approaches to evaluation of genomic relationship.
- In particular, an established threshold of 95% sequence identity by ANI was found to correspond to the traditional 'gold standard' prokaryotic species threshold of 70% identity by DDH.

Pritchard *et al.*,2016

Whole genome sequencing (WGS)
Next Generation Sequencing (NGS)
MiSI (Microbial Species Identifier) method

- Recently, the related MiSI (Microbial Species Identifier) method has been proposed as an approach for correction of inconsistent prokaryotic species classification, and for taxonomic assignment of newly-sequenced organisms.
- For MiSI, a common gene content of at least 60% of the complement, with at least 96.5% nucleotide sequence identity between reciprocally best-matching genes, is found to correspond well to species boundaries.
- The MiSI method was systematically applied to over 13000 sequenced prokaryotes in the NCBI public database.
- Both ANI and MiSI techniques identify misclassified genomes in the public repositories; up to 18% of all annotated species in public sequence databases were indicated to include potentially misclassified isolates, by MiSI.

Pritchard *et al.*,2016

Whole genome sequencing (WGS)
Microbial species delineation using whole genome sequences

A. Breakdown of genomes and species into cliques, clique groups and singletons. This figure shows the number of genomes and named species at each step of our method. The genomes/species in common are annotated between the boxes.

B. Breakdown of named species for 901 clusters. This graph shows how many clusters are populated by one species, two species and so on.

C. Breakdown of species into different categories. This graph shows how many species fall into the 'single homogeneous', 'multiple homogeneous', 'single heterogeneous' and 'multiple heterogeneous' species category.

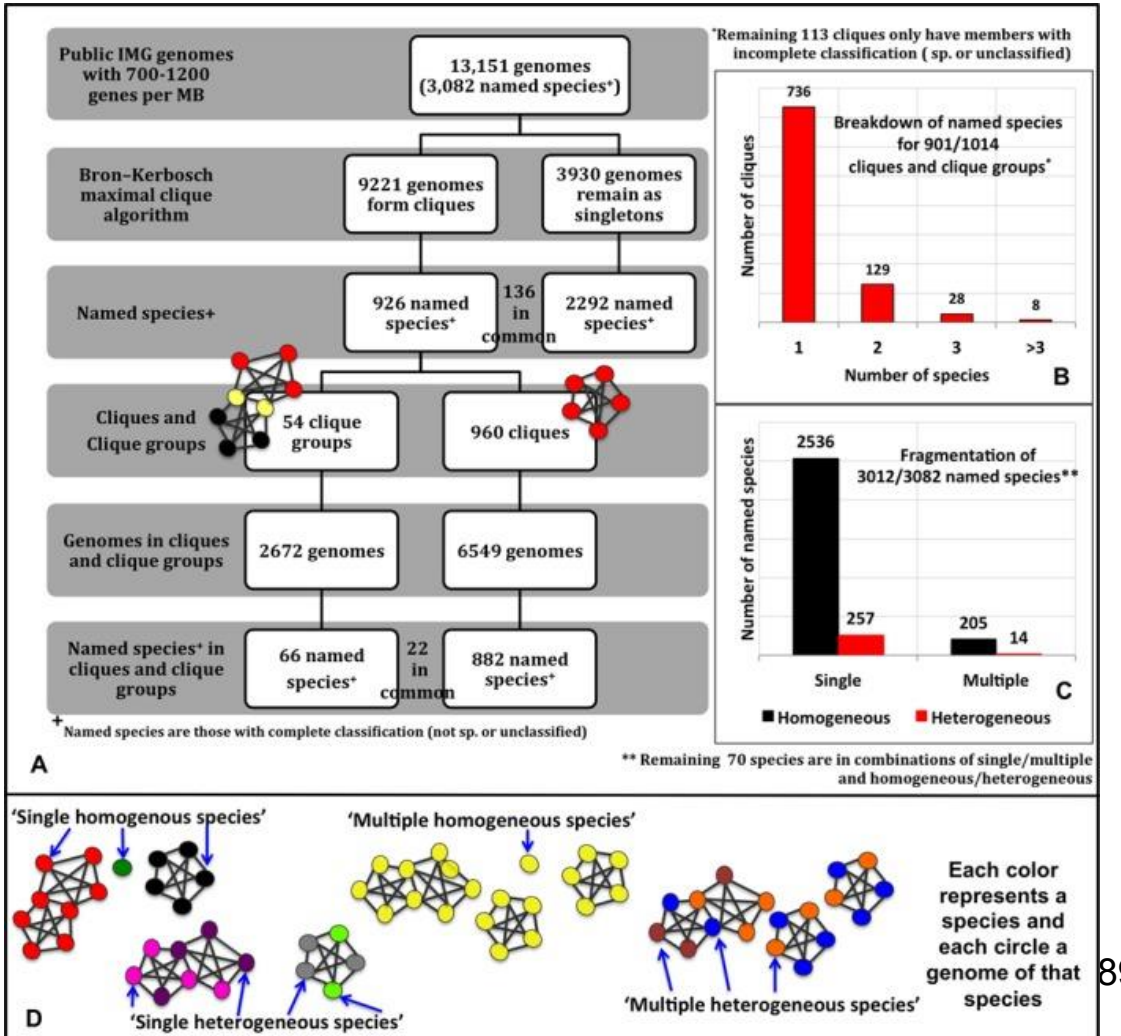
D. Pictorial description of the species categories. Each color represents a species and each circle a genome of that species. Species 'black' is present only in one clique and that clique only has genomes of that species, making it a 'single homogeneous species'. Species 'pink' has genomes in only one clique group, but that clique group has genomes from other species ('purple'), thus making it a 'single heterogeneous species'. The genomes of species 'yellow' are present in multiple clusters, but each cluster they belong to only have genomes of that species, making it a 'multiple homogeneous species'. Species 'blue' has genomes in multiple clusters and each cluster has genomes of other species (orange, brown), thus making species 'blue' a 'multiple heterogeneous' species.

Clique: a group of individuals who interact with one another and share similar interests. Individuality and cliques are common in bacterial communities.

Varghese *et al.*,2015

Whole genome sequencing (WGS)
Microbial species delineation using whole genome sequences

Microbial species delineation using whole genome sequences.
MiSI method was used for clustering of genomes and Species assignment.



Varghese *et al.*,2015

The Species Concept in Prokaryotes

New threshold for a species delineation

Genome-wide Average Nucleotide Identity (gANI) metric

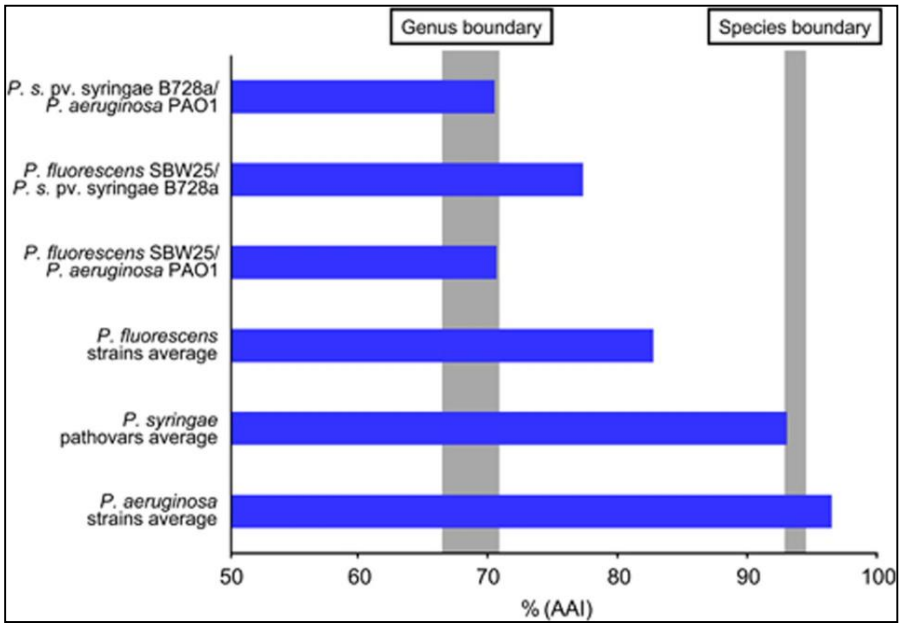
- Average nucleotide identity (ANI) and average amino acid identity (AAI) use pairwise comparisons of isolates based on genes and proteins conserved between particular pairs to determine relationships.
- Using MLST, gANI and AAI, and a gene-independent analysis of dinucleotide composition, recent studies have directly examined the relatedness of different bacterial species within the recognized genera such as *Pseudomonas* spp., *Erwinia* spp., *Rhizobium* spp., etc.

The Species Concept in Prokaryotes

New threshold for a species delineation

Comparison of ANI and DNA-DNA hybridization similarity

- AAI for pairs of *Pseudomonas* spp.
- While *Pseudomonas fluorescens* strains do not reach the species cutoff, they are more closely related than SBW25 is to PAO1.
- Genus and species boundaries are as defined.



In particular, an established threshold of 95% sequence identity by ANI was found to correspond to the traditional 'gold standard' prokaryotic species threshold of 70% identity by DDH.

Silby *et al.*,2011

The Species Concept in Prokaryotes

New threshold for a species delineation

Comparison of ANI and 16S rRNA gene sequence similarity

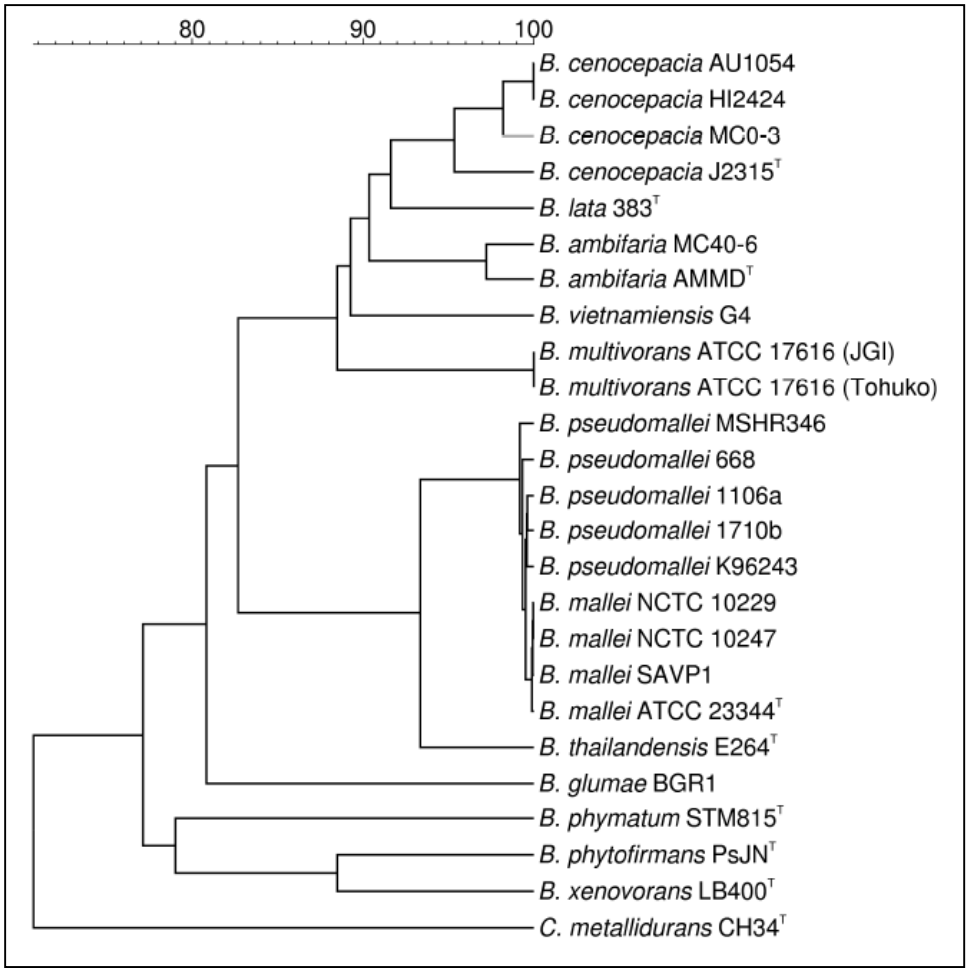
- Average nucleotide identity (ANI) is one of the most robust measurements of genomic relatedness between strains.
- Because there is a close correlation between ANI and 16S rRNA gene sequence similarity.
- The optimal threshold of 16S rRNA gene sequence similarity for species delineation that corresponds to 95-96% ANI.

Kim *et al.*,2014

The Species Concept in Prokaryotes

New threshold for a species delineation

ANI based phylogeny



Vandamme,2013

The Species Concept in Prokaryotes

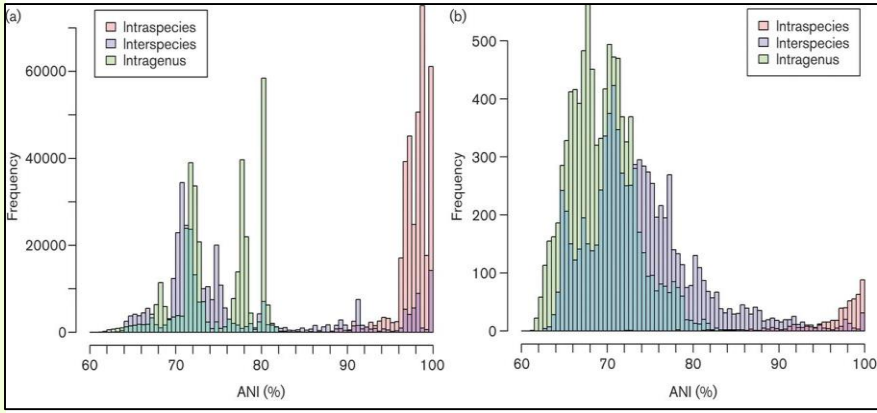
New threshold for a species delineation

Comparison of ANI and 16S rRNA gene sequence similarity

a) ANI distribution based on all genomes and
b) A normalized set of genomes.

- Each colour represents pairwise ANI calculations between strains belonging to different taxonomic ranks:

1. between two strains belonging to the same species (orange),
2. belonging to different species (green) and
3. belonging to different genera (purple).
4. Mean ANI values per species are represented in (b).



Using ANI as a substitute for DDH, our proposed threshold of 98.65% would greatly speed up the process of recognizing novel species.

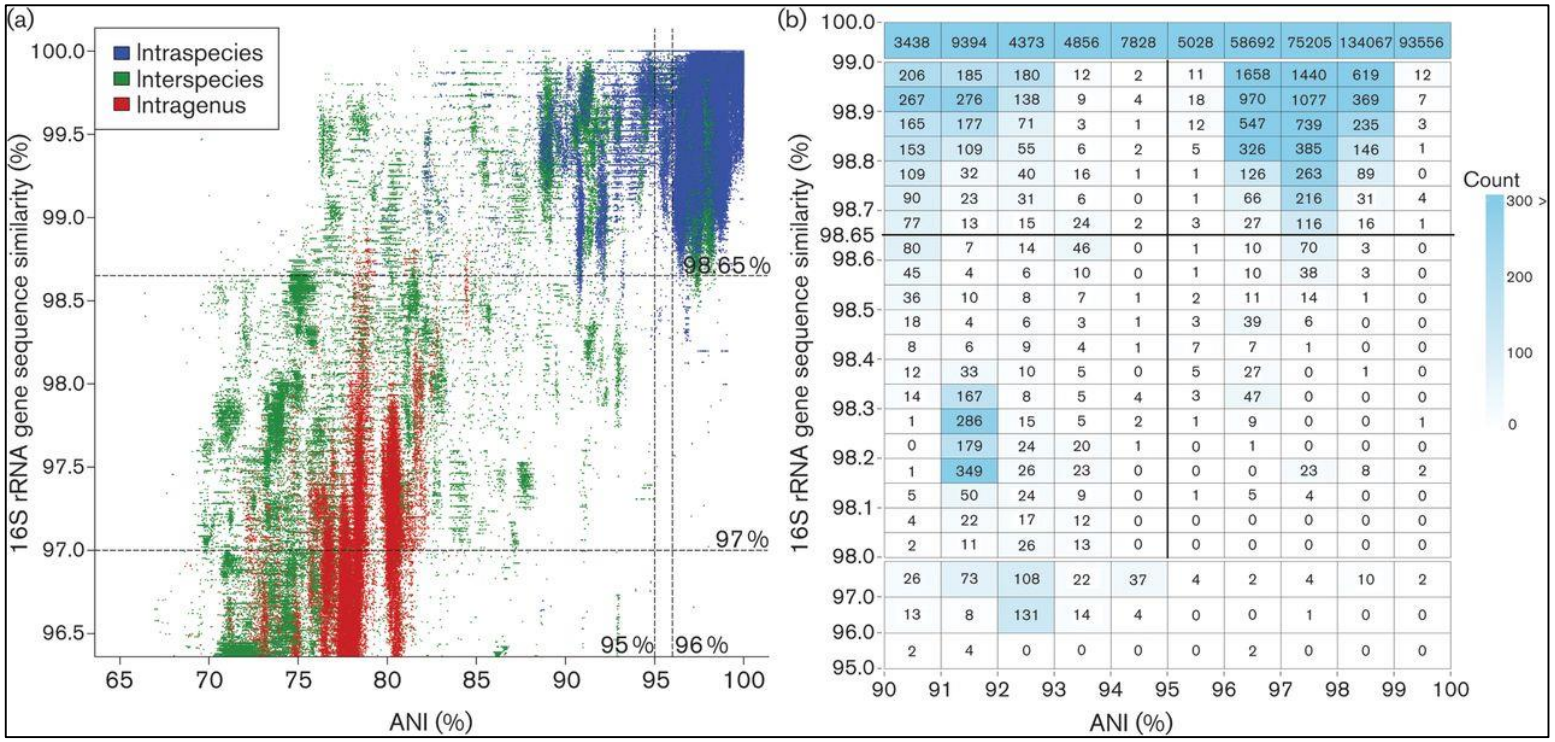
Kim *et al.*,2014

The Species Concept in Prokaryotes

New threshold for a species delineation

Comparison of ANI and 16S rRNA gene sequence similarity

- Association plot (a) and table (b) between ANI values and 16S rRNA gene sequence similarities. The number of strain pairs is displayed in each category square when divided by several intervals in 16S rRNA gene sequence similarity.



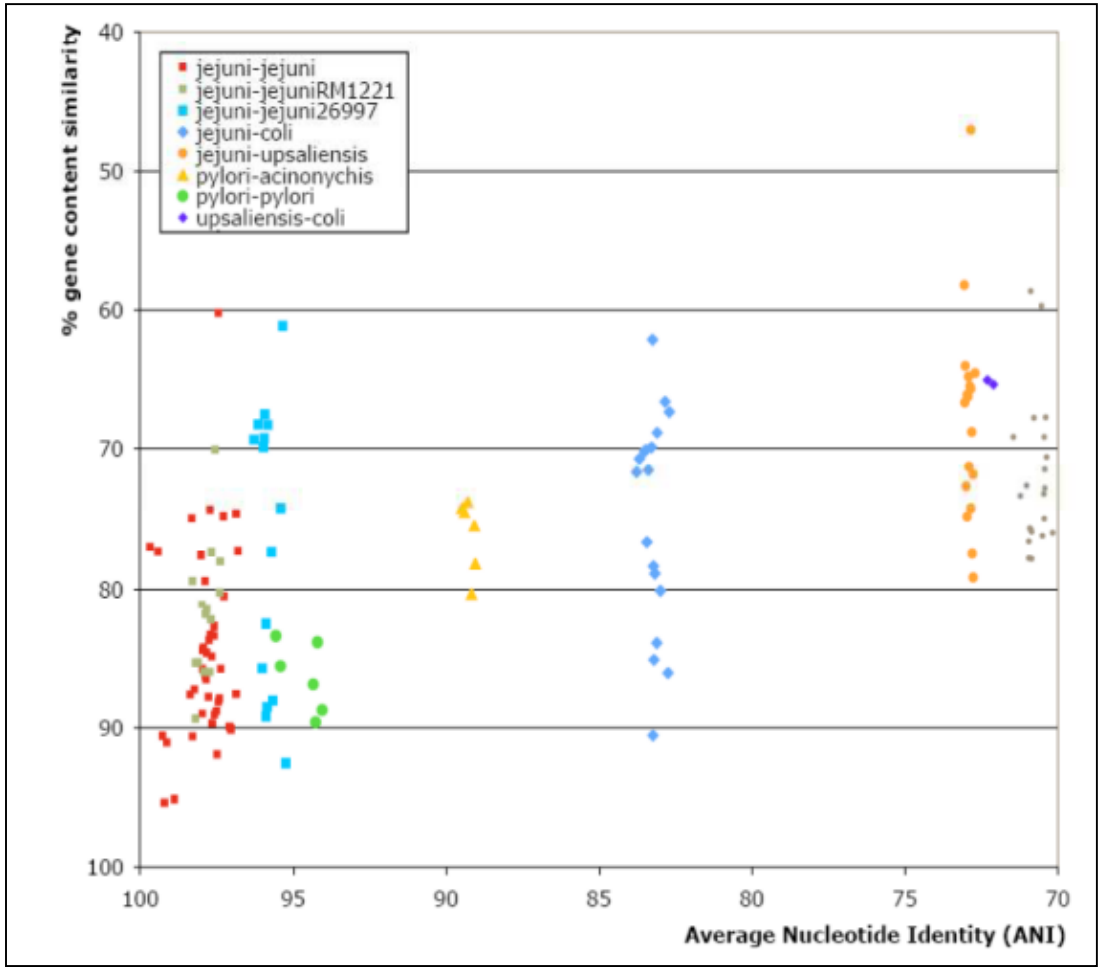
16S rRNA gene sequence similarity (%)	3438	9394	4373	4856	7828	5028	58692	75205	134067	93556
99.0	206	185	180	12	2	11	1658	1440	619	12
98.9	267	276	138	9	4	18	970	1077	369	7
98.8	165	177	71	3	1	12	547	739	235	3
98.7	153	109	55	6	2	5	326	385	146	1
98.65	109	32	40	16	1	1	126	263	89	0
98.6	90	23	31	6	0	1	66	216	31	4
98.5	77	13	15	24	2	3	27	116	16	1
98.4	80	7	14	46	0	1	10	70	3	0
98.3	45	4	6	10	0	1	10	38	3	0
98.2	36	10	8	7	1	2	11	14	1	0
98.1	18	4	6	3	1	3	39	6	0	0
98.0	8	6	9	4	1	7	7	1	0	0
97.0	12	33	10	5	0	5	27	0	1	0
96.0	14	167	8	5	4	3	47	0	0	0
95.0	1	286	15	5	2	1	9	0	0	1
	0	179	24	20	1	0	1	0	0	0
	1	349	26	23	0	0	0	23	8	2
	5	50	24	9	0	1	5	4	0	0
	4	22	17	12	0	0	0	0	0	0
	2	11	26	13	0	0	0	0	0	0
	26	73	108	22	37	4	2	4	10	2
	13	8	131	14	4	0	0	1	0	0
	2	4	0	0	0	0	2	0	0	0

Kim *et al.*,2014

The Species Concept in Prokaryotes

New threshold for a species delineation

Comparison of ANI and 16S rRNA gene sequence similarity



Vandamme,2013

Bacterial systematics, ecology, and evolution

Ecology-based approaches

Cohan Ecotype Concept

Frederick M. Cohan

Professor of biology at Wesleyan University

Email: fcohan@wesleyan.edu

Biographical Sketch

- Frederick Cohan studies the origins of ecological diversity in bacteria.
- He is a professor of biology and environmental studies at Wesleyan University, where he teaches classes in the origins of bacterial diversity, evolutionary bioinformatics, global change and infectious disease, and introductory biology.
- He graduated from Pasadena High School and earned his BS at Stanford in Biological Sciences.
- He was the first to earn a PhD from Harvard's Organismic and Evolutionary Biology Department.

His group has also developed an "ecotype simulation" approach to determine the sequence clusters that correspond to bacterial ecotypes (See ecotype simulation web site).

Bacterial systematics, ecology, and evolution

Speciation

Evolution of new species

- Biology features a lot of big questions, and how new species evolve is one of the biggest and most important.
- Speciation increases diversity and complexity, and, importantly, it allows organisms to explore new evolutionary paths.
- There is little in biology that isn't touched by speciation, and it is little wonder that Darwin himself referred to it as "that mystery of mysteries."

Whittaker,2013

Bacterial systematics, ecology, and evolution

Definition of species

The broad definition

- Are bacterial species real? Are there species in the world of bacteria?
- There is no doubt that species exist, and that it is useful to classify, name, and describe them.
- In bacterial systematics, species are defined as groups of close relatives that are separated from one another by large gaps in phenotypic and molecular characters (Vandamme *et al.*,1996).
- This broad definition practiced by bacterial systematics is not implemented and not totally acceptable by the microbial ecologists and evolutionary biologists.

Cohan,2006;Kopac and Cohan,2011

Bacterial systematics, ecology, and evolution

Definition of species

Cohan definition

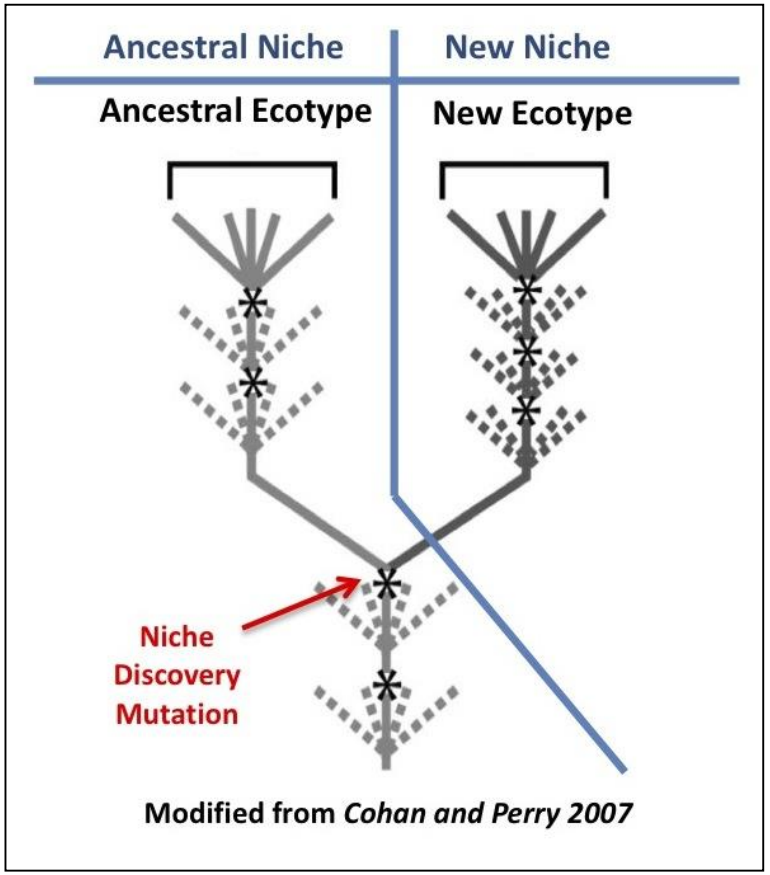
- We define ecotype here as a phylogenetic group of close relatives that are ecologically very similar, in that the members of an ecotype share genetic adaptations to a particular set of habitats, resources, and conditions.
- More specifically, different ecotypes are predicted to coexist indefinitely as a result of their ecological differences, while lineages within one ecotype are ecologically too homogeneous to allow indefinite coexistence.
- The present definition of ecotype implies no other species-like characteristics beyond ecological distinctness.

Cohan,2015

Bacterial systematics, ecology, and evolution

Frederick Cohan's Ecotype Species Concept (ESC)

- Frederick Cohan's Ecotype Species Concept (ESC), emphasizes the evolution of ecological differences.
- It argues that a bacterial species is born when a mutation grants access to a new niche, resulting in a new lineage called an ecotype.
- Because the new ecotype and its parent occupy difference niches, they are able to coexist and evolve independently.



Whittaker,2013

Bacterial systematics, ecology, and evolution
Variation in numbers

- Bacteria have been exploring evolutionary adaptations much longer than eukaryotes.
- There are ~10,000 described bacterial species.
- However, the vast majority of bacterial species have not been described.
- Estimates range considerably, but <1% of bacteria are believed to be able to be cultured.
- Estimates place the total number of bacterial species at 5-10 million. That's an almost unanswerable question.

Note: The number of recognized bacterial species is constantly updated by this site:
http://www.bacterio.cict.fr/number.html#total
The current number is 10,929.

Bacterial systematics, ecology, and evolution
Variation in numbers

- Genome sequencing has become the standard for the study of bacterial species.
- Of the described species:
- As of October 2013,
- 6851 bacterial genomes have been completed;
- Sequencing of 18679 are ongoing, and
- 1111 represent targeted genome projects (Gold Genomes OnLine Database, http://www.genomesonline.org)

Kishore *et al.*,2015

Bacterial systematics, ecology, and evolution
Variation in numbers and traits

- Re-analyses of the Torsvik work suggests tens of millions of species in that soil collection of Torsvik.
- Computational improvements reveal great bacterial diversity and high metal toxicity in soil and that there might be billions of bacterial species worldwide:
- Why are there so many species of bacteria?

Courtesy Cohan,2011

Bacterial systematics, ecology, and evolution
Variation in numbers and traits

- Molecular surveys suggest that bacterial species may number in the millions or even billions (Cohan,2011).
- According to pioneering work of Torsvik *et al.*,1998 and 2002, 30g of forest soil held about 10000 species.
- There is variation within a typical species recognized by systematics:

1. genome content,
2. DNA sequence, phenotype, and
3. ecology typically seen within the named bacterial species.

Courtesy Cohan,2006

Bacterial systematics, ecology, and evolution

Species demarcations under different criteria

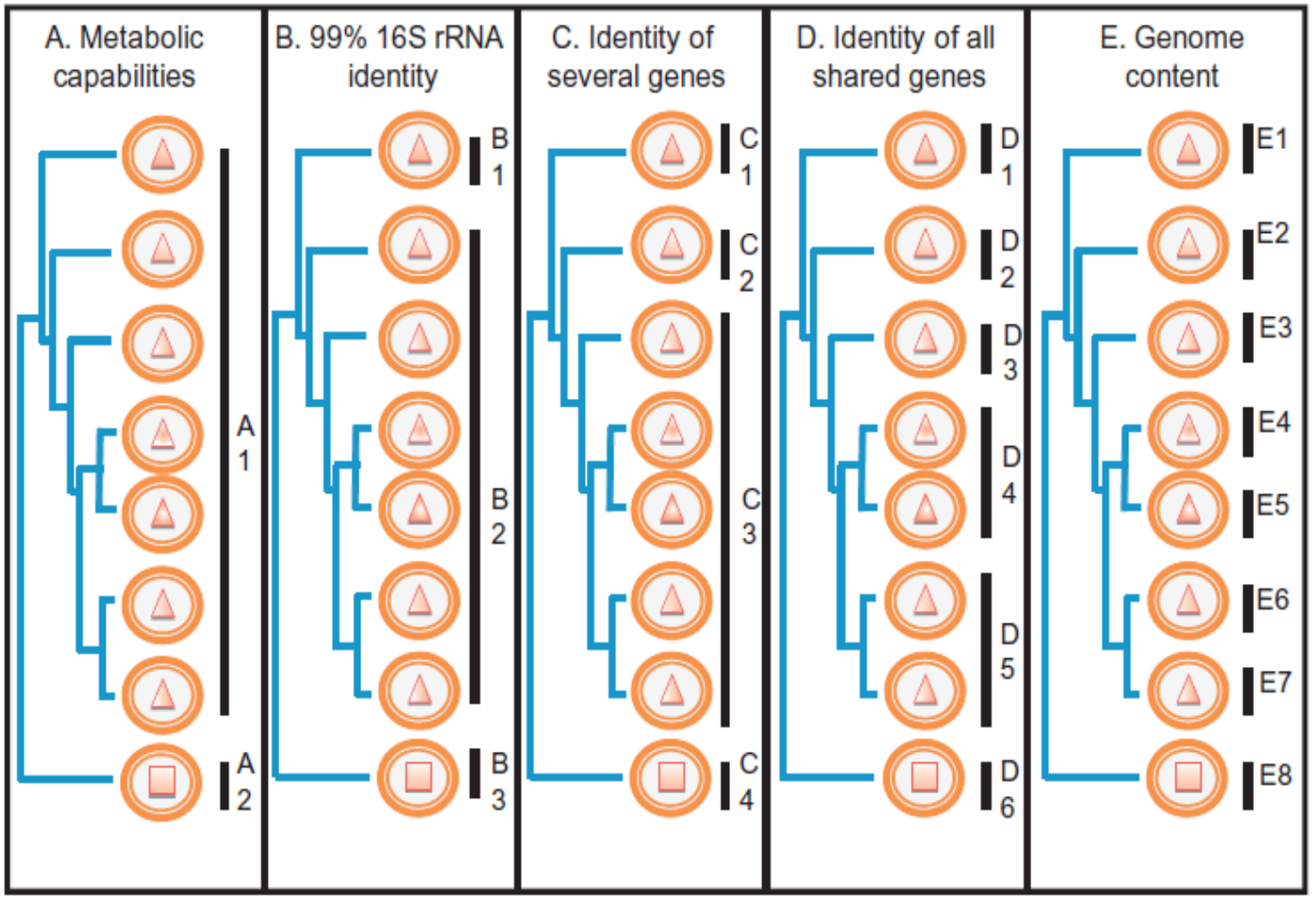
A. Variation in phenotypes
B. Variation in 16S rRNA gene
C. Variation in several genes
D. Variation in shared genes
E. Variation in genome content
F. Variation in ecological diversity

Kopac and Cohan,2011

Bacterial systematics, ecology, and evolution

Species demarcations under different criteria

Gene content reflects ecology (split close relatives that are ecologically identical into different species)



Kopac and Cohan,2011

Bacterial systematics, ecology, and evolution

Species demarcations under different criteria

A. Variation in phenotypes:

- Species were originally defined as groups that differ to a large extent in metabolic capability (indicated by triangle versus square), frequently with much metabolic diversity within each species (indicated here by shading differences within the triangles).

Kopac and Cohan,2011

Bacterial systematics, ecology, and evolution
Species demarcations under different criteria

B. Variation in 16S rRNA gene:

- Defining a species as a group of organisms sharing at least 99% 16S rRNA identity can split the metabolically defined species in the previous panel, as seen here by the splitting of species A1 into B1 and B2.

Kopac and Cohan,2011

Bacterial systematics, ecology, and evolution
Species demarcations under different criteria

C. Variation in several genes:

- Defining species as: clusters based on several protein-coding gene sequences can split a 16S-defined species into groups that are each more ecologically homogeneous.
- This is seen here by the splitting of species B2 into C2 and C3.

Kopac and Cohan,2011

Bacterial systematics, ecology, and evolution
Species demarcations under different criteria

D. Variation in all shared genes:

- Defining species as: clusters based on sequence identity for all shared genes can divide species even further with, for example, species C3 being split into D3, D4, and D5.
- This may be the most highly resolving method for identifying species based on sequences of shared genes.
- Within species D4, we can see the possibility that even with this level of resolution for species demarcation, there may still be ecological heterogeneity (indicated by the difference in shading between cells in species D4).
- Species D5 shows an alternative model where this high level of resolution finds clusters that are ecologically homogeneous, as noted by the same shading patterns among members of D5.

Kopac and Cohan,2011

Bacterial systematics, ecology, and evolution
Species demarcations under different criteria
Variability in gene content

E. Variation in genome contents:

- Genomes seem to be composed of:

1. A core set of genes that is conserved among strains of the same species and
2. Accessory genes that are strain specific.

- Content and size of core vary with species.

Variation in genome GC content, which varies in different organisms from as low as 17% to as high as 75%. This variation is generally ascribed to differences in the pattern of mutation between bacteria.

Konstantinidis and Tiedje,2005;Vandamme,2013

Bacterial systematics, ecology, and evolution
Species demarcations under different criteria

E. Variation in genome contents:

- Defining species by identity of genome content could spuriously split close relatives that are ecologically identical into different species.
- Note that the two organisms within D5, with the same ecology, are split on the basis of genome content into different species.
- In this case, E6 and E7 would most likely be different for phage or insertion sequence genes that do not specify ecological niche.

Kopac and Cohan,2011

Bacterial systematics, ecology, and evolution

Definition of species

Drawbacks of old approaches

- The systematics of bacteria does not aim to demarcate diversity at fine level of ecotypes, and indeed, we have shown that the named species of bacteria typically contain multiple ecotypes (a group of bacteria that are ecologically similar to one another).

Kopac and Cohan,2011

Genome sequencing

Definition of species

The new approaches

- The information emerging from genome sequencing, together with information from other approaches such as gene expression studies, should eventually converge to a more soundly based bacterial species definition.

Konstantinidis *et al.*,2006

Bacterial systematics, ecology, and evolution
Definition of species
The new approaches

- The molecular revolution has taken us far beyond the early days of systematics, when species demarcation was based entirely on metabolism and other phenotypic traits.
- Sequencing has now revealed ecologically distinct populations within the recognized species, yet we do not take advantage of this information to refine the demarcations of species.
- The time has come to incorporate the high resolution of molecular technology into our taxonomy, so that the physiological and ecological diversity we know to exist within the named species can be officially recognized.

Kopac and Cohan,2011

Bacterial systematics, ecology, and evolution
F. Variation in ecological diversity
Ecotype concept

- Even though bacterial species were originally demarcated as phenotypic (usually metabolic) clusters, most species recognized by bacterial systematics are highly diverse in their metabolic capabilities.
- The molecular methods suggest that a typical named species contains many ecotypes, each with the universal attributes of species.
- **A named bacterial species is thus more like a genus than a species.**

Cohan,2002;2006

Bacterial systematics, ecology, and evolution
Variation in ecological diversity
What is an ecotype?

- Although the ecotype is not recognized as an official rank of bacterial taxonomy (Brenner *et al.*,2000), the concept of ecotype is important for microevolutionary studies because it describes a collection of strains that shows some level of ecological distinctiveness within its species (Schloter *et al.*,2000; Cohan,2001).

Konstantinidis *et al.*,2006

Bacterial systematics, ecology, and evolution
Variation in ecological diversity
The challenge to microbial ecology

- The challenge to microbial ecology is:

1. To identify the ecologically distinct bacterial groups (ecotypes) within a community, and
2. To determine what differences allow them to coexist and to perform different ecosystem functions.

Cohan,2002

Bacterial systematics, ecology, and evolution
Differences in ecotypes of the same species

- Members of the same ecotype would therefore be expected to have only small gene content (or expression) differences compared with other ecotypes of the same species, or
- If larger differences exist, they would be carried by unstable parts of the genome such as plasmids.
- Our analysis reveals that justifiable ecotypes are observable among homogeneous groups of strains (e.g. *Streptococci*), even among almost identical strains of the *E. coli* and *Salmonella* groups.

Konstantinidis *et al.*,2006

Bacterial systematics, ecology, and evolution
HGT and the origins of bacterial species

- According to Doolittle and colleagues, there is a continuum of genetic diversity, with no two strains in nature having the same ecology.
- Cohan hypothesis is that there are many newly divergent ecotypes, each being ecologically homogeneous.
- But which idea is true (and whether there is a true continuum of diversity) is not clear yet.

Konstantinidis *et al.*,2006;…

Bacterial systematics, ecology, and evolution
HGT and the origins of bacterial species

- Species are cohesive, in the sense that genetic diversity within a species is limited by a force of evolution.
- The origin of bacterial species is accelerated by unique features of bacterial genetics, perhaps the most important being the ability of bacteria to readily acquire genes from other organisms.
- Early evidence of the importance of horizontal genetic transfer (HGT) in bacterial evolution was seen in the spread of penicillin resistance through plasmid transfer across the Enterobacteriaceae.

Wiedenbeck & Cohan,2011

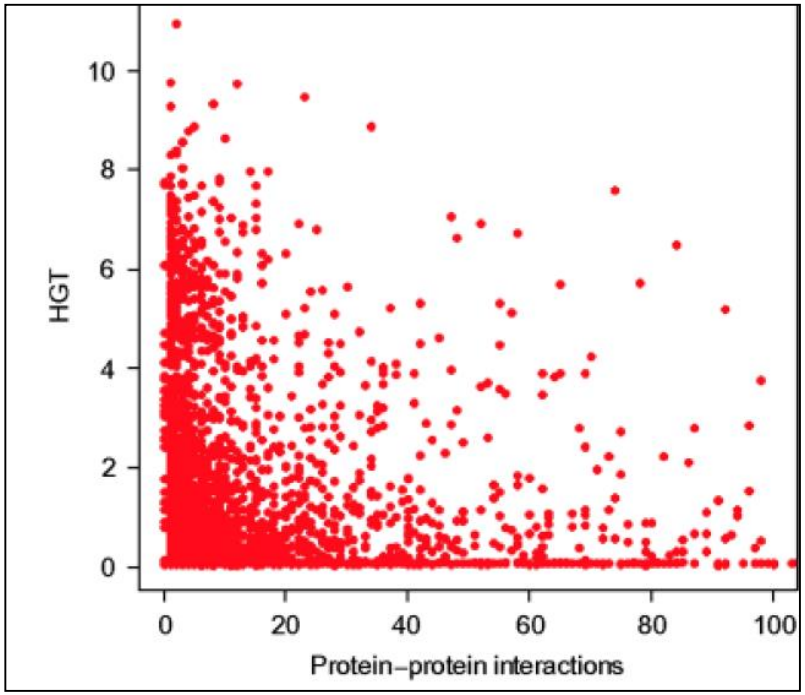
Bacterial systematics, ecology, and evolution
HGT and the origins of bacterial species

- More recent HGT events have resulted in important ecological differences between closely related species and between populations within a single recognized species taxon (Welch *et al.*,2002).
- For example, the virulence factors that distinguish *Salmonella* from *Escherichia coli* were largely acquired by HGT.

Wiedenbeck & Cohan,2011

Bacterial systematics, ecology, and evolution
HGT and the origins of bacterial species

- The relationship between the number of HGT events and the number of protein-protein interactions.
- Each dot represents a gene family, and the number of protein-protein interactions for a given family is quantified as the number of other gene families with which the given family interacts.



Cohen,2011;Oxford University Press

Ecology-based approaches

An Ecotype-Based Systematics

New approach in defining bacterial species

- Ecology-based approaches (ecological differences of bacterial populations) will help the bacterial systematics to define the species more precisely.
- Most importantly, a named bacterial species is typically an assemblage of ecologically distinct populations that are able to coexist in the same region (Scholer *et al.*,2000; Lopez-Lopez *et al.*,2005 & Smith *et al.*,2006).
- We have proposed a systematics for identifying ecotypes, the fundamental units of bacterial ecology and evolution.

