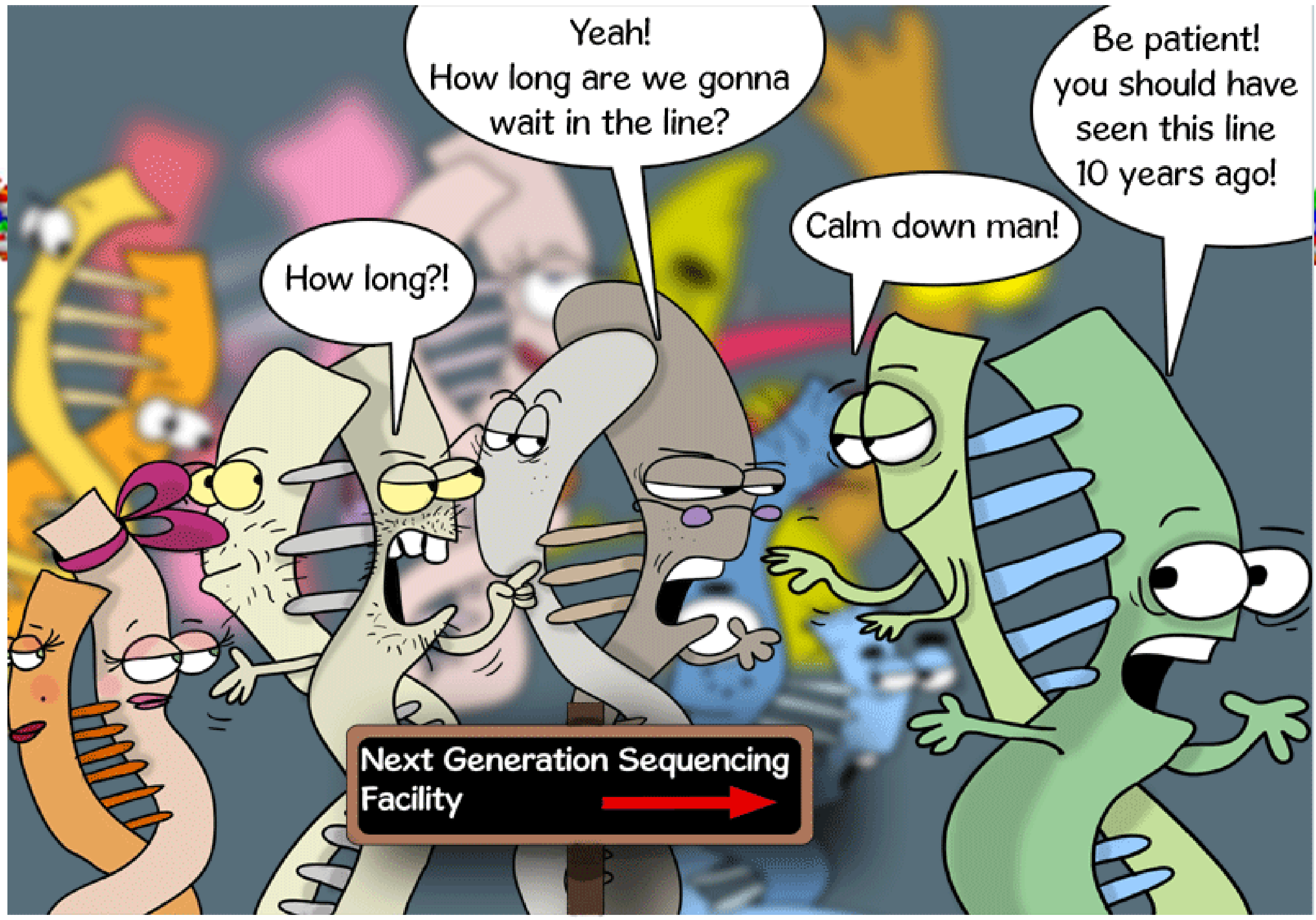
Phylogenomic Analysis of the *Flavobacteriaceae*

Jeff Newman,
Lycoming College, Williamsport PA

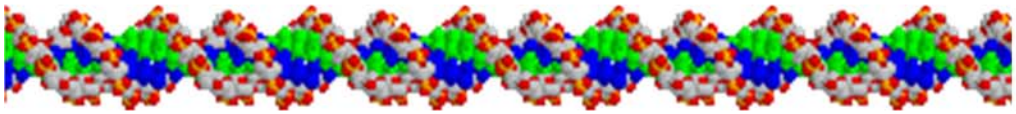
October 27, 2015



Human Genome 10th Anniversary

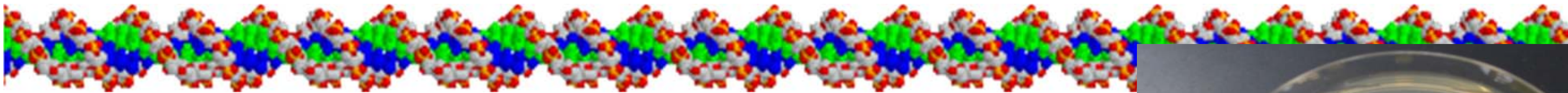
http://biocomicals.blogspot.com

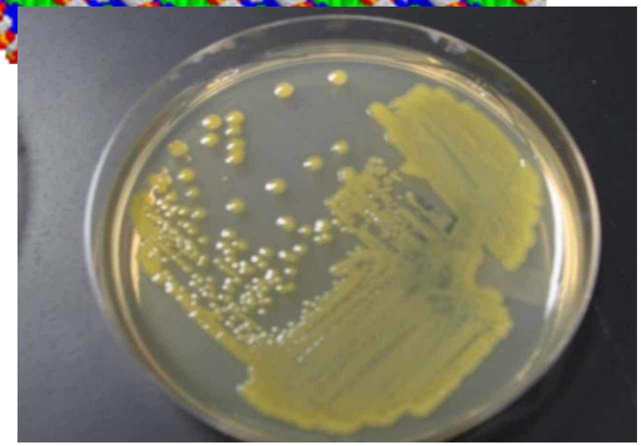
The Journey



- Starting point: Undergrad Micro UK's from freshwater creeks, 16S → novel species
- To publish w/o DDH→MLST→genomes
- Existing Phylogenomic Metrics (GGDC,ANI)
- RAST for genome analysis→ROSA
  - Create family *Chryseobacteriaceae*
  - Reinstate genus *Kaistella*
  - What is a *Flavobacterium*? (similar to problem with *Bacillus, Pseudomonas, Lactobacillus, Clostridium, Rhizobium*)
    - comparisons to *F. aquatile*?
    - Solutions to problem?

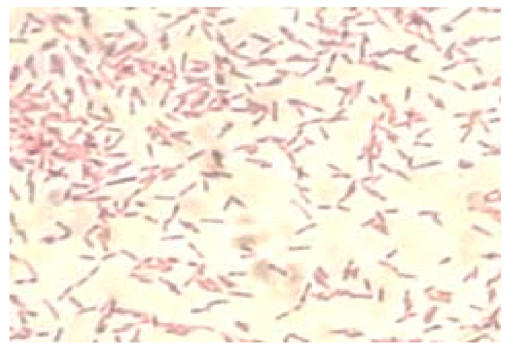
The LycoMicro Unknown Microbe Lab

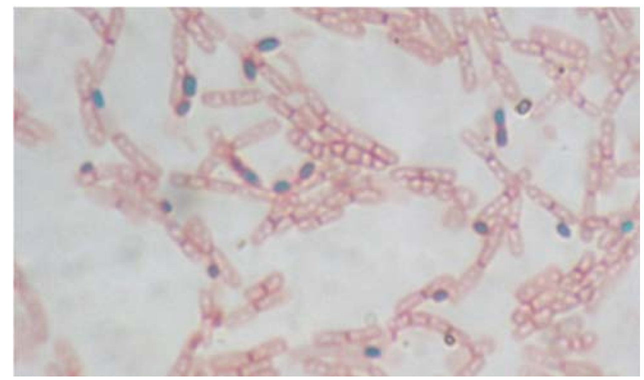




Week 1 - Aseptic technique/Inoculation

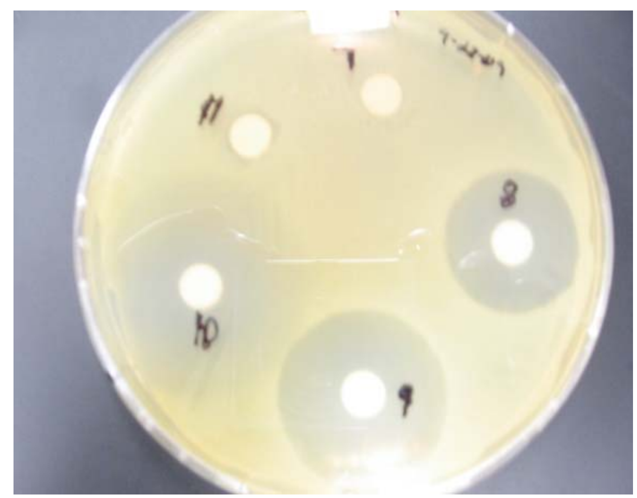
- pipetting sterile media
- selection of knowns and unknowns
- streak plates, inoculation of liquid
- preparation of frozen permanents





Week 2 – Staining & Microscopy

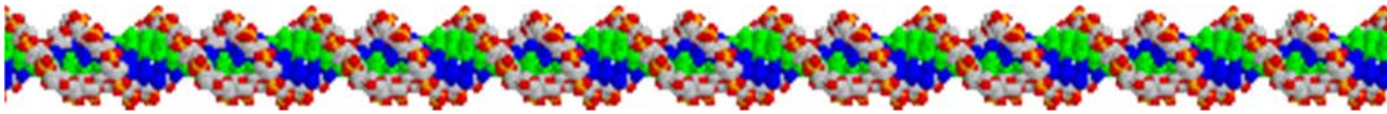
- Gram stain
- Endospore stain
- Wet mount



Week 3 – Antibiotic sensitivity, pH, [NaCl], Temperature, Oxygen requirements

- Kirby-Bauer Disk Diffusion assay (10)
- oxidase, catalase, Gas-Pak Jar

The LycoMicro Unknown Microbe Lab

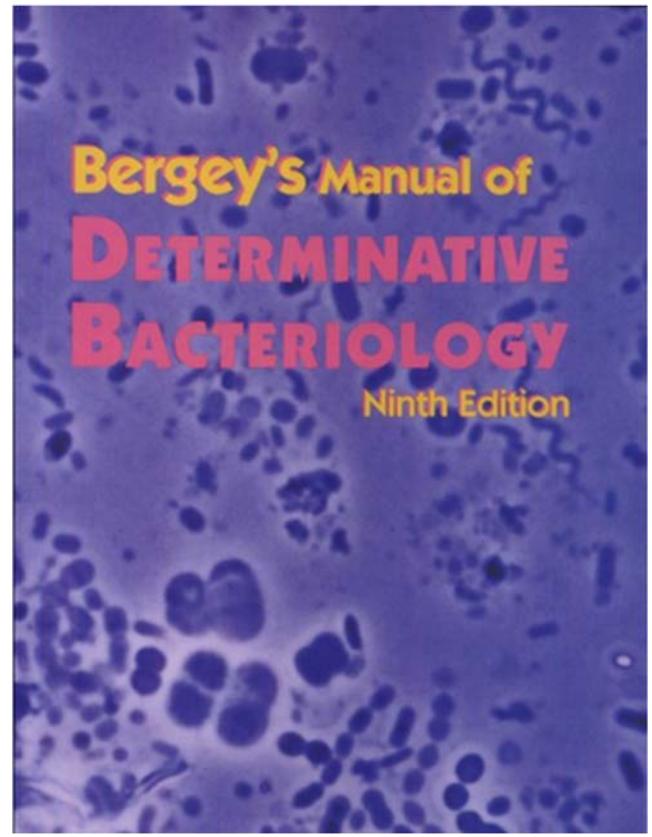


Week 4 – Carbohydrate & Nitrogen Metabolism

- MRVP, citrate, phenol red, TSI
- urease, nitrate reduction, SIM

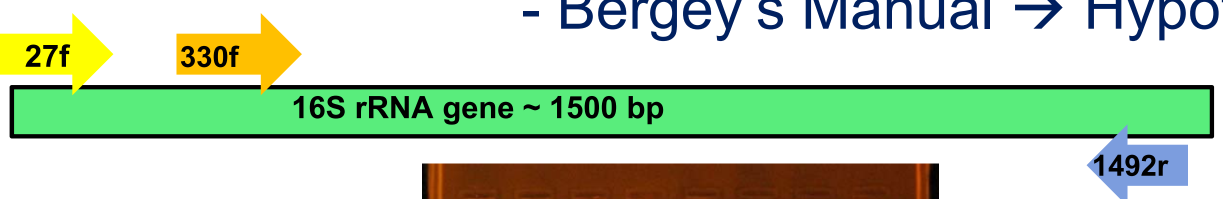
Week 5 – Exoenzymes, Differential/Selective media

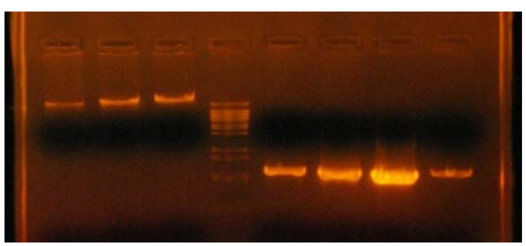
- caseinase, lipase, amylase, DNase,
- EMB, HEA, MSA, PhenylEthanol,
- Bile Esculin, Brilliant Green, EG Min.



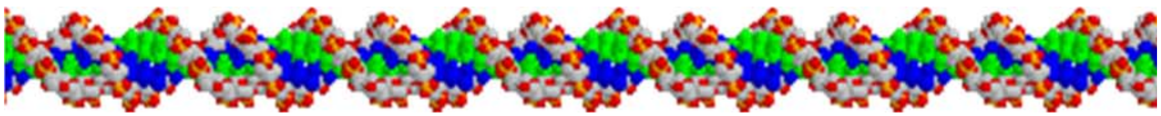
Week 6 - PCR of 16S rDNA, gel, send for sequencing.

- Bergey's Manual → Hypothesis





The LycoMicro Unknown Microbe Lab



Week 7 – MIDI – FAME Analysis

- BUG+Blood (Hemolysis) → **Biolog GenIII Plates**

Week 8 – Analyze DNA sequence @ http://www.ezbiocloud.net/eztaxon ,

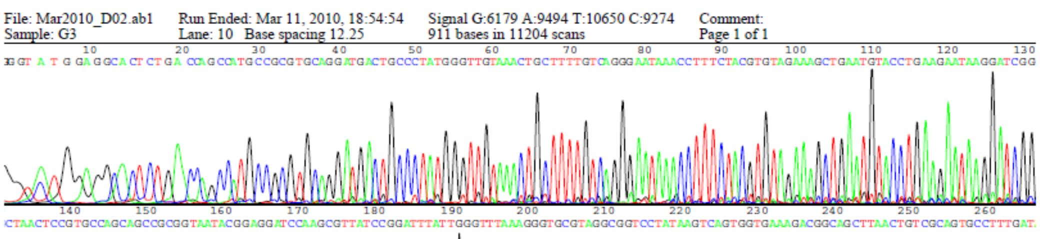
- Construct Phylogenetic Tree w/MEGA,
- Literature Research (IJSEM)

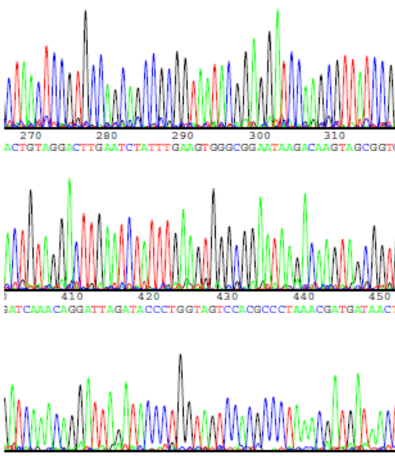
RT	Response	Ar/Ht	RFact	ECL	Peak Name	Percent	Comment1	Comment2
0.7087	997637	0.005	----	6.5830		----	< min rt	
0.7150	2.392E+9	0.025	----	6.6245	SOLVENT PEAK	----	< min rt	
0.9403	3636	0.019	----	8.1196		----	< min rt	
1.4058	344	0.009	----	10.9549	unknown 10.9525	----	ECL deviates 0.002	
1.6463	1040	0.011	1.060	12.0005	12:0	0.30	ECL deviates 0.001	Reference -0.009
2.1563	542	0.010	----	13.8325		----		
2.2050	5594	0.009	0.989	13.9985	14:0	1.52	ECL deviates -0.001	Reference -0.009
2.3660	482	0.009	----	14.5205	unknown 14.502	----	ECL deviates 0.004	
2.5146	1057	0.009	----	15.0020	15:0	----	ECL deviates 0.002	
2.6788	5932	0.009	0.954	15.5175	Sum In Feature 2	1.56	ECL deviates 0.002	14:0 3OH/16:1 iso I
2.7816	138178	0.009	0.948	15.8402	Sum In Feature 3	36.00	ECL deviates 0.000	16:1 w7c/16:1 w6c
2.8086	600	0.009	0.946	15.9253	16:1 w5c	0.16	ECL deviates -0.003	
2.8326	141842	0.009	0.945	16.0006	16:0	36.85	ECL deviates 0.001	Reference -0.007
3.1245	34913	0.009	0.932	16.9158	Sum In Feature 3	8.94	ECL deviates 0.001	17:0 cyclo
3.1518	1097	0.009	0.931	17.0014	17:0	0.28	ECL deviates 0.001	Reference -0.006
3.4195	52052	0.009	0.921	17.8442	Sum In Feature 8	13.19	ECL deviates -0.003	18:1 w7c
3.4686	3044	0.009	0.920	17.9989	18:0	0.77	ECL deviates -0.001	Reference -0.009
3.4964	918	0.009	0.919	18.0885	18:1 w7c 11-methyl	0.23	ECL deviates 0.002	
3.7579	773	0.010	0.912	18.9333	Sum In Feature 8	0.19	ECL deviates 0.001	19:0 cyclo w8c
----	5932	---	----	----	Summed Feature 2	1.56	16:1 iso I/14:0 3OH	14:0 3OH/16:1 iso I
----	173091	---	----	----	Summed Feature 3	44.95	16:1 w7c/16:1 w6c	16:1 w6c/16:1 w7c
----	-----	---	----	----		----	17:0 cyclo	
----	52826	---	----	----	Summed Feature 8	13.38	18:1 w7c	18:1 w6c
----	-----	---	----	----		----	19:0 cyclo w8c	

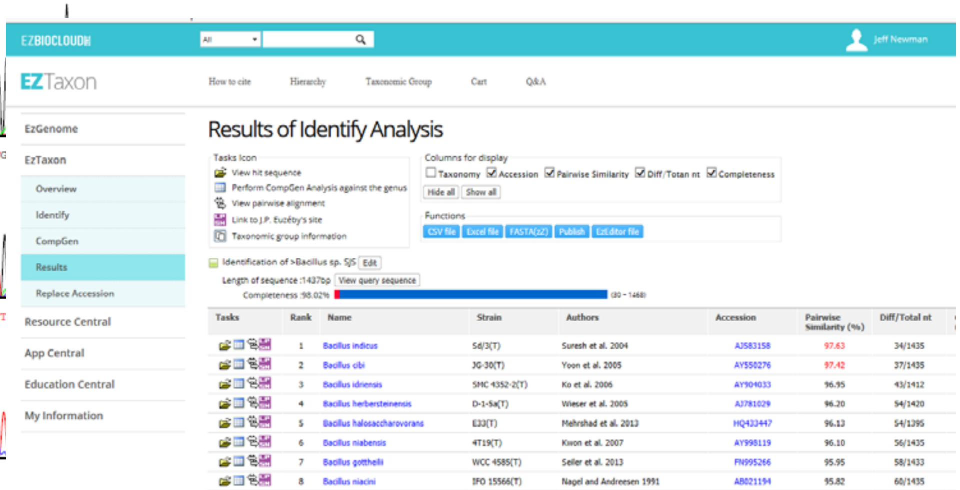
ECL Deviation: 0.002 Reference ECL Shift: 0.008 Number Reference Peaks: 5
Total Response: 386527 Total Named: 385984
Percent Named: 99.86% Total Amount: 363709

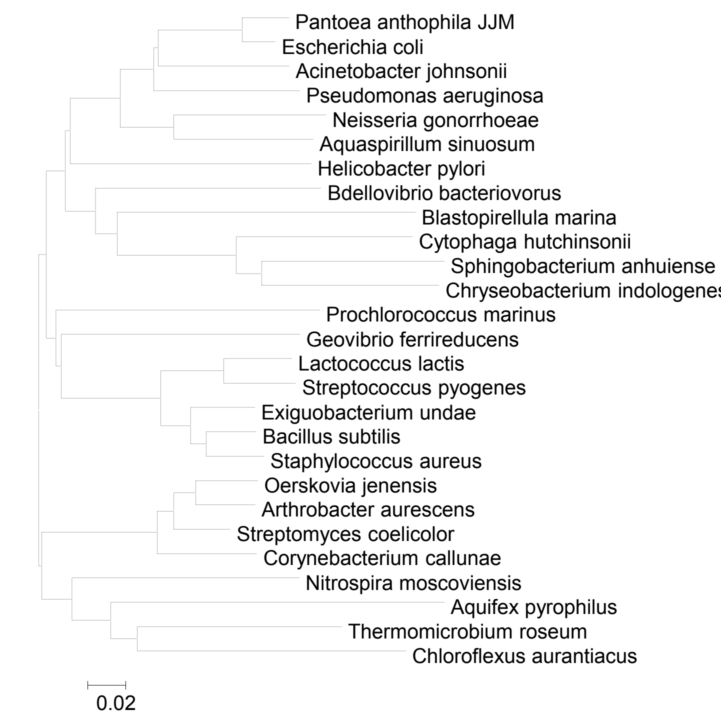
Matches:

Library	Sim Index	Entry Name
ITSA1 1.10	0.890	Serratia-liquefaciens (confirm with other tests)
	0.797	Pantoea-agglomerans-biogroup 1B (confirm with other tests)
	0.648	Ewingella-americana
	0.579	Edwardsiella-tarda
	0.537	Pseudomonas-putida-GC subgroup B
	0.537	Yersinia-frederiksenii/entercolitica/intermedia

