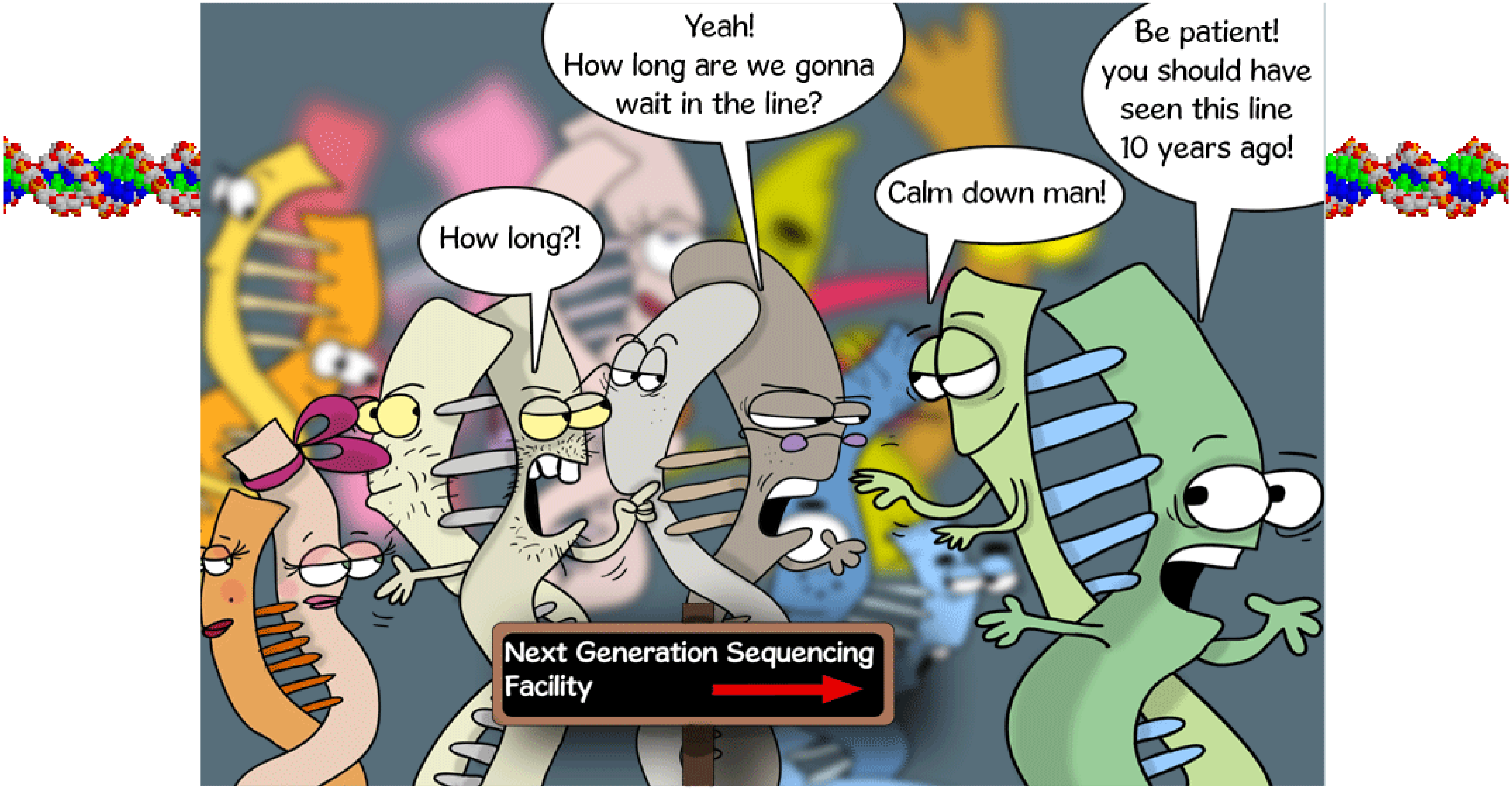
Integration of Average Amino Acid Identity (AAI) And Percentage of Orthologous Genes (BBH) in a Single Phylogenomic Metric, The Reciprocal Orthology Score Average (ROSA).

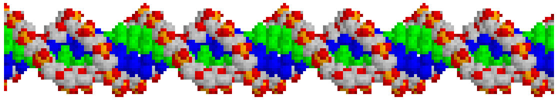
Jeff Newman
Lycoming College, Williamsport, PA USA

BISMIS September 12, 2016

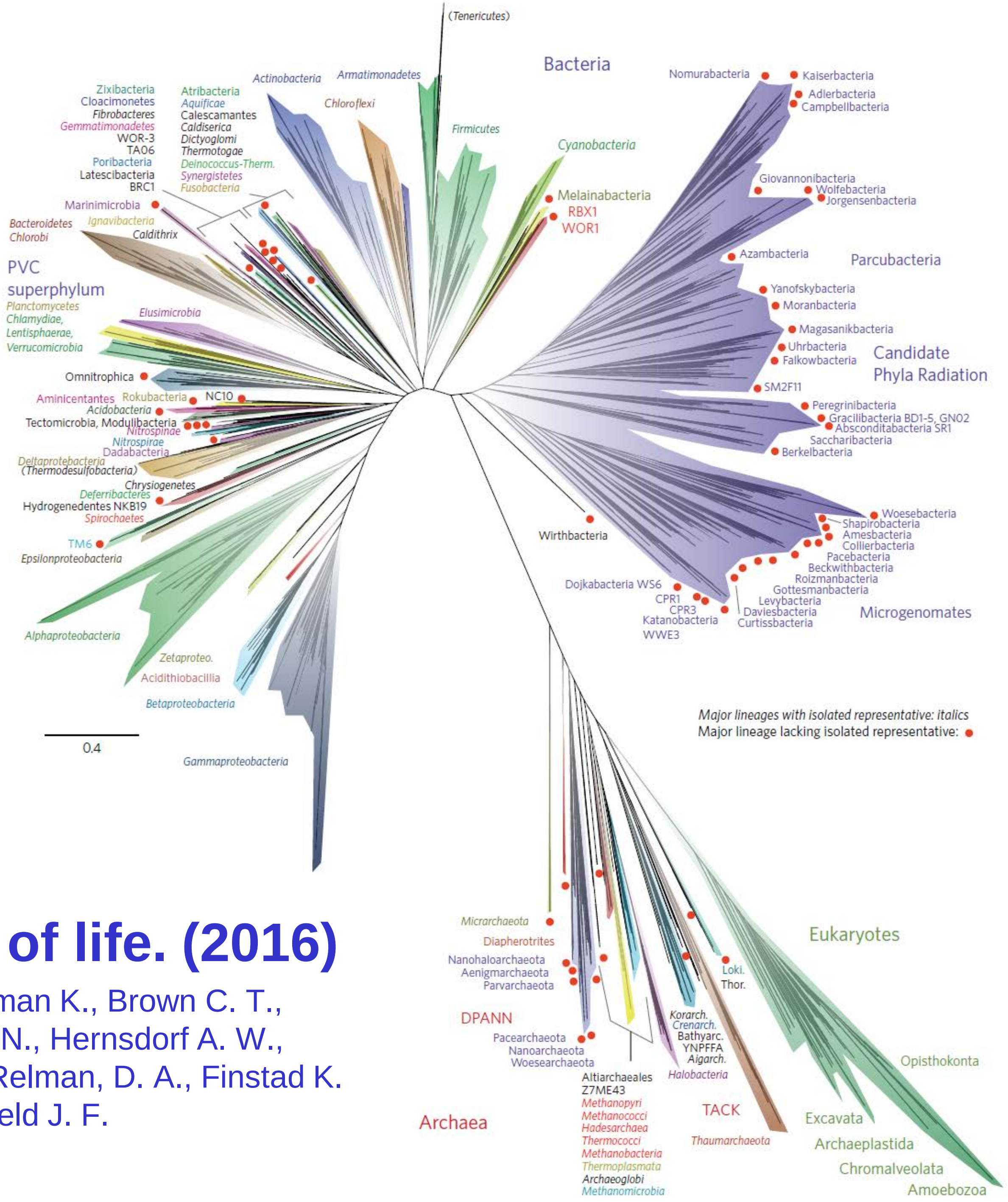


Human Genome 10th Anniversary

http://biocomicals.blogspot.com



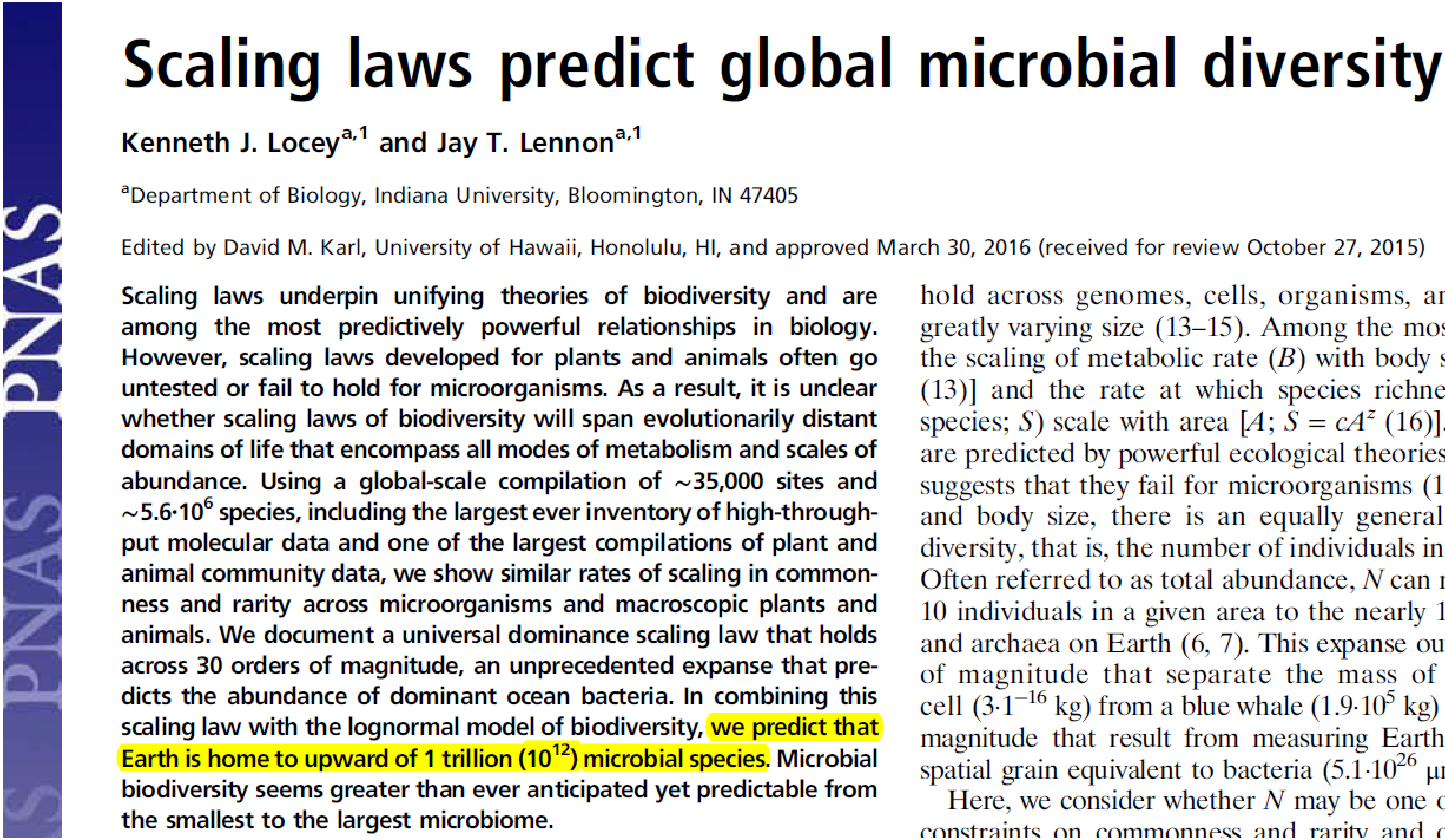
The Tree of Life is Broad!



A new view of the tree of life. (2016)

From Hug L. A., Baker B. J., Anantharaman K., Brown C. T., Probst A. J., Castelle C. J., Butterfield C. N., Hernsdorf A. W., Amano Y., Ise K., Suzuki Y., Dudek N., Relman, D. A., Finstad K. M., Amundson R., Thomas B. C., & Banfield J. F. *Nature Microbiol.* 1, 16048.

The Tree of Life is DEEP!