Cohan,2006;Cohan and Perry,2007

Ecology-based approaches

An Ecotype-Based Systematics

New approach in defining bacterial species

- Fortunately bacterial systematics like plant and animal systematics now aims to identify species that represented the fundamental units of ecology and evolution.
- But, the established institutions of bacterial systematics (e.g., the DSMZ) may have no intention of changing systematics to recognize ecotypes.

Cohan,2006; Cohan and Perry,2007;…

Bacterial systematics, ecology, and evolution
Definition of species
Alternative theories of bacterial speciation

- We are testing alternative theories of bacterial speciation:

1. The Stable Ecotype model (where ecotypes are long-standing and are recurrently purged of diversity by periodic selection),
2. The Species-Less model (where ecotypes are not subject to cohesive forces), and
3. The Nano-Niche model (where most-closely-related ecotypes are only quantitatively different in their ecological niches).

Cohan,2015

Periodic Selection and Ecological Diversity in Bacteria

The stable ecotype model

The dynamics of ecotype formation and periodic selection within an ecotype

- Diversity within an ecotype is only transient, awaiting its demise(the end of existence or activity) with the next periodic selection event.
- The question is what is the source of permanent divergence among closely related bacteria?
- Ecological diversity in bacteria is governed by different mutations (or recombination events).
- When a mutation (or recombination event) places the organism into a new ecological niche and the organism thereby founds a new ecotype.

Courtesy Cohan,2006

Periodic Selection and Ecological Diversity in Bacteria

The stable ecotype model

The dynamics of ecotype formation and periodic selection within an ecotype

- The adaptive mutation (indicated by an asterisk) originally occurs in Ecotype 1.
- After the selective sweep, a small region of chromosome around the adaptive mutation (between 1L and 1R) enters Ecotype 2, causing a selective sweep in that ecotype.
- Then a segment around the adaptive mutation (between2L and 2R) enters Ecotype 3, causes a selective sweep there, and so on.

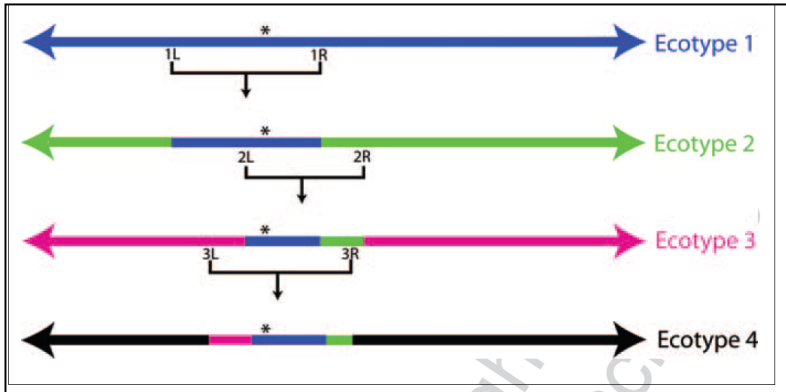


Figure 6. In the adapt globally, act locally model, the region that is homogenized is expected to differ between each pair of ecotypes. The adaptive mutation (indicated by an asterisk) originally occurs in Ecotype 1. After the selective sweep, a small region of chromosome around the adaptive mutation (between 1L and 1R) enters Ecotype 2, causing a selective sweep in that ecotype. Then, a segment around the adaptive mutation (between 2L and 2R) enters Ecotype 3, and causes a selective sweep there, and so on. The source of DNA along the chromosome is indicated by shade. The entire ensemble of ecotypes becomes homogenized for the sequence near the adaptive mutation, but the boundaries of the homogeneous region differ for each pair of ecotypes. For example, Ecotypes 1 and 2 are identical between 1L and 1R, while Ecotypes 2 and 4 are identical between 2L and 3R.

Here is showing how, even though there is genome-wide purging within each ecotype, there can be a sweep of a small chromosome segment across many different ecotypes.

Cohan,2005

Periodic Selection and Ecological Diversity in Bacteria

The stable ecotype model

The dynamics of ecotype formation and periodic selection within an ecotype

- Because the new ecotype is ecologically distinct from the parental ecotype, periodic selection events in the parental ecotype cannot extinguish (end or death of) the founding organism and its descendants (fig).

1. Thus, the new ecotype escapes the periodic selection of the parental ecotype, and
2. The two new ecotypes are free to diverge indefinitely.

Courtesy Cohan,2006

Periodic Selection and Ecological Diversity in Bacteria

The stable ecotype model

The dynamics of ecotype formation and periodic selection within an ecotype

a) Ecotype-formation event.

- A mutation or a recombination event allows the cell to occupy a new ecological niche, founding a new ecotype. A new ecotype can be formed only if the founding organism has undergone a fitness trade-off, whereby it cannot compete successfully with the parental ecotype in the old niche.

b) Periodic selection event.

- A periodic selection mutation improves the fitness of an individual such that the mutant and its descendants outcompete all other cells within the ecotype; these mutations do not affect the diversity within other ecotypes because ecological differences between ecotypes prevent direct competition.
- Periodic selection leads to the distinctness of ecotypes by purging (removing) the divergence within, but not between ecotypes.

Wiedenbeck & Cohan,2011

Periodic Selection and Ecological Diversity in Bacteria

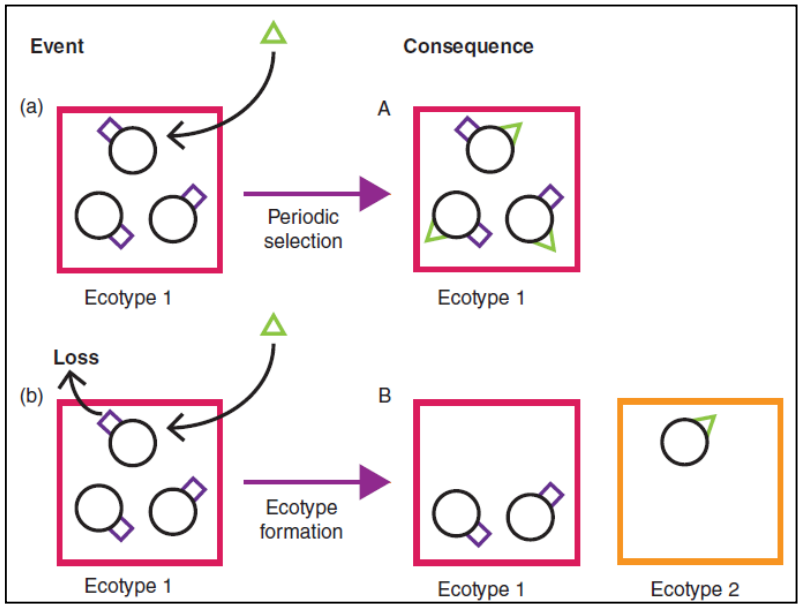
The stable ecotype model

The consequences of a change in ecological niche following an HGT event

- Acquisition of a new ecological function by HGT (indicated by the green triangle) either cause:

A. A periodic selection event (improves the fitness of an individual), or

B. An ecotype formation event (allows the cell to occupy a new ecological niche, founding a new ecotype).



Wiedenbeck & Cohan,2011

New species can arise as frequently as adaptations within a species

Identification of putative ecotypes

- We could identify putative ecotypes formed in different communities through differences in:

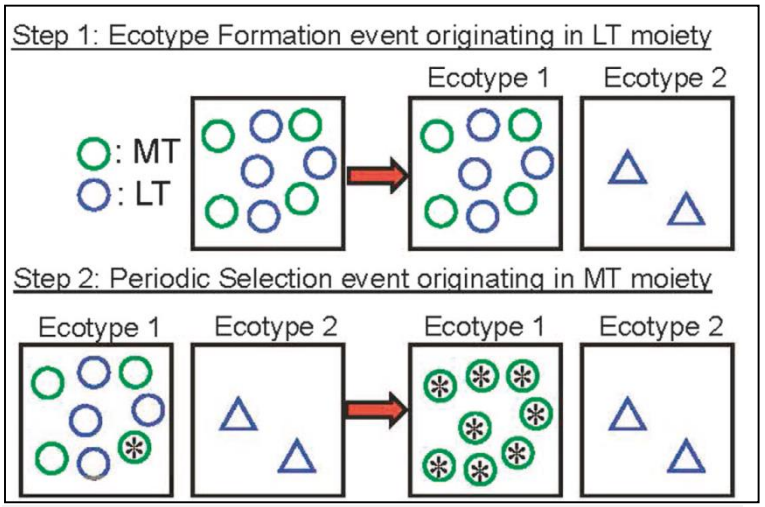
1. Genetic marker association,
2. Colony morphology, and
3. Microhabitat association.

Koeppel *et al.*,2013

Periodic Selection and Ecological Diversity in Bacteria

New species can arise as frequently as adaptations within a species

- Each box represents an ecotype;
- Each circle or triangle represents an individual organism within an ecotype;
- The different colors indicate the selectable markers methionine (MT) and lysine (LT);
- An asterisk indicates a mutant strain that is adaptive within its ecotype's niche.

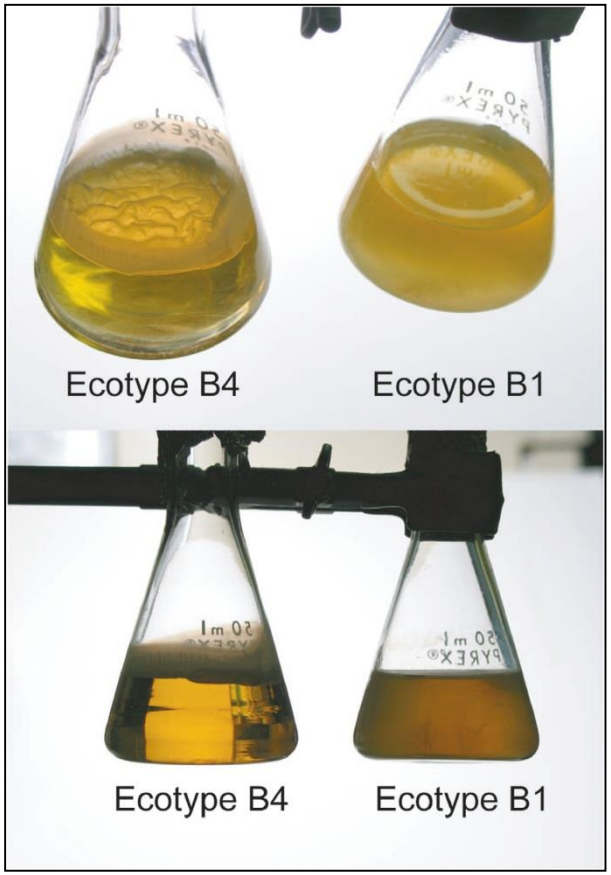


We generated prototrophic revertant mutants for lysine (MT) and methionine (LT) by selection on plates with minimal medium (including Spizizen salts, glucose and 10 mg^{ml1} required amino acids). Each colony was streaked for isolation three times on modified Luria broth plates and selectable markers (LT or MT) were reconfirmed on selective minimal media.

Koeppel *et al.*,2013

Evolution in a bottle
Mat formation

- Differences in mat-forming ability of putative ecotypes B4 and B1 within community B.
- Putative ecotype B4 (LT, morphotype 4) formed a dense mat on the surface of the medium,
- While putative ecotype B1 grew in the broth phase, suggesting niche partitioning on the basis of microhabitat specialization.



Koeppel *et al.*,2013

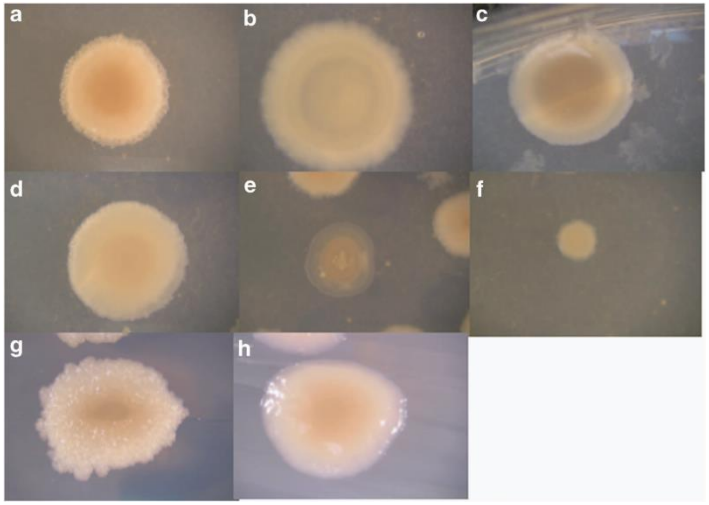
New species can arise as frequently as adaptations within a species

Colony morphology

- We could identify putative ecotypes through differences in colony morphology.

a. Morphotype 1 was the ancestral morphotype.

- All other morphotypes were unique to a particular marker moiety (LT or MT) and a single community.



As mentioned earlier, the prototrophic revertant mutants for lysine (MT) and methionine (LT) was generated by selection on plates with minimal medium Each colony was streaked for isolation three times on modified Luria broth plates and selectable markers (LT or MT) were reconfirmed on selective minimal media.

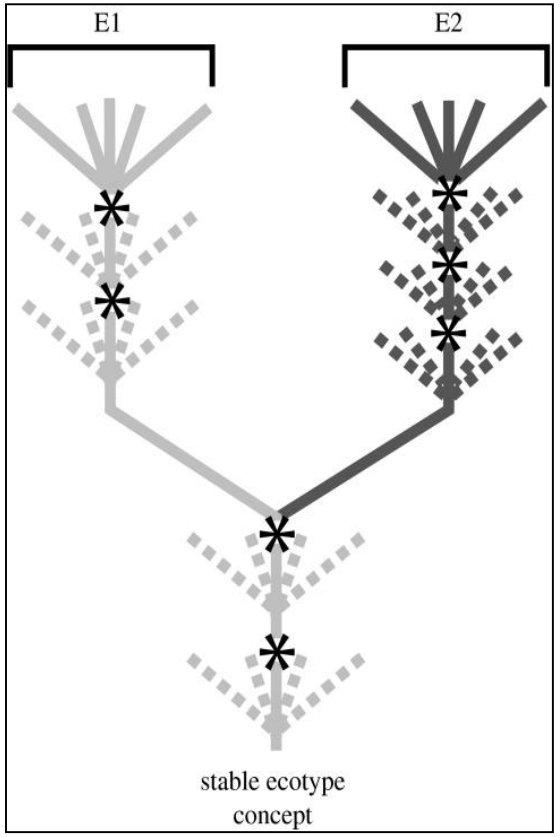
Koeppel *et al.*,2013

Periodic Selection and Ecological Diversity in Bacteria

The stable ecotype model

Source of permanent divergence among closely related bacteria

- The phylogenetic history of two closely related ecotypes under the stable ecotype model.
- After each periodic selection event, indicated by an asterisk, only one variant from an ecotype survives.
- After periodic selection, the descendants of the surviving variant diverge (indicated by dashed lines), but with the next periodic selection event, again only one variant survives.
- The sequence diversity within an ecotype is much less than sequence divergence between members of different ecotypes.



*, periodic selection event
---, descendants of the surviving variant diverge

Cohan,2006

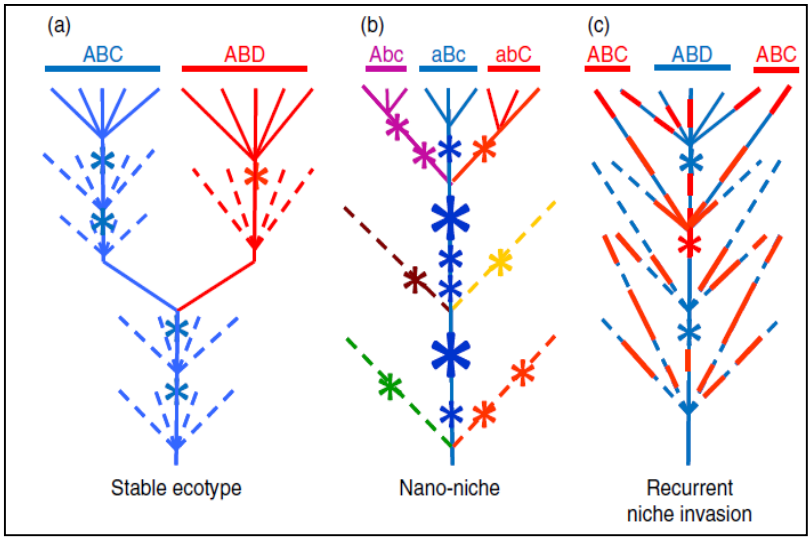
Periodic Selection and Ecological Diversity in Bacteria

The stable ecotype model

Models of bacterial speciation

1. Ecotypes are represented by different colors;
2. Periodic selection events are indicated by asterisks, and
3. Extinct lineages are represented by dashed lines.

- The Stable Ecotype model is marked by a much higher rate of periodic selection than ecotype formation.



Wiedenbeck & Cohan,2011

Periodic Selection and Ecological Diversity in Bacteria

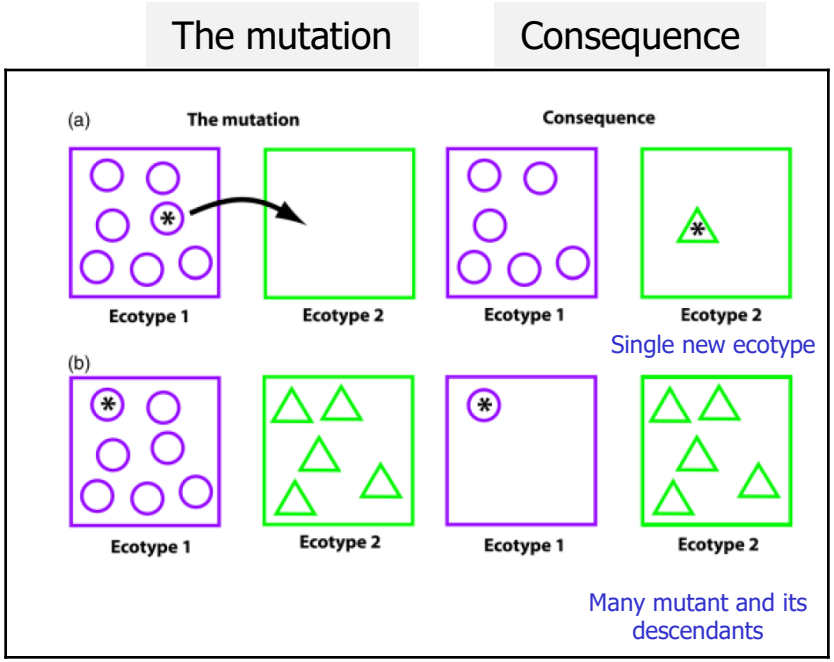
The stable ecotype model

The dynamics of ecotype formation and periodic selection within an ecotype

- Circles represent different genotypes, and asterisks represent adaptive mutations.

a) Ecotype-formation event: A mutation or a recombination event allows the cell to occupy a new ecological niche, founding a new ecotype.

b) Periodic selection event: A periodic selection mutation improves the fitness of an individual such that the mutant and its descendants outcompete all other cells within the ecotype.



Wiedenbeck & Cohan,2011

Periodic Selection and Ecological Diversity in Bacteria

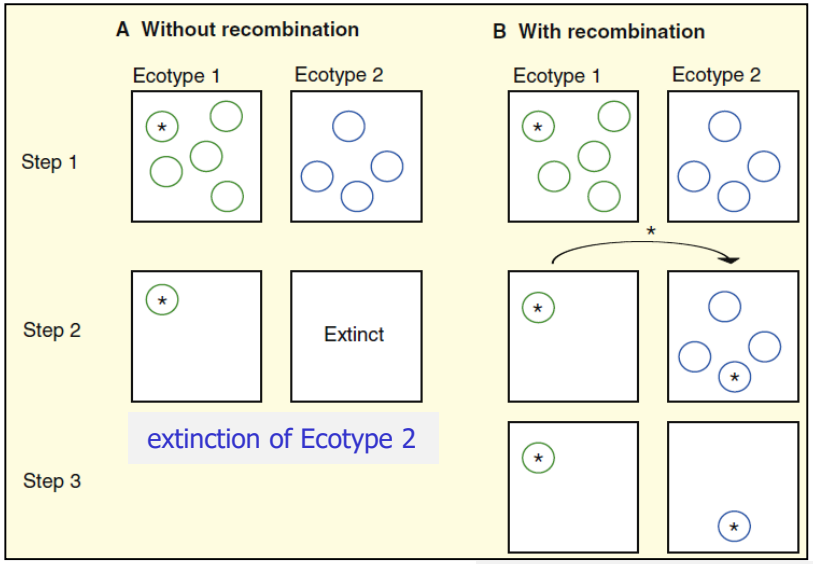
The stable ecotype model

Recombination between nascent species can promote their coexistence

- Recombination between nascent (recently formed or developed species) can promote their coexistence.
- Circles represent different genotypes within each ecotype.
- Asterisks represent a mutation giving an exceptionally competitive advantage to the genotype that possesses it.

A. When no recombination occurs, the adaptive mutant in Ecotype 1 causes the extinction of Ecotype 2, along with all other genotypes in Ecotype 1.

B. With recombination, the adaptive allele may be transferred into Ecotype 2, causing a periodic selection event in that ecotype, and allowing for the continued coexistence of both ecotypes.



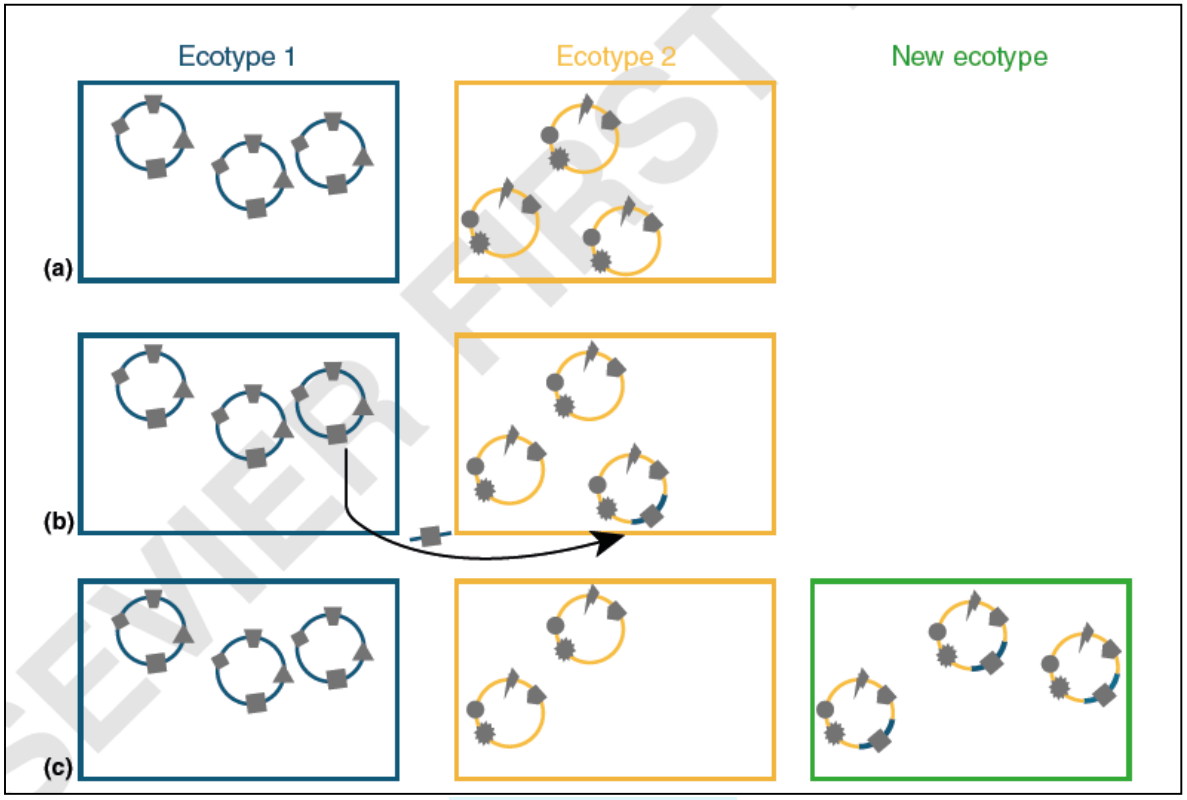
Cohan and Koeppel,2008

Periodic Selection and Ecological Diversity in Bacteria

The stable ecotype model

Recombination between nascent species can promote their coexistence

- The consequences of an adaptive genetic transfer between ecologically distinct populations when an adaptive transfer creates a new ecotype.



Cohan,2016

Periodic Selection and Ecological Diversity in Bacteria

Ecological theory

Reasons for coexistence of the species

- Ecological theory tells us that species cannot coexist for the long term unless:

1. They use different resources, or
2. Thrive (grow and develop) in different conditions, or
3. Respond differently to their predators and pathogens (Chase & Leibold,2003).

Cohan,2006

Periodic Selection and Ecological Diversity in Bacteria

Ecological theory

Reasons for coexistence of the species

- There are certainly many resource dimensions upon which bacteria could diversify:

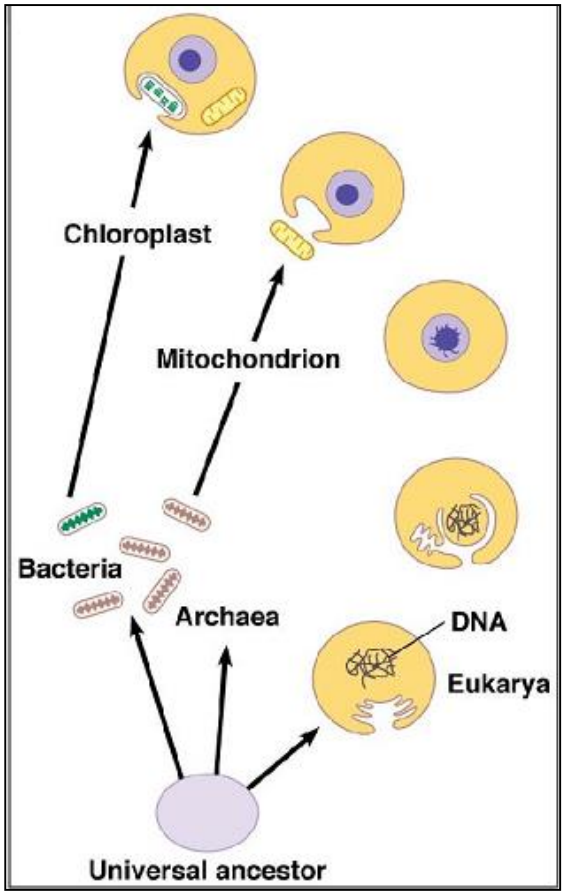
1. There are probably very few carbon sources that cannot be used by at least one bacterium; for example, some bacteria can use ordinary toxic compounds and ordinary insoluble macromolecules.
2. In addition, each of the many eukaryotic species may have its own exclusive bacterial endosymbionts and pathogens.
3. Bacteria can coexist on the basis of different unusual conditions under which they can grow, including an extreme range of salinity, temperature and radiation.

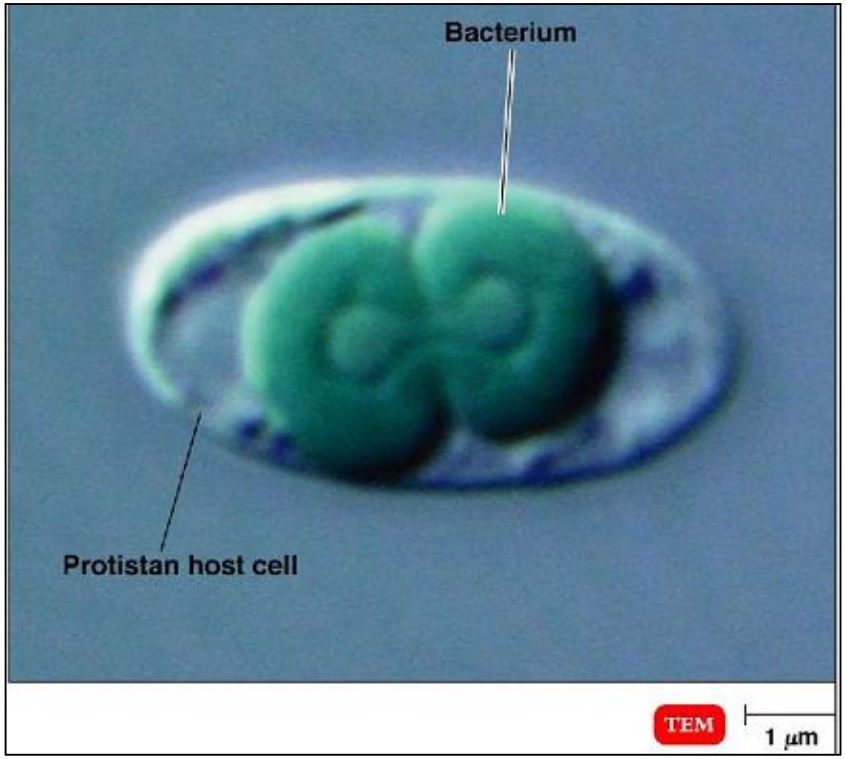
Courtesy Cohan,2006

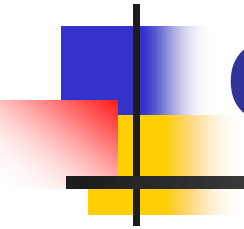
Periodic Selection and Ecological Diversity in Bacteria

Ecological theory

Reasons for coexistence of the species







Genome sequencing

rDNA sequencing

rDNA sequencing
Goal and Background

- **Goal**
- Learn and understand the steps necessary to make a phylogenetic assignment for a nucleotide sequence (here the bacterial 16S rDNA).
- **Background**
- The nucleotide sequence of extracted nucleic acid can be compared with a database of already assigned sequences.

DNA sequencing

- By the mid-1990s, sequencing of the small subunit (16S) rDNA genes had become commonplace, considered a standard tool of microbial taxonomists not only for elucidating phylogenetic relatedness but also as a means of bacterial identification.
- The automation of 16S rDNA gene sequencing with such instruments as the ABI Prism 377 DNA sequencer (Applied Biosystems, Foster City, Calif.) allowed for a quick comparative analysis of published sequences deposited in microbial genome databases.
- Today, bacterial strains that defy identification by conventional commercial methodologies are often subjected to 16S rDNA sequence analysis so that a useful label can be placed on the isolate in question.

Janda and Abbott,2002

DNA sequencing
rDNA sequences

- DNA sequencing has become established in the routine laboratory.
- Sequencing currently is being used for the:

1. Identification of organisms that are difficult to identify using conventional methods or
2. To detect and identify uncultivable organisms.

- These rDNA sequences are the cornerstone of studies on taxonomy and phylogenetic relationships between bacterial or fungal species.
- rDNA sequences of different organisms available in publicly-available databases are far greater than any other region of the genome (Ward *et al.,*2004).

DNA sequencing
16S rDNA sequences

- 16S rDNA is the normal target for sequencing.
- Some of the rDNA regions are well conserved throughout the evolutionary process and other parts are variable even within a species.
- However, the equivalent rDNA gene fragments detected in different organisms have sequence variations that can be exploited for the subsequent selection of pathogen-specific primer sequences allowing the identification of the pathogen in question.
- Other conserved regions in the microbial genome, such as *hrp* genes, 16-23S rDNA interspacer region, ribosomal spacer region, *rpoB* gene, *gyrB* gene, or elongation factor Tu, can be used (sequenced) as alternative targets to differentiation below the species level (pathovar level).

Steps of sequence analysis
Principles

1. Extract nucleic acid from isolates or amplify sequence in the polymerase chain reaction.
2. Purify pure culture or amplified DNA using a kit.
3. Send to sequencing facility for PCR based sequencing with an appropriate primer.
4. Edit sequence.
5. Using the web based database (The Ribosomal Database Project, RDP) find closest relative to the primary structure of this sequence with Sequence Match.
6. Align sequences with aligned sequences of the closest relatives from the RDP database using alignment software. Proof read the alignment: base the alignment of conserved and variable areas on secondary structure of a closely related organism (available from your instructor or specified web pages.
7. Create a distance matrix.
8. Test sequence for potential chimeric structures.
9. Export the aligned product and the aligned sequences of the closest relatives into software able to create a tree of phylogenetic relationship based on specific algorhythms.

- Now enough is known to design a probe for the analyzed sequence.

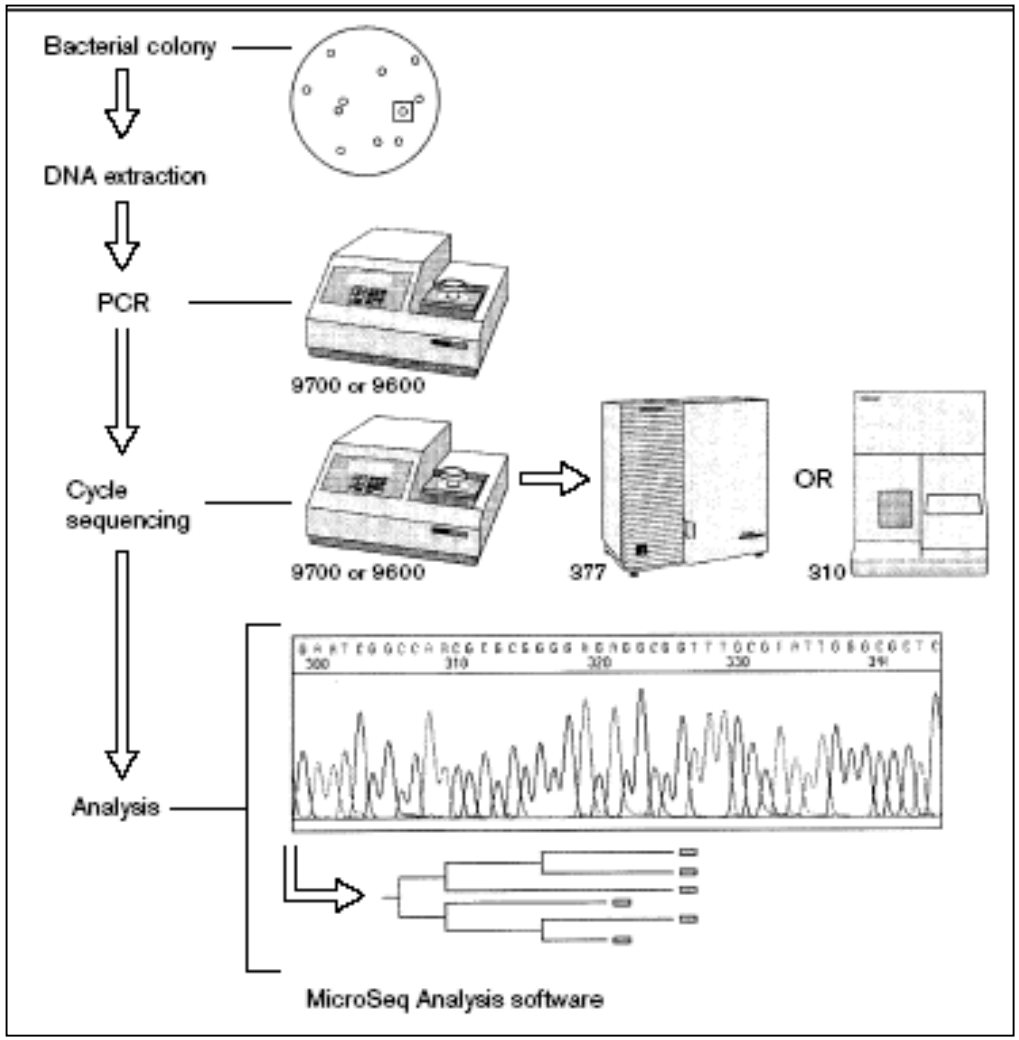
Steps of sequence analysis
Sequence analysis of 16S rRNA gene

- Genomic DNA is extracted directly from dead or alive bacterial colonies grown under any conditions.
- The 16S rRNA gene is amplified using universal primers and thermalcyclers.
- The amplified 16S rRNA gene product is sequenced using dye terminator sequencing.
- The sequence reactions are analyzed using automated DNA sequencers and software.
- Dependant upon the level of confidence, speed and cost desired, either the entire 1540 base pair 16S rRNA gene, or a smaller portion of the gene can be (500 bp) sequenced.
- A printout of the actual sequence data and the identification is included in each report.

Steps of sequence analysis

Sequence analysis of 16S rRNA gene

- Extract DNA from a colony, or from an environmental sample without growing the organism.
- PCR with primers for rRNA sequences.
- Automated DNA sequencer
- Coefficient of Similarity.



Ribosomal DNA sequencing as a tool for identification of bacterial pathogens
Christopher P Kolbert* and David H Persing†

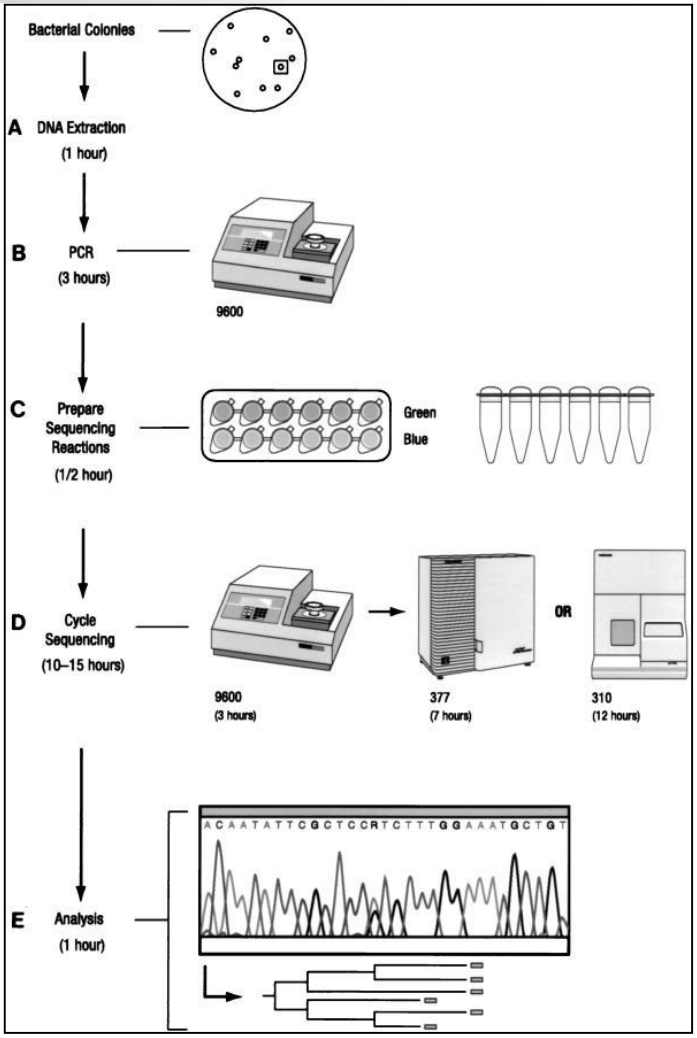
Steps of sequence analysis

Sequence analysis of 16S rRNA gene

- Flowchart of the MicroSeq process from culture to sequence
- The total elapsed time was 15.5 to 18.5 h, comprising:

A. Bacterial DNA extraction,
B. PCR,
C. Sequencing reaction preparation,
D. Cycle sequencing,
E. Analysis.

- The time required for each step is indicated.



Tang *et al.*,1998

Steps of sequence analysis

1. Extract DNA and set up a PCR reaction

1. Scrape a small blob of cells from the agar plate.
2. Stir the loop into a microcentrifuge tube containing 500 ml of sterile water. Close the tube and vortex very well (about 1 minute or until the cell clump is dispersed).
3. Place the tubes into a boiling water bath for 10 min. Use the special tube holder to prevent the tops of the tubes from opening during boiling.
4. Remove from the bath and centrifuge the tube contents for 10 min at 13,000 rpm in the microcentrifuge.
5. Transfer approximately half the supernatant to clean, labeled microcentrifuge tubes using an automatic pipettor.

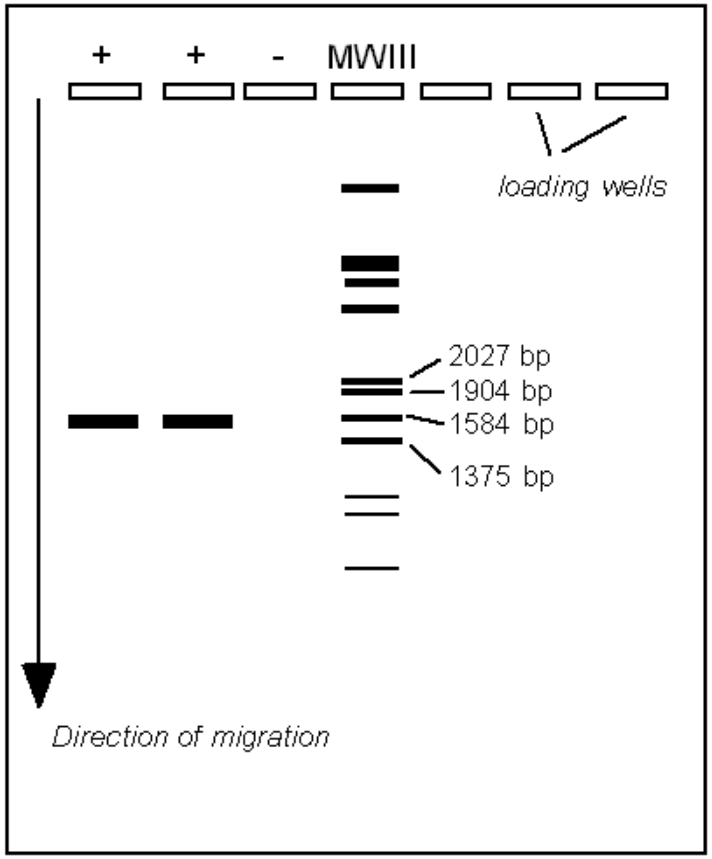
- This will serve as the DNA template for your PCR reaction.
- If you need to stop at this point, you can store the DNA preparation (i.e., supernatant) in the freezer.

Moran,2010

Steps of sequence analysis

2. Examine PCR product on agarose gel

- Agarose gel showing ~1500 bp PCR product from the bacterial 16S rRNA gene.
- The first two lanes (labeled '+') are PCR products from two bacterial isolates.
- The third lane (labeled '-') is the negative control.
- The fourth lane (labeled 'MWIII') is the molecular weight marker.



