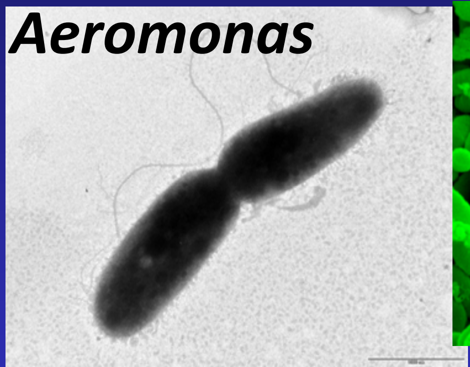
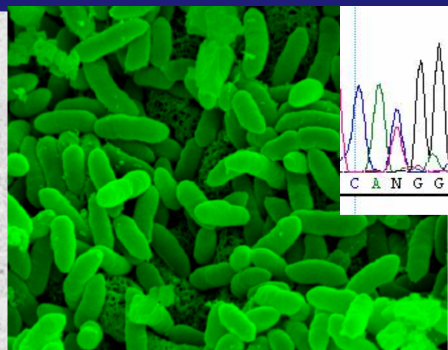
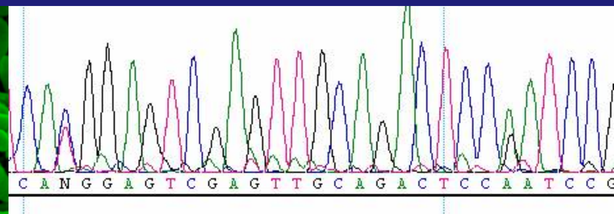
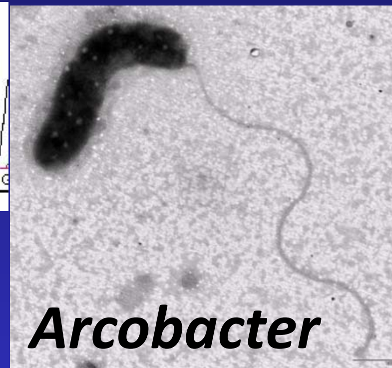
Interdisciplinary Center for Aquaculture Research (INCAR), Concepción, Chile
Interdisciplinary Center for Aquaculture Research (INCAR), Concepción, Chile

Taxonomic and toxicogenic potential derived from whole genome sequencing (WGS) information









Prof. Maria José Figueras
Unit of Microbiology
Faculty of Medicine

Reus, Spain

UNIVERSITAT ROVIRA I VIRGILI

OBJECTIVES

- To present the strategies to determine the species and strain identity using the genome information

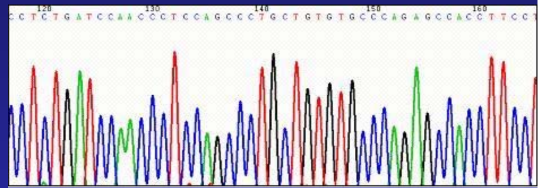
- To underline additional information such antibiotic resistance and virulence genes that can be found in the genomes

Standard requirements for defining new bacteria species (Stackebrandt et al., 2002)

1. **As many strains as possible** (different numbers proposed i.e. 5, 10, 25)

2. **Phenotypic characterisation**

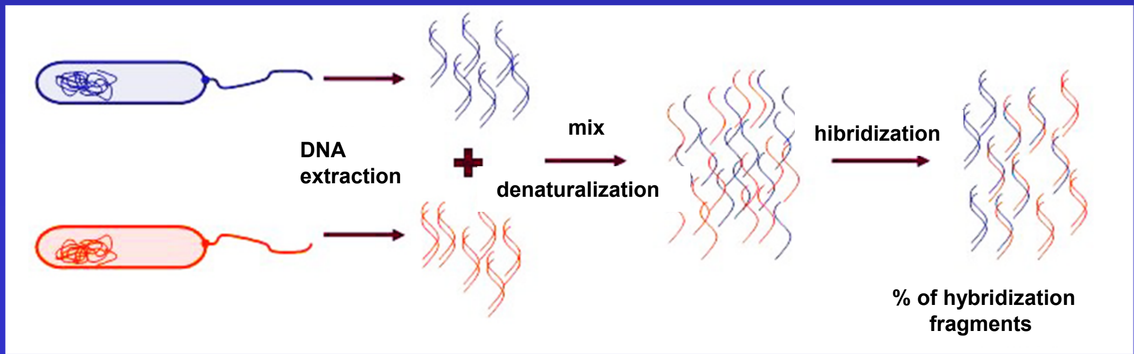
3. **Sequences of the 16S rRNA gene (>1300bp, <5% ambiguity)**



similarity > 97%
(Stackebrandt & Goebel 1994)

DNA-DNA hybridization

> 70% DNA corresponded to the same species



http://www.seimc.org/documentos/protocolos/microbiologia/cap37.asp

Genera and species of bacteria in which the 16S rRNA posses a poor resolution for their discrimination

Genus	Species
Aeromonas	*A. veronii, A. caviae, A. trota, A. salmonicida, A. bestiarum* **100% similarity**
Bacillus	*B. anthracis, B. cereus, B. thuringiensis, B. globisporus, B. psychrophilus*
Bordetella	*B. pertussis, B. parapertussis, B. bronchiseptica, B. holmesii*
Brucella	*B. melitensis, B. abortus, B. suis* y otros
Burkholderia	*B. mallei, B. pseudomallei, B. cocovenenans, B. gladioli, B. thailandensis B. cepacia, B vietnamiensis, B. multivorans, B. stabilis*
Campylobacter	*C. jejuni C.coli* **100% similarity**
Corynebacterium	*C. diphtheriae, C. pseudotuberculosis, C. ulcerans, C. kutscheri. C. afermentans*
Enterobacteriaceae	*E. coli, Shigella* spp./ *E. coli* enteroinvasivo (EIEC)
Streptococcus	*S. sinensis , S. gallolyticus, S. infantarius, S. pneumoniae, S. pseudopneumoniae, S. salivarius, S. mutans, S. suis, S. sanguinis, S. cristatus, S. sinensis, S. anginosus, S. intermedius, S. constellatus, S. mitis, S. infantis, S. peroris, S. oralis, S. oligofermentans* etc
Vibrio	***V. harveyi, V. campbelli***

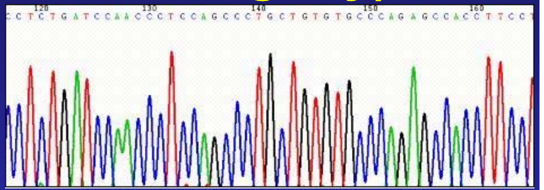
Standard requirements for defining new bacteria species (Stackebrandt et al., 2002)

1. **As many strains as possible** (different numbers proposed i.e. 5, 10, 25)

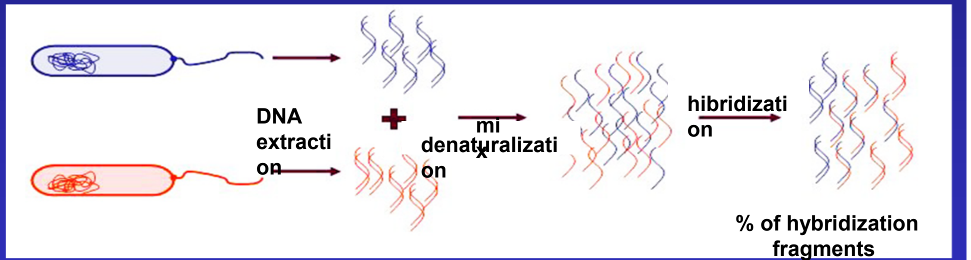
2. **Phenotypic characterisation**

3. **Sequences of the 16S rRNA gene (>1300bp, <5% ambiguity)**

similarity > 97% (Stackebrandt & Goebel 1994) **98.7-99%** (Stackebrandt & Ebers 2006)



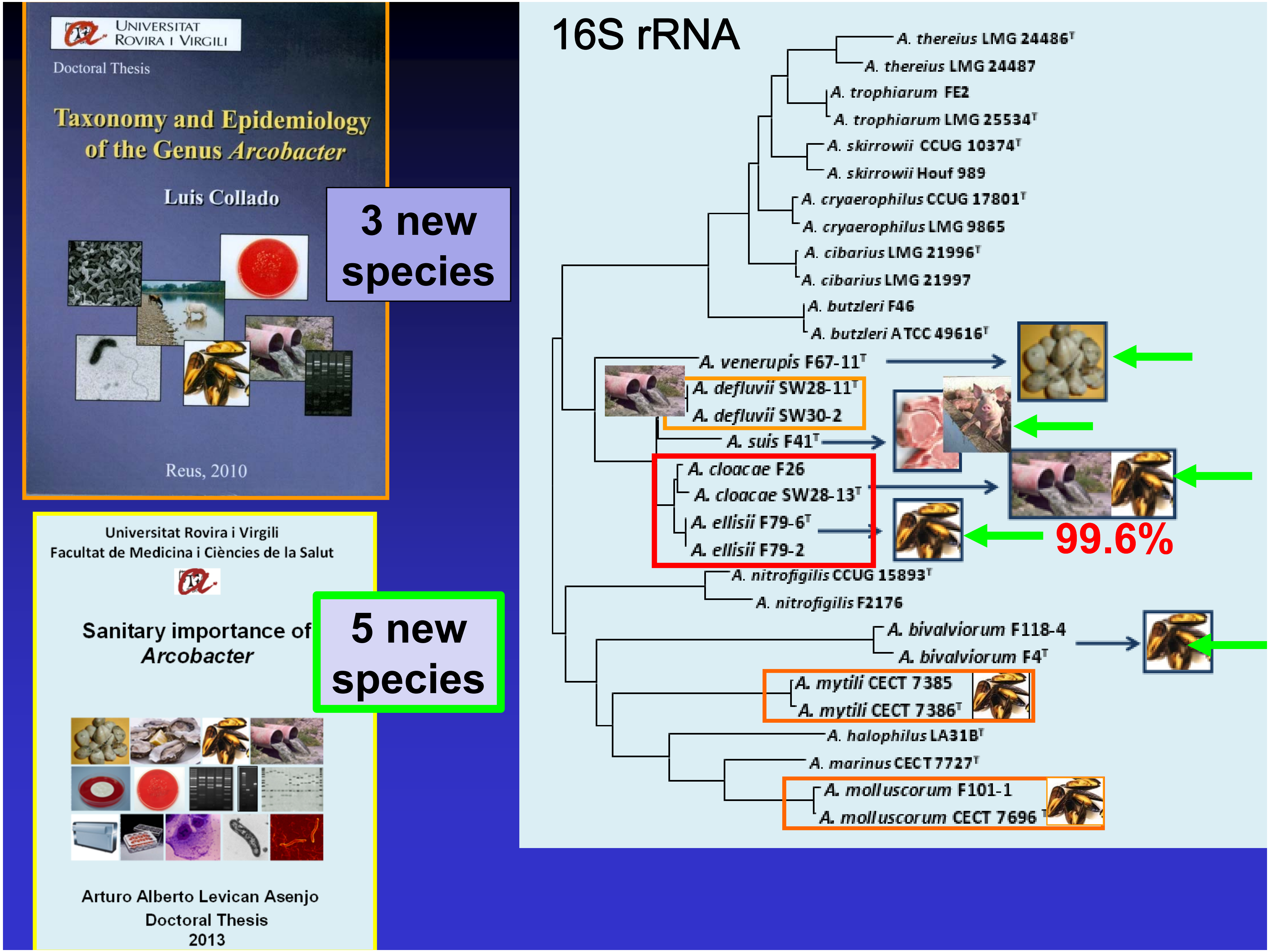
DNA-DNA hybridization > 70%



5. **Multilocus sequence analysis (MLSA) or phylogenetic analysis (MLPA) of a minimum of 5 housekeeping genes**

6. **Genotyping (ERIC-PCR, AFLP..)**

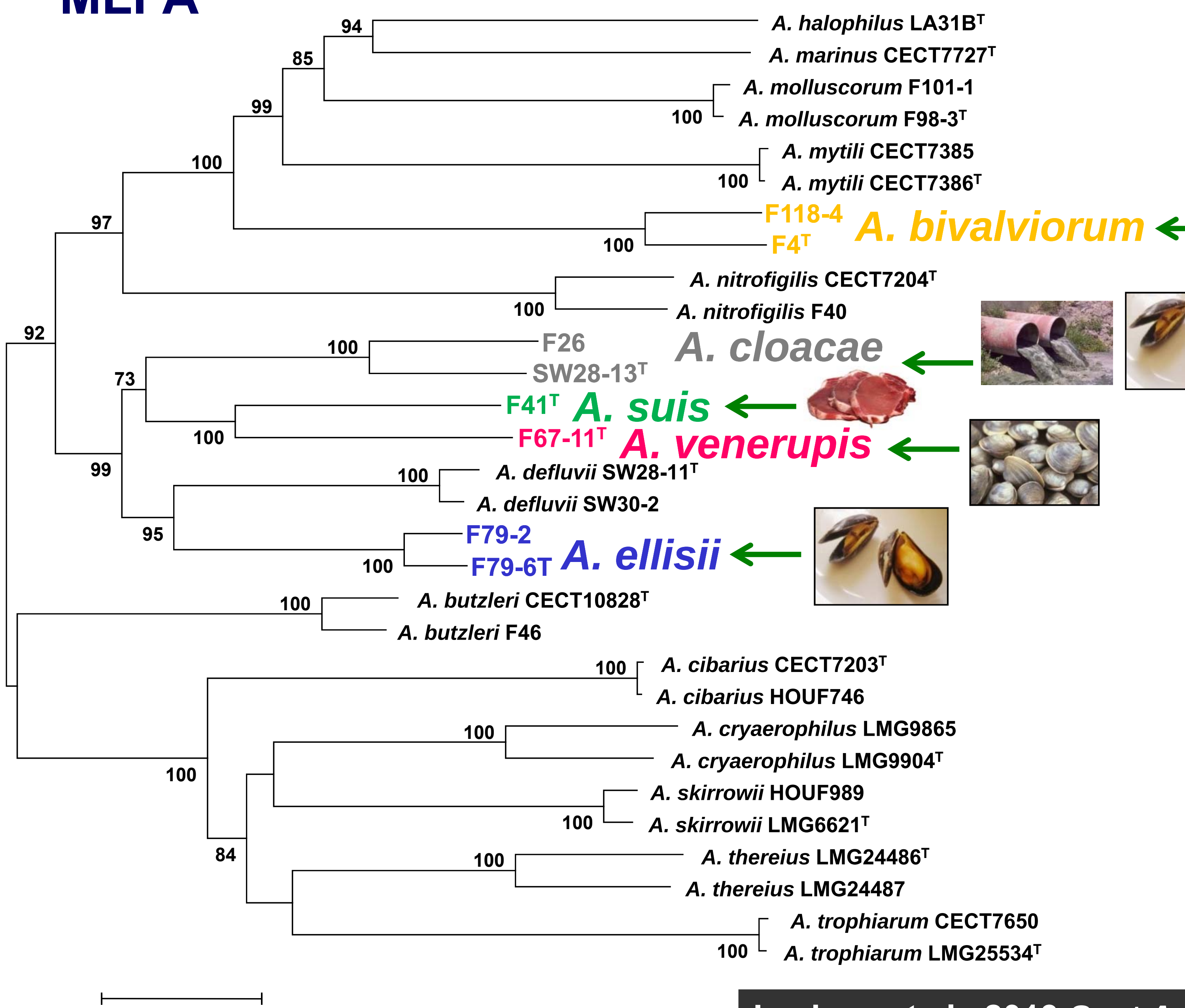
7. **Chemotaxonomic properties (cell wall composition, lipids, fatty acids …)**



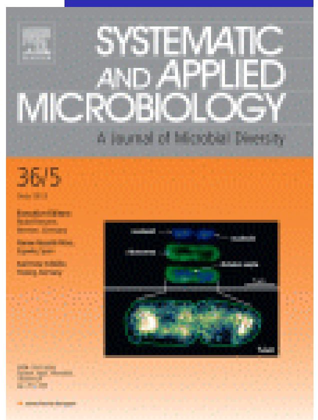
Universitat Rovira i Virgili
Doctoral Thesis
Taxonomy and Epidemiology of the Genus Arcobacter
Luis Collado
Reus, 2010
3 new species
Universitat Rovira i Virgili
Facultat de Medicina i Ciències de la Salut
Sanitary importance of Arcobacter
Arturo Alberto Levican Asenjo
Doctoral Thesis
2013
5 new species
16S rRNA
A. thereius LMG 24486T
A. thereius LMG 24487
A. trophiarum FE2
A. trophiarum LMG 25534T
A. skirrowii CCUG 10374T
A. skirrowii Houf 989
A. cryaerophilus CCUG 17801T
A. cryaerophilus LMG 9865
A. cibarius LMG 21996T
A. cibarius LMG 21997
A. butzleri F46
A. butzleri ATCC 49616T
A. venerupis F67-11T
A. defluvii SW28-11T
A. defluvii SW30-2
A. suis F41T
A. cloacae F26
A. cloacae SW28-13T
A. ellisii F79-6T
A. ellisii F79-2
99.6%
A. nitrofigilis CCUG 15893T
A. nitrofigilis F2176
A. bivalviorum F118-4
A. bivalviorum F4T
A. mytili CECT 7385
A. mytili CECT 7386T
A. halophilus LA31BT
A. marinus CECT 7727T
A. molluscorum F101-1
A. molluscorum CECT 7696T

Neighbour joining tree based on the concatenated sequences of 5 genes:

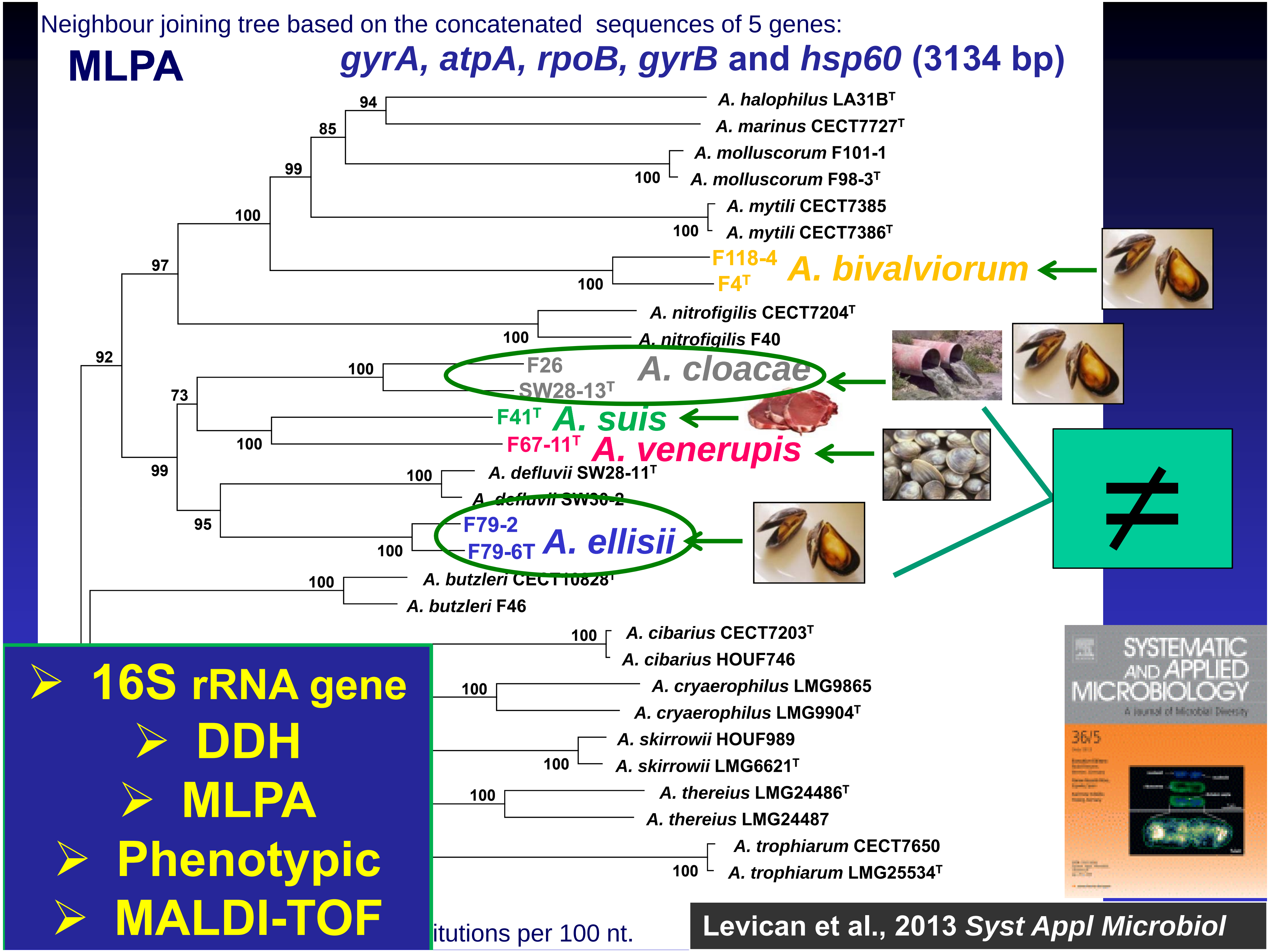
MLPA *gyrA, atpA, rpoB, gyrB* and *hsp60* (3134 bp)



. Bar, 2 substitutions per 100 nt.

