Bacterial taxonomy:
Current developments

George M. Garrity
“The greatest value of a picture is when it forces us to notice what we never expected to see.”

John Tukey
“Certainly many systematists (with the exception of microbiologists, who had few preconceptions about bacterial evolution) gave me the impression that because our proposal did not presuppose phylogenetic judgements, they must be anti-evolutionary. They seldom understood that these proposals could lead to techniques by which one could actively explore phylogeny.”

Peter Sneath
“Sam Cowan was determined to show that it was possible to distribute cultures that were authentic, pure and typical: the fact that this is now broadly accepted as the normal practice is a measure of his success.”

“The proposal to make a new start for bacterial names in 1980 gave Cowan much quiet satisfaction.”

“He had the unusual ability to make taxonomy interesting to other people.”

Peter Sneath
Crowdsourcing the global prokaryotic taxonomy

The current 16S rRNA based taxonomy

In development for 35 years
Work product of > 17,500 authors
> 20,600 taxonomic description
12,195 effective publications
Earliest taxa > 175 yrs
Average age 16.8 yrs
13.8% synonyms
7.8% explicitly emended
Dispelling commonly held perceptions

Taxonomic proposals and names
Hypotheses, not facts
Falsifiable

Valid taxa, valid names and an official taxonomy
Only nomenclature is governed
The concept of validly published names
Then and now: a systematic review of the systematics of prokaryotes in the last 80 years

Aharon Oren · George M. Garrity

Fig. 1 Then and now. A comparison of the accepted taxonomies in use at the time of the first and current issues of *Antonie van Leeuwenhoek*. Printed with permission. Copyright, NamesforLife, LLC 2013. All rights reserved.
Te r m

Symbolizes

Stands for

Term

Concept

Symbolizes

Refers to

Object

Taxonomic concept

Symbolizes

Refers to

Exemplar

Name

Stands for

Refers to

Stands for

Refers to

doi >

doi >

doi >
N4L Scribe

Annotation Options

- Embed semantic links
- Summarize Content
- Add Comments

DATA SOURCES

- Prokaryotic Names (NamesforLife)
- Plant Names (Kew, USDA, NCBI)
- Scientific Names (other) (NCBI Taxonomy)
- Sequence Accessions
- Culture Accessions

DOCUMENT (XML, Word, ODF)

Annotation

Comments

Comment [N4L2]: S. michiganensis current name for this taxon.  The current name is Ensifer
http://dx.doi.org/10.1609/nums.1328
Consider revising the document to use current name.

Comment [N4L3]: External link
http://www.medicagex.org/Products/Aluro3

Annotation Summary

<table>
<thead>
<tr>
<th>References</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATCC700748</td>
<td>E. meliloti [4, 5]</td>
</tr>
</tbody>
</table>

Taxonomy

- Prokaryotic Names (NamesforLife)
- Plant Names (Kew, USDA, NCBI)
- Scientific Names (other) (NCBI Taxonomy)

Sequence Data

- Sequence Accessions
- Culture Accessions

Strain Resolution

- Prokaryotic Names (NamesforLife)
- Plant Names (Kew, USDA, NCBI)
- Scientific Names (other) (NCBI Taxonomy)
- Sequence Accessions
- Culture Accessions

DOCUMENT (XML, Word, ODF)

Add Comments

Annotation

Annotation Summary

<table>
<thead>
<tr>
<th>Count</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ensifer Capra 1982</td>
</tr>
</tbody>
</table>
A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea

Dongying Wu¹,², Philip Hugenholtz³, Konstantinos Mavromatis¹, Rüdiger Pukall³, Eileen Dalin¹, Natalia N. Ivanova¹, Victor Kunin¹, Lynne Goodwin⁴, Martin Wu⁵, Brian J. Tindall³, Sean D. Hooper¹, Amrita Pati¹, Athanasios Lykidis¹, Stefan Spring³, Iain J. Anderson¹, Patrik D’haeseleer¹,⁶, Adam Zemla⁶, Mitchell Singer², Alla Lapidus¹, Matt Nolan¹, Alex Copeland¹, Cliff Han⁴, Feng Chen¹, Jan-Fang Cheng¹, Susan Lucas¹, Cheryl Kerfeld¹, Elke Lang³, Sabine Gronow³, Patrick Chain¹,⁴, David Bruce⁴, Edward M. Rubin¹, Nikos C. Kyrpides¹, Hans-Peter Klenk³ & Jonathan A. Eisen¹,²
A phylogeny-driven genomic encyclopaedia of Bacteria

Genomic Encyclopedia of Type Strains, Phase I: The one thousand microbial genomes (KMG-I) project