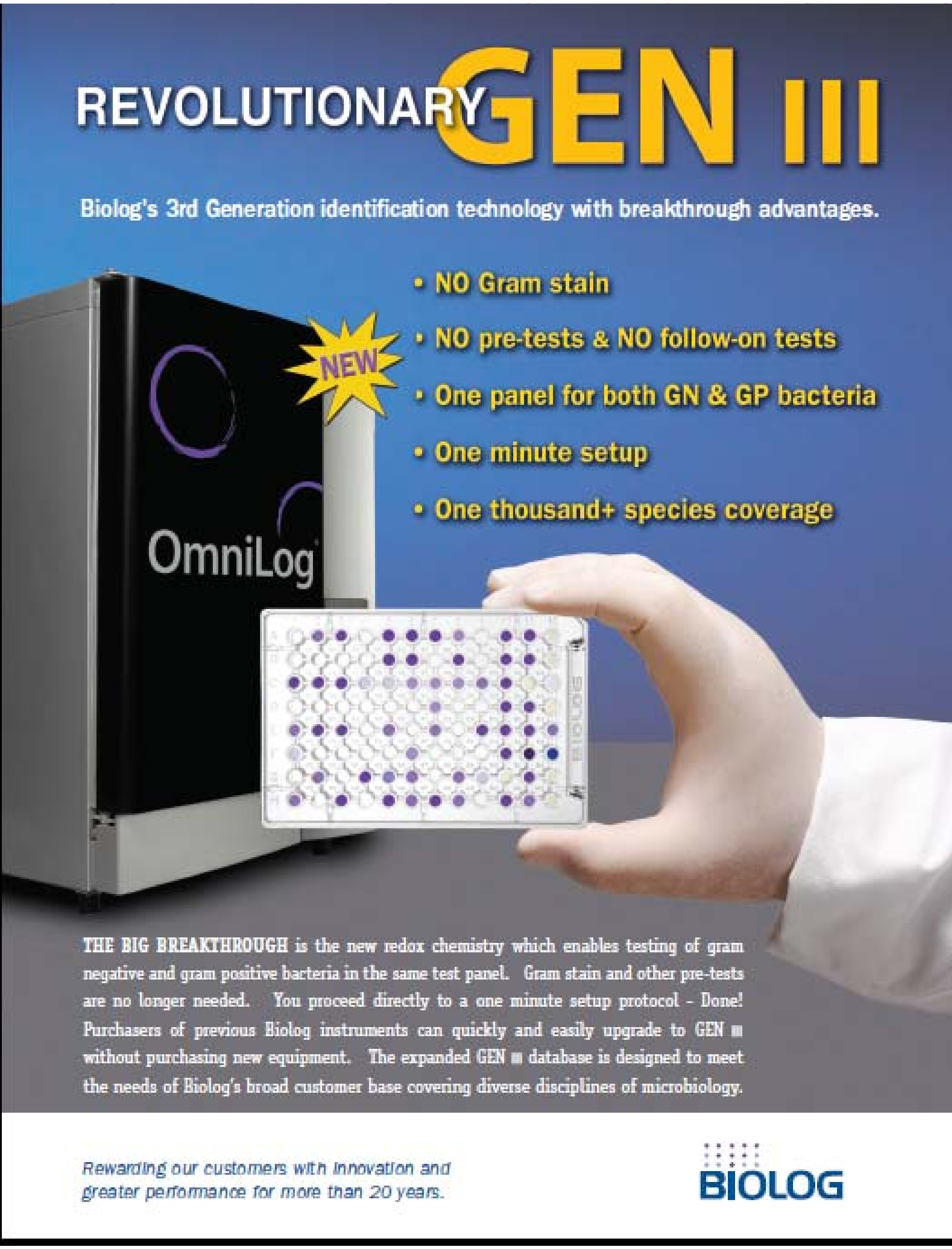


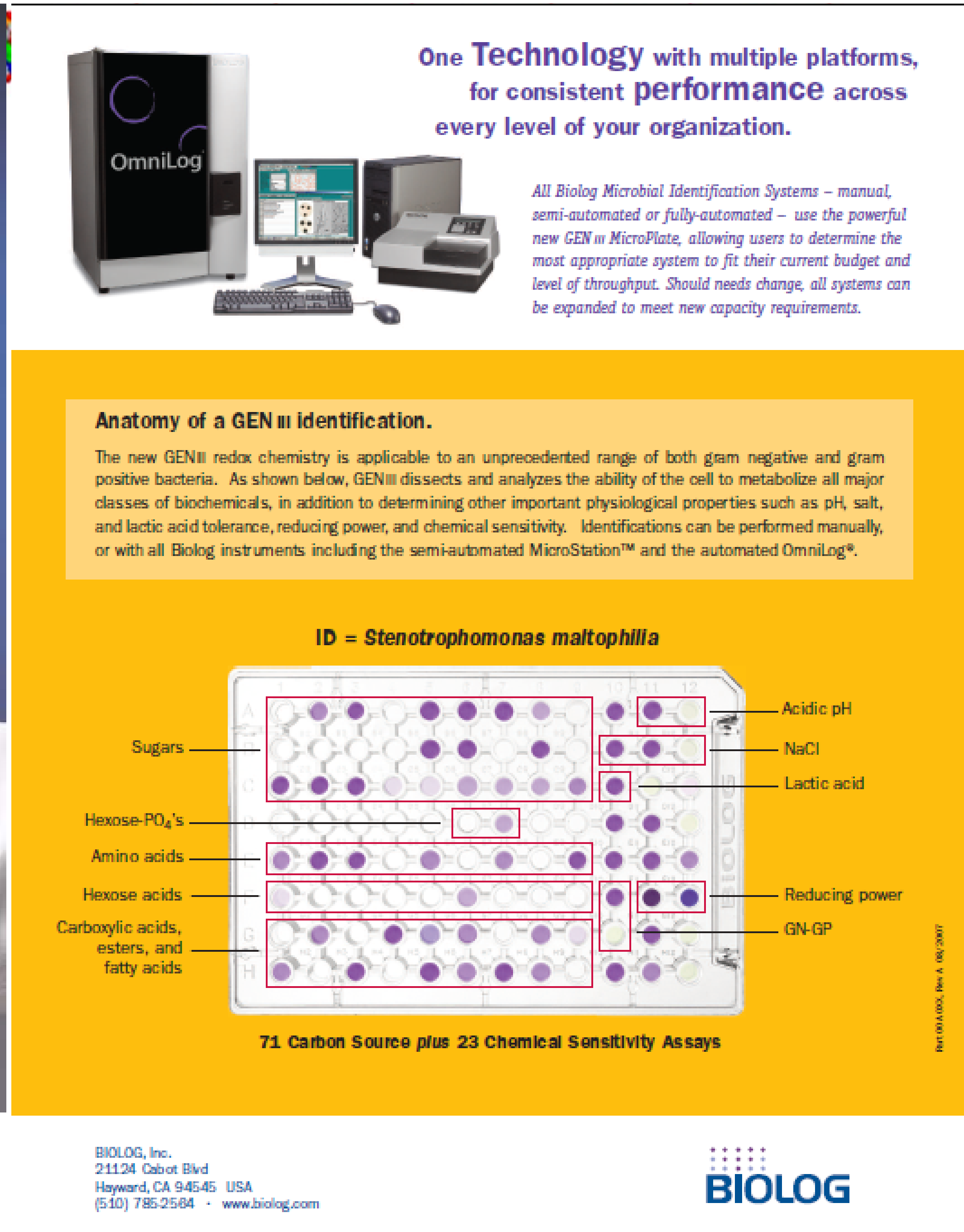




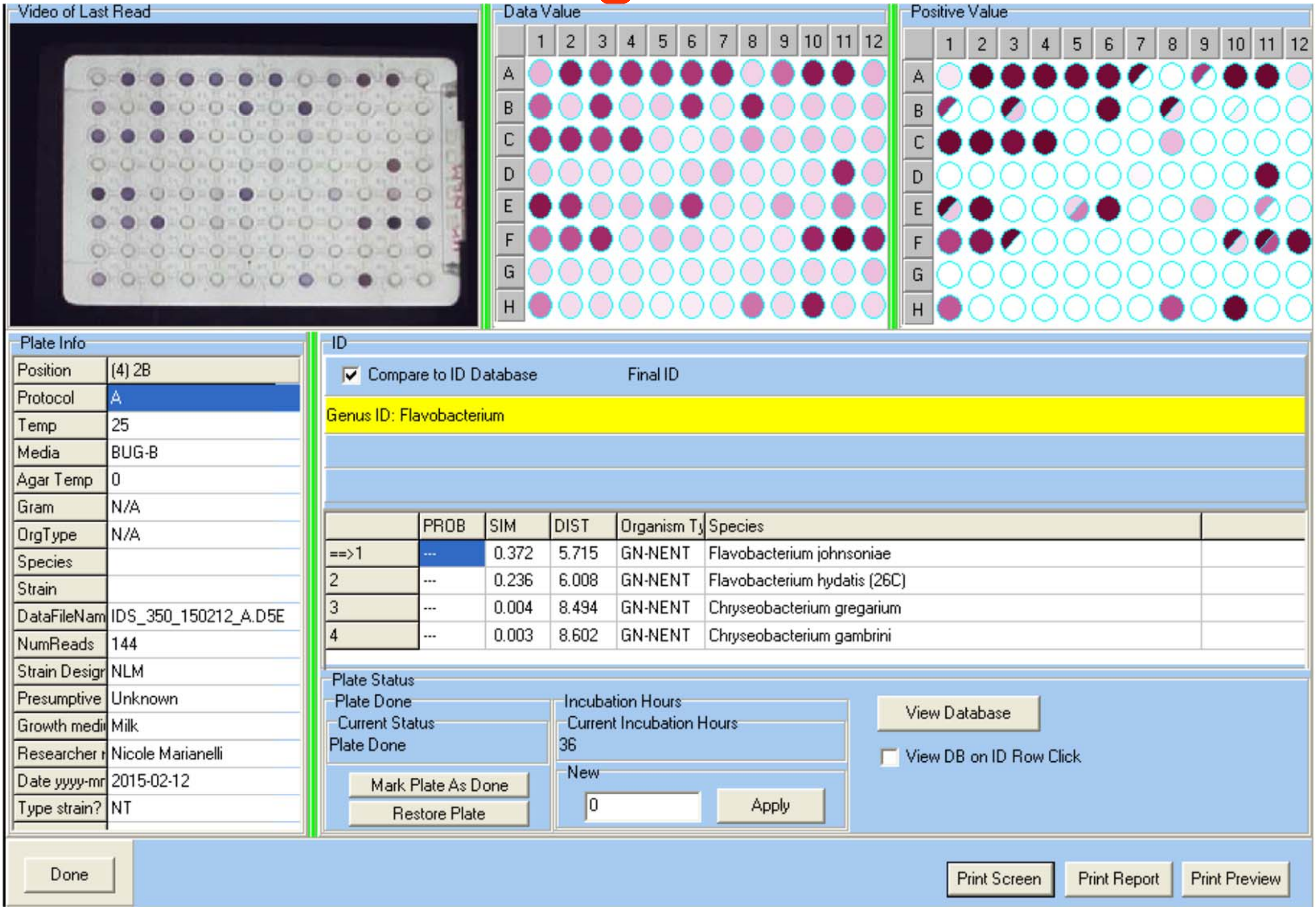


Biolog GenIII for ID (2009 NSF MRI grant)

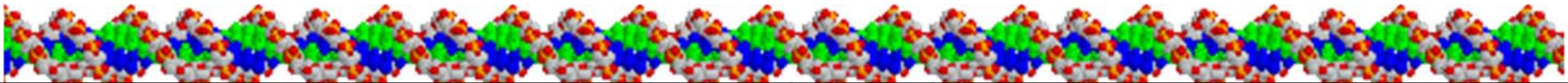




Biolog results

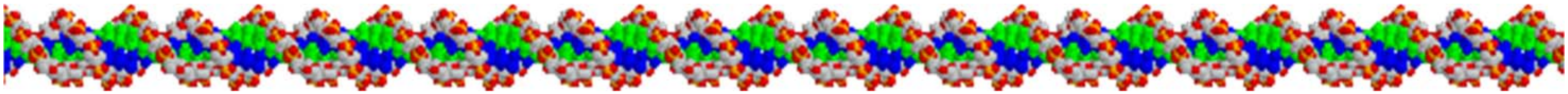


Biolog GenIII Plate

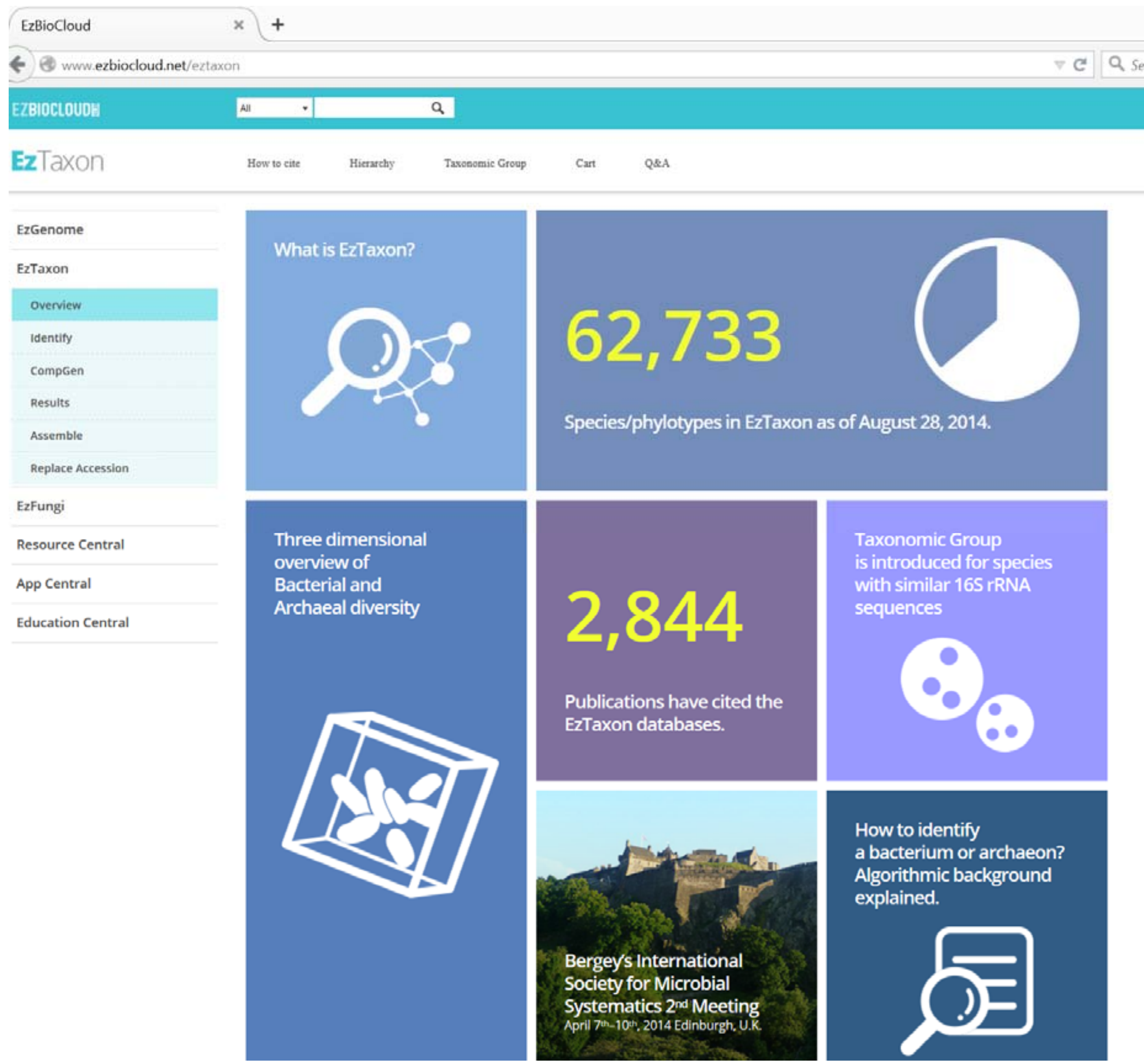


	1	2	3	4	5	6	7	8	9	10	11	12
A	negative control	**Dextrin** amylose glu-α1-4, α1-6 poly	**D-maltose** glu-α1-4-glu	**D-trehalose** glu-α1-α1-glu	**D-cellobiose** glu-β1-4-glu	**gentiobiose** glu-β1-6-glu	**Sucrose** glu-α1-β2-fru	**D-turanose** glu-α1-α3-fru	**Stachyose** gal(α1→6) gal(α1→6) glc(α1↔2β) fru	positive control	pH 6	pH 5
B	**D-raffinose** gal-α1-6-glu-α1-β2-fru	**α-D-lactose** gal-β1-4-glu (α)	**D-melibiose** gal-α1-6-glu	β-methyl-D-glucoside	**D-salicin** aspirin-β1-glu	N-acetyl-D-glucos-amine	N-acetyl-β-D-mannos-amine	N-acetyl-D-galactos-amine	N-acetyl neuraminic acid	1% NaCl	4% NaCl	8% NaCl
C	α-D-glucose	D-mannose	D-fructose	D-galactose	3-methyl glucose	D-fucose	L-fucose	L-rhamnose	inosine	1% Na-lactate	fusidic acid	D-serine
D	D-sorbitol	D-mannitol	D-arabitol	myo-inositol	glycerol	D-glucose-6-PO4	D-fructose-6-PO4	D-aspartic acid	D-serine	Troleando-mycin	rifamycin SV	minocycline
E	gelatin	glycyl-L-proline	L-alanine	L-arginine	L-aspartic acid	L-glutamic acid	L-histidine	L-pyro-glutamic acid	L-serine	lincomycin	guanidine HCl	niaproof 4
F	pectin	D-galact-uronic acid	L-galact-uronic acid lactone	D-gluconic acid	D-glucuronic acid	Glucuron-amide	mucic acid	quinic acid	D-saccharic acid	vancomycin	tetrazolium violet	tetrazolium blue
G	p-hydroxy-phenyl-acetic acid	methyl pyruvate	D-lactic acid methyl ester	L-lactic acid	citric acid	α-keto-glutaric acid	D-malic acid	L-malic acid	bromo-succinic acid	nalidixic acid	LiCl	K-tellurite
H	tween-40	γ-amino-butyric acid	α-hydroxy-butyric acid	β-hydroxy-D,L-butyric acid	α-keto-butyric acid	acetoacetic acid	propionic acid	acetic acid	formic acid	aztreonam	Na-butyrate	Na bromate

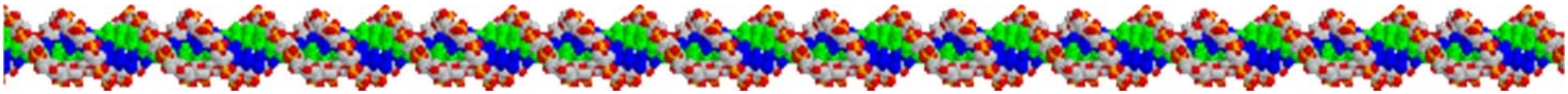
EZ-Taxon



- http://www.ezbiocloud.net/eztaxon
- Database contains type strains = official representatives of a species
  - Good pairwise scores
  - “CompGen” compares to all members of Genus, can easily download and format sequences for trees



EZ-Taxon



What is threshold for same vs different species?

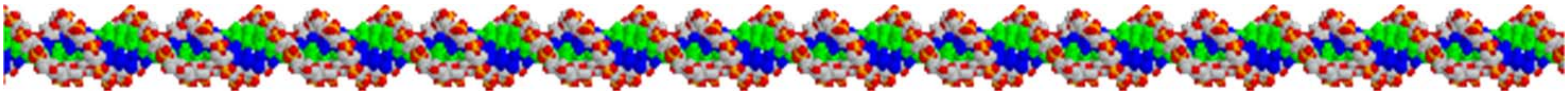
RESULT OF IDENTIFY ANALYSIS

Query: UK-AED

View all seq | View(acc) | View(zZ) | NJ Tree | Add query to cart | Add to cart

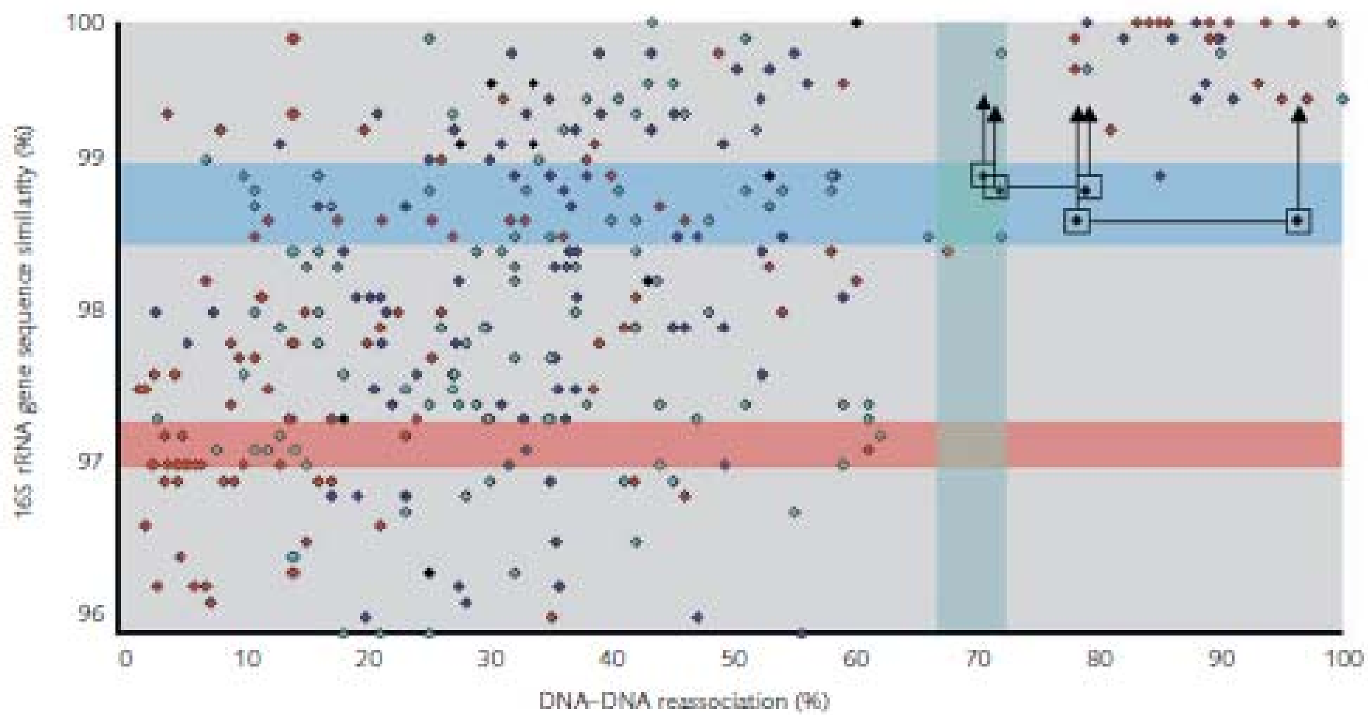
Rank	Name/Title	Authors	Strain	Accession	Pairwise Similarity	Diff/Total nt
1	Flavobacterium reichenbachii	Ali et al. 2009	WB 3.2-61(T)	AM177616	97.059	22/748
2	Flavobacterium sinopsychrotolerans	Xu et al. 2011	0533(T)	FJ654474	96.524	26/748
3	Flavobacterium limicola	Tamaki et al. 2003	ST-82(T)	AB075230	96.524	26/748
4	Flavobacterium glaciei	Zhang et al. 2006	0499(T)	DQ515962	95.722	32/748
5	Flavobacterium hibernum	McCammon et al. 1998	ATCC 51468(T)	L39067	95.576	33/746
6	Flavobacterium frigoris	Van Trappen et al. 2004	LMG 21922(T)	AJ557887	95.455	34/748
7	Flavobacterium succinicans	(Reichenbach 1989) Bernardet et al. 1996	DSM 4002(T)	AM230492	95.321	35/748
8	Flavobacterium resistens	Ryu et al. 2008	BD-b365(T)	EF575563	95.321	35/748
9	Flavobacterium xinjiangense	Zhu et al. 2003	JCM 11314(T)	AF433173	95.321	35/748
10	Flavobacterium xanthum	(ex Inoue and Komagata 1976) McCammon and Bowman 2000	ACAM 81(T)	AF030380	95.321	35/748
11	Flavobacterium chilense	Kämpfer et al. (in press)	LM-09-Fp(T)	FR774915	95.187	36/748
12	Flavobacterium urumqiense	Dong et al. (in press)	Sr25(T)	HQ436467	95.053	37/748
13	Flavobacterium tiangeerense	Xin et al. 2009	563(T)	EU036219	95.053	37/748
14	Flavobacterium banpakuense	Kim et al. 2011	15F3(T)	GQ281770	94.920	38/748
15	Flavobacterium omnivorum	Zhu et al. 2003	JCM 11313(T)	AF433174	94.786	39/748
16	Flavobacterium chungbukense	Lim et al. (in press)	CS100(T)	HM627539	94.786	39/748
17	Flavobacterium chungangense	Kim et al. 2009	CJ7(T)	EU924275	94.758	39/744
18	Flavobacterium araucananum	Kämpfer et al. (in press)	LM-19-Fp(T)	FR774916	94.652	40/748
19	Flavobacterium aquidurense	Cousin et al. 2007	WB-1.1.56(T)	AM177392	94.652	40/748
20	Flavobacterium hydatis	Bernardet et al. 1996	DSM 2063(T)	AM230487	94.652	40/748
21	Flavobacterium micromati	Van Trappen et al. 2004	LMG 21919(T)	AJ557888	94.652	40/748

How are bacterial species defined?