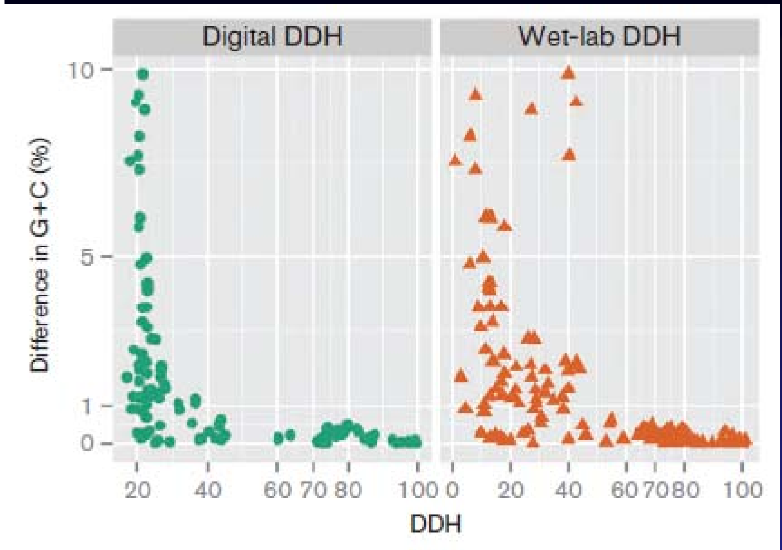


Levican et al., 2013 *Syst Appl Microbiol*

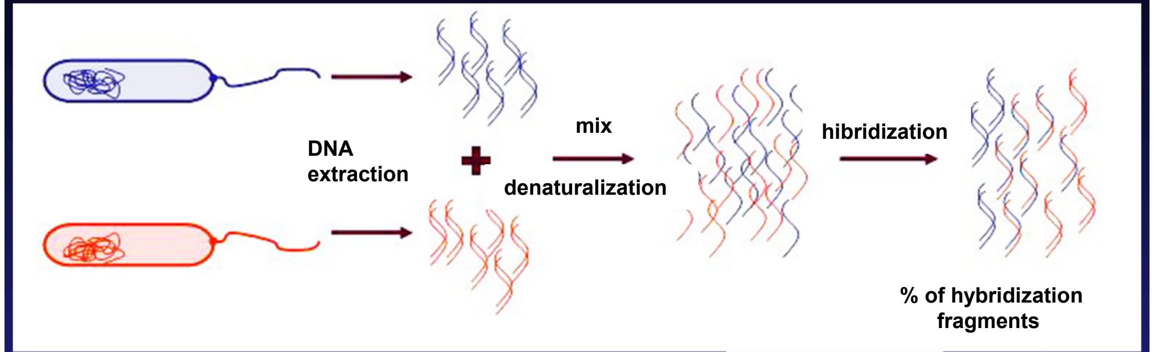


Neighbour joining tree based on the concatenated sequences of 5 genes:
MLPA
gyrA, atpA, rpoB, gyrB and hsp60 (3134 bp)
94
85
99
100
97
92
A. halophilus LA31BT
A. marinus CECT7727T
A. molluscorum F101-1
100
A. molluscorum F98-3T
A. mytili CECT7385
100
A. mytili CECT7386T
F118-4
F4T
100
A. bivalviorum
A. nitrofigilis CECT7204T
100
A. nitrofigilis F40
100
F26
SW28-13T
A. cloacae
73
F41T
A. suis
100
F67-11T
A. venerupis
99
100
A. defluvii SW28-11T
A. defluvii SW30-2
95
F79-2
100
F79-6T
A. ellisii
100
A. butzleri CECT10828T
A. butzleri F46
100
A. cibarius CECT7203T
A. cibarius HOUF746
100
A. cryaerophilus LMG9865
A. cryaerophilus LMG9904T
A. skirrowii HOUF989
100
A. skirrowii LMG6621T
100
A. thereius LMG24486T
A. thereius LMG24487
A. trophiarum CECT7650
100
A. trophiarum LMG25534T
≠
16S rRNA gene
DDH
MLPA
Phenotypic
MALDI-TOF
itutions per 100 nt.
SYSTEMATIC AND APPLIED MICROBIOLOGY
A Journal of Microbial Diversity
36/5
Levican et al., 2013 Syst Appl Microbiol

in silico method

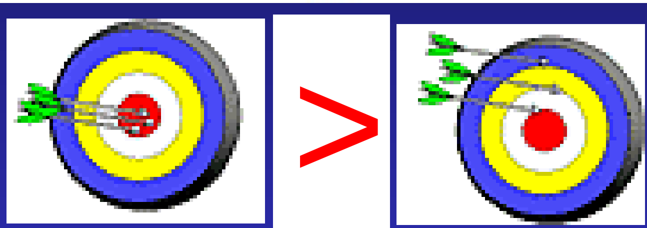


DNA-DNA hybridization (DDH)



GGDC
Genome-to-Genome Distance Calculator
Meier-Kolthoff et al., 2013

DSMZ



***is*DDH= *in silico* or digital= dDDH**

Precision

The species concept for *Bacteria* and *Archaea* is based on the 16S rRNA gene and on DNA-DNA hybridization (DDH), a method known to be tedious.

The GGDC is in silico method for genome-to-genome comparison, thus reliably mimicking conventional DDH

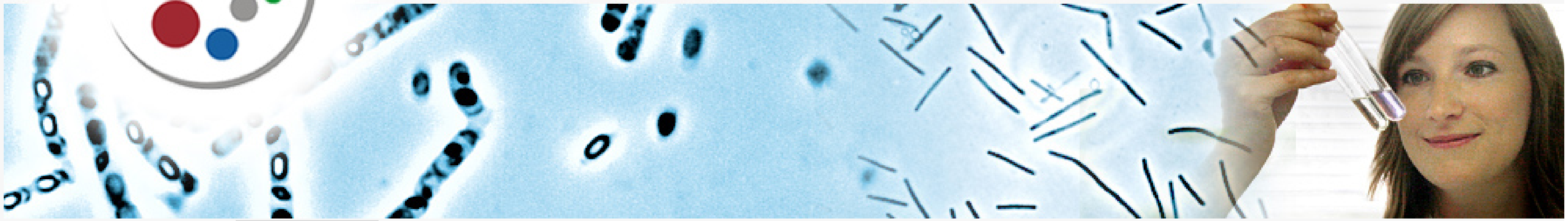
DDH = digital DDH >70%

Nueva pestaña

https://www.dsmz.de/research/microorganisms/projects/genome-to-genome-distance-calculator.html

Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH

Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures



HOME | ABOUT US | RESEARCH | BACTERIAL DIVERSITY | CATALOGUES | DEPOSIT | SERVICES | SHOP | SUPPORT | CONTACT | FAQ

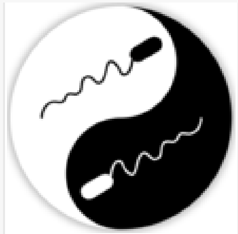
Website Search

to product search »

- Publications
- Microorganisms
  - Group Geomicrobiology
  - Curators
  - Projects
  - Publications
  - Scientific Staff
- Microbial Ecology and Diversity Research
- Human and Animal Cell Lines
- Plant Viruses
- Central Services

Research > Microorganisms > Projects > Genome-to-genome distance calculator

GGDC: Genome-To-Genome Distance Calculator



The pragmatic species concept for *Bacteria* and *Archaea* is ultimately based on **DNA-DNA hybridization (DDH)**. While enabling the taxonomist, in principle, to obtain an estimate of the overall similarity between the genomes of two strains, this technique is tedious and not easily be made reproducible between different labs and cannot be used to incrementally built up a comparative database. Recent technological progress in the area of genome sequencing calls for bioinformatics methods to replace the wet-lab DDH by in-silico genome-to-genome comparison.

The web service hosted at DSMZ offers state-of-the-art methods for inferring whole-genome distances which are well able to mimic DDH. Values calculated with GGDC yield a better correlation with wet-lab DDH values than alternative approaches such as "ANI". These distance functions can also cope with heavily reduced genomes and repetitive sequence regions. Some of them are also very robust against missing fractions of genomic information (due to incomplete genome sequencing). Thus, this web service can be used for genome-based species delineation. As of 2014 the GGDC also delineates subspecies.

Use the GGDC here.

SHOPPING CART

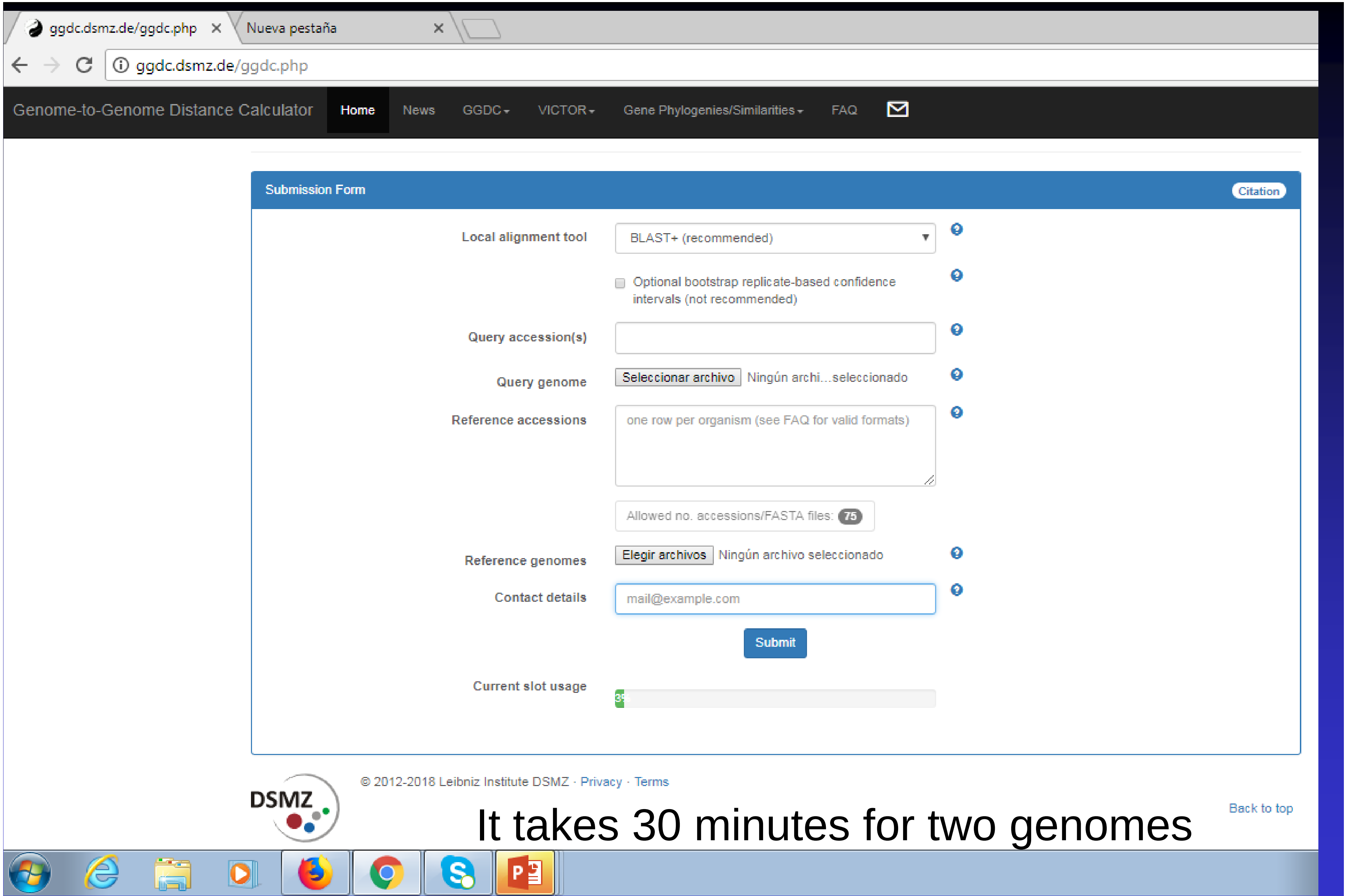
Please login!

To Shopping Cart »

To Login/Logout »

Quality Management DQS ISO 9001:2008

Zertifikat seit 2010 audit berufundfamilie

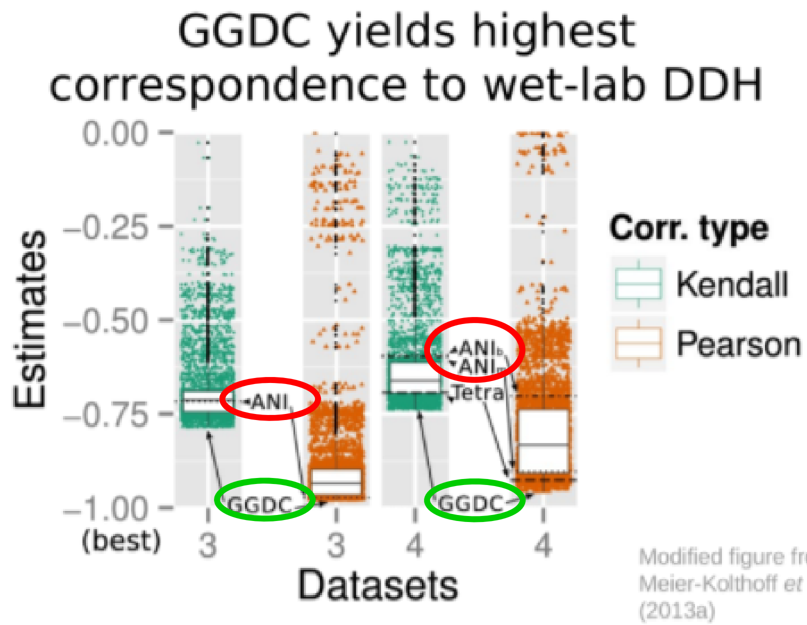


It takes 30 minutes for two genomes

Tools to compare genomes to determine their taxonomic relatedness

Digital DNA:DNA hybridization. Very reliable in silico method.

GGDC yielded higher correlations with wet-lab DDH (without mimicking its pitfalls) than other in silico approaches. GGDC uses statistical models that considerably improve on the linear models used by other approaches (e.g. ANI). A practical advantage of GGDC over ANI is that GGDC operates on the same scale than wet-lab DDH values, which makes comparisons much easier.

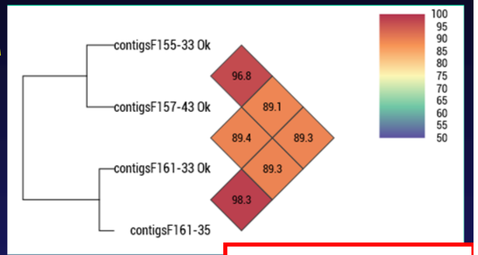


Modified figure from Meier-Kolthoff *et al.* (2013a)

ANI = Average Nucleotide Identity is considered an overall genome related index (OGRI) that provides a % of relatedness between the genomes compared

UNIVERSITAT ROVIRA I VIRGILI

Results of the comparison of the different available platforms that calculate the Average Nucleotide Identity (ANI)



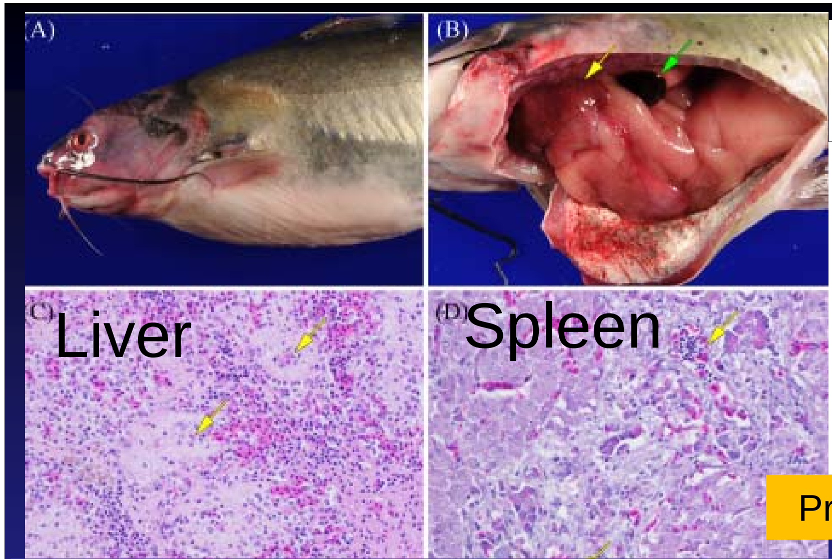
Yoon et al., 2017

ANIu

Caracteristics	Jspecies	ANI calculator	EzGenome	OrthoANI
Web server	✓	✓	✓	✗
Does not need internet connection	✗	✗	✗	✓
Alert when task is finished	✗	✓	✗	✗
Easy to handle	✓	✓	✓	✓
Construct Matrix comparisons	✓	✗	✗	✓
Direct results	✗	✓	✗	✓
+2 genomes/analysis	<12Mb ✓	✗	✗	10 gen. ✓

Oat OAT Orthologous Average Nucleotide Identity Tool a similarity measurement tool for genomes > JSpecies Taxonomic Thresholds > ANI Average Nucleotide Identity Kostas lab » Tools » ANI calculator > EZBIOCLOUD EzGenome

OAT-ANI is currently the best



frontiers in Microbiology

UNIVERSITAT ROVIRA I VIRGILI

ORIGINAL RESEARCH
published: 18 October 2016
doi: 10.3389/fmicb.2016.01615

Classification of a Hypervirulent *Aeromonas hydrophila* Pathotype Responsible for Epidemic Outbreaks in Warm-Water Fishes

Cody R. Rasmussen-Ivey[1], Mohammad J. Hossain[1], Sara E. Odom[1], Jeffery S. Terhune[2], William G. Hemstreet[3], Craig A. Shoemaker[4], Dunhua Zhang[4], De-Hai Xu[4], Matt J. Griffin[5], Yong-Jie Liu[6], Maria J. Figueras[7], Scott R. Santos[1], Joseph C. Newton[8*] and Mark R. Liles[1*]

Prof. Mark Liles

Goris *et al*., 2007

JSpecies
Taxonomic Thresholds
Richter & Rosselló-Mora, 2009

Average nucleotide identity of *A. hydrophila* genomes vs. genome of the epidemic strain ML09-119