Moran,2010

Steps of sequence analysis

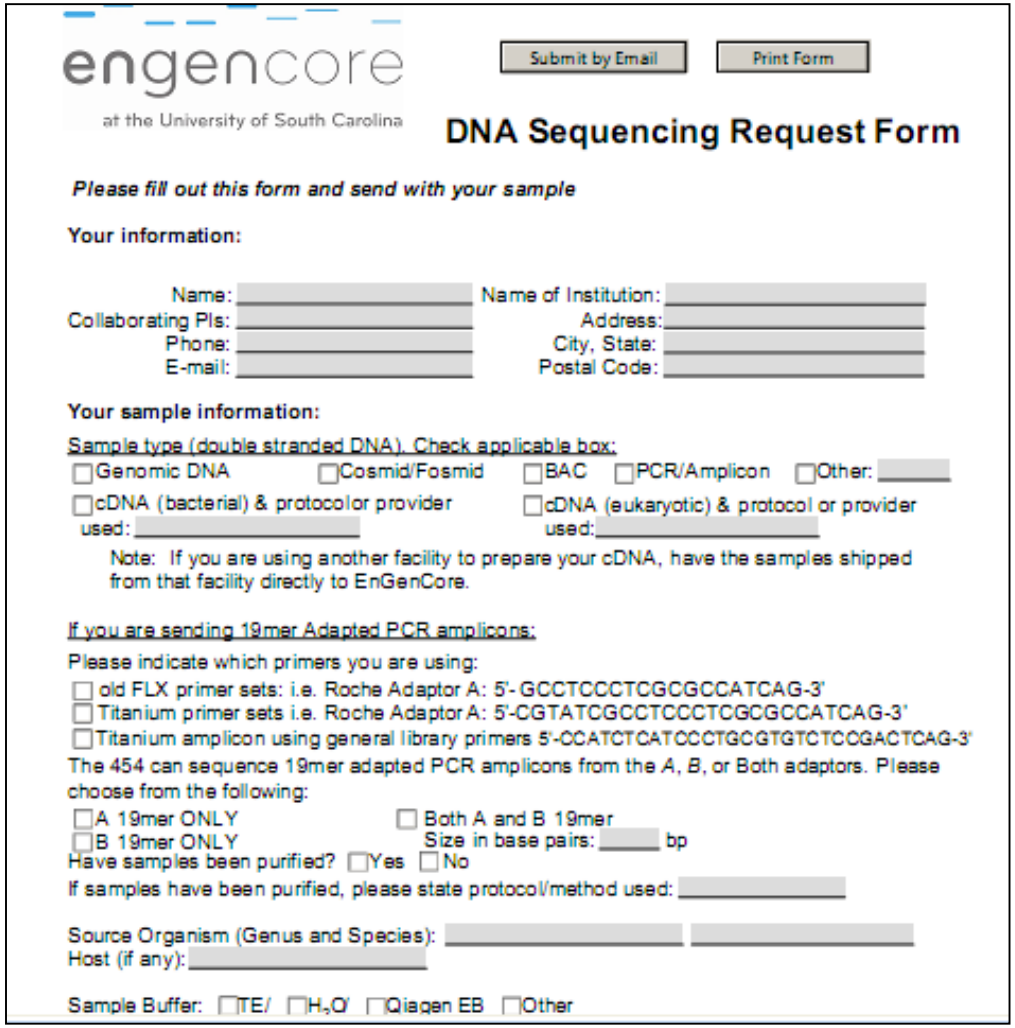
3. Take your PCR product for sequencing

- When you have a PCR product of the appropriate size and concentration, take the product for DNA sequencing.
- Choose your isolate with the best PCR product (i.e., strongest band) to continue working with.
- Fill out the "DNA Sequence Request Form" (provided in the laboratory handout) with the required information and send it with your sample.
- To keep rest of your PCR product cold during transport, put the tube in a plastic cup filled with ice.
- In approximately 1 week, DNA sequence result will received by e-mail.
- Save a copy of the sequence in a text file on a floppy disk or save the e-mail so you can access the file from another computer.

Moran,2010

Steps of sequence analysis

4. Fill out the DNA sequence request form



engencore

at the University of South Carolina

Submit by Email | Print Form

DNA Sequencing Request Form

Please fill out this form and send with your sample

Your information:

Name: _____ Name of Institution: _____
Collaborating PIs: _____ Address: _____
Phone: _____ City, State: _____
E-mail: _____ Postal Code: _____

Your sample information:

Sample type (double stranded DNA). Check applicable box:

☐ Genomic DNA ☐ Cosmid/Fosmid ☐ BAC ☐ PCR/Amplicon ☐ Other: _____

☐ cDNA (bacterial) & protocol or provider used: _____

☐ cDNA (eukaryotic) & protocol or provider used: _____

Note: If you are using another facility to prepare your cDNA, have the samples shipped from that facility directly to EnGenCore.

If you are sending 19mer Adapted PCR amplicons:

Please indicate which primers you are using:

☐ old FLX primer sets: i.e. Roche Adaptor A: 5'- GCCTCCCTCGCGCCATCAG-3'
☐ Titanium primer sets i.e. Roche Adaptor A: 5'-CGTATCGCCTCCCTCGCGCCATCAG-3'
☐ Titanium amplicon using general library primers 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3'

The 454 can sequence 19mer adapted PCR amplicons from the *A*, *B*, or Both adaptors. Please choose from the following:

☐ A 19mer ONLY ☐ Both A and B 19mer
☐ B 19mer ONLY Size in base pairs: _____ bp

Have samples been purified? ☐ Yes ☐ No

If samples have been purified, please state protocol/method used: _____

Source Organism (Genus and Species): _____ _____
Host (if any): _____

Sample Buffer: ☐ TE/ ☐ H_2O/ ☐ Qiagen EB ☐ Other

Steps of sequence analysis

4. Fill out the DNA sequence request form

Shipping Instructions

DNA Quality: DNA should be purified by column or gel purification protocols. Samples providing A260/A280 of ~1.8 normally provide optimal results.

A number of commercially available kits will yield DNA of sufficient quality and purity for sequencing. One µL of each DNA sample should be subjected to electrophoresis on a 2% agarose gel to evaluate for concentration and quality. A hardcopy image of the EtBr-stained gel should be included with the sample(s). Please include a 1 kb ladder as a reference and clearly identify the contents of each lane.

DNA Quantity: For genomic sequencing projects, a minimum of 500 ng of DNA per Pico Green (or other fluorometric methods) provided at a concentration of ≥30 ng/µL is normally adequate.
For a Titanium chemistry run, a minimum of 500 ng of DNA is required. For sequencing projects where the PI uses the Roche Rapid Library Preparation protocol to prepare the library, please send 50 microliters as per the Roche protocol.

Sample Submissions:
Samples can be provided in sterile water or TE buffer 50ng/µL (no more than 100 µL). Individual samples should be provided in a 1.5 mL microfuge tube properly labeled.

2. Preparation of microfuge sample tubes for shipping. Place up to six tubes inside of a 50 mL centrifuge tube. Secure the microfuge sample tubes by packing the tube with a paper towel or tissue paper.

Shipping:
Samples should be frozen and shipped on dry ice. We recommend placing the 50 mL tube(s) inside of a plastic bag prior to placing on dry ice. Samples should be shipped via an express (next day) shipping carrier to:

Environmental Genomics Core Facility
921 Assembly Street
Public Health Research Building, Room 413
Columbia, SC 29208
803-777-4338/3999
Attn: Toni Hammond

Please notify us by email of your tracking number! It has been our experience that **FedEx** is more reliable than **UPS** for shipping samples to us.

Steps of sequence analysis

5. Quality of PCR product for direct sequencing

- Sequencing includes all kinds of cloned DNAs (plasmids, cosmid, phages, BACs) as well as PCR-products.
- Direct PCR product sequencing means using the PCR product as a template for the sequencing reaction without first cloning it into a vector.
- There is no universal rule that will determine what is the "best" approach.
- However, below are some points to consider if you intend to try it:

1. One really needs a very clean PCR product that contains a single amplified product.
2. The product should be at least 200-300 bp in size.
3. The PCR reaction should give a high yield of product.

- Should your PCR product not meet the above requirements (either before or after purification), then direct sequencing is not likely to be very successful.

Moran,2010;..

Steps of sequence analysis

6. Lists of sequencing service providers

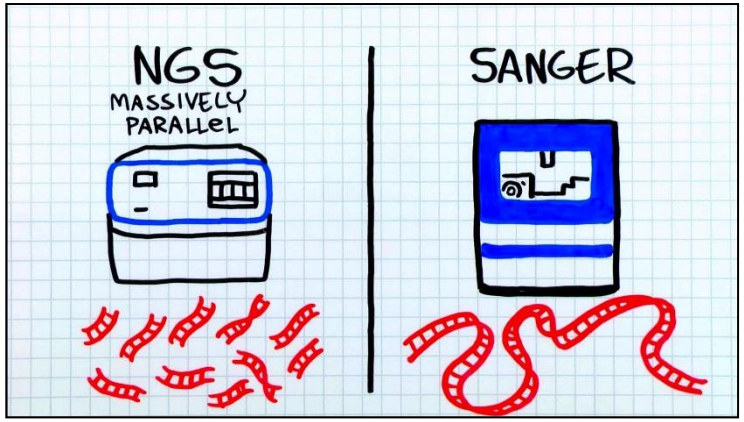
- There are numerous popular DNA sequencing service providers all around the world.
- Examples:
- Agencourt Bisosciences (USA)
- Biofidal (France)
- Cybergene (Sweden)
- Macrogen (Korea)
- Hokkaido System Science (Japan)
- First BASE Laboratories (Malaysia)
- AgGenomics (Australia)
- Inqaba Biotec (South Africa)
- DNAmind (United Kingdom)
- Genoma Sequencing (Italy)
- GATC Biotech (Germany)

Two most popular methods of DNA sequencing

Sanger sequencing method vs. Next-generation sequencing (NGS)

- The biggest difference between the two is sequencing volume.

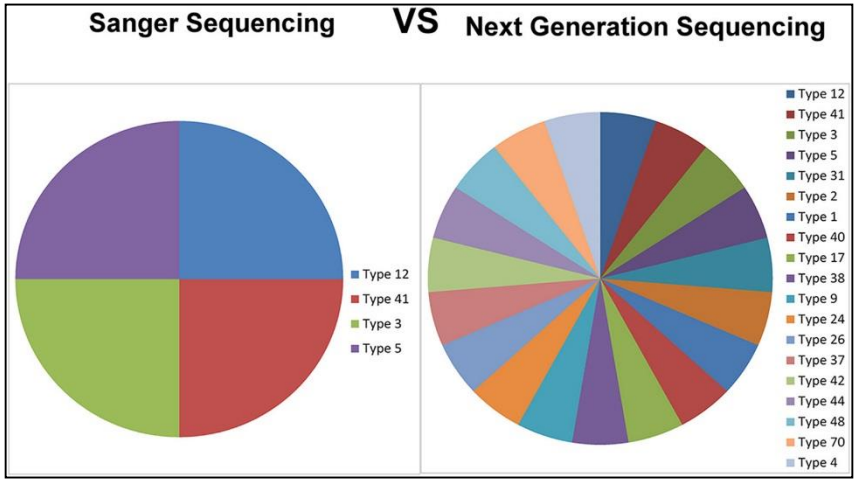
1. NGS, allows millions of fragments to be sequenced in a single run.
2. Sanger sequencing can only sequence one fragment at a time.



Two most popular methods of DNA sequencing

Sanger sequencing method vs. Next-generation sequencing (NGS)

- Next-generation sequencing (NGS) is the newer gene sequencing technology that rapidly performs sequencing from abundant small fragments of DNA.
- NGS results revealed a marked HAdV diversity with 16 additional types detected beyond the four types found by Sanger sequencing.

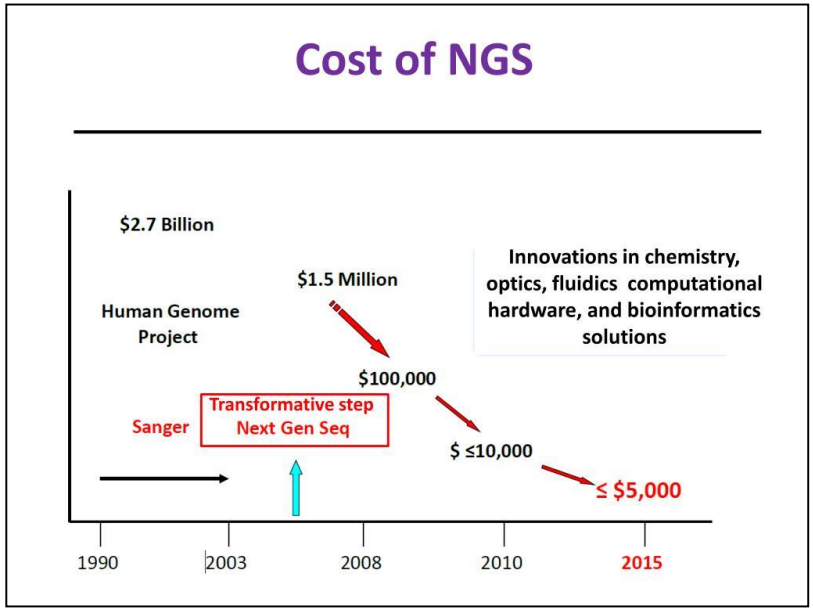


Dey,2018

Two most popular methods of DNA sequencing

Cost for metagenomic library, and cost for metagenomic analysis

- If you meant library, a gross estimate would be about $50-$100 per sample if you do the DNA isolations yourself.
- If you can select or enrich for bacteria that might help, but most likely you'll need to do some sequencing runs to get empirical data on the number/amount of bacteria, and then determine how much coverage you'll need when using only the bacterial data from your mixed samples (i.e. throwing out the Euk data).
- The cost of sequencing those libraries varies a lot, but another gross estimate is $15 to $35 per billion-base-pairs.
- Once you have an estimate of bacteria present, and coverage needed, then figure the bp needed per experiment, and pick a sequencer that can fit your samples together if possible, to minimize technical variability.



Gibson; Hoyt,2018

Steps of sequence analysis

Sanger sequencing method

Nucleotides (dNTP) are modified (dideoxynucleotides = ddNTP)

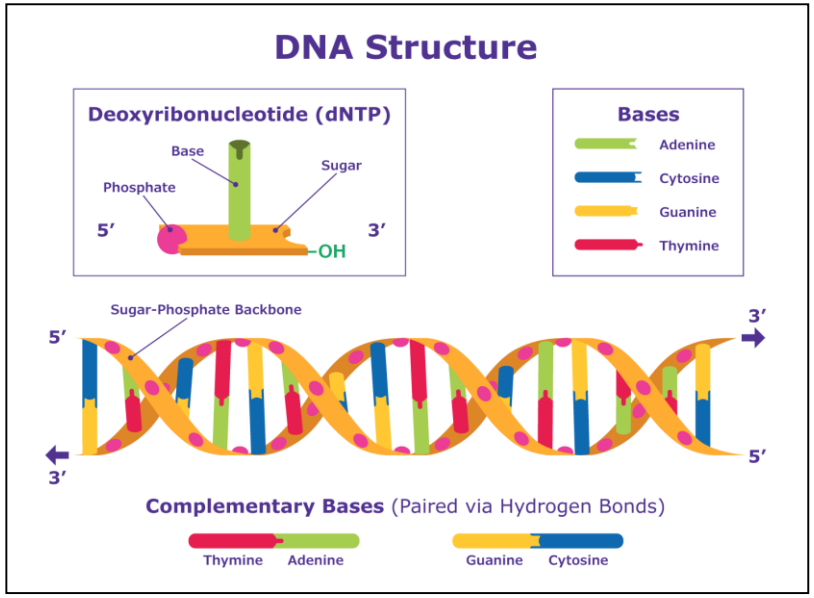
NO polymerisation after a dideoxynucleotide!

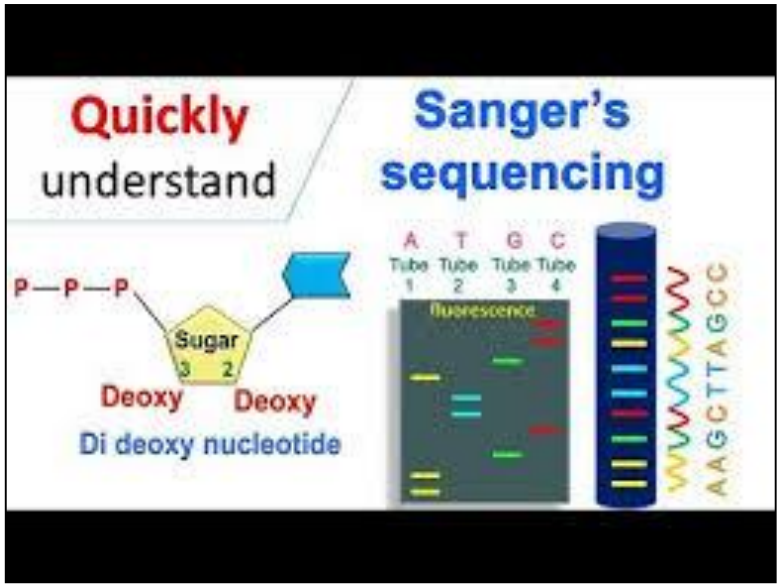
Fragments of DNA differing only by one nucleotide are generated

Nucleotides are either radioactive **or** fluorescent

Steps of sequence analysis

Sanger sequencing method





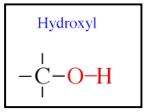
Dideoxynucleotides (ddNTPs) are chain-elongating inhibitors of DNA polymerase, used in the Sanger method for DNA sequencing.

Steps of sequence analysis
Sanger sequencing method

- Purified PCR products were used as templates and sent to the Bioneer Company, Korea for nucleotide sequence determination.
- The PCR products were directly sequenced using the forward and reverse primers used for the amplification step.
- PCR products were sequenced using the Sanger sequencing method by ABI3730XL sequence analyzer.
- Sanger sequencing is a method of DNA sequencing based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication (outside a cell).

Steps of sequence analysis
Sanger sequencing method

1. A primer is annealed to a sequence,
2. Reagents are added to the primer and template, including: DNA polymerase, dNTPs, and a small amount of all four dideoxynucleotides (ddNTPs) labeled with fluorophores.
3. During primer elongation, the random insertion of a ddNTP instead of a dNTP terminates synthesis of the chain because DNA polymerase cannot react with the missing hydroxyl. This produces all possible lengths of chains.



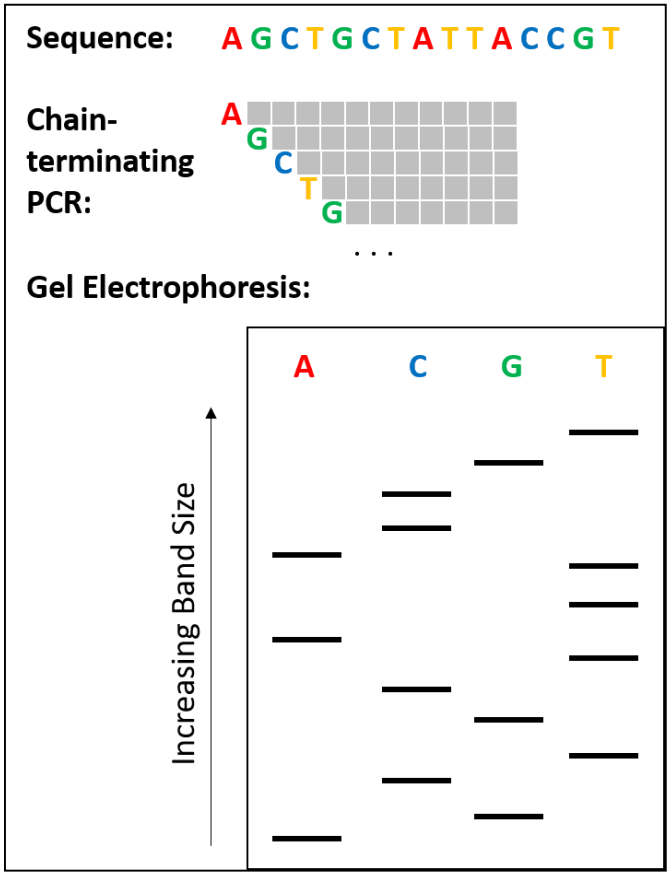
4. The products are separated on a single lane capillary gel, where the resulting bands are read by a imaging system.
5. The DNA sequence is read through the fluorescent emission of the di-deoxynucleotide as it flows through the gel.

Steps of sequence analysis

Sanger sequencing method

Early termination of sequencing reactions during PCR

- Mixing dNTPS with ddNTPs causes random early termination of sequencing reactions during PCR.
- Four reactions are run, each with the chain-terminating version of only one base (A, T, G or C).
- When visualized with gel electrophoresis, one reaction per lane, the fragments are sorted by length, allowing the DNA sequence to be read off base by base.

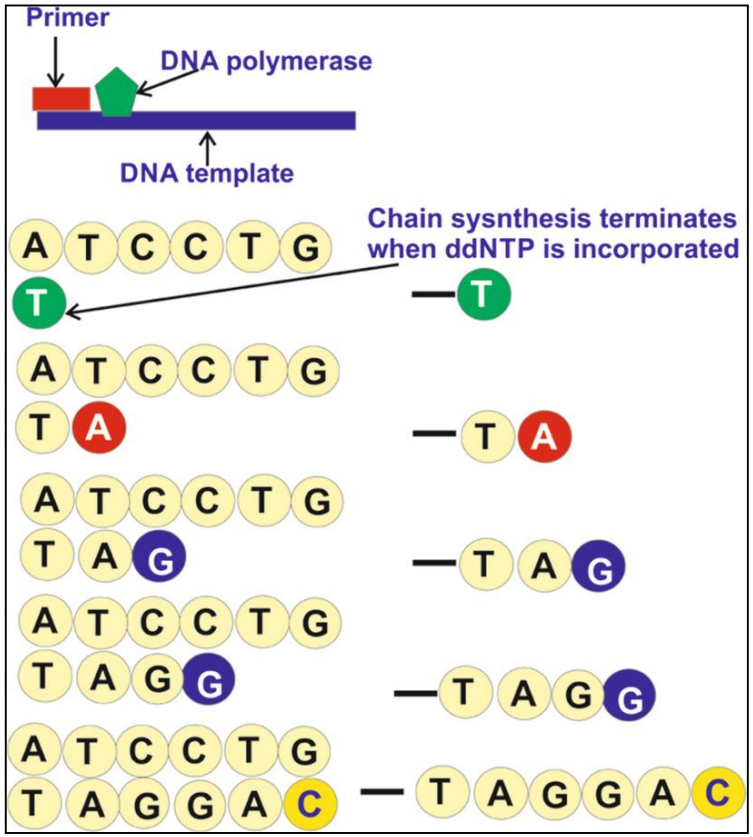


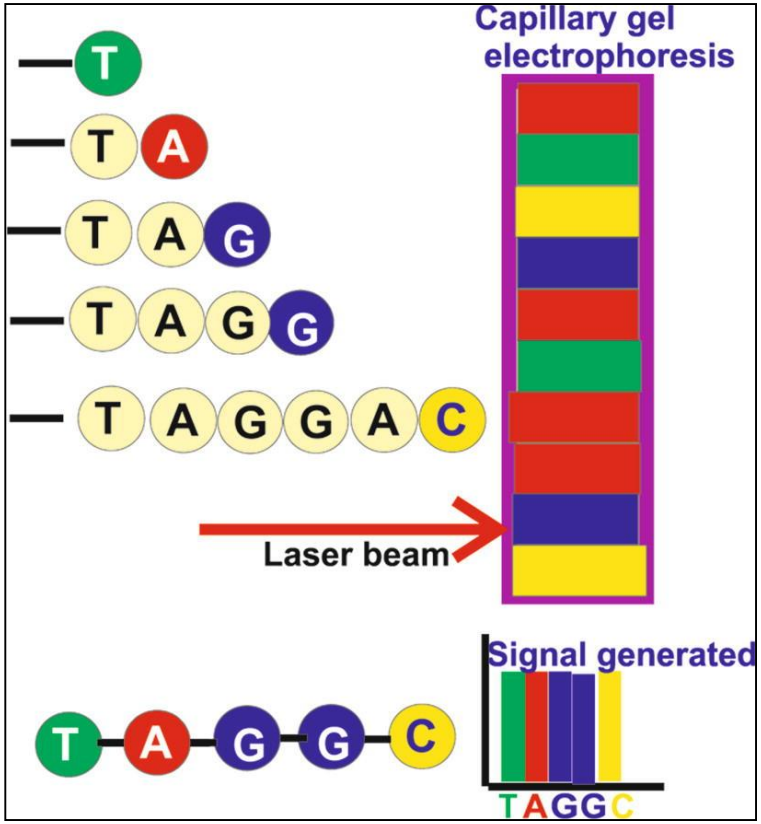
PACBIO

Steps of sequence analysis

Sanger sequencing method

Early termination of sequencing reactions during PCR



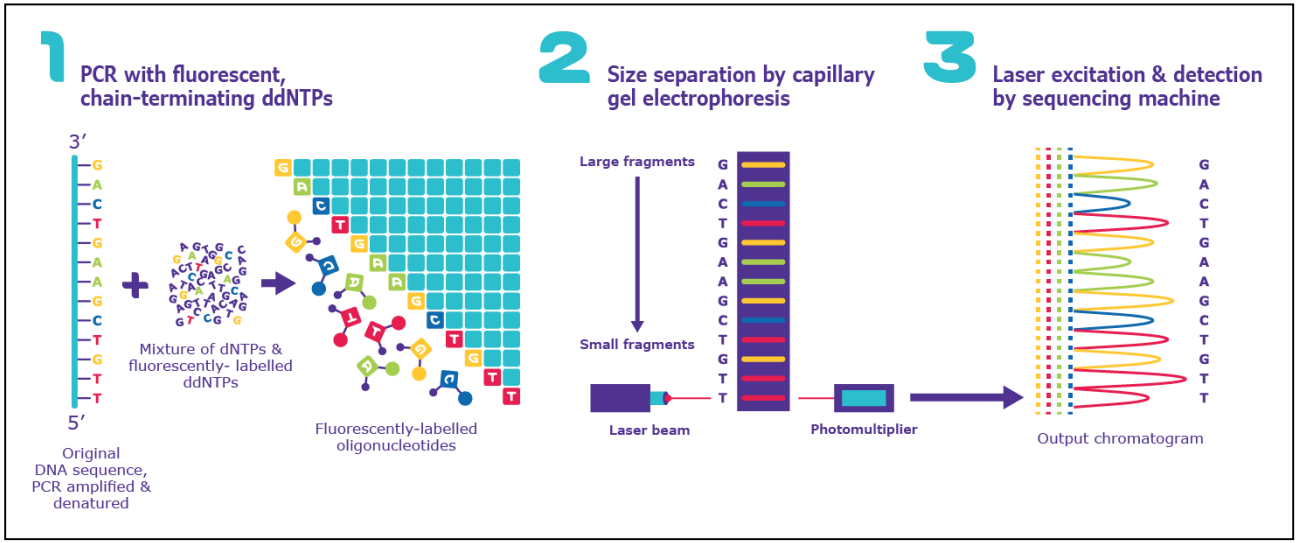


Dey,2018

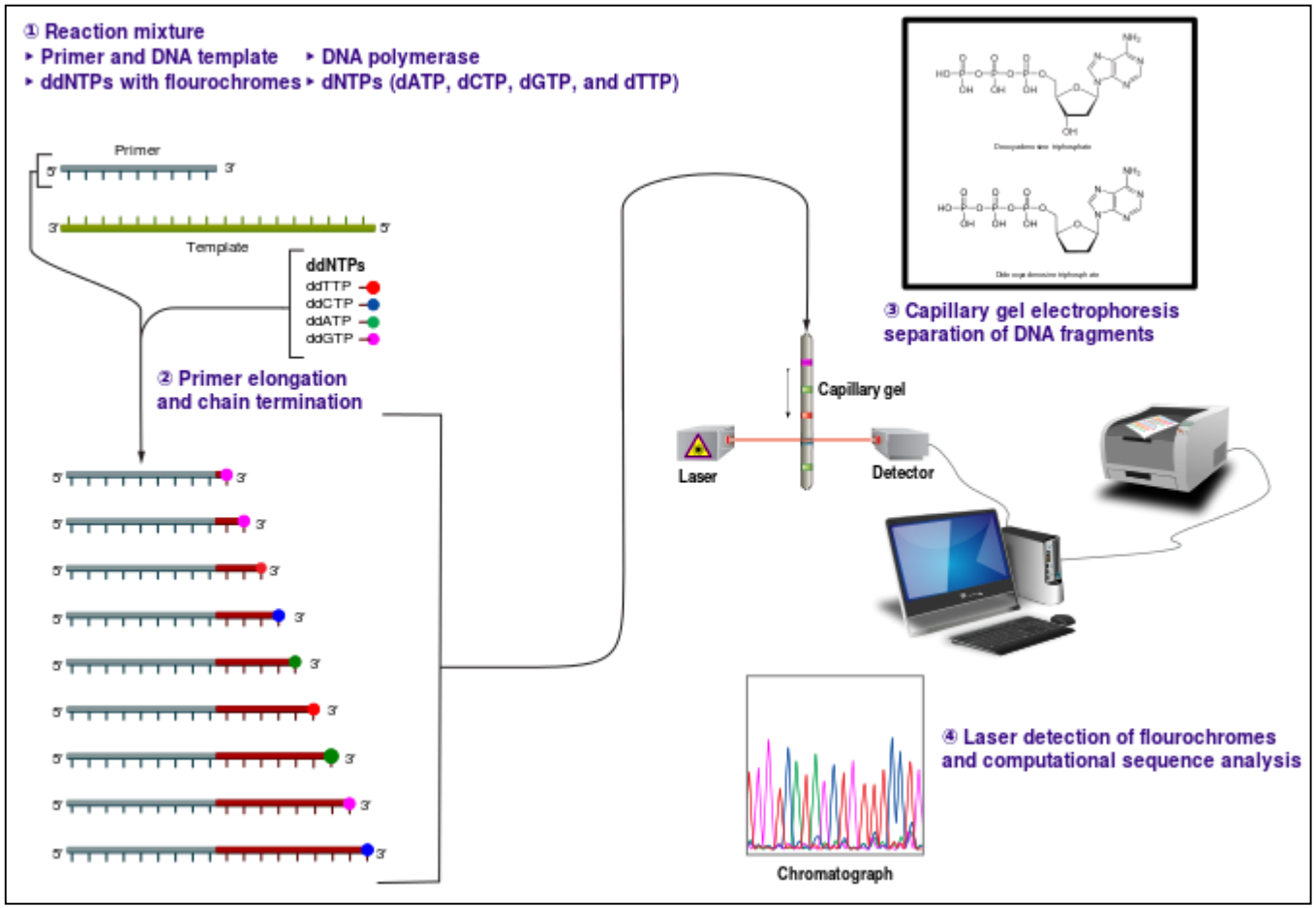
Steps of sequence analysis
Sanger sequencing method

- Sanger sequencing can be performed manually or, more commonly, in an automated fashion via sequencing machine.
- Each method follows three basic steps, as described below:

1. DNA Sequence For Chain Termination PCR
2. Size Separation by Gel Electrophoresis
3. Gel Analysis & Determination of DNA Sequence.



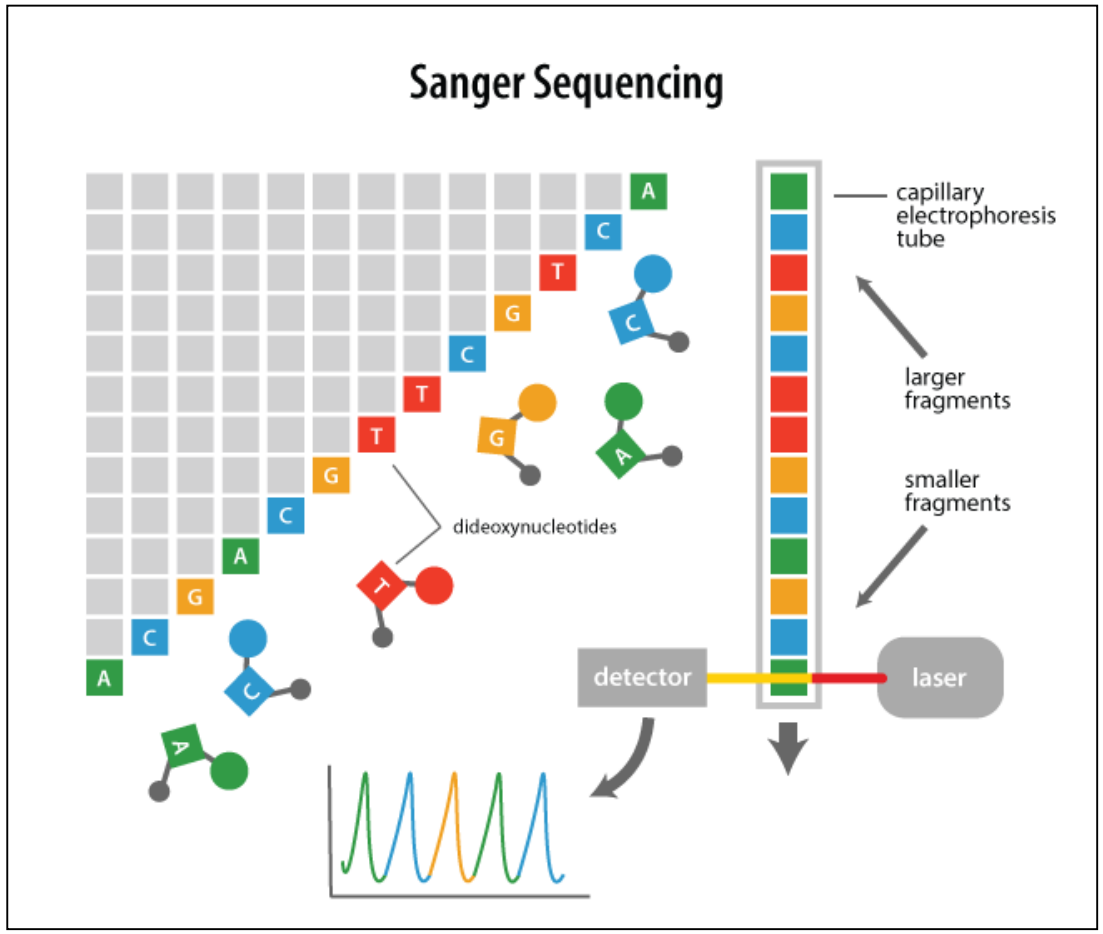
Steps of sequence analysis
Sanger sequencing method



Wikipedia,2016

Steps of sequence analysis

Sanger sequencing method



www.abmgood.com

Steps of sequence analysis

7. DNA-sequencers

Different sequencing methods

- Genetic information is stored in the order or sequence of nucleotides in DNA.
- The fluorescent dye-terminator cycle sequencing method (Chain termination sequencing or Dye Terminator Sequencing) is the standard method for the determination of nucleotide sequence.
- Alternatives to Dye Terminator Sequencing are:

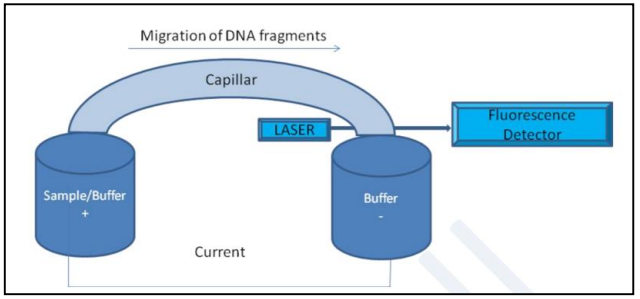
1. Pyrosequencing;
2. Microarray sequencing: aka sequencing by hybridization;
3. Nanopore sequencing.

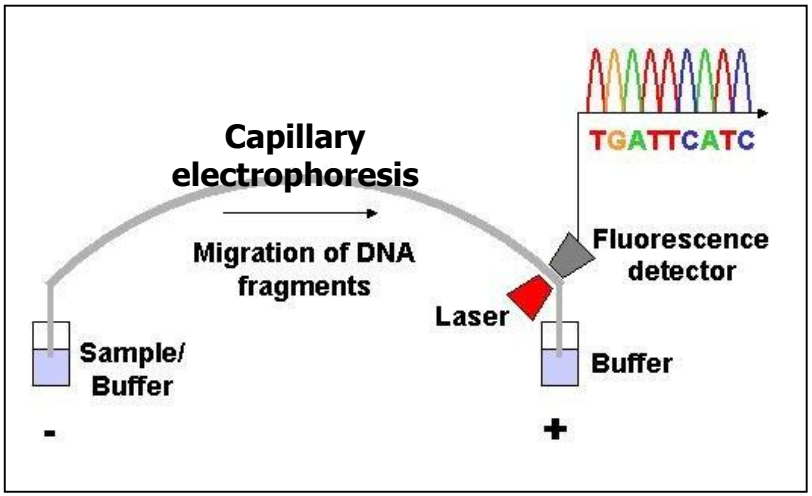
Steps of sequence analysis

DNA-sequencers

Capillary based DNA sequencing system

- DNA sequencers carry out capillary electrophoresis for size separation, detection and recording of dye fluorescence, and data output as fluorescent peak trace chromatograms.





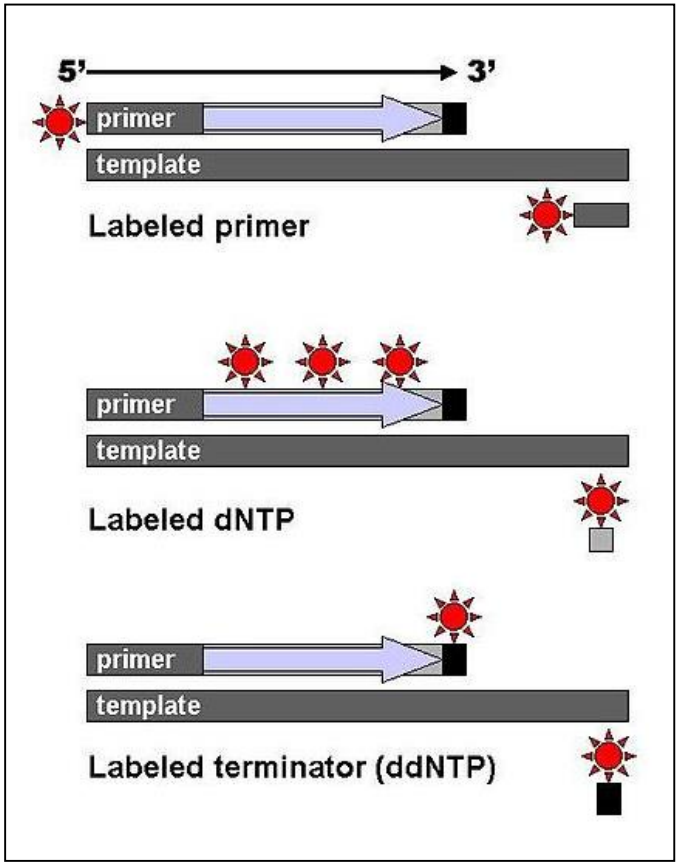
A chromatogram (electrogram, sometimes also called electropherogram) is the visual representation of a DNA sample produced by a sequencing machine.

Steps of sequence analysis

DNA-sequencers

Capillary based DNA sequencing system

- DNA fragments are labelled with fluorescent tag on
- The primer (1),
- In the new DNA strand with a labeled dNTP, or
- With a labeled ddNTP (whenever a ddNTP was added, the elongation stopped).



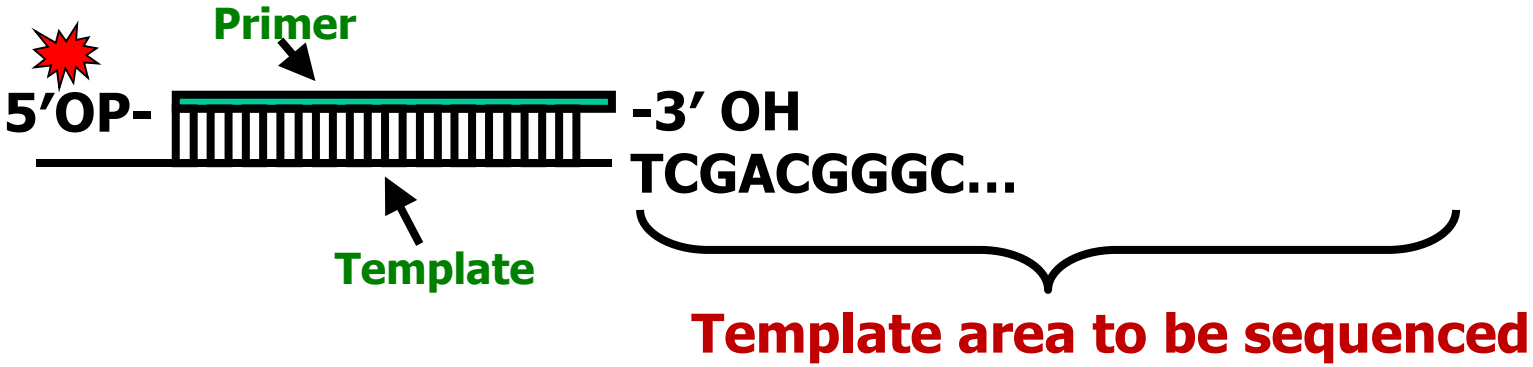
Deoxynucleotide triphosphates (dNTPs) are the essential building blocks of nucleic acid molecules, and as such are necessary components of PCR mixes as no new (amplified) DNA could be generated without them. Four Deoxynucleotides (dNTPs) are: dATP, dCTP, dGTP, and dTTP.

Steps of sequence analysis

DNA-sequencers

Capillary based DNA sequencing system

- A sequencing reaction mix includes labeled primer and template (primer and terminator sequencing are labelled with fluorescent dyes).



- Dideoxynucleotides (ddNTPs) are added separately to each of the four tubes. Each of four dideoxynucleotide chain terminators emit light at different wavelengths.

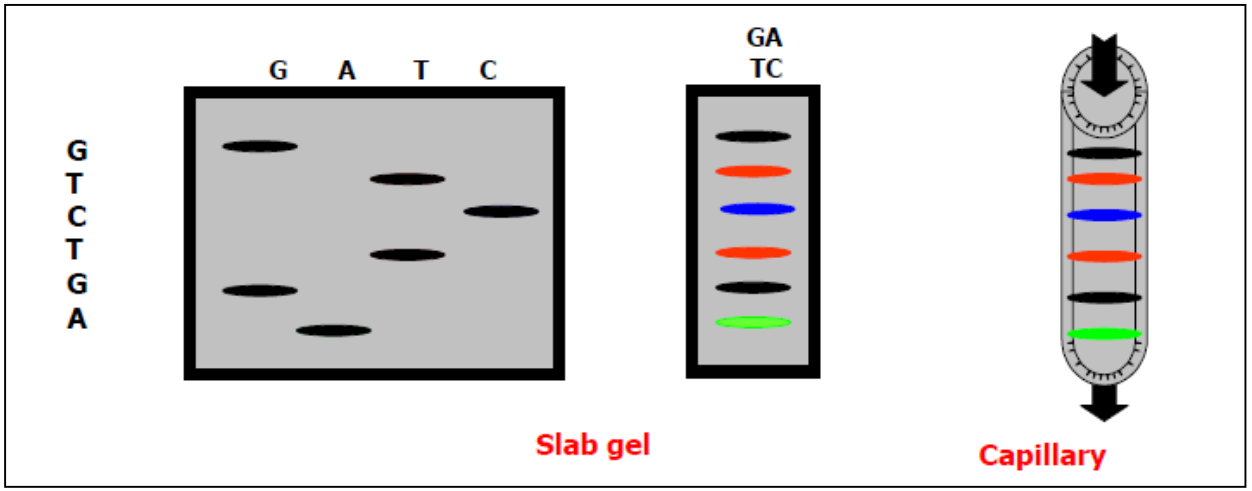
Rushids,2011;...