Scaling laws predict global microbial diversity

Kenneth J. Locey[a,1] and Jay T. Lennon[a,1]

[a]Department of Biology, Indiana University, Bloomington, IN 47405

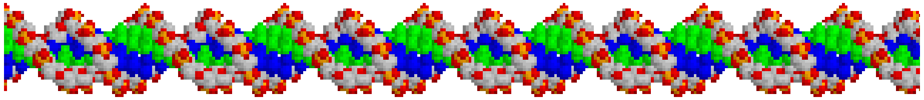
Edited by David M. Karl, University of Hawaii, Honolulu, HI, and approved March 30, 2016 (received for review October 27, 2015)

Scaling laws underpin unifying theories of biodiversity and are among the most predictively powerful relationships in biology. However, scaling laws developed for plants and animals often go untested or fail to hold for microorganisms. As a result, it is unclear whether scaling laws of biodiversity will span evolutionarily distant domains of life that encompass all modes of metabolism and scales of abundance. Using a global-scale compilation of ~35,000 sites and ~$5.6{\cdot}10^6$ species, including the largest ever inventory of high-throughput molecular data and one of the largest compilations of plant and animal community data, we show similar rates of scaling in commonness and rarity across microorganisms and macroscopic plants and animals. We document a universal dominance scaling law that holds across 30 orders of magnitude, an unprecedented expanse that predicts the abundance of dominant ocean bacteria. In combining this scaling law with the lognormal model of biodiversity, we predict that Earth is home to upward of 1 trillion (10^{12}) microbial species. Microbial biodiversity seems greater than ever anticipated yet predictable from the smallest to the largest microbiome.

hold across genomes, cells, organisms, ar
greatly varying size (13–15). Among the mos
the scaling of metabolic rate (*B*) with body s
(13)] and the rate at which species richne
species; *S*) scale with area [*A*; $S = cA^z$ (16)].
are predicted by powerful ecological theories
suggests that they fail for microorganisms (1
and body size, there is an equally general
diversity, that is, the number of individuals in
Often referred to as total abundance, *N* can r
10 individuals in a given area to the nearly 1
and archaea on Earth (6, 7). This expanse ou
of magnitude that separate the mass of
cell ($3{\cdot}1^{-16}$ kg) from a blue whale ($1.9{\cdot}10^5$ kg)
magnitude that result from measuring Earth
spatial grain equivalent to bacteria ($5.1{\cdot}10^{26}$ μ
Here, we consider whether *N* may be one c
constraints on commonness and rarity and c

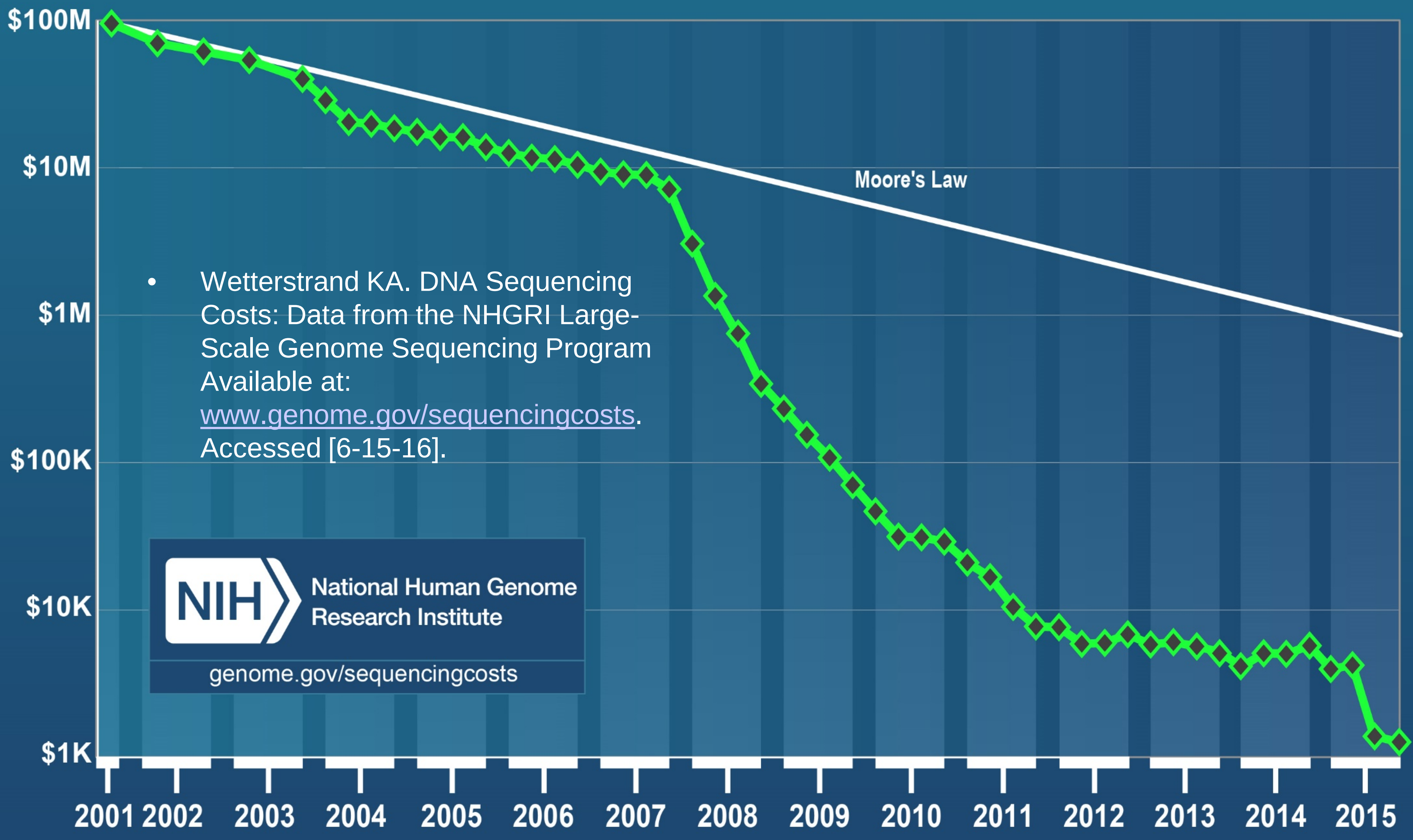
From Locey, K.J. and Lennon J.T. (2016) Scaling laws predict global microbial diversity. *Proc Natl Acad Sci USA* 113:5970-5975.

The Journey



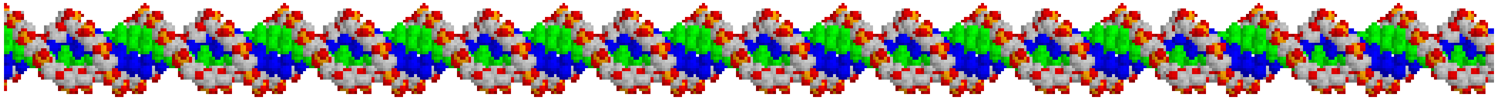
- Starting point: Micro UK's from freshwater creeks → Traditional tests → Bergey's MDB
- Biolog GenIII + FAME + 16S seq → ID
- Research Methods – Complete 16S Compare phenotypes to closest type strains
    - Literature → family/genus specific tests
- Genomes →Calculate Phylogenomic Metrics (GGDC, ANI, AAI, ROSA) (& %GC)
- RAST/SEED viewer for genome analysis, correlation with phenotypes
- Simple web/excel tools use RAST output to:
    - Calculate AAI, BBH, ROSA
    - ID unique and shared genes
    - Create Venn Diagram
    - Correlate with phenotype

Cost per Genome



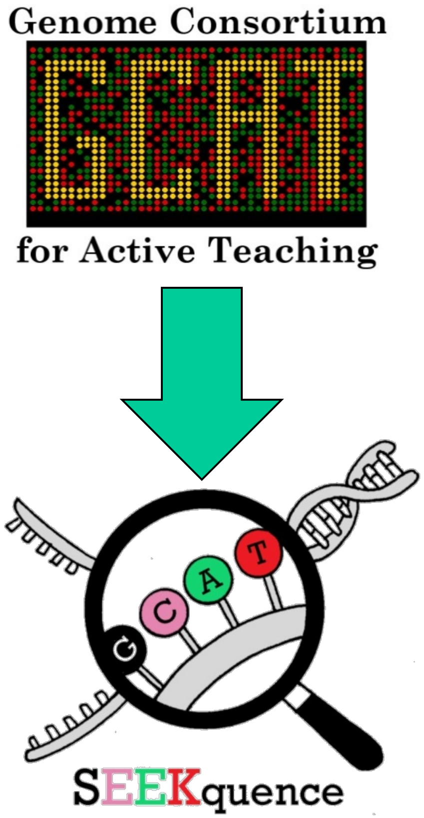
- Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program Available at: www.genome.gov/sequencingcosts. Accessed [6-15-16].

GCAT→ GCAT-SEEK

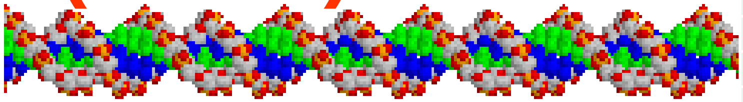


- Genome Consortium for Active Teaching (GCAT) founded in 2000 to bring Genomics (Microarrays) to the undergraduate curriculum.
- Multiple HHMI & NSF funded workshops

- 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 College’s HHMI Genomics Leadership Initiative



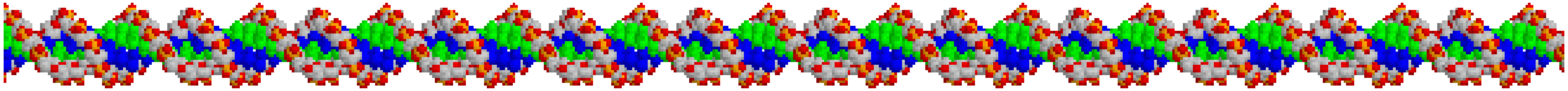
Shared MiSeq (2x300) Runs



- 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.

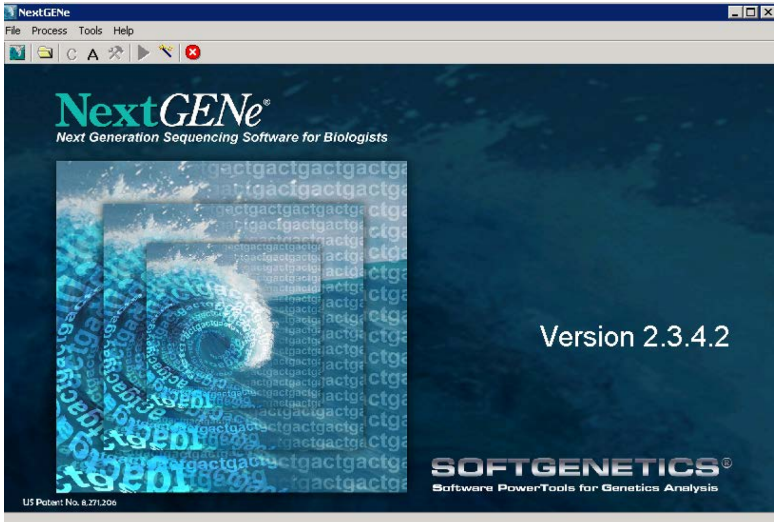
Sample	Reads est.	Bases est.
GSF665-1-E_coli-C06b	217,320	130,391,966
GSF665-2-Chryseobacterium-LO	1,317,872	790,723,170
GSF665-3-Linfield-KH	809,893	485,935,870
GSF665-4-Linfield-NH	301,171	180,702,758
GSF665-5-Exiguobacterium	794,482	476,689,384
GSF665-6-Plesiomonas_shigelloides	656,143	393,685,659
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



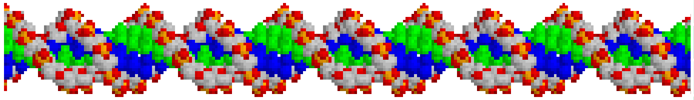
[SoftGenetics Assembler: Assembly Results Statistics Report]

- **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]

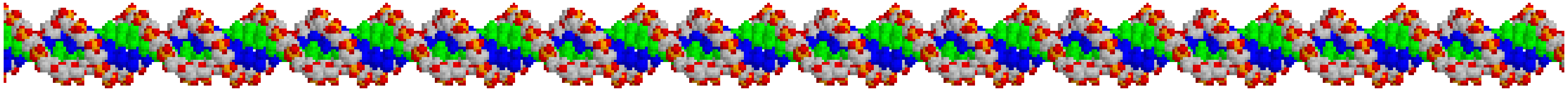
- **Final Contig Merge Sequences Number: 13**
- Final Contig Merge Average Sequence Length: 269063
- Final Contig Merge Minimum Sequence Length: 173
- Final Contig Merge Maximum Sequence Length: 856388
- **Final Contig Merge N50 Length: 586767**
- Matched Reads Count: 1977550
- Number of Matched Bases: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