Species	Isolates	ANI(%)
A. hydrophila	ML09-121	99,99
A. hydrophila	ML09-122	99,99
A. hydrophila	S04-690	99,99
A. hydrophila	ZC1	99,98
A. hydrophila	PB10-118	99,99
A. hydrophila	AL10-121	99,99
A. hydrophila	AL09-79	99,99
A. hydrophila	ATCC7966T	97,14
A. hydrophila	226	97,13
A. hydrophila	AL06-06	97,13
A. hydrophila	E1	97,12
A. hydrophila	E2	97,12
A. hydrophila	SNUFPCA8	97,11
A. hydrophila	MN98-04	97,09
A. hydrophila	AL97-91	97,08
A. hydrophila	AL06-01	97,08
A. hydrophila	TN97-08	97,07
A. hydrophila	145	93,84
A. hydrophila	277	93,76
A. hydrophila	SSU	93,75
A. hydrophila	173	93,72
A. hydrophila	187	93,7
A. hydrophila	259	93,7
A. hydrophila	GA97-22	89,94
A. hydrophila	4AKA	88,48

99.99%

Aeromonas hydrophila

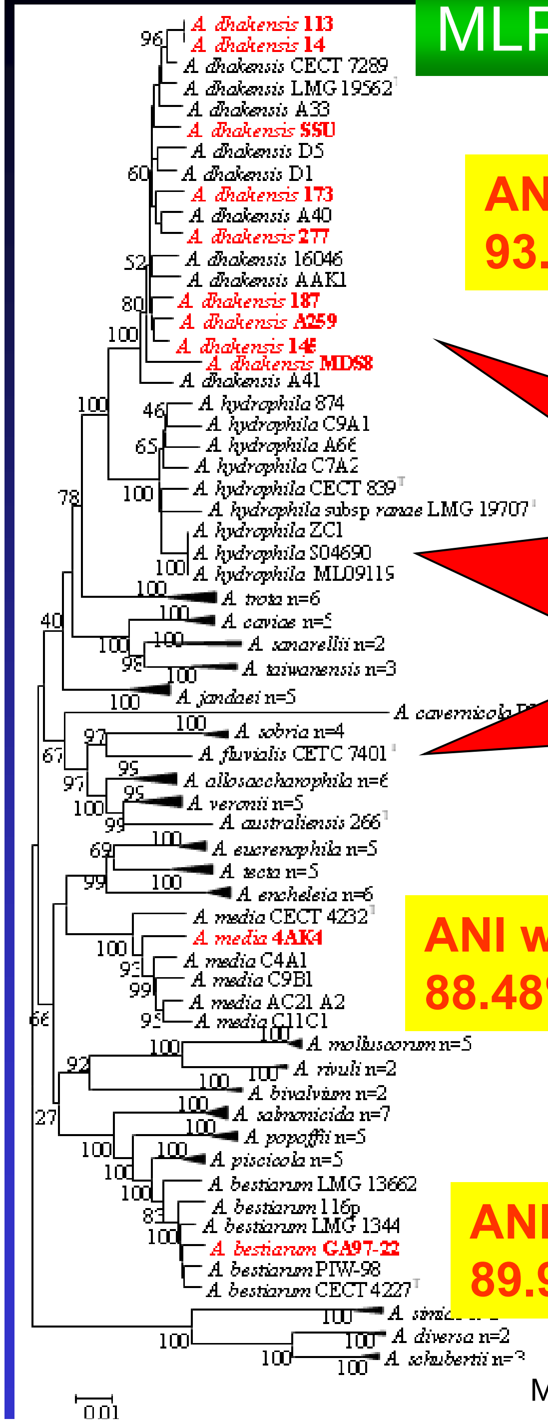
>97%

ANI < 96% different species

88.48 - 93.84%

~~***Aeromonas hydrophila***~~

MLPA (*gyrB, rpoD, recA, dnaJ, gyrA, dnaX*, 3882 bp)

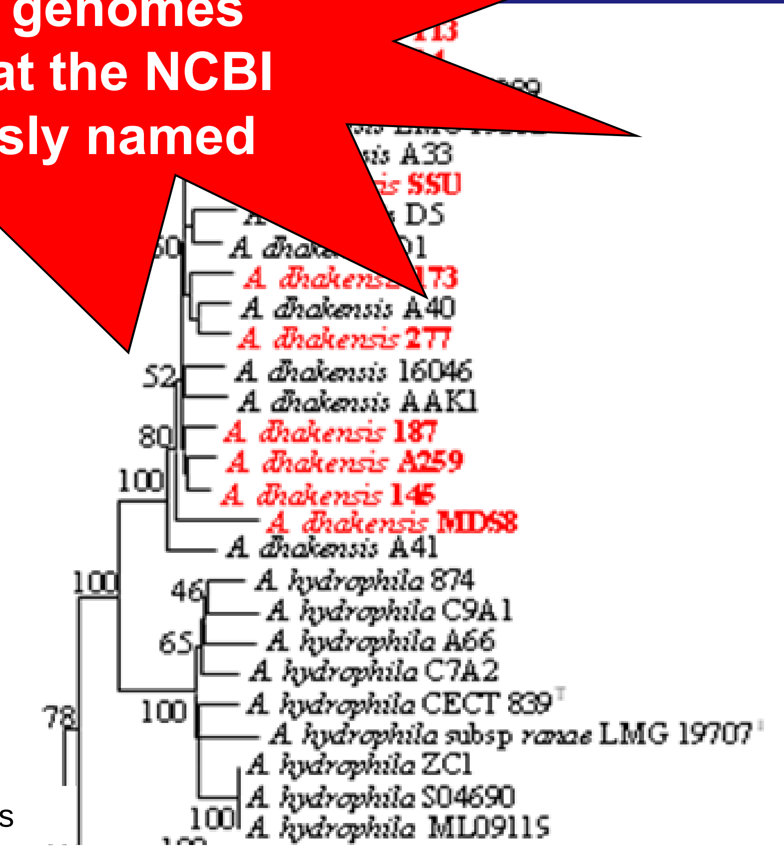


ANI values

A. hydrophila	145	93,84
A. hydrophila	277	93,76
A. hydrophila	SSU	93,75
A. hydrophila	173	93,72
A. hydrophila	187	93,7
A. hydrop	259	93,7
A. hyd	GA97-22	89,94
A. h	4AKA	88,48
A	AAK1	93,73
	113	93,74
	14	93,77
		93,71

ANI with *A. hydrophila* 93.71-93.84%

55% of the *A. hydrophila* strains whose genomes were available at the NCBI were erroneously named



ANI with *A. hydrophila* 88.48%

ANI with *A. hydrophila* 89.94%

MLPA = Multilocus Phylogenetic Analysis

UNIVERSITAT ROVIRA I VIRGILI

Prof. Mark Liles,
Dr. Jahangir Hossain

genomeA
Journals.ASM.org

2014

COMMENTARY

Taxonomic Affiliation of New Genomes Should Be Verified Using Average Nucleotide Identity and Multilocus Phylogenetic Analysis

María José Figueras,[a] Roxana Beaz-Hidalgo,[a] Mohammad J. Hossain,[b] Mark R. Liles[b]
Unitat de Microbiologia, Departament de Ciències Mèdiques Bàsiques, Facultat de Medicina i Ciències de la Salut, IISPV, Unive
Department of Biological Sciences, Auburn University, Auburn, Alabama, USA[b]

SERENDIPITY
(ser en dip i-te) n.
making fortunate discovery by accident

PLOS | ONE

2015

RESEARCH ARTICLE

Strategies to Avoid Wrongly Labelled Genomes Using as Example the Detected Wrong Taxonomic Affiliation for *Aeromonas* Genomes in the GenBank Database

Roxana Beaz-Hidalgo[1], Mohammad J. Hossain[2], Mark R. Liles[2], Maria-Jose Figueras[1]*
1 Unitat de Microbiologia, Departament de Ciències Mèdiques Bàsiques, Facultat de Medicina i Ciències de la Salut, IISPV, Universitat Rovira i Virgili, Reus, Spain, 2 Department of Biological Sciences, Auburn University, Auburn, Alabama, United States of America

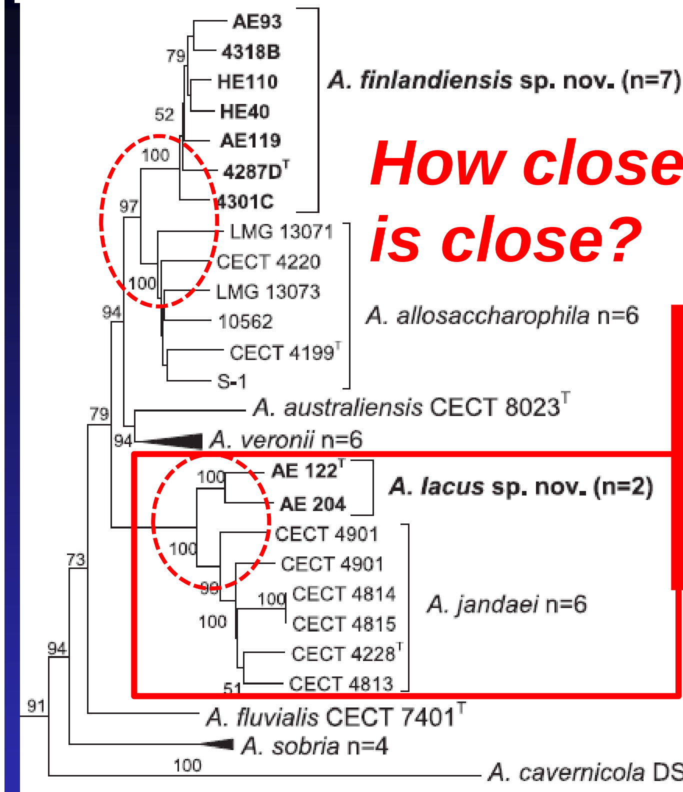
If ANI is < 96% = different species

1. **The closest neighbours can be determined with a Multilocus Phylogenetic Analysis (MLPA) extracting the genes from the genomes.**

2. **Then the genome (of the type or another) of these closest neighbour species should be used for the Average Nucleotide Identity (ANI) and *in silico* DNA-DNA (isDDH) hybridization calculations to determine the final identity.**

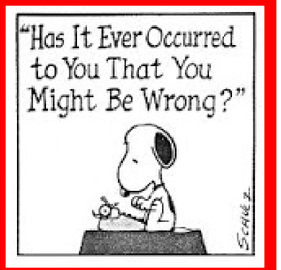
MLPA 7 genes (*gyrB, rpoD, recA, dnaJ, gyrA, dnaX and atpD*; 4093 bp)

Aeromonas

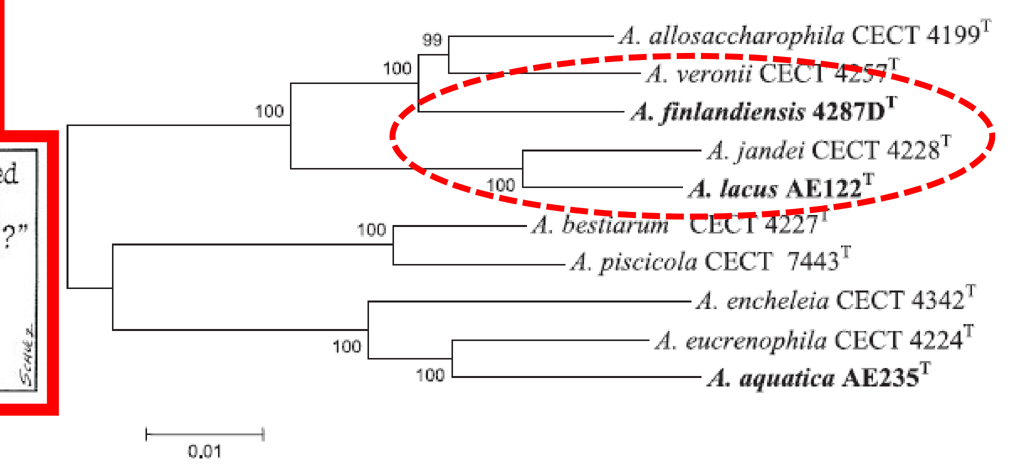


MLPA = Multilocus Phylogenetic Analysis

How close is close?



MLPA 15 genes (*cpn60, dnaK, gltA, mdh, radA, rpoB, tsf, zipA, gyrA, gyrB, rpoD, recA, dnaJ, atpD and dnaX*, 8751 bp)



ANI values seems to be more objective

A. lacus ≠ *A. jandaei* ANI < 96% and *is*DDH < 70%

Results of ANI calculations using ANI calculator, EzGenome tools and JSpecies software and *in silico* DDH of the 3 new species with respect to the most closely related species.

Species	ANI Calculator (ANIb)	EzGenome (ANIb)	JSpecies (ANIb)	JSpecies (ANIm)	GGDC
A. lacus AE122^T vs. *A. jandaei* CECT 4228^T	95.16	95.17	95.44	95.74	63.20

Federhen *et al. Standards in Genomic Sciences* (2016) 11:15
DOI 10.1186/s40793-016-0134-1

2016

Standards in Genomic Sciences

MEETING REPORT Open Access

CrossMark

Meeting report: GenBank microbial genomic taxonomy workshop (12–13 May, 2015)

Scott Federhen[1*], Ramon Rossello-Mora[2], Hans-Peter Klenk[3], Brian J. Tindall[4], Konstantinos T. Konstantinidis[5],

ANI and Proxytype (= genome designated by NCBI to serve as a proxy for the type, for species that do not yet have a genome from type)

Abstract

Many genomes are incorrectly identified at GenBank. We developed a plan to find and correct misidentified genomes using genomic comparison statistics together with a scaffold of reliably identified genomes from type. A workshop was organized with broad representation from the bacterial taxonomic community to review the proposal, the GenBank Microbial Genomic Taxonomy Workshop, Bethesda MD, May 12–13, 2015.

Keywords: GenBank, Genomic taxonomy, Misidentified sequence entries

ANI_cutoff values in Taxonomy

Acetobacter pasteurianus	92.5
Acinetobacter pittii	92.5
Aeromonas allosaccharophila	95.0
Aeromonas veronii	94.0
Klebsiella michiganensis	93.5
Listeria monocytogenes	92.4
Mycobacterium africanum	99.9
Mycobacterium bovis	99.9
Mycobacterium tuberculosis	99.9
Prochlorococcus marinus	78.0
Raoultella ornithinolytica	96.5
Raoultella planticola	96.5
Rhodococcus fascians	80.0

WRONG CONCLUSION

This is a wrong conclusion derived from considering that some wrongly-labelled strains belong to those species

Some current species concepts span much more (or much less) than the default rule-of-thumb 96% ANI. We can set this value on a ~~species-by-species~~ **genus** basis in the taxonomy database.

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

RESEARCH ARTICLE
Chun *et al.*, *Int J Syst Evol Microbiol* 2018;68:461–466
DOI 10.1099/ijsem.0.002516

January 2018

MICROBIOLOGY SOCIETY
OPEN MICROBIOLOGY

Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes

Jongsik Chun,[1,*] Aharon Oren,[2] Antonio Ventosa,[3] Henrik Christensen,[4] David Ruiz Arahal,[5] Milton S. da Costa,[6] Alejandro P. Rooney,[7] Hana Yi,[8] Xue-Wei Xu,[9] Sofie De Meyer[10] and Martha E. Trujillo[11,*]

Web-services and standalone software tools for taxonomic purposes

Algorithm	Function	Type	URL/Reference
OrthoANI with usearch	Calculation of ANI	Standalone	https://www.ezbiocloud.net/tools/orthoaniu [9]
OrthoANI with usearch	Calculation of ANI	Web service	https://www.ezbiocloud.net/tools/ani [9]
Genome-to-Genome Distance Calculator	Calculation of dDDH	Web service	http://ggdc.dsmz.de/ggdc.php/ [7]
ANI calculator	Calculation of ANI	Web service	http://enve-omics.ce.gatech.edu/ani/
JSpecies	Calculation of ANI	Standalone	http://imedea.uib-csic.es/jspecies/ [5]
JSpeciesWS	Calculation of ANI	Web service	http://jspecies.ribohost.com/ [30]
CheckM	Checking contamination	Standalone	http://ecogenomics.github.io/CheckM/ [29]
ContEst16S	Checking contamination	Web service	https://www.ezbiocloud.net/tools/contest16s [28]
BBMap	Calculation of sequencing depth of coverage	Standalone	https://sourceforge.net/projects/bbmap/
Amphora2	Phylogenomic treeing	Standalone	http://wolbachia.biology.virginia.edu/WuLab/Software.html [21]
BIGSdb	Phylogenomic treeing	Standalone	https://pubmlst.org/software/database/bigsdb/ [31]
bcgTree	Phylogenomic treeing	Standalone	https://github.com/iimog/bcgTree [32]
Phylophlan	Phylogenomic treeing	Standalone	https://huttenhower.sph.harvard.edu/phylophlan[22]
UBCG	Phylogenomic treeing	Standalone	https://www.ezbiocloud.net/tools/ubcg

TABLE 1. Web-services and standalone software tools for taxonomic purposes

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

RESEARCH ARTICLE

Chun *et al.*, *Int J Syst Evol Microbiol* 2018;68:461–466

DOI 10.1099/ijsem.0.002516

MICROBIOLOGY SOCIETY

OPEN MICROBIOLOGY

January 2018

Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes

Jongsik Chun,[1,*] Aharon Oren,[2] Antonio Ventosa,[3] Henrik Christensen,[4] David Ruiz Arahal,[5] Milton S. da Costa,[6] Alejandro P. Rooney,[7] Hana Yi,[8] Xue-Wei Xu,[9] Sofie De Meyer[10] and Martha E. Trujillo[11,*]

Make sure that the quality of a genome sequence is suitable for taxonomic purposes and include:

1. Genome size = is defined as the length sum of all contigs.

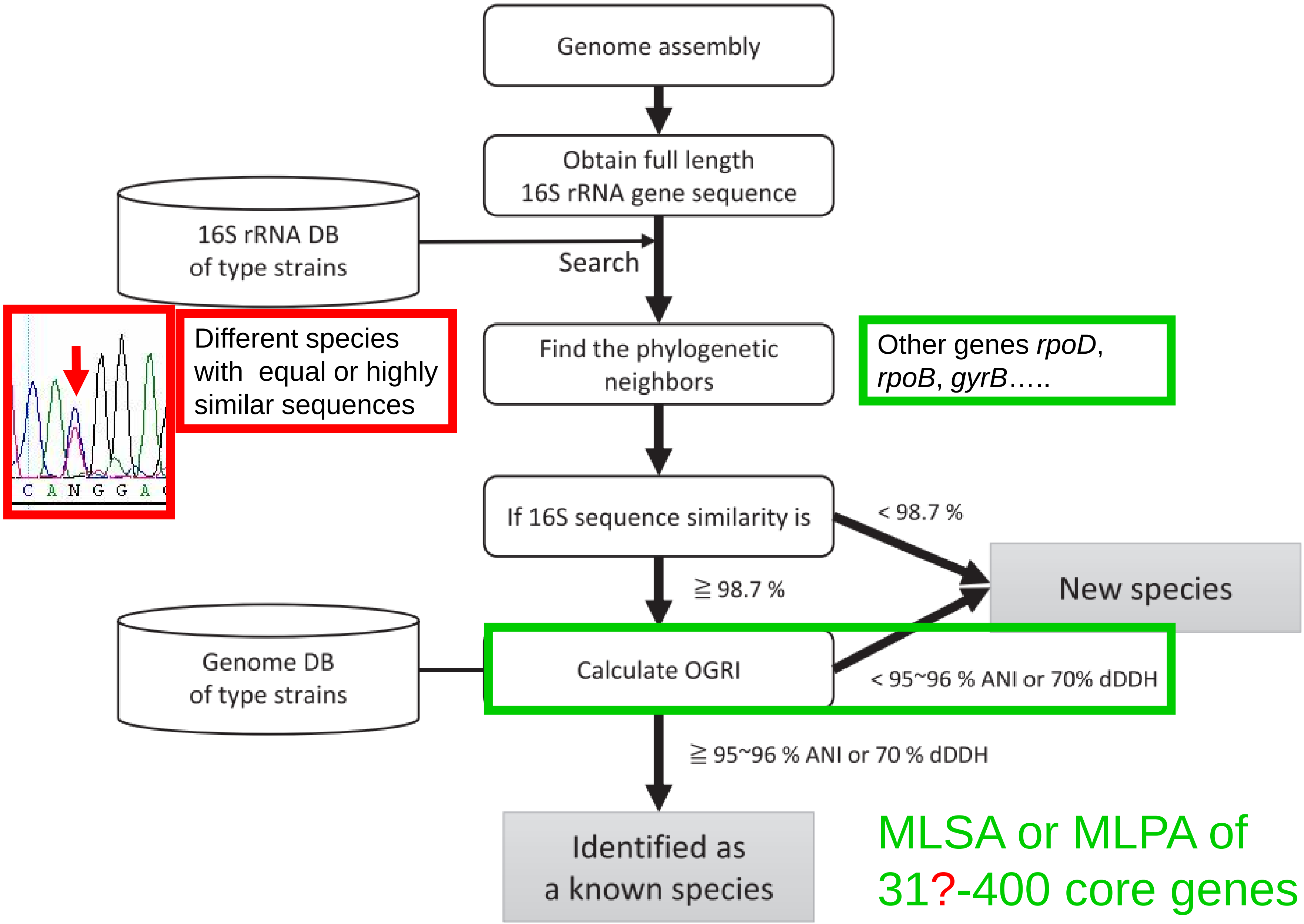
2. Number of contigs and N50.