Steps of sequence analysis

DNA-sequencers

Capillary based DNA sequencing system

- Dye primer or dye terminator sequencing on capillary instruments.
- The DNA ladder is resolved in one gel lane or in a capillary.



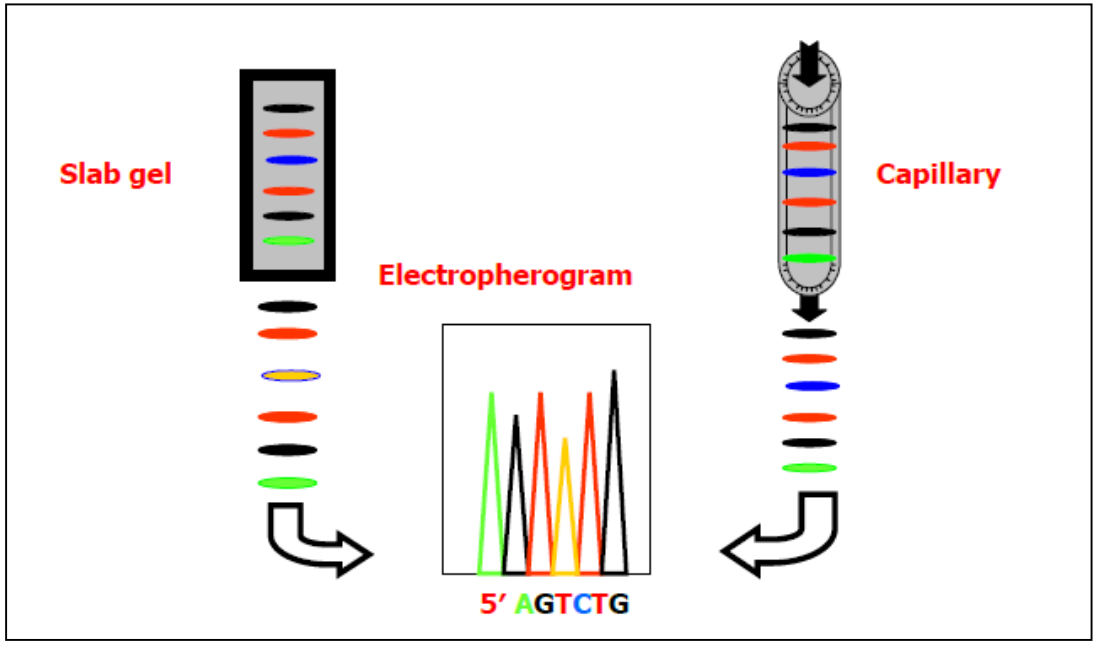
Rushids,2011

Steps of sequence analysis

DNA-sequencers

Capillary based DNA sequencing system

- The DNA ladder is read on an electropherogram.
- Base sequence: adenine (A, green), thymine (T, red), cytosine (C, orange), and guanine (G, blue).



Rushids,2011

Steps of sequence analysis

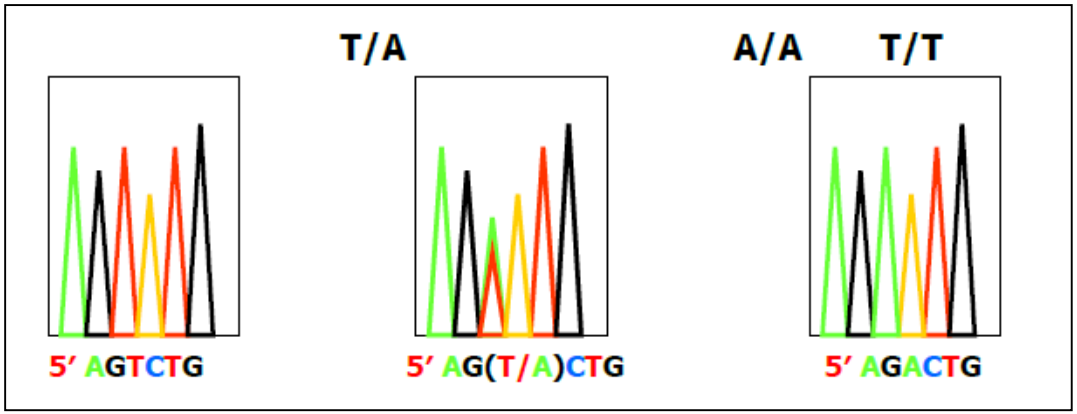
7. DNA-sequencers

Capillary based DNA sequencing system

- Sequence analysis software provides analyzed sequence:

1. In text, and
2. Electropherogram form.

- Peak patterns reflect mutations or sequence changes.



Rushids,2011

Steps of sequence analysis

Sequence Report

Nucleotide sequences and electropherogram of 16SrRNA

Sequence analysis software provides analyzed sequence: In text (.txt), and Electropherogram form (in .ab1 or .scf format). The raw output of a sequencing machines is known as a trace. If the software cannot determine which nucleotide is in a particular position it will assign the letter N to the unknown base.

Steps of sequence analysis

DNA-sequencers

Fluorescent dye-terminator cycle sequencing method

- When you load the entire PCR reaction into one lane of a polyacrylamine gel in a special laser sequencer (for example, the ABI 377).
- The sequencer can read the different dyes as they pass the laser at the bottom of the gel.
- Because the different fragments migrate toward the positively-charged pole at different rates, the shorter fragments migrate toward the bottom of the gel faster, so you read it bottom->top (5'->3').

Licohen,2011

Steps of sequence analysis

DNA-sequencers

Fluorescent dye-terminator cycle sequencing method

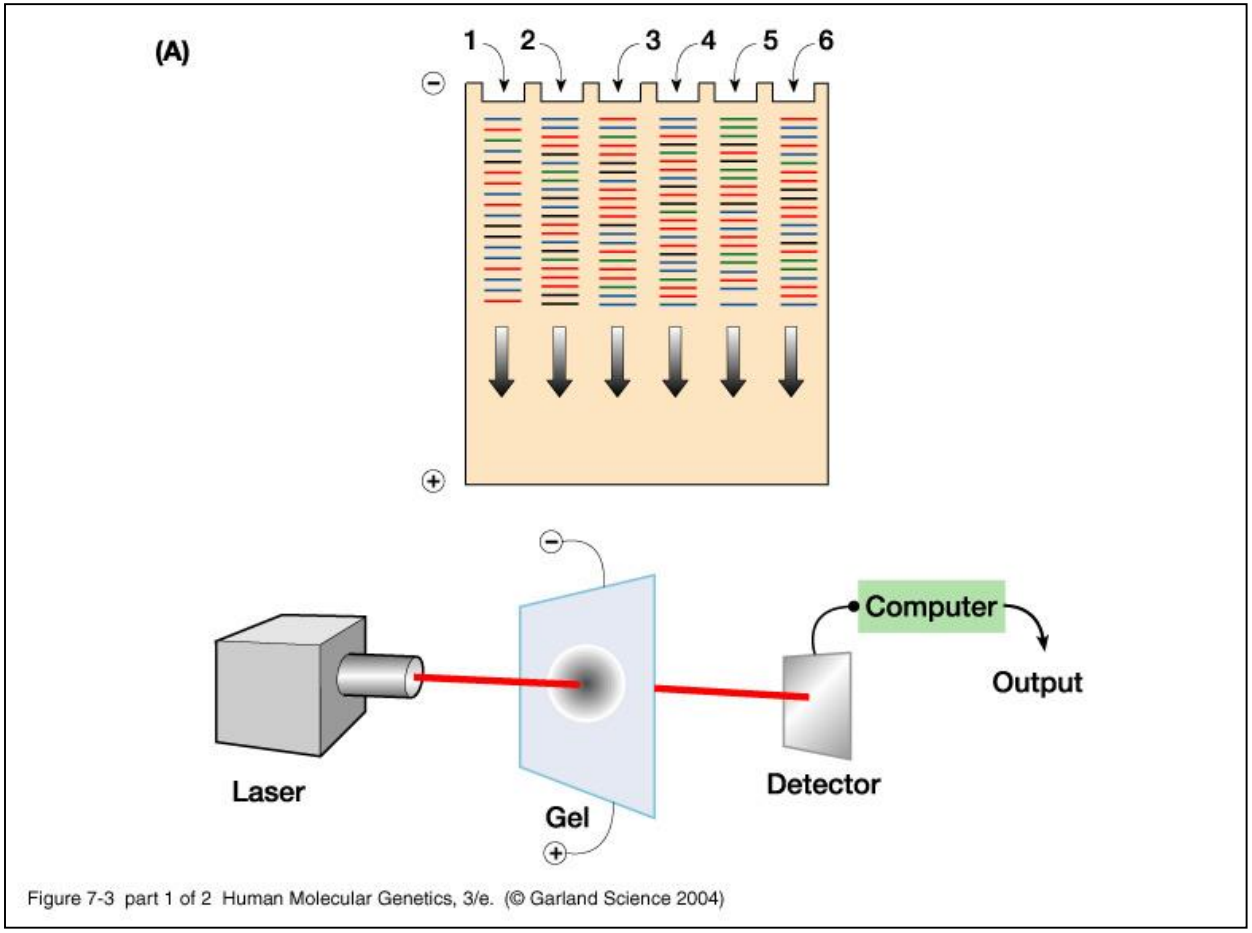


Figure 7-3 part 1 of 2 Human Molecular Genetics, 3/e. (© Garland Science 2004)

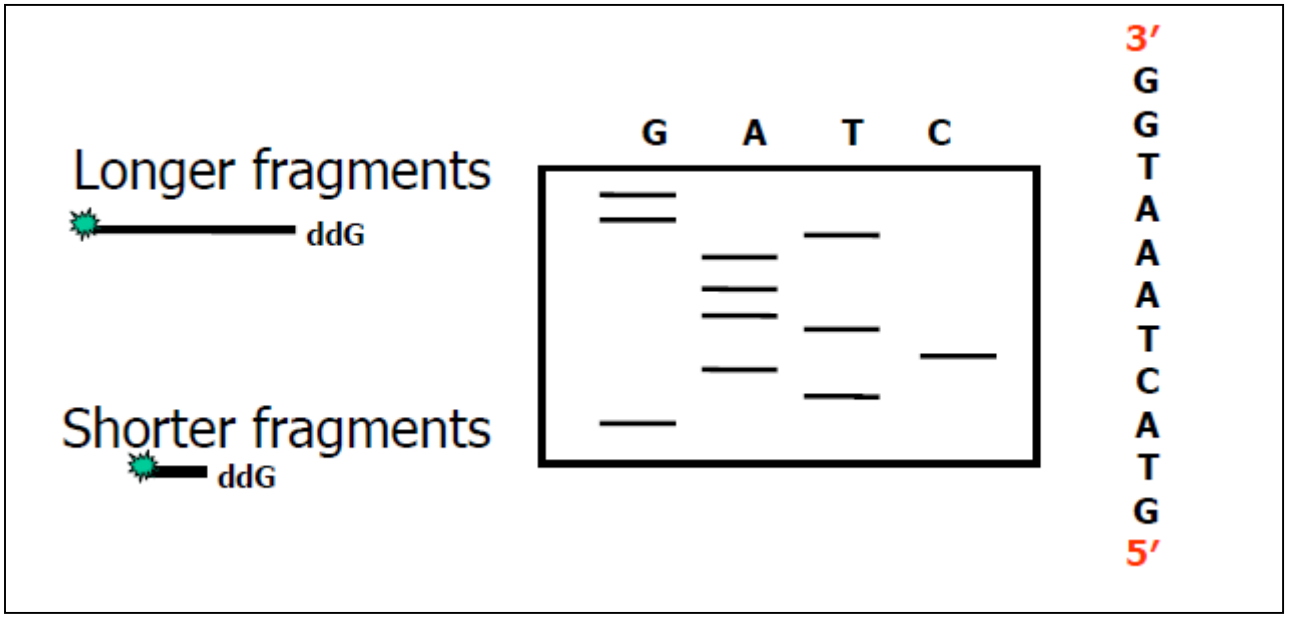
Jamison,2011

Steps of sequence analysis

DNA-sequencers

Fluorescent dye-terminator cycle sequencing method

- Sequencing gels are read from bottom to top (5′ to 3′).



Rushids,2011

Steps of sequence analysis

7. DNA-sequencers

Fluorescent dye-terminator cycle sequencing method

- The computer then compiles these data into the image of a gel.
- The colors are then read as bases to produce a chromatogram of the piece of DNA sequenced.
- The sequence can be taken right from the chromatogram with an extremely high level of accuracy, and the task is complete!

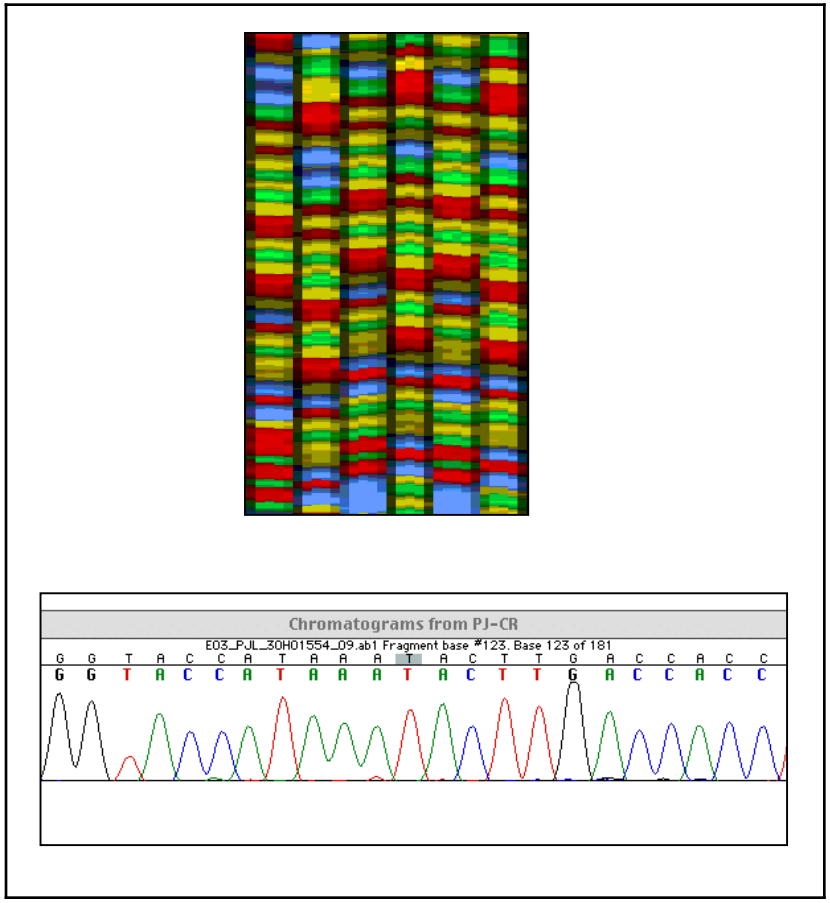
Licohen,2011

Steps of sequence analysis

DNA-sequencers

Fluorescent dye-terminator cycle sequencing method

1. Image of a gel: Trace files (dye signals) are analyzed and bases called to create chromatograms (electrophorgram).
2. Chromatogram of the piece of DNA sequenced: Chromatograms from opposite strands are reconciled with software to create double-stranded sequence data.

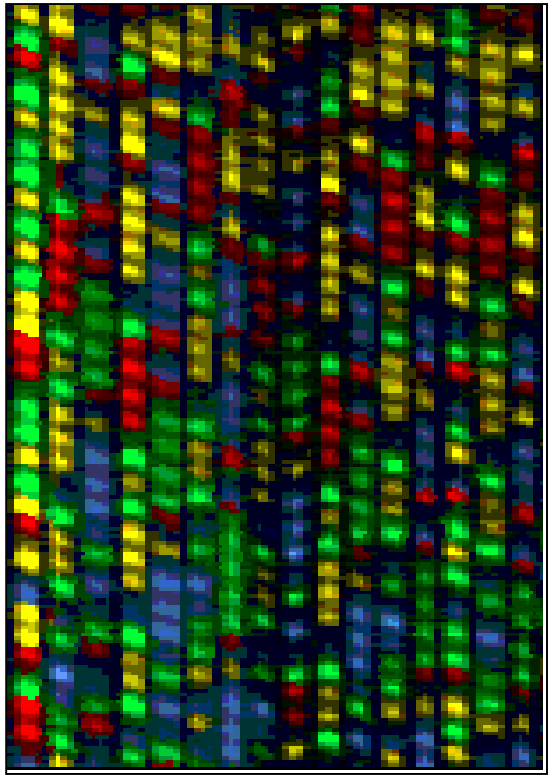


Steps of sequence analysis

DNA-sequencers

Fluorescent dye-terminator cycle sequencing method

1. An actual polyacrylamide sequencing gel with 16 lanes, each loaded with a different sequence sample.
2. Each one of the four colors represents a different base.
3. By translating the colors to bases, you can determine the sequence of the DNA that PCR amplified.



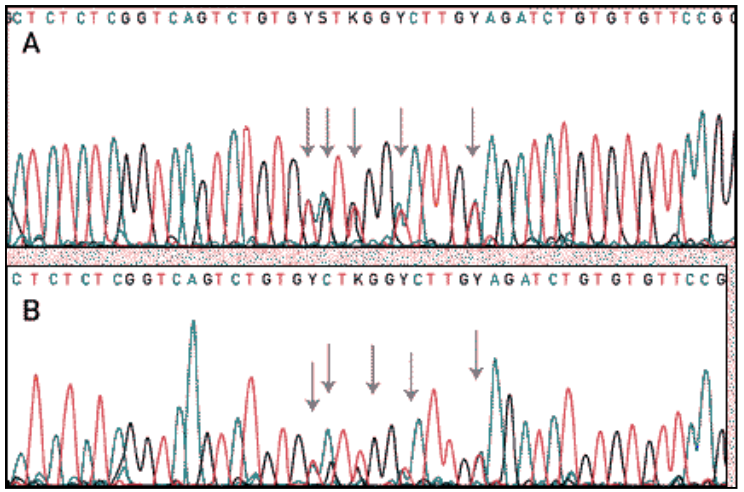
Licohen,2011

Steps of sequence analysis

DNA-sequencers

The final electrogram output of DNA

- Chromatogram of the DNA sequence.
- The computer reads the colored peaks as one of four bases, translating only at strong peaks.
- 'Y' represents a base that was ambiguous-either there were two conflicting signals or there was no signal when the computer was expecting one.
- The gray arrows show where the computer would expect to read a base.



Licohen,2011

Steps of sequence analysis

Sequence Report

BIONEER

Sequencing Report

Register date : MAR 11th 2011
Register No. : Q20110328-15
Orderer : UPM, plant protection
Name : EISA NAZERIAN

Bioneer messages for customers

Data Information					
No.	sample name	primer	Result	data group	comment
1	1sh	F	SUCCESS		
2		R	SUCCESS		

Data group	
A	reading length is short or noisy. Re-reaction is recommended.
B	Failed at analysing. Re-reaction is recommended.
C	There are some problem at analysis because of template structure. Re-reaction is not recommended.
D	Re-reaction is not recommended.
Re-reaction result is released after one day.	

Steps of sequence analysis

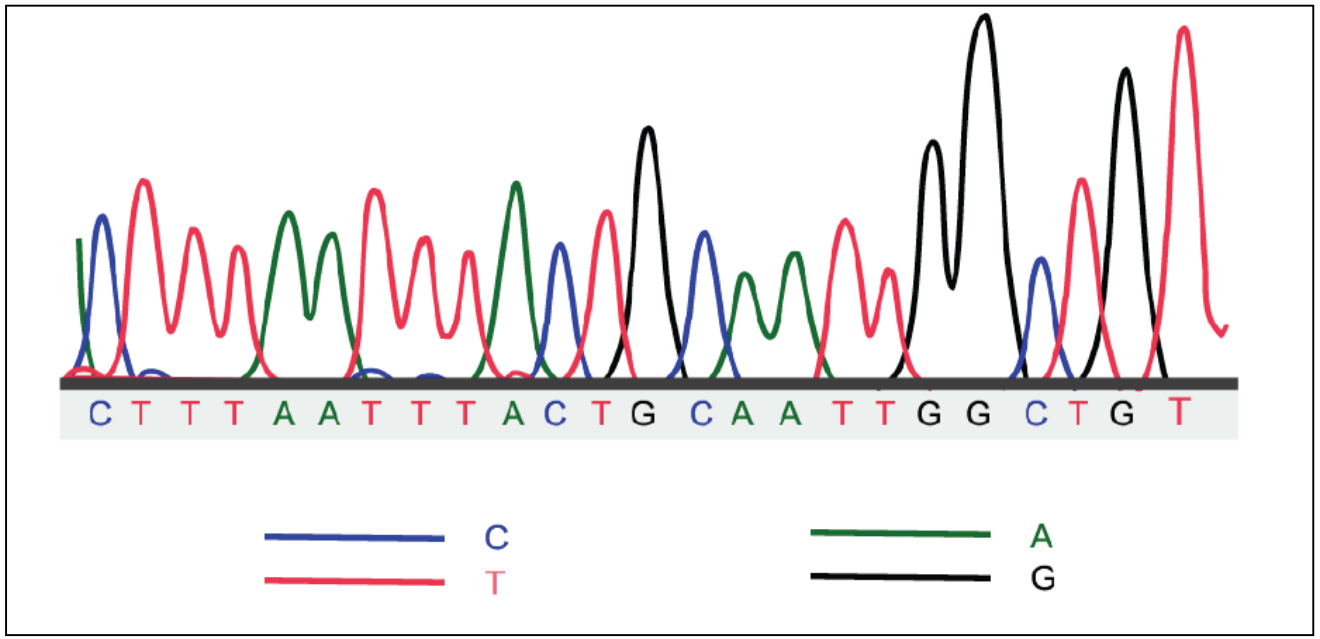
Interpretation of sequencing chromatograms

- The electropherogram (chromatogram or electrochromatogram) is a graphical representation of data received from a sequencing machine and is also known as a trace.
- Each line represents one of the four nucleotides, and the peaks in the lines indicate the strength of the signal given off from the laser beam as it hits the DNA fragment.
- In other words, each peak corresponds to a fluorescently labeled nucleotide base, and the order of the peaks is the nucleotide sequence (because the fragments have previously been ordered by gel electrophoresis).

Steps of sequence analysis

Interpretation of sequencing chromatograms

- Note that in a real electropherogram, guanine would be represented by yellow, rather than black.



Elgin and Flowers

Steps of sequence analysis

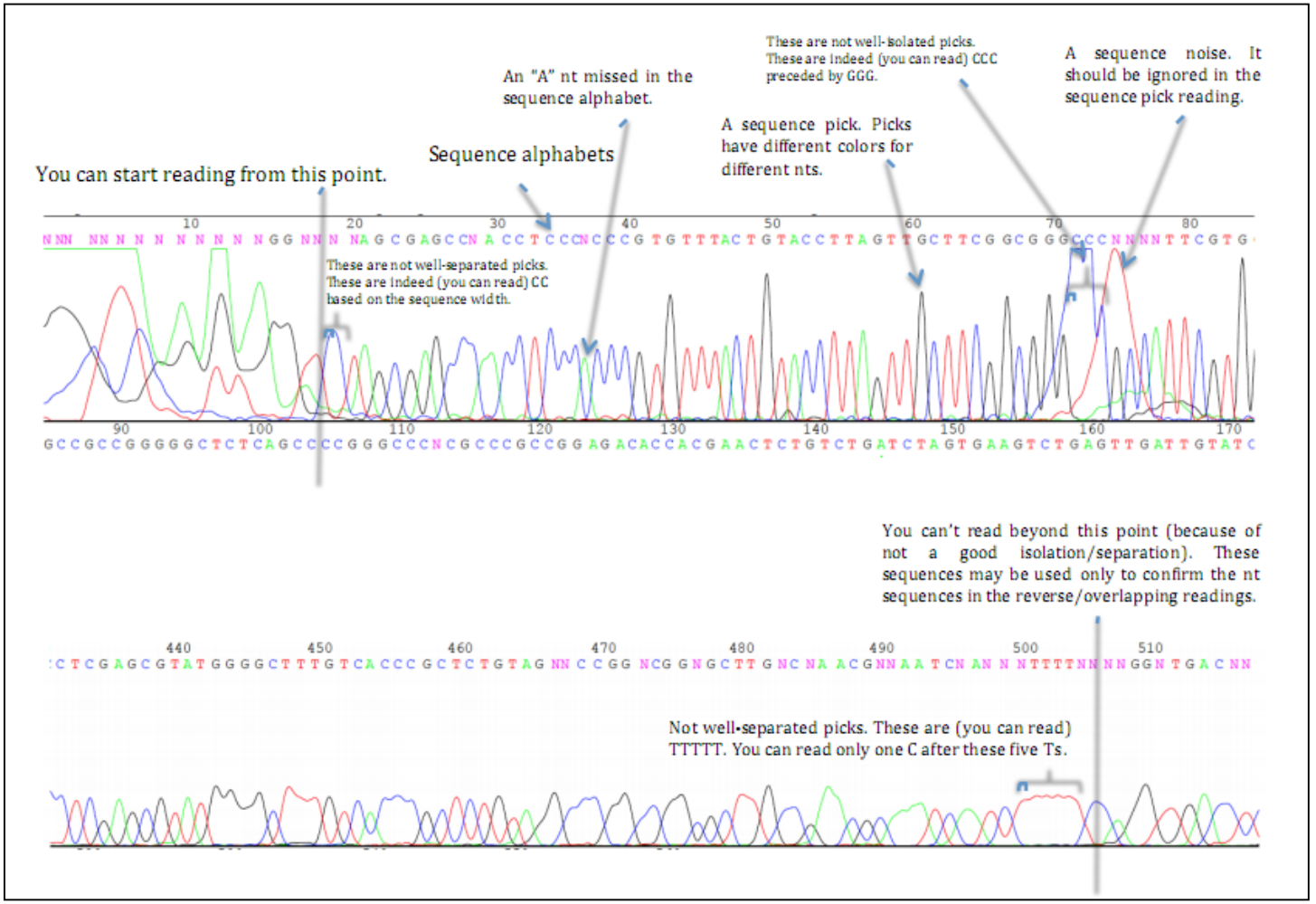
Interpretation of sequencing chromatograms

1. Open the .abl file by using BioEdit or Chromaslite programs (you can find both programs online, then download and install them easily).
2. Generate a text file from the sequence hologram (if this file is not included in the files sent by the company) by using "File Export as Text" or "Export" option of the above programs, respectively. Use this text file as a template for your reading.
3. Generate a PDF file from the sequence hologram (if this file is not included in the files sent by the company) and print out it in COLOR.
4. To start reading, first, remove the complementary nt sequences of the second amplification primer (i.e., the primer which was not used for the sequence reading) from the end of the sequence read (text file). Note: Based on the quality of the sequence reading, only some of the complementary nt sequences of the second primer may be found.
5. Remove all nt sequences appeared after the second primer from the text file as well.
6. Start reading from the upstream but please remember that reading should be started from the nt which its separation (a distinguishable pick) and isolation (an individual pick, not mixed picks) is suitable.

- Note: You should read each sequence three times. The first time is done based on the alphabets appeared above the sequence hologram. The second and the third (final) readings will be performed on the basis of "the colored picks. After finishing each round of reading, the template text file (e.g., 38-S6-R1-ITS1) should be saved as 38-S6-R1-ITS1-1st read, ……2nd read, and ……3rd read.
- Note: At the first and the end of a sequence reading, you can find many non-well isolated/separated picks, most of them may be clarified in the reverse reading (by a second primer).
- **Some other general notes: please see the next page pictures carefully.**

Steps of sequence analysis

Interpretation of sequencing chromatograms

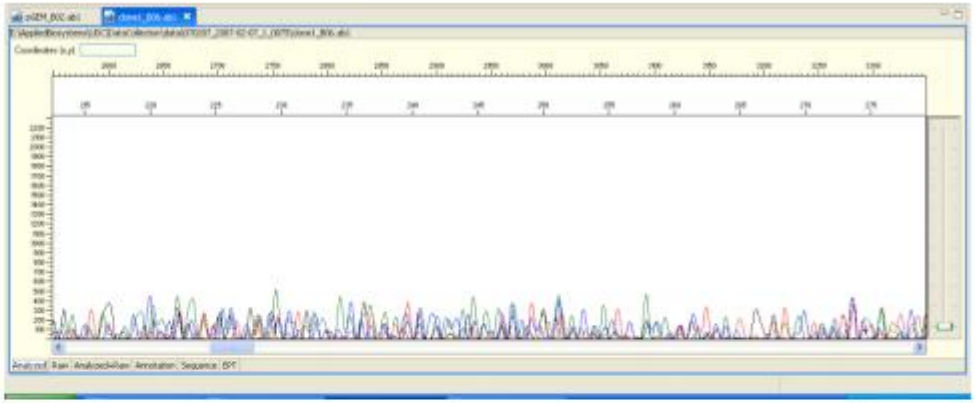


Raw and analyzed DNA sequence data (trace)

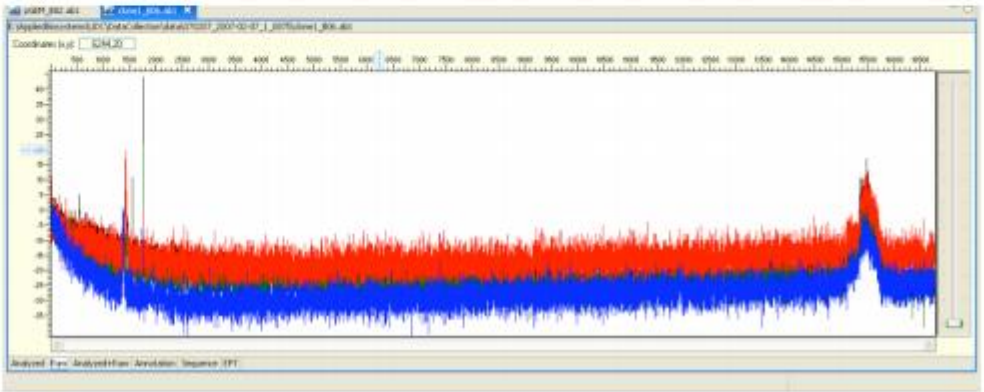
- Unfortunately, electropherograms are not always as high quality.
- The incoming signals may not be as strong and evenly spaced out.
- However, with the help of other computer programs and a finisher, the sequence often can be determined, even if the electropherogram isn't entirely clear.
- A finisher is someone whose job it is to analyze the raw sequence data presented by an electropherogram and create a high quality sequence by editing and calling for additional reactions to gather better data where needed.

Raw and analyzed DNA sequence data (trace)

Only back ground noise present in the analysed trace



Absence of peaks in the raw data profile

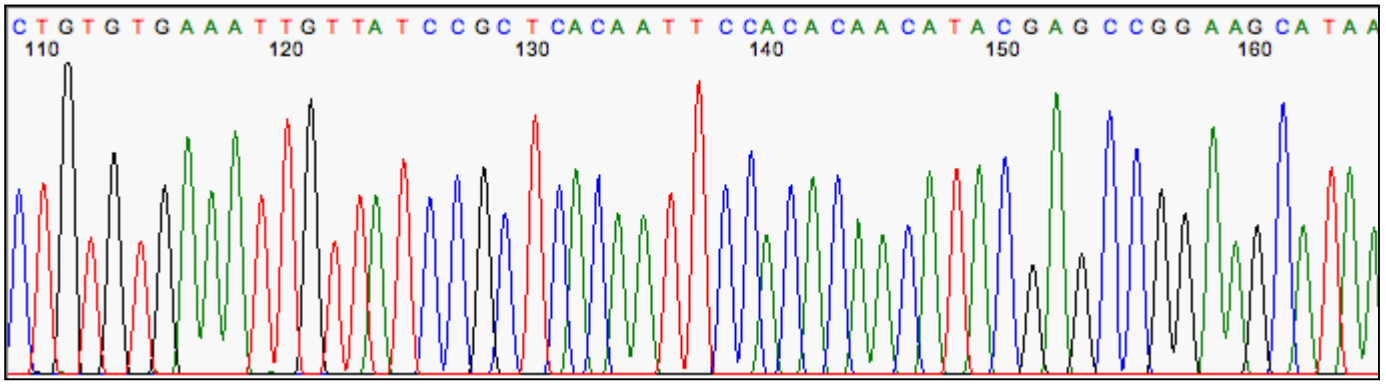


Steps of sequence analysis

Interpretation of sequencing chromatograms

No baseline noise (Good quality sequence)

- Here's an example of excellent sequence.
- Note the evenly-spaced peaks and the lack of baseline 'noise' (see further down for examples of higher baseline noise):



Notice how well defined and sharp the peaks are, with nice even spacing between them. Peak height is also significantly higher than the earliest fragments.

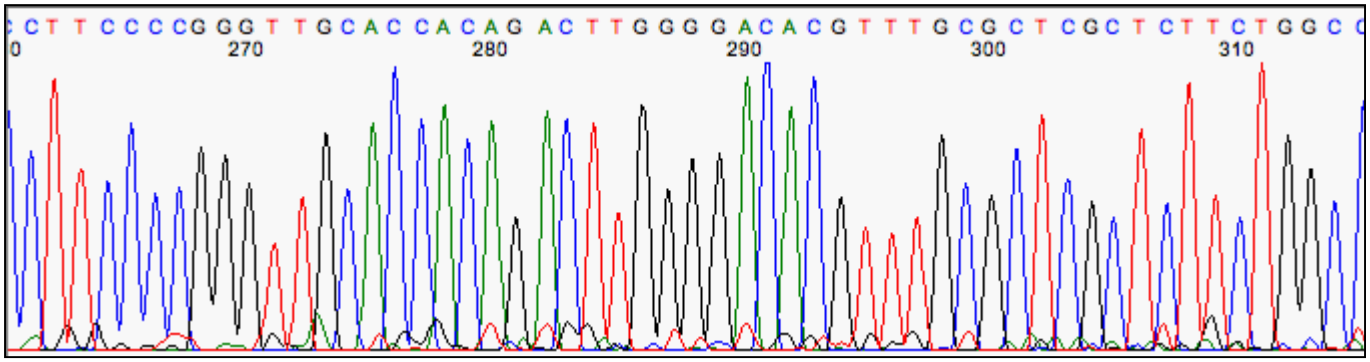
Michigan univ.DNA Sequencing Core

Steps of sequence analysis

Interpretation of sequencing chromatograms

A little baseline noise

- The next example has a little baseline noise, but the 'real' peaks are still easy to call, so there's no problem with this sample:



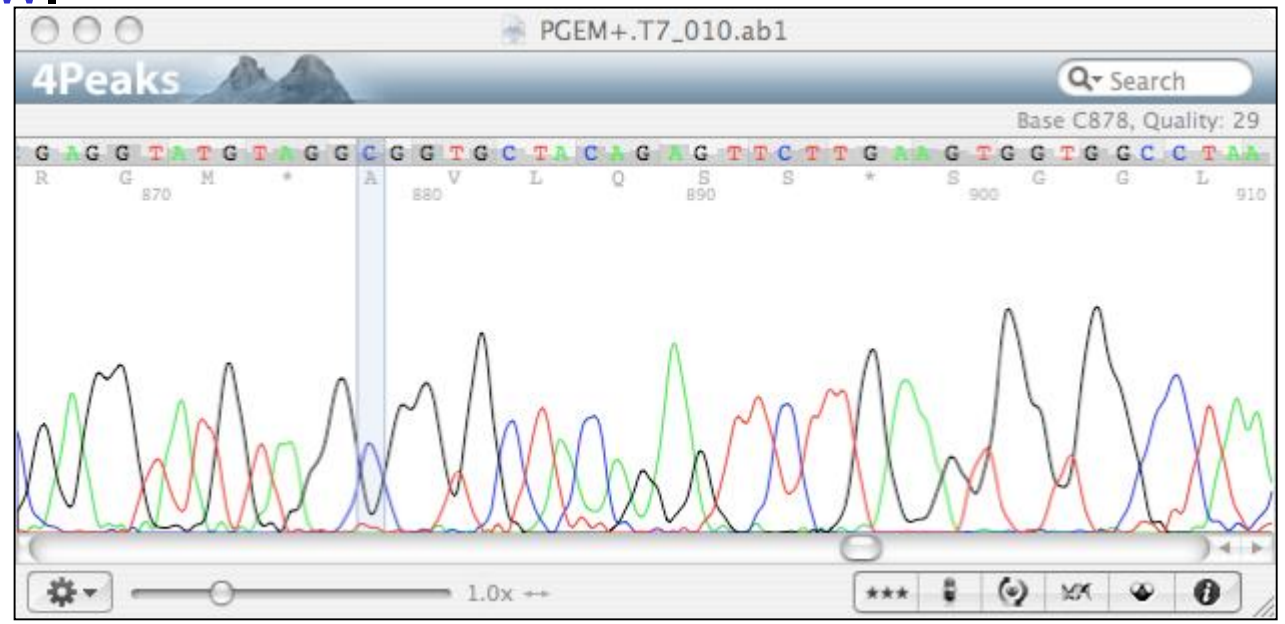
Michigan univ.DNA Sequencing Core

Steps of sequence analysis

Interpretation of sequencing chromatograms

A little baseline noise

- The peaks are no longer sharp but have become broad and more rounded in shape, especially where you have more than one of the same nucleotide in a row.



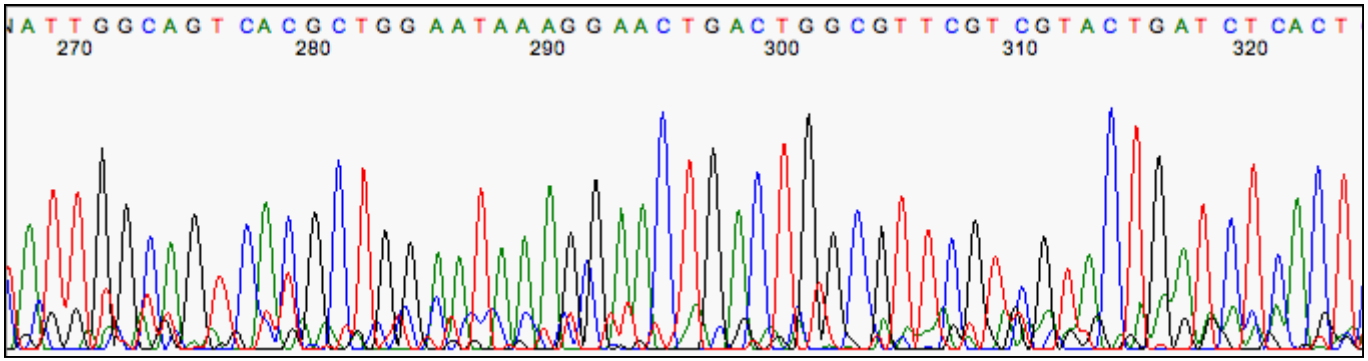
Steps of sequence analysis

Interpretation of sequencing chromatograms

A bit too much baseline noise

- Now we get to an example that has a bit too much noise. Note the:

1. Multicolored peaks at 271, 273 and 279,
2. The oddly-spaced interstitial peaks near 291, and 301,
3. And it is impossible to determine the real nucleotide is at 310.



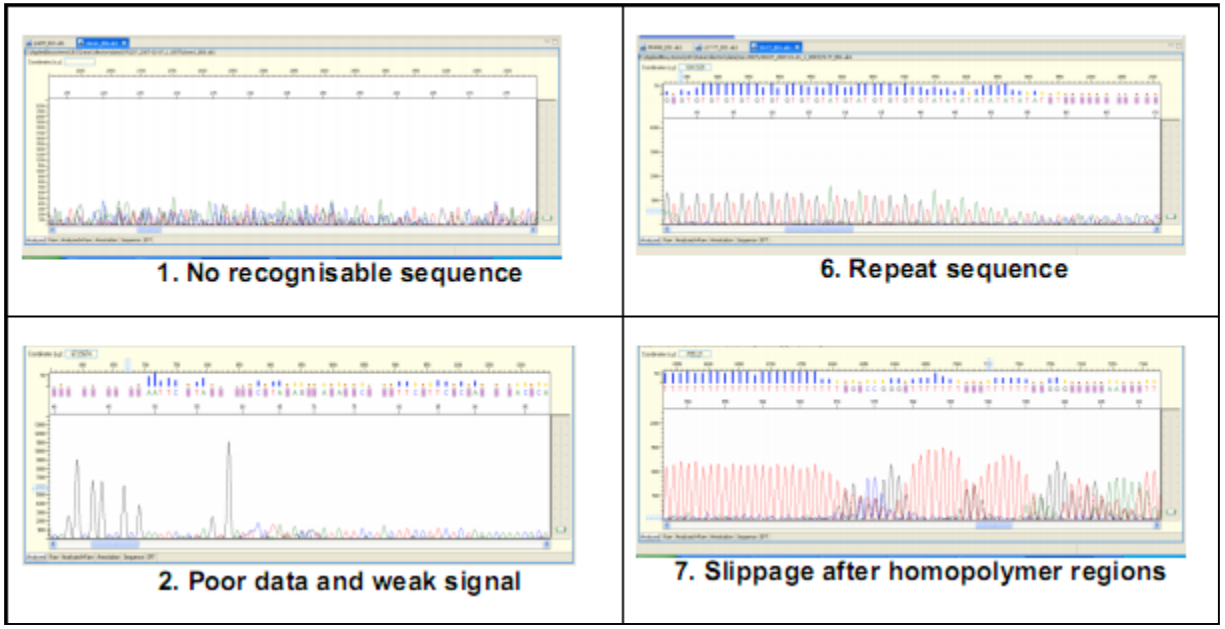
Michigan univ.DNA Sequencing Core

Steps of sequence analysis

Interpretation of sequencing chromatograms

Trouble-shooting of the sequence data

- When evaluating or trouble-shooting sequence data, it is important to look at the raw and analysed data traces in addition to data values (signal strength and start/end points) displayed in the annotation file.