Nikos C. Kyrpides¹, Tanja Woyke¹, Jonathan A. Eisen², George Garrity³,⁴, Timothy G. Lilburn⁵, Brian J. Beck⁵, William B. Whitman⁶, Phil Hugenholtz⁷, and Hans-Peter Klenk⁸

¹ DOE Joint Genome Institute, Walnut Creek, CA
² University of California, Davis, CA
³ Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, USA
⁴ NamesforLife, LLC, East Lansing, MI, USA
⁵ American Type Culture Collection, Manassas, VA
⁶ Department of Microbiology, University of Georgia, Athens, GA
⁷ Australian Centre for Ecogenomics, The University of Queensland, Brisbane QLD 4072, Australia
⁸ Leibniz Institute DSMZ — German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany
A phylogeny-driven genomic encyclopaedia of Bacteria

Genomic Encyclopedia of Type Strains, Phase I: The one thousand microbial genomes (KMG-I) project

Genomic Encyclopedia of Bacteria and Archaea: Sequencing a Myriad of Type Strains

A phylogeny-driven genomic encyclopaedia of Bacteria

Genomic Encyclopedia of Type Strains, Phase I: The one thousand microbial genomes (KMG-I) project

Genomic Encyclopedia of Bacteria and Archaea:

Genomic Encyclopedia of Bacterial and Archaeal Type Strains, Phase III: the genomes of soil and plant-associated and newly described type strains
Validly published named prokaryotes used in the analyses

<table>
<thead>
<tr>
<th>Rank</th>
<th>Validly Published</th>
<th>Coverage (16S)a</th>
<th>Genomes</th>
<th>Outliers</th>
<th>Singleton</th>
<th>2 to 3</th>
<th>4 to 9</th>
<th>≥10</th>
</tr>
</thead>
<tbody>
<tr>
<td>species</td>
<td>12,981</td>
<td>11,800</td>
<td>4092d</td>
<td>274</td>
<td>11,800</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>genus</td>
<td>2,716</td>
<td>2422</td>
<td>1333</td>
<td>131</td>
<td>1,166</td>
<td>616</td>
<td>437</td>
<td>203</td>
</tr>
<tr>
<td>family</td>
<td>389</td>
<td>451b</td>
<td>343</td>
<td>86</td>
<td>121</td>
<td>71</td>
<td>79</td>
<td>180</td>
</tr>
<tr>
<td>family</td>
<td>389</td>
<td>451b</td>
<td>343</td>
<td>86</td>
<td>121</td>
<td>71</td>
<td>79</td>
<td>180</td>
</tr>
<tr>
<td>order</td>
<td>174</td>
<td>202b</td>
<td>161</td>
<td>47</td>
<td>52</td>
<td>28</td>
<td>34</td>
<td>90</td>
</tr>
<tr>
<td>class</td>
<td>83</td>
<td>85b</td>
<td>69</td>
<td>19</td>
<td>23</td>
<td>13</td>
<td>14</td>
<td>35</td>
</tr>
<tr>
<td>phylum</td>
<td>0</td>
<td>34c</td>
<td>33</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>domain</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Source - NamesforLife, LLC, based on validly published named prokaryotic species, subspecies and higher taxa as of August 5, 2015; adjusted for synonyms, basonyms, orthographic corrections and rejected names arising from Judicial Opinions.
International Code of Nomenclature of Prokaryotes

Authors: Charles T Parker¹, Brian J Tindall², George M Garrity³

Published Ahead of Print: 20 November, 2015 International Journal of Systematic and Evolutionary Microbiology doi: 10.1099/ijsem.0.000778

Published Online: 20/11/2015
International Code of Nomenclature of Prokaryotes

Authors: Charles T Parker\textsuperscript{1}, Brian J Tindall\textsuperscript{2}, George M Garrity\textsuperscript{3}

Published Ahead of Print: 20 November, 2015 International Journal of Systematic and Evolutionary Microbiology doi: 10.1099/ijsem.0.000778

Published Online: 20/11/2015

International Journal of Systematic and Evolutionary Microbiology (2015), 65, 4284–4287

Proposal to include the rank of phylum in the International Code of Nomenclature of Prokaryotes

Aharon Oren,\textsuperscript{1} Milton S. da Costa,\textsuperscript{2} George M. Garrity,\textsuperscript{3} Fred A. Rainey,\textsuperscript{4} Ramon Rossello-Mora,\textsuperscript{5} Bernhard Schink,\textsuperscript{6} Iain Sutcliffe,\textsuperscript{7} Martha E. Trujillo\textsuperscript{8} and William B. Whitman\textsuperscript{9}

Taxonomic Note
**CRUCIAL QUESTIONS**

Here are five questions that anyone conducting or evaluating this research should ask to keep from getting carried away by hype.