3. Sequencing depth of coverage.

50X for the currently available NGS platforms (Illumina, Ion Torrent, and Pacific Biosciences) **is recommended**. BBMap estimate the sequencing depth of coverage.

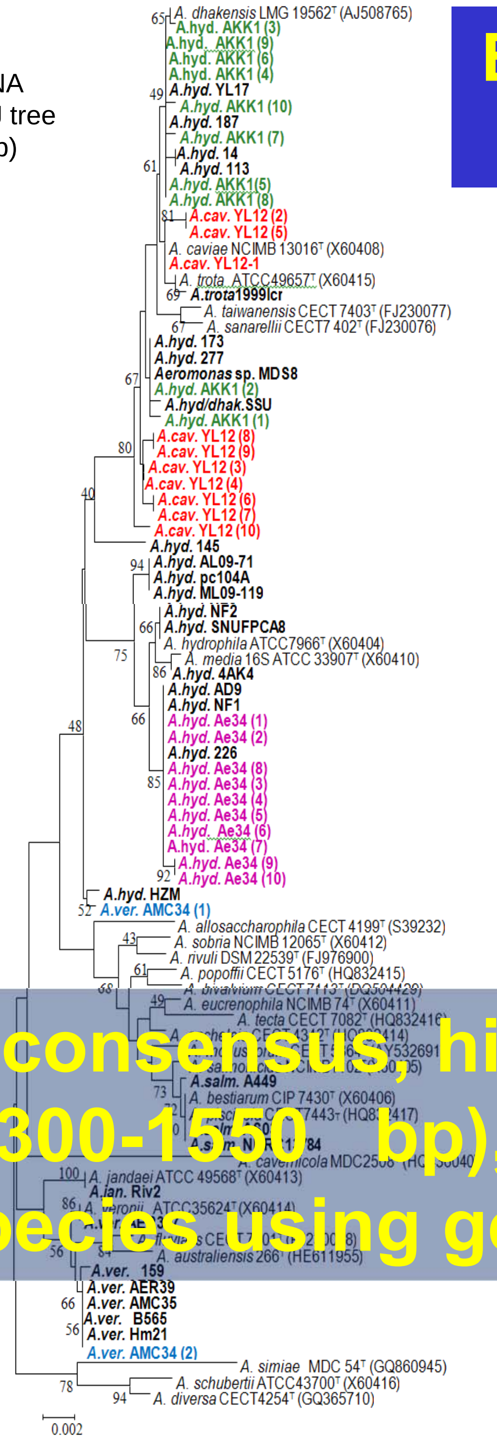
Chun *et al.*, *Int J Syst Evol Microbiol* 2018;68:461–466

Workflow of genome based classification at the species level.



Evaluation of the 16S rRNA gene copies in genomes of *Aeromonas*

16S rRNA gene NJ tree (1503 bp)



Sequenced mistakes or true variability????

Name at the NCBI	16S rRNA gene copies	Base pair size of the 165 rRNA copies (No)
A. veronii AER39	12	1503 (7), 1490, 1341, 1246, 1173
A. hydrophila ATCC7966T	10	1503
A. hydrophila AKK1*	10	1503
A. caviae YL12	10	1503
A. veronii B565	10	1503
A. hydrophila Ae34	10	1503
A. salmonicida A449	9	1503
A. veronii AMC35	9	1503
Aeromonas sp. nov. 4AK4	9	1503
A. dhakensis SSU	8	1503 (2), 1350, 1363, 440, 320, 233, 74
A. hydrophila AL09-71	7	1503
A. salmonicida A503	7	1503
Aeromonas sp. MDS8	6	1503, 956, 639, 439, 68, 28
A. hydrophila HZM*	6	1503, 86, 81, 80, 80, 62
A. veronii Phln2	5	916, 364, 327, 56, 29
A. veronii AMC 34*	4	1503 (2), 174 (2)
A. dhakensis 145	4	1502, 39 (2), 38
A. veronii Hm21	4	1503
A. hydrophila pc104A	3	1503
A. hydrophila ML09-119	3	345, 230, 226
A. hydrophila NF2	2	1503
A. hydrophila AD9	2	1503, 59
A. veronii AER 397	2	1503
A. dhakensis YL17	1	1503
A. hydrophila 187*	1	1503
A. hydrophila 14*	1	1503
A. hydrophila 113	1	1503
A. hydrophila 173*	1	1503
A. hydrophila 277	1	1503
A. hydrophila SNUFPCA8	1	1503
A. hydrophila NF1	1	1503
A. hydrophila 226	1	1503
A. trota 1999lcr	1	1503
A. salmonicida NBR 13784	1	1503
A. salmonicida 34melT	1	1503
A. veronii 159	1	1503
A. jandaei Riv2	1	1503
A. dhakensis AAK1	1	167

*Mislabelled genomes

A consensus, high quality 16S rRNA gen sequence (1300-1550 bp), when describing new bacteria species using genomes should be required

3 new species with complete genomes

2015

Systematic and Applied Microbiology 38 (2015) 161–168

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

ELSEVIER

Aeromonas aquatica sp. nov., *Aeromonas finlandiensis* sp. nov. and *Aeromonas lacus* sp. nov. isolated from Finnish waters associated with cyanobacterial blooms☆

R. Beaz-Hidalgo[a], F. Latif-Eugenín[a], M.J. Hossain[b], K. Berg[c], R.M. Niemi[d], J. Rapala[e], C. Lyra[c], M.R. Liles[b], M.J. Figueras[a,*]

2014

genomeA
Journals.ASM.org

CrossMark

Draft Genome Sequences of Two Novel *Aeromonas* Species Recovered in Association with Cyanobacterial Blooms

Mohammad J. Hossain,[a] Roxana Beaz-Hidalgo,[b] María J. Figueras,[b] Mark R. Liles[a]

Department of Biological Sciences, Auburn University, Auburn, Alabama, USA[a]; Unitat de Microbiologia. Departament de Ciències Mèdiques Bàsiques, Facultat de Medicina i Ciències de la Salut, IISPV, Universitat Rovira i Virgili, Reus, Spain[b]

Aeromonas aquatica and *Aeromonas lacus* are two new species that have been found in association with cyanobacterial blooms from recreational Finnish lakes where adverse human health effects have been recorded. Here, we present the draft genome sequences of their type strains.

Received 2 October 2014 Accepted 16 October 2014 Published 20 November 2014

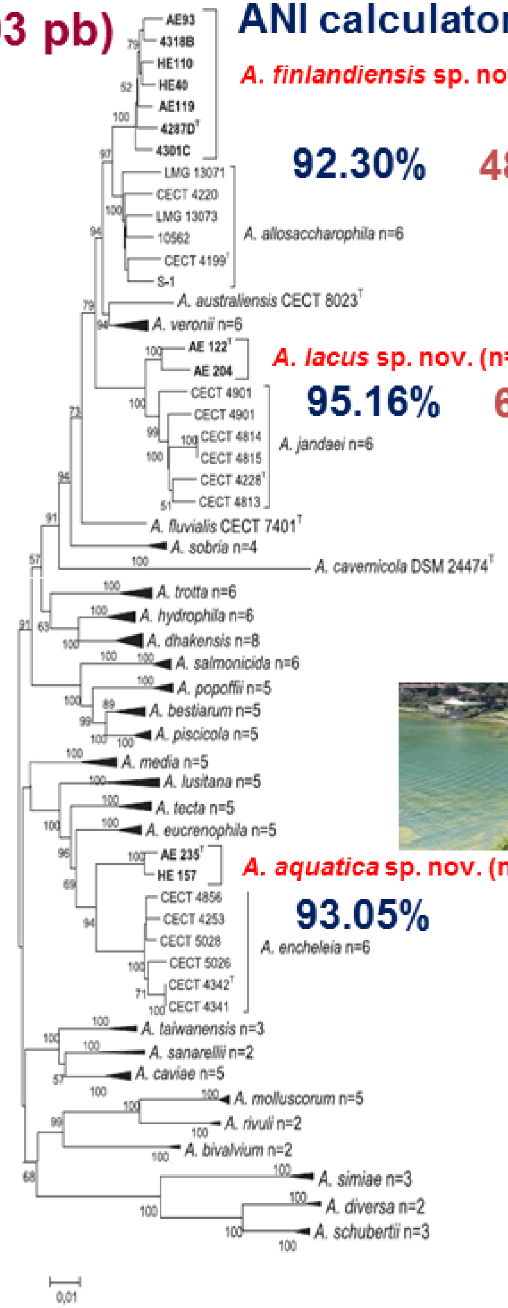
FADUA LEILA LATIF EUGENÍN

Tesis Doctoral

2015

Prof. Mark Liles

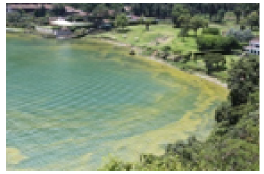
MLPA (4093 pb)



ANI calculator & *is*DDH

A. finlandiensis sp. nov. (n= 7)

92.30% 48.40%



A. lacus sp. nov. (n= 2)

95.16% 63.20%

A. aquatica sp. nov. (n= 2)

93.05% 50.50%

MLPA = Multilocus Phylogenetic Analysis

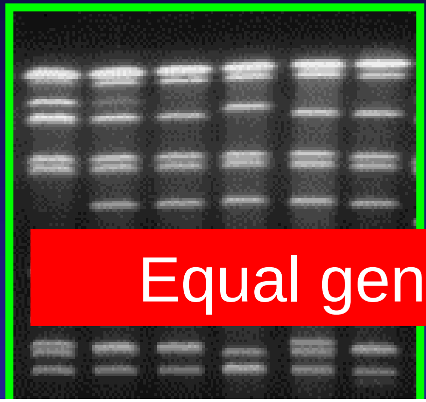
OBJECTIVES

- To present the strategies to determine the species and **strain identity** using the genome information

- To underline additional information such antibiotic resistance and virulence genes that can be found in the genomes

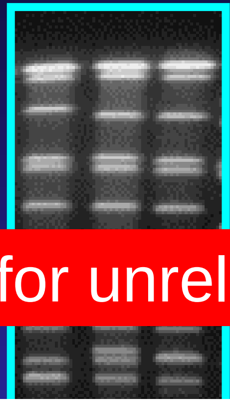
Limitations of the epidemiological typing methods

Environmental strains



=

Clinical strains



Pulsed-Field Gel Electrophoresis (PFGE)

Equal genotypes for unrelated strains

GOLD STANDARD

E·W·G·L·I

The European Working Group for Legionella Infections

Legionella pneumophila Sequence-Based Typing

Sequence Base Typing database (v3.0) http://www.ewgli.org

ST	Allelic profile[a]	N	%
ST47	5, 10, 22, 15, 6, 2, 6	24	27.9
ST1	1, 4, 3, 1, 1, 1, 1	17	19.8

PARIS

Equal ST in unrelated strains

Multilocus Sequence Typing (MLST) - 7 genes

virulence

housekeeping

Number of ***flaA*** alleles:	38
Number of ***pilE*** alleles:	53
Number of ***asd*** alleles:	73
Number of ***mip*** alleles:	85
Number of ***mompS*** alleles:	96
Number of ***proA*** alleles:	55
Number of ***neuA*** alleles:	65
Number of ***neuAh*** alleles:	30

Gaia *et al*., 2005; Ratzow *et al*., 2007; Qin *et al*., 2014; David *et al*., 2016

Whole-Genome Sequencing (WGS)

RESEARCH ARTICLE — Open Acces

Comparison of the *Legionella pneumophila* population structure as determined by sequence-based typing and whole genome sequencing

Anthony P Underwood[1*], Garan Jones[1,3], Massimo Mentasti[2], Norman K Fry[2] and Timothy G Harrison[2]

American Society for Microbiology | Journal of Clinical Microbiology

CrossMark

Evaluation of an Optimal Epidemiological Typing Scheme for *Legionella pneumophila* with Whole-Genome Sequence Data Using Validation Guidelines

Sophia David,[a,b] Massimo Mentasti,[b] Rediat Tewolde,[b] Martin Aslett,[a] Simon R. Harris,[a] Baharak Afshar,[b,c] Anthony Underwood,[b] Norman K. Fry,[b] Julian Parkhill,[a] Timothy G. Harrison[b]

Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom[a]; Public Health England, London, United Kingdom[b]; The European Programme for Public Health Microbiology Training, European Centre for Disease Prevention and Control, Stockholm, Sweden[c]

Clinical Infectious Diseases

MAJOR ARTICLE

IDSA Infectious Diseases Society of America | hivma hiv medicine association | OXFORD

Seeding and Establishment of *Legionella pneumophila* in Hospitals: Implications for Genomic Investigations of Nosocomial Legionnaires' Disease

Sophia David,[1,2] Baharak Afshar,[2,3] Massimo Mentasti,[2] Christophe Ginevra,[4,5] Isabelle Podglajen,[6] Simon R. Harris,[1] Victoria J. Chalker,[2] Sophie Jarraud,[4,5] Timothy G. Harrison,[2] and Julian Parkhill[1]

[1]Pathogen Genomics, Wellcome Trust Sanger Institute, Cambridge, and [2]Respiratory and Vaccine Preventable Bacteria Reference Unit, Public Health England, London, United Kingdom; [3]European

American Society for Microbiology | Applied and Environmental Microbiology

CrossMark

Genomic Resolution of Outbreak-Associated *Legionella pneumophila* Serogroup 1 Isolates from New York State

Brian H. Raphael,[a] Deborah J. Baker,[b] Elizabeth Nazarian,[b] Pascal Lapierre,[b] Dianna Bopp,[b] Natalia A. Kozak-Muiznieks,[a] Shatavia S. Morrison,[a] Claressa E. Lucas,[a] Jeffrey W. Mercante,[a] Kimberlee A. Musser,[b] Jonas M. Winchell[a]

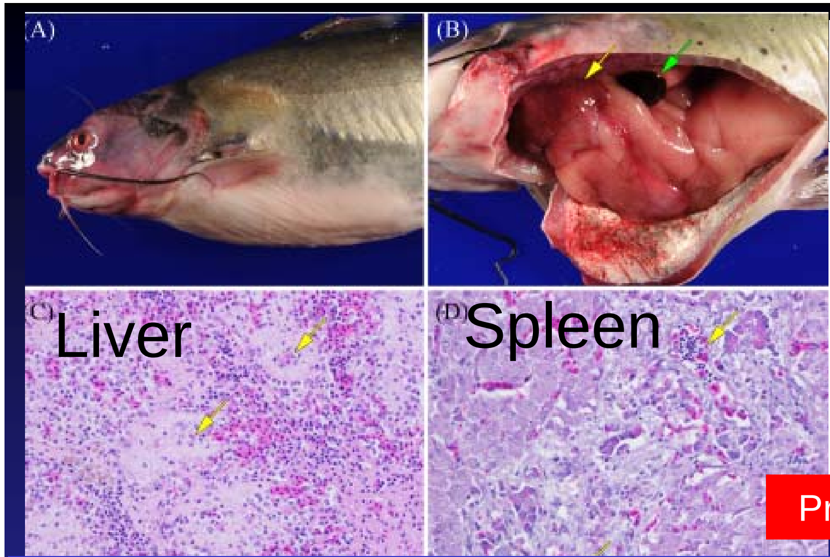
- WGS is more discriminatory than SBT and PFGE
- Extended MLST scheme with ± 50 genes provides optimal epidemiological concordance
- Phylogenetic analyses of whole genome (wg) or core genome (cg) Single-Nucleotide Polymorphism (SNP)

There is a need for efficient, easy to handle and to interpret bioinformatic tools

Salipante *et al*., 2015; Raphael *et al*., 2016; David *et al*., 2016; 2017;

Core genome = shared genome

- **ANI can provide a similar resolution as other tools based on the WGS for recognizing strains involved in outbreaks**



frontiers in Microbiology

UNIVERSITAT ROVIRA I VIRGILI

ORIGINAL RESEARCH
published: 18 October 2016
doi: 10.3389/fmicb.2016.01615

2016

Classification of a Hypervirulent *Aeromonas hydrophila* Pathotype Responsible for Epidemic Outbreaks in Warm-Water Fishes

Cody R. Rasmussen-Ivey[1], Mohammad J. Hossain[1], Sara E. Odom[1], Jeffery S. Terhune[2], William G. Hemstreet[3], Craig A. Shoemaker[4], Dunhua Zhang[4], De-Hai Xu[4], Matt J. Griffin[5], Yong-Jie Liu[6], Maria J. Figueras[7], Scott R. Santos[1], Joseph C. Newton[8] and Mark R. Liles[1]**

Prof. Mark Liles

Average Nucleotide Identity of *A. hydrophila* genomes vs. genome of the epidemic strain ML09-119

JSpecies Taxonomic Thresholds

Goris *et al*., 2007

Richter & Rosselló-Mora, 2009

Species	Isolates	ANI(%)
A. hydrophila	ML09-121	99,99
A. hydrophila	ML09-122	99,99
A. hydrophila	S04-690	99,99
A. hydrophila	ZC1	99,98
A. hydrophila	PB10-118	99,99
A. hydrophila	AL10-121	99,99
A. hydrophila	AL09-79	99,99
A. hydrophila	ATCC7966T	97,14
A. hydrophila	226	97,13
A. hydrophila	AL06-06	97,13
A. hydrophila	E1	97,12
A. hydrophila	E2	97,12
A. hydrophila	SNUFPCA8	97,11
A. hydrophila	MN98-04	97,09
A. hydrophila	AL97-91	97,08
A. hydrophila	AL06-01	97,08
A. hydrophila	TN97-08	97,07
A. hydrophila	145	93,84
A. hydrophila	277	93,76
A. hydrophila	SSU	93,75
A. hydrophila	173	93,72
A. hydrophila	187	93,7
A. hydrophila	259	93,7
A. hydrophila	GA97-22	89,94
A. hydrophila	4AKA	88,48

99.99% clone

Aeromonas hydrophila

>97%

ANI < 96% different species

88.48 - 93.84%

~~*Aeromonas hydrophila*~~

OAT
Orthologous Average Nucleotide Identity Tool
a similarity measurement tool for genomes