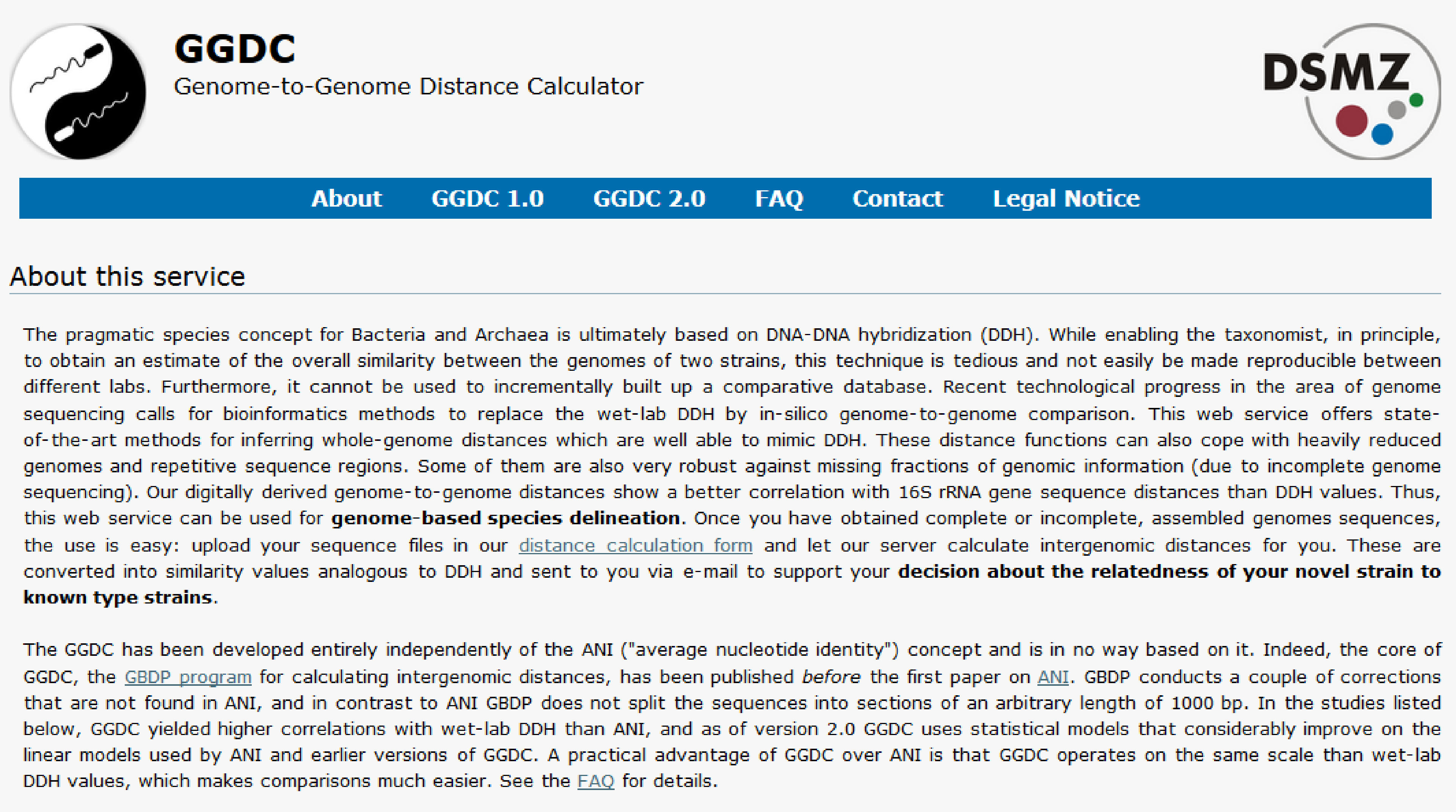


Genomes deposited in GenBank

species	**strain**	**WGS**	**genome size**	**contigs**
Bacillus cibi	DSM 16189	JNVC02	4,072,175	24
Bacillus indicus	LMG 22858	JGVU02	4,129,127	22
Bacillus sp SJS (colbertis)	SJS	JAQV02	4,148,893	36
Chryseobacterium angstadtii	KM	LFND01	5,202,773	11
Chryseobacterium antarcticum	LMG 24720	JPEP01	3,123,663	4
Chryseobacterium aquaticum	KCTC 12483	LLYZ01	3,813,178	21
Chryseobacterium formosense	LMG 24722	JPRP01	4,364,663	10
Chryseobacterium greenlandense	UMB34	LMAI01	4,036,521	21
Chryseobacterium hispalense	DSM 25574	JARQ01	4,363,762	27
Chryseobacterium jeonii	DSM 17048	JSYL01	3,261,282	35
Chryseobacterium luteum	DSM 18605	JPRO01	4,718,546	49
Chryseobacterium piperi	CTM	JPRJ01	4,340,594	89
Chryseobacterium soli	DSM 19298	JPRH01	4,754,478	26
Chryseobacterium solincola	DSM 22468	JSYK01	2,345,236	14
Chryseobacterium sp BLS98	BLS98	LFNF01	4,412,600	8
Chryseobacterium sp FH2	FH2	LFNE01	3,988,758	24
Chryseobacterium sp. JM1	JM1	JPRN01	5,256,156	31
Chryseobacterium vrystaatense	LMG 22846	JPRI01	5,487,842	18
Epilithonimonas diehli FH1	FH1	JPLZ01	3,975,884	12
Epilithonimonas lactis	LMG 24401	JPLY01	4,236,390	22
Flavobacterium aquatile	LMG 4008	JRHH01	3,490,856	7
Flavobacterium chilense	LMG 26360	LSYT01	5,924,510	21
Flavobacterium chungangense	LMG 26729	JASY01	5,591,549	45
Flavobacterium hibernum	DSM 12611	JPRK01	5,283,672	30
Flavobacterium hydatis	DSM 2063	JPRM01	5,877,671	99
Flavobacterium reichenbachii	LMG 25512	JPRL01	5,501,401	12
Flavobacterium sp AED (franzi)	AED	JSYM01	3,925,153	15
Flavobacterium sp JRM	JRM	JSYO01	5,380,719	169
Flavobacterium sp KJJ	KJJ	JARP01	4,603,121	10
Flavobacterium sp KMS	KMS	JSYP01	5,597,129	58
Flavobacterium sp VMW	VMW	LHVT01	5,627,095	59
Flavobacterium sp. ABG (douthatii)	ABG	LEKS01	5,270,010	89
Flavobacterium succinicans	LMG 10402	JATV01	3,652,806	48
Kaistella koreensis	CCUG	LFNG01	3,154,203	64
Pedobacter borealis	DSM 19626	JAUG01	5,544,917	216
Pedobacter ginsenosidimutans	KACC 14530	LMZQ01	6,517,553	86
Pedobacter kyungheensis	KACC 16221	JSYN01	6,358,642	67
Pedobacter sp 20-19	R20-19	JCKI01	5,006,523	15
Pedobacter sp. BMA	BMA	LECU01	5,040,634	20

Genome-Genome Distance Calculator (GGDC) at the DSMZ calculates digital DDH based on “high scoring segment pairs”





GGDC

Genome-to-Genome Distance Calculator

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.

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

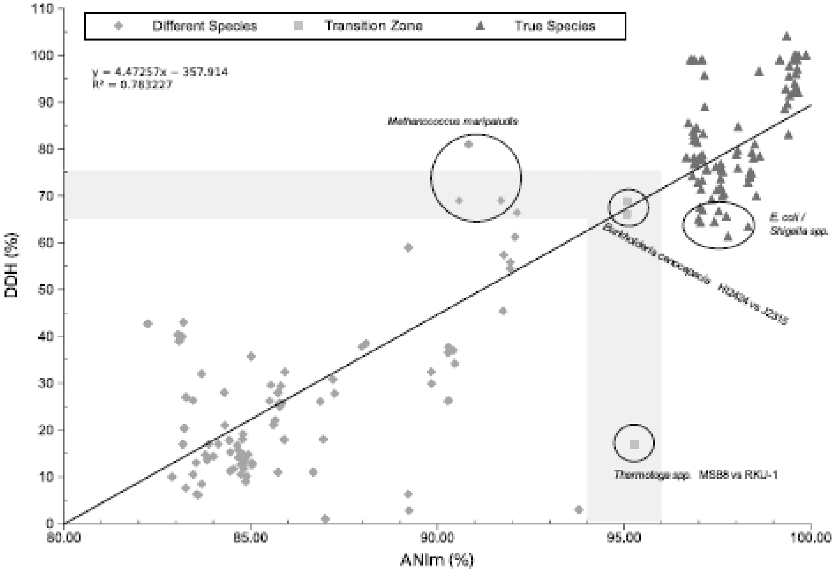


Fig. 2. Plotted values of DDH versus ANIm. Triangles show values that correspond to what taxonomists consider as ''true'' species according to the DDH values traditionally applied and that have previously been classified. Squares indicate values that appear to be in the transition zone.

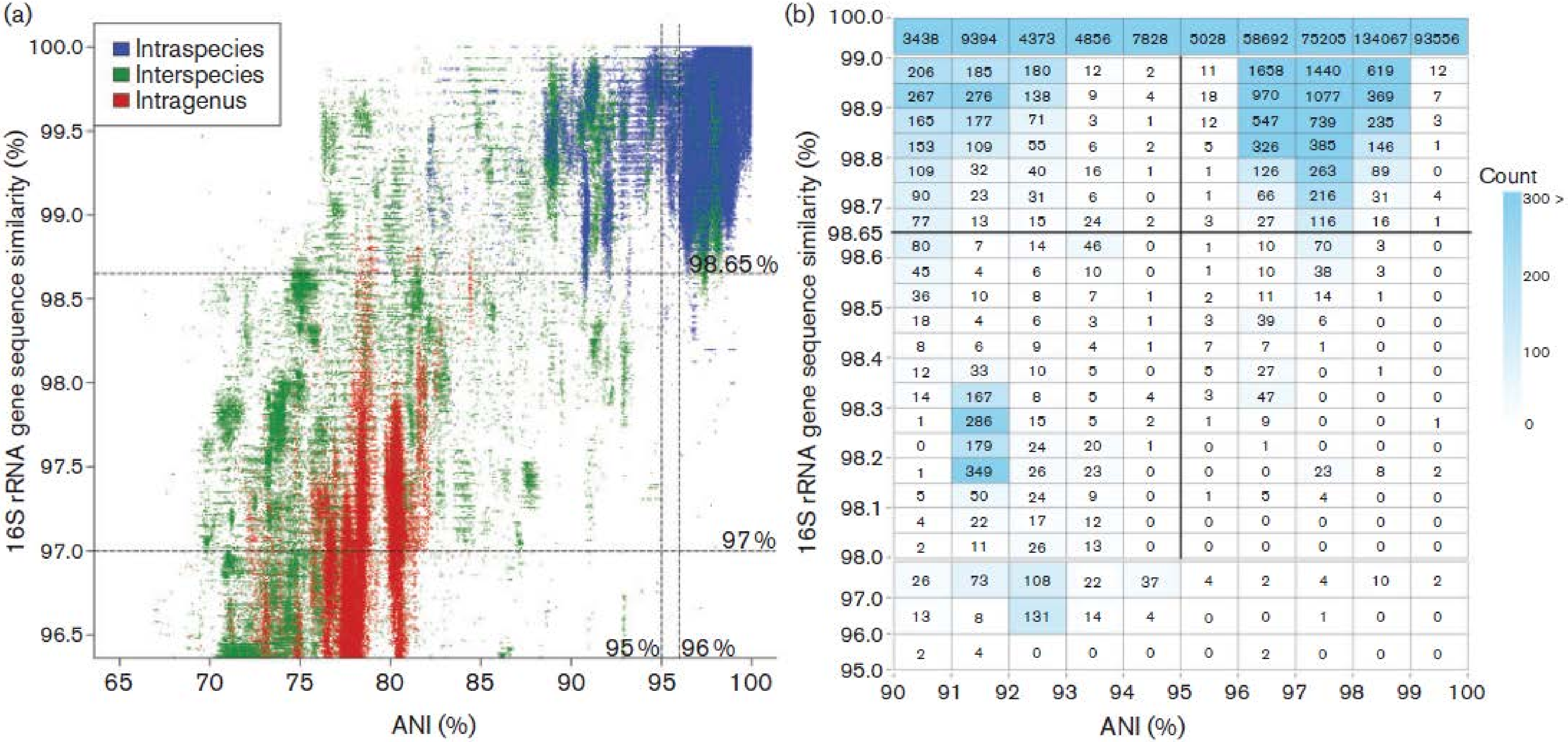
19126–19131 | PNAS | **November 10, 2009** | vol. 106 | no. 45

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

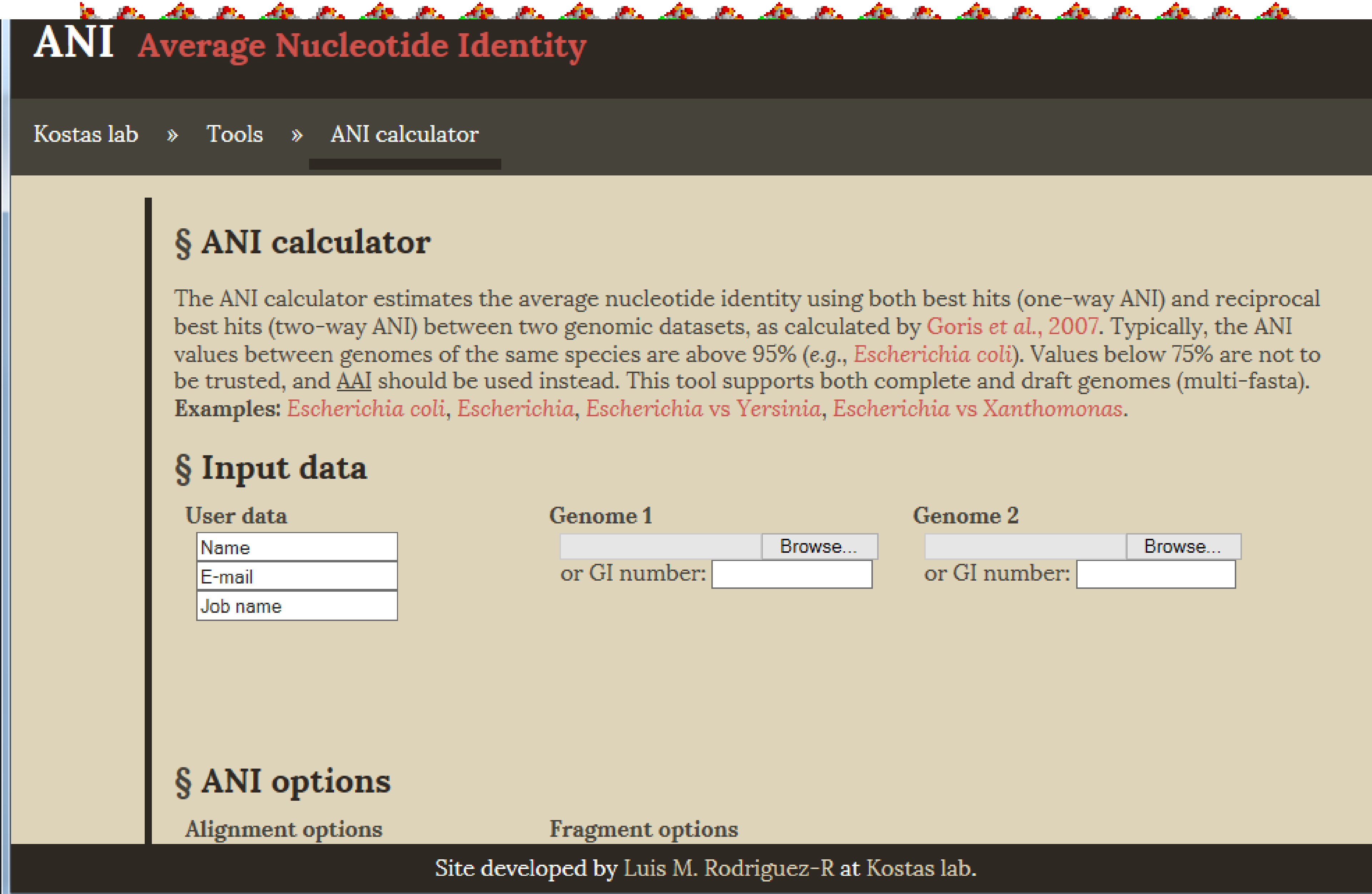
Mincheol Kim,[1] Hyun-Seok Oh,[2] Sang-Cheol Park[2] and Jongsik Chun[1,2]



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

§ 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

Genome 1

or GI number:

Genome 2

or GI number:

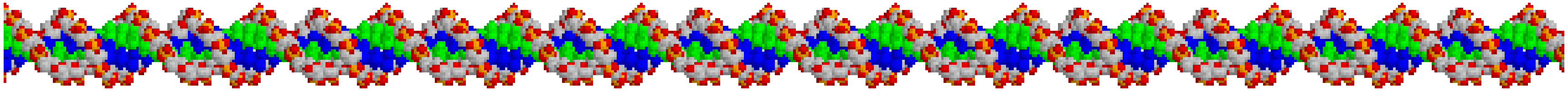
§ ANI options

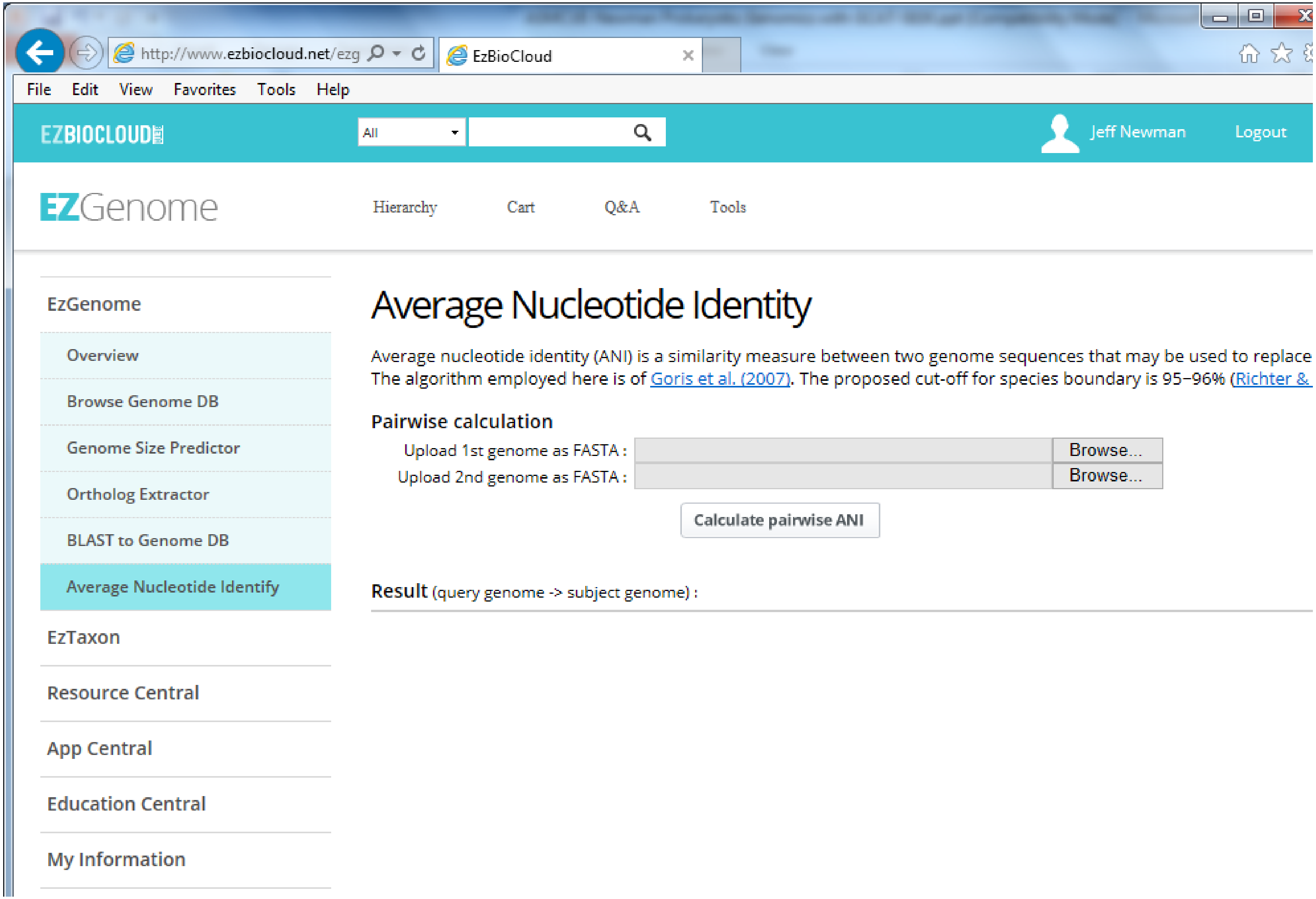
Alignment options

Fragment options

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

Determine Phylogenomic Metrics – Chun Lab EzGenome ANI Calculator





Average Nucleotide Identity

Average nucleotide identity (ANI) is a similarity measure between two genome sequences that may be used to replace
The algorithm employed here is of Goris et al. (2007). The proposed cut-off for species boundary is 95~96% (Richter &

Pairwise calculation

Upload 1st genome as FASTA :

Upload 2nd genome as FASTA :

Result (query genome -> subject genome) :

ANI & GGDC

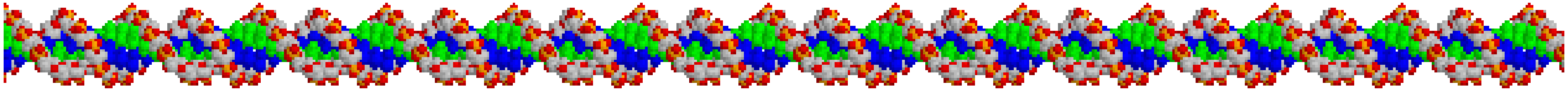
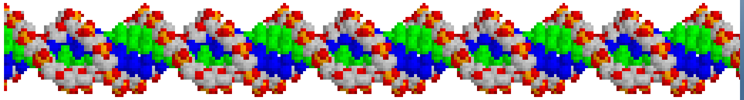


Table 2. Estimated DNA-DNA Hybridization (species threshold = 70%) calculated by GGDC (Meier-Kolthoff. et al, 2013) and Average Nucleotide Identity (ANI) (species threshold = 95%) calculated by Kostas lab ANI Calculator http://enve-omics.ce.gatech.edu/ani/

	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**	**-**

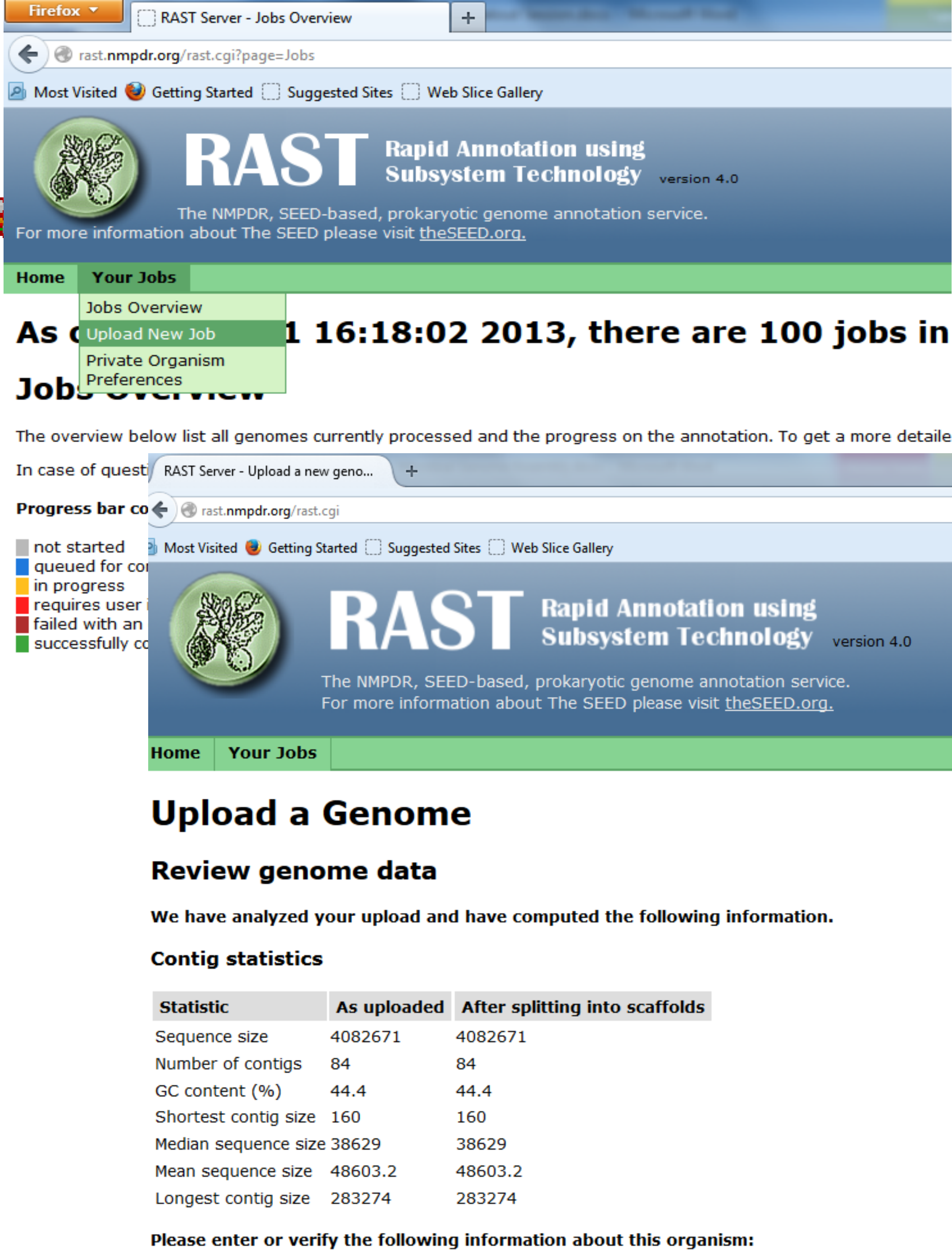
Annotation with RAST



- http://rast.nmpdr.org/

- Upload genome as a .fasta file,

- Get GC mol% composition
- 1-24 hrs later…



RAST 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.

As o… 1 16:18:02 2013, there are 100 jobs in

Job…

The overview below list all genomes currently processed and the progress on the annotation. To get a more detaile

In case of questi

Progress bar co

- ⬜ not started
- 🟦 queued for co
- 🟨 in progress
- 🟥 requires user
- 🟥 failed with an
- 🟩 successfully co

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.

Upload a Genome

Review genome data

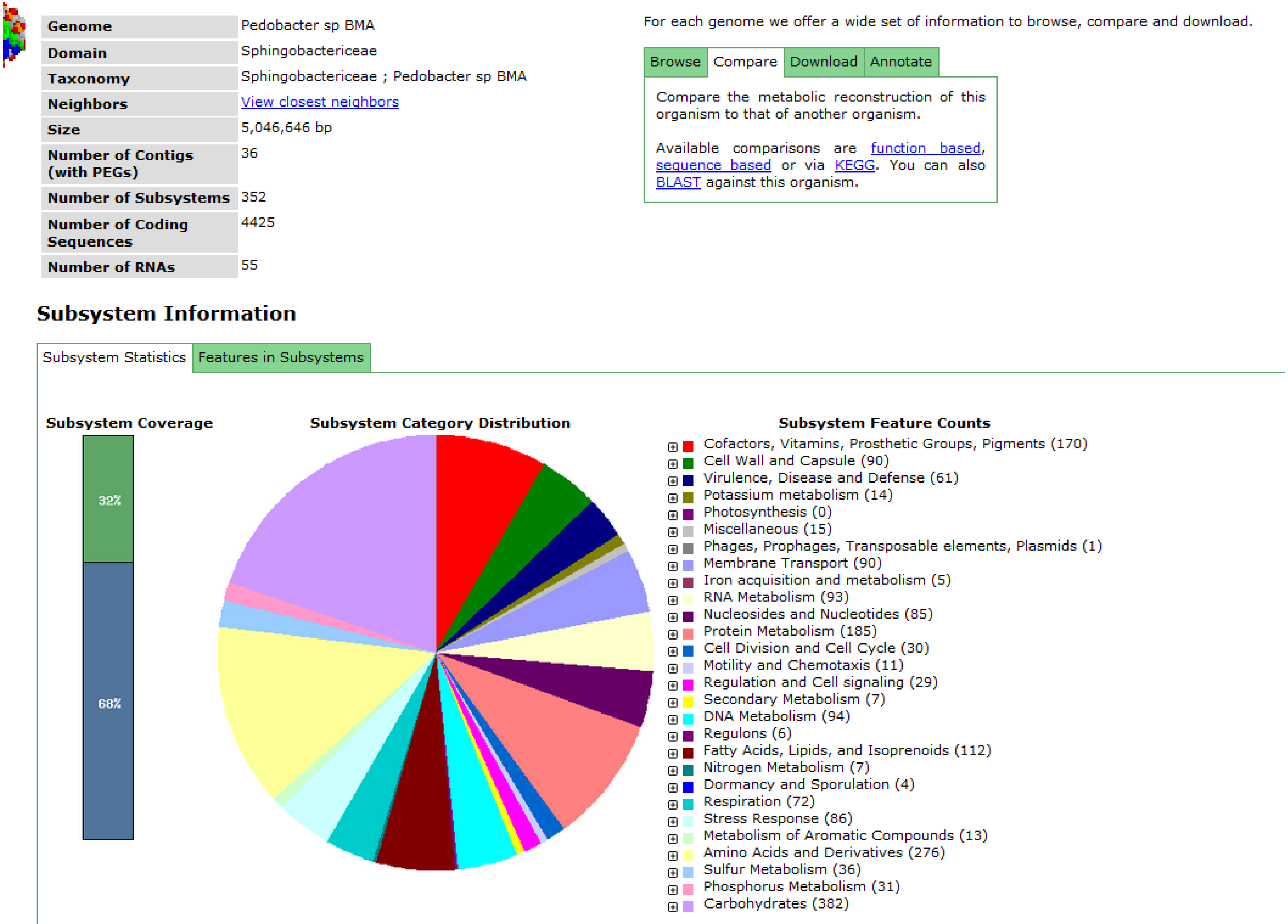
We have analyzed your upload and have computed the following information.

Contig statistics

Statistic	As uploaded	After splitting into scaffolds
Sequence size	4082671	4082671
Number of contigs	84	84
GC content (%)	44.4	44.4
Shortest contig size	160	160
Median sequence size	38629	38629
Mean sequence size	48603.2	48603.2
Longest contig size	283274	283274

Please enter or verify the following information about this organism:

RAST Annotation
Genes are identified and classified by function



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.