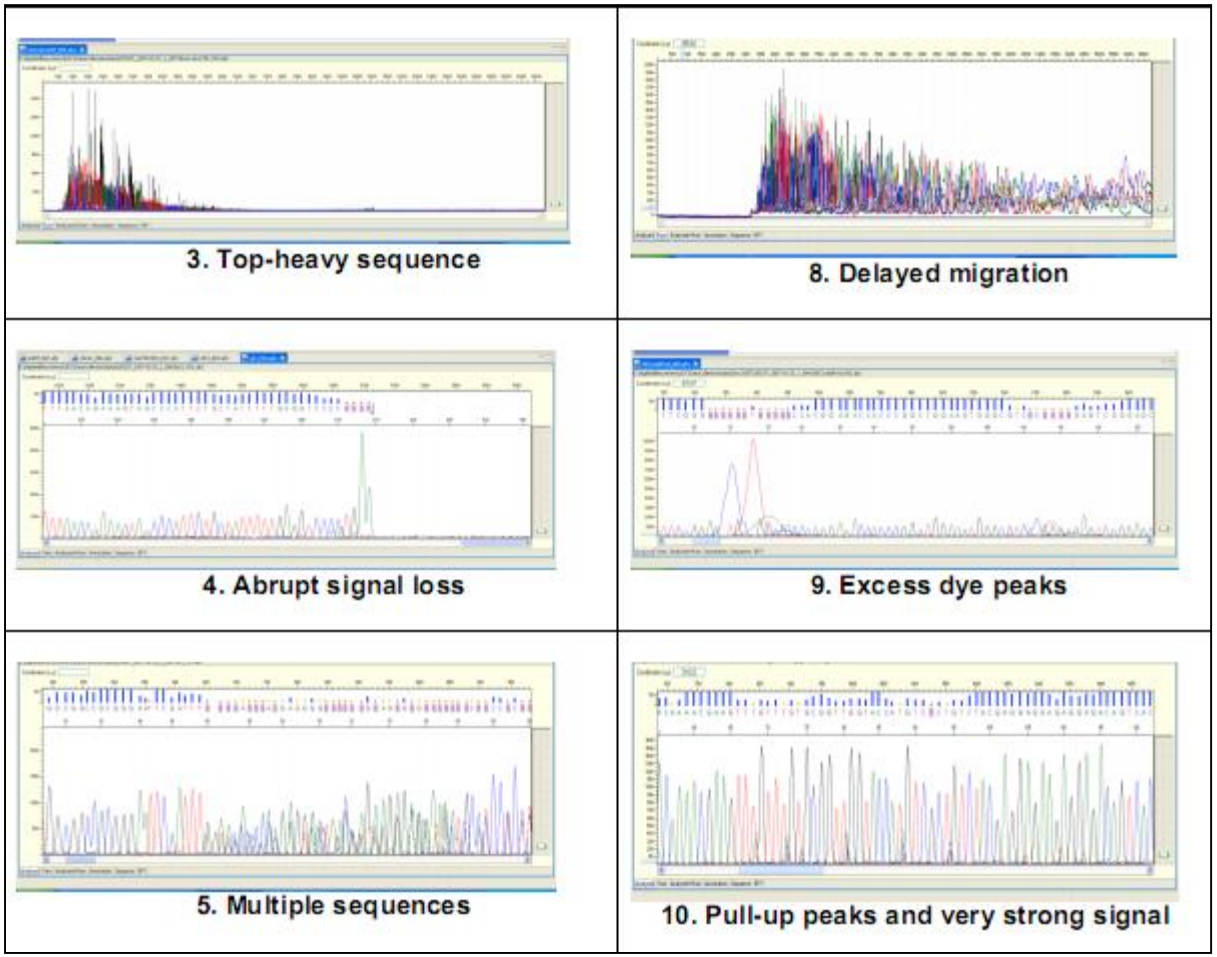


Vasic,2011

Steps of sequence analysis

Interpretation of sequencing chromatograms

Trouble-shooting of the sequence data



Vasic,2011

Steps of sequence analysis

Interpretation of sequencing chromatograms

Trouble-shooting of the sequence data

- A number of commercial and non-commercial software packages can trim low-quality DNA traces automatically.
- These DNA sequence assembly programs such as DNASTAR's SeqMan Ngen software score the quality of each peak and remove low-quality base peaks (generally located at the ends of the sequence).

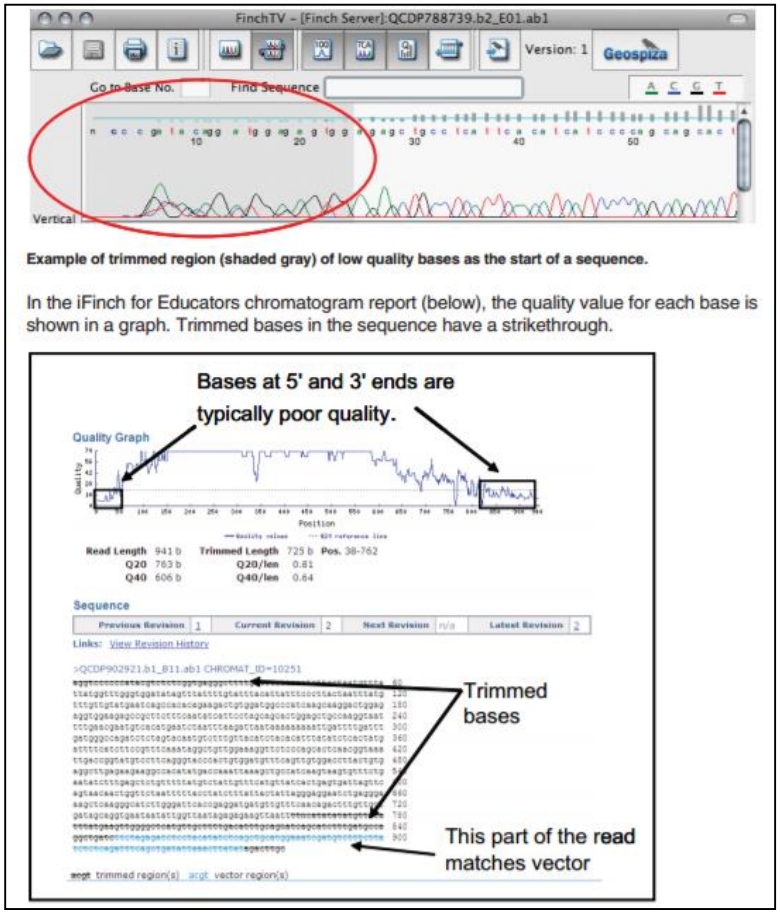
SeqMan is a sequence assembly tool that lets you assemble anything from two sequences to tens of thousands of sequences into contigs. DNASTAR software (Lasergene) first gained popularity in the 1980s and 1990s for its sequence assembly and analysis capabilities of Sanger sequencing assembly data. Lasergene 15.3 was released in 2018.

Steps of sequence analysis

Interpretation of sequencing chromatograms

Trouble-shooting of the sequence data

- Quality trimming: The sequence of bases from each chromatogram is called a read.
- Low quality bases at the beginning of the sequence were trimmed and are shown as having gray shading over them.
- Since the data at the 5'- and 3'-ends of reads are often of poor quality, a standard step in DNA sequencing is quality trimming.
- In this process, a program examines the quality of each base at either end of the read.
- In FinchTV, the trimmed regions are indicated by a gray shadow.

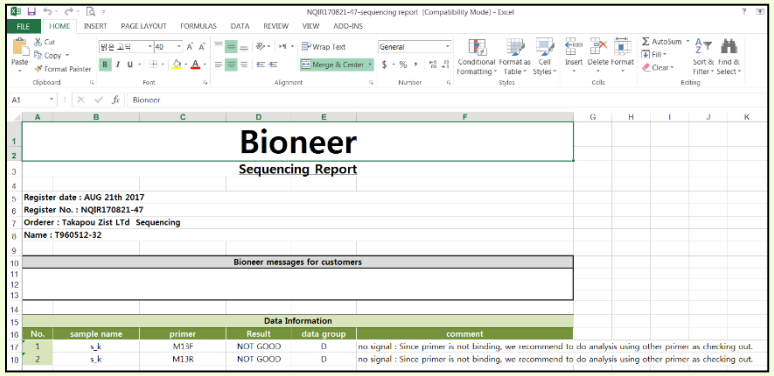


Sequence Report

Nucleotide sequences and Electropherogram

Failed at analysing. Re-reaction is recommended

- no signal: Since primer is not binding, we recommend to do analysis using other primer as checking out.



Bioneer

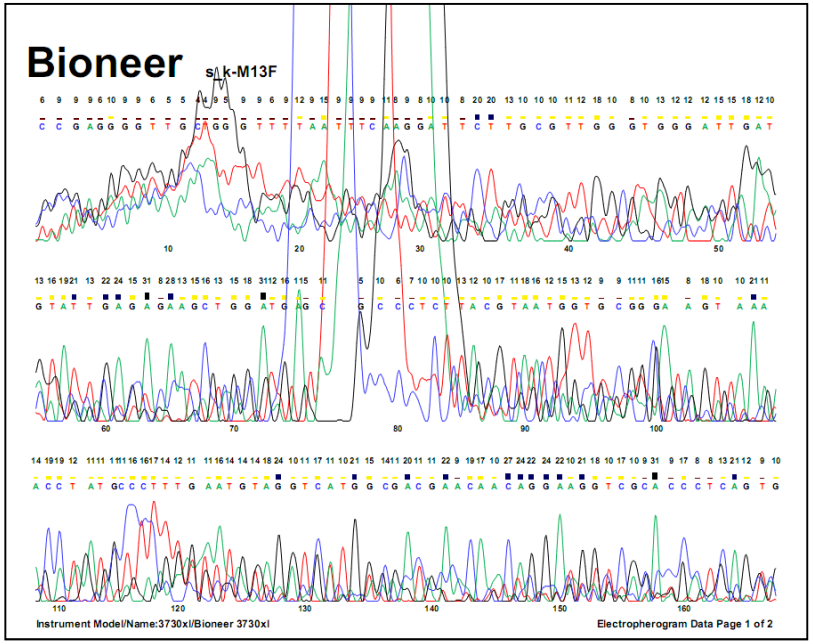
Sequencing Report

Register date : AUG 21th 2017
Register No. : NQIR170821-47
Orderer : Takapou Zist LTd Sequencing
Name : T960512-32

Bioneer messages for customers

Data Information

No.	sample name	primer	Result	data group	comment
1	s_k	M13F	NOT GOOD	D	no signal : Since primer is not binding, we recommend to do analysis using other primer as checking out.
2	s_k	M13R	NOT GOOD	D	no signal : Since primer is not binding, we recommend to do analysis using other primer as checking out.



Steps of sequence analysis

Interpretation of sequencing chromatograms

What is DNAstar Lasergene?

- DNAstar Lasergene is a suite of 7 programs:

1. SeqBuilder - for sequence editing and annotation, automated virtual cloning, and primer design;
2. SeqMan Pro- for contig assembly and analysis including SNP discovery, coverage evaluation, and project annotation.
3. MegAlign - for DNA and protein sequence alignments and analysis;
4. GeneQuest - for gene discovery and annotation;
5. Protean - for protein structure analysis and prediction;
6. PrimerSelect - for primer design;
7. EditSeq - for importing and editing unusual file types.

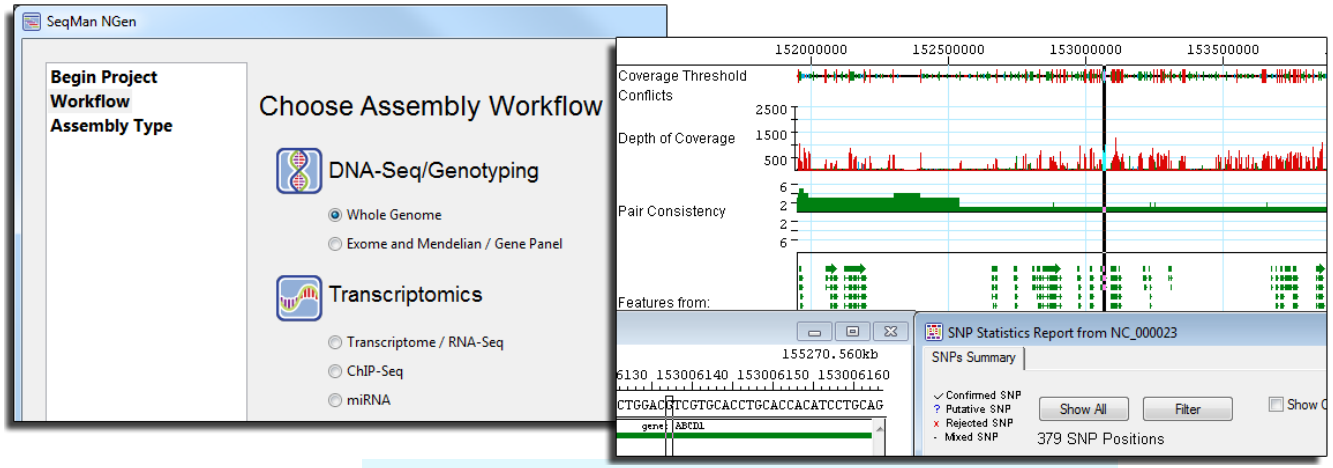
Note: SeqMan Pro analyzes the shape and intensity of peaks found in trace data to calculate quality scores, Q, and averaged quality scores, Q/n. A sequence contig is a contiguous, overlapping sequence read resulting from the reassembly of the small DNA fragments generated by bottom-up sequencing strategies.
Single nucleotide polymorphisms (SNPs) is the measurement of genetic variations of single nucleotide polymorphisms (SNPs) between members of a species.

Steps of sequence analysis

Interpretation of sequencing chromatograms

SeqMan NGen is fully integrated with SeqMan Pro

- SeqMan Pro is a great application for identifying gene mutations/ polymorphisms.
- SeqMan NGen is fully integrated with SeqMan Pro, so that once your assembly is complete, you can continue with downstream analysis including discovering SNPs and small indels, identifying large insertions and deletions, evaluating coverage, and annotating your consensus sequence.



Steps of sequence analysis

Interpretation of sequencing chromatograms

Reasons for trouble-shooting of the sequence data

1. **No recognisable sequence**
2. **Poor data and weak signal**
3. **Top-heavy sequence**

Sequence Problems

Below the automatic alignment, the Registry software will display Sequence Problems, specific nucleotide discrepancies between the reads and their target part.

Part	204	c a c a t a c t a g a g a a a
20982	265	c a c a t a c c a g a g a a a
20914	145	c a c a t a c c a g a g a a a

The nucleotide in question is colored red, and its locations on the documented sequence of the part, and the two reads are noted.

Steps of sequence analysis

Reasons for trouble-shooting of the sequence data

1. No recognisable sequence

- No analysed data is present because the signal-to-noise level is below the threshold for bases to be called.
- **Failed reactions are characterised by:**

1. The absence of clearly defined peaks in the raw data trace
2. The absence of base calls in the analysed electropherogram
3. Very low signal-to-noise ratios S/N G:<25 A:<25 T:<25 C:<25

- **The cause of a failed reaction may be due to:**

1. Insufficient or poor quality template and/or primer
2. Absence of primer annealing site or mutation in primer binding site
3. Failed sequencing reaction or clean-up.

- **Recommended actions include:**

1. Check template and/or primer concentrations and quality
2. Check primer binding site and primer design
3. Check ethanol concentrations and centrifugation speed and times

Vasic,2011

Steps of sequence analysis

Reasons for trouble-shooting of the sequence data

2. poor data and weak signal

- Poor data, weak signal and the presence of dye blobs are often a result of low template concentration.
- **Reactions displaying low signal strength are characterised by:**

1. Very low peak height in the raw data trace and the presence of dye blobs
2. In the analysed electropherogram, base calls fade off before the end of the read
3. Very low signal-to-noise ratios S/N G:<50 A:<50 T:<50 C:<50

- **Low signal strength may be the result of:**

1. Insufficient or poor quality template and/or primer
2. Poor primer design (low Tm) or mutation in primer binding site
3. Inferior reagents used or poor clean-up

- **Recommended actions include:**

1. Check template and/or primer concentrations and quality
2. Check primer binding site and primer design
3. Check sequencing reagents and clean-up protocol

Vasic,2011

Steps of sequence analysis

Trouble-shooting of the sequence data

3. Top-heavy sequence

- Sample set up with too much template. Template and primers are exhausted at the beginning of cycle sequencing creating an excess of short fragments
- **Reactions displaying top-heavy sequence are characterised by:**

1. Very high peaks in the raw data trace that fade off abruptly
2. In the analysed electropherogram, base calls fade off before the end of the read
3. An excess of short fragments are generated that are preferentially injected into the capillary

- **Top-heavy sequence may be the result of:**

1. Too much template used in the sequencing reaction
2. Too much primer used in the sequencing reaction

- **Recommended actions include:**

1. Check template concentration.
2. Check primer concentration. Use 3.2 pmol

Vasic,2011

Steps of sequence analysis

Reasons for trouble-shooting of the sequence data

4. Abrupt signal loss

- Abrupt signal loss is due to the presence of secondary structure.
- **Abrupt signal loss is often characterised by:**

1. Very high peaks in the raw data trace that stop abruptly
2. In the analysed electropherogram, base calls suddenly stop before the end of the read

- **Abrupt signal loss may be the result of:**

1. Secondary structure in the template
2. High GC content
3. Primer dimer contamination

- **Recommended actions include:**

1. Sequence complementary strand
2. Use a primer that anneals at a different position
3. Incubate the reaction at 96 degrees C for 10 minutes before cycling
4. Increase the extension temperature by 2 to 3 degrees C
5. Increase denaturation temperature to 98 degrees C
6. Add DMSO to a final concentration of 5%
7. Double all reaction components and incubate at 98 degrees for 10 minutes before cycling
8. Linearise the DNA with a restriction enzyme
9. Shear the insert into smaller fragments (<200bp) and subclone
10. Redesign primer to avoid primer dimer formation

Vasic,2011

Steps of sequence analysis

Reasons for trouble-shooting of the sequence data

5. Multiple sequences

- The presence of more than one PCR template in a reaction will result in overlapping sequences being generated.
- **Reactions displaying multiple sequences are characterized by:**

1. Lower peaks in the raw data trace
2. More than one sequence trace in the analysed data trace
3. More than one sequence commencing after base 50 to 100 (MCS)

- **Multiple sequences may be the result of:**

1. Mixed plasmid preparation
2. Multiple PCR products
3. Frame shift mutation
4. Primer-dimer contamination
5. Multiple priming sites
6. Multiple primers in reaction
7. Primer with N-1 contamination
8. Slippage after homopolymer or repeat regions in the template

- **Recommended actions include:**

1. Re-isolate the DNA from a pure colony and re-sequence
2. Check PCR template on gel for single band
3. Use a different primer after the mutation or sequence the complementary strand
4. Optimise PCR amplification or redesign primer
5. Make sure primer only has one priming site
6. Ensure only one primer has been used

Vasic,2011

Steps of sequence analysis

Reasons for trouble-shooting of the sequence data

6. Repeat sequences

- Signal loss is due to the presence of a repetitive sequence.
- **Reactions displaying repeat sequences are characterised by:**

1. The gradual decrease of peak height in the raw data trace after the repeat region.
2. In the analysed electropherogram, base calls fade off after the repeat region.

- **Recommended actions include:**

1. Sequence the complementary strand.
2. Use a primer that anneals at a different position.

Vasic,2011

Steps of sequence analysis

Reasons for trouble-shooting of the sequence data

7. Homopolymer regions

- Long homopolymer T regions (or A regions) can cause problems due to enzyme slippage.
- **Sequence data containing homopolymer regions display:**

1. Overlapping sequence following a homopolymer region due to slippage of the enzyme.

- **Recommended actions include:**

1. Sequence the complementary strand.
2. Use a primer that anneals at a different position.
3. Use an anchored primer (i.e., a sequencing primer that is polyT containing a A, C, or G base at the 3′ end of a poly A region).
4. The 3′ base will anchor the primer into place at the end of the homopolymer region.

Vasic,2011

Steps of sequence analysis

Reasons for trouble-shooting of the sequence data

8. Delayed migration

- Delayed samples often result from an excess of salt in the sample.
- **Sequence data displaying delayed migration show:**

1. Peaks commence after the usual start point of 600 to 800 in the raw data trace.
2. Peaks are not evenly spaced in the raw and analysed traces.
3. Poor base calls in the analysd electropherogram.

- **Delayed migration may be the result of:**

1. Contaminating negative ions (salts or other contaminants) in the sample being preferentially injected to the labeled fragments
2. Heavily overloaded samples. Excess of template used during sequencing.

- **Recommended actions include:**

1. Diluting the sample in deionized formamide and rerunning the sample can often correct this problem and yield good data.

Vasic,2011

Steps of sequence analysis

Reasons for trouble-shooting of the sequence data

9. Excess dye peaks at beginning of sequence

- Incomplete removal of excess dyes during the post cycle sequencing cleanup can obscure data at the beginning of the sequence.
- **Sequence data displaying excess dye peaks have:**

1. Peaks of excess dye present in the raw data trace
2. Dye blobs in the analysed data trace at positions 80, 120 and 190
3. Low signal to noise ratios S/N G:<50 A:<50 T:<50 C:<50

- **The presence of dye blobs in the data may be the result of:**

1. Incorrect estimation of template concentration (i.e. insufficient used)
2. Poor removal of unincorporated dye terminators

- **Recommended actions include:**

1. Check template concentration by agarose gel
2. Use fresh ethanol and sodium acetate (at room temp.) and use correct concentrations
3. With microfuge tubes, aspirate the supernatant rather than decanting
4. Do not use denatured alcohol
5. Do not leave reactions precipitating over night

Vasic,2011

Steps of sequence analysis

Reasons for trouble-shooting of the sequence data

10. Pull-up peaks and very strong signal

- Pull-up peaks and very high signal may result from use of too much template during cycle sequencing.
- **Sequence data displaying pull-up peaks are characterized by:**

1. Very high peaks in the raw data trace
2. very high peaks in the analysed data trace with pull up peaks and poor base calls
3. very high signal to noise ratios S/N G:>750 A:>750 T:>750 C:>750

- **Pull-up peaks may be the result of:**

1. Incorrect estimation of template concentration (i.e. too much used)

- **Recommended actions include:**

1. Diluting the sample in deionised formamide and rerunning the sample can often correct this problem and yield good data
2. Reduce the amount of template used in sequencing.

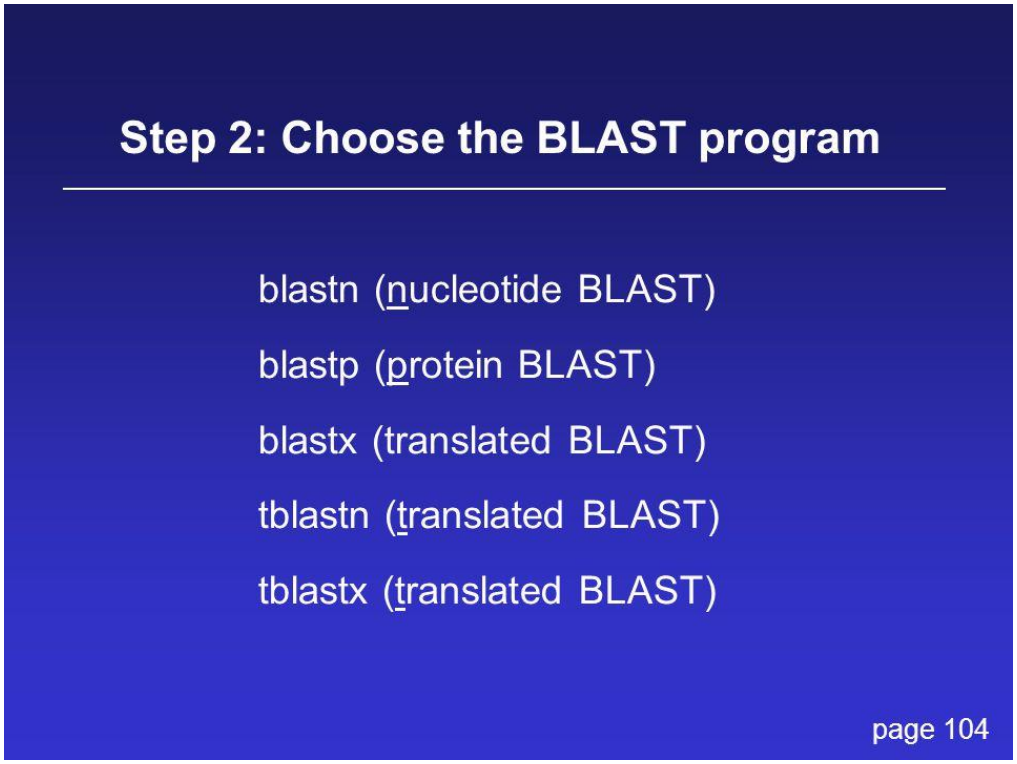
Vasic,2011

Steps of sequence analysis

BLAST search at GenBank

blastn, blastp, blastx, tblastn, tblastx

- BLAST (Basic Local Alignment Search Tool) search to identify similar sequence from the database.

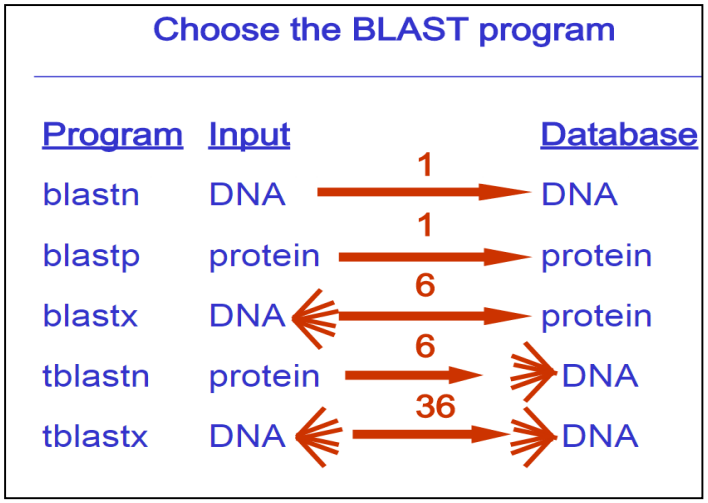


Steps of sequence analysis

BLAST search at GenBank

blastn, blastp, blastx, tblastn, tblastx

- BLAST (Basic Local Alignment Search Tool) search to identify similar sequence from the database.



BLAST program	*Query*	*Database*
Nucleotide blast (blastn)	Nucleotide	Nucleotide
Protein blast (blastp)	Protein	Protein
blastx	Translated Nucleotide	Protein
tblastn	Protein	Translated Nucleotide
tblastx	Translated Nucleotide	Translated Nucleotide

Steps of sequence analysis
BLAST search at GenBank

- To explore biodiversity you will work with the 16S rRNA sequence of your bacterium using databases and programs available on the Internet.
- This exercise will tell you whether your isolate corresponds to a previously characterized species, or whether it is something new.
- If it is new, you can determine which organisms are its closest relatives, and perhaps predict some of the physiological properties of your unknown based on what is known about its relatives.

Moran,2010;…

Steps of sequence analysis
BLAST search at NCBI GenBank

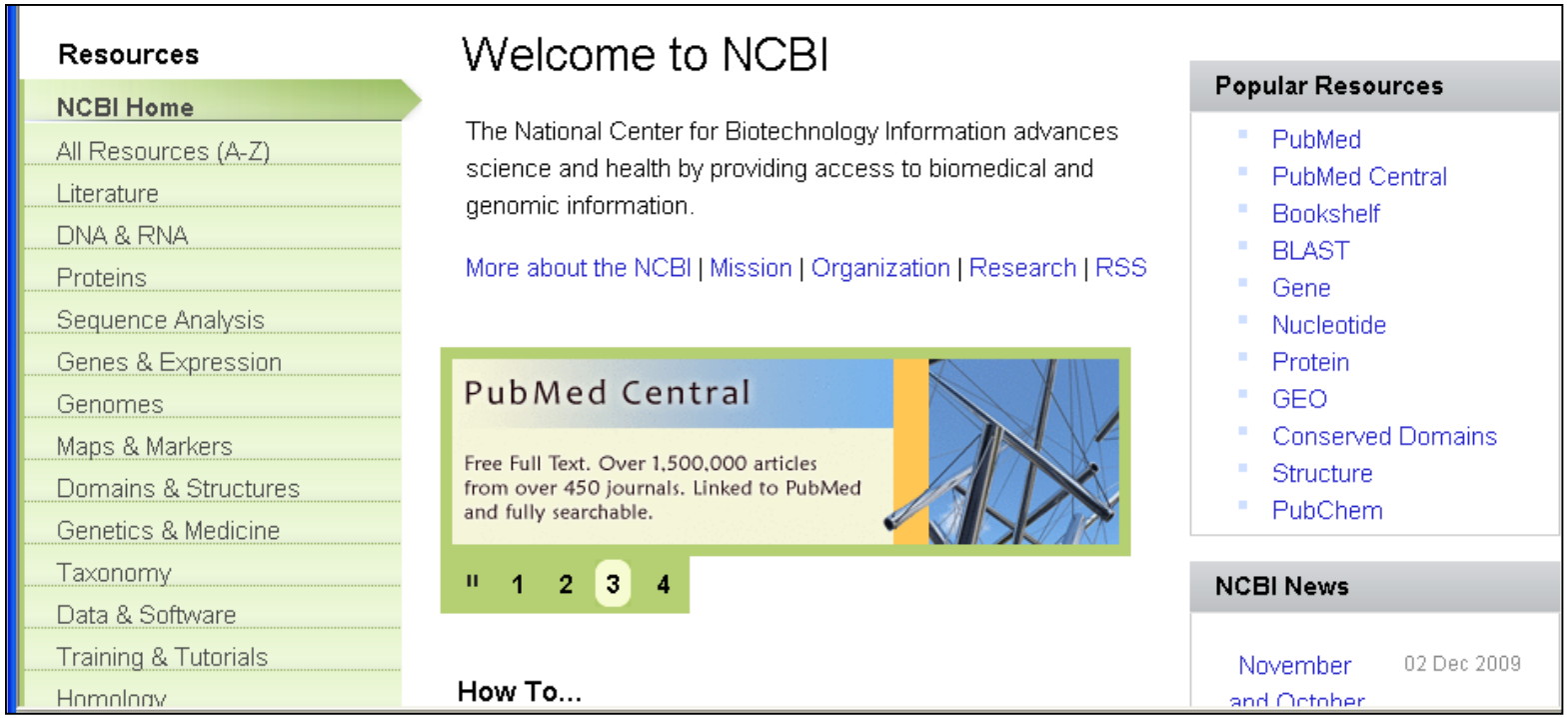
- GenBank is the National Institutes of Health (NIH) genetic sequence database, a collection of all publicly available DNA sequences.
- Scientists who determine the sequences of genes (or sometimes of entire genomes) deposit the sequence information in GenBank.
- Aside from doing this as a service to the scientific community, many journals require submission of new sequence data to GenBank as a precondition to publishing a manuscript referring to the sequence.

Note: GenBank is a comprehensive sequence database. It is advised to edit sequences prior depositing in GenBank database.

Moran,2010

NCBI

National Center for Biotechnology Information



Steps of sequence analysis

BLAST search at GenBank or other servers

- Choose one of the public servers:
  - NCBI www.ncbi.nlm.nih.gov/blast
  - EBI www.ebi.ac.uk/blast
  - EMBNet www.expasy.ch/blast
- Select a database to search:
  - NR to find any protein sequence
  - Swiss-Prot to find proteins with known functions
  - PDB to find proteins with known structures
- Cut and paste your sequence
- Click the **BLAST** button

Spiegel,2007

Steps of sequence analysis
BLAST search at GenBank

- GenBank has approximately 13,543,000,000 bases in 12,814,000 sequence records as of August 2001, derived from viruses, bacteria, and complex eukaryotic organisms.
- Included among these many sequences are a number of 16S rRNA gene sequences from Bacteria and Archaea.
- The current number of 16S rRNA sequences is approximately 12,000.
- Some of these are 'complete' (they cover the entire 1500+ bases of the 16S rRNA gene), while other are 'partial' (like yours will be) and cover only a portion of the gene (your sequence includes only ~400 bases near the beginning of the gene).

Moran,2010

Steps of sequence analysis
BLAST search at GenBank

- A sophisticated computer program called BLAST compares the query sequence you submit to all of the sequences previously deposited in the GenBank database.
- BLAST finds sequences in the database that are most similar to the submitted sequence, and then returns the names and a brief description of the matching sequences in the output from the program.

In typical usage, the query sequence is much smaller than the Refseq database, e.g., the query may be one thousand nucleotides while the database is several billion nucleotides.

Moran,2010

Steps of sequence analysis

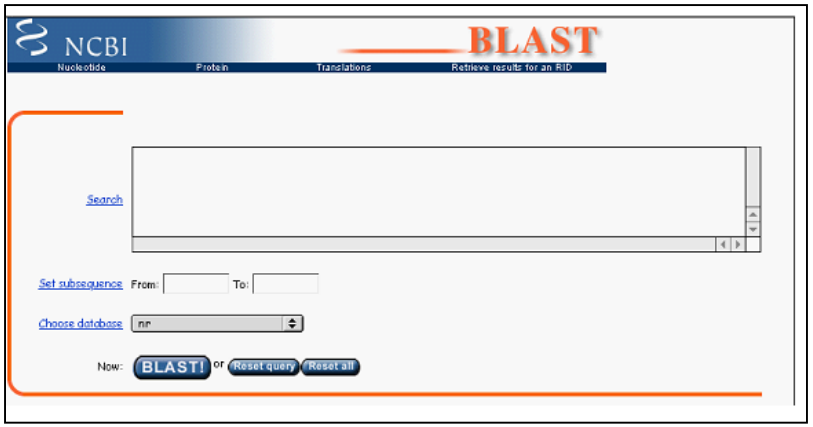
BLAST search at GenBank

- BLAST (Basic Local Alignment Search Tool) is a tool for comparing one sequence with all the other sequences in a database.
- BLAST can compare:
  - DNA sequences
  - Protein sequences
- BLAST is more accurate for comparing protein sequences than for comparing DNA sequences.

Spiegel,2007

Steps of sequence analysis
BLAST search at GenBank

- View of NCBI BLAST page.
- Paste your sequence into the top box labeled 'Search'.
- Start the query by clicking the 'BLAST!' button.

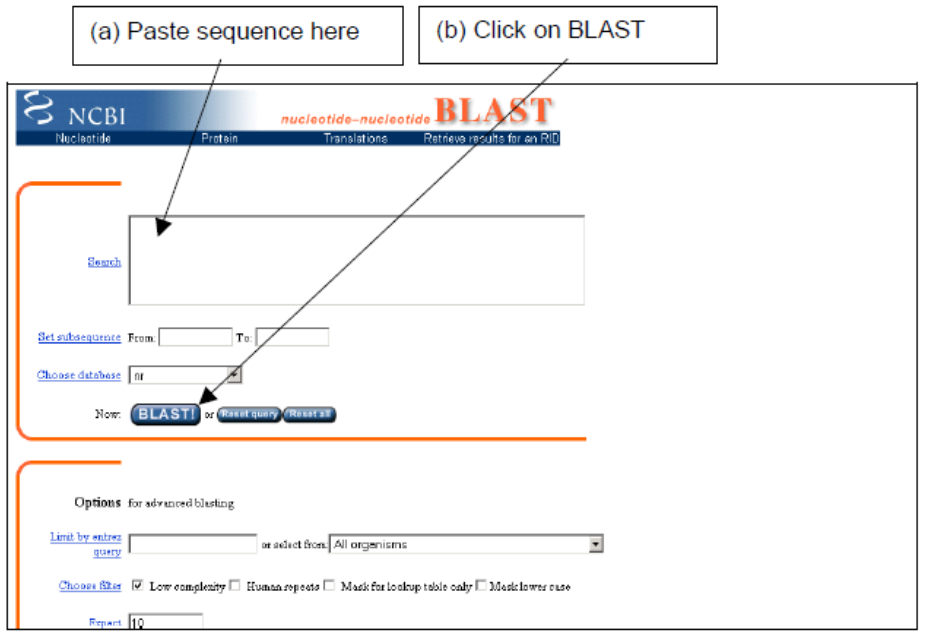


Searching the databases with short sequences may result in an error. The minimum length is 11 bases, however the recommended minimum sequence length is 22 bases (nucleotides/blastn) or 6 amino acids (proteins/blastp). Usually short queries can match both long and short sequences, producing many answers.

Moran,2010;..

Steps of sequence analysis
BLAST search at GenBank

- View of NCBI BLAST page.
- Paste your raw or edited sequences into the top box labeled 'Search'.
- Start the query by clicking the 'BLAST!' button.



Moran,2010;Boonham,2008

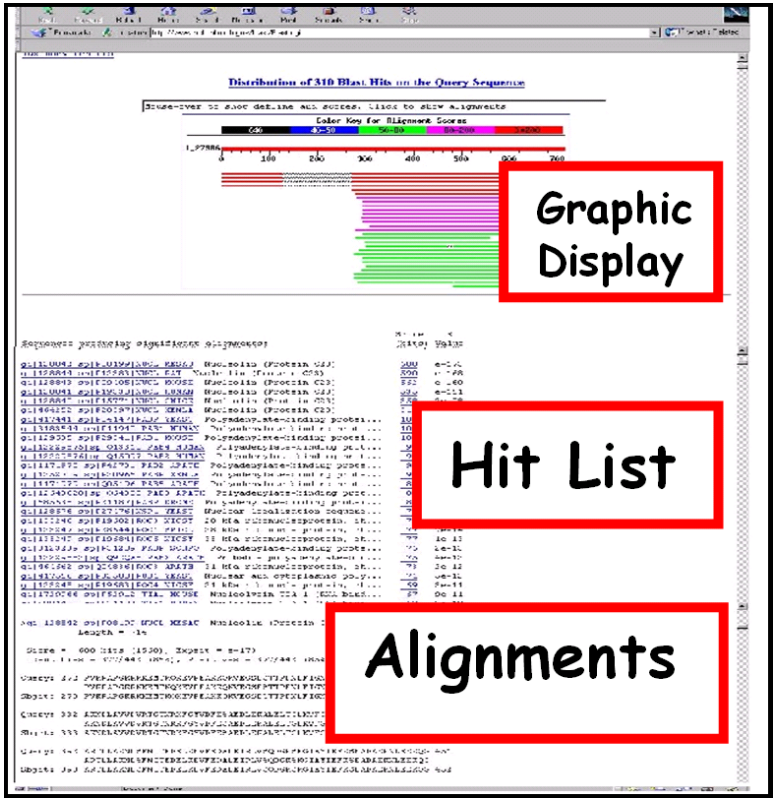
Interpreting BLAST result
Reading BLAST Output

- The Graphic Summary shows alignments (as colored boxes) of database matches to our Query sequence (solid red bar under the color key).
- The query coordinates and length correspond to the numbered scale across the top.
- Sequences found by BLAST, "hits," are represented as horizontal bars below this scale.
- These will vary in length, position, and color-coded scoring.
- Generally, the higher the alignment score, the more significant the hit.
- Each line represents a hit from your blast search.

Interpreting BLAST result

Reading BLAST Output

- Graphic Display
  - Overview of the alignments
- Hit List
  - Gives the score of each match
- Alignments
  - Details of each alignment



A high-scoring segment pair, or HSP, is a subsegment of a pair of sequences, either nucleotide or amino acid, that share high level of similarity. Each HSP is drawn as a line, and is aligned with the query sequence and each hit may contain one or more high scoring pairs (HSPs). Hit names and P-values are displayed at the left side of the graph.

Interpreting BLAST result

Reading BLAST Output

Hits are Color Coded

- The hits are color coded according to their P value. A set of five fixed ranges is used to determine a color for each hit. These ranges, from "worst" to "best" are:

  1.0 to 1e-10
  1e-10 to 1e-50
  1e-50 to 1e-100
  1e-100 to 1e-200
  1e-200 to 0.0

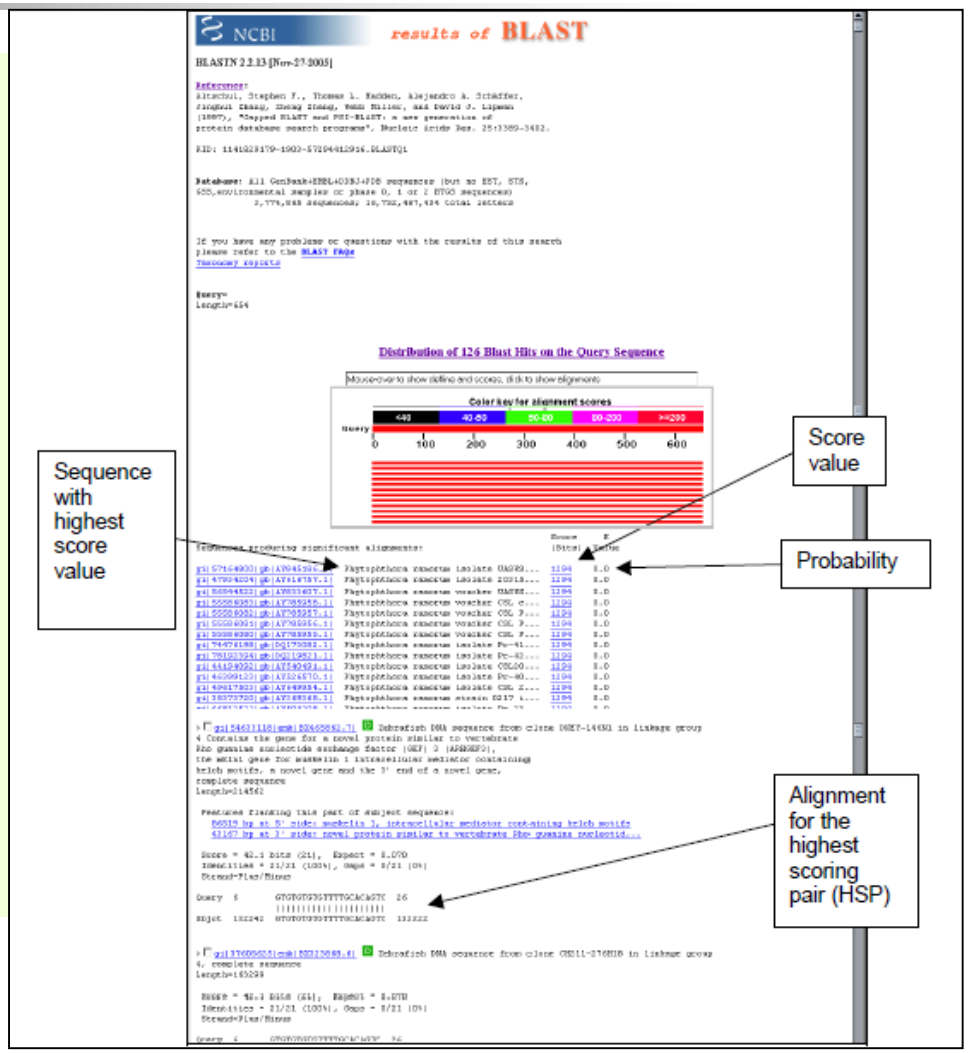
- The key shows these colors, and notes the negative value of the exponents in each range. It progresses from "worst" on the left to "best" on the right.

- Note that ranges might not contain any hits, since the ranges are fixed while the hit P-values are not. When ranges share a boundary value (e.g.: 1e-50), that value falls in the "better" range and will be colored thus (e.g.: green).