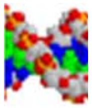
- The Gold Standard
  - Less than 70% DNA-DNA hybridization (green line)
- % Identity with 16S rDNA sequence
  - 97% (red line) vs. 98.5% (blue line)… but only 1 gene

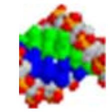
- Fig from Stackebrandt & Ebers, 2006, Microbiol.Today 33:152



95% ANI ~ 70% DDH ~ 69% Conserved (>90%)

International Journal of Systematic and Evolutionary Microbiology (2007), **57**, 81–91 **DOI** 10.1099/ijs.0.64483-0





DNA–DNA hybridization values and their relationship to whole-genome sequence similarities

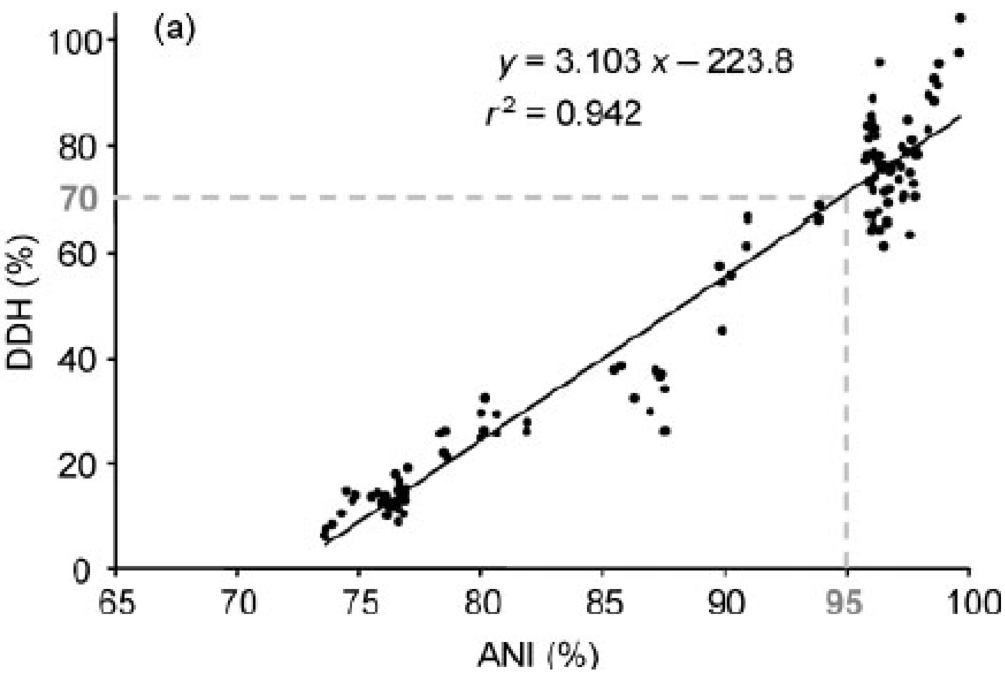
Johan Goris,[1]† Konstantinos T. Konstantinidis,[1]‡ Joel A. Klappenbach,[1] Tom Coenye,[2] Peter Vandamme[2] and James M. Tiedje[1]

[1]Center for Microbial Ecology, Michigan State University, East Lansing, MI 48824, USA

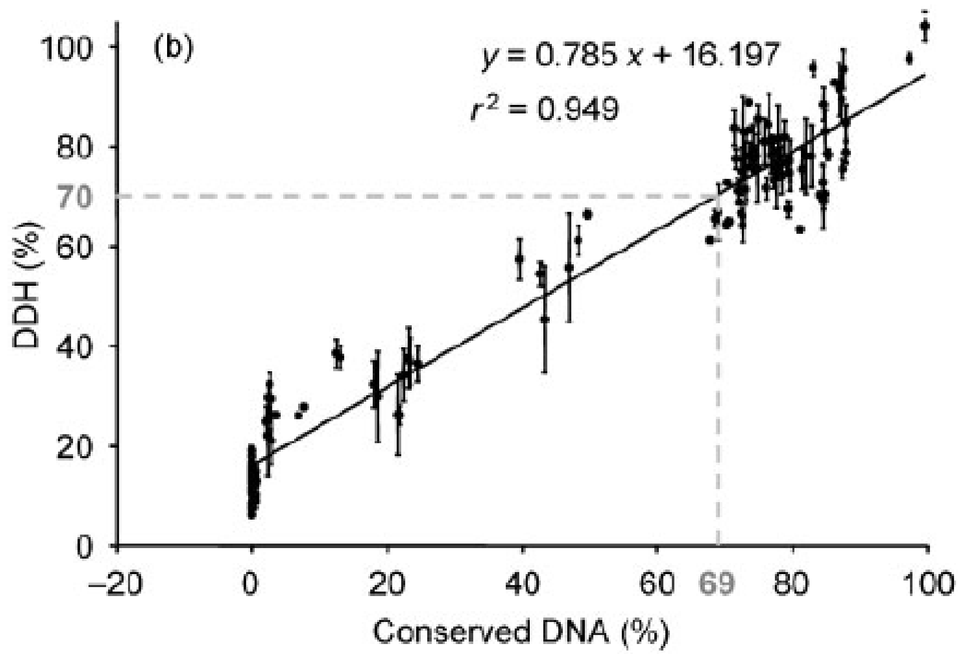
[2]Laboratory for Microbiology, Gent University, K. L. Ledeganckstraat 35, B-9000 Gent, Belgium

Correspondence
Johan Goris
johan_goris@applied-maths.com

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. Since the extent of hybridization between a pair of strains is ultimately governed



que
enc
e p
hicl
f ba
The
/bri
off
DN
DH
ene
the
fere
ely



Shifting the genomic gold standard for the prokaryotic species definition

Michael Richter and Ramon Rosselló-Móra[1]

Marine Microbiology Group, Institut Mediterrani d'Estudis Avançats (CSIC-UIB), E-07190 Esporles, Spain

Edited by James M. Tiedje, Center for Microbial Ecology, East Lansing, MI, and approved September 16, 2009 (received for review June 11, 2009)

DNA-DNA hybridization (DDH) has been used for nearly 50 years as the gold standard for prokaryotic species circumscriptions at the genomic level. It has been the only taxonomic method that offered a numerical and relatively stable species boundary, and its use has had a paramount influence on how the current classification has been constructed. However, now, in the era of genomics, DDH appears to be an outdated method for classification that needs to be substituted. The average nucleotide identity (ANI) between two genomes seems the most promising method since it mirrors DDH closely. Here we examine the work package JSpecies as a user-friendly, biologist-oriented interface to calculate ANI and the correlation of the tetranucleotide signatures between pairwise genomic comparisons. The results agreed with the use of ANI to substitute DDH, with a narrowed boundary that could be set at ≈95–96%. In addition, the JSpecies package implemented the tetranucleotide signature correlation index, an alignment-free parameter that generally correlates with ANI and that can be of help in deciding when a given pair of organisms should be classified in the same species. Moreover, for taxonomic purposes, the analyses can be produced by simply randomly sequencing at least 20% of the genome of the query strains rather than obtaining their full sequence.

average nucleotide identity | DNA-DNA hybridization | genome-based taxonomy | tetranucleotide regression

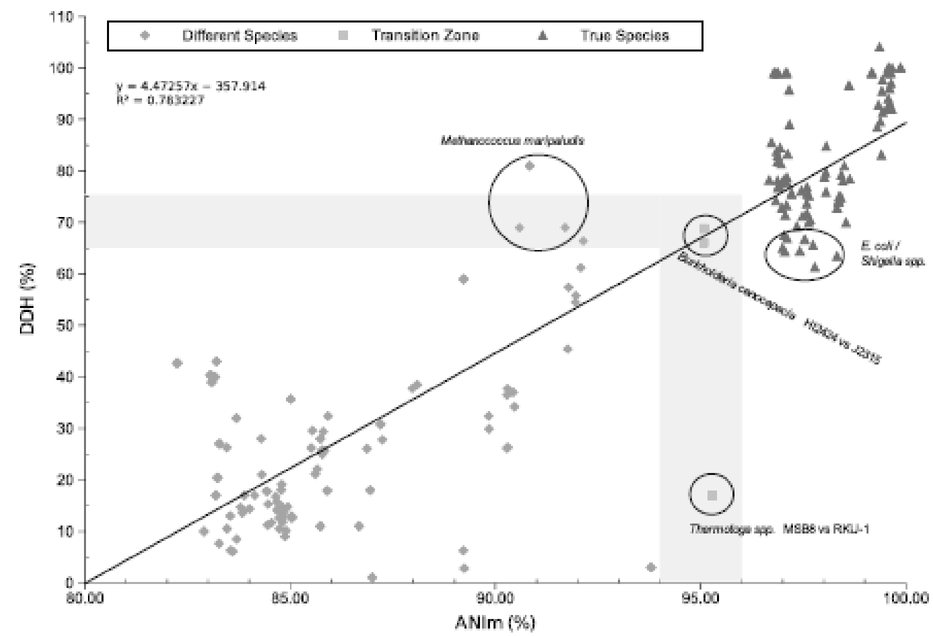


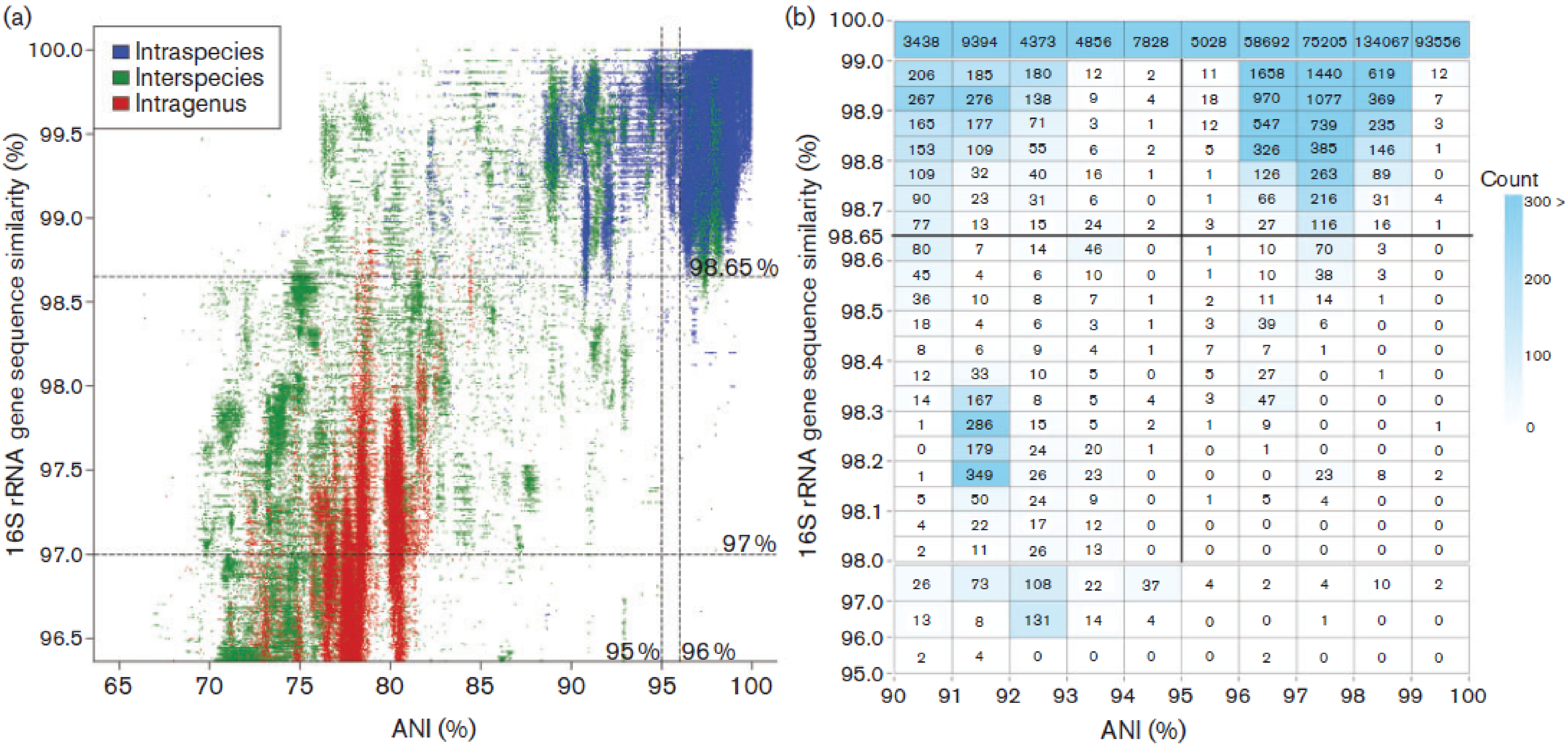
Fig. 2. Plotted values of DDH versus ANIm. Triangles show values that correspond to what taxonomists consider as "true" species according to the DDH values traditionally applied and that have previously been classified. Squares indicate values that appear to be in the transition zone.

www.pnas.org/cgi/doi/10.1073/pnas.0906412106

International Journal of Systematic and Evolutionary Microbiology (2014), 64, 346–351 DOI 10.1099/ijs.0.059774-0

Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes

Mincheol Kim,[1] Hyun-Seok Oh,[2] Sang-Cheol Park[2] and Jongsik Chun[1,2]



3438	9394	4373	4856	7828	5028	58692	75205	134067	93556
206	185	180	12	2	11	1658	1440	619	12
267	276	138	9	4	18	970	1077	369	7
165	177	71	3	1	12	547	739	235	3
153	109	55	6	2	5	326	385	146	1
109	32	40	16	1	1	126	263	89	0
90	23	31	6	0	1	66	216	31	4
77	13	15	24	2	3	27	116	16	1
80	7	14	46	0	1	10	70	3	0
45	4	6	10	0	1	10	38	3	0
36	10	8	7	1	2	11	14	1	0
18	4	6	3	1	3	39	6	0	0
8	6	9	4	1	7	7	1	0	0
12	33	10	5	0	5	27	0	1	0
14	167	8	5	4	3	47	0	0	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

Fig. 3. 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 and 1.0 in ANI.

Determine Phylogenomic Metrics – Kostas Lab ANI Calculator

ANI Average Nucleotide Identity

Kostas lab » Tools » ANI calculator

§ ANI calculator

The ANI calculator estimates the average nucleotide identity using both best hits (one-way ANI) and reciprocal best hits (two-way ANI) between two genomic datasets, as calculated by Goris *et al.*, 2007. Typically, the ANI values between genomes of the same species are above 95% (*e.g.*, *Escherichia coli*). Values below 75% are not to be trusted, and AAI should be used instead. This tool supports both complete and draft genomes (multi-fasta).
Examples: *Escherichia coli*, *Escherichia*, *Escherichia* vs *Yersinia*, *Escherichia* vs *Xanthomonas*.

§ Input data

User data

Name
E-mail
Job name

Genome 1

Browse...
or GI number:

Genome 2

Browse...
or GI number:

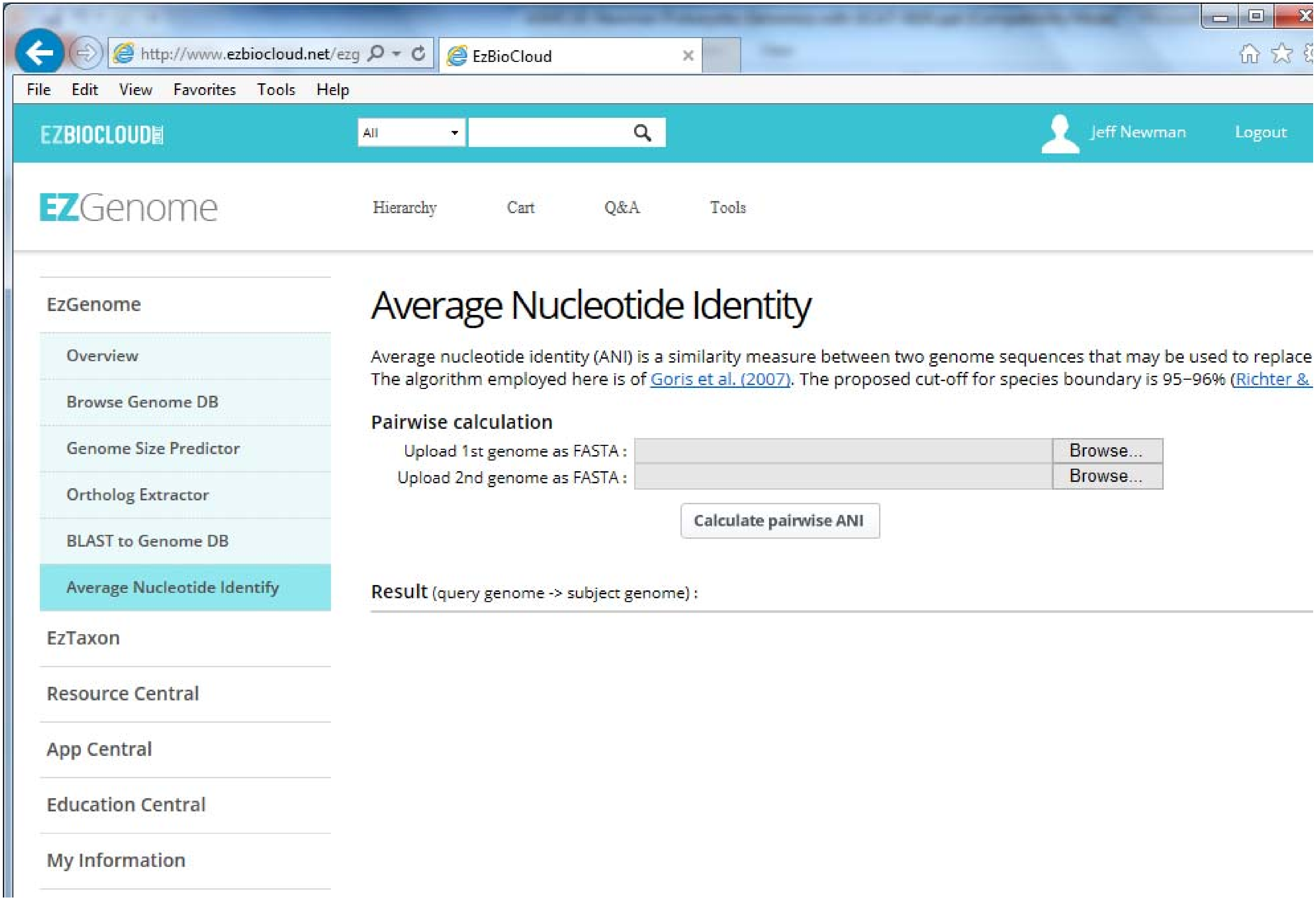
§ ANI options

Alignment options

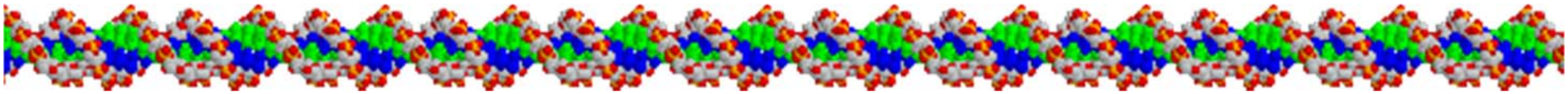
Fragment options

Site developed by Luis M. Rodriguez-R at Kostas lab.

Determine Phylogenomic Metrics – Chun Lab EzGenome ANI Calculator



Genome-Genome Distance Calculator (GGDC) at the DSMZ calculates digital DDH



GGDC

Genome-to-Genome Distance Calculator

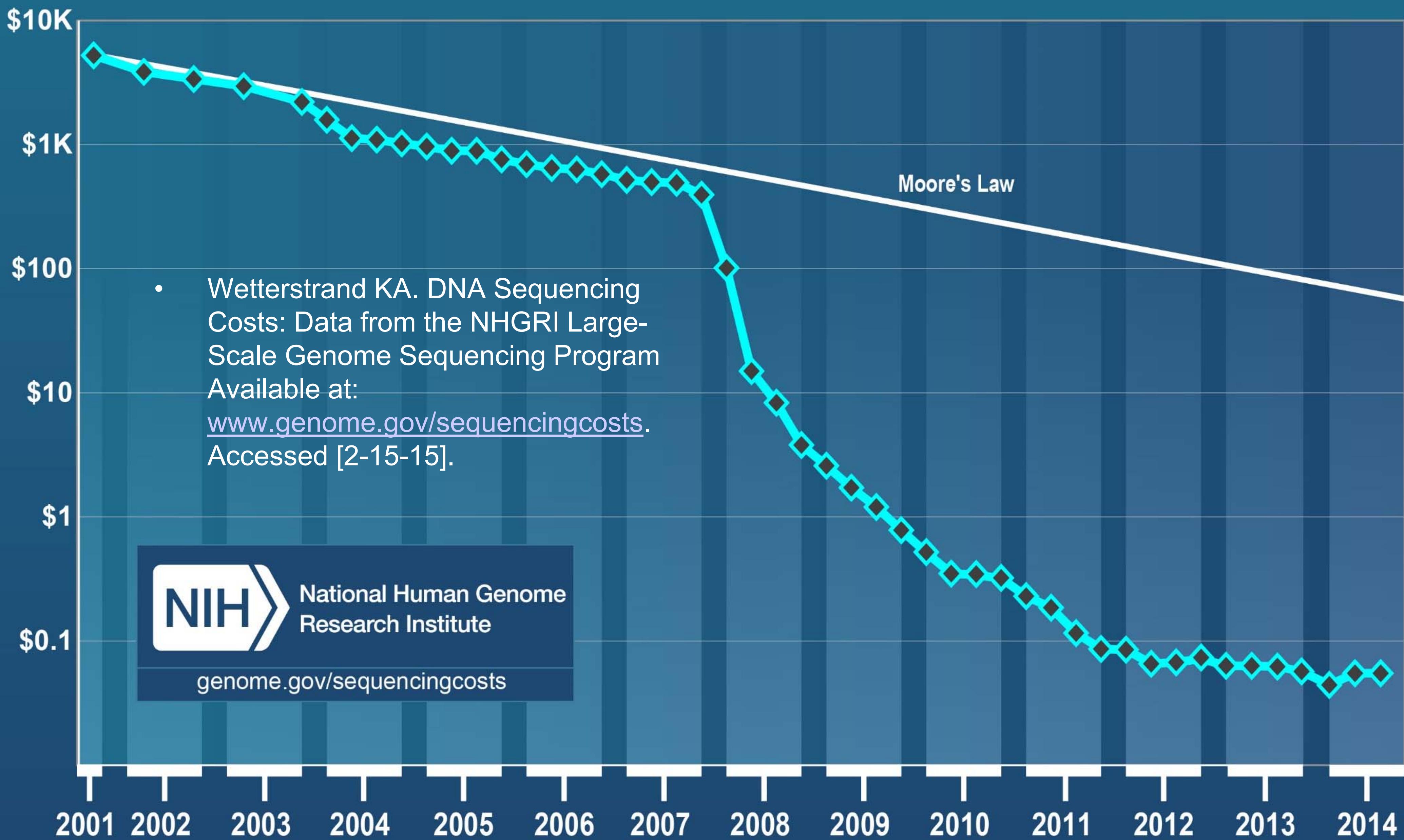
DSMZ

About this service

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. Furthermore, it 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. This web service offers state-of-the-art methods for inferring whole-genome distances which are well able to mimic DDH. 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). Our digitally derived genome-to-genome distances show a better correlation with 16S rRNA gene sequence distances than DDH values. Thus, this web service can be used for **genome-based species delineation**. Once you have obtained complete or incomplete, assembled genomes sequences, the use is easy: upload your sequence files in our distance calculation form and let our server calculate intergenomic distances for you. These are converted into similarity values analogous to DDH and sent to you via e-mail to support your **decision about the relatedness of your novel strain to known type strains**.

The GGDC has been developed entirely independently of the ANI ("average nucleotide identity") concept and is in no way based on it. Indeed, the core of GGDC, the GBDP program for calculating intergenomic distances, has been published *before* the first paper on ANI. GBDP conducts a couple of corrections that are not found in ANI, and in contrast to ANI GBDP does not split the sequences into sections of an arbitrary length of 1000 bp. In the studies listed below, GGDC yielded higher correlations with wet-lab DDH than ANI, and as of version 2.0 GGDC uses statistical models that considerably improve on the linear models used by ANI and earlier versions of GGDC. 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. See the FAQ for details.

Cost per Raw Megabase of DNA Sequence



- Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program Available at: www.genome.gov/sequencingcosts. Accessed [2-15-15].

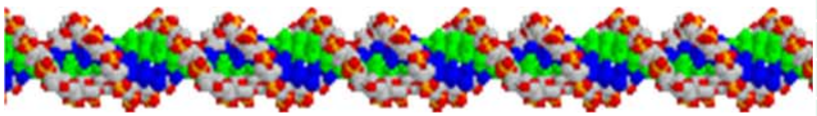
GCAT→ GCAT-SEEK

- Genome Consortium for Active Teaching (GCAT) founded in 2000 to bring Genomics (Microarrays) to the undergraduate curriculum.
- Multiple HHMI & NSF funded workshops
- GCAT-SEEKquence “spin-off” to bring NextGen sequencing to the undergraduate curriculum.
- 3 genomes (Ion Torrent & 454 as part of pilot)
- NSF Research Collaboration Network, Juniata’s HHMI Genomics Leadership Initiative

Genome Consortium GCAT for Active Teaching

SEEKquence

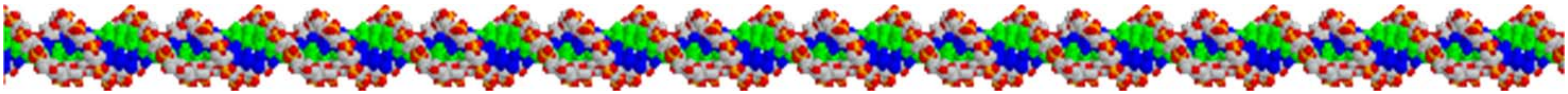
Shared MiSeq Runs



- NextGen Instruments generate more data than most UG faculty can use or afford.
- November 2013 – 27 bacteria @$200 each (including *Flavobacterium aquatile)*
- April, 2014 – Opened to Microedu Listserv → 35 Bacteria and Phage from 16 institutions @$190/sample
- October 2014 – 30 phage, viruses and bacteria @$175/sample.