Can experiments detect differences that matter? Profiling a microbiome could produce a catalogue at the level of phyla, species or genes. Much work relies on analysis of 16S rRNA, an ancient gene that tolerates little variation and so is reliably found across the bacterial kingdom. But this allows only a coarse sorting. For example, microbiomes associated with obesity have been distinguished by different ratios of bacterial phyla, which encompass a staggering range of diversity. If this criterion were used to characterize animal communities, an avairy of 100 birds and 25 snails would be considered identical to an aquarium with 8 fish and 2 squid, because each has four times as many vertebrates as molluscs. Even within a single species, strains often differ greatly in the genes they contain.

**William P. Hanage**

Nature 512, 247–248 (21 August 2014)
doi:10.1038/512247a
Firmicutes

Bacteria: Firmicutes
Planctomycetes

Bacteria: Planctomycetes
Chlamydiae

Bacteria: Chlamydiae
Fibrobacteres

Bacteria: Fibrobacteres
Incertae sedis 172

Bacteria: Incertae sedis 172
Recovering shapes from data
16S rRNA similarity of type strains of Bacteria and Archaea 1980
16S rRNA similarity of type strains of Bacteria and Archaea 1981
16S rRNA similarity of type strains of Bacteria and Archaea 1982
16S rRNA similarity of type strains of Bacteria and Archaea 1983
16S rRNA similarity of type strains of Bacteria and Archaea 1984
16S rRNA similarity of type strains of Bacteria and Archaea 1985
16S rRNA similarity of type strains of Bacteria and Archaea 1986
16S rRNA similarity of type strains of Bacteria and Archaea 1987
16S rRNA similarity of type strains of Bacteria and Archaea 1988
16S rRNA similarity of type strains of Bacteria and Archaea 1989
16S rRNA similarity of type strains of Bacteria and Archaea 1990
16S rRNA similarity of type strains of Bacteria and Archaea 1991
16S rRNA similarity of type strains of Bacteria and Archaea 1992
16S rRNA similarity of type strains of Bacteria and Archaea 1993
16S rRNA similarity of type strains of Bacteria and Archaea 1994
16S rRNA similarity of type strains of Bacteria and Archaea 1995
16S rRNA similarity of type strains of Bacteria and Archaea 1996
16S rRNA similarity of type strains of Bacteria and Archaea 1997
16S rRNA similarity of type strains of Bacteria and Archaea 1998
16S rRNA similarity of type strains of Bacteria and Archaea 1999
16S rRNA similarity of type strains of Bacteria and Archaea 2000
16S rRNA similarity of type strains of Bacteria and Archaea 2002
16S rRNA similarity of type strains of Bacteria and Archaea 2004
16S rRNA similarity of type strains of Bacteria and Archaea 2005

Bacteroidetes
Spirochaetes
Actinobacteria
Firmicutes
Proteobacteria
Deltaproteobacteria
Euryarchaeota
Crenarchaeota
16S rRNA similarity of type strains of Bacteria and Archaea 2006
16S rRNA similarity of type strains of Bacteria and Archaea 2007
16S rRNA similarity of type strains of Bacteria and Archaea 2008
16S rRNA similarity of type strains of Bacteria and Archaea 2011
16S rRNA similarity of type strains of Bacteria and Archaea 2013
16S rRNA similarity of type strains of Bacteria and Archaea 2015
Refinement of the underlying taxonomy

Improved version of the Self-Organizing Self Correcting Classifier

Goal - smooth classification while protecting taxonomic structure

Input

13,125 HQ full-length 16S rRNA sequences
Aligned to Greengenes alignment (Release 13.5)
Jukes-Cantor model, computed using Mothur

Output filtered to N4L-exemplar DOIs to select for validly published named species/subspecies

Outliers are identified at species level for members with four or more members based on within group distribution as represented in the lower triangle of each corresponding sub-matrix (2 $\sigma$), output visualized as heatmaps, phylograms, (APE) and histograms.