Interpreting BLAST result

Reading BLAST Output

- Graphic Display
  - Overview of the alignments
- Hit List
  - Gives the score of each match
- Alignments
  - Details of each alignment

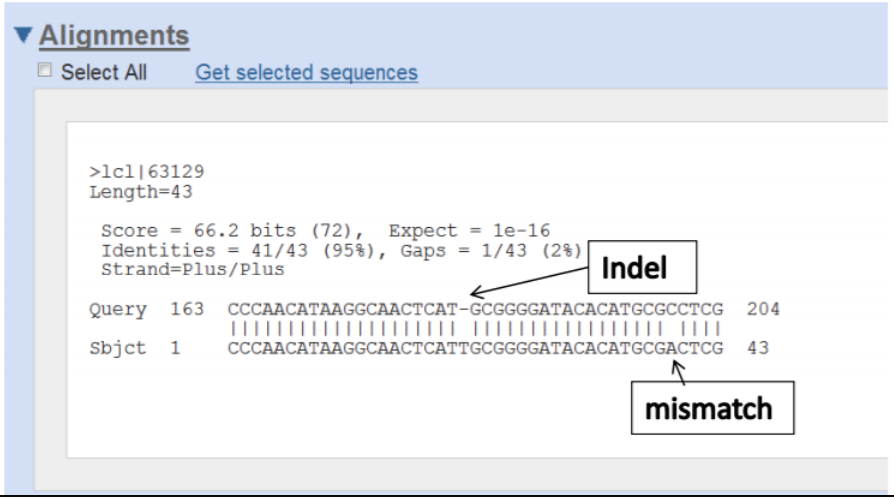


Spiegel,2007;Boonham,2008

Interpreting nucleotide blast results
Reading BLAST Output

Example 2: Local alignment of a 300bp sequence and an internal fragment containing a single insertion and a single mismatch using BLAST

>Random seq 1
TCGAAGGCGCTCGGTAGAGTACGTGTCCCAACTGTTGCCTAAGCGCGCGTACAGTAGGGCGAGGCACGCTACTG
TTACGAGATTCCTACCGAAGAAAAGTTAAGCCCCTCGAAAGGTAACCATCGGAGCCCGTGATCTGGCATGAAATA
CTACGGGCCTTCCCCCAACATAAGGCAACTCATTGCGGGGATACACATGCGACTCGGTCCGATATGATTGCCGCA
TTTTCACGGTTGCCTCATCAAGCCCGCCAACGGGTTAGTGGAACGAATATGAGGCAGACTCTCACATCGCTATCT
GT



Tadmor,2010

Interpreting BLAST result
Interpreting the score values

- The first format (under the heading "Distribution of Blast Hits on the Query Sequence") is a graphical representation of the results, showing the matches to your input sequence as a stack of colored bars with the best matches at the top.
- The color of the bar indicates the match score (i.e., degree of similarity) between your sequence and the database matches.
- Red indicates the best matches,
- Black indicates the poorest.
- The length of the bar shows the length of the sequence that was noticeably similar to the query.

Moran,2010

Interpreting BLAST result
Interpreting the score values

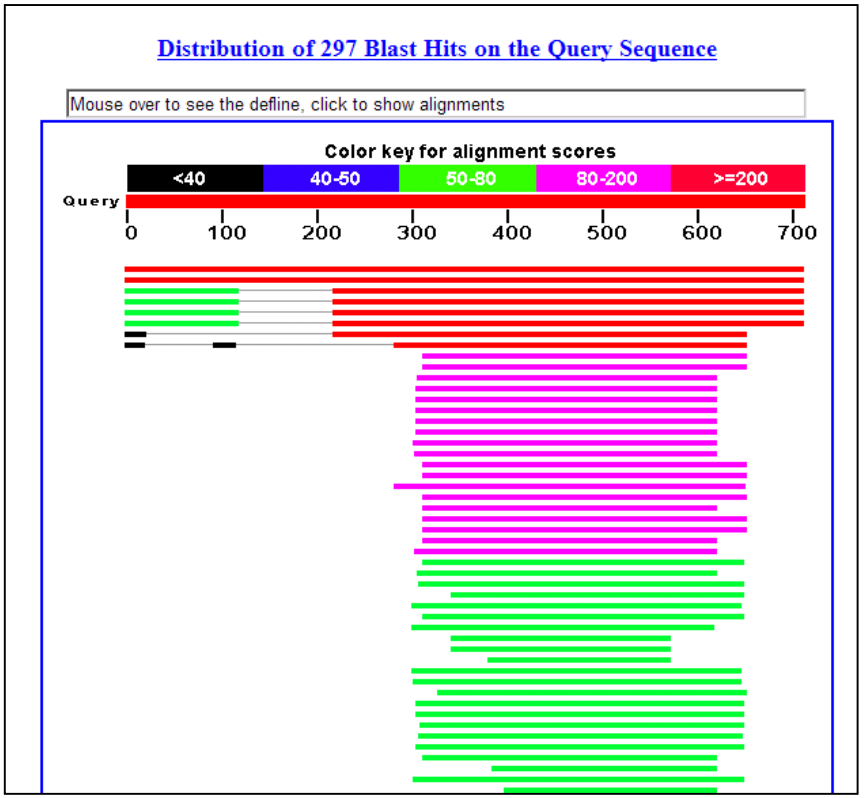
- As its name suggests, BLAST (Basic Local Alignment Search Tool) is designed to identify local regions of sequence similarity.
- This means that BLAST may report multiple discrete regions of sequence similarity between a query sequence and a subject sequence in a database.
- The color of the boxes corresponds to the score (S) of the alignment, with red representing the highest alignment scores.

Interpreting BLAST result

Reading BLAST Output

The graphic display

- The Horizontal Axis (0-700) corresponds to your protein (query).
- Color codes indicate that match's quality
  - Red: very good
  - Green: acceptable
  - Black: bad
- The top two (solid red bar) hits match much better to our sequence than the remaining BLAST hits.
- Thin lines join independent matches on the same sequence.



Spiegel,2007

Interpreting BLAST result
The Graphic Display
List of Significant BLAST Hits

- The BLAST algorithm calculates similarity scores for local alignments (i.e., the most similar regions between 2 sequences) between the query sequence and subject sequences using specific scoring matrices, and returns a table of the best matches ("hits") from the database.
- The hit table includes several useful pieces of information, including:

1. The similarity score, query coverage (percent of the query sequence that overlaps the subject sequence),
2. E-value, and
3. Max identity (percent similarity between the query and subject sequences over the length of the coverage area).

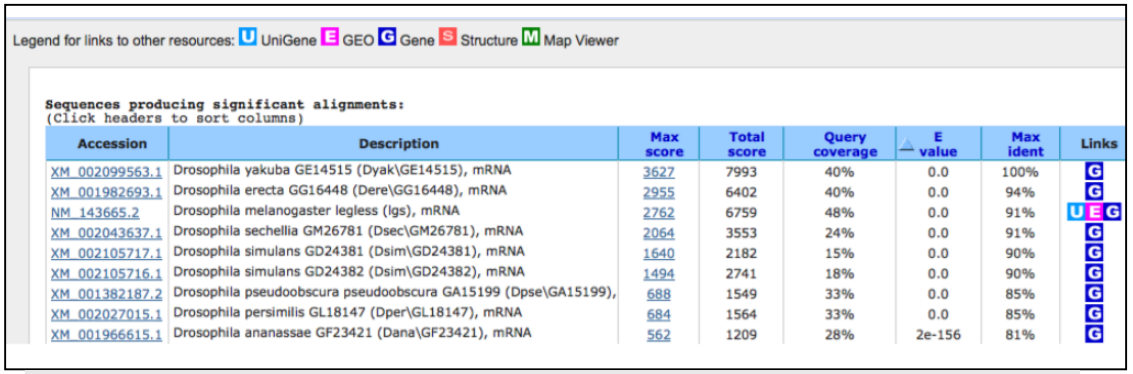
Superimposed sequences: degenerated, overlapping) sequences.

Interpreting BLAST result

The Graphic Display

List of Significant BLAST Hits

- List of blastn hits that produce significant alignments with our query sequence. Scrolling further down the output, we find a summary table that shows all the sequences in the Refseq database that show significant sequence homology to our sequence.
- By default, the results are sorted according to the Expect value (E-value) in ascending order. We can click on the column headers to sort the results by different categories.



Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

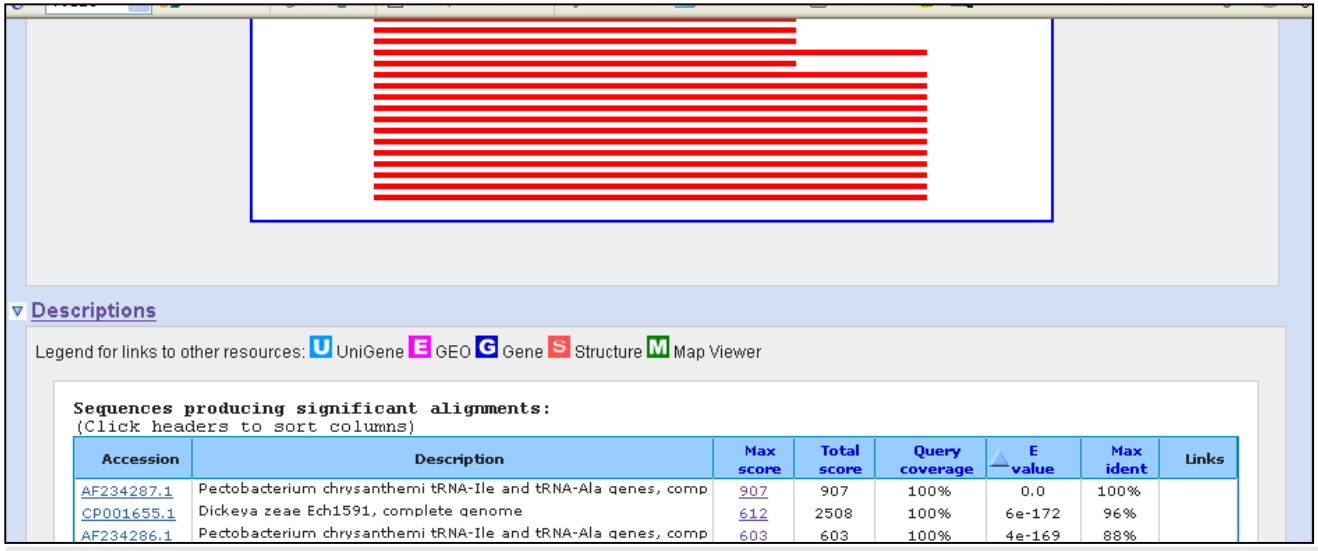
Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
XM_002099563.1	Drosophila yakuba GE14515 (Dyak\GE14515), mRNA	3627	7993	40%	0.0	100%	G
XM_001982693.1	Drosophila erecta GG16448 (Dere\GG16448), mRNA	2955	6402	40%	0.0	94%	G
NM_143665.2	Drosophila melanogaster legless (lgs), mRNA	2762	6759	48%	0.0	91%	U E G
XM_002043637.1	Drosophila sechellia GM26781 (Dsec\GM26781), mRNA	2064	3553	24%	0.0	91%	G
XM_002105717.1	Drosophila simulans GD24381 (Dsim\GD24381), mRNA	1640	2182	15%	0.0	90%	G
XM_002105716.1	Drosophila simulans GD24382 (Dsim\GD24382), mRNA	1494	2741	18%	0.0	90%	G
XM_001382187.2	Drosophila pseudoobscura pseudoobscura GA15199 (Dpse\GA15199),	688	1549	33%	0.0	85%	G
XM_002027015.1	Drosophila persimilis GL18147 (Dper\GL18147), mRNA	684	1564	33%	0.0	85%	G
XM_001966615.1	Drosophila ananassae GF23421 (Dana\GF23421), mRNA	562	1209	28%	2e-156	81%	G

Comparative sequence analysis revealed 99.0-100% sequence similarities for the 16S rRNA genes between the reference strains and target strains.

Interpreting nucleotide blast results
The Graphic Display
List of Significant BLAST Hits

1. query coverage (percent of the query sequence that overlaps the subject sequence),
2. max identity (percent similarity between the query and subject sequences over the length of the coverage area).



In this case, we notice that the top three hits match much better to our sequence than the remaining BLAST hits. We also see that these three database matches span almost the entire length of our Query sequence.

Interpreting nucleotide blast results

What is the the minimum percentage of identity required by BLAST to tentatively determine new species of bacteria?

- **Qs.** Is a similarity of 16S rRNA of the unknown isolate with reference species above 97% represents a new species?
- **Ans.** It might be a new species. It might not be a new species.
- 16S rRNA sequences by themselves don't make it easy to define a new species as our idea of what is a species or a strain is mostly artificial anyway. If there is other evidence to suggest that your species is somehow different (like other genomic or phenotypic differences, for example) then it may complement the different 16S sequence.
- I've seen varying thresholds for what constitutes a different species or genus, anyway. I have linked a paper which suggests that 98.7 percent is the usual species cutoff. They do also note "These genetic criteria should always be accompanied by a discriminant phenotypic property".

Harry Caufield

Interpreting nucleotide blast results

What is the the minimum percentage of identity required by BLAST to tentatively determine new species of bacteria?

- **Qs.** Is a similarity of 16S rRNA of the unknown isolate with reference species above 97% represents a new species?
- **Ans.**
- If you have complete sequence of 16S rRNA gene (approximately 1500 bp) and it is showing 96% similarity with type strain (not with other strain), then it should be a new species.
- The 16S RNA sequence similarity is not enough by itself to talk about a new species. To define a new species DNA-DNA hybridization is done with all the Type strains which have less than 97% similarity with your strain.

Interpreting nucleotide blast results

Reading BLAST Output

The Hit List or Sequence List

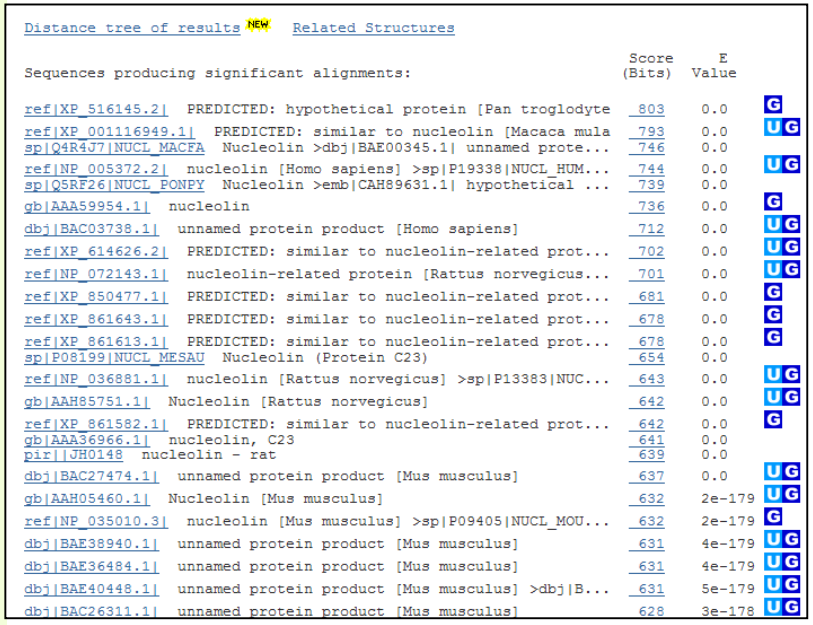
- Below the graphical output is a list of the top 100 sequences that matched your query, giving:
- Their reference number in the database,
- A very brief description,
- A score (the higher the score, the more similar the sequences), and
- An E value (the odds that the similarity between the query sequence and the match might have occurred purely by chance; the lower the score, the more similar the sequences).

E value decreases exponentially as the Score (S) of the match increases.

Moran,2010

Interpreting nucleotide blast results
Reading BLAST Output
The Hit List

- Sequence accession number
  - Depends on the database
- Description
  - Taken from the database
- Bit score
  - **High** bit score = **good** match
- E-Value (Expect value)
- **Low** E-value = **good** match
- Links
  - Genome
  - Uniref, database of transcripts



Distance tree of results NEW Related Structures

Sequences producing significant alignments:	Score (Bits)	E Value	
ref|XP_516145.2| PREDICTED: hypothetical protein [Pan troglodyte	803	0.0	G
ref|XP_001116949.1| PREDICTED: similar to nucleolin [Macaca mula	793	0.0	UG
sp|Q4R4J7|NUCL_MACFA Nucleolin >dbj|BAE00345.1| unnamed prote...	746	0.0	
ref|NP_005372.2| nucleolin [Homo sapiens] >sp|P19338|NUCL_HUM...	744	0.0	UG
sp|Q5RF26|NUCL_PONPY Nucleolin >emb|CAH89631.1| hypothetical ...	739	0.0	
gb|AAA59954.1| nucleolin	736	0.0	G
dbj|BAC03738.1| unnamed protein product [Homo sapiens]	712	0.0	UG
ref|XP_614626.2| PREDICTED: similar to nucleolin-related prot...	702	0.0	UG
ref|NP_072143.1| nucleolin-related protein [Rattus norvegicus...	701	0.0	UG
ref|XP_850477.1| PREDICTED: similar to nucleolin-related prot...	681	0.0	G
ref|XP_861643.1| PREDICTED: similar to nucleolin-related prot...	678	0.0	G
ref|XP_861613.1| PREDICTED: similar to nucleolin-related prot...	678	0.0	G
sp|P08199|NUCL_MESAU Nucleolin (Protein C23)	654	0.0	
ref|NP_036881.1| nucleolin [Rattus norvegicus] >sp|P13383|NUC...	643	0.0	UG
gb|AAH85751.1| Nucleolin [Rattus norvegicus]	642	0.0	UG
ref|XP_861582.1| PREDICTED: similar to nucleolin-related prot...	642	0.0	G
gb|AAA36966.1| nucleolin, C23	641	0.0	
pir||JH0148 nucleolin - rat	639	0.0	
dbj|BAC27474.1| unnamed protein product [Mus musculus]	637	0.0	UG
gb|AAH05460.1| Nucleolin [Mus musculus]	632	2e-179	UG
ref|NP_035010.3| nucleolin [Mus musculus] >sp|P09405|NUCL_MOU...	632	2e-179	G
dbj|BAE38940.1| unnamed protein product [Mus musculus]	631	4e-179	UG
dbj|BAE36484.1| unnamed protein product [Mus musculus]	631	4e-179	UG
dbj|BAE40448.1| unnamed protein product [Mus musculus] >dbj|B...	631	5e-179	UG
dbj|BAC26311.1| unnamed protein product [Mus musculus]	628	3e-178	UG

Uniref, also known as UniProt NREF, is a set of comprehensive protein databases curated by the Universal Protein Resource consortium.

Spiegel,2007

Interpreting nucleotide blast results
Reading BLAST Output
E-value

- The Expect value (E) is a parameter that describes the number of hits one can "expect" to see by chance when searching a database of a particular size.
- It decreases exponentially as the Score (S) of the match increases.
- Essentially, the E value describes the random background noise.
- The lower the E-value, the more "significant" a match to a database sequence is (i.e. there is a smaller probability of finding a match just by chance).
- For example, an E value of 1 assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see 1 match with a similar score simply by chance.

Interpreting nucleotide blast results
Reading BLAST Output
E-value

- e-values in blast results represent the probability of the alignment occurring by chance.
- It is a statistical calculation based on the quality of alignment (the score) and the size of the database.
- The lower the E-value, or the closer it is to zero, the more "significant" the match is.

1. An e value of 0 is actually a rounded down probability (maybe 1e-250 or something), and is simply saying that there is (almost) no chance that alignment can occur by chance.
2. An e-value of 1e-3 is saying that there is a 0.001 chance that that alignment would exist in the database by chance, that is, if the database contains 10000 sequences, then you might expect that alignment to occur maybe 10 times (10000/1000=10).

Interpreting nucleotide blast results
Reading BLAST Output
E-value

- However, keep in mind that virtually identical short alignments have relatively high E values.
- This is because the calculation of the E value takes into account the length of the query sequence.
- These high E values make sense because shorter sequences have a higher probability of occurring in the database purely by chance.
- For more details please see the calculations in the BLAST Course.

Steps of protein sequence analysis

BLAST search at GenBank

blastp (blasting protein sequences)

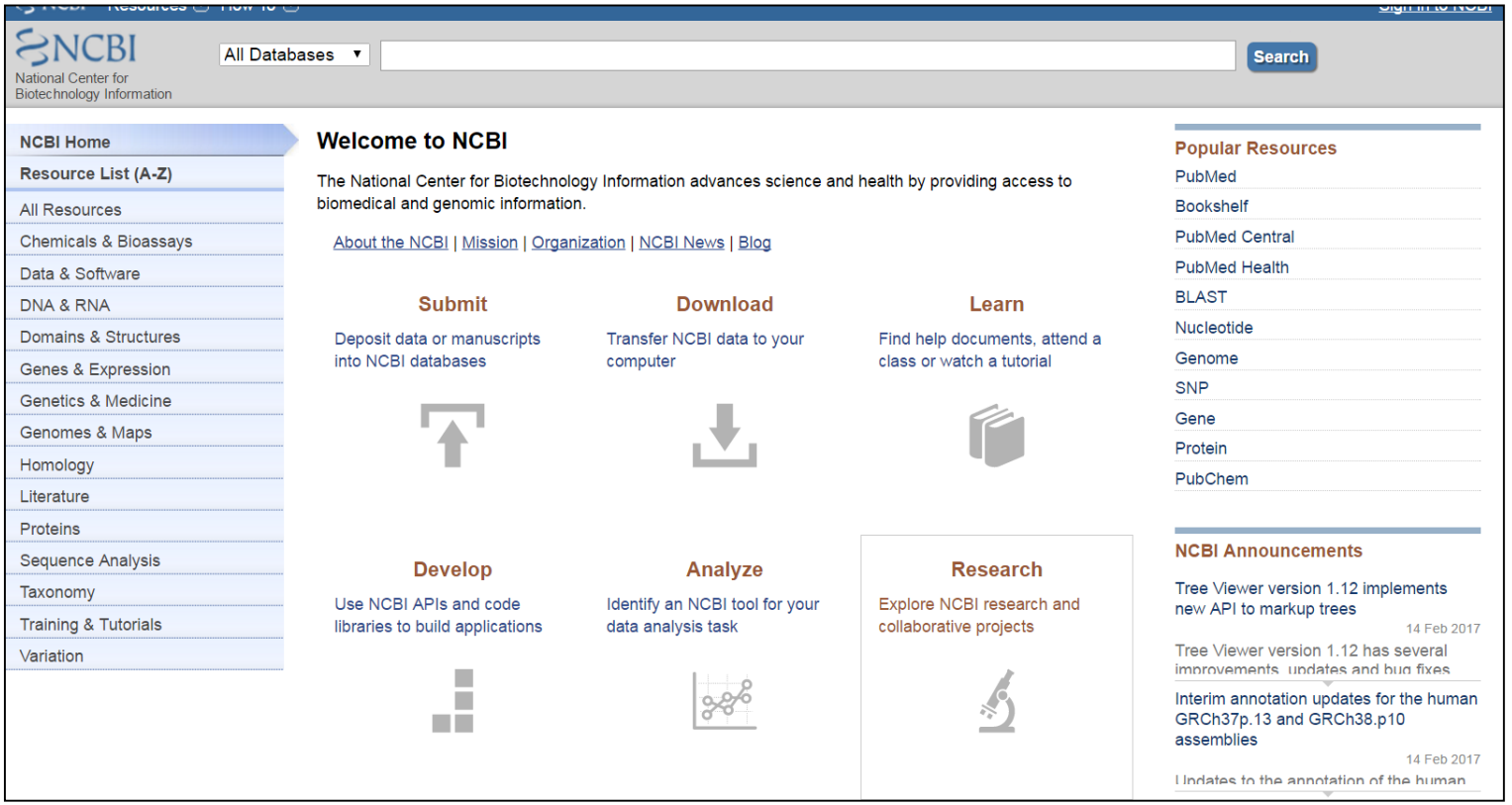
- blastp = Compares a protein sequence with a protein database.
- If you want to find something about the function of your protein, use blastp to compare your protein with other proteins contained in the databases.
- Two of the most popular blastp online services:

1. NCBI (National Center for Biotechnology Information) server
2. Swiss EMBnet server (European Molecular Biology network)

Bordoli,2003

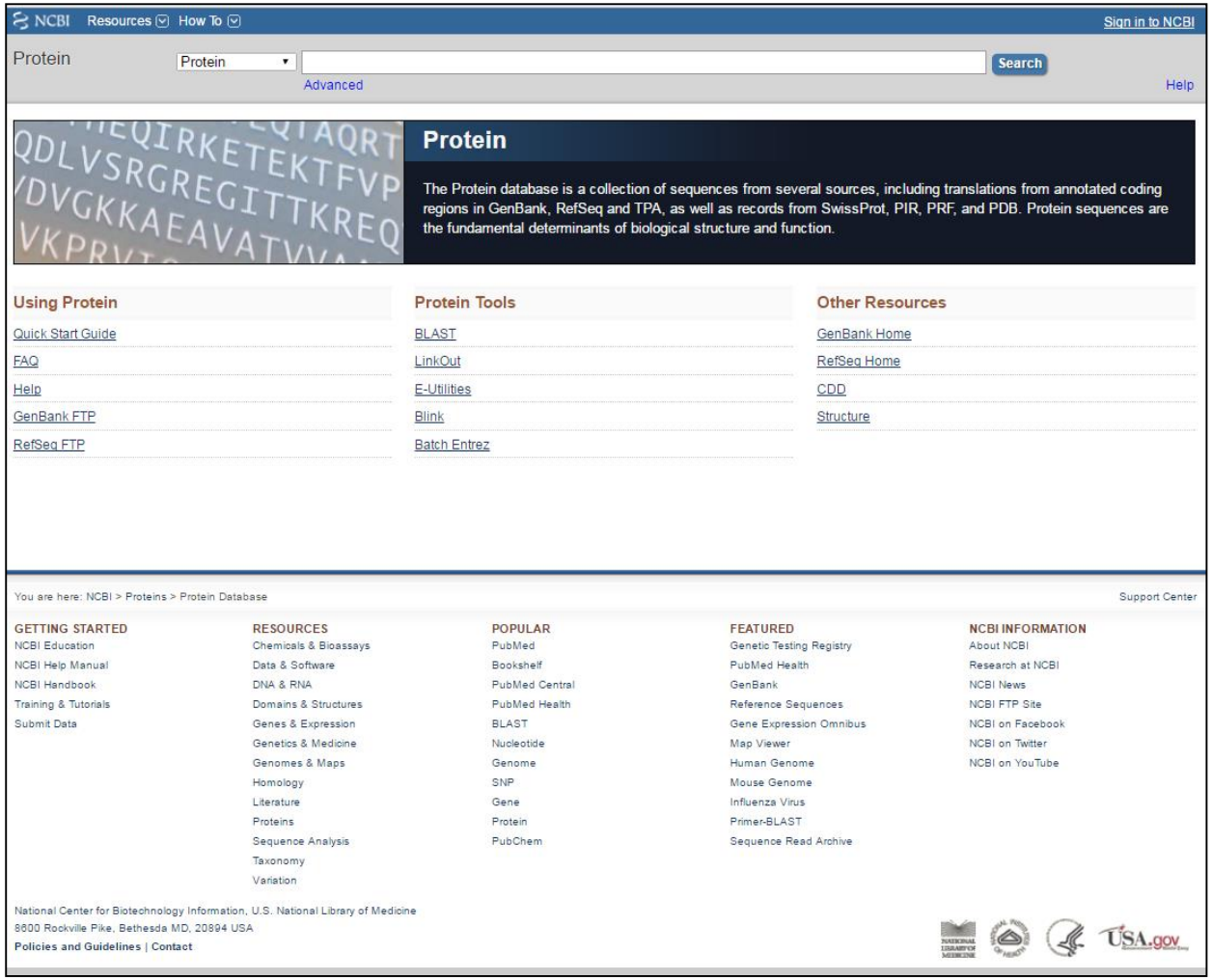
Steps of protein sequence analysis

BLAST search at GenBank



Steps of protein sequence analysis

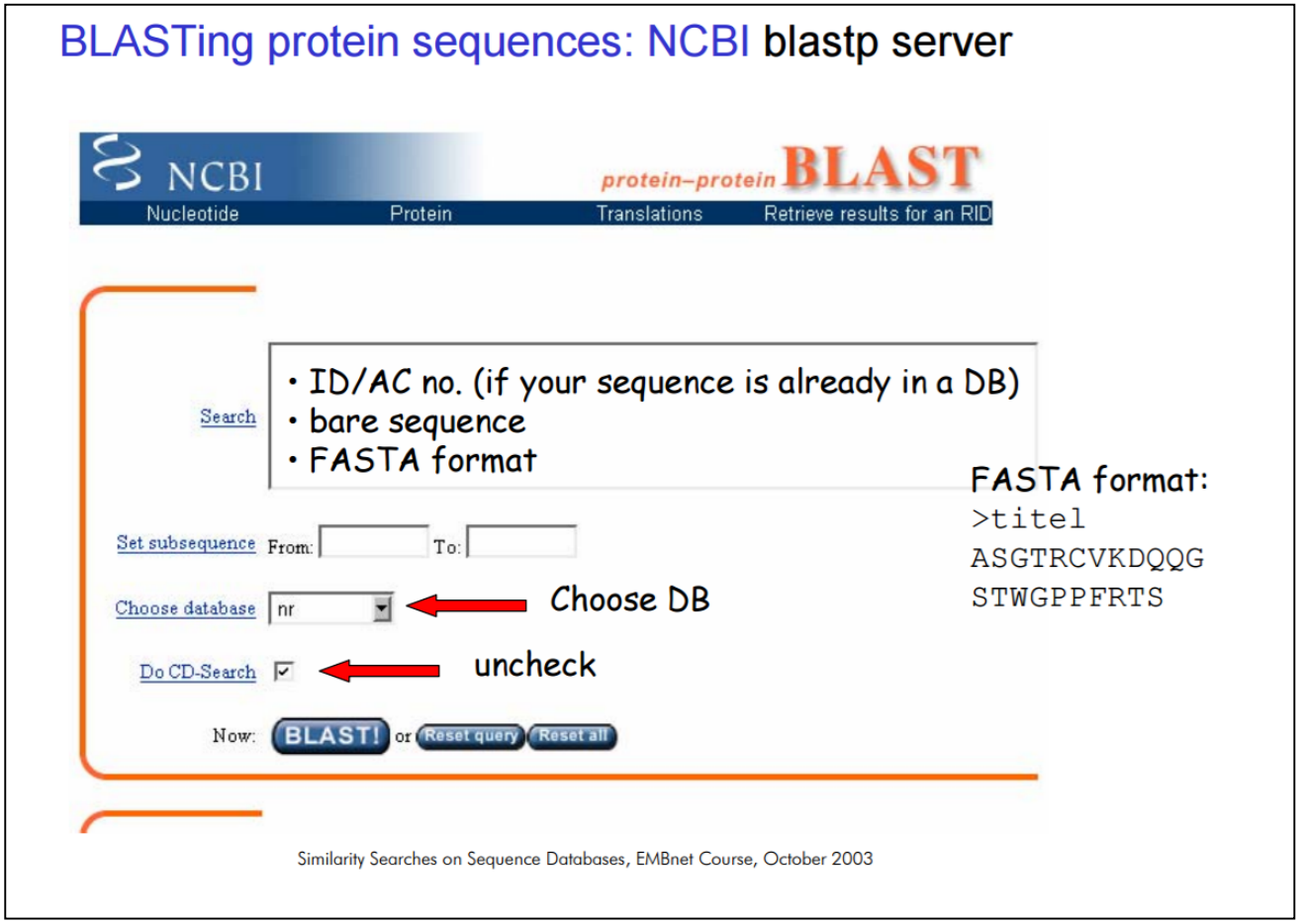
BLAST search at GenBank



Steps of protein sequence analysis

BLAST search at GenBank

blastp (blasting protein sequences)



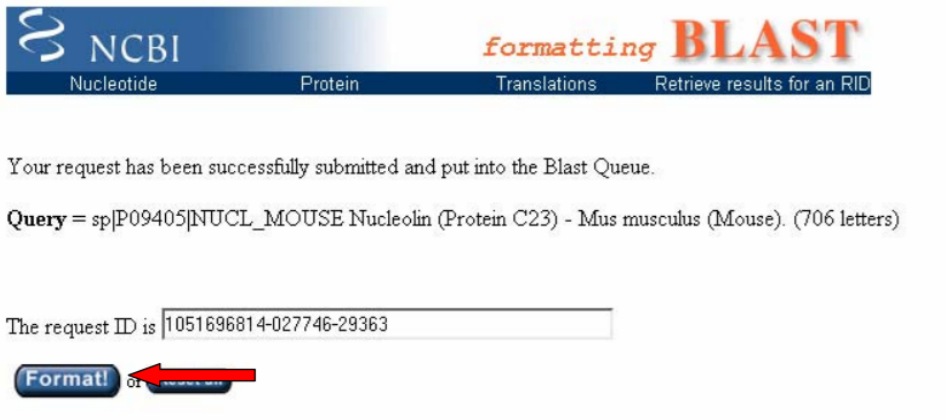
Bordoli,2003

Steps of protein sequence analysis

BLAST search at GenBank

blastp (blasting protein sequences)

BLASTing protein sequences: NCBI blastp server



If you get no reply, DO NOT resubmit the same query several times in a row - it will only make things worse for everybody (including you)!

Similarity Searches on Sequence Databases, EMBnet Course, October 2003

Bordoli,2003

Steps of protein sequence analysis

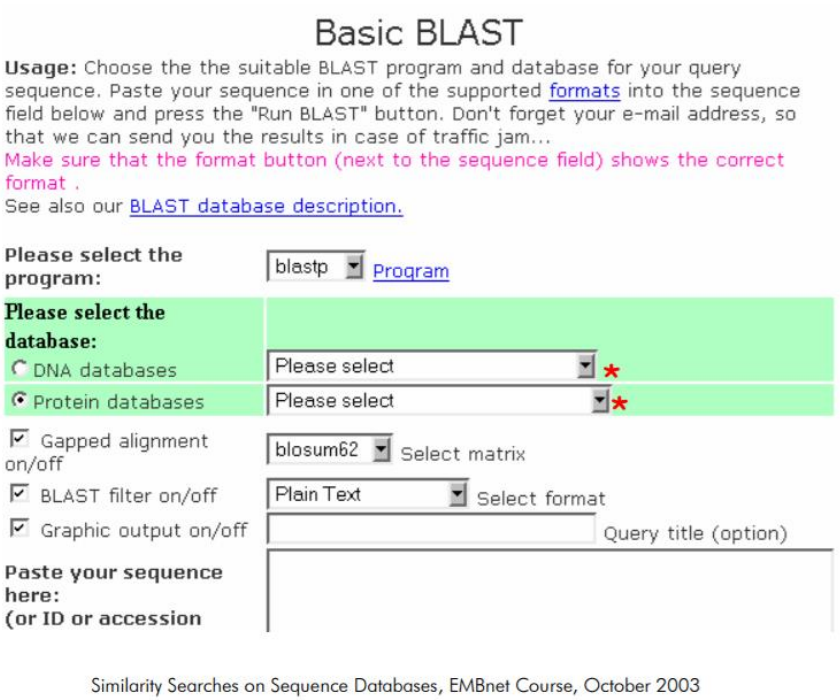
BLAST search at GenBank

blastp (blasting protein sequences)

BLASTing protein sequences: Swiss EMBnet blastp server

- URL: http://www.ch.embnet.org/software/bBLAST.html

The EMBnet interface gives you many more choices *:



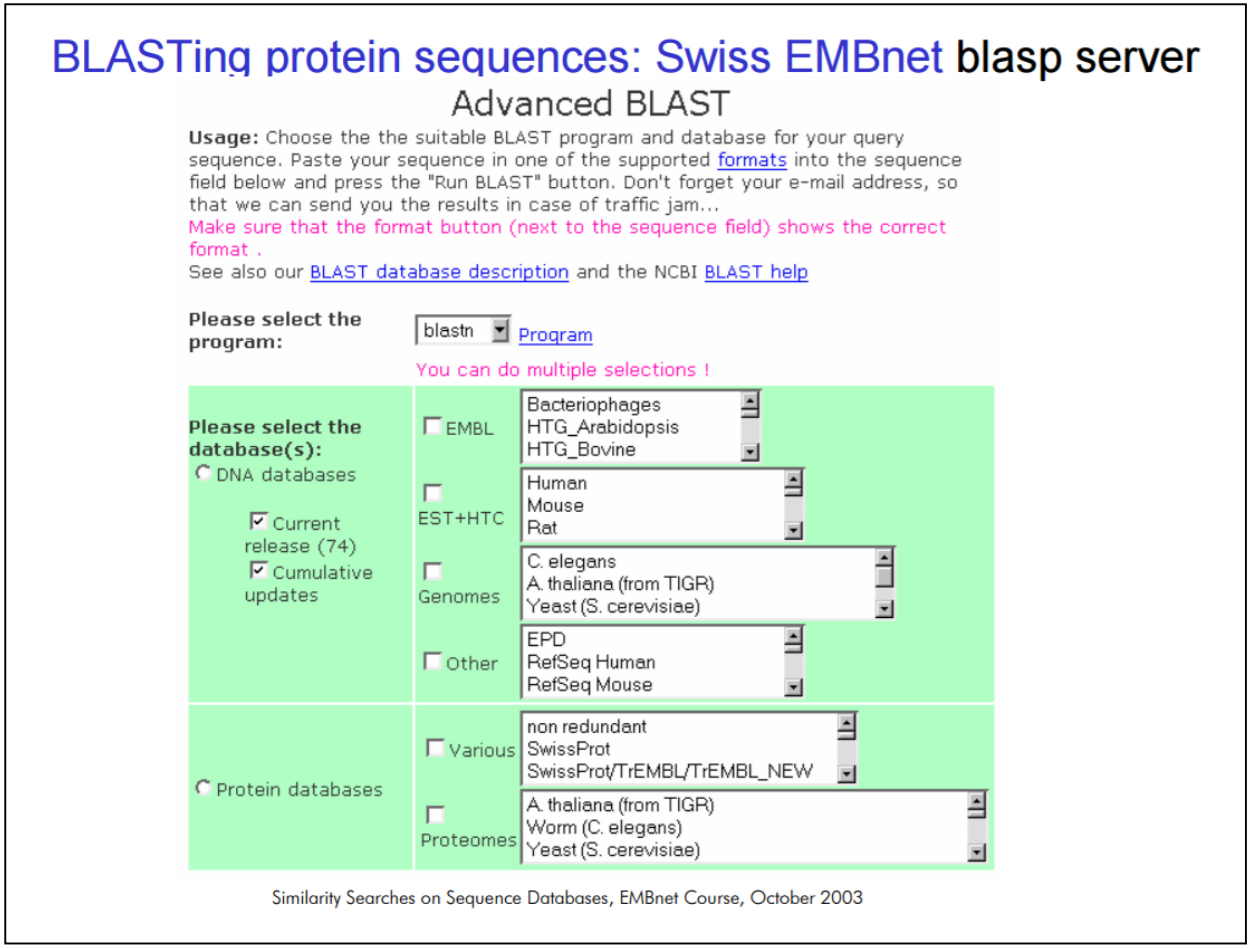
Similarity Searches on Sequence Databases, EMBnet Course, October 2003

Bordoli,2003

Steps of protein sequence analysis

BLAST search at GenBank

blastp (blasting protein sequences)



Bordoli,2003

Steps of protein sequence analysis
BLAST search at GenBank

- Understanding your BLAST output:

1. Graphic display: shows you where your query is similar to other sequences.
2. Hit list: the name of sequences similar to your query, ranked by similarity.
3. The alignment: every alignment between your query and the reported hits.
4. The parameters: a list of the various parameters used for the search

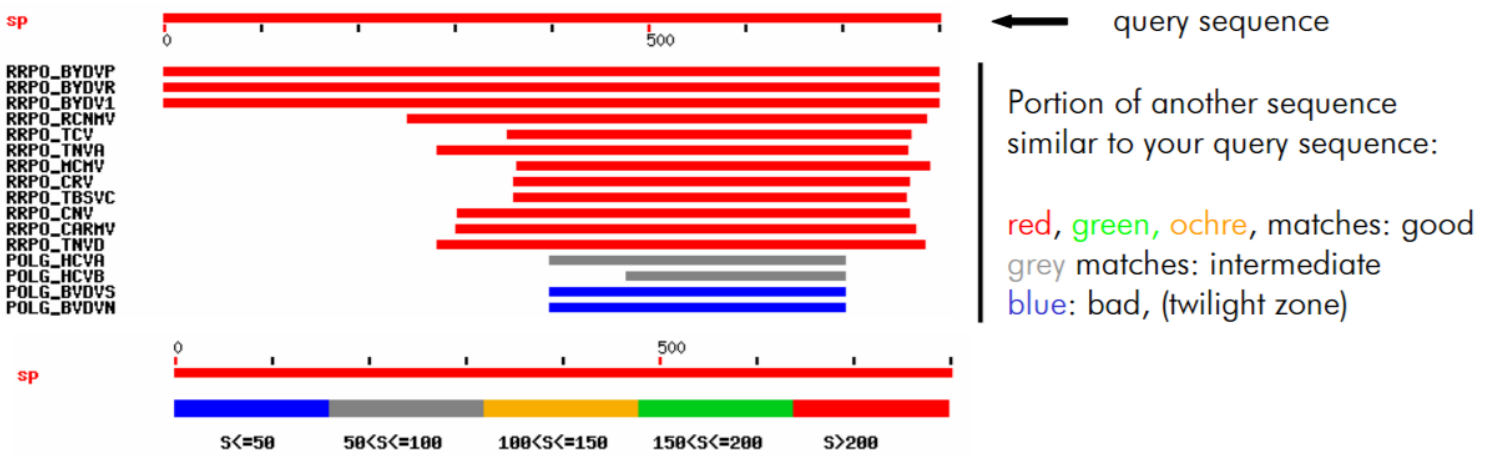
Bordoli,2003

Interpreting protein blast results

BLAST search at GenBank

blastp (blasting protein sequences)

Understanding your BLAST output: 1. Graphic display



The display can help you see that some matches do not extend over the entire length of your sequence => useful tool to discover domains.