Sample	Reads est.	Bases est.
GSF665-1-E_coli-C06b	217,320	130,391,966
GSF665-2-Chryseobacterium-LO	1,317,872	790,723,170
GSF665-3-Linfield-KH	809,893	485,935,870
GSF665-4-Linfield-NH	301,171	180,702,758
GSF665-5-Exiguobacterium	794,482	476,689,384
GSF665-6-Plesiomonas_shigelloides	656,143	393,685,659
GSF665-7-Halosimplex_carlsbadense	595,655	357,393,201
GSF665-8-Phage_Eapen	573,447	344,068,354
GSF665-9-Phage_Aspire	170,895	102,536,927
GSF665-10-strain_3572	593,179	355,907,159
GSF665-11-Gracilibacillus_dipsosauri	986,925	592,154,880
GSF665-12-Serratia_S12	827,533	496,519,794
GSF665-13-Rhodococcus_T1Sofl-14	297,153	178,292,067
GSF665-14-Janthinobacterium-BJB1	823,488	494,092,592
GSF665-15-Janthinobacterium-BJB349	883,287	529,972,260
GSF665-16-Janthinobacterium-BJB304	1,098,516	659,109,346
GSF665-17-Janthinobacterium-BJB317	549,616	329,769,324
GSF665-18-Iodobacter-BJB302	206,973	124,183,611
GSF665-19-Asaia_bogorensis	1,096,204	657,722,373
GSF665-20-Asaia_siamensis	820,818	492,490,968
GSF665-21-Asaia_astilbes	783,447	470,068,239
GSF665-22-Asaia_platycodi	808,325	484,994,710
GSF665-23-Asaia_krungthepensis	1,152,811	691,686,698
GSF665-24-Asaia_prunellae	1,035,414	621,248,288
GSF665-27-Serratia -DL	129,258	77,554,903
GSF665-28-Phage-KitKat	53,773	32,263,632
GSF665-29-Cyanobacterium-RC610	909,265	545,559,194
GSF665-30-Serratia_marcescens-RH	307,886	184,731,584
GSF665-31-Bacillus_cibi	693,101	415,860,714
GSF665-32-Pedobacter-BMA	1,200,365	720,218,713
GSF665-33-Flavobacterium-KMS	185,975	111,585,274
GSF665-34-Flavobacterium_hibernum	1,432,517	859,510,422
GSF665-36-Flavobacterium_hydatis	744,893	446,935,512
GSF665-39-Kaistella_koreensis	1,238,892	743,334,928
GSF665-40-Kaistella_haifense	1,067,969	640,781,490
Total	**25,364,460**	**15,218,675,963**
Average	**724,699**	**434,819,313**

Assembly statistics



[SoftGenetics Assembler: Assembly Results Statistics Report]

- **Total Reads Number: 2056329**
- Matched Reads Number: 1983986
- Unmatched Reads Number: 72343
- **Assembled Sequences Number: 61**
- Average Sequence Length: 57497
- Minimum Sequence Length: 158
- Maximum Sequence Length: 641985
- **N50 Length: 366076**

[Final Contig Merge Results Statistics Report]

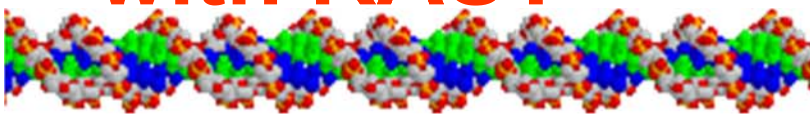
- **Final Contig Merge Sequences Number: 13**
- Final Contig Merge Average Sequence Length: 269063
- Final Contig Merge Minimum Sequence Length: 173
- Final Contig Merge Maximum Sequence Length: 856388
- **Final Contig Merge N50 Length: 586767**
- Matched Reads Count: 1977550
- Number of Matched Bases: 562514128
- **Average Read Length: 285**
- **Average Coverage: 161**
- **Reference Length: 3507364**

Genomic Comparisons

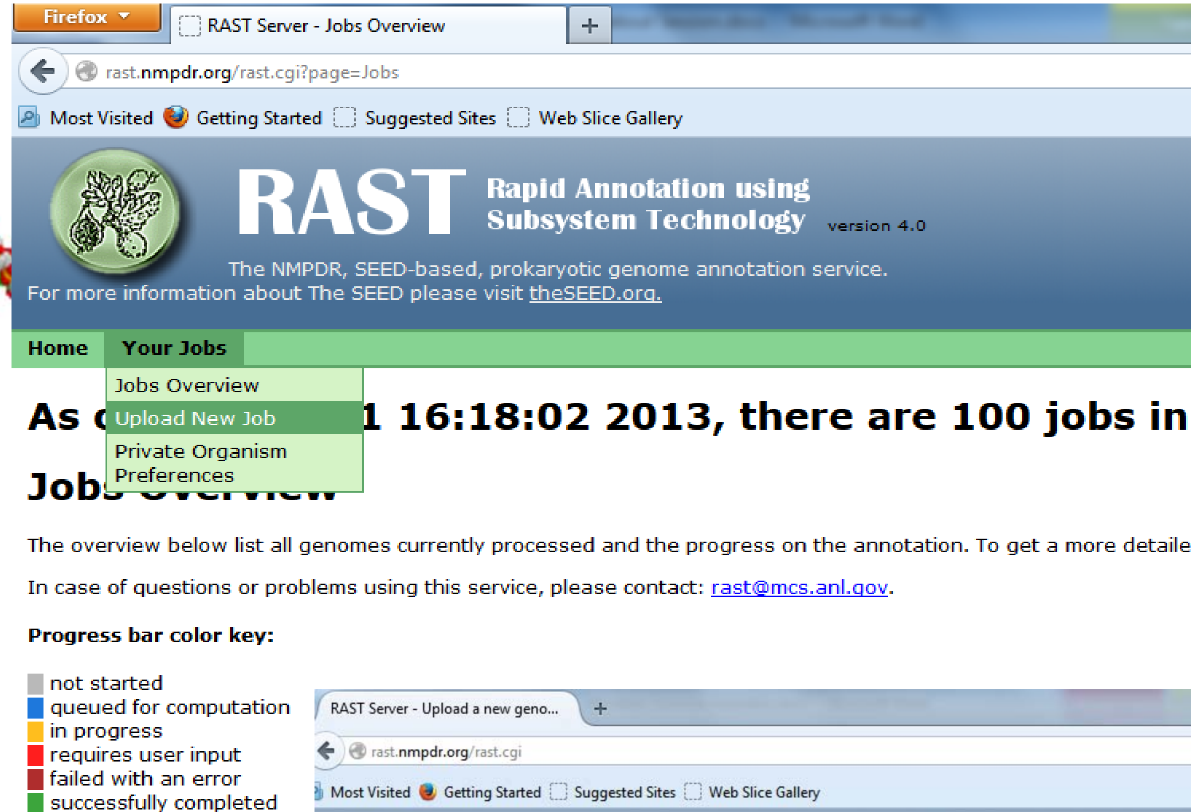
SEEKquence	Table 1. Genome Sequencing Statistics (MiSeq v3 2x300 PE)			
	Flavobacterium falloni sp. JRM	*Flavobacterium falloni sp. KMS*	*Flavobacterium hydatis*	*Flavobacterium hibernum*
Total Reads	826,175	725,291	1,489,773	2,865,011
Contigs	170	58	99	29
Average Coverage	27X	28X	54X	109X
Assembly Length	5,386,118 bp	5,620,217 bp	5,877,671 bp	5,283,662 bp
Accession	JSYO01	JSPY01	JRHH01	JPRK01

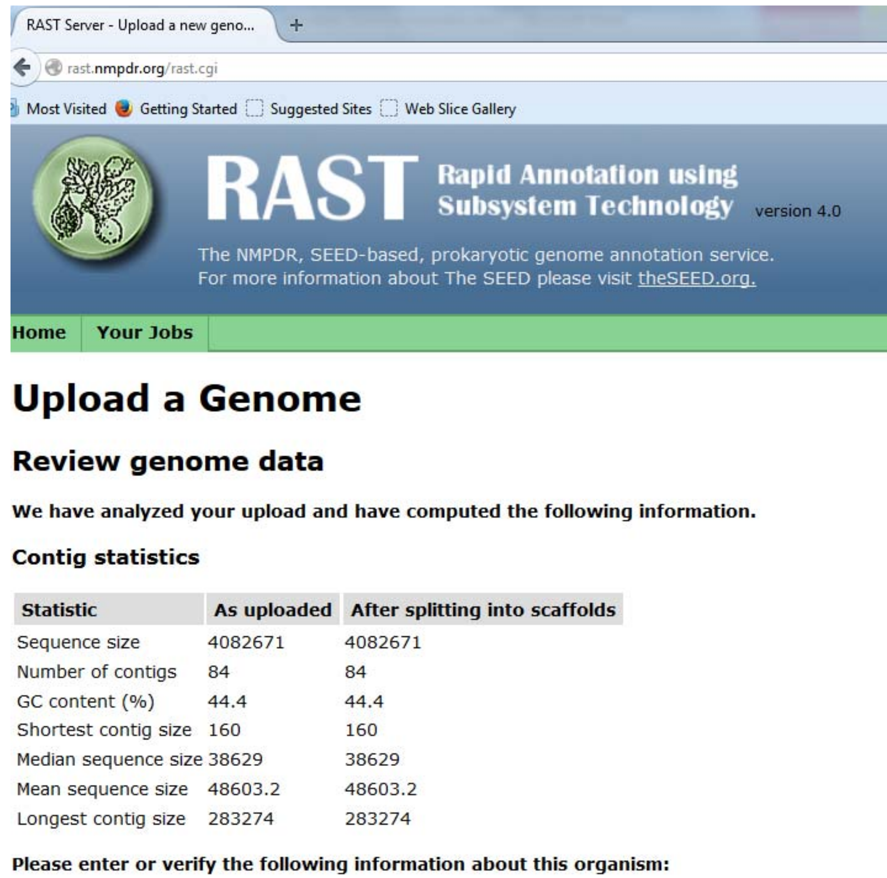
	Table 2. Estimated DNA-DNA Hybridization (species threshold = 70%) calculated by GGDC (Meier-Kolthoff. et al, 2013) and Average Nucleotide Identity (ANI) (species threshold = 95%) calculated by Kostas lab ANI Calculator http://enve-omics.ce.gatech.edu/ani/			
	Flavobacterium JRM	***Flavobacterium KMS***	***Flavobacterium hydatis***	***Flavobacterium hibernum***
Flavobacterium sp. JRM	-	74.2	41.7	23.4
Flavobacterium sp. KMS	96.95	-	40.8	23.3
Flavobacterium hydatis	90.54	90.22	-	23.3
Flavobacterium hibernum	78.36	78.25	78.38	-

Annotation with RAST



- http://rast.nmpdr.org/
- Upload genome as a .fasta file,
- Get GC mol% composition
- 1-24 hrs later…





RAST Screens

RAST Server - Jobs Overview

rast.nmpdr.org/?page=Jobs

Most Visited · Getting Started · Suggested Sites · Web Slice Gallery

As of Wed May 14 06:09:02 2014, there are 566 jobs in the RAST queue

Jobs Overview

The overview below list all genomes currently processed and the progress on the annotation. To get a more detailed report on an annotation job, please click on the progress bar graphic in the overview.

In case of questions or problems using this service, please contact: rast@mcs.anl.gov.

Progress bar color key:

- not started
- queued for computation
- in progress
- requires user input
- failed with an error
- successfully completed

Jobs you have access to :

display 50 items per page

displaying 1 - 50 of 854 next» last»

Job	Owner	ID	Name	Num contigs	Size (bp)	Creation Date	Annotation Progress	Status
								comp
151899	Newman, Jeffrey	6666666.68633	Pedobacter sp BMA	36	5046646	2014-05-04 22:38:01	[view details]	complete
151885	Newman, Jeffrey	991.3	Flavobacterium hydatis DSM 2063	187	5927954	2014-05-04 18:48:32	[view details]	complete
151881	Newman, Jeffrey	37752.3	Flavobacterium hibernum DSM 12611	62	5314106	2014-05-04 14:52:53	[view details]	complete
151879	Newman, Jeffrey	265729.9	Bacillus cibi DSM 16189	84	4082671	2014-05-04 08:42:34	[view details]	complete
151878	Newman, Jeffrey	6666666.68616	Chryseobacterium sp. LO	85	5489991	2014-05-04 07:24:17	[view details]	complete
151877	Newman, Jeffrey	445961.3	Chryseobacterium soli DSM 19298	44	4774668	2014-05-04 05:48:58	[view details]	complete

RAST Server - Job Details

rast.nmpdr.org/?page=JobDetails&job=151899

Most Visited · Getting Started · Suggested Sites · Web Slice Gallery

RAST Rapid Annotation using Subsystem Technology version 4.0

The NMPDR, SEED-based, prokaryotic genome annotation service.
For more information about The SEED please visit theSEED.org.

Home | Your Jobs | Manage Job #151899 | Jeffrey Newman

Job Details #151899

» Browse annotated genome in SEED Viewer

» Available downloads for this job: Genbank · Download · Update download files

» Share this genome with selected users

» Back to the Jobs Overview

Genome Upload has been successfully completed.

Genome ID - Name:	6666666.68633 - Pedobacter sp BMA
Job:	#151899
User:	newmanlab
Date:	Sun May 4 22:38:01 2014
Sequencing method:	unknown
Coverage:	unknown
Number of contigs:	unknown
Read length:	
Genetic code:	11
Include into SEED:	no
Preserve gene calls:	no
Automatically fix errors:	yes

RASTs – Many genes assigned to expandable subsystems

Genome	Pedobacter sp BMA
Domain	Sphingobactericeae
Taxonomy	Sphingobactericeae ; Pedobacter sp BMA
Neighbors	View closest neighbors
Size	5,046,646 bp
Number of Contigs (with PEGs)	36
Number of Subsystems	352
Number of Coding Sequences	4425
Number of RNAs	55

For each genome we offer a wide set of information to browse, compare and download.

Browse | Compare | Download | Annotate

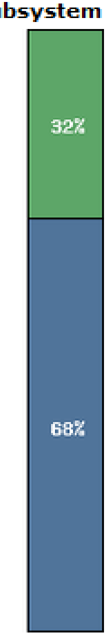
Compare the metabolic reconstruction of this organism to that of another organism.

Available comparisons are function based, sequence based or via KEGG. You can also BLAST against this organism.

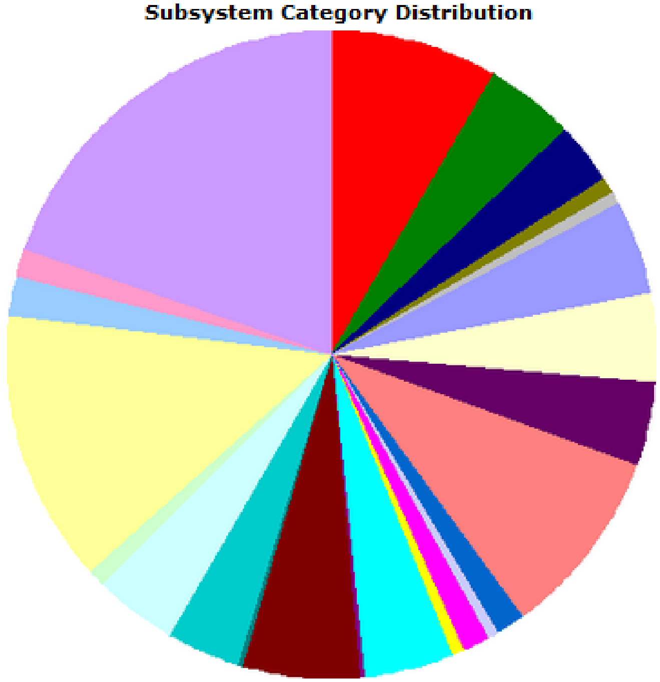
Subsystem Information

Subsystem Statistics | Features in Subsystems

Subsystem Coverage

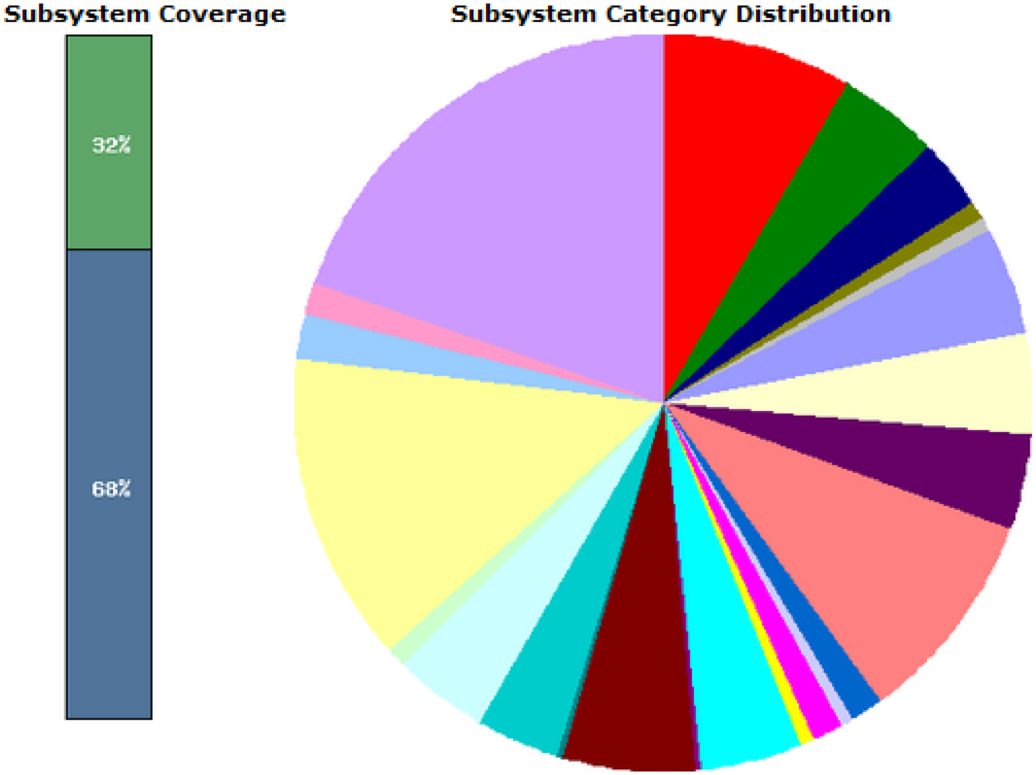


Subsystem Category Distribution



Subsystem Feature Counts

- Cofactors, Vitamins, Prosthetic Groups, Pigments (170)
- Cell Wall and Capsule (90)
- Virulence, Disease and Defense (61)
- Potassium metabolism (14)
- Photosynthesis (0)
- Miscellaneous (15)
- Phages, Prophages, Transposable elements, Plasmids (1)
- Membrane Transport (90)
- Iron acquisition and metabolism (5)
- RNA Metabolism (93)
- Nucleosides and Nucleotides (85)
- Protein Metabolism (185)
- Cell Division and Cell Cycle (30)
- Motility and Chemotaxis (11)
- Regulation and Cell signaling (29)
- Secondary Metabolism (7)
- DNA Metabolism (94)
- Regulons (6)
- Fatty Acids, Lipids, and Isoprenoids (112)
- Nitrogen Metabolism (7)
- Dormancy and Sporulation (4)
- Respiration (72)
- Stress Response (86)
- Metabolism of Aromatic Compounds (13)
- Amino Acids and Derivatives (276)
- Sulfur Metabolism (36)
- Phosphorus Metabolism (31)
- Carbohydrates (382)



Subsystem Feature Counts

- Cofactors, Vitamins, Prosthetic Groups, Pigments (170)
- Cell Wall and Capsule (90)
- Virulence, Disease and Defense (61)
  - Adhesion (0)
  - Toxins and superantigens (0)
  - Bacteriocins, ribosomally synthesized antibacterial peptides (0)
  - Resistance to antibiotics and toxic compounds (46)
    - Copper homeostasis (3)
    - Cobalt-zinc-cadmium resistance (17)
    - Multidrug Resistance, Tripartite Systems Found in Gram Negative Bacteria (6)
    - Zinc resistance (2)
    - Mercuric reductase (1)
    - Arsenic resistance (4)
    - Resistance to fluoroquinolones (4)
    - Beta-lactamase (8)
    - Resistance to chromium compounds (1)
  - Virulence, Disease and Defense - no subcategory (0)
  - Detection (0)
  - Invasion and intracellular resistance (15)
- Potassium metabolism (14)
- Photosynthesis (0)
- Miscellaneous (15)
- Phages, Prophages, Transposable elements, Plasmids (1)
  - Phage family-specific subsystems (0)
  - Transposable elements (0)
  - Phages, Prophages (1)
  - Phages, Prophages, Transposable elements, Plasmids - no subcategory (0)
  - Pathogenicity islands (0)
  - Gene Transfer Agent (GTA) (0)
  - Plasmid related functions (0)
- Membrane Transport (90)
- Iron acquisition and metabolism (5)
- RNA Metabolism (93)
- Nucleosides and Nucleotides (85)
- Protein Metabolism (185)
- Cell Division and Cell Cycle (30)
- Motility and Chemotaxis (11)
- Regulation and Cell signaling (29)
- Secondary Metabolism (7)
- DNA Metabolism (94)
- Regulons (6)
- Fatty Acids, Lipids, and Isoprenoids (112)
  - Phospholipids (20)
  - Triacylglycerols (0)
  - Fatty acids (36)
  - Fatty Acids, Lipids, and Isoprenoids - no subcategory (11)
  - Isoprenoids (45)
    - Myxoxanthophyll biosynthesis in Cyanobacteria (1)
    - Carotenoids (11)
    - Isoprenoinds for Quinones (6)
    - Isoprenoid Biosynthesis (14)
    - Polyprenyl Diphosphate Biosynthesis (4)
    - Nonmevalonate Branch of Isoprenoid Biosynthesis (7)

Annotation = Genes identified and classified by function

RAST – Many genes assigned to expandable subsystems

Genome	Pedobacter sp BMA
Domain	Sphingobactericeae
Taxonomy	Sphingobactericeae ; Pedobacter sp BMA
Neighbors	View closest neighbors
Size	5,046,646 bp
Number of Contigs (with PEGs)	36
Number of Subsystems	352
Number of Coding Sequences	4425
Number of RNAs	55

For each genome we offer a wide set of information to browse, compare and download.

Browse | Compare | Download | Annotate

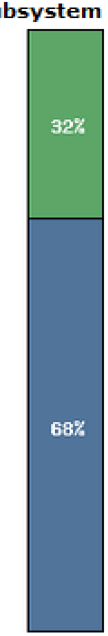
Compare the metabolic reconstruction of this organism to that of another organism.

Available comparisons are function based, sequence based or via KEGG. You can also BLAST against this organism.

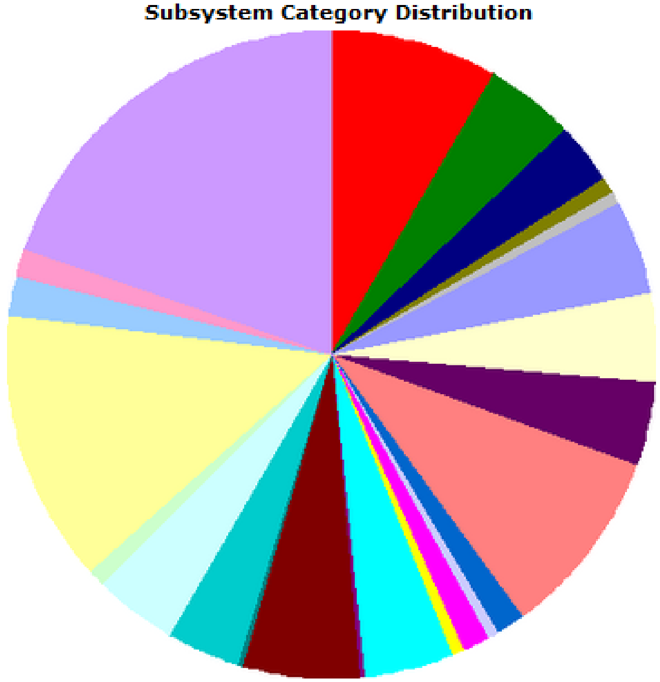
Subsystem Information

Subsystem Statistics | Features in Subsystems

Subsystem Coverage



Subsystem Category Distribution



Subsystem Feature Counts

- Cofactors, Vitamins, Prosthetic Groups, Pigments (170)
- Cell Wall and Capsule (90)
- Virulence, Disease and Defense (61)
- Potassium metabolism (14)
- Photosynthesis (0)
- Miscellaneous (15)
- Phages, Prophages, Transposable elements, Plasmids (1)
- Membrane Transport (90)
- Iron acquisition and metabolism (5)
- RNA Metabolism (93)
- Nucleosides and Nucleotides (85)
- Protein Metabolism (185)
- Cell Division and Cell Cycle (30)
- Motility and Chemotaxis (11)
- Regulation and Cell signaling (29)
- Secondary Metabolism (7)
- DNA Metabolism (94)
- Regulons (6)
- Fatty Acids, Lipids, and Isoprenoids (112)
- Nitrogen Metabolism (7)
- Dormancy and Sporulation (4)
- Respiration (72)
- Stress Response (86)
- Metabolism of Aromatic Compounds (13)
- Amino Acids and Derivatives (276)
- Sulfur Metabolism (36)
- Phosphorus Metabolism (31)
- Carbohydrates (382)

RAST – Sequence based comparison tool to ID orthologs

http://rast.nmpdr.org/seedviewer.cgi | About this service | Taylor Anspach - Outlook Web... | RAST Server - Job Details | Seed Viewer - Multi-Genom... | Newman Lab Orthology Calcul...

change organism selection

You chose to compute data for the following organisms:

Reference	Sphingomonas melonis C3 (1090316.6)
Comparison Organism 1	Novosphingobium aromaticivorans DSM 12444 (279238.21) BlastDotPlot
Comparison Organism 2	Novosphingobium nitrogenifigens DSM 19370 (983920.4) BlastDotPlot
Comparison Organism 3	Novosphingobium nitrogenifigens Y88, DSM 19370 (983920.5) BlastDotPlot
Comparison Organism 4	Sphingomonas melonis C3 (1090316.6) BlastDotPlot
Comparison Organism 5	Sphingomonas melonis DAPP-PG 224 (1090320.4) BlastDotPlot
Comparison Organism 6	Sphingomonas melonis FR1 (1090317.3) BlastDotPlot
Comparison Organism 7	Novosphingobium aromaticivorans (48935.1) BlastDotPlot
Comparison Organism 8	Sphingomonas wittichii RW1 (392499.4) BlastDotPlot
Comparison Organism 9	Sphingopyxis alaskensis RB2256 (317655.9) BlastDotPlot