Lee *et al*., (2015)

http://www.ezbiocloud.net/tools/orthoani

Orthologous Average Nucleotide Identity Tool

EZ BioCloud DASHBOARD IDENTIFY TOOLS RESOURCES HOW TO CITE ABOUT SUPPORT Log In

Requirements

- Java Runtime Environment Version 8 (**Java Download**)
- NCBI BLAST is required if you are using the Runnable JAR version of OAT. We recommend ncbi-blast-2.2.30+ or higher since OAT was tested with ncbi-blast-2.2.30+ (**BLAST+ executables**)

Download OAT

- OAT standalone

OAT Runnable JAR	Download 64 bit
OAT for Windows OS	Download 64 bit
OAT User Manual	Download PDF

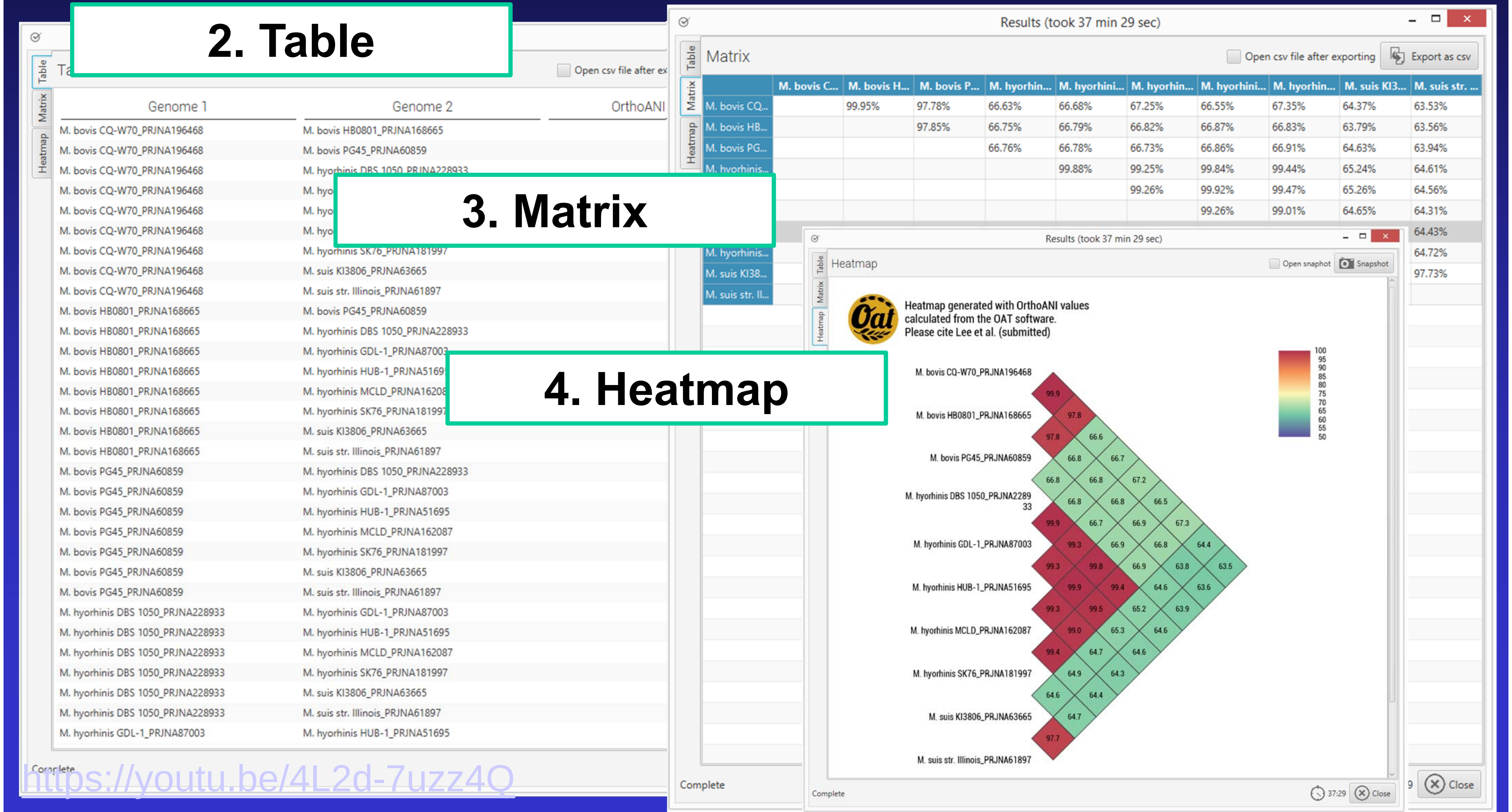
- OAT command line

Lee et al., (2015)

OAT
Orthologous Average Nucleotide Identity Tool
a similarity measurement tool for genomes

http://www.ezbiocloud.net/tools/orthoani

Orthologous Average Nucleotide Identity Tool



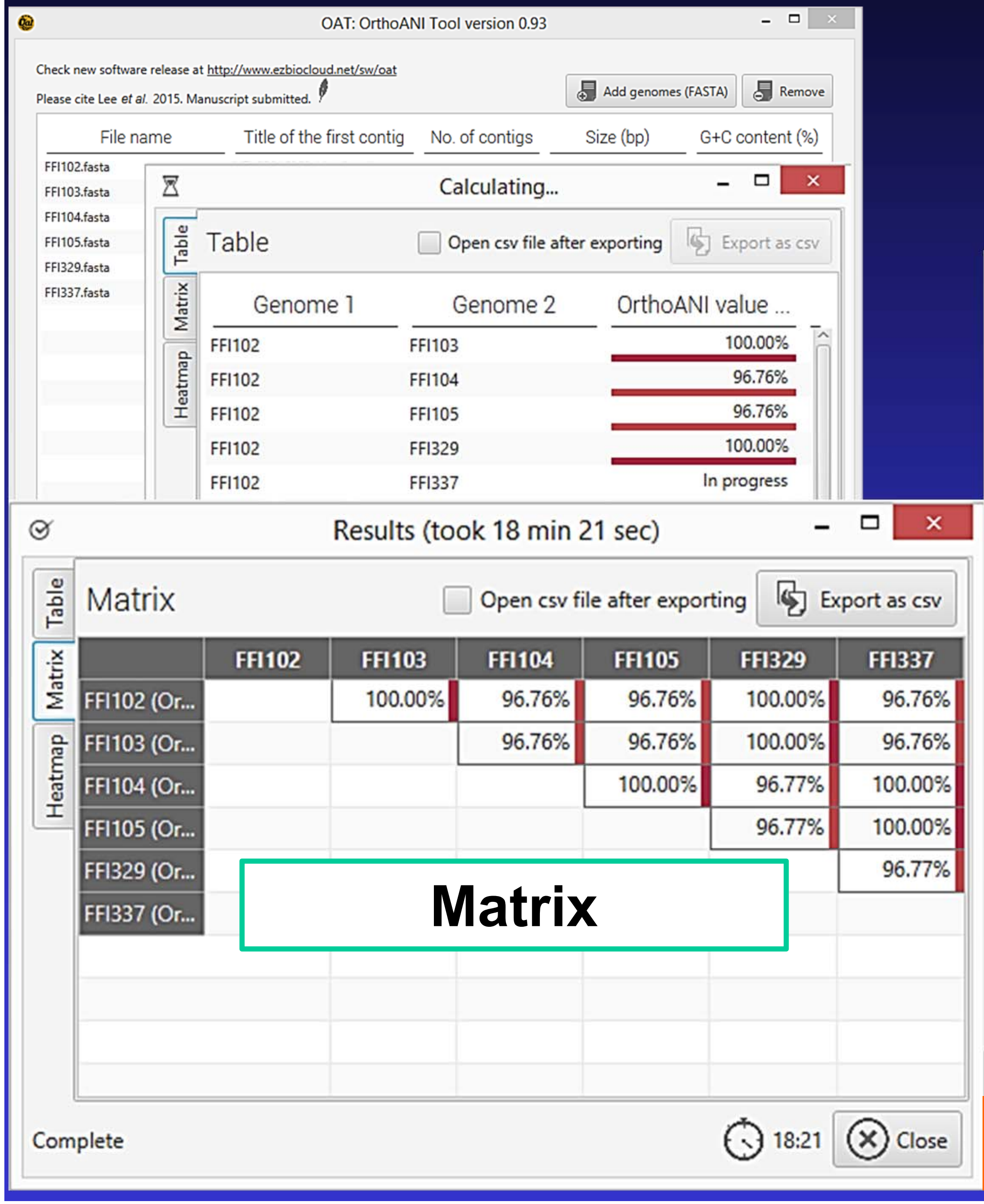
https://youtu.be/4L2d-7uzz4Q

Lee *et al.*, (2015)

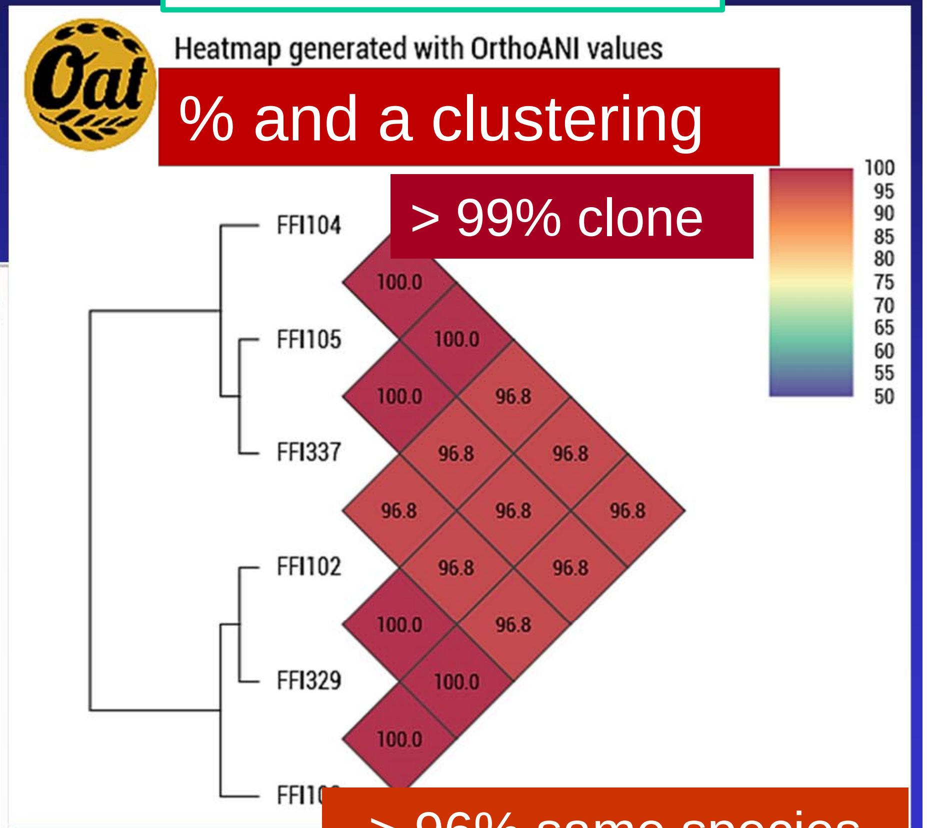
OAT
Orthologous Average Nucleotide Identity Tool
a similarity measurement tool for genomes

http://www.ezbiocloud.net/tools/orthoani

Orthologous Average Nucleotide Identity Tool



Heatmap



% and a clustering

> 99% clone

> 96% same species

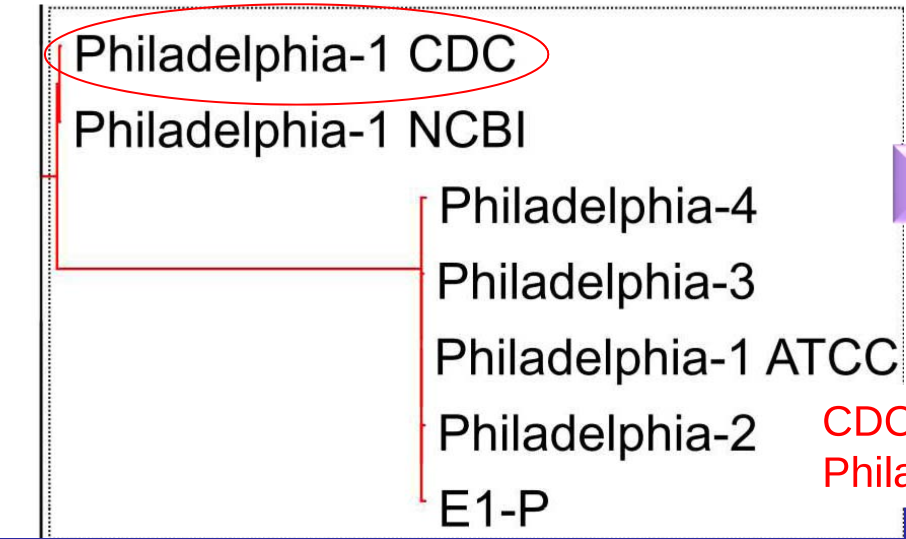
ANI < 96% different species

Outbreak Philadelphia ST36

PLOS ONE

Mercante *et al*., 2016

Maximum-likelihood tree based on 11,356 core SNPs



ST36

RESEARCH ARTICLE

Genomic Analysis Reveals Novel Diversity among the 1976 Philadelphia Legionnaires' Disease Outbreak Isolates and Additional ST36 Strains

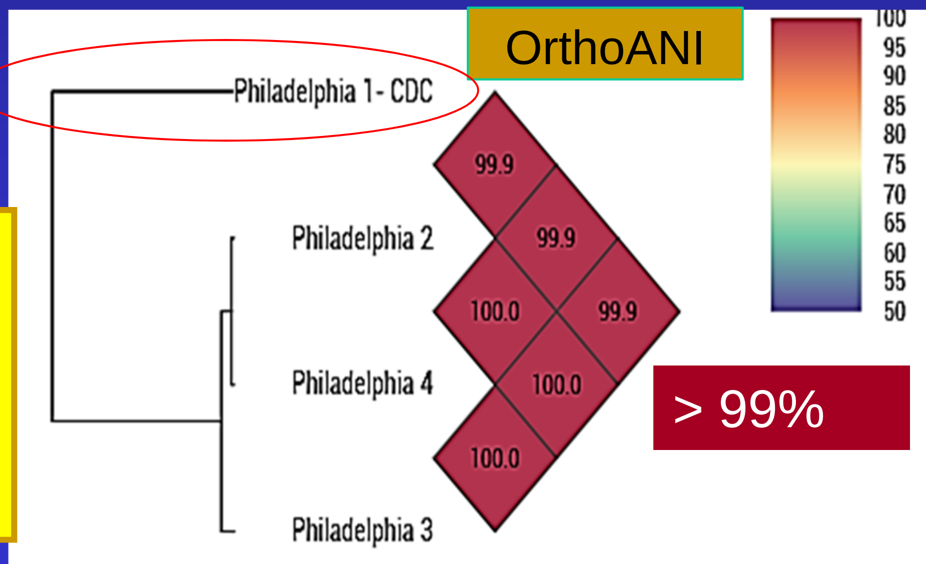
Jeffrey W. Mercante, Shatavia S. Morrison, Heta P. Desai, Brian H. Raphael, Jonas M. Winchell*

Pneumonia Response and Surveillance Laboratory, Respiratory Diseases Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America

* jwinchell@cdc.gov

CDC genome set apart from the historical outbreak Philadelphia 2, 3 and 4 genomes

ANI results could distinguish the small genetic differences among ST36 Philadelphia outbreak strains



> 99%

frontiers in Microbiology

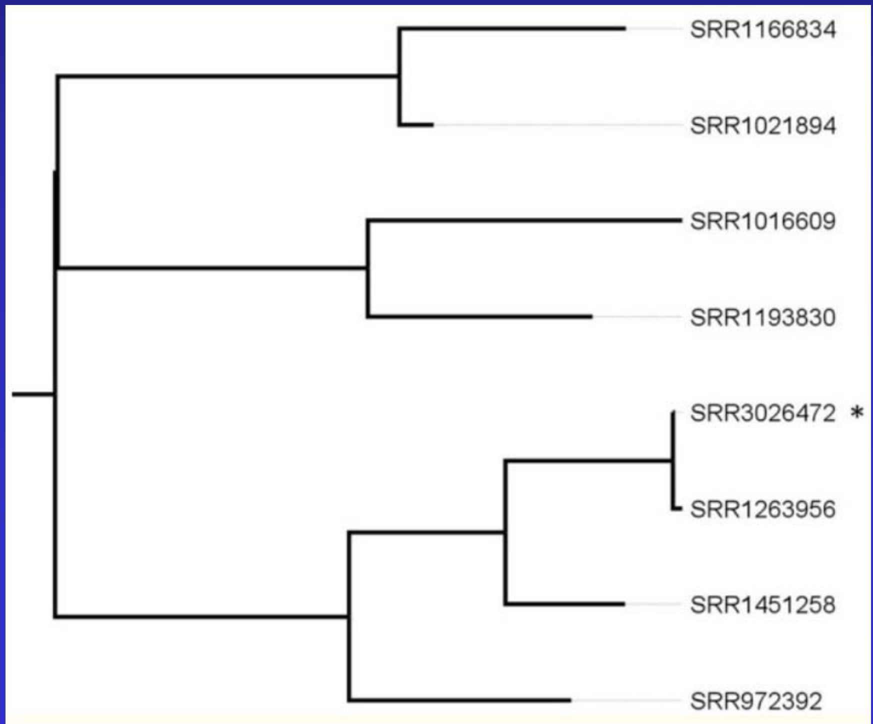
Front Microbiol. 2018; 9: 446.
Published online 2018 Mar 14. doi: 10.3389/fmicb.2018.00446

PMCID: PMC5861
PMID: 29593

A Validation Approach of an End-to-End Whole Genome Sequencing Workflow for Source Tracking of *Listeria monocytogenes* and *Salmonella enterica*

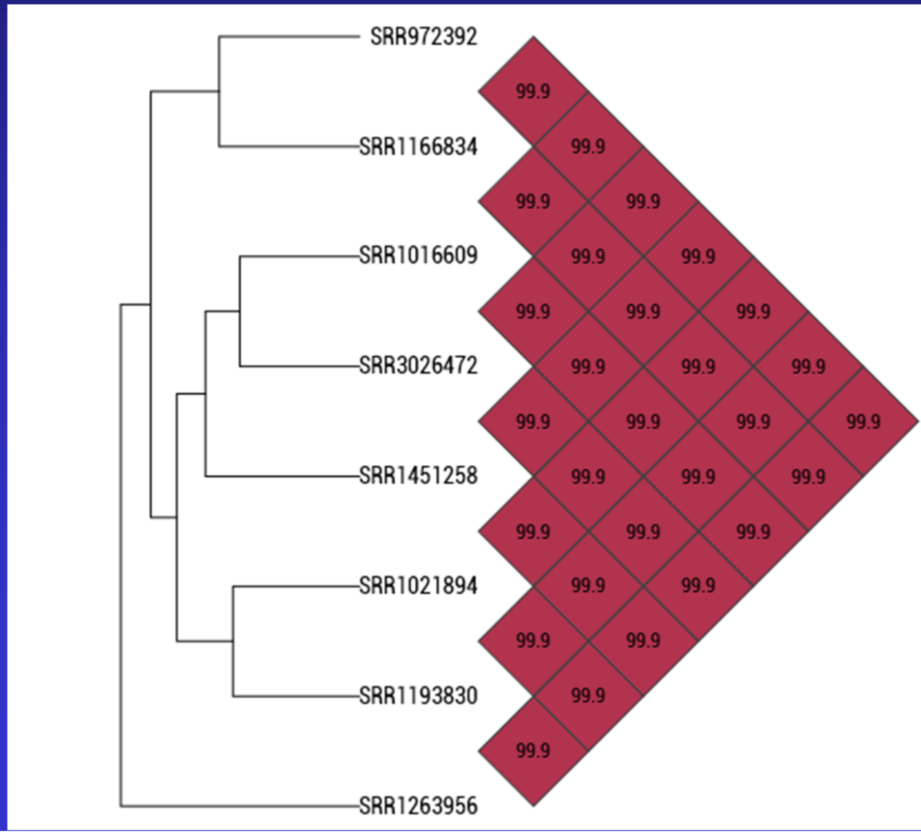
Anne-Catherine Portmann,[1] Coralie Fournier,[2] Johan Gimonet,[1] Catherine Ngom-Bru,[1] Caroline Barretto,[1] and Leen Baert[1,*]