Bordoli,2003

Interpreting nucleotide blast results

Reading BLAST Output

Example 3: Global alignment of two random 300 residue amino acid sequences

>Random seq 1

EMQSSVHIKTADYYITYFGHHFIVGEWLPNIRFPYFFWIITTDARNDMAIFNCQDETQSKKPSYNSDANNNYQYMW
GCDLQEAKVTAMGNLHLWNHRGPRFQKDHACQLCEPHRGITETKRQKIDCSMNPHIPHARKHYRGLNYMYMAE
NMRFIELPTEEEFWSWWDWVSWRMEMWGSDLMPEQYMRMDSWENSEQCRKHSIGRCLHYRHLNLWDDRFA
QVSCFNMWWEIFPIGQRHDGYLYRVRESMIQNQDENTVCPAMFAANWQLLKEHHVRGSKEYREWFFINVWHTE
GSRVAQAH

>Random seq 2

SYLTKSAIQEVCLCKVNVNDMNRFAVLGPYGFMSKGWCSVQPFHIYVPGAKKGWMMQRQCETDDLMDMTSQE
DHEQYGGKRCPKCTHLLDRIAHKMAPDKMSRWGGEKQGEMFVYGDYQKYRQHKWCVHSLEHPYWNNWFALW
GQCCGKQMTNPMIYRKCAKTKTCMDQAPVPSLQCQVCLCHNGSTYLTPANCCDCQVEQHESGNMGGRWIRYQ
MFCVFLWKAITPKAPHFSTASKRQNKLRVQEEALQHYYNKGPLQIWPDDGWFMNRWHIILQCWYMGKFWRLH
MKCNARESEMVML

>lcl|31039 unnamed protein product
Length=300

28% similarity

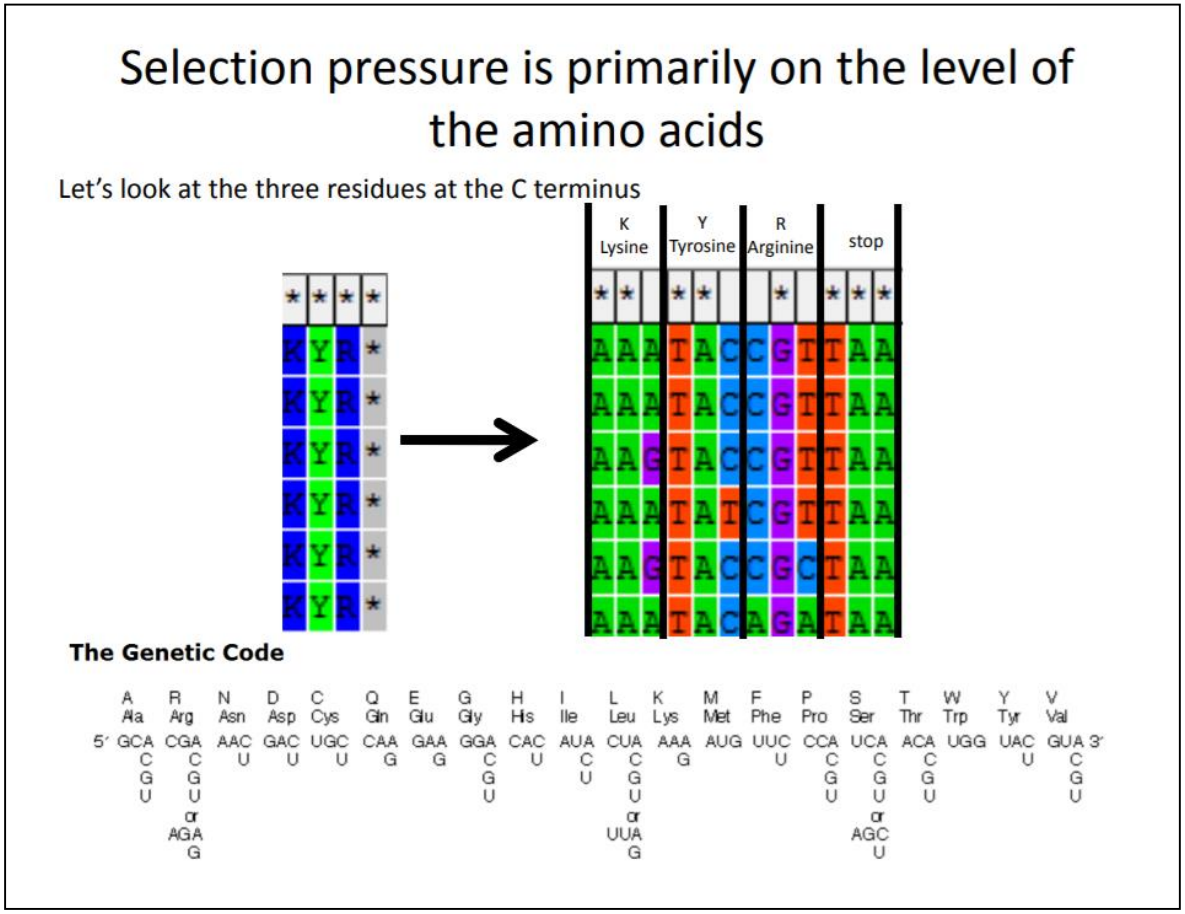
only 12% identity

```
 NW Score = -118
 Identities = 46/354 (12%), Positives = 100/354 (28%), Gaps = 108/354 (30%)

Query  1    EMQSS----------VHIKTADYYITYFGHHFIVGEWLPNIRFPYFF------WIITTDA  44
            +          V++   + +     + F+   W     F  +      W++
Sbjct  1    SYLTKSAIQEVCLCKVNVNDMNRFAVLGPYGFMSKGWCSVQPFHIYVPGAKKGWMMQRQC  60

Query  45   RND--MAIFNCQDETQ---SKKPSYNSDANNNYQYM-------WGCDLQEAKVTAMGNLH  92
              D  M + + +D  Q     + P       +    M       WG + Q  ++  G+
Sbjct  61   ETDDLMDMTSQEDHEQYGGKRCPKCTHLLDRIAHKMAPDKMSRWGGEKQ-GEMFVYGDYQ  119

Query  93   ---------------LWNHRGPRFQKDHACQLCEP--HRGITETKRQKIDCSMNPHIPHA  135
                           WN+   + +    Q+  P  +R    +IK     C      +P
Sbjct  120  KYRQHKWCVHSLEHPYWNNWFALWGQCCGKQMTNPMIYRKCAKTKT----CMDQAPVPSL  175

Query  136  RKH----YRGLNYMYMAENMRFIELPTEEEFWSWWDWVSWRME---MWGSDLMPE--QYM  186
            +      + G  Y+  A N      ++     W+ ++M    +W + + P+    +
Sbjct  176  QCQVCLCHNGSTYLTPA-NCCDCQVEQHESGNMGGRWIRYQMFCVFLWKA-ITPKAPHFS  233

Query  187  RMDSWENSEQCRKHSIGRCLHYRHLNLWDDRFAQVSCFNMWWEIFPIGQRHDGYLYRVRE  246
              +N  + ++ ++      +   L +W D                    DG+
Sbjct  234  TASKRQNKLRVQEEALQHYYNKGPLQIWPD--------------------DGW------  266

Query  247  SMIQNQDENTVCPAMFAANWQLLKEHHVRGSKEYREWFFINVWHTEGSRVAQAH  300
                        F   W ++ +    G K +R     N    +E   +
Sbjct  267  ----------FMNRWHIILQCWYMG-KFWRLHMKCNARESEMVML      300
```

For >100 aa, >25% identity is required to say with almost certainty that an alignment is not the result of chance

Tadmor,2010

Interpreting nucleotide blast results
Reading BLAST Output



Tadmor,2010

Interpreting protein blast results

BLAST search at GenBank

blastp (blasting protein sequences)

Understanding your BLAST output: 2. Hit list

```
                                                                  Score    E
Sequences producing significant alignments:                       (bits) Value

sp|P09505|RRPO_BYDVP Putative RNA-directed RNA polymerase (EC 2....  1652   0.0
sp|P29045|RRPO_BYDVR Putative RNA-directed RNA polymerase (EC 2....  1635   0.0
sp|P29044|RRPO_BYDV1 Putative RNA-directed RNA polymerase (EC 2....  1625   0.0
sp|P22956|RRPO_RCNMV Putative RNA-directed RNA polymerase (EC 2....   367   e-101
sp|P17460|RRPO_TCV Probable RNA-directed RNA polymerase (EC 2.7....   286   1e-76
sp|P22958|RRPO_TNVA RNA-directed RNA polymerase (EC 2.7.7.48) [C...   280   1e-74
```

Sequence ac number and name | Description | Bit score | E-value

- Sequence ac number and name: Hyperlink to the database entry: useful annotations
- Description: better to check the full annotation

- Bit score (normalized score) : A measure of the similarity between the two sequences: the higher the better (matches below 50 bits are very unreliable)

- E-value: The lower the E-value, the better. Sequences identical to the query have an E-value of 0. Matches above 0.001 are often close to the twilight zone. As a rule-of-thumb an E-value above 10-4 (0.0001) is not necessarily interesting. If you want to be certain of the homology, your E-value must be lower than 10^{-4}

Similarity Searches on Sequence Databases, EMBnet Course, October 2003

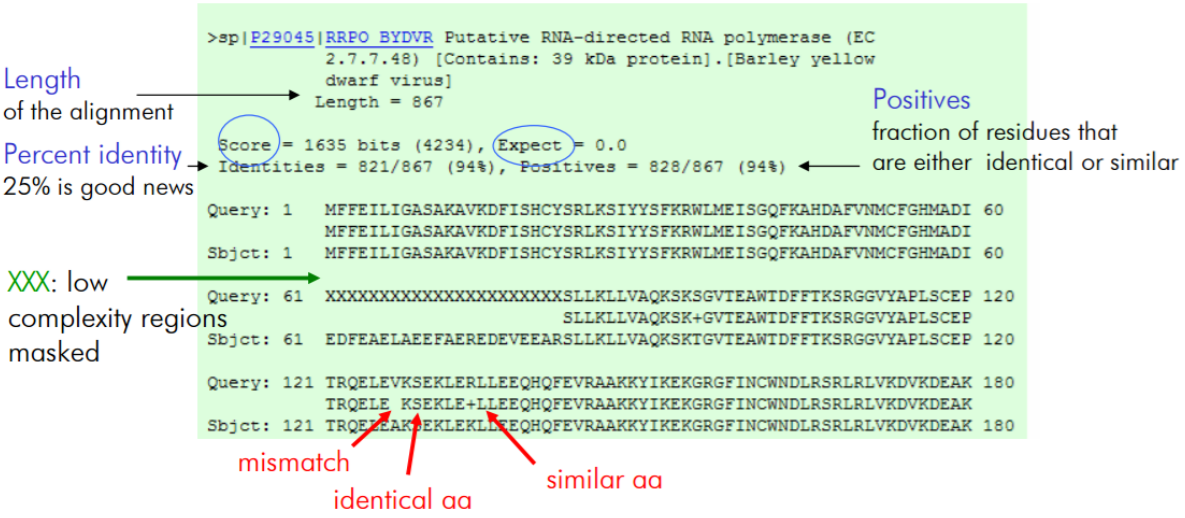
Bordoli,2003

Interpreting protein blast results

BLAST search at GenBank

blastp (blasting protein sequences)

Understanding your BLAST output: 3. Alignment



A good alignment should not contain too many gaps and should have a few patches of high similarity, rather than isolated identical residues spread here and there

Similarity Searches on Sequence Databases, EMBnet Course, October 2003

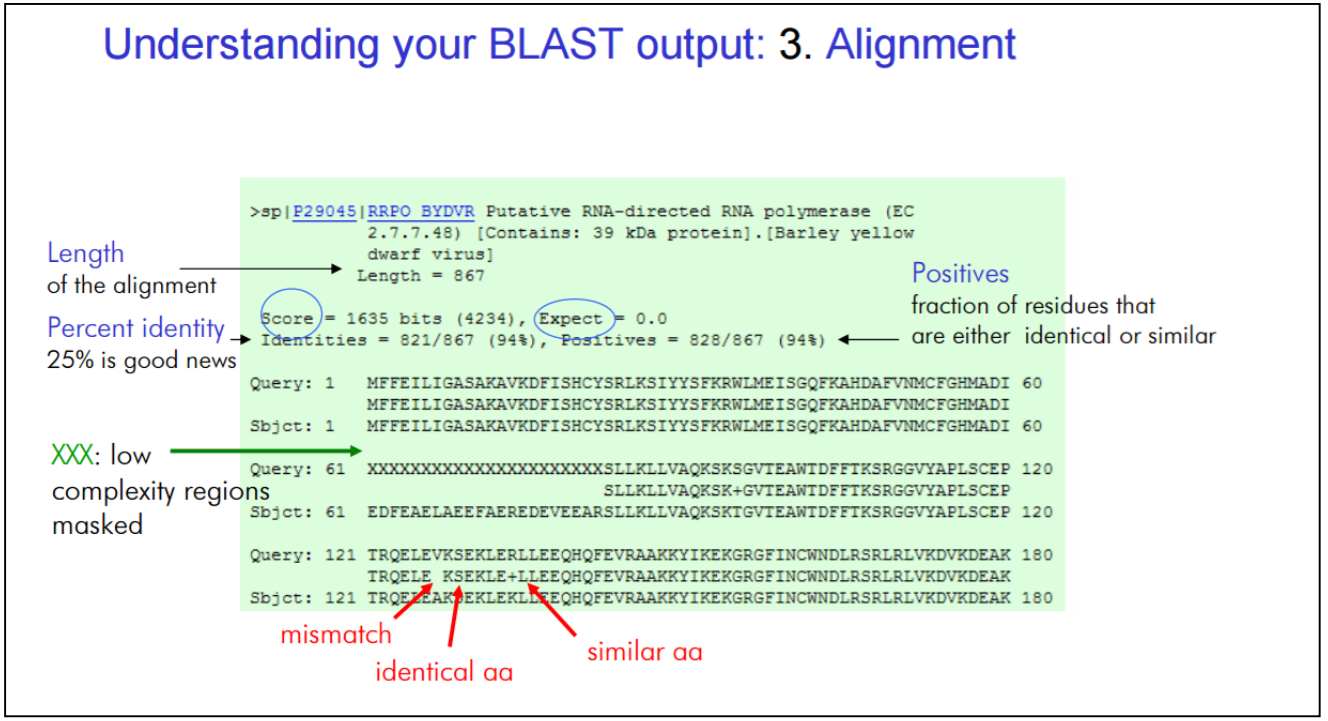
Bordoli,2003

Interpreting protein blast results

BLAST search at GenBank

blastp (blasting protein sequences)

- A good alignment should not contain too many gaps and should have a few patches of high similarity, rather than isolated identical residues spread here and there.



Bordoli,2003

Alignments

- Finally, the third data format provides a base-to-base alignment of the query sequence with each of its matches, starting from the "best match" and progressing down the list.
- This format repeats the alignment score and E value given above, but also has a measurement called "Identities".
- The identity give the number of exact matches between the your sequence and the sequence from GenBank and is presented as a fraction (the number of positions in the sequences that matched divided by the number of positions that the two sequences had in common) and as a percent.

Moran,2010

Types of alignments

Global vs. Local alignments

Based on pairwise alignments

- Alignments can be global or local:

1. Local alignment algorithms: Finds the region (or regions) of highest similarity between two sequences and build the alignment outward from there.
2. Global alignment algorithms: Start at the beginning of two sequences and add gaps to each until the end of one is reached.

- **Global alignment** is useful when you want to force two sequences to align over their entire length.

Programs used in global and Local alignments

Pairwise alignment programs

- Local Alignment Programs:

1. FASTA
2. BLAST

- e.g. of Fasta program: LALIGN

- Global Alignment Programs:

1. Fasta program Align
2. Needle. Its improved version is called Stretcher.

Multiple alignments
Progressive pairwise alignments

- In theory, making an optimal alignment between two sequences is computationally straightforward, but aligning a large number of sequences using the same method is almost impossible.
- The problem increases exponentially with the number of sequences involved

  (the product of the sequence lengths)

Programs used in multiple alignments

Common multiple alignment programs

- Most of the available multiple alignment programs use some sort of incremental or progressive method that makes pairwise alignments, then adds new sequences one at a time to these aligned groups.
- CLUSTAL is the most popular alignment program that uses a progressive pairwise algorithm.
- PILEUP is the multiple alignment program in the GCG package.

BLAST

Makes local sequence alignments

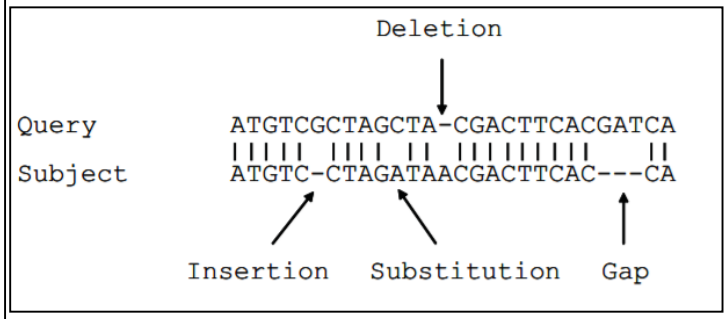
- BLAST is a local alignment algorithm.
  - It only aligns what can be aligned
  - It ignores the rest
- BLAST calculates local alignments, for databank searches and to find pairwise similarities local alignments are preferred (Pairwise Alignment).

Spiegel,2007;Peter Gogarten

BLAST
Local (pairwise)alignment

- BLAST programs compare a user-entered (query) sequence with subject sequences in a database.
- It scores the match depending on the sequence identity and the number of differences between sequences such as deletions, insertions, substitutions, and gaps.

```
and tRNA-Ile and tRNA-Ala genes, complete sequence
Length=498

 Score =  289 bits (156),  Expect = 1e-74
 Identities = 337/417 (80%), Gaps = 42/417 (10%)
 Strand=Plus/Plus

Query  28   CACACAGATTGTCTGATGAAGAAAACGAGCAGTAA--AACCT-TATAGGCTTGTAGCTCA  84
            ||||||||||||||||||| || | |||||||||  || || |  ||||||||||||||
Sbjct  28   CACACAGATTGTCTGATGAA-AATA-GAGCAGTAAGTAATCTCTGCAGGCTTGTAGCTCA  85

Query  85   GGTGGTTAGAGCGCACCCCTGATAAAGGTGAGGTCGGTGGTTCAAGTCCACTCAGGCCTA  144
            ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  86   GGTGGTTAGAGCGCACCCCTGATAAGGGTGAGGTCGGTGGTTCAAGTCCACTCAGGCCTA  145

Query  145  CCAA-TT-GCTTCCCATGCTGTGTTGCGTCTCCGCTCGCATACTG-ATGTATGCGTCGCG  201
            |||| || ||| | | |||| ||||| | |   ||||||||||  | ||||||| ||||
Sbjct  146  CCAAATTCGCT-CTCGTGCTTTGTTGTGGCAAAGCTCGCATACTTCA-GTATGCTTCGCT  203

Query  202  A-ATCCGCGCCTT-GCCTGGAAAGCCA-TTGATGTGGAAG-GTTTC-TAC--TGGGGCTA  254
              | || ||| |  ||| | ||| | | || | ||  ||| | ||| | |  ||||||||
Sbjct  204  TCA-CCACGC-TGCGCCCGAAAA-CGAATTAAGGTTAAAGAGATTCATTCGATGGGGCTA  260

Query  255  TAGCTCAGCTGGGAGAACGCCTGCTTTGCACGCAGGAGGTCTGCGGTTCGATCCCGCATA  314
            ||||||||||||||| ||||||||||||||||||| |||||||||||||||||||||||||
Sbjct  261  TAGCTCAGCTGGGAGAGCGCCTGCTTTGCACGCAGGCGGTCTGCGGTTCGATCCCGCATA  320

Query  315  GCTCCACCAT-TT-AC-G--AAT----GA---CTTCAGAGTGTAC-TGGCGACAGTATGC  361
            ||||||||||| || || | |  |    ||   |||||||||||| || || | || |||
```



Software analysis
Phylogenetic tree building at RDP

- Each bacterial sequence was subjected to software analysis (www.ebi.ac.uk and http://itol.embl.de/,....) to draw phylogenic tree.
- The Ribosomal Database Project (RDP) hosts a web site dedicated to the analysis of bacterial 16S rRNA sequences.
- You will use the tools at the RDP site to build a phylogenetic tree showing the relationship of your bacterium to some better-known and better-studied bacteria.
- The tree will group 16S rRNA genes according to sequence similarity, providing information on how your bacterium fits into the existing taxonomic framework for prokaryotes.

Moran,2010

CLUSTAL
Makes multiple sequence alignments

- The most commonly used algorithm is **CLUSTAL** and can be performed at several websites (e.g. http://www.ebi.ac.uk/clustraw/

1. Multiple sequence alignments allows you to identify nucleotides (nucleotide bases) with identity to other sequences.
2. It is also the first part of the process for generating phylogenies for deriving evolutionary relationships.
3. It is also a useful tool to use prior to performing diagnostic PCR primers, since it allows you to identify regions of sequence that have conservations (perhaps between members of closely related species.

CLUSTAL

Makes multiple sequence alignments

1. **Clustral** first calculates approximate pairwise similarity scores between all sequences to be aligned, and they are clustered into a dendrogram (tree structure).
2. Then the most similar pairs of sequences are aligned.
3. Averages (similar to consensus sequences) are calculated for the aligned pairs.
4. New sequences and clusters of sequences are added one by one, according to the branching order in the dendrogram.

CLUSTAL
Input and output formats

- Clustal requires that sequences be input as a single multi-sequence Fasta file.

1. Input format is a list of sequences in Fasta formats.
2. Output can be in many formats including:
3. GCG/MSF,
4. Clustal, and
5. Multi-Fasta.

- Ouputted to text and pasted into the alignment tool.

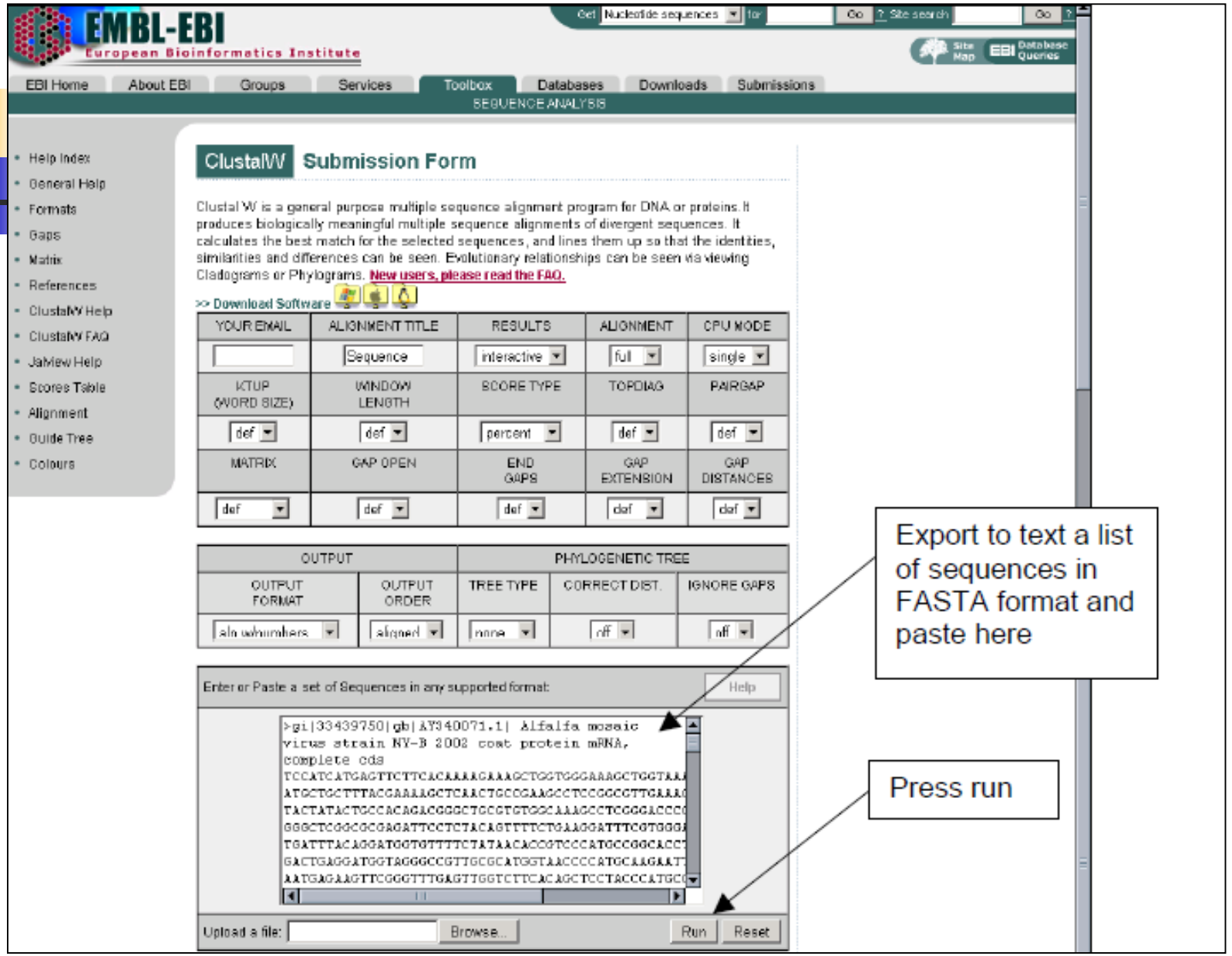
S.M. Brown;..

CLUSTAL
Input files

>PYCDA07TF input_file_1
ATGCCCATACTACTCTTCTGGTAGTTGGAATGAAGCCCAAAATATGATAAAACCTTTTCT
TACTAAAGTTTGTCAGGAAGTAGAAAGAATTGCTCATTGTGGAAAATGGGAAGAATGGAG
TGAATGTTCTACTACTTGT
>PYCDA08TR input_file_2
TATAGAAATAAAACTCCATTAAAAAATATTTTCCTTTTTCCTAATTATTTCTCTAAAATA
TAACAATCTAATTCATATAATATCATTACAATCACATATATATCTCTTTAAATTTTGTTC
CCTTTTTCCCTACGAGTTGTATCAGCAATAATCTCCTACAAGGTTAGACGTTGCTTCAAG
TTATTTTCAACAAATTTGGTCATTTTCAGCAAATTTTGCCATTTTCAGCAAATTTTGCCA
TTTTCAACAAATTTTGCCATTTTCAACAGATTTTGCCATTTTCAACAGATTTTGCCATTT
TCAGCAAATTTTGCCATATTCAACAAATTTTGCCATTTTCAGCAAATTTTACCATTTTTA
GCAAATTAGTATACCGTGTTAT
>PYCDA09TRB input_file_3
GCGGGAATATAGAAATAAAACTCCATTAAAAAATATAAACCTTTTTTTTAATTATCACCC
TAAAACATAACAATCTAATTCATATAATATCATTACAATCACATATATATCTCTTTAAAT
AATGATCCCTTTTTCCCTACGAGTTGTATCAGCAATAATCTCCTACAACGGATAGACGTT
GCTTCAAGTTCTTTTCAACAAATTGGGTCATTTTCAGCGAATTTTGCCATTTTCAGCAAA
TTTGGCCATACTCAACAAATTTTGCCATTGGCAACAGATTTTGCCATTTTCAACAGATTT
TGCCGTTGTCAGCAAATTTTGCCATATTCAACAAATTTTGCCAATCTCAGCAAATTTTAC
CATTTTCAGCAAATTAGCATACCGTTTTATCATTATCGCCTTCACTAATG

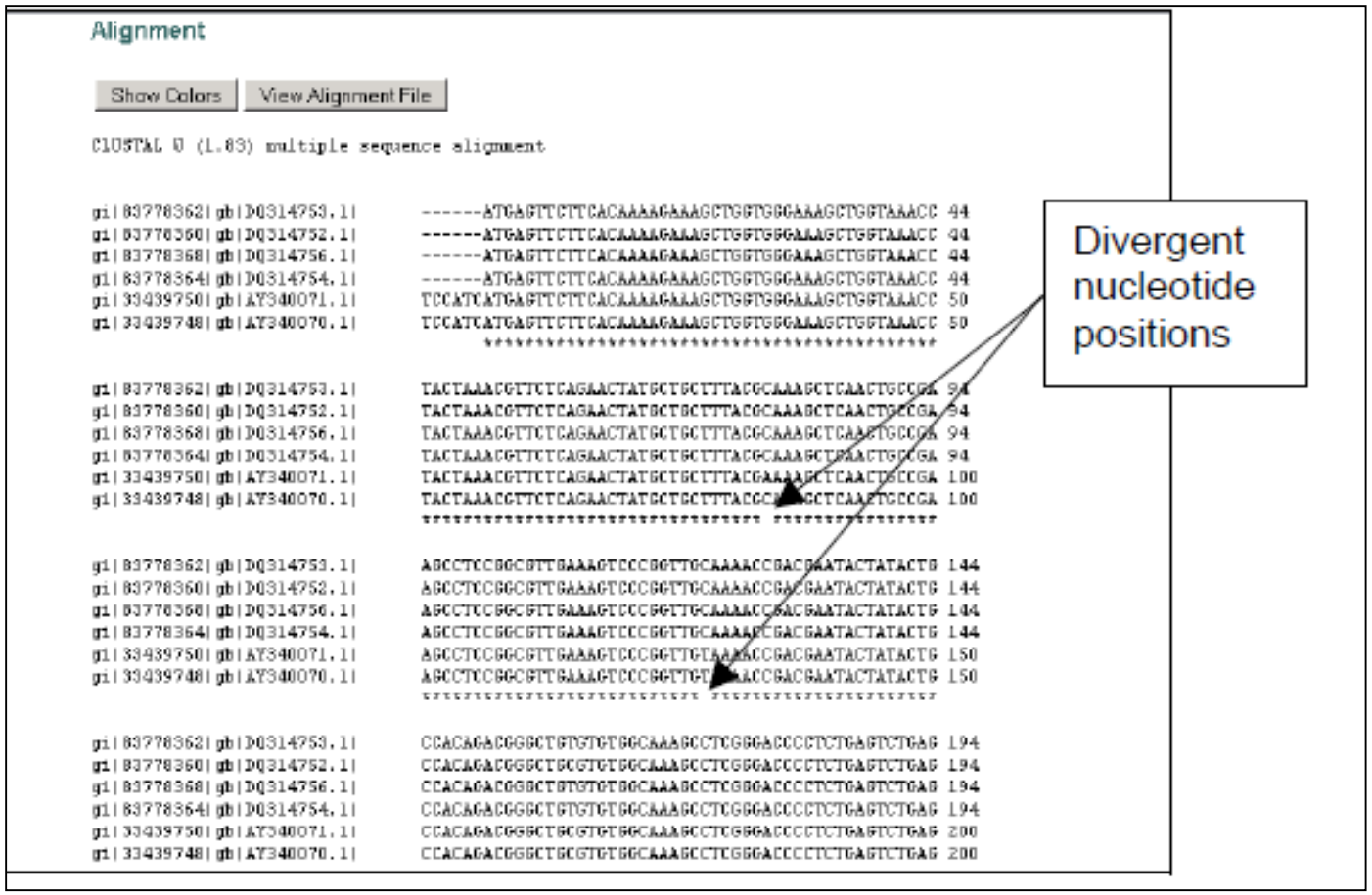
S.M. Brown

CLUSTAL: Submission form



The results of alignment

Allows to identify conserved and divergent sequence regions



Editing Multiple Alignments

- There are a variety of tools that can be used to modify a multiple alignment.
- These programs can be very useful in formatting and annotating an alignment for publication.
- An editor can also be used to make modifications by hand to improve biologically significant regions in a multiple alignment created by one of the automated alignment programs.

Editing Multiple Alignments
BOXSHADE

- Shades by similarity.
- The alignments can be highlighted using a program such as BOXSHADE on:

1. http://www.ch.embnet.org/software/BOX_form.html
2. http://bioweb.pasteur.fr/seqanal/interfaces/boxshade.html
3. http://huge.eng.uiowa.edu/~tscheetz/sequence-analysis/examples/BoxShade/BOX_form.html

Partial 16S rDNA sequence alignment *Xanthomonas* and *Stenotrophomonas* spp.

- Partial 16S rDNA sequence alignment of 13 *Xanthomonas* and *Stenotrophomonas* type-strains and seven *X. translucens* pv. *graminis* (X.t.g.) isolates.
- Shading indicates sequence differences to the X.t.g. type-strain.
- Bars mark the diagnostic PCR primer site characteristic for the X.t.g. group.
- Numbers on top denote position in the *E. coli* reference sequence.

Kölliker *et al.*,2006

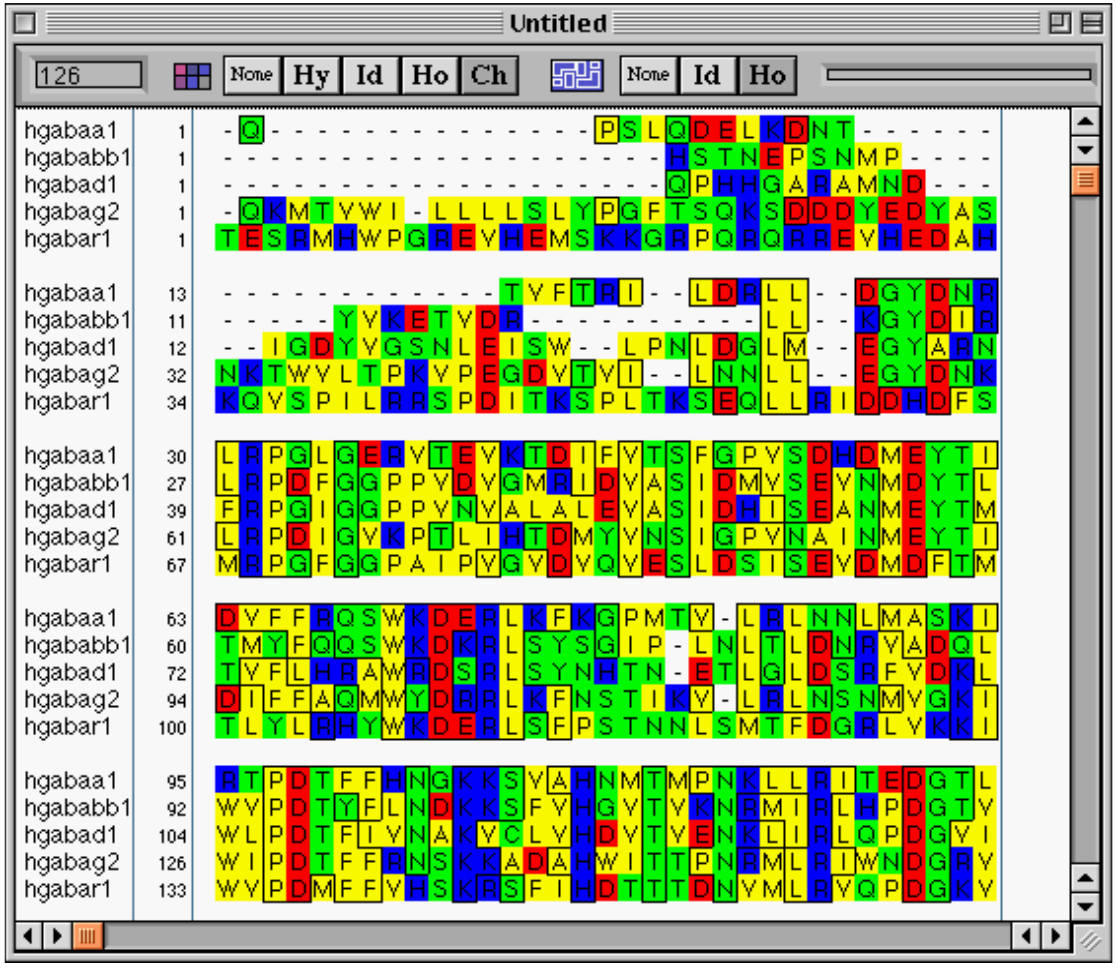
Editing multiple alignments
Other editors

- The MACAW and SeqVu program for Macintosh and GeneDoc and DCSE for PCs are free and provide excellent editor functionality.
- Many "comprehensive" molecular biology programs include multiple alignment functions:
- MacVector, OMIGA, Vector NTI, and GeneTool/PepTool all include a built-in version of CLUSTAL.

Editing Multiple Alignments

Other editors

SeqVu



Example of multiple sequence alignment

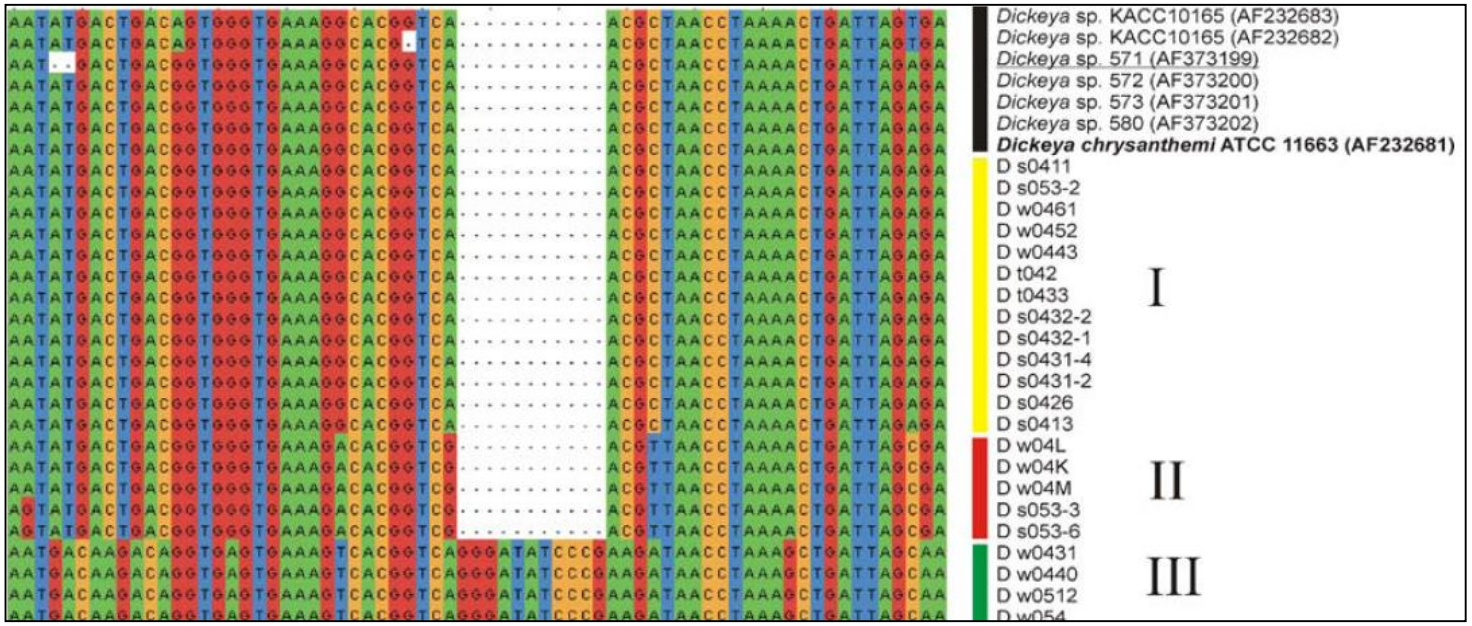
16S-23S rDNA intergenic spacer region of *Dickeya* strains

- Multiple sequence alignment of the most variable part of 16S-23S rDNA intergenic spacer region of *Dickeya* strains.
- Reference sequences obtained from GenBank are marked with a black bar, the *D. chrysanthemi* type strain is marked with bold and the reference strain obtained from potato is underlined.

Laurila *et al.*,2008

Multiple sequence alignment

16S-23S rDNA intergenic spacer region of *Dickeya* strains



See also bacterial diagnosis file-Part 1

Laurila *et al.*,2008

Software analysis
Phylogenetic Tree Building at RDP

- Use a browser to access the RDP analysis page (http://rdp.cme.msu.edu/html/analyses.html);
- Find the 'Phylip Interface' row and click on the 'run' triangle under the Small Subunit heading.
- You are now on the 'Start' page for the Phylip analysis tools.
- Brings your sequence into RDP and formats it for tree building……
- Choose the 'Distance Matrix' button from the top of the page. This takes you to a tool that will compare the sequences.
- Click on the 'Calculate Matrix' button to generates a table showing relatedness for all pairs of sequences, and these are the values that will be used to make the tree.
- Next click on 'Phylogenetic Tree'…..
- Your phylogenetic tree is generated in the main window.

Moran,2010

PHYLIP
Version 3.6

- It includes programs to carry out parsimony, distance matrix methods, maximum likelihood, and other methods on a variety of types of data, including DNA and RNA sequences, protein sequences, restriction sites, 0/1 discrete characters data, gene frequencies, continuous characters and distance matrices.
- It is the most widely-distributed phylogeny package, with over 20,000 registered users, some of them satisfied.
- It competes with PAUP* to be the program responsible for the most published trees.
- It has been distributed since October, 1980 and is now celebrating its 25th anniversary, as the oldest distributed phylogeny package.
- PHYLIP is distributed at the PHYLIP web site at http://evolution.gs.washington.edu/phylip.html.

See also bacterial phylogeny file

Taxonomic characterization of bacteria
PCR, DNA sequencing, sequence alignment

1. Sequence alignment of the 16S rRNA gene region of each isolate was performed using CLUSTALW option of the software Molecular Evolutionary Genetics Analysis (MEGA) software v4.0. (Tamura *et al.*, 2007).
   - The sequences were deposited at GenBank and were subjected to a NCBI BLAST search.
2. In a second alignment, the 16S rRNA gene region sequences were aligned using CLUSTALW together with the homologous sequences (retrieved from GenBank) of closely related species of bacteria.

Construction of a phylogenetic tree from the aligned sequences was created in MEGA software.

MEGA

Molecular evolutionary genetics analysis

- MEGA has been updated and expanded several times and currently all these versions are available from the MEGA website. The latest release, MEGA7, has been optimized for use on 64-bit computing systems.
- Tree-making methods:
- Neighbor joining
- Minimum evolution (ME)method
- UPGMA
- Maximum parsimony
- Maximum likelihood
- Bootstrap phylogeny test
- Confidence probability test
- Distance matrix viewer

The **minimum-evolution** (ME) method of phylogenetic inference is based on the assumption that the tree with the smallest sum of branch length estimates is most likely to be the true one. Maximum parsimony and minimum evolution are methods that try to minimize branch lengths by either minimizing distance (minimum evolution) or minimizing the number of mutations (maximum parsimony). The major problem with these methods is that the fail to take into account many factors of sequence evolution (e.g. reversals, convergence, and homoplasy). Thus, the deeper the divergence times that more likely these methods will lead to erroneous or poorly supported groupings. ME method perform best in the group of distance-based methods, but they work much more slowly than NJ.

MEGA

Molecular evolutionary genetics analysis

Estimates phylogeny for sequence data

1. Estimates phylogeny for sequence data by computing matrix of pair-wise distances between OTUs (OTUs= aligned sequences in this case).
2. Phylogenetic relationships are then estimated from pair-wise distances estimates using various algorithms.
3. Assumes phylogenetic divergence parallels sequence divergence.
4. Many different types of distances available.
5. Distance based algorithms include:

a. UPGMA (and variants)
b. Fitch–Margoliash least squares
c. Minimum evolution
d. Neighbor-joining
e. Weighted neighbor-joining

MEGA

Estimates phylogeny for sequence data

Parsimony, likelihood and Distance-matrix methods

1. When we infer a tree from genetic data using parsimony we minimize the amount of change along the branches of the tree.
2. Similarly when we use the likelihood principle we minimize change conditional on a specific mutation model.
3. An alternative to likelihood or parsimony is an approach based on evolutionary distances between a pair of sequences where the distance is accounting for all unseen events, for example using similar mutation models as likelihood.