Percent protein sequence identity

Bidirectional best hit	100	99.9	99.8	99.5	99	98	95	90	80	70	60	50	40	30	20	10
Unidirectional best hit	100	99.9	99.8	99.5	99	98	95	90	80	70	60	50	40	30	20	10

export table | clear all filters

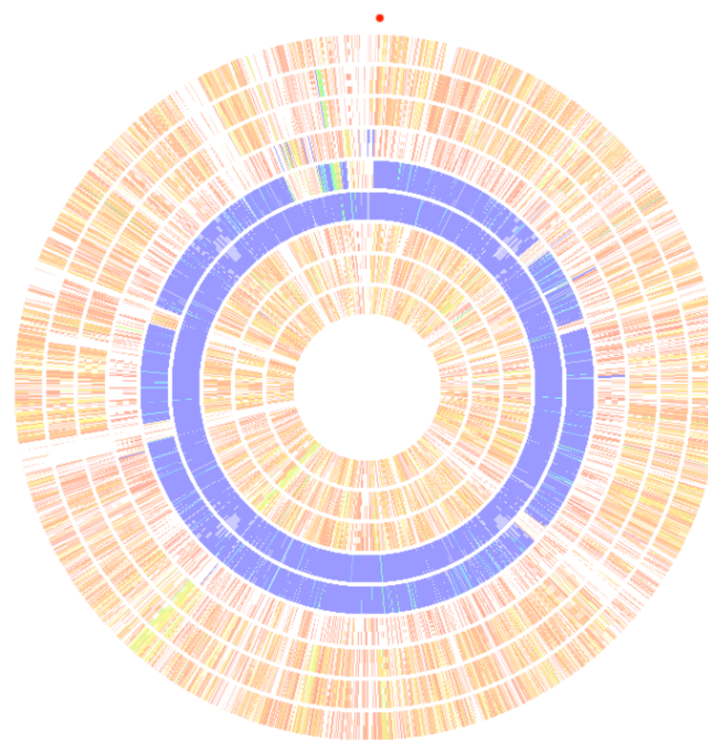
display 30 items per page

«first «prev | displaying 2 - 31 of 3706 | next» last»

percent identity 279238.21 <
percent identity 983920.5 <
percent identity 1090320.4 <
percent identity 48935.1 <
percent identity 317655.9 <
percent identity 983920.4 <
percent identity 1090316.6 <
percent identity 1090317.3 <
percent identity 392499.4 <

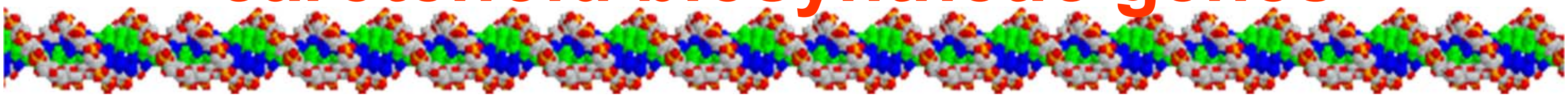
1090316.6			279238.21			983920.4			983920.5			1090316.6		
Contig (all)	Gene	Length	Hit	Contig (all)	Gene	Hit	Contig (all)	Gene	Hit	Contig (all)	Gene	Hit	Contig (all)	Gene
1	2	97	-			-			-			bi	1	16
1	3	184	uni	2	3527	-			-			bi	1	15
1	4	155	-			-			-			bi	1	14
1	5	510	-			-			-			-		
1	6	234	uni	1	1499	uni	9	1438	uni	61	3817	uni	2	3556
1	7	504	uni	1	1498	uni	1	1	uni	61	3816	uni	2	3555
1	8	1040	-			bi	1	292	bi	30	1865	-		
1	9	178	-			-			-			-		
1	10	292	-			bi	1	290	bi	30	1867	-		
1	11	591	-			bi	1	289	bi	30	1868	-		
1	12	612	-			bi	1	288	bi	30	1869	bi	1	764
1	13	55	-			-			-			-		
1	14	155	-			-			-			bi	1	4
1	15	184	uni	2	3527	-			-			bi	1	3
1	16	97	-			-			-			bi	1	2
1	17	90	-			-			-			bi	1	1
1	18	493	-			bi	4	1156	bi	8	207	bi	1	1320
1	19	310	bi	1	3176	bi	5	1244	bi	29	1618	uni	1	749
1	20	266	bi	1	3284	bi	43	3562	bi	44	3313	-		
1	21	478	bi	1	3274	bi	43	3551	bi	44	3302	uni	1	973
1	22	250	-			-			-			-		
1	23	1514	bi	1	3275	bi	43	3552	bi	44	3303	-		
1	24	64	-			-			-			-		
1	25	225	bi	1	3175	bi	5	1243	bi	29	1617	uni	1	1675
1	26	121	-			bi	2	742	bi	34	2391	-		
1	27	788	bi	1	712	bi	13	1972	bi	20	1089	-		
1	28	372	bi	1	711	bi	13	1973	bi	20	1088	-		
1	29	113	bi	1	837	uni	12	1606	uni	23	1446	uni	1	1516
1	30	122	bi	1	709	bi	13	1974	bi	20	1087	-		
1	31	68	bi	1	708	bi	13	1975	bi	20	1086	-		

1090316.6		1090320.4			1090317.3			48935.1			392499.4			317655.9		
Contig (all)	Gene	Hit	Contig (all)	Gene	Hit	Contig (all)	Gene	Hit	Contig (all)	Gene	Hit	Contig (all)	Gene	Hit	Contig (all)	Gene
1	2	-			uni	1	1542	-			-			-		
1	3	uni	3	3836	uni	1	1543	uni	1	26	uni	2	332	-		
1	4	-			uni	1	1544	-			-			-		
1	5	-			bi	1	1545	-			-			-		
1	6	-			uni	1	1546	uni	114	1298	uni	1	14	uni	1	394
1	7	-			uni	1	1547	uni	114	1299	uni	1	13	uni	1	395
1	8	-			bi	1	1549	-			-			-		
1	9	-			bi	1	1550	-			-			-		
1	10	-			bi	1	1551	-			-			-		
1	11	-			bi	1	1552	-			-			-		
1	12	-			bi	1	1553	-			-			-		
1	13	-			bi	1	1554	-			-			-		
1	14	-			bi	1	1544	-			-			-		
1	15	uni	3	3836	bi	1	1543	uni	1	26	bi	2	332	-		
1	16	-			bi	1	1542	-			-			-		
1	17	-			uni	1	1541	-			-			-		
1	18	uni	1	3373	bi	1	1559	-			bi	3	2664	uni	1	48
1	19	bi	1	666	bi	1	1560	bi	49	431	bi	3	987	bi	1	2062
1	20	bi	1	665	bi	1	1561	bi	167	3004	bi	3	986	bi	1	2063
1	21	bi	1	664	bi	1	1562	bi	167	3014	bi	3	985	bi	1	2064
1	22	bi	1	663	bi	1	1563	-			-			bi	1	2065
1	23	bi	1	662	bi	1	1564	bi	167	3013	bi	3	983	bi	1	2066
1	24	bi	1	661	bi	1	1565	-			-			-		
1	25	bi	1	660	bi	1	1566	bi	49	430	bi	3	4732	bi	1	2068
1	26	bi	1	659	bi	1	1567	-			-			bi	1	1159
1	27	bi	1	658	bi	1	1568	bi	140	1928	bi	3	2636	bi	1	2648
1	28	bi	1	657	bi	1	1569	bi	140	1927	bi	3	2637	bi	1	2647
1	29	bi	1	656	bi	1	1570	bi	174	3470	bi	3	2638	bi	1	684
1	30	bi	1	655	bi	1	1571	bi	140	1925	bi	3	2639	bi	1	2796
1	31	bi	1	654	bi	1	1572	bi	140	1924	bi	3	2640	bi	1	2797



10:03 PM 11/10/2013

C.populense lacks carotenoid biosynthetic genes

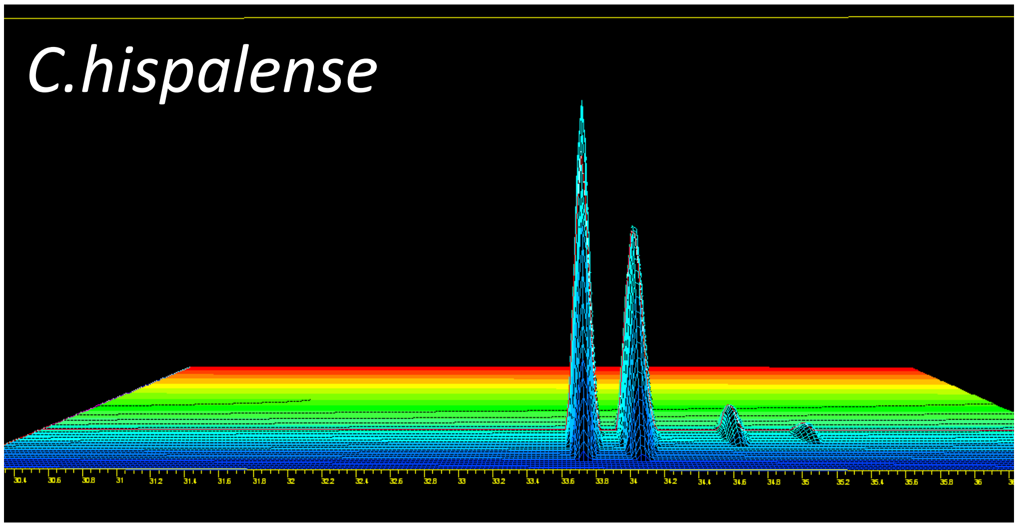


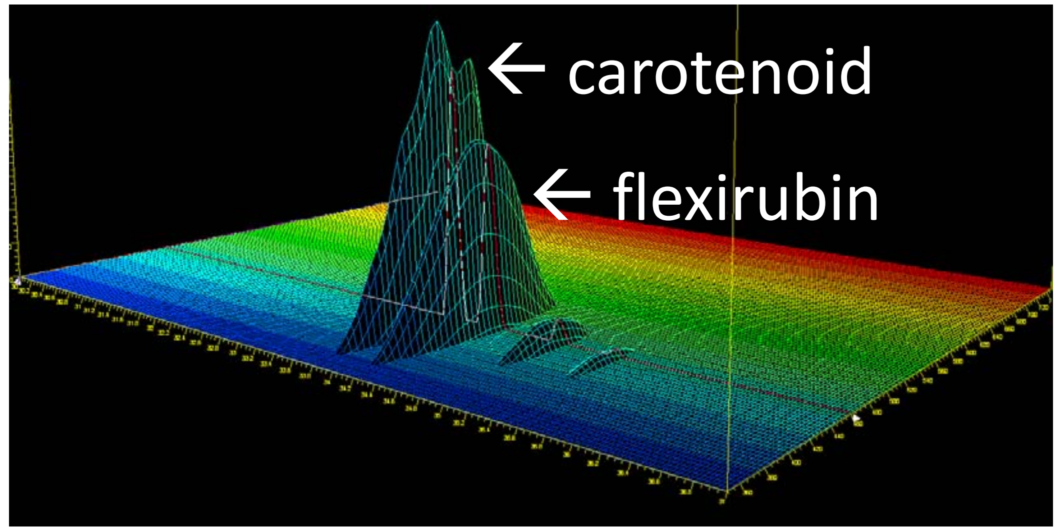
C.hispalense | *C.populense*

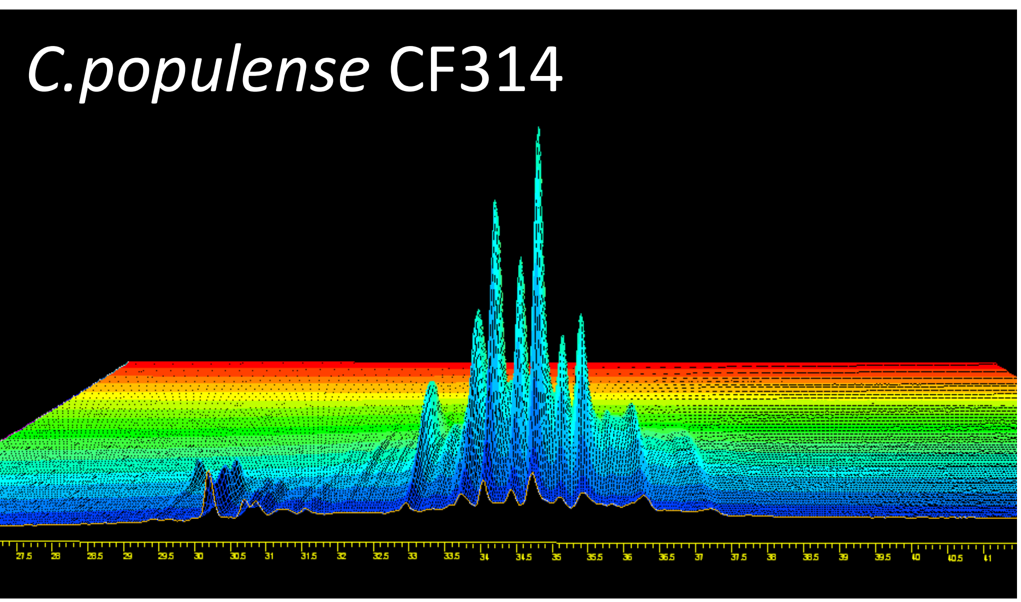
491205.4			558151.4			1121286.3			236814.3			525257.7			1121287.3			1218103.4			558152.3			445961.3			59732.8			307480.3		
Contig	Gene	Length	Hit	Contig	Gene	Hit	Contig	Gene	Hit	Contig	Gene	Hit	Contig	Gene	Hit	Contig	Gene	Hit	Contig	Gene	Hit	Contig	Gene	Hit	Contig	Gene	Hit	Contig	Gene	Hit	Contig	Gene
all				all			all			all			all			all			all			all			all			all			all	
8	991	794	-			bi	4	1671	bi	5	858	bi	38	3567	bi	4	1495	-			-			-			bi	108	4042	-		
8	992	637	-			bi	4	1670	bi	5	859	bi	38	3568	bi	4	1496	-			-			bi	2	270	bi	108	4041	-		
8	993	482	-			bi	4	1669	bi	5	860	bi	38	3569	bi	4	1497	-			-			bi	2	271	bi	108	4040	-		
8	994	362	-			bi	4	1668	bi	5	861	bi	38	3570	bi	4	1498	-			-			bi	2	272	bi	108	4039	-		
8	995	324	-			bi	4	1667	bi	5	862	bi	38	3571	bi	4	1499	-			-			bi	2	273	bi	108	4038	-		
8	996	678	-			bi	4	1666	bi	5	863	bi	38	3572	bi	4	1501	-			-			bi	2	274	bi	108	4036	-		
8	997	199	-			-			bi	5	1990	bi	38	3573	bi	4	1502	-			-			-			-			-		
8	998	291	bi	1	523	bi	5	2037	bi	5	1991	bi	38	3574	bi	4	1503	bi	19	4242	bi	37	2234	bi	11	2704	bi	34	1232	bi	27	4951
8	999	396	bi	1	50	bi	4	1664	bi	5	867	bi	38	3575	bi	4	1504	-			bi	82	3758	-			bi	108	4032	-		
8	1000	351	bi	1	51	bi	4	1663	bi	5	868	bi	38	3576	bi	4	1505	-			bi	82	3759	-			bi	108	4031	-		
8	1001	150	-			bi	17	3765	bi	5	114	bi	7	907	bi	4	1512	bi	7	1747	-			bi	13	303	-					
8	1002	223	-			bi	17	3766	bi	5	2174	bi	7	970	bi	4	1513	bi	7	1762	-			bi	17	341	-				19	4081
8	1003	490	uni	11	3029	bi	17	3767	bi	5	2173	bi	7	953	bi	4	1514	bi	7	1563	uni	24	1021	bi	17	341	-				19	4080
8	1004	279	-			bi	17	3768	bi	5	2172	bi	7	954	bi	4	1515	bi	7	1564	-			bi	17	341	-				19	4079
8	1005	152	-			bi	17	3769	bi	5	2170	bi	7	956	bi	4	1516	bi	7	1566	-			bi	17	340	-				19	4077
8	1006	238	-			bi	17	3770	bi	5	2169	bi	7	957	bi	4	1517	bi	7	1567	-			bi	17	340	-					4076
8	10[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		-			-			-			-			-			-					
8	10[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2234	bi	1	511	bi	4	1403	bi	2	273	bi	75	3573	bi	5	753	bi	16	370	bi	2	1192
8	10[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2241	bi	10	1314	bi	4	1396	bi	10	2705	uni	36	2185	bi	17	3369	uni	102	3947	bi	5	1891
8	1010	160	-			bi	4	1579	-			bi	42	3938	bi	4	1519	bi	10	2602	bi	28	1303	-			-			bi	5	1956
8	1011	78	-			bi	4	1578	bi	5	97	bi	4	651	bi	4	1520	-			-			bi	13	2986	-			bi	5	1967
8	1012	966	uni	3	1058	-			bi	5	877	-			bi	4	1521	uni	11	2751	uni	21	825	bi	17	3444	-			bi	5	1982
8	1013	204	uni	1	225	bi	4	1577	uni	5	2523	bi	27	2599	bi	4	1522	bi	10	2634	bi	28	1304	bi	17	3431	uni	20	644	bi	5	1970
8	1014	73	-			bi	17	3764	bi	5	886	-			bi	4	1523	-			-			bi	13	3025	bi	25	886	-		
8	1015	52	-			-			-			-			-			-			-			-			-			-		
8	1016	77	-			bi	17	3771	bi	5	1964	bi	21	2217	bi	4	1525	bi	19	4260	-			-			bi	72	2999	bi	5	1968
8	1017	332	uni	12	4297	bi	17	3772	bi	5	2229	bi	42	3977	bi	4	1526	bi	10	2555	bi	7	264	bi	2	249	bi	2	21	uni	11	3903

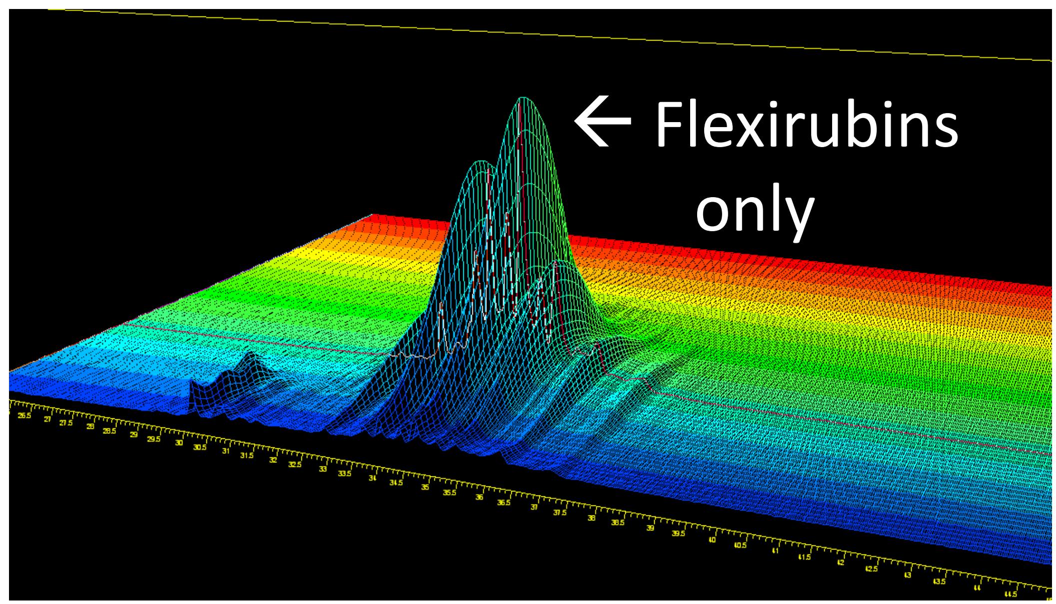
fig|49120[illegible]4.peg.1006
location: 11+0+27+18+48_length:875329 110864 111577
function: Lycopene cyclase

Explain phenotypic differences – e.g. Pigment "Landscapes"

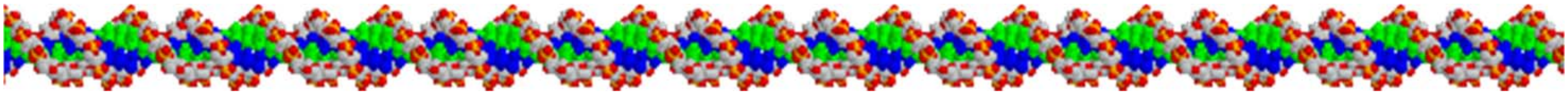








Sequence-Based Comparison to Calculate AAI



You chose to compute data for the following organisms:

Reference	Chryseobacterium hispalense DSM 25574 (491205.4)
Comparison Organism 1	Chryseobacterium gleum F93, ATCC 35910 (525257.7) BlastDotPlot
Comparison Organism 2	Chryseobacterium sp. CF314 (1144316.4) BlastDotPlot

Percent protein sequence identity

Bidirectional best hit	100	99.9	99.8	99.5	99	98	95	90	80	70	60	50	40	30	20	10
Unidirectional best hit	100	99.9	99.8	99.5	99	98	95	90	80	70	60	50	40	30	20	10

export table   clear all filters

display 30 items per page

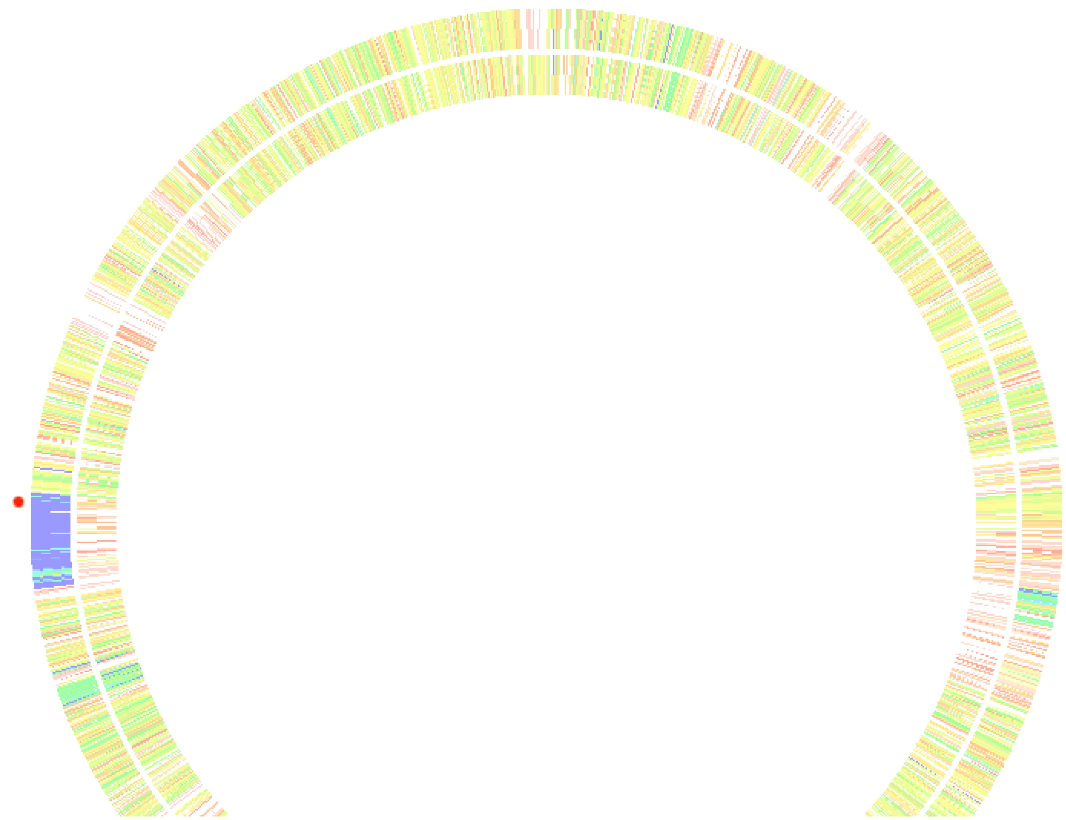
«first «prev   displaying 3028 - 3057 of 4006   next» last»

percent identity 525257.7 <   percent identity 1144316.4 <

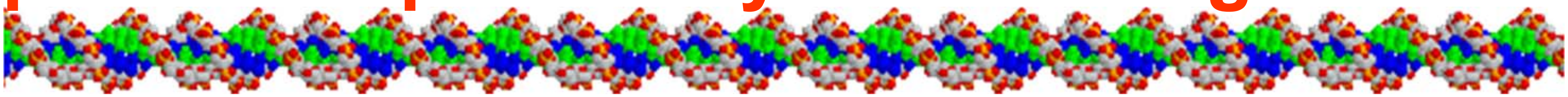
491205.4			525257.7			1144316.4		
Contig all	Gene	Length	Hit all	Contig all	Gene	Hit all	Contig all	Gene
21	3028	360	bi	1	241	bi	22	1525
21	3029	270	bi	1	240	uni	51	2777
21	3030	680	bi	1	239	uni	2	181
21	3031	92	bi	1	238	uni	14	1009
21	3032	580	bi	1	237	uni	49	2655
21	3033	298	bi	1	236	bi	2	183
21	3034	293	bi	1	235	bi	2	184
21	3035	249	bi	1	234	uni	1	1
21	3036	440	bi	1	233	bi	2	185
21	3037	1455	bi	1	232	bi	2	186
21	3038	41	bi	1	231	-		
21	3039	208	bi	1	230	bi	69	3277
21	3040	348	bi	1	229	uni	3	200
21	3041	92	bi	1	228	uni	59	2913
21	3042	258	bi	1	227	-		
21	3043	158	bi	1	226			
21	3044	110	bi	1	22[illegible]			
21	3045	408	bi	1	22[illegible]			
21	3046	106	bi	1	22[illegible]			
21	3047	392	bi	1	22[illegible]			

fig|525257.7.peg.224
location: NZ_ACKQ01000001 253632 254855
length: 407
identity: 1
function: Tyrosine type site-specific recombinase

rast.nmpdr.org/seedviewer.cgi?page=Annotation&feature=fig|525257.7.peg.224



Sequence Based Comparison provides protein seq similarity for shared genes…..