Phylogenetic tree based on SNP differences from selected patient isolates and the lettuce reference isolate*



Lee *et al.*, (2015)

OAT
Orthologous Average Nucleotide Identity Tool
a similarity measurement tool for genomes

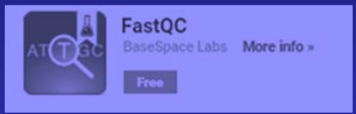


CHIMERIC OR CONTAMINATED GENOMES

Genome Analysis

Quality control of identity and possible contamination

FastQ → Assembly → Annotation



MSR

RAST server/sftw

Search for HK genes for a MLPA (BLASTn)
Comparison with the previously sequenced genes

WGS

Alba Pérez-Cataluña

HK = housekeeping genes

MLPA = Multilocus Phylogenetic Analysis

Method

CheckM

CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes

Donovan H. Parks,[1] Michael Imelfort,[1] Connor T. Skennerton,[1] Philip Hugenholtz,[1,2] and Gene W. Tyson[1,3]

2017

OPEN

The ISME Journal (2016) 10, 269–272

www.nature.com/ismej

SHORT COMMUNICATION

ProDeGe: a computational protocol for fully automated decontamination of genomes

Kristin Tennessen[1], Evan Andersen[1], Scott Clingenpeel[1], Christian Rinke[1], Derek S Lundberg[2], James Han[1], Jeff L Dangl[3], Natalia Ivanova[1], Tanja Woyke[1], Nikos Kyrpides[1] and Amrita Pati[1]

2016

Maruyama et al. BMC Bioinformatics (2017) 18:152
DOI 10.1186/s12859-017-1572-5

BMC Bioinformatics

SOFTWARE

SAG-QC: quality control of single amplified genome information by subtracting non-target sequences based on sequence compositions

Toru Maruyama[1,2], Tetsushi Mori[3], Keisuke Yamagishi[1] and Haruko Takeyama[1,2,3*]

2017

EZBioCloud

ContEst16s:

Contamination Estimator by 16S

http://tool.ezbiocloud.net/contest16s/

ContEst16S

2017

RESEARCH ARTICLE

Lee et al., Int J Syst Evol Microbiol 2017;67:2053–2057
DOI 10.1099/ijsem.0.001872

MICROBIOLOGY SOCIETY

ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences

Imchang Lee,[1] Mauricio Chalita,[2] Sung-Min Ha,[1,3] Seong-In Na,[2] Seok-Hwan Yoon[1,3] and Jongsik Chun[1,2,3,*]

Genome: *Arcobacter anaerophilus* DSM24636T

ContEst16s:

Contamination Estimator by 16S

http://tool.ezbiocloud.net/contest16s/

EZBioCloud ContEst16S

1

Decision

Contaminated

By comparing 4 16S rRNA gene fragments, this project has been confirmed for contamination.
Maximum difference between fragments is 27.3%.

2

Taxonomic identification of extracted fragments

Fragment#	Length(bp)	Top hit	Similarity(%)	Difference / Compared (bp)	Taxonomy
Fragment 1	1,517	Arcobacter anaerophilus JC84(T)	99.4	9/1385	Bacteria;Proteobacteria;Epsilonproteobacteria;Campylobacterales;Campylobacteraceae;Arcobacter;Arcobacter anaerophilus
Fragment 2	1,517	Arcobacter molluscorum F98-3(T)	99.9	2/1401	Bacteria;Proteobacteria;Epsilonproteobacteria;Campylobacterales;Campylobacteraceae;Arcobacter;Arcobacter molluscorum
Fragment 3	810	Bacillus thuringiensis ATCC 10792(T)	100	0/775	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;Bacillus thuringiensis
Fragment 4	763	Bacillus thuringiensis ATCC 10792(T)	100	0/763	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;Bacillus thuringiensis

Fragment 4
73.6
72.7
Insufficient alignment
0

A dash "-" indicates insufficient alignment.

WGS

Alba Pérez-Cataluña

DTU Food
National Food Institute

Technical Univeristy of Denmark

Center for Genomic Epidemiology

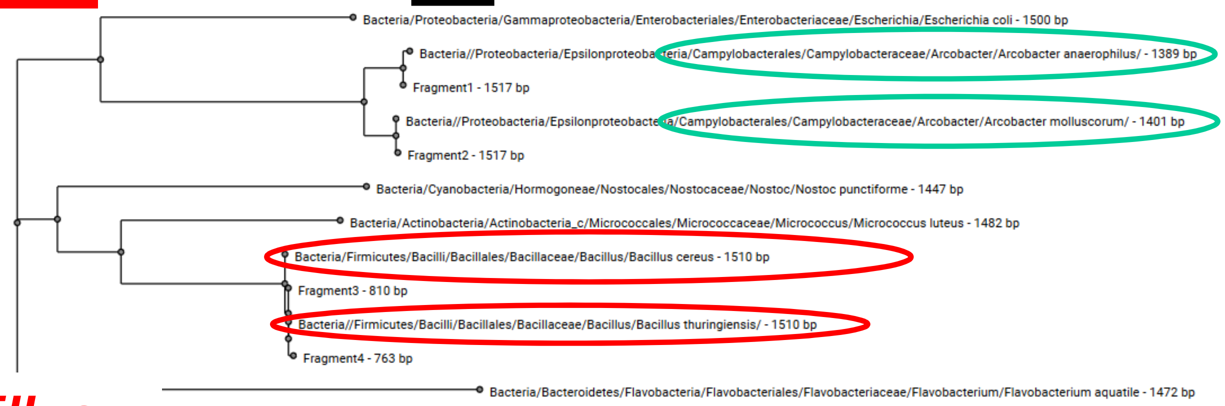
Home Services Instructions Output

KmerFinder 2.0 Usage Instructions

Hit	Score	z-score	Query Coverage [%]	Template Coverage [%]	Depth	Total Query Coverage [%]	Total Template Coverage [%]	Total Depth
Bacillus thuringiensis, Bacillus thuringiensis HD-771 get sequence	239	44.1	6.48	2.73	0.03	6.48	2.73	0.03
Arcobacter nitrofigilis, Arcobacter nitrofigilis DSM 7299 get sequence	120	39.8	3.26	3.26	0.03	3.36	3.37	0.03
Arcobacter sp. L get sequence	101	41.2	2.74	3.74	0.04	3.36	4.59	0.05

Maximum likelihood phylogenetic tree

3



Contamination with *Bacillus*

Dr. J Chun CHUNLAB

Seoul National University

Seoul, South Korea

Bergey's International Society for Microbial Systematics

BISMiS 2016

third meeting on

Microbial Systematics and Metagenomics

at

MICROBIAL CULTURE COLLECTION, PUNE

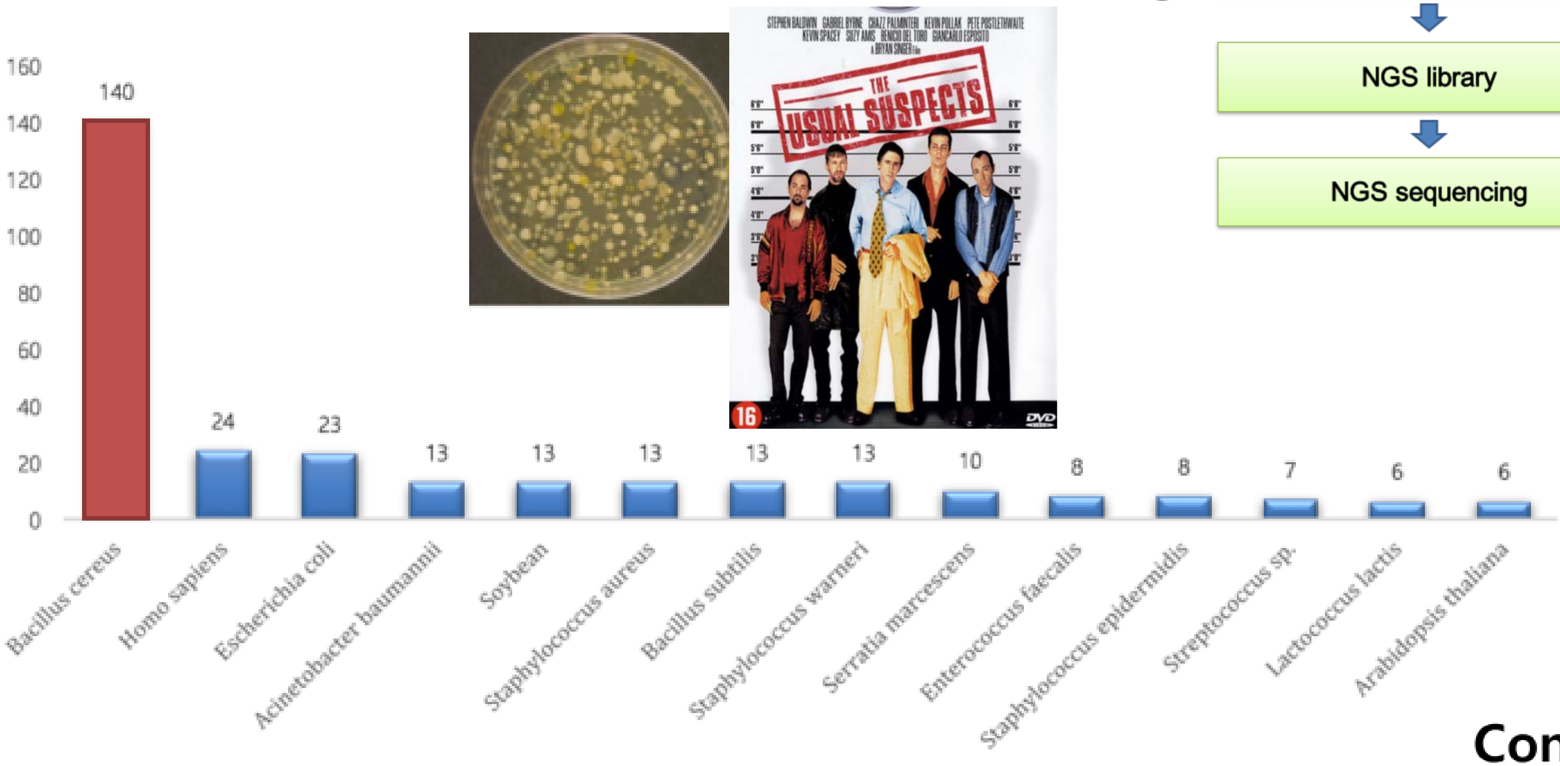
Pune पुणे

September 12- 15, 2016

Contaminated genomes in public database

Bacillus cereus (13.6%)

Common contaminant while isolation & culturing



When was it contaminated?

Isolation & Culturing	Microbiology Lab.
↓ DNA extraction	Microbiology Lab.
↓ NGS library	Sequencing Lab.
↓ NGS sequencing	Sequencing Lab.

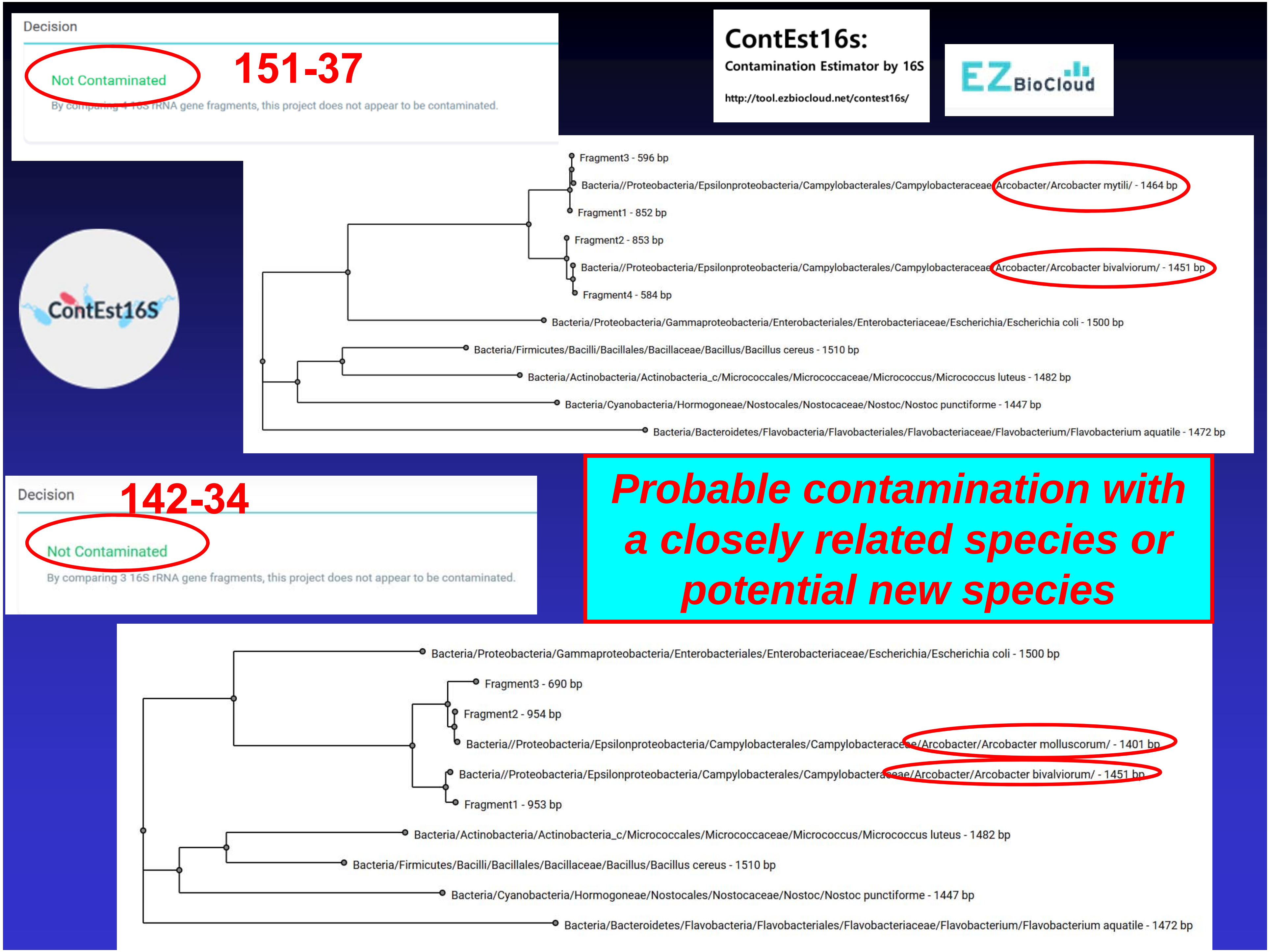
EZBioCloud

ContEst16S

ContEst16s:

Contamination Estimator by 16S

http://tool.ezbiocloud.net/contest16s/



Decision
Not Contaminated
151-37
By comparing 4 16S rRNA gene fragments, this project does not appear to be contaminated.
ContEst16s:
Contamination Estimator by 16S
http://tool.ezbiocloud.net/contest16s/
EZBioCloud
ContEst16S
Fragment3 - 596 bp
Bacteria//Proteobacteria/Epsilonproteobacteria/Campylobacterales/Campylobacteraceae/Arcobacter/Arcobacter mytili/ - 1464 bp
Fragment1 - 852 bp
Fragment2 - 853 bp
Bacteria//Proteobacteria/Epsilonproteobacteria/Campylobacterales/Campylobacteraceae/Arcobacter/Arcobacter bivalviorum/ - 1451 bp
Fragment4 - 584 bp
Bacteria/Proteobacteria/Gammaproteobacteria/Enterobacteriales/Enterobacteriaceae/Escherichia/Escherichia coli - 1500 bp
Bacteria/Firmicutes/Bacilli/Bacillales/Bacillaceae/Bacillus/Bacillus cereus - 1510 bp
Bacteria/Actinobacteria/Actinobacteria_c/Micrococcales/Micrococcaceae/Micrococcus/Micrococcus luteus - 1482 bp
Bacteria/Cyanobacteria/Hormogoneae/Nostocales/Nostocaceae/Nostoc/Nostoc punctiforme - 1447 bp
Bacteria/Bacteroidetes/Flavobacteria/Flavobacteriales/Flavobacteriaceae/Flavobacterium/Flavobacterium aquatile - 1472 bp
Decision
142-34
Not Contaminated
By comparing 3 16S rRNA gene fragments, this project does not appear to be contaminated.
Probable contamination with a closely related species or potential new species
Bacteria/Proteobacteria/Gammaproteobacteria/Enterobacteriales/Enterobacteriaceae/Escherichia/Escherichia coli - 1500 bp
Fragment3 - 690 bp
Fragment2 - 954 bp
Bacteria//Proteobacteria/Epsilonproteobacteria/Campylobacterales/Campylobacteraceae/Arcobacter/Arcobacter molluscorum/ - 1401 bp
Bacteria//Proteobacteria/Epsilonproteobacteria/Campylobacterales/Campylobacteraceae/Arcobacter/Arcobacter bivalviorum/ - 1451 bp
Fragment1 - 953 bp
Bacteria/Actinobacteria/Actinobacteria_c/Micrococcales/Micrococcaceae/Micrococcus/Micrococcus luteus - 1482 bp
Bacteria/Firmicutes/Bacilli/Bacillales/Bacillaceae/Bacillus/Bacillus cereus - 1510 bp
Bacteria/Cyanobacteria/Hormogoneae/Nostocales/Nostocaceae/Nostoc/Nostoc punctiforme - 1447 bp
Bacteria/Bacteroidetes/Flavobacteria/Flavobacteriales/Flavobacteriaceae/Flavobacterium/Flavobacterium aquatile - 1472 bp

Tools to compare genomes to determine their taxonomic relatedness

ANI and isDDH produce consistent results

Genome 1	Genome 2	ANI	isDDH
Contaminated	Type strain	97.9%	81.7%
Correct	Type strain	97.8%	81.7%

OBJECTIVES

- To present the strategies to determine the species and strain identity using the genome information

- To underline additional information such antibiotic resistance and virulence genes that can be found in the genomes

Genome Analysis

Virulence genes were searched by BLASTn:

- Virulence Factors of Pathogenic Bacteria Database (VFDB) (Chen et al., 2005)
- **Victors Database** (University of Michigan, USA)
- **PATRIC_VF** (Wattam et al., 2017).
- **Search for specific genes by BLASTp analysis**

RAST Rapid Annotation using Subsystem Technology version 2.0
The NMPDR, SEED-based, prokaryotic genome annotation service.
For more information about The SEED please visit theSEED.org.

Antibiotic resistance genes:

- Antibiotic Resistance Database (ARDB) (Liu and Pop, 2009)
- Comprehensive Antibiotic Resistance Database (CARD) (Jia et al., 2017).
- Antibiotic Resistance Gene-Annotation database (ARG-ANNOT) (Gupta et al., 2014)

frontiers in Microbiology

ORIGINAL RESEARCH
published: xx April 2018
doi: 10.3389/fmicb.2018.00805

A Polyphasic and Taxogenomic Evaluation Uncovers *Arcobacter cryaerophilus* as a Species Complex That Embraces Four Genomovars

Alba Pérez-Cataluña[1], Luis Collado[2], Oscar Salgado[2,3], Violeta Lefiñanco[2] and María J. Figueras[1*]*

a) RAST/PATRIC results, b) ARG–ANNOT results, c) BLASTn of virulence genes results, d,b)-lactamase class D, e) Phospholipase A and C.