Beerli,2005

Evolutionary distances between a pair of sequences

Distance methods

- A variety of distance algorithms are available to calculate pairwise distance. for example Proportional (p) distances.
- Distance analysis compares two aligned sequences at a time, and builds a matrix of all possible sequence pairs.
- During each comparison, the number of changes (base substitutions and insertion/deletion events) are counted and presented as a proportion of the overall sequence length.
- These final estimates of the difference between all possible pairs of sequences are known as pairwise distances.

https://www.ncbi.nlm.nih.gov/Class/NAWBIS/Modules/Phylogenetics/phylo12.html

Evolutionary distances between a pair of sequences

Pairwise distance methods

- Distance matrices are used in phylogeny as non-parametric distance methods and were originally applied to phenetic data using a matrix of pairwise distances.
- Pairwise methods evaluate all pairs of sequences and transform the differences into a distance.
- Pairwise distance methods are not so popular anymore because the are outperformed by likelihood methods.
- Distance is often defined as the fraction of mismatches at aligned positions, with gaps either ignored or counted as mismatches.

Wikipedia,2018;Beerli,2005

Evolutionary distances between a pair of sequences

Pairwise distance methods

- Example of a problematic data sets for distance methods, the used distance is simply counting sites that are different between pairs.

Individual	Sequence
one	ATTAGC
two	ATTGGC
three	ATGGGC
four	GTGGGC

→

Individual	one	two	three	four
one	-	1	2	3
two		-	1	2
three			-	1
four				-

Beerli,2005

Evolutionary distances between a pair of sequences

Pairwise distance methods

- The diagram above summarizes the calculation of pairwise distances between the gene sequences for four hypothetical species.
  - The coefficients provide a simple summary of how similar (or different) each sequence is from the other.
  - Sequence I and II are more alike to each other than either is to III.
  - In this example, we calculated the distances across the length of the whole sequence (10 bases); distances can be calculated for different sections of a sequence to see if some parts are more conserved than others.
  - Such a comparison would combine the pairwise distance approach with the rule of functional constraint.
- The above example illustrates the initial calculation of the distance. Most investigators would use additional statistical techniques to correct for hidden differences (parallel and back-mutations) using a variety of evolutionary models.

Distance Methods

Proportional (*p*) Distance

	DNA Site									
Species	1	2	3	4	5	6	7	8	9	10
I	A	T	A	T	A	C	G	T	A	T
II	A	T	G	T	A	C	G	T	A	T
III	G	T	A	-	A	C	G	T	G	C
IV	G	C	G	T	A	T	G	C	A	C

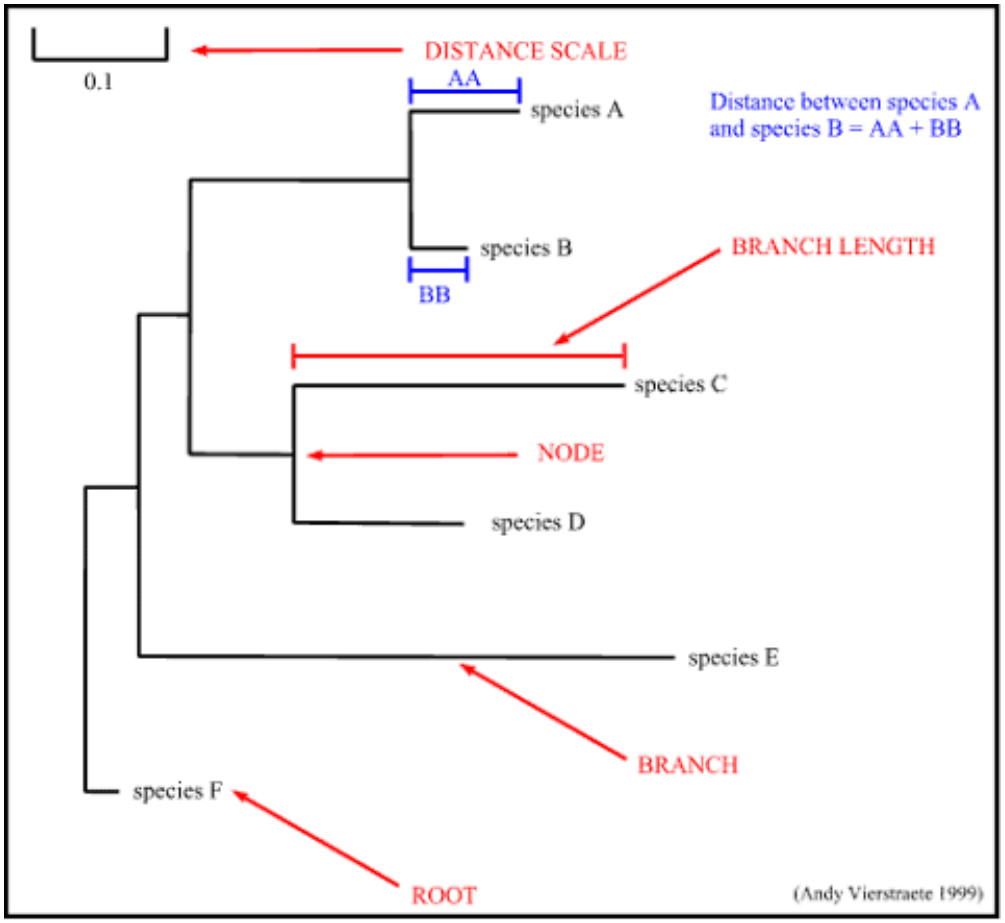
$$p = \frac{\text{\# differences}}{\text{\# sites}}$$

	I	II	III	IV
I	-	0.1	0.4	0.6
II		-	0.5	0.5
III			-	0.6
IV				-

https://www.ncbi.nlm.nih.gov/Class/NAWBIS/Modules/Phylogenetics/phylo12.html

Phylogenetic tree

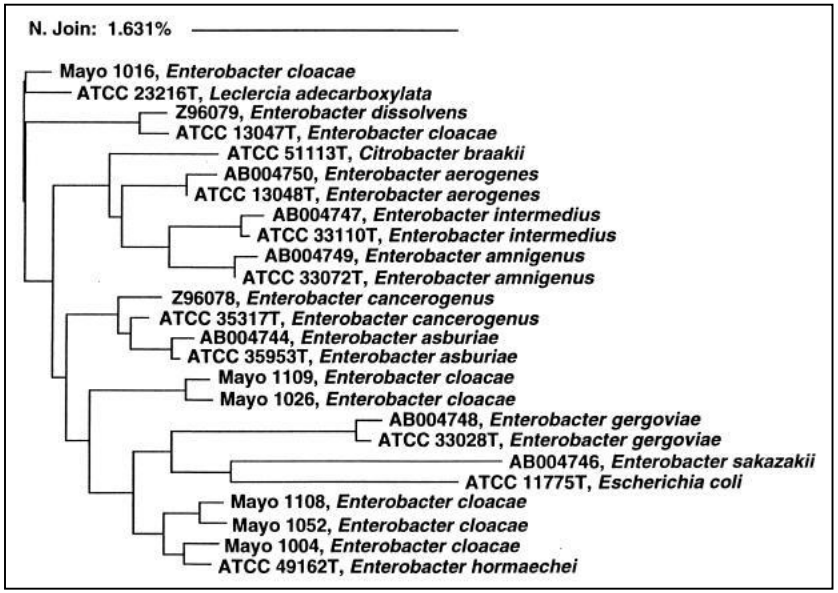
A typical rooted tree with scaled branches



Phycs498BIO Assignment2

Identification of *Enterobacter* spp.
Based on sequence analysis of different regions of the 16S rRNA gene

- Neighbor-joining analysis of DNA sequences from several *Enterobacter* spp.
- Phylogenetic analysis was based on full 16S rRNA gene sequences, and the scale reflects relative phylogenetic distance.
- Isolates with names beginning with Mayo were evaluated in this study.
- Isolates with names beginning with accession numbers were retrieved from GenBank.
- The remaining isolates, whose names begin with ATCC numbers, were type strains stored in the MicroSeq database.



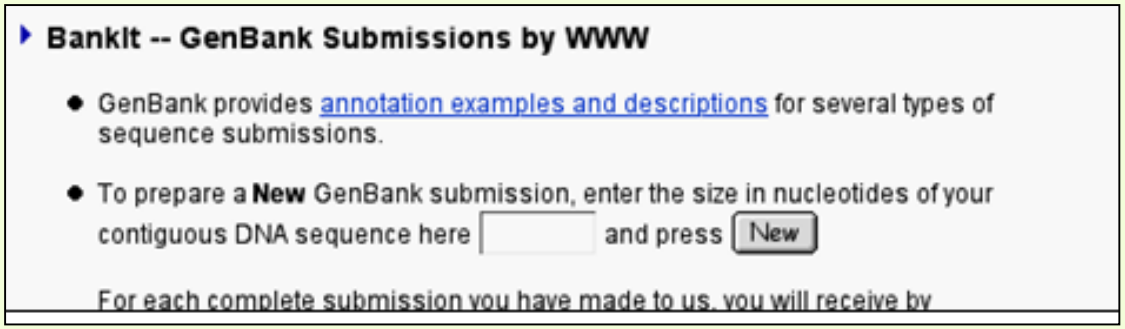
Second alignment with the homologous sequences (retrieved from GenBank) of closely related species of bacteria.

Tang *et al.*,1998

Sequence Submission to GenBank

Portion of the GenBank submission page showing how to enter the BankIt page

- By adding your new sequence to the existing databases, scientists that discover similar 16S rRNA sequences in the future will be able to find out about your bacterium, as well as the habitat and environmental conditions under which your bacterium was isolated.
- To submit your sequences use a web browser to go to the GenBank entry page at http://www.ncbi.nlm.nih.gov/BankIt.
- Scroll down the page until you see the section shown in next Fig. showing Portion of the GenBank submission page showing how to enter the BankIt page.
- Enter the number of nucleotides in your trimmed sequence in the box and click the 'New' button.



Moran,2010

Steps of sequence analysis

GenBank

Depositing sequences in GenBank

- Before submitting a request, remove any numerical digits in the query sequence or replace them with the appropriate letter codes (e.g., N for an unknown nucleotide residue or X for an unknown amino acid residue).

A	Adenine
G	Guanine
C	Cytosine
T	Thymine
U	Uracil
R	Purine (A or G)
Y	Pyrimidine (C or T)
Z	2-Aminoadenine
N	**Any nucleotide**
W	Weak (A or T)
S	Strong (G or C)
M	Amino (A or C)
K	Keto (G or T)
B	Not A (G or C or T)
H	Not G (A or C or T)
D	Not C (A or G or T)
V	Not T (A or G or C)

GenBank
Depositing sequences in GenBank

- GenBank is the National Institutes of Health (NIH) genetic sequence database, a collection of all publicly available DNA sequences.
- Scientists who determine the sequences of genes (or sometimes of entire genomes) deposit the sequence information in GenBank.
- Aside from doing this as a service to the scientific community, many journals require submission of new sequence data to GenBank as a precondition to publishing a manuscript referring to the sequence.

Moran,2010

Sequence Submission to GenBank
How do we get accession number?

- Accession number (bioinformatics), a unique identifier given to a biological polymer sequence (DNA, protein) when it is submitted to a sequence database.
- Accession number (library science), the unique number given to each new acquisition as it is entered in the catalog of a library or museum.

Sequence Submission to GenBank

Portion of the GenBank submission page showing how to enter the BankIt page

- The web page that appears next is a BankIt entry form that assigns a unique number to your entry.
- Scroll down through the form, entering information exactly as shown in the appendix.
- After you've finished filling in the BankIt form, click the "Validate and Continue" button at the bottom.
- A series of pages will appear asking you to check for possible errors.
- Ignore the warning message about not providing a recognized organism name (your isolate doesn't have a one); just scroll to the bottom of the form and click 'Validate and Continue' again.

Moran,2010

Sequence Submission to GenBank

Portion of the GenBank submission page showing how to enter the BankIt page

- When asked if you want to modify your submission regarding coding regions or other featuers, don't make any changes; select 'Submit to Genbank'.
- And when asked if you forgot to indicate a coding sequence interval in your submission, say you didn't (your gene doesn't code for a protein).
- Read these error pages carefully to make sure your submission isn't cancelled without you knowing it.

Moran,2010

Sequence Submission to GenBank

Portion of the GenBank submission page showing how to enter the BankIt page

- When you the get to the final "Thank you for using Bankit" page (see what it looks like in the Appendix), your submission is complete.
- GenBank will contact you by e-mail to:

1. Give you a copy of your submission,
2. Give you the GenBank submission number assigned to your sequence, and
3. Show you your final submission after GenBank staff have checked/corrected the format.

Moran,2010

Terminology

- **Addendum:** It includes article giving additional information on earlier published research paper of the author.
- **Basonym:** The original epithet of a name retained in a new combination.
- **Bioinformatics:** Computer based analysis of data on biological sequencing of the genome of an organism to predict gene function, protein and RNA structure, genome organization and molecular bases in relation to responses of plants to microbial pathogens.
- **Biovars:** Stains that are differentiated by biochemical or other serological means.
- **Chemotype** (= chemovar or chemoform): Chemically characterized portions of morphologically indistinguishable populations.
- **Chimeric sequences:** Chimeras are sequences formed from two or more biological sequences joined together. Amplicons with chimeric sequences can form during PCR. Chimeras are rare with shotgun sequencing, but are common in amplicon sequencing when closely related sequences are amplified.

Terminology

- **Chronogram:** An ultrametric tree or chronogram is a phylogenetic tree that explicitly represents evolutionary time through its branch lengths. The edge lengths represent time (so current taxa are equidistant from the root).
- **Cladogram:** branching diagram without branch length estimates. Cladogram displays branching information to explain the evolutionary relationships. The edge lengths do not represent anything.
- **Dendrogram:** a broad term for the diagrammatic representation of a phylogenetic tree.
- **Ecotype:** (not nomenclatural). A group of populations within a species genetically adapted to a particular ecological niche; sometimes used preceded by an hyphen to indicate a particular category of ecotype, e.g. agro-ecotype, geo-ecotype.
- The named ecotype (population adapted to a local habitat) from which sequence was obtained customarily applied to populations of *Arabidopsis thaliana*.

Terminology

- **Epithet:** The part of a scientific name identifying the species, variety, or other subunit within a genus. see specific epithet
- **Specific epithet:** The Latin or latinized noun or adjective that follows the genus name in a taxonomic binomial.
- **Holotype**: The single specimen designated as the type of a species by the original author at the time the species name and description was published. Isotype: A duplicate specimen of the holotype.
- **Holopathotype:** The single strain designated by the original author as the type strain of a newly described pathovar.

Terminology

- **Genospecies/genomospecies/genomic species**
- Groups that are phenotypically similar but genotypically different have been referred to as "genospecies," "genomospecies," or "genomic species."
- A genospecies has been defined in bacteriology as a species that can be discerned only by comparison of nucleic acids.
- If a specific genospecies cannot be differentiated from another genospecies on the basis of any known phenotypic trait, it should not be named until such a differentiating trait is found.
- Brenner *et al.*,2001, proposed that the term "genospecies" be replaced by "genomospecies."
- This would avoid confusion with the earlier definition of genospecies, which was a group of strains able to exchange genetic materials.
- The term "genomic species" is also in use: it is a group of strains with high DNA-DNA hybridization values.

Terminology

- **Genomospecies (genospecies):**
- The phylogenetic definition of a species (genomospecies) generally would include strains with approximatively 70% or greater DNA-DNA relatedness.
- Change in rank
- All strains clustered in these six genomic species(genomospecies 1-6) be assigned to the six species:
- *Dickeya zeae* sp. nov.
- *Dickeya dadantii* sp. nov.
- *Dickeya chrysanthemi* comb. nov.
- *Dickeya dieffenbachiae* sp. nov.
- *Dickeya dianthicola* sp. nov.
- *Dickeya paradisiaca* comb. nov., respectively.

Terminology

- **Genomovars (Multiple genetic types):**
- The term genomovar has been suggested to accommodate different DNA similarity groups within a nomenspecies.
- This term has been positively accepted by some taxonomists because it indicates that a genomic species is an integral part of a nomenspecies.
- It is suggested that genomovars encompassed in species should be numbered, and not named.
- Ultimately, genomovars could be given a formal name when a determinative phenotype is described.

Terminology

- **Genomovars (Multiple genetic types):**
- Genomovars are phenotypically similar, genotypically distinct groups of strains that share a low level of DNA hybridization and are considered distinct species, but are not given an official species name until differential diagnostic tests have been established.
- To date there are ten phylogenetically similar but genomically distinct species (genomovars) which are known as the *Burkholderia cepacia* complex.
- *B. cepacia* (genomovar I)
- *B. multivorans* (genomovar II)
- *B. cenocepacia* (genomovar III)
- *B. stabilis* (genomovar IV)
- *B. vietnamiensis* (genomovar V)
- *B. dolosa* (genomovar VI)
- *B. ambifaria* (genomovar VII)
- *B. anthina* (genomovar VIII)
- *B. pyrrocinia* (genomovar IX)
- *B. ubonensis* (genomovar IX)

Terminology

- **Isolate:** An isolate is a pure culture derived from a heterogeneous, wild population of microorganisms.
- **Isolate:** A living culture of bacterium, fungus, virus or other microorganism obtained from a single source of infection or site, and usually accorded a unique number; when shown to be distinctive, an isolate may be considered as a strain or accorded other infraspecific designation.
- **Lithotrophy:** 'lithos' (rock) and 'troph' (consumer), meaning "eaters of rock". Microorganisms that use inorganic compounds as electron donors to conserve energy for growth.
- **Lysotype (= lysovar):** Phagotype. Formerly used for infrasubspecific populations of bacteria based on reactions to bacteriophages; a phagovar.
- **Macroevolution:** Evolution that occurs above the level of the species.
- **Metagenomics:** the study of genetic material recovered directly from environmental samples. The broad field may also be referred to as env ironmental genomics, ecogenomics or community genomics (See the file microbial ecology).

Terminology

- **Microcosm:** Anything that is regarded as a world in miniature.e.g. bacterial microcosm.
- **Microevolution:** Evolution that occurs below the level of species.
- **Morphovar (morphotype):** A morphovar is a strain which is differentiated on the basis of morphological distinctions.
- **Monotypic:** Having only one immediately subordinate taxon; e.g. of a genus with a single validly published species.
- **Monotypic genus:** A genus for which a single binomial is validly published.
- **Morphovar (morphotype):** A morphovar is a strain which is differentiated on the basis of morphological distinctions.
- **Monotypic:** Having only one immediately subordinate taxon; e.g. of a genus with a single validly published species.
- **Monotypic genus:** A genus for which a single binomial is validly published.
- **Nomenclatural type:** The entity to which the name is permanently associated. The nomenclatural type is a strain for a species of subspecies, a species for a genus, a genus for higher taxa.

Terminology

- **Nomenspecies** as a group that bears a binomial name.
- **Neopathotype:** Proposed as the pathotype only when none of the strains that were used in preparing the original description are no longer extant.
- **Neotype:** Proposed as a type strain only when none of the strains that were used in preparing the original description are no longer extant.
- **Organotrophy:** The condition of an organism obtaining its energy from organic compounds.
- **Patent name:** A name cited only in patent application; such names are treated as if not effectively published.
- **Pathotype:** Equivalent to Race. Pathologically differentiated races of fungi or bacteria (= "pathodeme"). Pathotype strains that are unsuitable because they do not reflect the pathogenic or other characteristics given in the description of the pathovar (Young *et al.*,1991a).
- **Pathovarietas Nova:** Abbreviated pv. nov. and used to unequivocally indicate the proposal of a new pathovar.

Terminology

- **Phenospecies:** A group of bacteria whose phenotypic description is known and which is judged to have a rank corresponding to species; Compared with nomenspecies.
- **Phylogram or phylogenetic tree:** branching diagram with branch length estimates. A phylogram has additional information: the length of branches according to the amount of changes (evolutionary process. Edge lengths represent time or genetic distance.
- **Phylotype:** A group that is part of the same phylogenetic cluster based on the analysis of three loci/Determination uses 16S-23S ITS sequence.
- **Proposal:** A suggested addition or amendment to the Code or a proposal for the conservation or rejection of a name.
- **Race:** All genotypes of a pathogen that carry the same set of avirulence genes. Shows levels of pathogenic specialization.
- **A ribosomal binding site (RBS):** is a sequence on mRNA that is bound by the ribosome when initiating protein translation.
- **Rejected Name:** A name which must not be used to designate a taxon, listed by the Judicial Commission of the ISCP for various reasons generally related to ambiguity and confusion.

Terminology

- **Serovar (serotype):** A serovar is a stain differentiated by serological means.
- **Sequevar:** A group of strains with identical or almost identical DNA sequence in three loci. E.g. in *R. solanacerum* endoglucanase gene sequence was analyzed to determine its sequevars.
- **Sequence type:** A group with the same allelic profile based on a MLST analysis. An allele is a variant of the DNA sequence at a given locus.
- **Species:** A bacterial species is "a population of cells with similar characteristics." In practice, strains that share certain type properties can be called the same species even if they differ by up to 30% in DNA variation/ molecularly two prokaryotes whose 16S rRNA sequences are greater than 97% identical are likely to be the same species.

Terminology

- **Strain:** A strain is "a population of organisms that descends from a single organism (arisen from a single cell) or pure culture isolate.
- A group of bacteria derived from a single cell is called a strain;
- Closely related strains constitute a bacterial species.
- Strains within a species may differ slightly from one another in many ways.
- Different isolates may be same species but are different strains; often have slight differences.
- **Lab strains** may have a defined genetic difference.
- Strain is equivalent to a clone.
- The taxonomic term for any object to be classified is operational taxonomic unit (OTUs) and for most bacteriological work the OTUs will be strains.
- The term strain is also applicable to eucaryotic microorganisms , as well as to viruses.

Terminology

- **Translational Taxonomy**
- Translational Taxonomy of Bacterial Plant Pathogens, Carolee Bull,2017:
- The taxonomy of bacterial plant pathogens is complicated by the microscopic nature of bacteria, misconceptions concerning host specificity, changes in the rules of nomenclature, a lack of access to reference materials and methods, and the iterative nature of classification.
- Furthermore, differences in bacterial plant pathogens important for disease management often occur at taxonomic levels below species or subspecies and may have an additional layer of codified nomenclature.
- We use the term translational taxonomy to refer to basic research in pathogen taxonomy that informs disease management tactics.
- Translational taxonomic approaches taken to develop crop rotation strategies, deploy host resistance, and strategize biological control research will be discussed.

Terminology

- **Type genus:** The generic name which is the nomenclatural type of a name in a rank between genus and family.
- **Type culture:** A strain designated as the nomenclatural type of a bacterial species or infraspecific taxon.
- **Type species:** The species that is the name-bearing type of a genus, or a taxon between the ranks of species and genus and which fixes the application of the name.
- **Type strain:** The living or permanently preserved culture which is the nomenclatural type of a species or infraspecific taxon. It is often the first strain isolated or best characterized. Kept in collections e.g., ATCC (American Type Culture Collection) maintains as frozen or freeze-dried stocks.

Selected References

- Anonymous, 1958. International Code of Nomenclature of Bacteria and Viruses. Iowa State College Press, Ames, Iowa, USA.
- Anonymous,1996. International Code of Nomenclature of Bacteria. Int. J. Syst. Bact. 16: 459-490.
- Balows, A., H. G. Trüper, M. Dworkin, and K.-H. Schleifer (Eds.) 1992. The Prokaryotes: A handbook on the biology of bacteria: ecophysiology, isolation, identification, applications. 2nd ed. Springer-Verlag. New York, NY.
- Barbieri, M. 1981. The Ribotype Theory on the Origin of Life. Journal of Theoretical Biology, 91, 545-601.
- Barbieri, M. 1985. The Semantic Theory of Evolution. Harwood Academic Publishers, London and New York.
- Bedone, S.T. Power point presentations and lecture notes.THE OHIO STATE UNIVERSITY.
- Bull, C.T., De Boer, S.H., Denny, T.P., Firrao, G., Fischer-Le Saux, M., Saddler, G.S., Scortichini, M., Stead, D.E., Takikawa, Y. and Young, J.M. 2006. Guidance for plant bacteriologists on rules for bacterial nomenclature: how to avoid common pitfalls. Plant Pathogenic Bacteria. 11th International Conference Proceedings. Edinburgh, Scotland, July 10-14, 2006.
- Bull, C.T., S.H. De Boer, T.P. Denny, G. Firrao, M. Fischer-Le Saux, G.S. Saddler, M. Scortichini, D.E. Stead and Y. Takikawa. 2008. DEMYSTIFYING THE NOMENCLATURE OF BACTERIAL PLANT PATHOGENS (Invited Review). Journal of Plant Pathology 90 (3), 403-417.(http://www.isppweb.org/about_tppb.asp).

Selected References

- Bull C.T., De Boer S.H., Denny T.P., Firrao G., Fischer-LeSaux M., Saddler G.S., Scortichini M., Stead D.E. and Takikawa Y., 2008. Demystifying Nomenclature of Bacterial Plant Pathogens. Journal of Plant Pathology 90: 403-417.
- Bull, C.T., S.H. De Boer, T.P. Denny, G. Firrao, M. Fischer-Le Saux, G.S. Saddler, M. Scortichini, D.E. Stead and Y. Takikawa. 2010. Comprehensive List of Names of Plant Pathogenic Bacteria, 1980-2007. Journal of Plant Pathology, 92 (3), 551-592 (LETTER TO THE EDITOR).
- Bull, C.T., S.H. De Boer, T.P. Denny, G. Firrao, M. Fischer-Le Saux, G.S. Saddler, M. Scortichini, D.E. Stead and Y. Takikawa. 2012. Comprehensive List of Names of Plant Pathogenic Bacteria (2008-2010). Journal of Plant Pathology 94 (1), 21-27.
- CAB International 2002. Plant Pathologist's Pocketbook (eds J.M. Waller, J.M. Lenné and S.J. Waller), 504 pp.
- Cohan, F.M.2002. What are bacterial species? Annu. Rev. Microbiol. 56:457-87.
- Cohan, F.M.2006. Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. Philos Trans R Soc Lond B Biol Sci. 361 (1475): 1985-1996.
- Cohan, F.M. and E.B. Perry. 2007. A systematics for Discovering the Fundamental Units of Bacterial Diversity. Current Biology 17, R373-R386.
- Cohen, O., Gophna, U. and T. Pupko. 2011. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. Mol. Biol. Evol. 28, 1481-1489.
- De Queiroz, K. and J. Gauthier. 1994. Toward a phylogenetic system of biological nomenclature. Tree 9: 27-31.

Selected References

- Dworkin, M., Stanley Falkow, Eugene Rosenberg, Karl-Heinz Schleifer, Erko Stackebrandt (Eds.).2006. The Prokaryotes: A Handbook on the Biology of Bacteria. 7 Volume -Set- Part 1-4. Springer Science+ Business Media, Inc.
- Dye, D.W., Bradbury, J.F., Goto, M., Hayward, A.C., Lelliot, R.A., Schroth, M.N. 1980. International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. Review of Plant Pathology 59, 153-168.
- Euzéby J. 2005. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. Validation list no. 106. International Journal of Systematic and Evolutionary Microbiology 55: 2235-2238.
- Euzéby, J.P. 1997. List of Bacterial Names with Standing in Nomenclature: a Folder Available on the Internet. International Journal of Systematic Bacteriology 47, 590-592; doi: 10.1099/00207713-47-2-590.
- Fuch *et al.*, 2001. Siderotyping – A powerful tool for the characterization of pyoverdines. Current Topics in Medicinal Chemistry 1: 31-57.
- Garrity, G.M. (Editor- in- chief) 2005. Bergey's Manual of Systematic Bacteriology, second ed. Vol. 2 (Parts A, B & C; Three-Volume Set). The first volume of the 2nd edition appeared in 2001.
- Brenner DJ, Krieg NR, Staley JT, Garrity GM (eds).2005. Bergey's Manual of Systematic Bacteriology, second edition, vol. 2 (The Proteobacteria), part B (The Gammaproteobacteria), Springer, New York, 2005, p. 1108.
- Gde Ley, J. The Proteobacteria: Ribosomal RNA Cistron Similarities and Bacterial Taxonomy.

Selected References

- Hacker, J. and Kaper, J.B. 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. 54:641-79.
- Holt, J. G. 1994. Bergey's Manual of Determinative Bacteriology. One vol. 9th edition, Lippincott Williams & Wilkins, 787 pages.
- Holt, J. G. (Editor-in-Chief) 1984-1989. Bergey's Manual of Systematic Bacteriology. 1st edition, 4 volumes, Williams & Wilkins, Baltimore, MD.
- Janda, J.M. and S. L. Abbott.2002. Bacterial Identification for Publication: When Is Enough Enough? J. Clin. Microbiol. 40(6): 1887-1891.
- Konstantinidis, K.T., A. Ramette, and J. M. Tiedje.2006. The bacterial species definition in the genomic era. Philos Trans R Soc Lond B Biol Sci. 361(1475): 1929-1940.
- Kopac,S. and F.M. Cohan.2011. A Theory-Based Pragmatism for Discovering and Classifying Newly Divergent Bacterial Species. In : S. M. Tibayrenc (ed.) Genetics and Evolution of Infectious Diseases. Pp.21-41. Elsevier Inc.
- Lapage, S.P., P.H.A. Sneath, E.F. Lessel, V.B.D. Skerman, H.P.R. Seeliger and W.A. Clark (1976 Revision). 1975. International Code of Nomenclature of Bacteria. American Society for Microbiology, Washington, DC, 180 pp.
- Lapage, S.P., Sneath P.H.A., Lessel, E.F., Skerman, V.B.D., Seeliger, H.P. R. and Clark, W.A. 1992. International Code of Nomenclature of Bacteria 1990 Revision Bacteriological code. American Society for Microbiology, Washington, DC, USA.
- Matzanke, F. Lecture Notes: Siderophores and Iron Metabolism - Structures, Functions, Role in Infection and Potential as a Novel Class of Antibiotics.47pp.

Selected References

- Microbial Taxonomy Last revised: Monday, March 31,2003. Reading: Ch. 17 in text: for further details, see Microbial Diversity class notes from N. C. State Univ.
- Murray, R.G.E. and Stackebrandt, E. 1995. Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described procaryotes. International Journal of Systematic Bacteriology 45: 186-187.
- Oren, A. and G. M. Garrity. 2021. Valid publication of the names of forty- two phyla of prokaryotes. Int. J. Syst. Evol. Microbiol. 2021;71:005056 DOI 10.1099/ijsem.0.005056.
- Parkinson, N. Identifying Relatedness Between Bacterial Plant Pathogens. Parkinson_MolID_CSL_1.pdf.
- Ramos, Juan-Luis and Roger, C. Levesque (Eds.) 2006. *Pseudomonas*. Volume 4, Springer. Printed in the Netherlands.
- Ray, T.S. 1990. Thoughts on the Synthesis of Life. 16 pp.
- Rescott *et al.*, 2005. Microbiology- Unit Four - Phylogeny of Bacteria, Prokaryotes: Bacterial Genetic Systems.
- Rossello. 2001. The species concept for prokaryotes. FEMS Microbiology Reviews 25(1): 39-67.
- Rosselló-Mora, R. 2005. Updating Prokaryotic Taxonomy. Journal of Bacteriology, September 187: 6255-6257.
- Parte, A.C. 2013. LPSN - list of prokaryotic names with standing in nomenclature. Nucleic Acids Research, 42, Issue D1, D613-D616; doi: 10.1093/nar/gkt1111.

Selected References

- Parte, A.C. 2018. LPSN — List of of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. International Journal of Systematic and Evolutionary Microbiology, 68, 1825-1829; doi: de10.1099/ijsem.0.002786.
- Palleroni N. J. 1984. Genus I. Pseudomonas Migula 1894, in Bergey's Manual of Systematic Bacteriology, Vol. 1, eds Krieg N. R., Holt J. G. (Baltimore, MD: Williams & Wilkins;), 141-199.
- Roves, 2002. MICROBIAL EVOLUTION and SYSTEMATICS.
- Ruggiero MA, Gordon DP, Orrell TM, Bailly N, Bourgoin T, Brusca RC, et al. 2015. A Higher Level Classification of All Living Organisms. PLoS ONE 10(4): e0119248. https://doi.org/10.1371/journal.pone.0119248.
- Ruggiero, M.A., D.P. Gordon, T.M. Orrell, N. Bailly, T. Bourgoin, R.C. Brusca, T. Cavalier-Smith, M.D. Guiry and P.M. Kirk.2015. Correction: A Higher Level Classification o f All Living Organisms. PLoS ONE 10(6): e0130114. doi:10.1371/journal. pone.0130114.
- Saddler G.S., 2005. International Committee on Systematics of Prokaryotes. Xth International (IUMS) Congress of Bacteriology and Applied Microbiology; Minutes of the meetings, 28 and 30 July 2002, Paris, France. International Journal of Systematic and Evolutionary Microbiology 55: 533-537.
- Salemi, A. and A-M. Vandamme. 2003. The Phylogenetic Handbook- A practical approach to DNA and protein phylogeny. Cambridge Uni. Press, 399 pp.

Selected References

- Schaad N.W., Postnikova E., Lacy G., M'Barck F. and Chang C.- J. 2004b. *Xylella fastidiosa* subspecies: *X. fastidiosa* subsp. *piercei*, subsp. nov., *X. fastidiosa* subsp. *multiplex*, subsp. nov., and *X. fastidiosa* subsp. *pauca* subsp. nov. - Erratum. Systematic and Applied Microbiology 27: 763.
- Schaad N.W., Postnikova E., Lacey G.H., Sechler A., Agarkova I., Stromberg P.E., Stromberg V.K. and Vidaver A.K. 2005. Reclassification of *Xanthomonas campestris* pv. *citri* (*ex* Hasse 1915) Dye 1978 forms A, B/C/D, and E as *X. smithii* subsp. *citri* (*ex* Hasse) sp. nov. nom. rev. comb. nov., *X. fuscans* subsp. *aurantifolii* (*ex* Gabriel 1989) sp. nov. nom. rev. comb. nov., and *X. alfalfae* subsp. *citrumelo*(*ex* Riker and Jones) Gabriel *et al.* 1989 sp. nov. nom. rev., *X. campestris* pv. *malvacearum* (*ex* Smith 1901) Dye 1978 as *X. smithii* subsp. *smithii* nov. comb. nom. nov., *X. campestris* pv. *alfalfae* (*ex* Riker and Jones, 1935) Dye 1978 as*X. alfalfae* subsp. *alfalfae* (*ex* Riker *et al.* 1935) sp. nov. nom. rev., and "var. fuscans" of *X. campestris* pv. *phaseoli* (*ex* Smith, 1897) Dye 1978 as *X. fuscans* subsp. *fuscans* sp. nov. Systematic and Applied Microbiology 28: 494-518.
- Schaad N.W., Postnikova E., Lacey G., Sechler A., Agarkova I., Stromberg P.E., Stromberg V.K. and Vidaver A.K. 2006. Emended classification of xanthomonad pathogens on citrus - Erratum. Systematic and Applied Microbiology 29: 690-695.
- Schalk, Isabelle J. 2006. *Pseudomonas*. Volume 4, Ramos, Juan-Luis and Roger, C. Levesque (Eds.).
- Schuster, M.L., Vidaver, A.K. and Mandel, M. 1968. A purple-pigment-producing bean wilt bacterium, *Corynebacterium flaccumfaciens* var. *violaceum*, n. var. Canadian Journal of Microbiology, 14(4), 423-427.

Selected References

- Skerman, V.B.D., McGowan V. and Sneath P.H.A. 1980. Approved lists of bacterial names. International Journal of Systematic Bacteriology 30: 225-420.
- Sneath, P.H. and Sokal, R.R. 1973. Numerical Taxonomy. San Fransisco, USA: Freeman & Co.
- Sneath, P.H.A. 1984. Bacterial nomenclature. In: Krieg N.R., Holt J.G. (eds). Bergey's Manual of Systematic Bacteriology, Vol.1, pp 19-23. Williams & Wilkins, Baltimore, MD, USA.
- Sneath, P.H.A.1992. International code of nomenclature of bacteria. American Society for Microbiology, Washington, D.C.
- Sneath, P.H.A. and Brenner, D.J. 1992. "Official" Nomenclature Lists. ASM News, 58: 175.
- Sneath, P.H.A. 2003. A short history of the bacteriological code. http:/www.the-icsp.org/Code-history.html.
- Sneath, P.H.A. 2005. The preparation of the approved lists of bacterial names. Int. J. Syst. Evol. Microbiol. 55: 2247-2249.
- Stackebrandt, E., Frederiksen, W., Garrity, G. M. & other authors (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol., 52, 1043-1047.
- Tindall, B.J., J. Sikorski, R. A. Smibert, and N.R. Krieg. 2000. Phenotypic Characterization and the Principles of Comparative Systematics. Methods for General and Molecular Microbiology, 3rd Edition. Editors: C. A. Reddy, Terry J. Beveridge, John A. Breznak, George Marzluf, Thomas M. Schmidt, Loren R. Snyder. Book ISBN or Item Number: 978-1-55581-223-2.

Selected References

- Tindall, B.J., Kämpfer, P., Euzéby, J.P., Oren, A. 2006. Valid publication of names of prokaryotes according to the rules of nomenclature: past history and current practice. International Journal of Systematic and Evolutionary Microbiology 56: 2715-2720.
- Tindall, B.J. 2008. Are the concepts of legitimate and illegitimate names necessary under the current International Code of Nomenclature of Bacteria? A proposal to make changes to the Code. International Journal of Systematic and Evolutionary Microbiology. 58: 1979-1986.
- Vandamme, P., B. Pot, M. Gillis, P. DE Vos, K. Kersters, and J. Swinges. 1996. Polyphasic Taxonomy, a Consensus Approach to Bacterial Systematics. Microbiological Reviews, 60: 407-438.
- Vanneste, J. L. 2000. Fire Blight:The Disease and its Causative Agent, *Erwinia amylovora*. CABI Publishing, New Zealand, 370 pp.
- Van Niel, C.B. 1946. The classification and natural relationships of bacteria. Cold Spring Harbor Symp 11, 285-301.
- Vidaver A.K. 1991. Nomenclatural revisions of plant pathogenic bacteria and list of names 1980-1988. Review of Plant Pathology 70: 211-221.
- Vinatzer, B.A., Bull, C.T. 2009. The impact of genomic approaches on our understanding of diversity and taxonomy of plant pathogenic bacteria. In: Jackson R.W. (ed). Plant Pathogenic Bacteria: Genomics and Molecular Biology, pp. 37-61. Caister Academic Press, Norfolk, UK.

Selected References

- Young, J.M., Dye D.W., Bradbury, J.F., Panagopoulos, C.G., Robbs, C.F., 1978. A proposed nomenclature and classification for plant pathogenic bacteria. New Zealand Journal of Agricultural Research 21: 153-177.
- Young, J.M., Takikawa, Y., Gardan, L., Stead, D.E. 1992. Changing concepts in the taxonomy of plant pathogenic bacteria. Annual Review of Phytopathology 30, 67-105.
- Young, J. M., Saddler, G.S., Takikawa, Y., De Boer, S.H., Vauterin, .L, Gardan, L., Gvozdyak, R.I., Stead, D.E. 1996. Names of plant pathogenic bacteria 1864-1995. ISPP Subcommittee on Taxonomy of Plant Pathogenic Bacteria. Review of Plant Pathology 75, 721-763.
- Young J.M., Kuykendall L.D., Martînez-Remero E., Kerr A., Sawada H., 2001a. A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie *et al*. 1998 as new combinations combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola*, and *R. vitis*. International Journal of Systematic and Evolutionary Microbiology 51: 89-103.

Selected References

- Young J.M., Bull C.T., De Boer S.H., Firrao G., Gardan L., Saddler G.E., Stead D.E., Takikawa Y., 2001c. International standards for naming pathovars of phytopathogenic bacteria. http://www.isppweb.org/about_tppb_naming.asp.
- Young J.M., Bull C.T., De Boer S.H., Firrao G., Gardan L.,Saddler G.E., Stead D.E., Takikawa Y., 2004a. Names of plant pathogenic bacteria published since 1995. Report of the Taxonomy of Bacterial Plant Pathogens Committee of the International Society of Plant Pathology. http://www.isppweb.org/names_bacterial_new2004. asp.
- Young, J.M., Watson, D.R.W., Dye, D.W. 2004b. Reconsideration of *Arthrobacter ilicis* (Mandel *et al.*,1961) Collins *et al.* 1982 as a plant-pathogenic species. Proposal to emend the authority and description of the species. Request for an Opinion. International Journal of Systematic and Evolutionary Microbiology 54:303-305.
- Young J.M., Bull C.T., De Boer S.H., Firrao G., Gardan L.,Saddler G.E., Stead D.E., Takikawa Y. 2005. Names of Plant Pathogenic Bacteria, 1864-2004. Report of the Taxonomy of Bacterial Plant Pathogens Committee of the International Society of Plant Pathology. http://www.isppweb.org/names_bacterial_new2004.asp.

Selected References

- Wiedenbeck,J. & F. M. Cohan. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev. 35: 957-976.
- Woese, C. R., Fox, G. E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., Lewis, B. J. & Stahl, D. 1975. Conservation of primary structure in 16S ribosomal RNA. Nature 254 (5495):83-86.
- Woese, C.R. and Fox, G.E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. U.S.A. 74 (11): 5088-90.
- Woese CR, Fox GE. 1977. The concept of cellular evolution. J. Mol. Evol. 1977 Sep 20;10 (1):1-6.
- Woese CR, Magrum LJ. and Fox GE.1978. Archaebacteria. J. Mol. Evol. 11:245-252.
- Woese CR, Gutell R, Gupta R, HF. Noller. 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiol Rev. Dec; 47 (4): 621-669.
- Woese C, Kandler O. and M. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87 (12): 4576-9.

Selected References

- Woese, C.R. 2002. On the evolution of cells. Proc Natl Acad Sci USA 99(13): 8742-7.
- Woese, C. R. 2004. A New Biology for a New Century. MMBR. June 68(2):173-86.
- Yabuuchi, E., Kosako, Y., Naka T., Suzuki, S., Yano, I., 1999. Proposal of *Sphingomonas suberifaciens* (van Bruggen, Jochimsen, and Brown 1990) comb. nov., *Sphingomonas natatoria* (Sly 1985) comb. nov., *Sphingomonas ursincola* (Yurkov *et al.*,1997) comb. nov., and emendation of the genus *Sphingomonas*. Microbiology and Immunology 43: 339-349.