Microsoft Excel - Fsucc Venn for poster.xlsx

Contig	Length	Gene id	Gene	function	Hit	Contig	Gene	Gene id	percent	function	Hit	Contig	Gene	Gene id	percent	function
22	149	fig|1450525.4.pe	3118	hypothetical protein	-				0		bi	1	2298	fig|37668	61.9	hypothetical protein
23	162	fig|1450525.4.pe	3124	Pectate lyase (EC 4.2.2.2)	-				0		bi	1	2696	fig|37668	73.91	Pectate lyase (EC 4.2.2.2)
25	72	fig|1450525.4.pe	3173	Glucosamine-6-phosphate deaminase (EC 3.5.99.6)	-				0		bi	1	4817	fig|37668	91.55	Glucosamine-6-phosphate deaminase (EC 3.5.99
26	408	fig|1450525.4.pe	3177	Gluconate permease, Bsu4004 homolog	-				0		bi	1	700	fig|37668	75.68	Gluconate permease, Bsu4004 homolog
26	156	fig|1450525.4.pe	3178	Endoribonuclease L-PSP	-				0		bi	1	701	fig|37668	92.26	Endoribonuclease L-PSP
26	216	fig|1450525.4.pe	3180	4-Hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16) @	-				0		bi	1	703	fig|37668	73.36	4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16)
26	372	fig|1450525.4.pe	3181	low-specificity D-threonine aldolase	-				0		bi	1	704	fig|37668	64.42	low-specificity D-threonine aldolase
26	346	fig|1450525.4.pe	3182	Membrane dipeptidase (EC 3.4.13.19)	-				0		bi	1	705	[illegible]	87.54	Membrane dipeptidase (EC 3.4.13.19)
26	259	fig|1450525.4.pe	3183	Transcriptional repressor of the fructose operon, D	-				0		bi	1	706	[illegible]	86.77	Transcriptional repressor of the fructose operon
27	245	fig|1450525.4.pe	3188	Transcriptional regulator, AraC family	-				0		bi	1	4563	[illegible]	37.68	regulatory protein; PcrR
28	68	fig|1450525.4.pe	3197	hypothetical protein	-				0		bi	1	1909	fig|37668	36.67	hypothetical protein
29	127	fig|1450525.4.pe	3205	hypothetical protein	-				0		bi	1	2828	fig|37668	40.54	hypothetical protein
32	228	fig|1450525.4.pe	3232	hypothetical protein	-				0		bi	1	769	fig|37668	34.87	hypothetical protein
38	38	fig|1450525.4.pe	3260	Ribonucleotide reductase of class Ia (aerobic), bet	-				0		bi	1	4302	fig|37668	89.19	Ribonucleotide reductase of class Ia (aerobic), b
39	94	fig|1450525.4.pe	3261	hypothetical protein	-				0		bi	1	4967	fig|37668	61.96	hypothetical protein
1	258	fig|1450525.4.pe	38	Possible restriction endonuclease	-				0		-				0	
1	274	fig|1450525.4.pe	56	Mobile element protein	-				0		-				0	
1	484	fig|1450525.4.pe	171	Predicted transcriptional regulator containing an I	-				0		-				0	
1	232	fig|1450525.4.pe	190	Sll8048 protein	-				0		-				0	
1	1568	fig|1450525.4.pe	191	Type I restriction-modification system, M subunit	-				0		-				0	
1	317	fig|1450525.4.pe	202	COG1242: Predicted Fe-S oxidoreductase	-				0		-				0	
1	257	fig|1450525.4.pe	246	probable integral membrane protein Cj1166c	-				0		-				0	
1	163	fig|1450525.4.pe	247	FIG001826: putative inner membrane protein	-				0		-				0	
1	829	fig|1450525.4.pe	262	Succinoglycan biosynthesis protein	-				0		-				0	
2	558	fig|1450525.4.pe	303	Formate--tetrahydrofolate ligase (EC 6.3.4.3)	-				0		-				0	
2	364	fig|1450525.4.pe	373	Cyanophycinase (EC 3.4.15.6)	-				0		-				0	
2	331	fig|1450525.4.pe	384	Homoserine kinase (EC 2.7.1.39)	-				0		-				0	
2	461	fig|1450525.4.pe	402	Type I restriction-modification system, specificity s	-				0		-				0	
2	76	fig|1450525.4.pe	432	Helix-turn-helix motif	-				0		-				0	
2	319	fig|1450525.4.pe	434	HipA protein	-				0		-				0	
2	68	fig|1450525.4.pe	448	DNA-binding domain of ModE	-				0		-				0	
2	172	fig|1450525.4.pe	451	Ubiquinol-cytochrome C reductase iron-sulfur sub	-				0		-				0	
2	350	fig|1450525.4.pe	454	Nitrate/nitrite transporter	-				0		-				0	
2	501	fig|1450525.4.pe	460	Cytochrome c552 precursor (EC 1.7.2.2)	-				0		-				0	
2	193	fig|1450525.4.pe	461	Cytochrome c nitrite reductase, small subunit NrfH	-				0		-				0	
2	79	fig|1450525.4.pe	475	B12 binding domain / kinase domain / Methylmalc	-				0		-				0	
2	119	fig|1450525.4.pe	490	Helix-turn-helix motif	-				0		-				0	

JOURNAL OF BACTERIOLOGY, Sept. 2005, p. 6258–6264 Vol. 187, No. 18
0021-9193/05/$08.00+0 doi:10.1128/JB.187.18.6258–6264.2005

Towards a Genome-Based Taxonomy for Prokaryotes

Konstantinos T. Konstantinidis[1,2] and James M. Tiedje[1,2,3]*

Center for Microbial Ecology[1] and Departments of Crop and Soil Sciences[2] and Microbiology and Molecular Genetics,[3] Michigan State University, East Lansing, Michigan

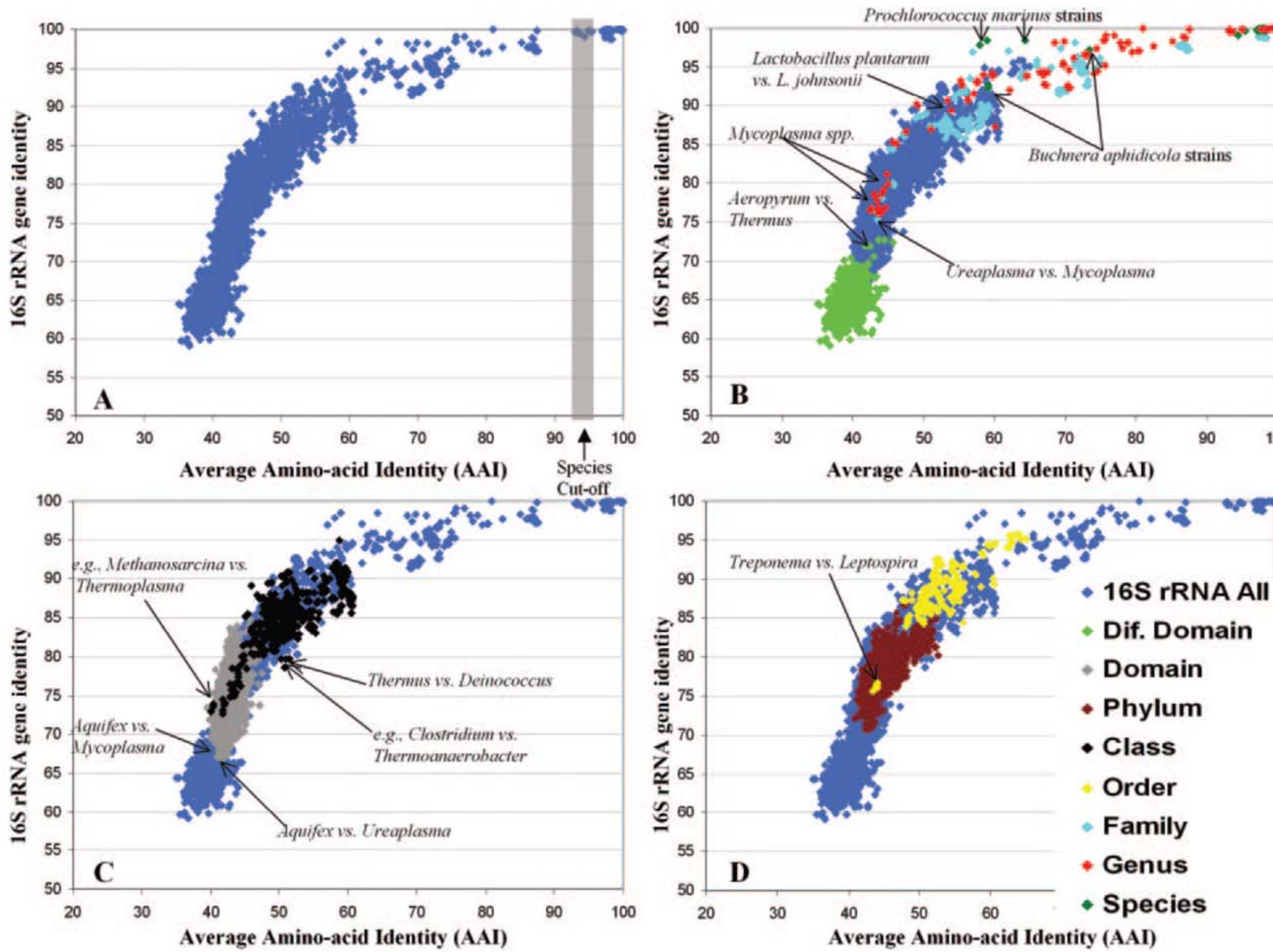
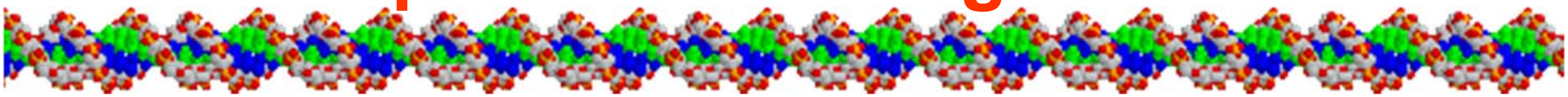


FIG. 3. Relationships between 16S rRNA, AAI, and taxonomic information for the 175 sequenced genomes. Each dot represents a comparison

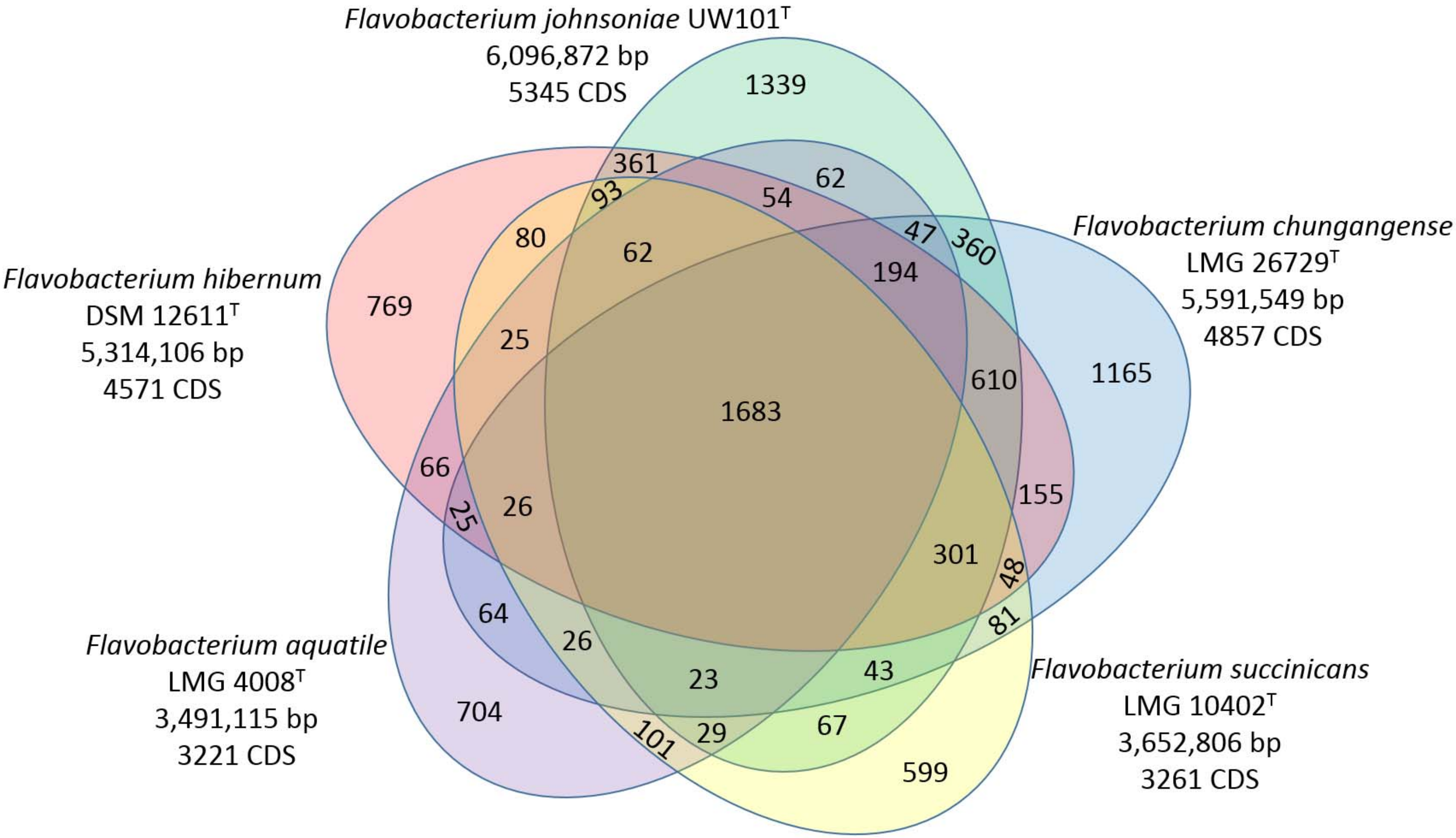
Sequence Based Comparison can ID unique and shared genes…..



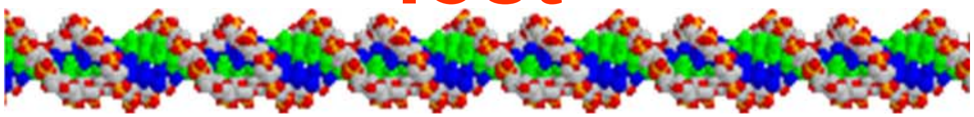
Microsoft Excel - Fsucc Venn for poster.xlsx

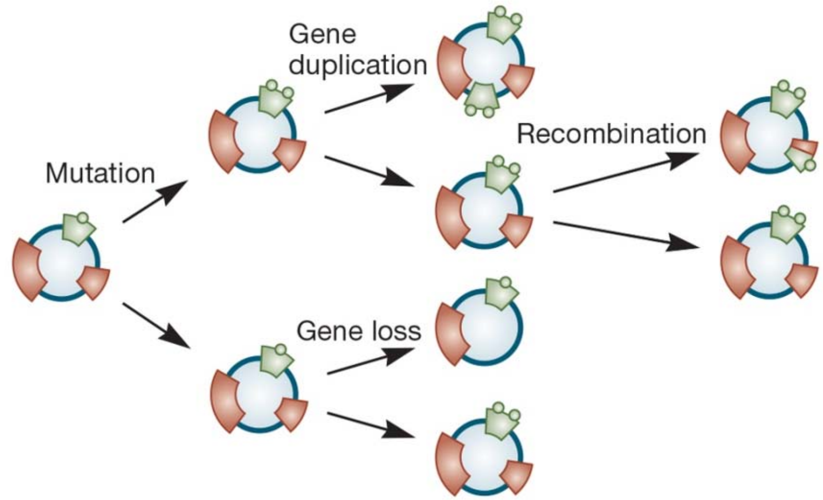
Contig	Length	Gene id	Gene	function	Hit	Contig	Gene	Gene id	percent	function	Hit	Contig	Gene	Gene id	percent	function
22	149	fig|1450525.4.pe	3118	hypothetical protein	-				0		bi	1	2298	fig|37668	61.9	hypothetical protein
23	162	fig|1450525.4.pe	3124	Pectate lyase (EC 4.2.2.2)					0		bi	1	2696	fig|37668	73.91	Pectate lyase (EC 4.2.2.2)
25	72	fig|1450525.4.pe	3173	Glucosamine-6-phosphate deaminase (EC 3.5.99.6)					0		bi	1	4817	fig|37668	91.55	Glucosamine-6-phosphate deaminase (EC 3.5.99
26	408	fig|1450525.4.pe	3177	Gluconate permease, Bsu4004 homolog	-				0		bi	1	700	fig|37668	75.68	Gluconate permease, Bsu4004 homolog
26	156	fig|1450525.4.pe	3178	Endoribonuclease L-PSP	-				0		bi	1	701	fig|37668	92.26	Endoribonuclease L-PSP
26	216	fig|1450525.4.pe	3180	4-Hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16) @	-				0		bi	1	703	fig|37668	73.36	4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16)
26	372	fig|1450525.4.pe	3181	low-specificity D-threonine aldolase	-				0		bi	1	704	fig|37668	64.42	low-specificity D-threonine aldolase
26	346	fig|1450525.4.pe	3182	Membrane dipeptidase (EC 3.4.13.19)	-				0		bi	1	705	fig|37668	87.54	Membrane dipeptidase (EC 3.4.13.19)
26	259	fig|1450525.4.pe	3183	Transcriptional repressor of the fructose operon, D	-				0		bi	1	706	fig|37668	86.77	Transcriptional repressor of the fructose operon
27	245	fig|1450525.4.pe	3188	Transcriptional regulator, AraC family	-				0		bi	1	4563	fig|37668	37.68	regulatory protein; PcrR
28	68	fig|1450525.4.pe	3197	hypothetical protein	-				0		bi	1	1909	fig|37668	36.67	hypothetical protein
29	127	fig|1450525.4.pe	3205	hypothetical protein	-				0		bi	1	2828	fig|37668	40.54	hypothetical protein
32	228	fig|1450525.4.pe	3232	hypothetical protein	-				0		bi	1	769	fig|37668	34.87	hypothetical protein
38	38	fig|1450525.4.pe	3260	Ribonucleotide reductase of class Ia (aerobic), bet	-				0		bi	1	4302	fig|37668	89.19	Ribonucleotide reductase of class Ia (aerobic), b
39	94	fig|1450525.4.pe	3261	hypothetical protein	-				0		bi	1	4967	fig|37668	61.96	hypothetical protein
1	258	fig|1450525.4.pe	38	Possible restriction endonuclease	-				0		-				0	
1	274	fig|1450525.4.pe	56	Mobile element protein	-				0		-				0	
1	484	fig|1450525.4.pe	171	Predicted transcriptional regulator containing an I	-				0		-				0	
1	232	fig|1450525.4.pe	190	Sll8048 protein	-				0		-				0	
1	1568	fig|1450525.4.pe	191	Type I restriction-modification system, M subunit	-				0		-				0	
1	317	fig|1450525.4.pe	202	COG1242: Predicted Fe-S oxidoreductase	-				0		-				0	
1	257	fig|1450525.4.pe	246	probable integral membrane protein Cj1166c	-				0		-				0	
1	163	fig|1450525.4.pe	247	FIG001826: putative inner membrane protei	-				0		-				0	
1	829	fig|1450525.4.pe	262	Succinoglycan biosynthesis protein	-				0		-				0	
2	558	fig|1450525.4.pe	303	Formate--tetrahydrofolate ligase (EC 6.3.4.3	-				0		-				0	
2	364	fig|1450525.4.pe	373	Cyanophycinase (EC 3.4.15.6)	-				0		-				0	
2	331	fig|1450525.4.pe	384	Homoserine kinase (EC 2.7.1.39)	-				0		-				0	
2	461	fig|1450525.4.pe	402	Type I restriction-modification system, specificity s	-				0		-				0	
2	76	fig|1450525.4.pe	432	Helix-turn-helix motif	-				0		-				0	
2	319	fig|1450525.4.pe	434	HipA protein	-				0		-				0	
2	68	fig|1450525.4.pe	448	DNA-binding domain of ModE	-				0		-				0	
2	172	fig|1450525.4.pe	451	Ubiquinol-cytochrome C reductase iron-sulfur sub	-				0		-				0	
2	350	fig|1450525.4.pe	454	Nitrate/nitrite transporter	-				0		-				0	
2	501	fig|1450525.4.pe	460	Cytochrome c552 precursor (EC 1.7.2.2)					0		-				0	
2	193	fig|1450525.4.pe	461	Cytochrome c nitrite reductase, small subunit NrfH	-				0		-				0	
2	79	fig|1450525.4.pe	475	B12 binding domain / kinase domain / Methylmalc	-				0		-				0	
2	119	fig|1450525.4.pe	490	Helix-turn-helix motif	-				0		-				0	

Identify Core, Genus or Family-Specific Genes

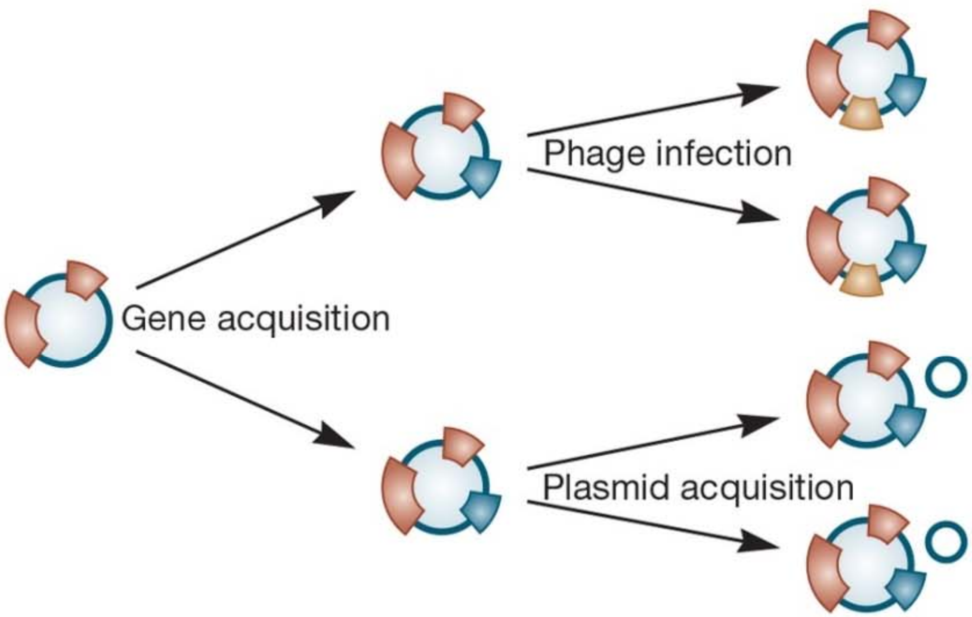


Genes can be gained and lost

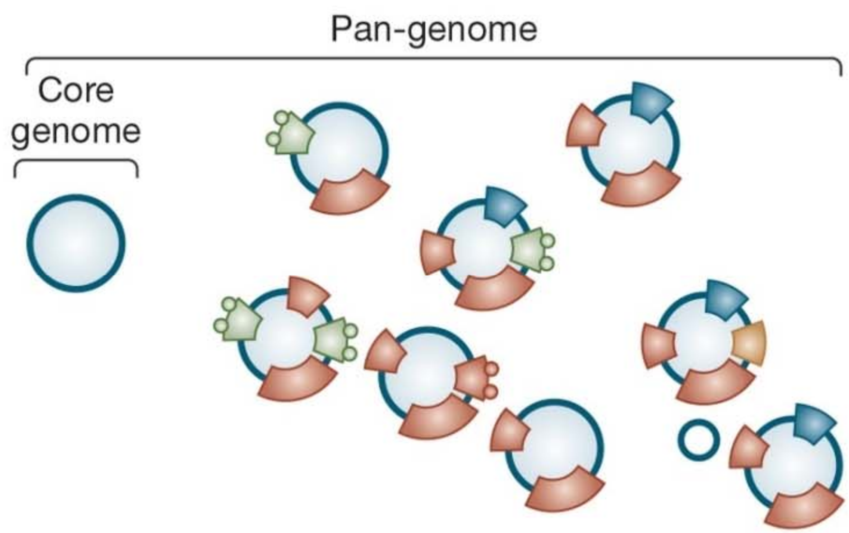




(a) Mechanisms of genetic variation within a homogeneous population

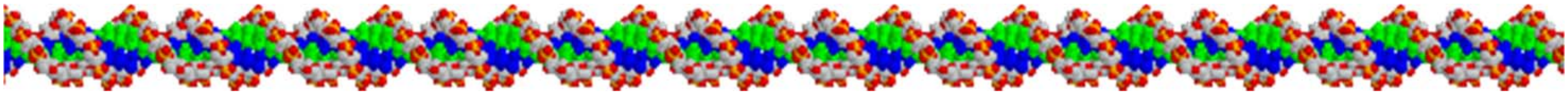


(a) Mechanisms of HGT



(b) Core and pan-genome

Reciprocal Orthology Score Average (ROSA)

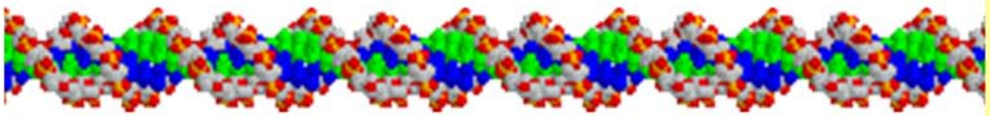


- ROSA takes into consideration both
  - similarity of orthologs at the protein level (AAI)
  - % of genome composed of orthologs (%BBH)
    - Differences in reference genome size create different %BBH values in reciprocal comparisons
- **Orthology Score (OS) = AAI^2 * %BBH**
- **ROSA = $(OS_{AB} + OS_{BA})/2$**

>~65=same species; 35-65=same genus diff species; 15-35=same family diff genus; <15=diff family

Staphylococcus (3) no type comparisons	**ROSA**	**93062.4**	**367830.3**	**93061.3**	**282459.1**	**158878.1**	**273036.3**	**282458.1**	**176279.3**	**176280.1**	**279808.3**	**342451.4**
Staphylococcus aureus subsp. aureus COL	**93062.4**											
Staphylococcus aureus subsp. aureus USA300	**367830.3**	94.8										
Staphylococcus aureus subsp. aureus NCTC 8325	**93061.3**	93.97	92.73									
Staphylococcus aureus subsp. aureus MSSA476	**282459.1**	90.39	90.1	90.85								
Staphylococcus aureus subsp. aureus Mu50	**158878.1**	88.01	88.25	88.15	87.5							
Staphylococcus aureus RF122	**273036.3**	86.78	85.6	86.89	86.02	86.18						
Staphylococcus aureus subsp. aureus MRSA252	**282458.1**	85.95	86.56	85.23	87.8	84.86	83.67					
Staphylococcus epidermidis RP62A	**176279.3**	43.49	43.38	42.5	43.55	43.37	43.55	43.97				
Staphylococcus epidermidis ATCC 12228	**176280.1**	43.25	43.25	42.62	43.7	42.69	43.71	43.22	86.44			
Staphylococcus haemolyticus JCSC1435	**279808.3**	41	41.28	40.73	41.15	40.63	41.32	41.24	44.94	44.9		
Staphylococcus saprophyticus ATCC 15305^T	**342451.4**	38.35	38.33	38.01	38.62	37.65	38.63	38.35	39.6	39.79	40.18	

Reciprocal Orthology Score Average (ROSA)



Newman Lab Orthology Score Calculator (for single .tsv files)

INSTRUCTIONS:
1) "Export file" from RAST, Sequence Comparison Tool output.
2) Browse for the file below.
3) Click the "Submit" button.
4) Copy and Paste Results Table to a Separate Spreadsheet or Word Processor Document.