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.

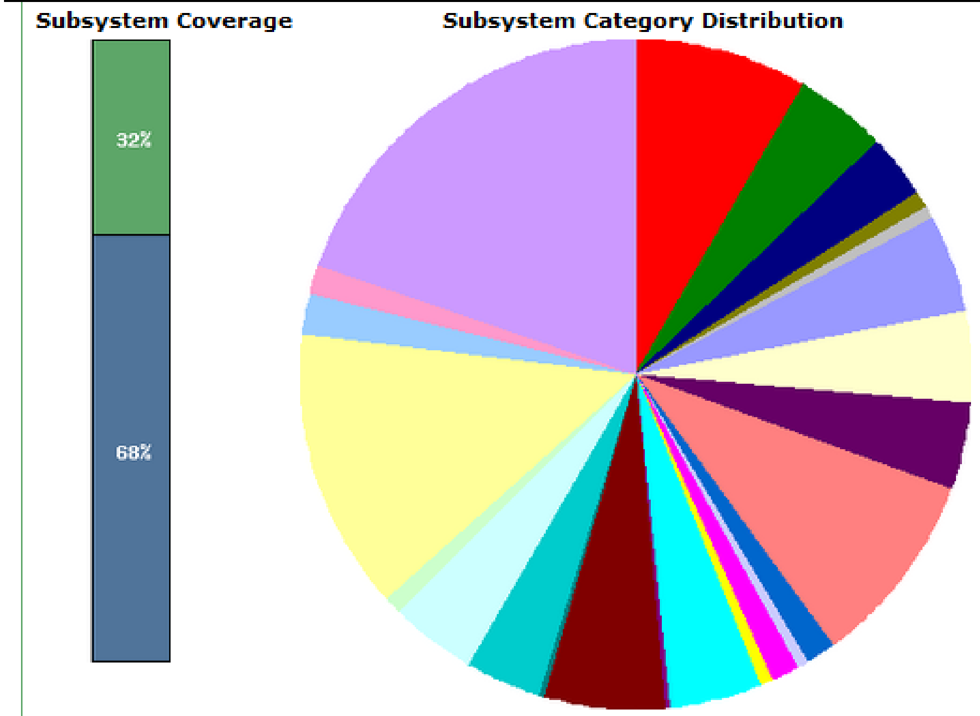
Subsystem Information

Subsystem 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 Coverage

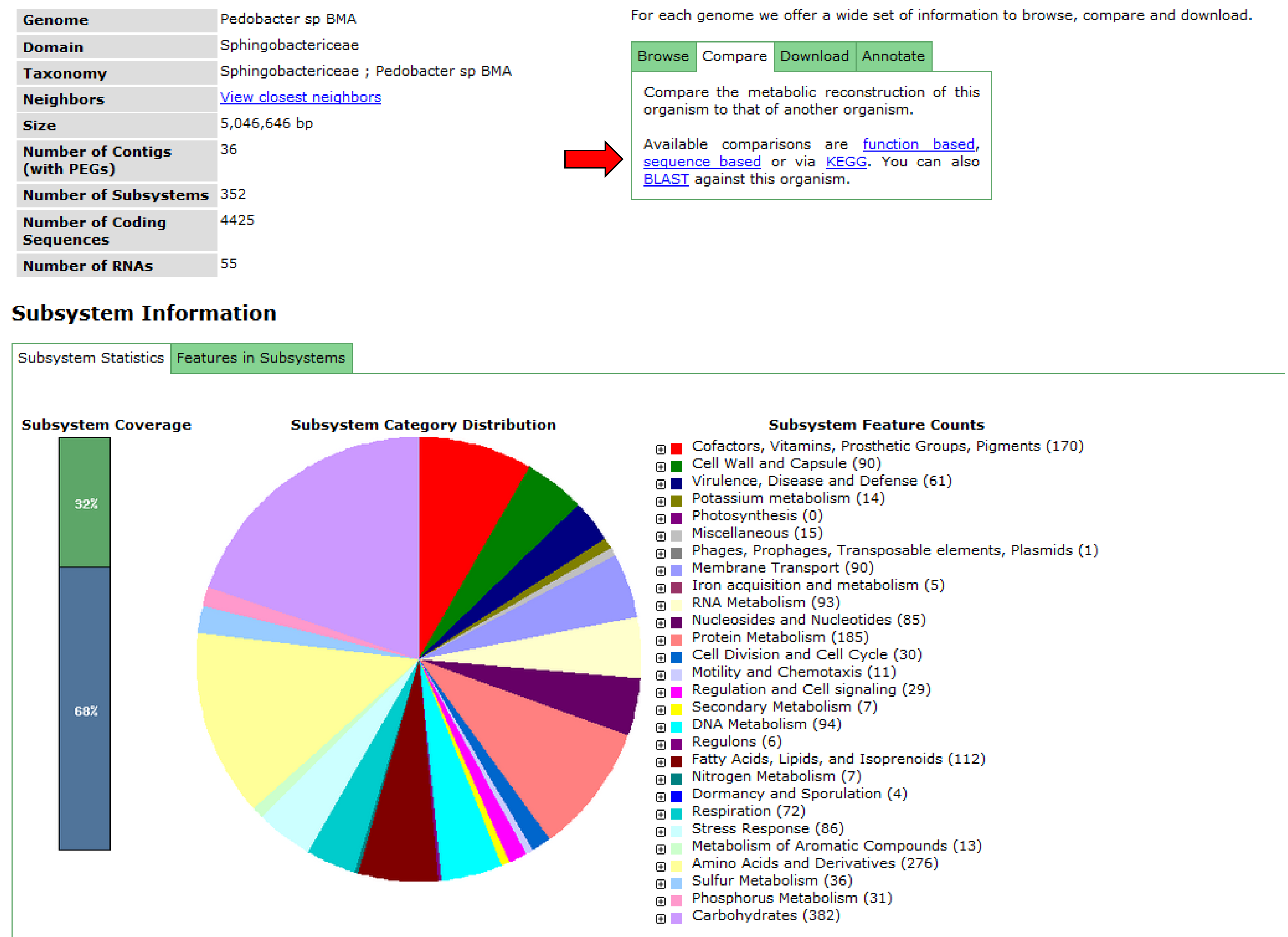
Subsystem Category Distribution

Many genes assigned to expandable subsystems

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)

Seed Viewer Sequence Based Comparison Tool



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.

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

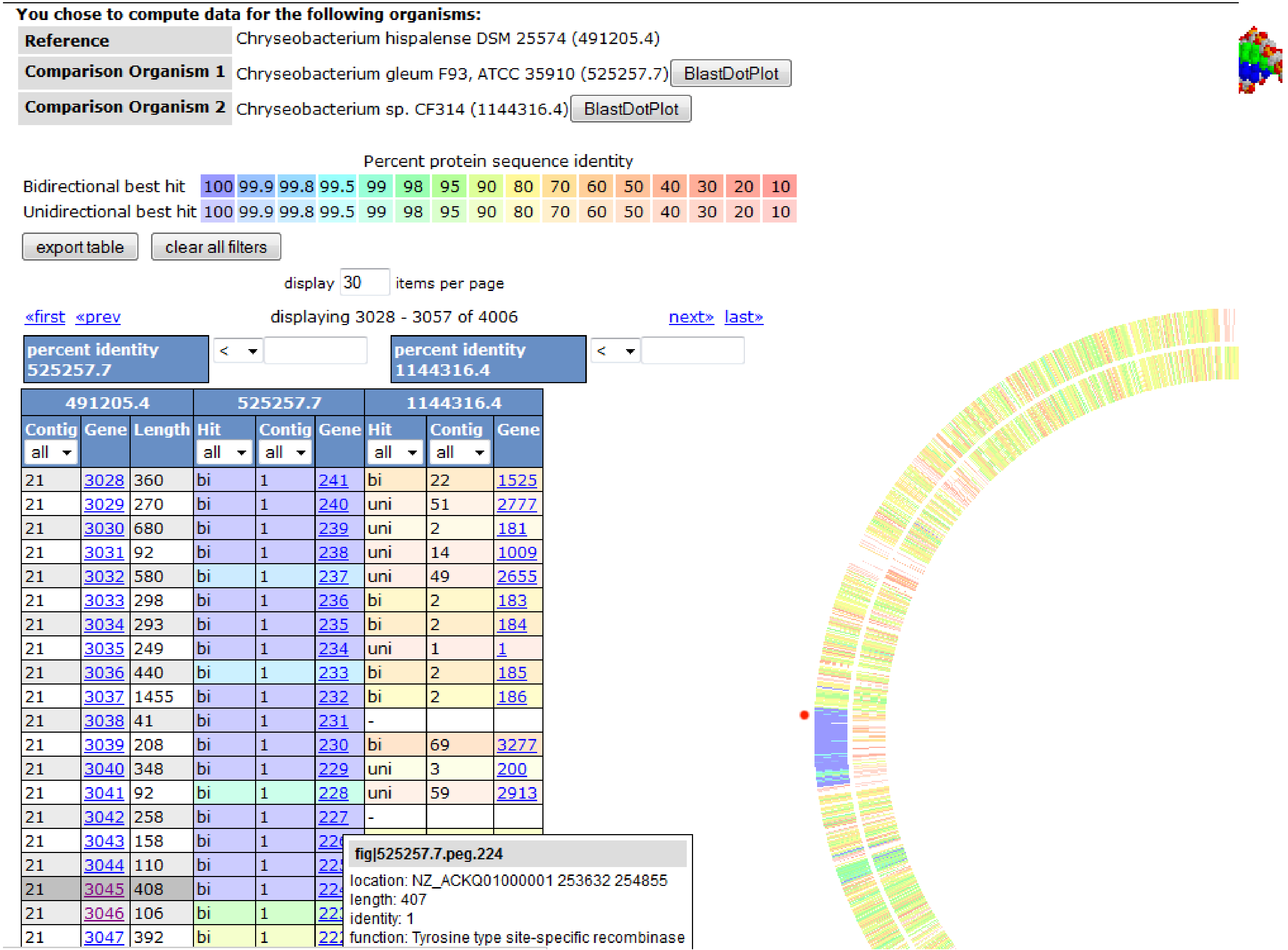
Subsystem Coverage

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)

Sequence-Based Comparison color codes similarity, "bi"-directional best hits identify orthologs



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)
Comparison Organism 2	Chryseobacterium sp. CF314 (1144316.4)

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

display 30 items per page

displaying 3028 - 3057 of 4006

percent identity 525257.7

percent identity 1144316.4

491205.4			525257.7			1144316.4		
Contig	Gene	Length	Hit	Contig	Gene	Hit	Contig	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	22			
21	3044	110	bi	1	22			
21	3045	408	bi	1	22			
21	3046	106	bi	1	22			
21	3047	392	bi	1	22			

fig|525257.7.peg.224

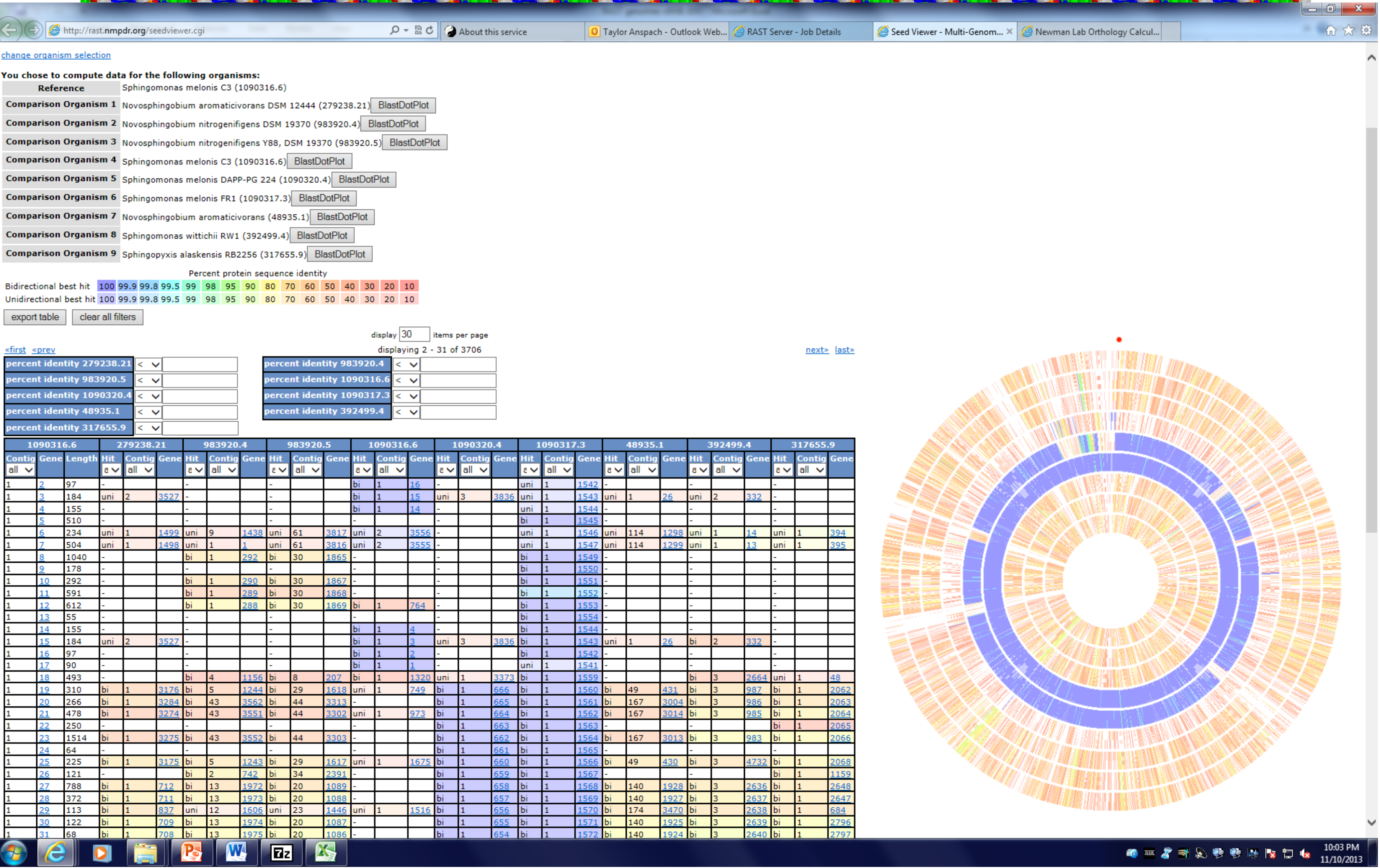
location: NZ_ACKQ01000001 253632 254855

length: 407

identity: 1

function: Tyrosine type site-specific recombinase

Sequence based comparison tool can compare 1 reference to up to 10 comparison genomes