	CLUSTERS I								II		III	IV	
	1	2	3	4	5	6	7	8	9	10	11	12	13
ANTIBIOTIC RESISTANCE													
Multidrug efflux pumps													
CmeABC system[a]	+	+	+	+	+	+	+	+	+	+	+	+	+
MFS Superfamily[a]	−	+	−	+	+	+	+	+	+	−	−	−	−
Macrolids													
MacAB–TolC[a]	+	+	+	+	+	+	+	+	+	+	+	+	+
Quinolones													
gyrA mutation	−	−	−	−	−	−	−	−	−	−	−	−	−
23S rRNA mutations	−	−	−	−	−	−	−	−	−	−	−	−	−
OqxB[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
β-lactamics													
β-lactamase[a]	−	−	−	−	−	−	−	+[d]	−	−	−	−	−
Colistin													
Mcr–1[b]	+	−	+	+	+	+	+	+	+	−	+	−	+
Mcr–2[b]	+	−	+	+	+	+	+	+	+	−	+	−	+
Acriflavin resistance[a]	+	+	+	+	+	+	+	+	+	+	+	+	+
Streptomycin/Spectinomycin[a]	−	−	−	−	−	−	−	−	−	−	−	+	−

	CLUSTERS I								II		III	IV	
	1	2	3	4	5	6	7	8	9	10	11	12	13
VIRULENCE FACTORS													
Invasion													
ciaB[c]	+	+	+	+	+	+	+	+	+	+	+	+	+
mviN[c]	+	+	+	+	+	+	+	+	+	+	+	+	+
Adhesion													
cj1349[c]	+	+	+	+	+	+	+	+	+	+	+	+	+
cadF[c]	−	−	−	−	−	−	−	−	−	−	−	−	−
Filamentous hemmaglutinin													
hecA[c]	−	−	−	−	−	−	−	−	−	−	−	−	−
Hemolysis													
hecB[c]	−	−	−	−	−	−	−	−	−	−	−	−	−
tlyA[c]	−	−	−	−	−	−	−	−	+	−	+	−	−
Outer membarne protein													
irgA[c]	−	−	−	+	−	−	+	−	−	−	−	−	−
Phospholipase													
pldA[c]	+	+	+	+	+	+	+	+	+	+	+	+[e]	+

PulseNet International

The International Molecular Subtyping Network for Foodborne Disease Surveillance

Africa, Asia Pacific, Canada, Europe, Latin America & Carribbean, Middle East, USA

http://www.pulsenetinternational.org/

PulseNet International

Foodborne illnesses do not respect any borders. As a result of increasing international trade, food produced in one country may be consumed in a different part of the world and cause disease if contaminated with a foodborne pathogen.

Similarly, international travel is increasing and it is possible to get to almost any destination from almost any place in the world in a matter of hours. Therefore, a disease contracted in one part of the world may first become apparent thousands of miles away.

PulseNet International is a network of National and regional laboratory networks dedicated to tracking foodborne infections world-wide. Each laboratory utilizes standardized genotyping methods, sharing information in real-time.

The resulting surveillance provides early warning of food and waterborne disease outbreaks, emerging pathogens, and acts of bioterrorism.

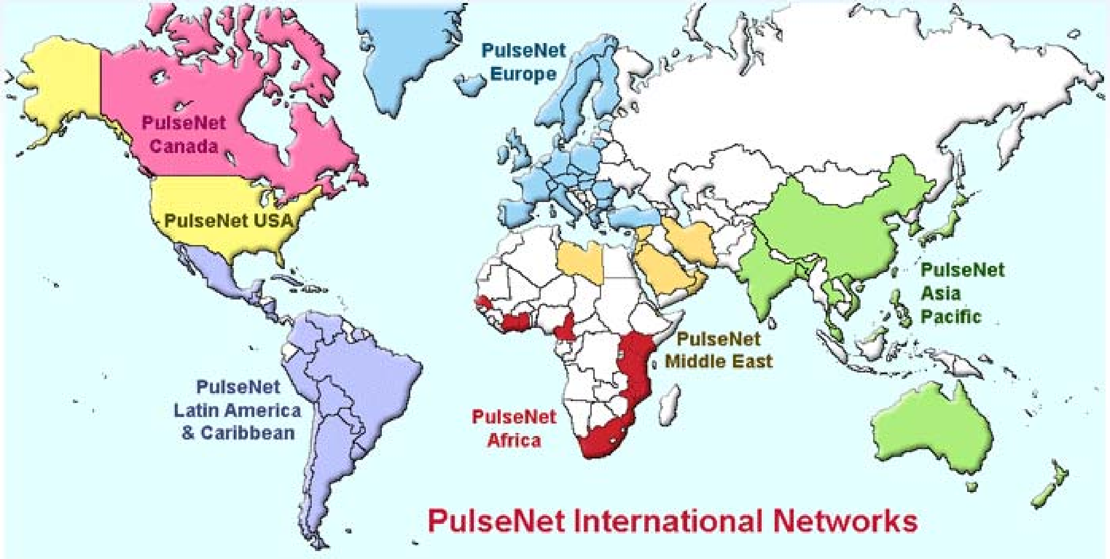
PulseNet International

The International Molecular Subtyping Network for Foodborne Disease Surveillance

USA, Canada, Europe, Asia Pacific, Latin America, Middle East

CDC

New!! Whole Genome Sequencing & PulseNet



What's available on this site?

Key tools & collaborators

We are looking to bring together the best tools and approaches from around the world, and implement them within the PulseNet network. Key collaborators or tools we are exploring include:

- Center for Genomic Epidemiology (Denmark)
- Integrated Rapid Infectious Disease Analysis – IRIDA(Canada)
- Whole genome multi locus sequence typing (wgMLST) (Applied Maths, Belgium)

Frequently Asked Questions

What is Whole Genome Sequencing (WGS)?

1. WGS is the output and the process of generating the full DNA sequence of the genome of a microorganism. For foodborne bacteria, the genome includes the chromosome and any extrachromosomal genomic material such as plasmids. The actual process is also called next generation sequencing (NGS) and is performed by sequencing the DNA in multiple (10- >100 x) small random fragments (‘**reads**’) that typically vary in size between less than 100 to several 1000 DNA basepairs (bp) (‘**massive parallel sequencing**’). The average number of times the genome is sequenced is called the **coverage**. Before the data can be analyzed, it must be cleaned and assessed for quality and often assembled into as few contiguous pieces (**contigs)** as possible. A completely assembled genome is in one contig for the chromosome and the extrachromosomal elements in each one piece but most often a genome will be assembled in 5- 200 contigs. If a genome is not fully assembled, we do not know the actual sequence of the whole genome but rather 97- 99 % of it. Assembling genomes is a computer intensive process that can be done by aligning the raw sequences against a well assembled sequence of a closely related strain (**reference based assembly**) or simply by aligning overlapping sequences from different reads without the need of a reference genome (**de novo assembly**). However, some comparisons of genomes may be performed little assembly (‘**assembly free**’) with minimal processing. For example, if you want to check if a specific gene, e.g, rpoB for species identification, or a specific set of genes, e.g., those used for multi locus sequence typing (MLST), for which the sequence(s) are known, the raw reads of the strain in question may be queried without assembly for the presence of this gene or these genes.

Center for Genomic Epidemiology

Home | Organization | Project | Services | Contact

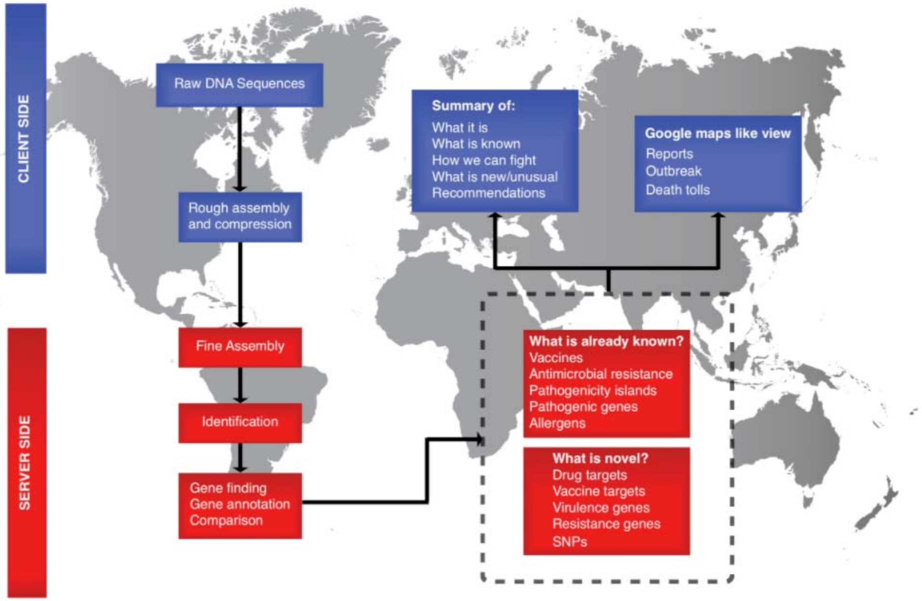
Services

Phenotyping:

- Identifcation of acquired antibiotic resistance genes. ResFinder
- Identifcation of functional metagenomic antibiotic resistance determinants. ResFinderFG
- Identifcation of acquired antibiotic resistance genes using Kmers. KmerResistance
- Prediction of a bacteria's pathogenicity towards human hosts. PathogenFinder
- Identifcation of acquired virulence genes. VirulenceFinder
- Determination of Restriction-Modification sites (based on REBASE .) Restriction-ModificationFinder
- SPIFinder identifies Salmonella Pathogenicity Islands SPIFinder

Typing:

- Multi Locus Sequence Typing (MLST) from an assembled genome or from a set of reads MLST
- PlasmidFinder identifies plasmids in total or partial sequenced isolates of bacteria. PlasmidFinder
- Multi Locus Sequence Typing (MLST) from an assembled plasmid or from a set of reads pMLST
- Prediction of bacterial species using a fast K-mer algorithm. KmerFinder
- Prediction of bacterial species using the S16 ribosomal DNA sequence. SpeciesFinder
- Fast prediction of bacterial



Welcome to the Center for Genomic Epidemiology

The cost of sequencing a bacterial genome is $50 and is expected to decrease further in the near future and the equipment needed cost less than $150 000. Thus, within a few years all clinical microbiological laboratories will have a sequencer in use on a daily basis. The price of genome sequencing is already so low that whole genome sequencing will also find worldwide application in human and veterinary practices as well as many other places where bacteria are handled. In Denmark alone this equals more than 1 million isolates annually in 15-20 laboratories and globally up to 1-2 billion isolates per year. The limiting factor will therefore in the future not be the cost of the sequencing, but how to assemble, process and handle the large amount of data in a standardized way that will make the information useful, especially for diagnostic and surveillance.

The aim of this center is to provide the scientific foundation for future internet-based solutions where a central database will enable simplification of total genome sequence information and comparison to all other sequenced including spatial-temporal analysis. We will develop algorithms for rapid analyses of whole genome DNA-sequences, tools for analyses and extraction of information from the sequence data and internet/web-interfaces for using the tools in the global scientific and medical community. The activity is being expanded to also include other microorganisms, such as vira and parasites as well as metagenomic samples.

News

What Can We Learn from a Metagenomic Analysis of a Georgian Bacteriophage Cocktail?
December 2015
Link to article....

WGS typing is a superior alternative to conventional typing strategies
August 2015
In combination with other available WGS typing tools, E. coli serotyping can be performed solely from WGS data, providing faster and cheaper typing than current routine procedures. Link to article....

Introduction to microbial whole genome sequencing and analysis for clinical microbiologist
April 2015
We offer clinical microbiologists the possibility to learn how to use the tools for e.g. typing, identifying plasmids, antibiotic resistance and virulence genes and for phylogenetic analysis. Sign up....

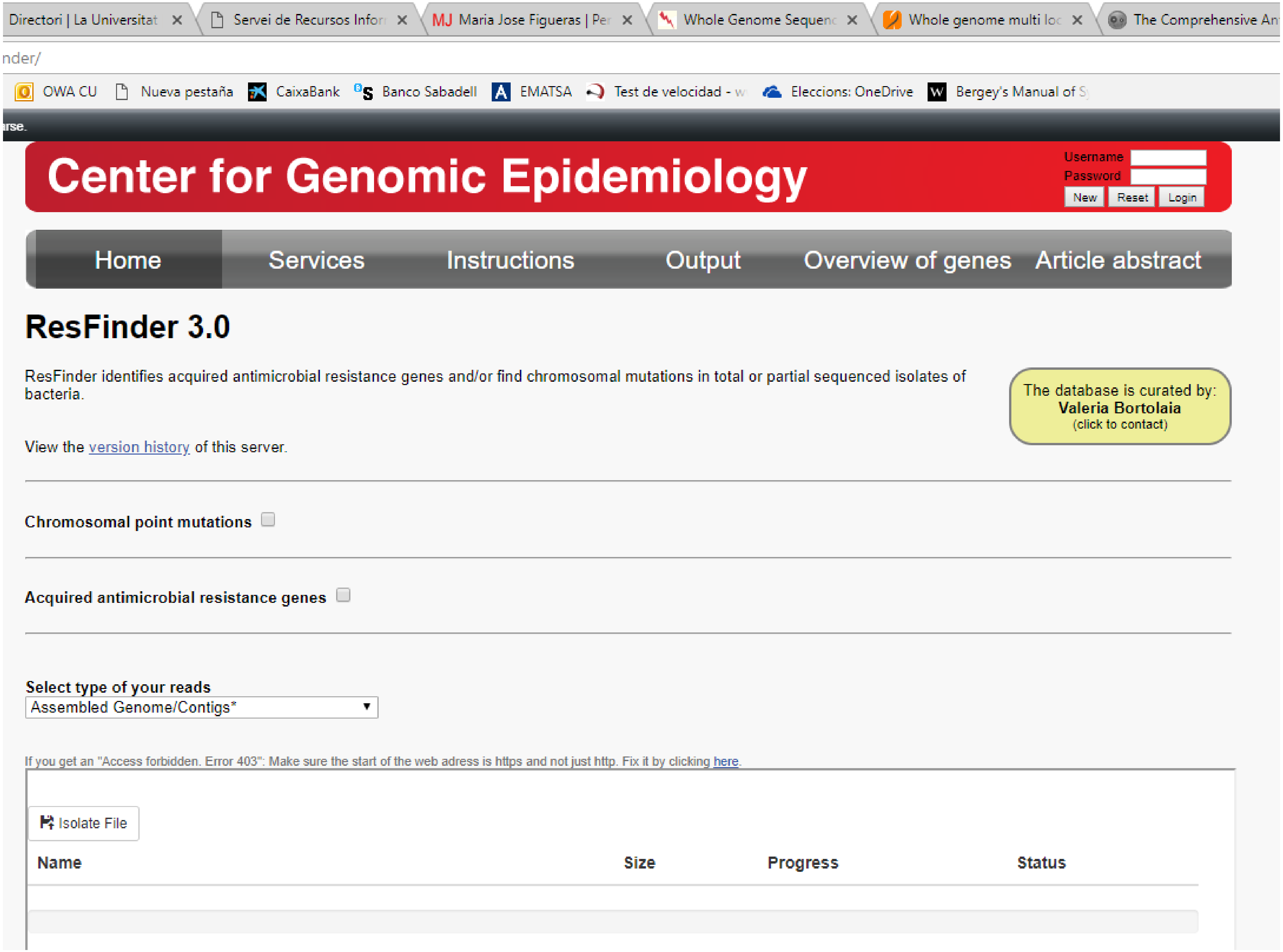
Consortium to combat infectious disease outbreaks
January 2015
The COMPARE project has been funded with 20 million Euros from the EU. The Consortium consists of 29 partners with multidisciplinary expertise in human health, animal health and food safety. Read more....

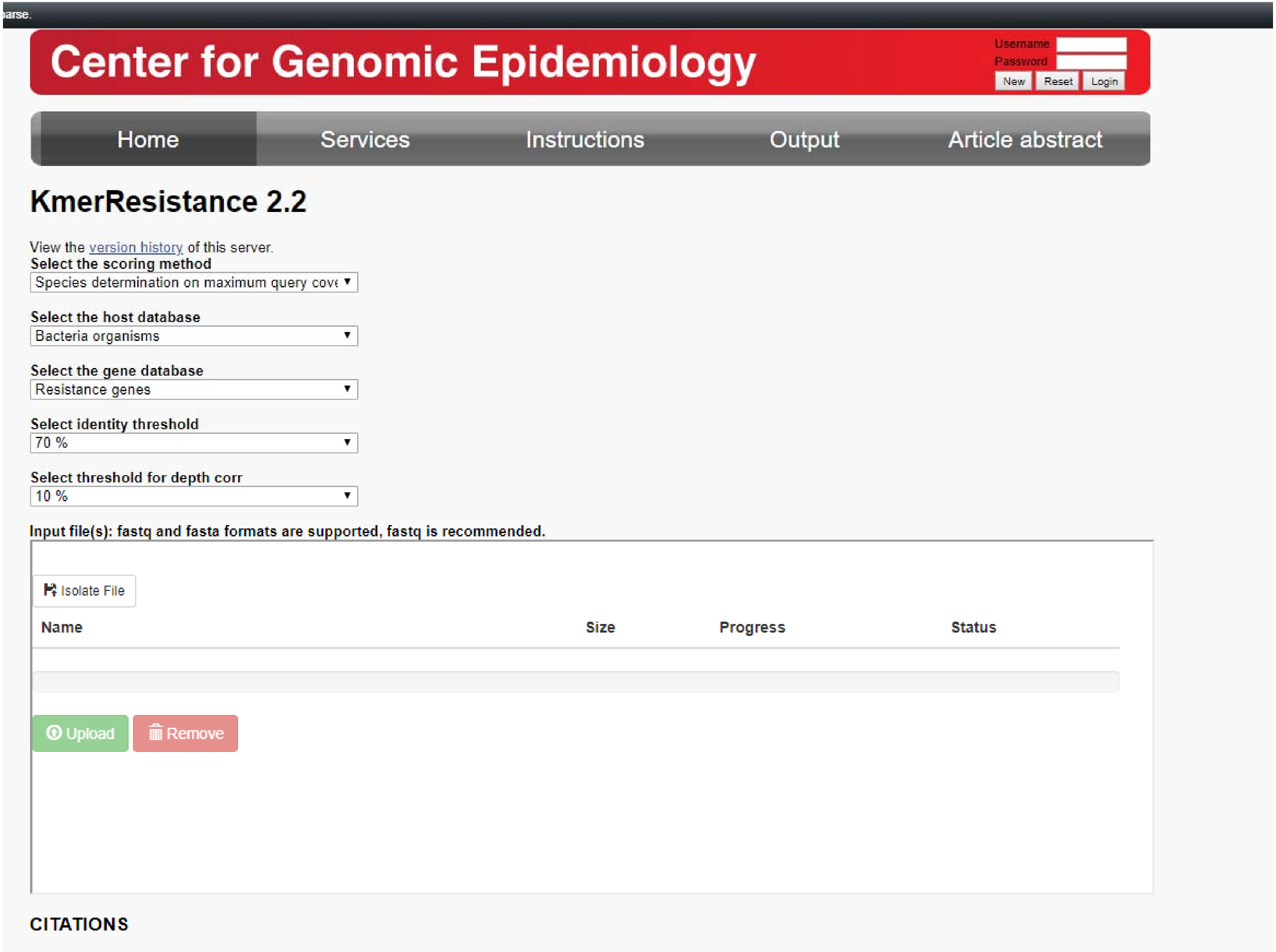
Benchmarking of Methods for Genomic Taxonomy
April 2014
How to optimally determine taxonomy from whole genome sequences. Link to article....