Browse... No file selected.

Submit Reset

A Perl script for this analysis that takes the filename on the command line and produces the results above (tab-delimited on standard output) is available here.

This file is a set of sample data to test the Orthology Score calculator.

A calculator to determine the Reciprocal Orthology Score Average (**ROSA**) from a set of up to 11 files using the same genomes but with each as a reference is available here.

When using this website or the Perl script for your study, please cite the following article:
Newman J.D, Krebs, J.E., Gale, A.N., Anspach, T.J., Kirk, K.E., Sontag, T.C., Keyser, V.K., and Peluso, E.M. (2013). Integration of Average Amino Acid Identity (AAI) and Percentage of Orthologous Genes in a Single Phylogenomic Metric, the Reciprocal Orthology Score Average (ROSA). Manuscript in Preparation.

Newman Lab ROSA Calculator

INSTRUCTIONS:
1) "Export file" from RAST Sequence Comparison Tool output for up to and including 11 files.
2) Browse for the files below. Holding the CTRL key down will allow you to select all the files at once by clicking on each of them; or click on the first file, then hold down the SHIFT key and click on the last file to select the range.
3) Hit the "Submit" button.
4) Copy and Paste Results Table to a Separate Spreadsheet or Word Processor Document.

Browse... 11 files selected.

Submit

ROSA Analysis

Orthology Analysis for Each Reference Genome

	Genome ID	AAIr	% BBH	OS
ref	317655.9			
1	1122612.3	58.348	61.747	21.022
2	279238.21	60.160	67.551	24.448
3	983920.5	57.863	62.671	20.983
4	1123240.3	56.027	55.375	17.382
5	452662.10	61.114	67.480	25.203
6	1090316.6	59.866	64.463	23.103

Poster #436

Integration of Average Amino Acid Identity (AAI) and Percentage of Orthologous Genes in a Single Phylogenomic Metric, the Reciprocal Orthology Score Average (ROSA).

Andrew Gale[1], Jordan Krebs[1], Eileen Peluso[2] ; Jeff Newman[1;]

Department of Biology[1] and Mathematical Sciences [2], Lycoming College, Williamsport, PA.

Abstract

With the decreasing cost of NextGen sequencing and the subsequent increase in the availability of microbial genome sequences, it has been suggested that the prokaryotic species definition should change from physical measurements of DNA-DNA hybridization (DDH) to computationally determined genome-wide metrics. The Reciprocal Orthology Score Average (ROSA) metric described here is calculated using Average Amino Acid Identity (AAI) and percent bi-directional best-hit (%BBH) genes at its core. We have developed a JavaScript-based tool (http://lycofs01.lycoming.edu/~newman/rosa/) that calculates AAI, %BBH, and ROSA using the output from the "Sequence-based comparison" tool on the Rapid Annotation with Subsystems Technology (RAST) service (rast.nmpdr.org). The ROSA metric has a range from <3 when comparing genomes from different domains to >98 when comparing closely related strains. Organisms at every level from subspecies to domain are clustered more accurately using ROSA thresholds than with any existing published metric because it takes into consideration both similarity of orthologs as well as percentage of the genome composed of orthologs.

Background/Review of Existing Phylogenomic Metrics

70% DNA-DNA Hybridization (DDH)-Official Species definition (Wayne *et al.* 1987)
- Physical integration of ortholog similarity and percentage of orthologs
- Examines entire genome
- Few labs have capability, error prone, applicable only at species level
- Requires physical reciprocal experiment for each organism, cannot database

70% estimated DDH using Genome-Genome Distance Calculator (GGDC
- Easily accessible web-based tool (Meier-Kolthoff *et al.* , 2013)
- Examines entire genome sequence, range similar to DDH, statistical support
- "Black box" calculation derived from unclear metric and complex statistics
- No widely accepted thresholds above species level

98.65% 16s rRNA similarity (Stackebrandt & Ebers, 2006; Kim *et al.*, 2014)
- Simple, very inexpensive, large databases.
- Easily accessible stand alone and web-based tools (Kim *et al.*, 2012)
- Examines only a single gene, many separate species have >99% similarity
- No widely accepted thresholds above species level

95-96% Average Nucleotide Identity (ANI) (Goris et al. , 2007)
- Accessible tools (Jspecies - Richter & Rossello´-Mora, 2009;
  Kostas Lab http://enve-omics.ce.gatech.edu/ani/;
  ChunLab http://www.ezbiocloud.net/ezgenome)
- Metric is moderately clear and intuitively understood at a basic level
- Value reflects average similarity of only highly similar (>60%) sequences
- Does not consider distant or non-orthologous sequences (min ~60%)
- No widely accepted thresholds above species level

95% Average Amino Acid Identity (AAI) (Konstantinidis & Tiedje, 2005)
- Higher amino acid sequence conservation detects more orthologs than ANI
- Metric is clear and intuitively understood, reflects unit of selection/conservation (proteins rather than randomly broken DNA seq)
- Applicable above species level, but no widely accepted thresholds (yet)
- Does not consider non-orthologous sequences
- No easily accessible tools

Goal: Create a genome-based metric to differentiate bacterial species.
- Incorporate orthology between distantly-related organisms (AAI)
- Incorporate core genome percentage due to HGT/reductive evolution (%BBH)
- Comparable to DDH using full range from 0 – 100
- Accessible via web using meaningful components, intuitive to understand

Rapid Annotation with Subsystem Technology (RAST)

http://rast.nmpdr.org/ (Overbeek *et al.*, 2014)
- Users upload multi-sequence fasta files
- RAST calculates GCmol%, genome size, identifies CDSs & RNAs, Integrates into subsystems.
- Sequence-Based Comparison tool identifies unidirectional and bidirectional best hits between a "Reference Genome" and up to ten "Comparison Genomes"
- Exportable color-coded table and graphic shows best matches.
- Screen "mouseover" or columns in exported table provide protein name and AA sequence identity of each.

Figure 1. Screenshot of Sequenced-Based Comparison tool results.

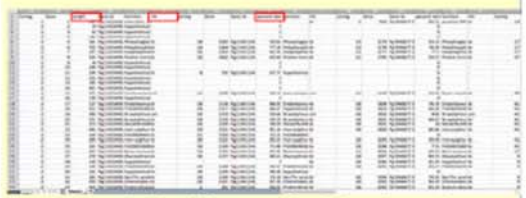


Figure 2. Exported table with Sequenced-Based Comparison tool results. Important columns are indicated with a red box.

Separate comparisons must be conducted using each organism as a reference and the others (up to 10) as comparison organisms.
- Can be used to construct Venn Diagrams

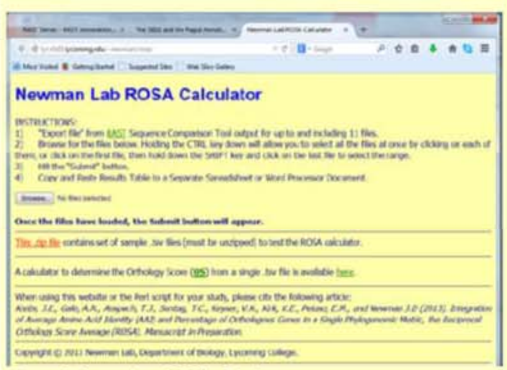


Figure 3. Screenshot of ROSA Calculator.

Logic of Phylogenomic Metric Design

- AAI → Due to variation in percent identity of CDS, (see fig. 2) length included as a factor.

$$AAI = \frac{\Sigma(\%identity * length)_{bbh}}{\Sigma length_{bbh}}$$

- Because Horizontal Gene Transfer (HGT) and reductive evolution would decrease DDH values between two organisms, we calculated this as %BBH.

$$\%BBH = \frac{\Sigma length_{bbh\,(comparison)}}{\Sigma length_{all\,(reference)}}$$

- Because 95% AAI corresponds to 70% DDH (the species barrier), we concluded that decreasing AAI would decrease DDH faster than a decrease in %BBH

Orthology Score (OS) = AAI^2 * %BBH

- Differences in reference genome size create different %BBH values in reciprocal comparisons

Reciprocal Orthology Score Average (ROSA)

$$ROSA = (OS_{AB} + OS_{BA})/2$$

Example ROSA Clusters

Table 1. Same species >65; Same genus, different species 35-65

Table 2. Same family, different genus 15-35; different family <15

Table 3. Same domain, different phylum 3-6; different domain <3

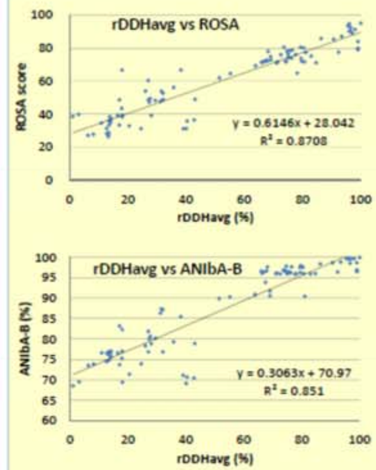


Figure 4. Relationship between reciprocal DDH and ROSA or ANI

Level	Same	Different	expected range	min	max	mean	n=	below range	above range
8	species	strain	>65	49	99.59	85.58	312	3	N/A
7	genus	species	35-65	6.85	95.8	34.63	521	294	38
6	family	genus	15-35	5.8	54.99	18.86	524	125	19
5	order	family	10-15	4.75	23.19	11.58	314	79	22
4	class	order	8-10	4.15	14.99	8.22	286	121	58
3	phylum	class	6-8	4.3	11.4	6.62	123	42	14
2	domain	phylum	3-6	1.85	7.66	4.45	211	10	6
1		domain	<3	1.36	4.44	2.42	47	0	9

Table 4. ROSA Values at Different Phylogenetic levels.

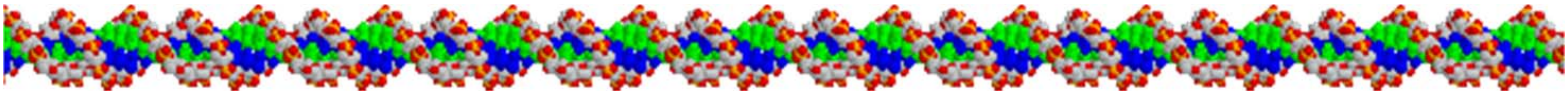
Conclusions:

- ROSA metric correlates well with DDH because it incorporates level of similarity and percentage of orthologs in a genome.
- If Microbial Systematics is to become based on phylogenomics, much taxonomic revision will be necessary.
- Comparison of shared and unique orthologs can allow prediction of differentiating traits between groups

References

- Auch, AF, Klenk HP, and Göker M (2010). Standard Operating Procedure for Calculating Genome-to-genome Distances Based on High-scoring Segment Pairs. [illegible]
- Goris, J., K.T. Konstantinidis, J.A. Klappenbach, T. Coenye, P. Vandamme, and J.M. Tiedje. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. [illegible]
- Kim, Mincheol, Hyun-Seok Oh, Sang-Cheol Park, and Jongsik Chun. Towards a Taxonomic Coherence between Average Nucleotide Identity and 16S rRNA Gene Sequence Similarity for Species Demarcation of Prokaryotes. IJSEM 64 (2014) [illegible]
- Konstantinidis, K.T., and J.M. Tiedje. 2005. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. USA. [illegible]
- Konstantinidis, K.T. and J.M. Tiedje. 2005. Towards a genome-based taxonomy for prokaryotes. J. Bacteriol. [illegible]
- Meier-Kolthoff, JP, Auch AF, Klenk H-P, and Göker M. "Genome Sequence-based Species Delimitation with Confidence Intervals and Improved Distance Functions." BMC Bioinformatics 14.1 (2013): 60.
- Overbeek et al., (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucl Acids Res. [illegible]
- Richter, M. and R. Rosselló-Móra. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. USA. [illegible]
- Stackebrandt, E., and J. Ebers. 2006. Taxonomic parameters revisited: tarnished gold standards. Microbiol. Today [illegible]
- Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Truper. "Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics." IJSB 37.4 (1987): 463-64.

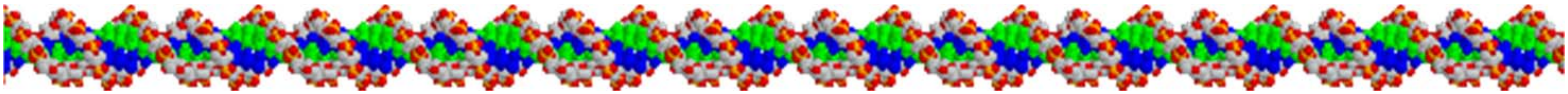
ROSA Thresholds



Level	Same	Different	expected range	min	max	mean	n=	below range	above range
8	species	strain	>65	49	99.59	85.98	312	3	N/A
7	genus	species	35-65	6.85	95.8	36.63	521	294	38
6	family	genus	15-35	5.8	54.99	18.86	524	125	19
5	order	family	10-15	4.75	23.19	11.58	314	79	22
4	class	order	8-10	4.15	14.99	8.22	286	121	58
3	phylum	class	6-8	4.3	11.4	6.62	123	42	14
2	domain	phylum	3-6	1.85	7.66	4.45	211	10	6
1		domain	<3	1.36	4.44	2.42	47	0	9

Table 4. ROSA Values at Different Phylogenetic levels.

Chryseo vs Kaistella genomes



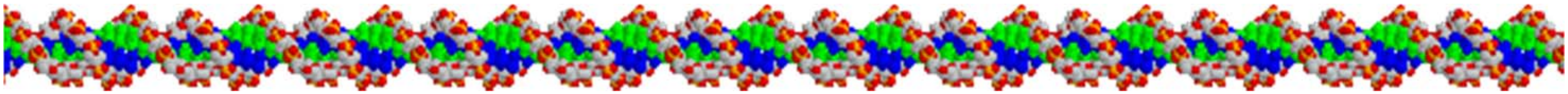
ROSA (sorted)		512012.1	531844.8	510955.3	421525.1	266749.3	1121288	266748.5	232216.7	1121870	525257.7	121696
Flavobacteriaceae bacterium JJC (512012.10)	**512012.1**											
Flavobacteriaceae bacterium 3519-10 (531844.8)	**531844.8**	50.305										
Chryseobacterium solincola DSM 22468 (510955.3)	**510955.3**	48.614	46.447									
Chryseobacterium haifense DSM 19056 (421525.12)	**421525.12**	47.947	41.886	42.263								
Chryseobacterium jeonii DSM 17048 (266749.3)	**266749.3**	47.332	42.154	46.796	38.772							
Chryseobacterium palustre DSM 21579 (1121288.3)	**1121288.3**	47.141	43.741	47.319	39.36	47.852						
Chryseobacterium antarcticum LMG 24720 (266748.5)	**266748.5**	44.819	41.226	49.764	37.835	50.213	44.893					
Chryseobacterium koreense CCUG 49689 (232216.7)	**232216.7**	44.653	39.149	41.074	38.487	39.248	41.354	37.91				
Epilithonimonas tenax DSM 16811 (1121870.3)	**1121870.3**	32.207	30.889	32.079	29.75	30.303	29.797	29.403	28.804			
Chryseobacterium gleum F93, ATCC 35910 (525257.7)	**525257.7**	30.406	28.241	30.002	28.223	27.728	27.799	27.315	28.573	29.996		
Elizabethkingia meningoseptica ATCC 13253 (1216967.8)	**1216967.8**	26.552	25.025	26.966	24.894	24.972	24.24	24.567	23.845	25.521	27.671	

- Top 8 organisms…
  - Show high genome similarity with each other (ROSA range of 38-50)
  - Show lower genome similarity with three genera (ROSA range of 24-32)
- *Kaistella koreensis* was first published in 2004 as a new genus, has priority

Comparison of Genome Size "True" Chryseobact. vs Kaistella

Species	Contigs	Genome Size		
Chryseobacterium gleum F93, ATCC 35910	69	5562004		**Chryseobacterium**
Chryseobacterium sp. LO	**85**	**5489991**		**Kaistella**
Chryseobacterium vrystaatense LMG 22846	**18**	**5487842**		
Chryseobacterium sp. JM1	**31**	**5256156**		
Chryseobacterium angstadtii KM	**16**	**5212770**		
Chryseobacterium sp. UNC8MFCol	35	4985241		
Chryseobacterium soli DSM 19298	**26**	**4754478**		
Chryseobacterium indologenes NBRC 14944	33	4753942		
Chryseobacterium luteum DSM 18605	**49**	**4718546**		
Chryseobacterium sp. CF314	132	4488346		
Chryseobacterium oranimense G311	15	4457049		
Chryseobacterium gregarium DSM 19109	44	4438326		
Chryseobacterium sp. BLS98	**10**	**4414965**		
Chryseobacterium formosense LMG 24722	**10**	**4364663**		
Chryseobacterium hispalense DSM 25574	**27**	**4363762**		
Chryseobacterium piperi CTM	**89**	**4340594**		
Chryseobacterium daeguense DSM 19388	27	4214400		
Chryseobacterium caeni DSM 17710	21	4115882		
Chryseobacterium sp. FH2	**25**	**3988967**		
Chryseobacterium jeonii DSM 17048	**42**	**3268780**		
Chryseobacterium koreense CCUG 49689	**75**	**3179892**		
Chryseobacterium antarcticum LMG 24720	**4**	**3123663**		
Chryseobacterium palustre DSM 21579	20	2981097		
Flavobacteriaceae bacterium JJC	**90**	**2853712**		
Chryseobacterium haifense DSM 19056	**505**	**2853560**		
Flavobacteriaceae bacterium 3519-10	1	2768102		
Chryseobacterium solincola DSM 22468	**16**	**2345734**		

Flexirubin Pigments



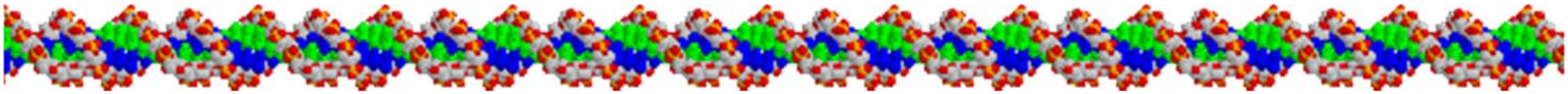
Conti	Gene	Lengt	C. gleum I	function	Hit	C. sp. KM	Hit	C. daegue	Hit	C. formos	Hit	C. gregari	Hit	C. hispale	Hit	C. indolog	Hit	C. luteum	Hit	C. piperi (	Hit	C. soli DSI	Hit	C. vrystaat
12	1517	305	fig|52525	Dialkylrecorsinol condensing enzyme	bi	fig|55815	bi	fig|11212	bi	fig|23681	bi	fig|11212	bi	fig|49120	bi	fig|12181	bi	fig|42153	bi	fig|55815	bi	fig|44596	bi	fig|30748

Conti	Gene	Lengt	C. Gleum	function	Hit	C. antarcti	Hit	C. haifens	Hit	C. jeonii	Hit	K. koreen	Hit	C. palustr	Hit	C. Solinco	Hit	F. 351910	Hit	F. JCC	Hit	Elizabethl	Hit	Epilithoni
12	1517	305	fig|52525	Dialkylrecorsinol condensing enzyme																			bi	fig|11218

Chyseobacterium | Kaistella | Elizabethkingia | Epilithonimonas | bi = Bi-directional best Hit

- Dialkylresorcinol condensing enzyme = Flexirubin Pigments
- All of the Chryseobacterium (yellow) have flexirubin pigments.
- None of the proposed Kaistella (green) have flexirubin pigments.

Analyzing the RAST Comparisons



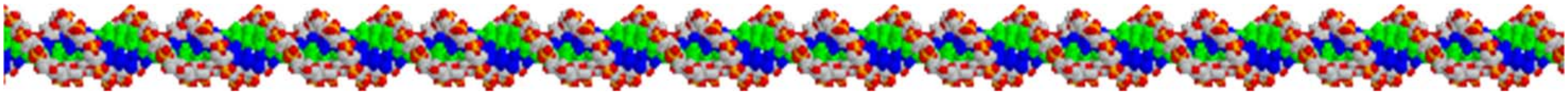
- Emphasis was given to genes which were present in all of the proposed Kaistella and none of the three other type-strains.

E1435 | Nitrousoxide reductase (EC 1.7.99.6)

1	Contig	Gene	Length	K. koreensis	function	Hit	C. antarcti	Hit	C. haifens	Hit	C. jeonii	Hit	C. palustr	Hit	C. Solinco	Hit	Flavobact	Hit	F. JCC	Hit	Elizabethl	Hit	C. Gleum	Hit	Epilithonimonas tenax
1430	41	2614	256	fig|232216.7	Nitrous oxide reductase maturation transmembrane protein NosY	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|53184	bi	fig|512012.10						
1431	41	2615	238	fig|232216.7	Nitrous oxide reductase maturation protein NosF (ATPase)	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|53184	bi	fig|512012.10		fig|1216967.8		fig|525257.7.		fig|1121870.3.peg.437
1432	41	2616	409	fig|232216.7	Nitrous oxide reductase maturation protein NosD	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|53184	bi	fig|512012.10						
1433	41	2617	137	fig|232216.7	Nitrous oxide reductase maturation protein, outermembrane lipoprotein NosL	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|53184	bi	fig|512012.10						
1434	41	2618	175	fig|232216.7	FIG01019293: hypothetical protein	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|53184	bi	fig|512012.10						
1435	41	2619	655	fig|232216.7	Nitrousoxide reductase (EC 1.7.99.6)	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|53184	bi	fig|512012.10						
1436	41	2620	166	fig|232216.7	FIG00618450: hypothetical protein	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|53184	bi	fig|512012.10						
1437	4	139	282	fig|232216.7	Lmo0572 protein	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|531844.8.		fig|51201	bi	fig|12169	bi	fig|52525	bi	fig|1121870.3.peg.528
1438	13	495	141	fig|232216.7	hypothetical protein	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|531844.8.		fig|51201	bi	fig|12169	bi	fig|52525	bi	fig|1121870.3.peg.2409
1439	28	1452	706	fig|232216.7	Translation elongation factor G	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|531844.8.		fig|51201	bi	fig|12169	bi	fig|52525	bi	fig|1121870.3.peg.1562
1440	31	1667	562	fig|232216.7	TPR repeatcontaining protein	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|531844.8.			bi	fig|12169	bi	fig|52525	bi	fig|1121870.3.peg.468
1441	22	920	93	fig|232216.7	hypothetical protein	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|531844.8.		fig|51201	bi	fig|12169	bi	fig|525257.7.		fig|1121870.3.peg.1140
1442	28	1428	72	fig|232216.7	Translation initiation factor 1	bi	fig|26674	bi	fig|42152	bi	fig|26674	bi	fig|11212	bi	fig|51095	bi	fig|531844.8.			bi	fig|12169	bi	fig|525257.7.		