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

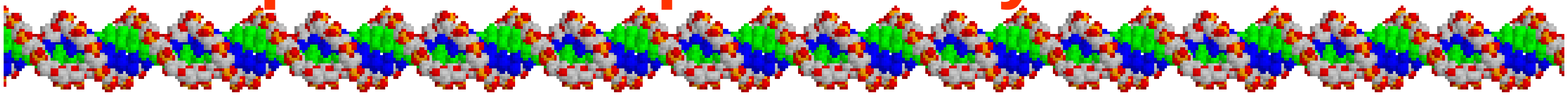
display 30 items per page

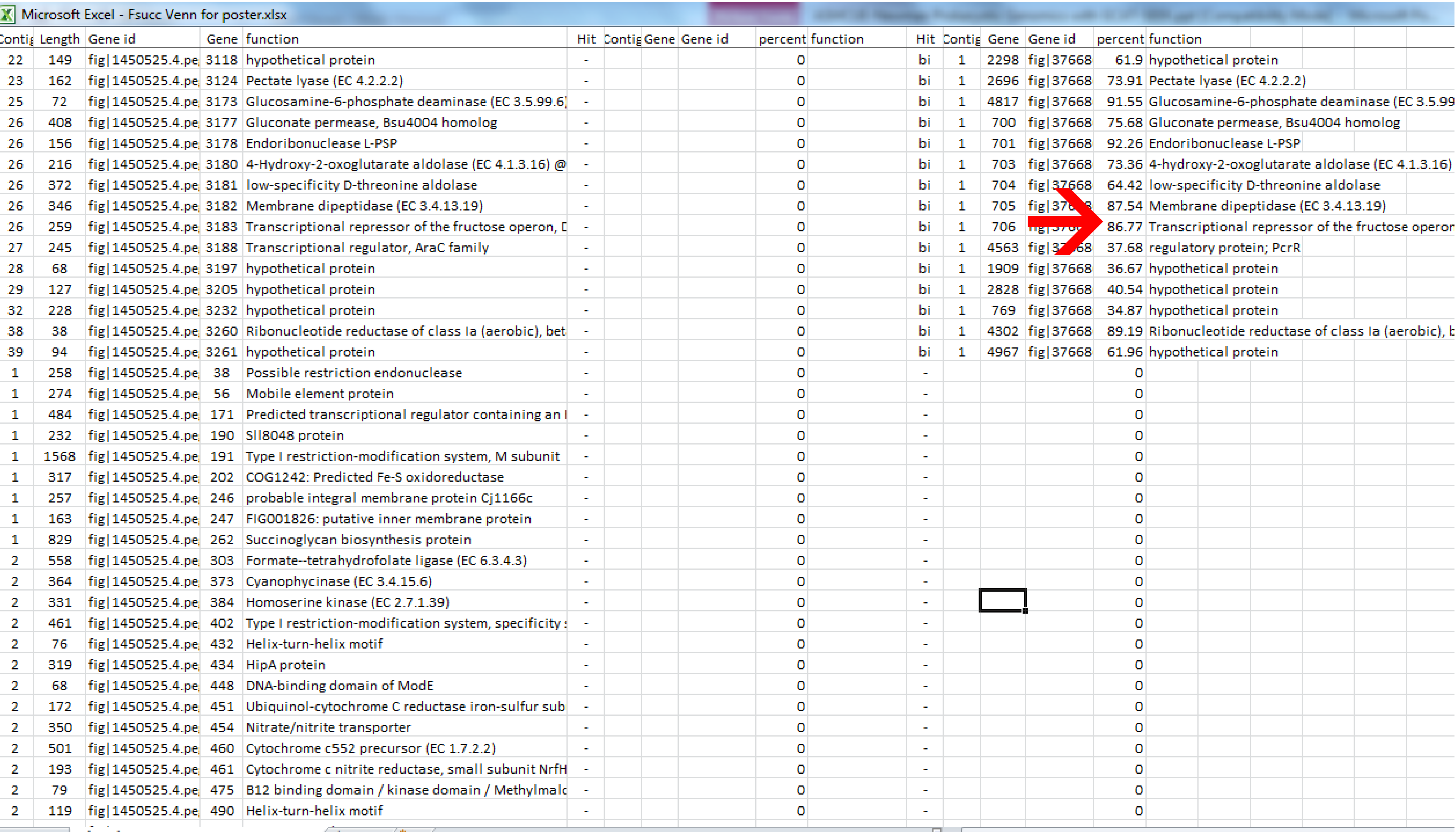
«first «prev　displaying 2 - 31 of 3706　next» last»

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			1090320.4			1090317.3			48935.1			392499.4			317655.9		
Contig all	Gene	Length	Hit a	Contig all	Gene	Hit a	Contig all	Gene	Hit a	Contig all	Gene	Hit a	Contig all	Gene	Hit a	Contig all	Gene	Hit a	Contig all	Gene	Hit a	Contig all	Gene	Hit a	Contig all	Gene	Hit a	Contig all	Gene
1	2	97	-			-			-			bi	1	16	-			uni	1	1542	-			-			-		
1	3	184	uni	2	3527	-			-			bi	1	15	uni	3	3836	uni	1	1543	uni	1	26	uni	2	332	-		
1	4	155	-			-			-			bi	1	14	-			uni	1	1544	-			-			-		
1	5	510	-			-			-			-			-			bi	1	1545	-			-			-		
1	6	234	uni	1	1499	uni	9	1438	uni	61	3817	uni	2	3556	-			uni	1	1546	uni	114	1298	uni	1	14	uni	1	394
1	7	504	uni	1	1498	uni	1	1	uni	61	3816	uni	2	3555	-			uni	1	1547	uni	114	1299	uni	1	13	uni	1	395
1	8	1040	-			bi	1	292	bi	30	1865	-			-			bi	1	1549	-			-			-		
1	9	178	-			-			-			-			-			bi	1	1550	-			-			-		
1	10	292	-			bi	1	290	bi	30	1867	-			-			bi	1	1551	-			-			-		
1	11	591	-			bi	1	289	bi	30	1868	-			-			bi	1	1552	-			-			-		
1	12	612	-			bi	1	288	bi	30	1869	bi	1	764	-			bi	1	1553	-			-			-		
1	13	55	-			-			-			-			-			bi	1	1554	-			-			-		
1	14	155	-			-			-			bi	1	4	-			bi	1	1544	-			-			-		
1	15	184	uni	2	3527	-			-			bi	1	3	uni	3	3836	bi	1	1543	uni	1	26	bi	2	332	-		
1	16	97	-			-			-			bi	1	2	-			bi	1	1542	-			-			-		
1	17	90	-			-			-			bi	1	1	-			uni	1	1541	-			-			-		
1	18	493	-			bi	4	1156	bi	8	207	bi	1	1320	uni	1	3373	bi	1	1559	-			bi	3	2664	uni	1	48
1	19	310	bi	1	3176	bi	5	1244	bi	29	1618	uni	1	749	bi	1	666	bi	1	1560	bi	49	431	bi	3	987	bi	1	2062
1	20	266	bi	1	3284	bi	43	3562	bi	44	3313	-			bi	1	665	bi	1	1561	bi	167	3004	bi	3	986	bi	1	2063
1	21	478	bi	1	3274	bi	43	3551	bi	44	3302	uni	1	973	bi	1	664	bi	1	1562	bi	167	3014	bi	3	985	bi	1	2064
1	22	250	-			-			-			-			bi	1	663	bi	1	1563	-			-			bi	1	2065
1	23	1514	bi	1	3275	bi	43	3552	bi	44	3303	-			bi	1	662	bi	1	1564	bi	167	3013	bi	3	983	bi	1	2066
1	24	64	-			-			-			-			bi	1	661	bi	1	1565	-			-			-		
1	25	225	bi	1	3175	bi	5	1243	bi	29	1617	uni	1	1675	bi	1	660	bi	1	1566	bi	49	430	bi	3	4732	bi	1	2068
1	26	121	-			bi	2	742	bi	34	2391	-			bi	1	659	bi	1	1567	-			-			bi	1	1159
1	27	788	bi	1	712	bi	13	1972	bi	20	1089	-			bi	1	658	bi	1	1568	bi	140	1928	bi	3	2636	bi	1	2648
1	28	372	bi	1	711	bi	13	1973	bi	20	1088	-			bi	1	657	bi	1	1569	bi	140	1927	bi	3	2637	bi	1	2647
1	29	113	bi	1	837	uni	12	1606	uni	23	1446	uni	1	1516	bi	1	656	bi	1	1570	bi	174	3470	bi	3	2638	bi	1	684
1	30	122	bi	1	709	bi	13	1974	bi	20	1087	-			bi	1	655	bi	1	1571	bi	140	1925	bi	3	2639	bi	1	2796
1	31	68	bi	1	708	bi	13	1975	bi	20	1086	-			bi	1	654	bi	1	1572	bi	140	1924	bi	3	2640	bi	1	2797

Sequence Based Comparison provides protein seq similarity → AAI

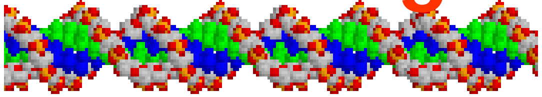




Microsoft Excel - Fsucc Venn for poster.xlsx

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|376	87.54	Membrane dipeptidase (EC 3.4.13.19)
26	259	fig|1450525.4.pe	3183	Transcriptional repressor of the fructose operon,	-				0		bi	1	706		86.77	Transcriptional repressor of the fructose operon
27	245	fig|1450525.4.pe	3188	Transcriptional regulator, AraC family	-				0		bi	1	4563	fig|3 68	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),
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	-				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	-				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 / Methylmal	-				0		-				0	
2	119	fig|1450525.4.pe	490	Helix-turn-helix motif	-				0		-				0	

AAI has broader range



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

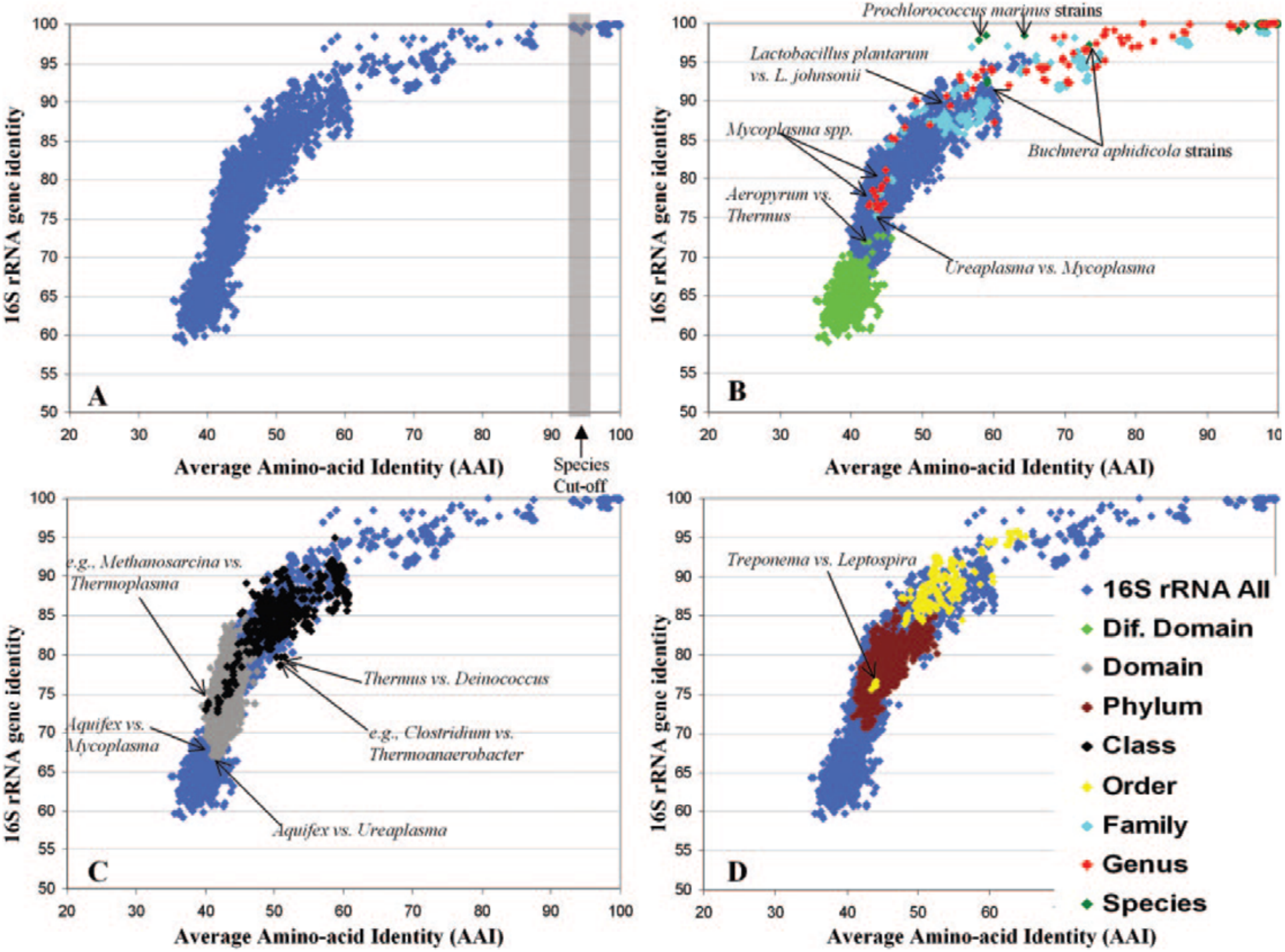
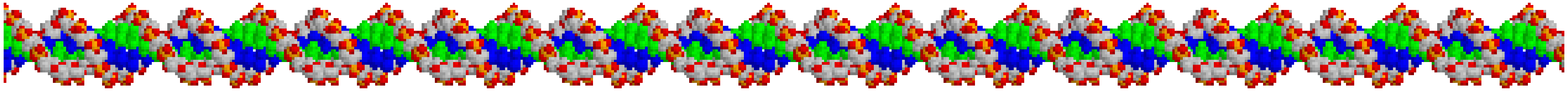
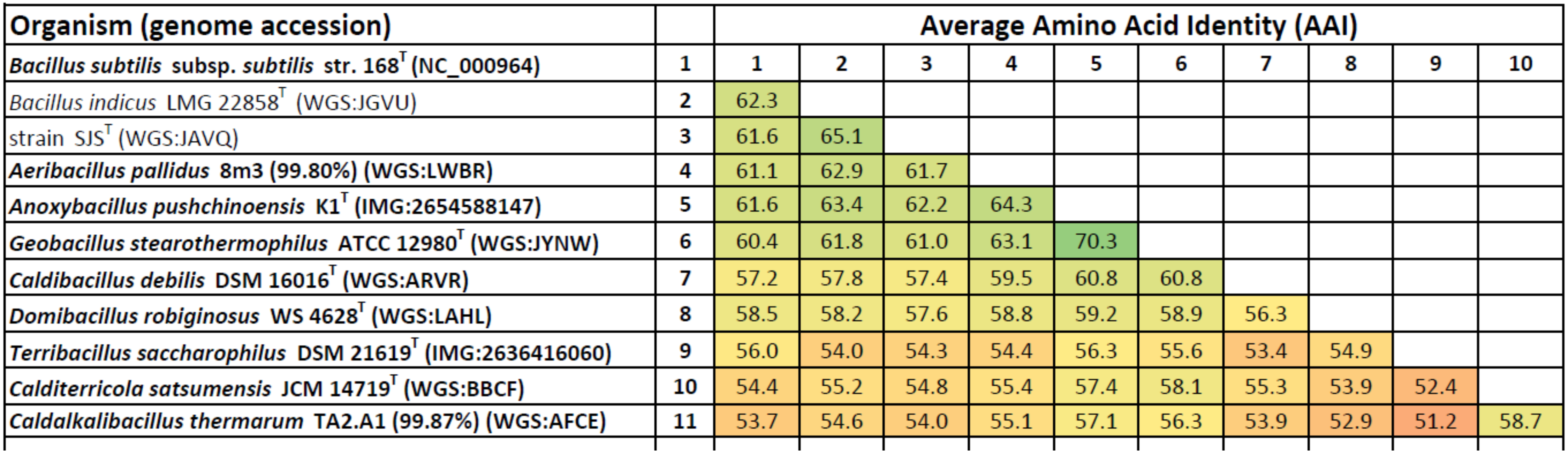