207/11,800 HQ 16S sequences were were flagged as outliers and excluded from further analysis
Domain level rearrangement
Challenges

Taxonomic ranks in current use based on 16S rRNA similarity
Significantly overlap
Validly named taxa may be over-specified
Rank of order may not be justifiable by a single measure
Reclassification and re-naming of at least species may be justified

One or more additional measures could help resolve taxa above the genus level
Projection along orthogonal axis or plane

Obvious candidates
Methods that classify on other genes, sequences, amino acids or proteins
Candidates
Average nucleotide identity
Average amino acid identity
K-mers
Latent semantic analysis of DNA sequences

Operational limits
Experimental plan
Assemble a reasonable sized collections of type strains with sequenced genomes
Strains representing the type (genus) of 170/202 validly named orders
Varying quality (finished, high quality draft, draft)
Sequenced/assembled by different groups
Collaborators perform analysis using preferred method for all strains
Visualized against 16S sequence similarly data for same strains
NamesforLife - Charles Parker
DOE JGI - Nikos Kyrpides, Neha Varghese
ORNL - Dave Ussery, Miriam Land, Se-Ran Jun, Intewat Nookaew, Visanu Wanchai
Univ. Nebraska, Lincoln - Khalid Sayood, Ufuk Ubantagalu
Trends between gene content and genome size in prokaryotic species with larger genomes

Konstantinos T. Konstantinidis*† and James M. Tiedje*‡§

*Center for Microbial Ecology and Departments of †Crop and Soil Sciences and ‡Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824-1325

Contributed by James M. Tiedje, December 24, 2003|
Trends between gene content and genome size

Genomic insights that advance the species definition for prokaryotes

Konstantinos T. Konstantinidis*† and James M. Tiedje*‡§

*Center for Microbial Ecology, and Departments of †Crop and Soil Sciences and ‡Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824

Contributed by James M. Tiedje, December 24, 2004

To help advance the species definition for prokaryotes, we have compared the gene content of 70 closely related and fully se-
Towards a Genome-Based Taxonomy for Prokaryotes

Konstantinos T. Konstantinidis\textsuperscript{1,2} and James M. Tiedje\textsuperscript{1,2,3,*}

Center for Microbial Ecology\textsuperscript{1} and Departments of Crop and Soil Sciences\textsuperscript{2} and Microbiology and Molecular Genetics,\textsuperscript{3} Michigan State University, East Lansing, Michigan
Trends between gene content and genome size

Genomic insights that advance the species definition

Vol. 187, No. 18

Towards a Genome-Based Taxonomy for Prokaryotes

Vol. 72, No. 1

Toward a More Robust Assessment of Intraspecies Diversity, Using Fewer Genetic Markers

Konstantinos T. Constantinidis,* Alban Ramette,§● and James M. Tiedje

Center for Microbial Ecology, Michigan State University, East Lansing, Michigan

Received 16 June 2006/Accepted 5 September 2006
Towards a Genome-Based Taxonomy for Prokaryotes

Toward a More Robust Assessment of Intraspecies Diversity.

The bacterial species definition in the genomic era

Konstantinos T. Konstantinidis*, Alban Ramette† and James M. Tiedje

Center for Microbial Ecology, Michigan State University, East Lansing, MI 48824, USA
Trends between gene content and genome size

Genomic insights that advance the species definition

Towards a Genome-Based Taxonomy for Prokaryotes

Toward a More Robust Assessment of Intraspecies Diversity.

The bacterial species definition in the genomic era

Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison

Alexander F. Auch¹, Mathias von Jan², Hans-Peter Klenk²*, Markus Göker²
16S rRNA sequence subset

ANI, Oak Ridge National Laboratories

166 type strains, with conditioning of AAI signal

\[ \text{cor} = -0.46807 \]
16S rRNA sequence subset

AAI, Oak Ridge National Laboratories

166 type strains, with conditioning of AAI signal

\[ \text{cor} = -0.6631114 \]
16S rRNA sequence subset

Latent Semantic analysis, 50 words.
Univ. Nebraska and NamesforLife

166 type strains, with conditioning of LSA signal
$\text{cor} = -0.02849482$
Type strains reordered according to LSA classification

Type strain genomes reordered according to LSA
Results

Taxonomic resolving power of 16S rRNA

Taxonomic resolving power of ANI, AAI and LSA

Operational details

Adequacy of metadata for downstream analyses/interpretation

Surprises

Standards

Naming and taxon calling

Persistence and linked to relevant literature

Data quality

Weighting schemes

Validation methods

Models

Software
This work was funded through the Small Business technology transfer program of the United States Department of Energy under grants number DE-FG02-07ER86321 and DE-SC0006191. Funding for business development was provided through grants and loans from the Michigan Economic Development Corporation and the Michigan Universities Commercialization Initiative. NamesforLife semantic resolution technology is covered under US Patent 7,925,444 B2. The SOSCC systems and methods is covered under US Patent 8,036,997. Semiotic fingerprinting is covered under US 8,903,824. Semantic tagging, indexing, equivalency mapping, and latent semantic analysis of molecular sequence data are the subjects of pending US and EPO patent applications.