- **The group of genes related to Nitrous Oxide Reductase appear <u>only</u> in the Kaistella's.**

Denitrification and the Nitrous Oxide Reductase cluster

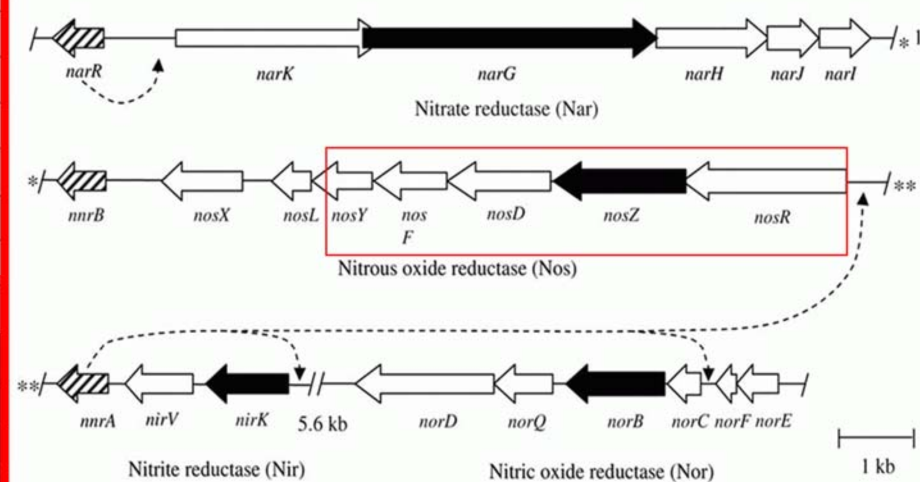


Kaistella

Chryseobacterium

Organism ▲▼	Domain all	Variant (?) ▲▼ = all	active yes	NirS	NirK	NirV	qNor	cNor-B	cNor-C	NorE	NorD	Nor	NosZ	NosD	NosF	NosY	NosL	osX	NosR	NnrS	NnrR	DNR	NnrU	NnrT
Chryseobacterium palustre DSM 21579 (1121288.3)	Bacteria	10	yes		1847								1966	1965	1962	1963								
Chryseobacterium koreense CCUG 49689 (232216.7)	Bacteria	10	yes		1340				1042				2619	2616	2615	2614	2617							
Chryseobacterium antarcticum LMG 24720 (266748.5)	Bacteria	10	yes		1574								1684	1681		1679	1682							
Chryseobacterium haifense DSM 19056 (421525.12)	Bacteria	3	yes		808		801						2892	668	3547	3535								
Chryseobacterium jeonii DSM 17048 (266749.3)	Bacteria	10	yes		926								815	818	819	820	812							
Chryseobacterium solincola DSM 22468 (510955.3)	Bacteria	10	yes		398								293	292		299	295							
Chryseobacterium gregarium DSM 19109 (1121287.3)	Bacteria	*-1	no																					
Chryseobacterium sp. KM (558151.5)	Bacteria	*-1	no																					
Chryseobacterium angstadtii KM (558151.4)	Bacteria	*-1	no																					
Chryseobacterium oranimense G311 (1300143.3)	Bacteria	*-1	no																					
Chryseobacterium vrystaatense LMG 22846 (307480.4)	Bacteria	*-1	no																					
Chryseobacterium soli DSM 19298 (445961.4)	Bacteria	*-1	no																					
Chryseobacterium indologenes NBRC 14944 (1218103.4)	Bacteria	-1	no				615, 616																	
Chryseobacterium hispalense DSM 25574 (491205.4)	Bacteria	8	yes		1105		1110	1132																
Chryseobacterium daequense DSM 19388 (1121286.6)	Bacteria	*-1	no																					
Chryseobacterium luteum DSM 18605 (421531.4)	Bacteria	*-1	no																					
Chryseobacterium formosense LMG 24722 (236814.4)	Bacteria	*-1	no																					
Chryseobacterium piperi CTM (558152.4)	Bacteria	*-1	no																					
Chryseobacterium gleum F93, ATCC 35910 (525257.7)	Bacteria	8	yes		3493		3499																	

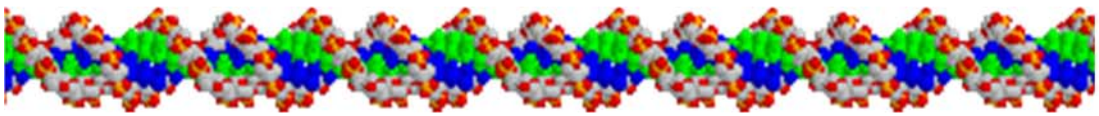
Denitrification Pathway



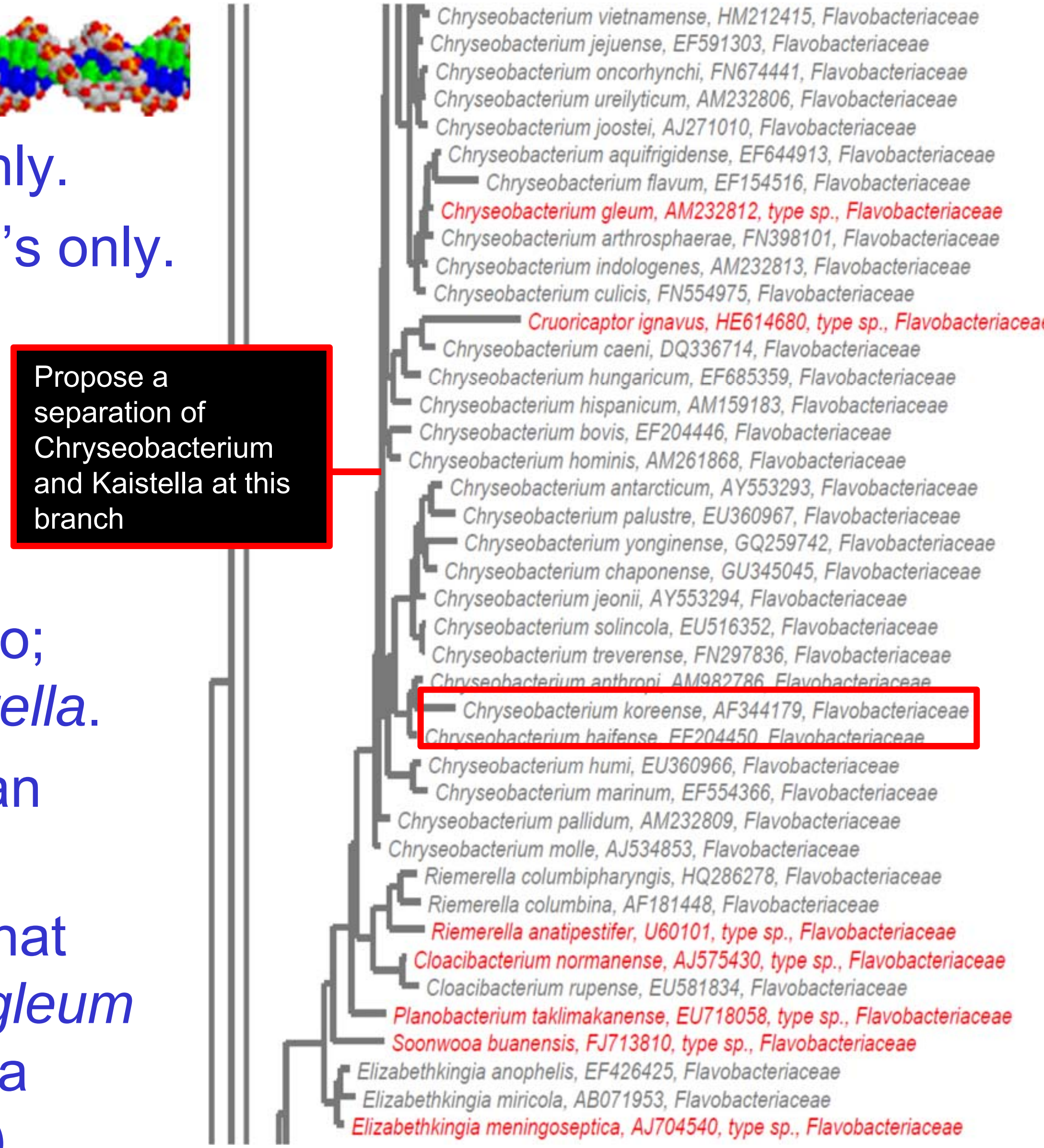
Nitrate (NO_3^-) —Nar→ Nitrite (NO_2^-) —Nir→ Nitric oxide (NO) —Nor→ Nitrous oxide (N_2O) —Nos→ Nitrogen (N_2)

Haine V et al. J. Bacteriol. 2006;188:1615-1619

Differences between Chryseo's & Kaistella's

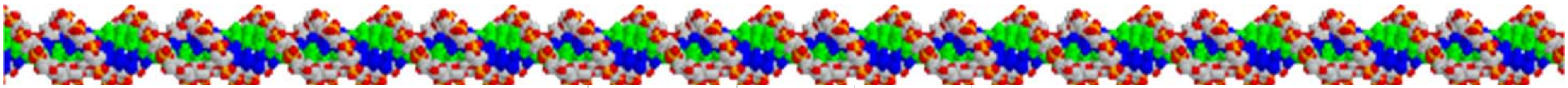


- Flexirubin pigments in Chryseo's only.
- Nitrous oxide reductase in *Kaistella*'s only.
- *Kaistella*'s are more sensitive to several antibiotics



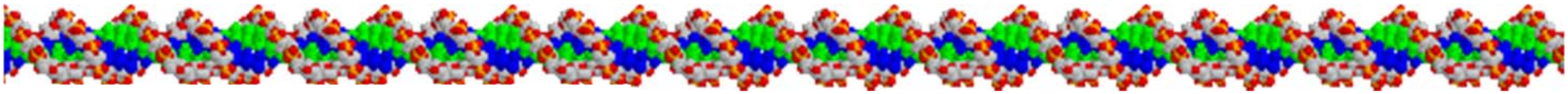
- Low (<10%) 15:0 anteiso in Chryseo; higher (>10%) 15:0 anteiso in *Kaistella*.
- *Kaistella* have smaller genomes than *Chryseobacterium*
- The ROSA and AAIr scores show that *Kaistella*'s are as different from *C. gleum* as they are from neighboring genera (*Epilithonimonas* & *Elizabethkingia*).

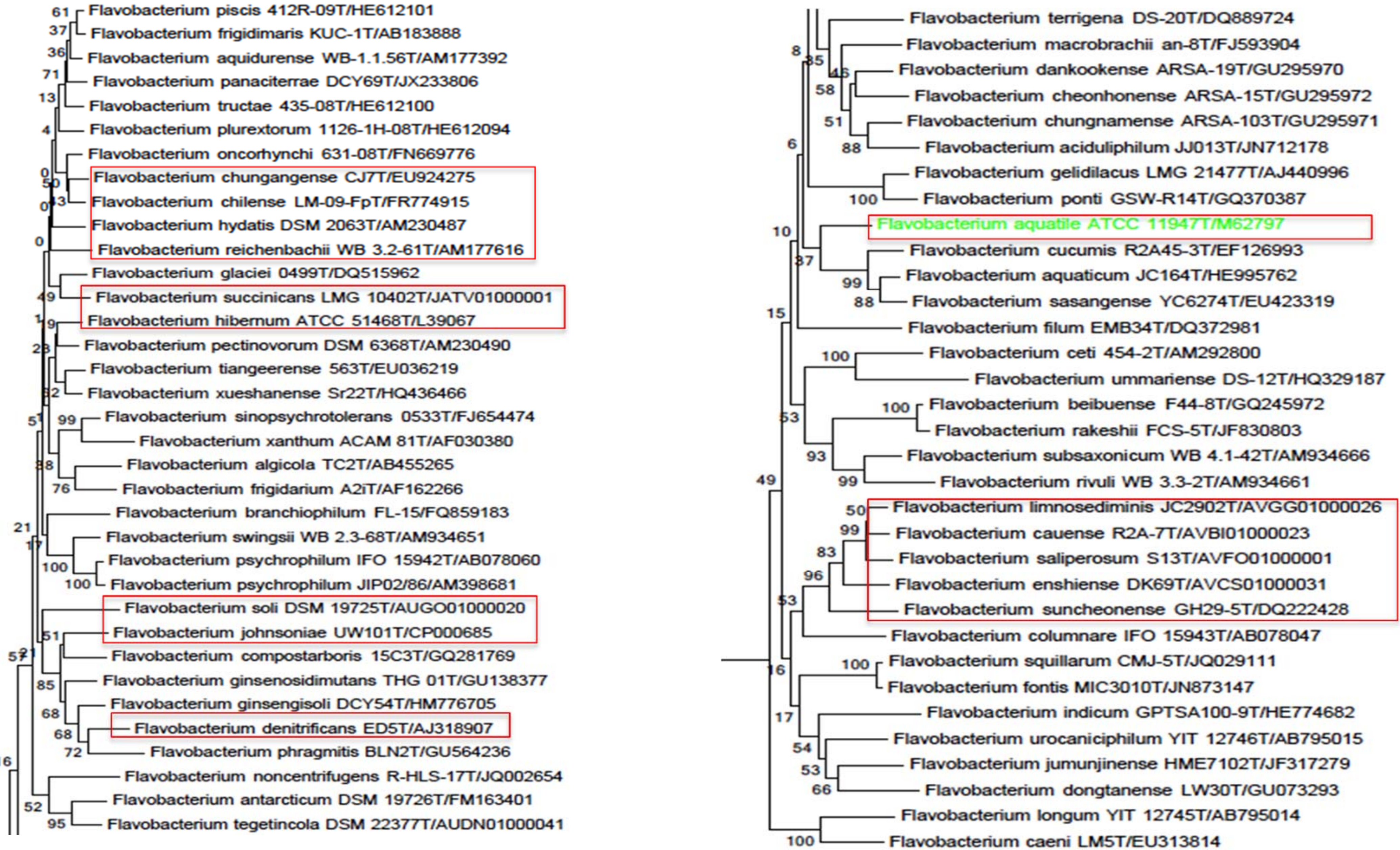
Create the Family Chryseobacteriaceae



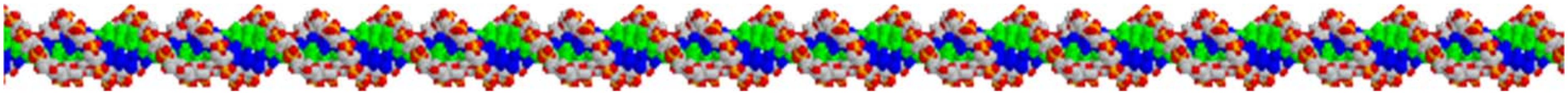
	ROSA	245.3	525257.7	
Flavobacterium aquatile LMG 4008	245.3		12.454	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Mesoflavibacter zeaxanthinifaciens DSM 18436	1122225.3	21.773	11.327	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Gaetbulibacter saemankumensis DSM 17032	1121909.3	20.272	10.949	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Croceibacter atlanticus HTCC2559	216432.3	19.939	10.797	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Psychroserpens burtonensis DSM 12212	1123037.3	19.343	9.644	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Gramella echinicola DSM 19838	1121931.3	18.244	10.443	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Gillisia limnaea R-8282, DSM 15749	865937.4	18.138	10.175	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Cellulophaga lytica DSM 7489	867900.5	18.073	10.452	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Joostella marina En5, DSM 19592	453852.3	17.552	10.385	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Myroides odoratus DSM 2801	929704.5	17.51	11.963	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Muricauda ruestringensis B1, DSM 13258	886377.5	17.356	10.057	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Eudoraea adriatica DSM 19308	1121875.4	16.699	9.473	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Robiginitalea biformata HTCC2501	313596.3	15.872	9.341	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Zunongwangia profunda SM-A87	655815.5	15.837	10.443	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Psychroflexus torquis ATCC 700755	313595.1	15.63	9.008	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Capnocytophaga ochracea DSM 7271	521097.9	15.089	10.518	Flavobacteriia; Flavobacteriales; Flavobacteriaceae
Weeksella virosa 9751, DSM 16922	865938.6	14.418	15.492	Flavobacteriia; Flavobacteriales; Chryseobacteriaceae
Empedobacter brevis ATCC 43319	1218108.4	13.437	16.693	Flavobacteriia; Flavobacteriales; Chryseobacteriaceae
Bergeyella zoohelcum ATCC 43767	883096.4	13.198	23.351	Flavobacteriia; Flavobacteriales; Chryseobacteriaceae
Chryseobacterium antarcticum LMG 24720	266748.3	14.234	27.012	Flavobacteriia; Flavobacteriales; Chryseobacteriaceae
Chryseobacterium koreense CCUG 49689	232216.5	13.354	27.375	Flavobacteriia; Flavobacteriales; Chryseobacteriaceae
Elizabethkingia meningoseptica ATCC 13253	1216967.8	12.662	27.671	Flavobacteriia; Flavobacteriales; Chryseobacteriaceae
Epilithonimonas tenax DSM 16811	1121870.3	13.764	29.996	Flavobacteriia; Flavobacteriales; Chryseobacteriaceae
Chryseobacterium gleum F93, ATCC 35910	525257.7	12.454		Flavobacteriia; Flavobacteriales; Chryseobacteriaceae
Ornithobacterium rhinotracheale DSM 15997	867902.4	11.92	12.574	Flavobacteriia; Flavobacteriales; Ornithobacteriaceae
Owenweeksia hongkongensis DSM 17368	926562.5	12.129	9.101	Flavobacteriia; Flavobacteriales; Cryomorphaceae
Fluviicola taffensis DSM 16823	755732.4	11.088	8.465	Flavobacteriia; Flavobacteriales; Cryomorphaceae
Sphingobacterium spiritivorum ATCC 33861	525373.5	10.297	11.417	Sphingobacteriia; Sphingobacteriales; Sphingobacteriaceae
Cytophaga hutchinsonii ATCC 33406	269798.18	9.593	7.952	Cytophagia; Cytophagales; Cytophagaceae

Flavobacterium 16S Tree





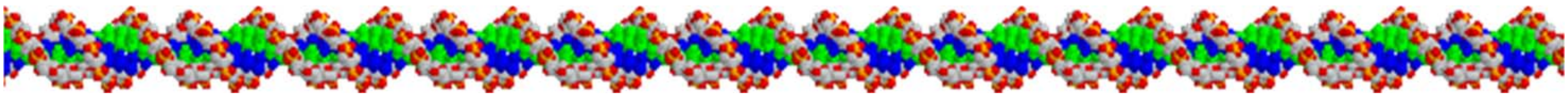
1 Cluster of Species



	ROSA (sorted)	1341155.4	1341154	1341182	1121899	1107311	1453498	402612.1	1041826	1121901	1121896	376686.2
Flavobacterium saliperosum S13 (1341155.4)	**1341155.4**											
Flavobacterium cauense R2A-7 (1341154.4)	**1341154.4**	66.422										
Flavobacterium limnosediminis JC2902 (1341181.6)	**1341181.6**	56.717	55.312									
Flavobacterium suncheonense DSM 17707 (1121899.4)	**1121899.4**	53.612	52.735	47.758								
Flavobacterium enshiense DK69 (1107311.4)	**1107311.4**	53.072	52.119	57.609	47.206							
Flavobacterium aquatile LMG 4008 (1453498.10)	**1453498.1**	34.026	33.597	31.612	32.482	30.579						
Flavobacterium psychrophilum JIP02/86 (402612.11)	**402612.11**	33.604	33.083	30.935	32.11	30.661	32.033					
Flavobacterium columnare ATCC 49512 (1041826.25)	**1041826.25**	30.585	30.126	27.938	30.722	28.575	26.755	32.335				
Flavobacterium tegetincola DSM 22377 (1121901.4)	**1121901.4**	29.221	28.559	27.029	28.376	26.234	29.126	28.153	23.76			
Flavobacterium rivuli DSM 21788 (1121895.6)	**1121895.6**	26.355	25.893	24.492	26.241	24.122	24.153	23.577	20.624	23.541		
Flavobacterium johnsoniae UW101 (376686.16)	**376686.16**	25.057	24.884	23.249	24.992	23.167	25.487	25.397	21.586	23.13	21.939	

		Average Amino Acid Identity (AAIr) Genus Threshold ~75					
		1	**2**	**3**	**4**	**5**	**6**
Flavobacterium saliperosum S13	**1**		88.8	84.9	80.4	82.6	69.1
Flavobacterium cauense R2A-7	**2**	66.4		84.6	80.9	82.5	69.1
Flavobacterium limnosediminis JC2902	**3**	56.7	55.3		79.3	85.4	68.5
Flavobacterium suncheonense DSM17707	**4**	53.6	52.7	47.7		79.1	68.2
Flavobacterium enshiense DK69	**5**	53.0	52.1	57.6	47.2		68.3
Flavobacterium aquatile LMG 4008	**6**	34.0	33.5	31.6	32.4	30.5	
		Reciprocal Orthology Score Average (ROSA) Genus Threshold ~35					

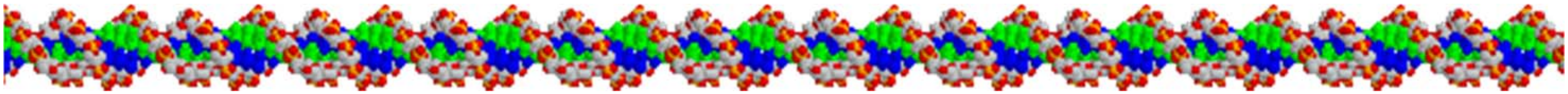
3 Clusters of Species



ROSA (sorted)		946677.6	37752.6	376686.2	362418.8	1453506	991.1	386300.3	1450526	1453498	1041826	1121896
Flavobacterium chilense (946677.6)	**946677.6**											
Flavobacterium hibernum DSM 12611 (37752.6)	**37752.6**	54.496										
Flavobacterium johnsoniae UW101 (376686.16)	**376686.16**	51.639	49.486									
Flavobacterium reichenbachii (362418.8)	**362418.8**	48.307	48.06	49.95								
Flavobacterium chungangense LMG 26729 (1453505.5)	**1453505.5**	43.429	45.362	44.112	46.498							
Flavobacterium hydatis DSM 2063 (991.10)	**991.1**	41.759	42.322	38.737	36.916	35.61						
Flavobacterium glaciei CGMCC 1.5380 (386300.3)	**386300.3**	33.463	34.678	32.37	33.248	33.194	34.512					
Flavobacterium succinicans LMG 10402 (1450525.7)	**1450525.7**	32.196	33.116	30.64	30.47	31.012	32.679	39.976				
Flavobacterium aquatile LMG 4008 (1453498.10)	**1453498.1**	25.927	27.093	25.487	26.16	25.68	26.214	33.887	30.568			
Flavobacterium columnare ATCC 49512 (1041826.25)	**1041826.25**	22.108	22.403	21.586	22.108	22.283	22.809	27.303	25.882	26.755		
Flavobacterium rivuli DSM 21788 (1121895.6)	**1121895.6**	21.53	22.688	21.939	22.868	22.587	21.058	25.448	22.426	24.153	20.624	

ROSA (sorted)		1121888.4	376686.2	1121887	1121898	1453498	1406840	1121889	1121898	1121896	1094466	1041826
Flavobacterium denitrificans DSM 15936 (1121888.4)	**1121888.4**											
Flavobacterium johnsoniae UW101 (376686.16)	**376686.16**	51.95										
Flavobacterium daejeonense DSM 17708 (1121887.4)	**1121887.4**	31.541	29.444									
Flavobacterium soli DSM 19725 (1121897.6)	**1121897.6**	28.846	26.516	25.819								
Flavobacterium aquatile LMG 4008 (1453498.10)	**1453498.1**	27.58	25.487	25.764	32.218							
Flavobacterium beibuense F44-8 (1406840.4)	**1406840.4**	24.266	23.133	22.001	28.826	25.892						
Flavobacterium filum DSM 17961 (1121889.4)	**1121889.4**	24.248	22.279	23.904	29.377	32.804	25.66					
Flavobacterium subsaxonicum DSM 21790 (1121898.4)	**1121898.4**	24.212	22.454	20.708	26.68	24.396	33.386	23.403				
Flavobacterium rivuli DSM 21788 (1121895.6)	**1121895.6**	23.857	21.939	20.73	26.18	24.153	32.049	22.888	44.082			
Flavobacterium indicum GPTSA100-9 (1094466.12)	**1094466.12**	23.578	22.043	22.462	26.105	29.944	25.122	28.494	23.202	22.897		
Flavobacterium columnare ATCC 49512 (1041826.25)	**1041826.25**	23.415	21.586	22.032	24.079	26.755	23.079	25.648	20.918	20.624	27.763	

More Clusters and Singletons



- ROSA is the smoke to point out genomic differences
- Examination of shared and unique genes and phenotypes can allow creation of evolutionarily and phenotypically coherent groups.
- It seems that much revision is necessary, due to power of genome sequences to reveal relationships

	ROSA (sorted)	1086011.9	1121890	1453498	1111731	376686.2	402612.1	1121892	1121896	1034807	1041826	1121896
Flavobacterium frigoris PS1 (1086011.9)	**1086011.9**											
Flavobacterium frigidarium DSM 17623 (1121890.4)	**1121890.4**	44.281										
Flavobacterium aquatile LMG 4008 (1453498.10)	**1453498.1**	30.109	28.818									
Flavobacterium antarcticum DSM 19726 (1111730.6)	**1111730.6**	30.08	29.654	31.201								
Flavobacterium johnsoniae UW101 (376686.16)	**376686.16**	29.768	29.928	25.487	24.928							
Flavobacterium psychrophilum JIP02/86 (402612.11)	**402612.11**	29.335	29.313	32.033	30.279	25.397						
Flavobacterium gelidilacus DSM 15343 (1121891.5)	**1121891.5**	27.347	26.178	31.118	29.332	21.431	28.085					
Flavobacterium sasangense DSM 21067 (1121896.4)	**1121896.4**	27.022	25.776	33.034	29.533	22.592	29.81	34.823				
Flavobacterium branchiophilum FL-15 (1034807.12)	**1034807.12**	26.536	25.533	27.35	25.134	22.407	27.922	23.747	25.77			
Flavobacterium columnare ATCC 49512 (1041826.25)	**1041826.25**	23.662	24.183	26.755	25.425	21.586	32.335	24.95	27.359	24.49		
Flavobacterium rivuli DSM 21788 (1121895.6)	**1121895.6**	23.328	23.412	24.153	24.905	21.939	23.577	22.42	23.302	20.414	20.624	

Thank You!