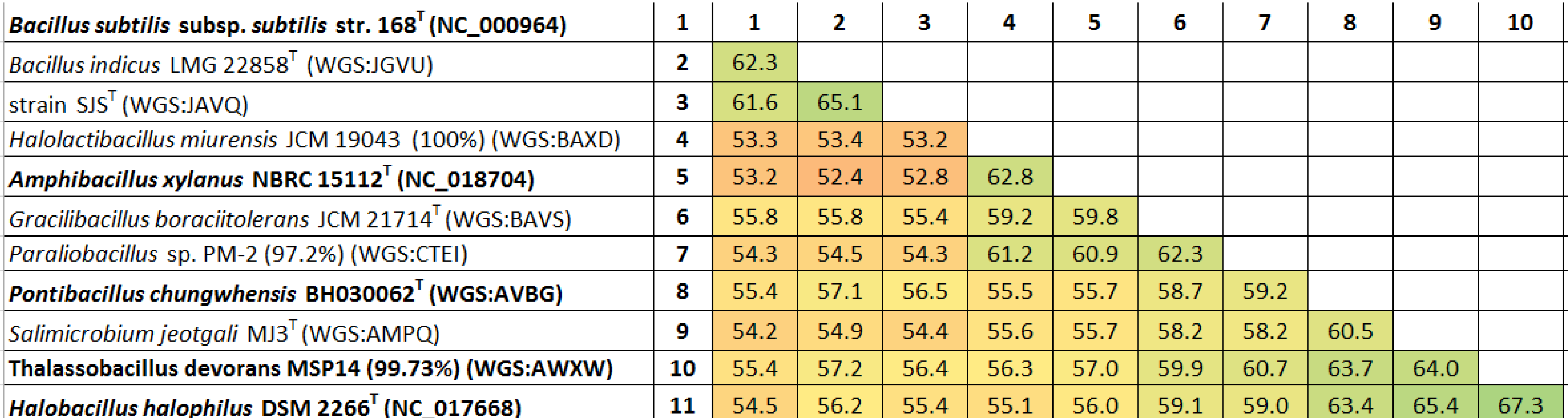
FIG. 3. Relationships between 16S rRNA, AAI, and taxonomic information for the 175 sequenced genomes. Each dot represents a compa

Different genera in same family have AAI of ~51 to ~71-75



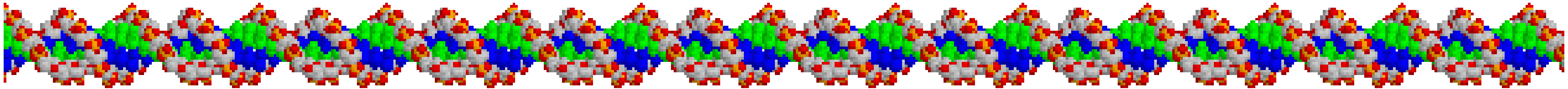


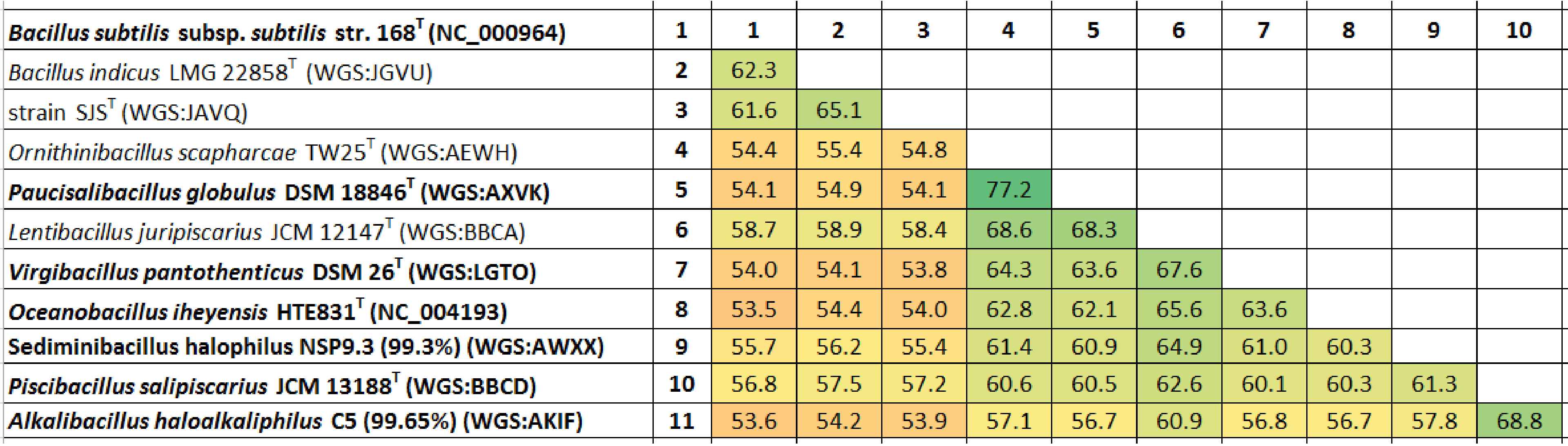
Organism (genome accession)		**Average Amino Acid Identity (AAI)**									
***Bacillus subtilis* subsp. *subtilis* str. 168^T (NC_000964)**	**1**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Bacillus indicus LMG 22858^T (WGS:JGVU)	**2**	62.3									
strain SJS^T (WGS:JAVQ)	**3**	61.6	65.1								
***Aeribacillus pallidus* 8m3 (99.80%) (WGS:LWBR)**	**4**	61.1	62.9	61.7							
***Anoxybacillus pushchinoensis* K1^T (IMG:2654588147)**	**5**	61.6	63.4	62.2	64.3						
***Geobacillus stearothermophilus* ATCC 12980^T (WGS:JYNW)**	**6**	60.4	61.8	61.0	63.1	70.3					
***Caldibacillus debilis* DSM 16016^T (WGS:ARVR)**	**7**	57.2	57.8	57.4	59.5	60.8	60.8				
***Domibacillus robiginosus* WS 4628^T (WGS:LAHL)**	**8**	58.5	58.2	57.6	58.8	59.2	58.9	56.3			
***Terribacillus saccharophilus* DSM 21619^T (IMG:2636416060)**	**9**	56.0	54.0	54.3	54.4	56.3	55.6	53.4	54.9		
***Calditerricola satsumensis* JCM 14719^T (WGS:BBCF)**	**10**	54.4	55.2	54.8	55.4	57.4	58.1	55.3	53.9	52.4	
***Caldalkalibacillus thermarum* TA2.A1 (99.87%) (WGS:AFCE)**	**11**	53.7	54.6	54.0	55.1	57.1	56.3	53.9	52.9	51.2	58.7



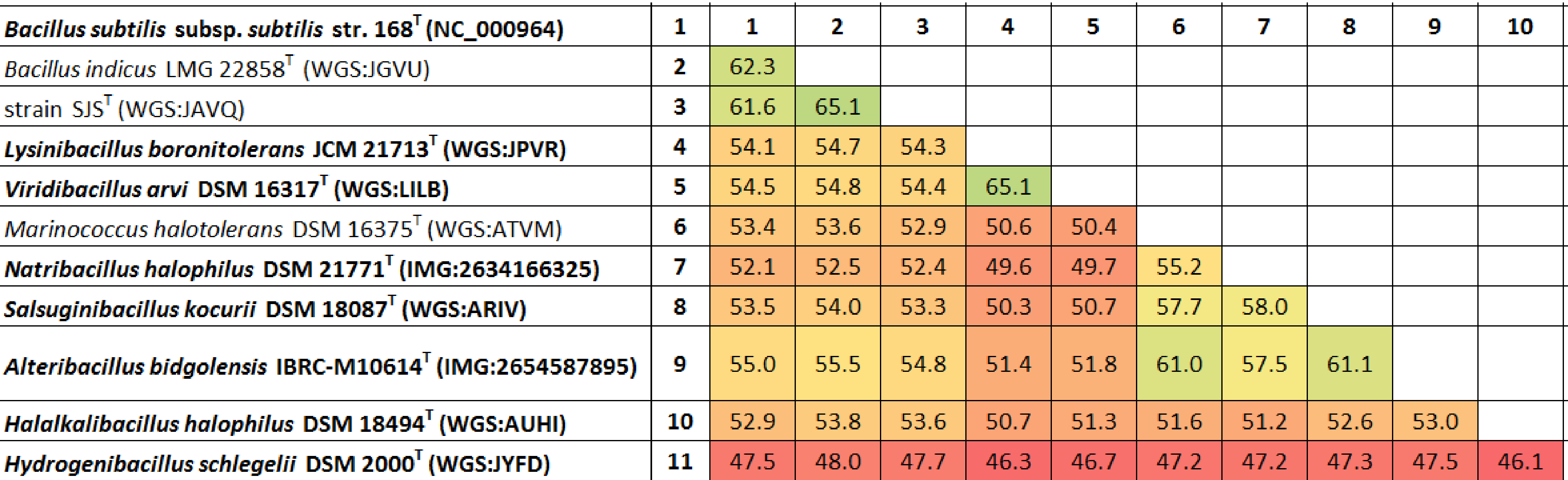
***Bacillus subtilis* subsp. *subtilis* str. 168^T (NC_000964)**	**1**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Bacillus indicus LMG 22858^T (WGS:JGVU)	**2**	62.3									
strain SJS^T (WGS:JAVQ)	**3**	61.6	65.1								
Halolactibacillus miurensis JCM 19043 (100%) (WGS:BAXD)	**4**	53.3	53.4	53.2							
***Amphibacillus xylanus* NBRC 15112^T (NC_018704)**	**5**	53.2	52.4	52.8	62.8						
Gracilibacillus boraciitolerans JCM 21714^T (WGS:BAVS)	**6**	55.8	55.8	55.4	59.2	59.8					
Paraliobacillus sp. PM-2 (97.2%) (WGS:CTEI)	**7**	54.3	54.5	54.3	61.2	60.9	62.3				
***Pontibacillus chungwhensis* BH030062^T (WGS:AVBG)**	**8**	55.4	57.1	56.5	55.5	55.7	58.7	59.2			
Salimicrobium jeotgali MJ3^T (WGS:AMPQ)	**9**	54.2	54.9	54.4	55.6	55.7	58.2	58.2	60.5		
Thalassobacillus devorans MSP14 (99.73%) (WGS:AWXW)	**10**	55.4	57.2	56.4	56.3	57.0	59.9	60.7	63.7	64.0	
***Halobacillus halophilus* DSM 2266^T (NC_017668)**	**11**	54.5	56.2	55.4	55.1	56.0	59.1	59.0	63.4	65.4	67.3