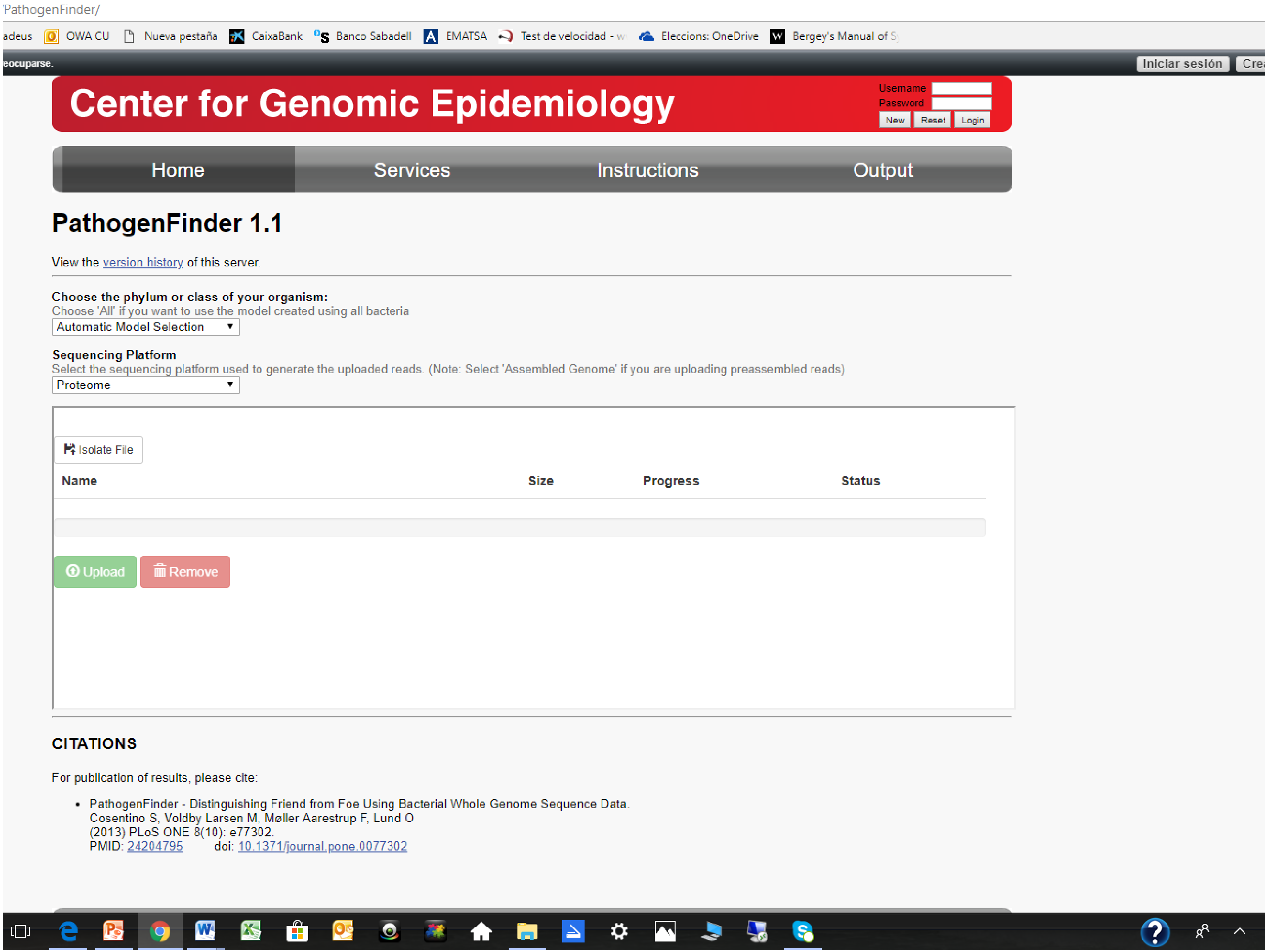
CGE tools applied for bacteriophage characterization
March 2014
Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and E. coli virulence genes in bacteriophage and prophage nucleotide sequences. Link to article....

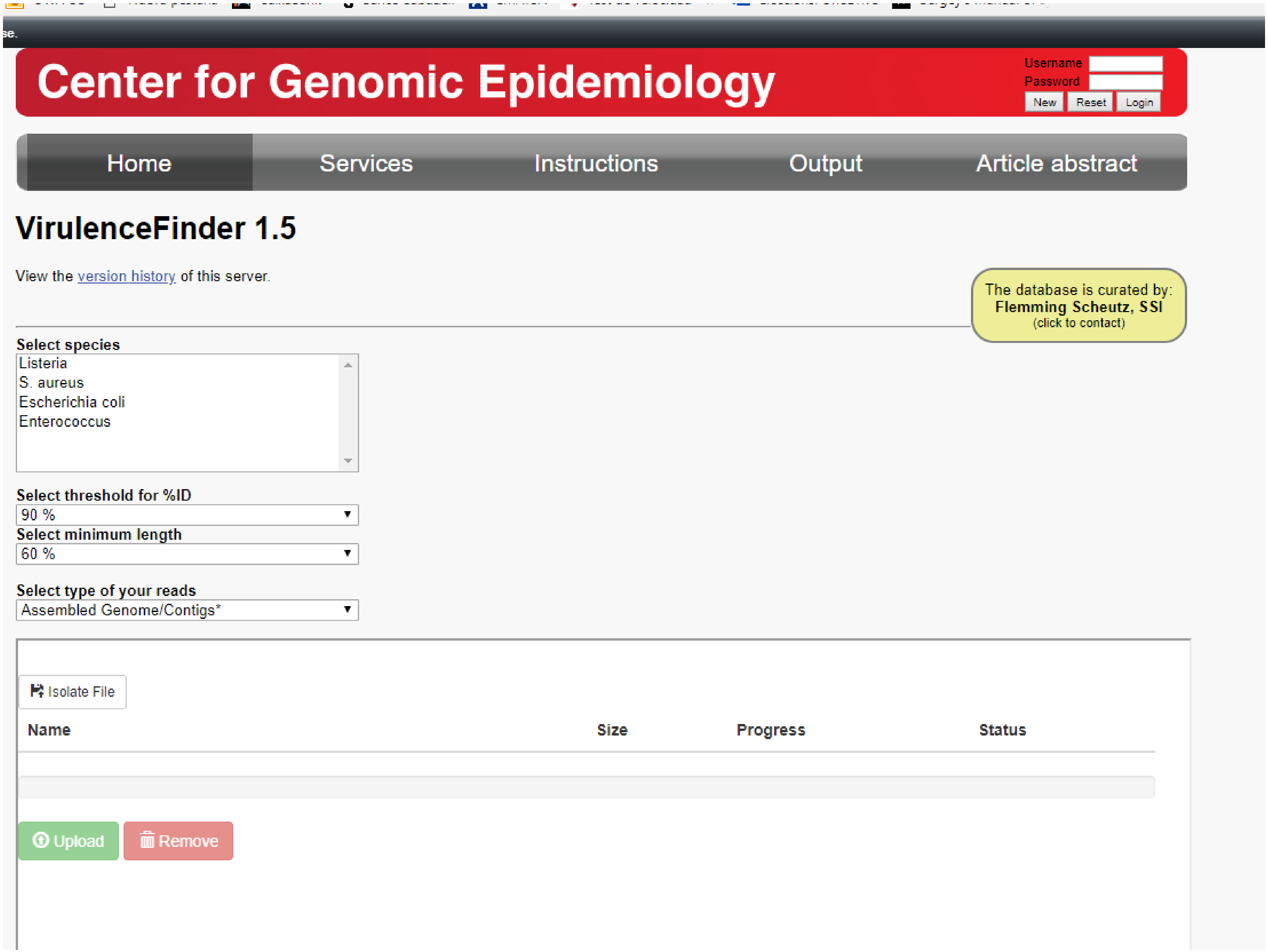
Evaluation of Whole Genome Sequencing for Outbreak Detection of Salmonella enterica
March 2014
We evaluated WGS for outbreak detection of Salmonella enterica including different approaches for analyzing and comparing with a traditional typing, PFGE. Link to article....

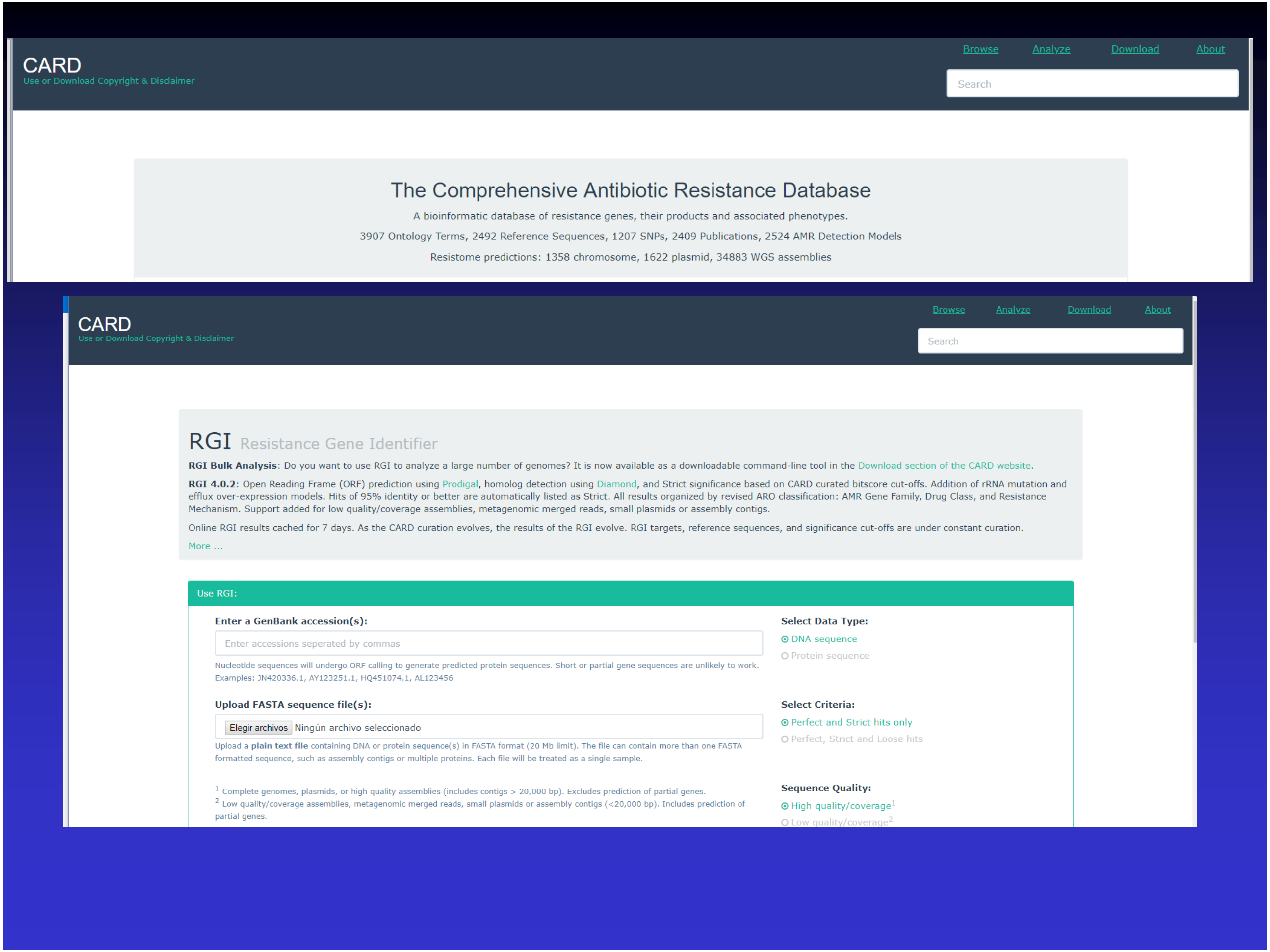
http://www.genomicepidemiology.org/


Directori | La Universitat
Servei de Recursos Infor
Maria Jose Figueras | Per
Whole Genome Sequenc
Whole genome multi loc
The Comprehensive An
nder/
OWA CU
Nueva pestaña
CaixaBank
Banco Sabadell
EMATSA
Test de velocidad - w
Eleccions: OneDrive
Bergey's Manual of S
Center for Genomic Epidemiology
Username
Password
New
Reset
Login
Home
Services
Instructions
Output
Overview of genes
Article abstract
ResFinder 3.0
ResFinder identifies acquired antimicrobial resistance genes and/or find chromosomal mutations in total or partial sequenced isolates of bacteria.
View the version history of this server.
The database is curated by:
Valeria Bortolaia
(click to contact)
Chromosomal point mutations
Acquired antimicrobial resistance genes
Select type of your reads
Assembled Genome/Contigs*
If you get an "Access forbidden. Error 403": Make sure the start of the web adress is https and not just http. Fix it by clicking here.
Isolate File
Name
Size
Progress
Status


Center for Genomic Epidemiology
Username
Password
New
Reset
Login
Home
Services
Instructions
Output
Article abstract
KmerResistance 2.2
View the version history of this server.
Select the scoring method
Species determination on maximum query cove
Select the host database
Bacteria organisms
Select the gene database
Resistance genes
Select identity threshold
70 %
Select threshold for depth corr
10 %
Input file(s): fastq and fasta formats are supported, fastq is recommended.
Isolate File
Name
Size
Progress
Status
Upload
Remove
CITATIONS


PathogenFinder/
adeus OWA CU Nueva pestaña CaixaBank Banco Sabadell EMATSA Test de velocidad - w Eleccions: OneDrive Bergey's Manual of S
eocuparse.
Iniciar sesión
Center for Genomic Epidemiology
Username
Password
New Reset Login
Home Services Instructions Output
PathogenFinder 1.1
View the version history of this server.
Choose the phylum or class of your organism:
Choose 'All' if you want to use the model created using all bacteria
Automatic Model Selection
Sequencing Platform
Select the sequencing platform used to generate the uploaded reads. (Note: Select 'Assembled Genome' if you are uploading preassembled reads)
Proteome
Isolate File
Name Size Progress Status
Upload Remove
CITATIONS
For publication of results, please cite:
PathogenFinder - Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data.
Cosentino S, Voldby Larsen M, Møller Aarestrup F, Lund O
(2013) PLoS ONE 8(10): e77302.
PMID: 24204795 doi: 10.1371/journal.pone.0077302


Center for Genomic Epidemiology
Username
Password
New
Reset
Login
Home
Services
Instructions
Output
Article abstract
VirulenceFinder 1.5
View the version history of this server.
The database is curated by:
Flemming Scheutz, SSI
(click to contact)
Select species
Listeria
S. aureus
Escherichia coli
Enterococcus
Select threshold for %ID
90 %
Select minimum length
60 %
Select type of your reads
Assembled Genome/Contigs*
Isolate File
Name
Size
Progress
Status
Upload
Remove


CARD
Use or Download Copyright & Disclaimer
Browse
Analyze
Download
About
Search
The Comprehensive Antibiotic Resistance Database
A bioinformatic database of resistance genes, their products and associated phenotypes.
3907 Ontology Terms, 2492 Reference Sequences, 1207 SNPs, 2409 Publications, 2524 AMR Detection Models
Resistome predictions: 1358 chromosome, 1622 plasmid, 34883 WGS assemblies
CARD
Use or Download Copyright & Disclaimer
Browse
Analyze
Download
About
Search
RGI Resistance Gene Identifier
RGI Bulk Analysis: Do you want to use RGI to analyze a large number of genomes? It is now available as a downloadable command-line tool in the Download section of the CARD website.
RGI 4.0.2: Open Reading Frame (ORF) prediction using Prodigal, homolog detection using Diamond, and Strict significance based on CARD curated bitscore cut-offs. Addition of rRNA mutation and efflux over-expression models. Hits of 95% identity or better are automatically listed as Strict. All results organized by revised ARO classification: AMR Gene Family, Drug Class, and Resistance Mechanism. Support added for low quality/coverage assemblies, metagenomic merged reads, small plasmids or assembly contigs.
Online RGI results cached for 7 days. As the CARD curation evolves, the results of the RGI evolve. RGI targets, reference sequences, and significance cut-offs are under constant curation.
More ...
Use RGI:
Enter a GenBank accession(s):
Enter accessions seperated by commas
Nucleotide sequences will undergo ORF calling to generate predicted protein sequences. Short or partial gene sequences are unlikely to work.
Examples: JN420336.1, AY123251.1, HQ451074.1, AL123456
Upload FASTA sequence file(s):
Elegir archivos Ningún archivo seleccionado
Upload a plain text file containing DNA or protein sequence(s) in FASTA format (20 Mb limit). The file can contain more than one FASTA formatted sequence, such as assembly contigs or multiple proteins. Each file will be treated as a single sample.
1 Complete genomes, plasmids, or high quality assemblies (includes contigs > 20,000 bp). Excludes prediction of partial genes.
2 Low quality/coverage assemblies, metagenomic merged reads, small plasmids or assembly contigs (<20,000 bp). Includes prediction of partial genes.
Select Data Type:
DNA sequence
Protein sequence
Select Criteria:
Perfect and Strict hits only
Perfect, Strict and Loose hits
Sequence Quality:
High quality/coverage1
Low quality/coverage2

Main limitations of WGS

1. **Absence of genomes of many type strains**
   Missing in ca. 50% of the bacteria species with validly published names (Chun et al., 2018)

2. **Presence of mislabeled genomes**
   27% (16701/62362) of the Whole Genome Assemblies (WGAs) studied by Yoon et al. (2017)

3. **Absence of 16S rRNA genes in 6% (4285/69745) of the WGAs (Lee et al., 2017)**

4. **Contaminated genomes or chimeras**
   0.9% (597/69745) detected using the ContEst16 (Lee et al., 2017)

Conclusions

- The isDDH, ANI and the MLPA are excellent tools to verify the identity of existing genomes and are extremely useful for defining new species and for recognizing strains.
- The Orto ANI seems to be the best platform.
- The 16S rRNA gene should be sequence ca. 1500 bp using conventional sequencing approaches.
- The use of several methodologies in parallel show their individual limitations and help to determine with more precision the similarity of the genomes.
- Quality control measure for genomes are essential.
- Update databases of virulence genes and ARG are needed.

GOBIERNO DE ESPAÑA | MINISTERIO DE CIENCIA E INNOVACIÓN

NEWMICRORISK

2012-16

Universitat de Barcelona

Laboratory of Virus Contaminants of Water and Food (VIRCONT)
Rosina Gironés Llop

Subproject 1:

VIRRECRISK

2012-16

UNIVERSITAT ROVIRA I VIRGILI

MicroAqua - FMCS

Subproject 1:

BACTRECRISK

AQUAVALENS

SEVENTH FRAMEWORK PROGRAMME

2013-17

Water JPI

2015-2018

METAWATER

GOBIERNO DE ESPAÑA | MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

About ASM

- **Oldest and largest** life science organization in the world
- **>47,000 members worldwide**
  - Membership starting at $22
  - Student discounts available
- **27 disciplines** of microbiological specialization including a division for microbiology educators
- **Meetings**: ASM Microbe, BioDefense, and specialize conferences
- **Professional Development:** Workshops, Student Travel Awards, Fellowships, and more
- **Publications**: *13 journals* – 42% of all citations in microbiology. 100s of book titles in multiple languages

María José Figueras

Universidad Rovira i Virgili

Spanish ASM ambassador

AMERICAN SOCIETY FOR MICROBIOLOGY

The mission of the ASM is to promote & advance the microbial sciences

"ASM is a member driven society"

Victor DiRita, Ph.D.

2016 JOIN or RENEW

AMERICAN SOCIETY FOR MICROBIOLOGY