Different genera in same family have AAI of ~51 to ~71-75



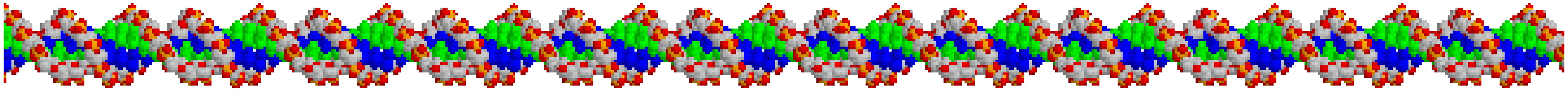


***Bacillus subtilis* subsp. *subtilis* str. 168^T (NC_000964)**	1	1	2	3	4	5	6	7	8	9	10
Bacillus indicus LMG 22858^T (WGS:JGVU)	2	62.3									
strain SJS^T (WGS:JAVQ)	3	61.6	65.1								
Ornithinibacillus scapharcae TW25^T (WGS:AEWH)	4	54.4	55.4	54.8							
***Paucisalibacillus globulus* DSM 18846^T (WGS:AXVK)**	5	54.1	54.9	54.1	77.2						
Lentibacillus juripiscarius JCM 12147^T (WGS:BBCA)	6	58.7	58.9	58.4	68.6	68.3					
***Virgibacillus pantothenticus* DSM 26^T (WGS:LGTO)**	7	54.0	54.1	53.8	64.3	63.6	67.6				
***Oceanobacillus iheyensis* HTE831^T (NC_004193)**	8	53.5	54.4	54.0	62.8	62.1	65.6	63.6			
Sediminibacillus halophilus NSP9.3 (99.3%) (WGS:AWXX)	9	55.7	56.2	55.4	61.4	60.9	64.9	61.0	60.3		
***Piscibacillus salipiscarius* JCM 13188^T (WGS:BBCD)**	10	56.8	57.5	57.2	60.6	60.5	62.6	60.1	60.3	61.3	
***Alkalibacillus haloalkaliphilus* C5 (99.65%) (WGS:AKIF)**	11	53.6	54.2	53.9	57.1	56.7	60.9	56.8	56.7	57.8	68.8

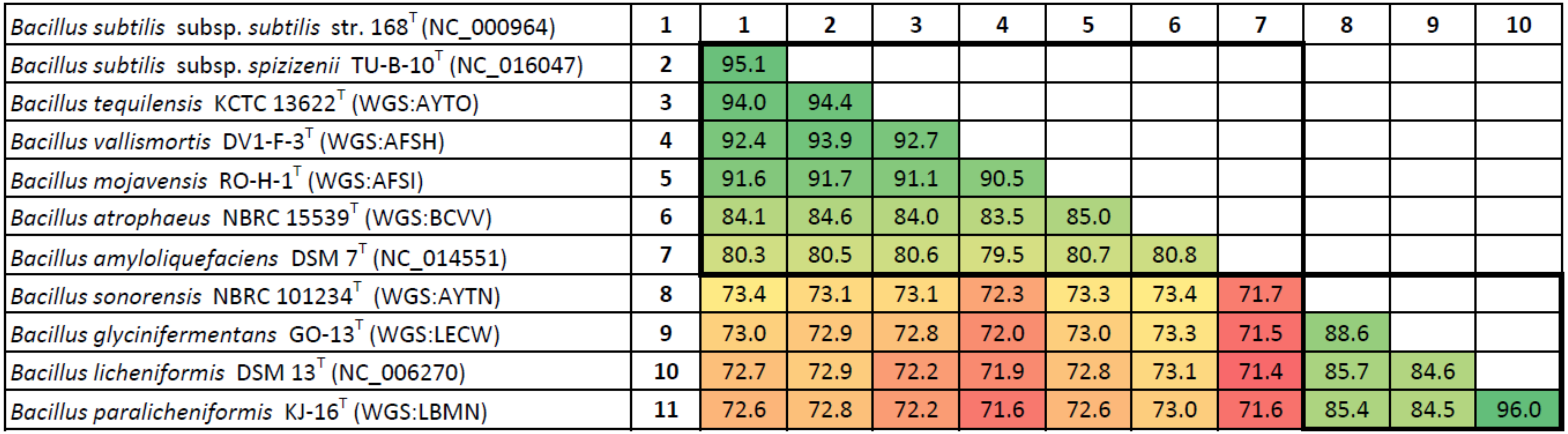


***Bacillus subtilis* subsp. *subtilis* str. 168^T (NC_000964)**	1	1	2	3	4	5	6	7	8	9	10
Bacillus indicus LMG 22858^T (WGS:JGVU)	2	62.3									
strain SJS^T (WGS:JAVQ)	3	61.6	65.1								
***Lysinibacillus boronitolerans* JCM 21713^T (WGS:JPVR)**	4	54.1	54.7	54.3							
***Viridibacillus arvi* DSM 16317^T (WGS:LILB)**	5	54.5	54.8	54.4	65.1						
Marinococcus halotolerans DSM 16375^T (WGS:ATVM)	6	53.4	53.6	52.9	50.6	50.4					
***Natribacillus halophilus* DSM 21771^T (IMG:2634166325)**	7	52.1	52.5	52.4	49.6	49.7	55.2				
***Salsuginibacillus kocurii* DSM 18087^T (WGS:ARIV)**	8	53.5	54.0	53.3	50.3	50.7	57.7	58.0			
***Alteribacillus bidgolensis* IBRC-M10614^T (IMG:2654587895)**	9	55.0	55.5	54.8	51.4	51.8	61.0	57.5	61.1		
***Halalkalibacillus halophilus* DSM 18494^T (WGS:AUHI)**	10	52.9	53.8	53.6	50.7	51.3	51.6	51.2	52.6	53.0	
***Hydrogenibacillus schlegelii* DSM 2000^T (WGS:JYFD)**	11	47.5	48.0	47.7	46.3	46.7	47.2	47.2	47.3	47.5	46.1

Different genera in same family have AAI of ~51 to ~71-75



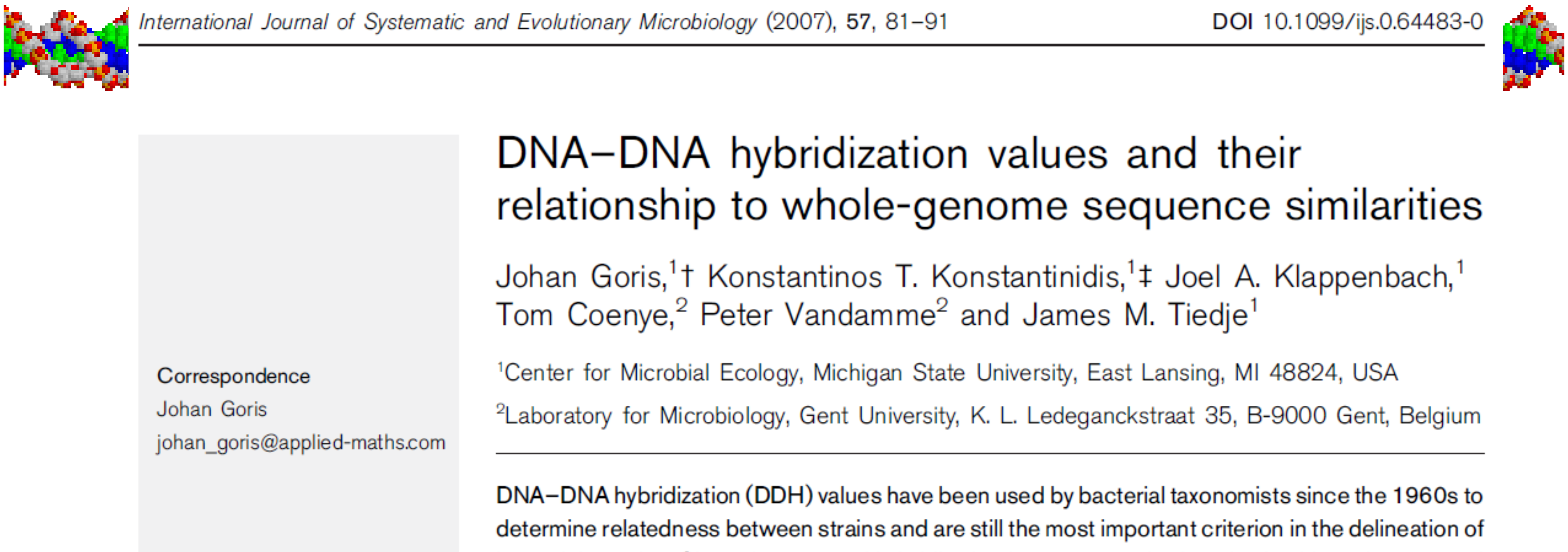
- *Bacillus* sensu stricto
  - *subtilis* group, *licheniformis* groups in 71-75 range
  - Borderline values – should they be separate genera?
  - Need distinguishing **phenotypes** near border to make a case for a separate genus



Bacillus subtilis subsp. *subtilis* str. 168^T (NC_000964)	1	1	2	3	4	5	6	7	8	9	10
Bacillus subtilis subsp. *spizizenii* TU-B-10^T (NC_016047)	2	95.1									
Bacillus tequilensis KCTC 13622^T (WGS:AYTO)	3	94.0	94.4								
Bacillus vallismortis DV1-F-3^T (WGS:AFSH)	4	92.4	93.9	92.7							
Bacillus mojavensis RO-H-1^T (WGS:AFSI)	5	91.6	91.7	91.1	90.5						
Bacillus atrophaeus NBRC 15539^T (WGS:BCVV)	6	84.1	84.6	84.0	83.5	85.0					
Bacillus amyloliquefaciens DSM 7^T (NC_014551)	7	80.3	80.5	80.6	79.5	80.7	80.8				
Bacillus sonorensis NBRC 101234^T (WGS:AYTN)	8	73.4	73.1	73.1	72.3	73.3	73.4	71.7			
Bacillus glycinifermentans GO-13^T (WGS:LECW)	9	73.0	72.9	72.8	72.0	73.0	73.3	71.5	88.6		
Bacillus licheniformis DSM 13^T (NC_006270)	10	72.7	72.9	72.2	71.9	72.8	73.1	71.4	85.7	84.6	
Bacillus paralicheniformis KJ-16^T (WGS:LBMN)	11	72.6	72.8	72.2	71.6	72.6	73.0	71.6	85.4	84.5	96.0

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

Both ortholog similarity and percentage of orthologs contribute to DDH



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

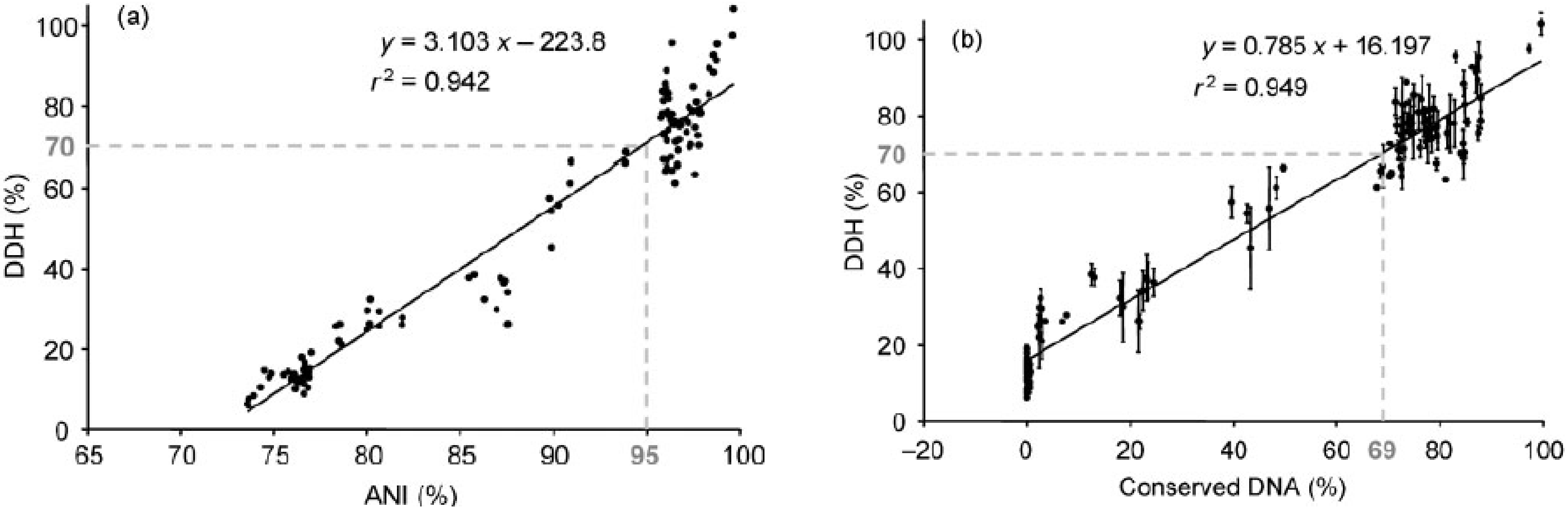
Johan Goris,[1]† Konstantinos T. Konstantinidis,[1]‡ Joel A. Klappenbach,[1]
Tom Coenye,[2] Peter Vandamme[2] and James M. Tiedje[1]

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

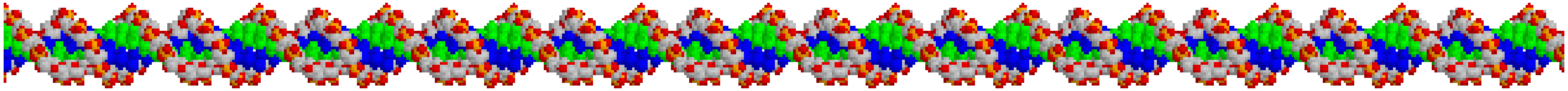
[2]Laboratory for Microbiology, Gent University, K. L. Ledeganckstraat 35, B-9000 Gent, Belgium

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



Reciprocal Orthology Score Average (ROSA)

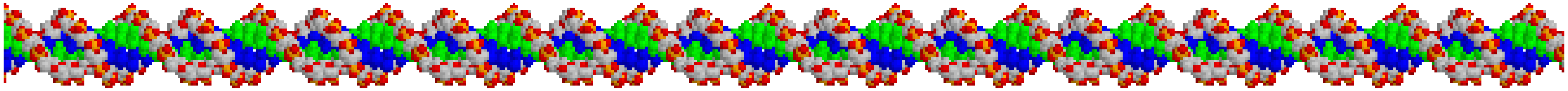


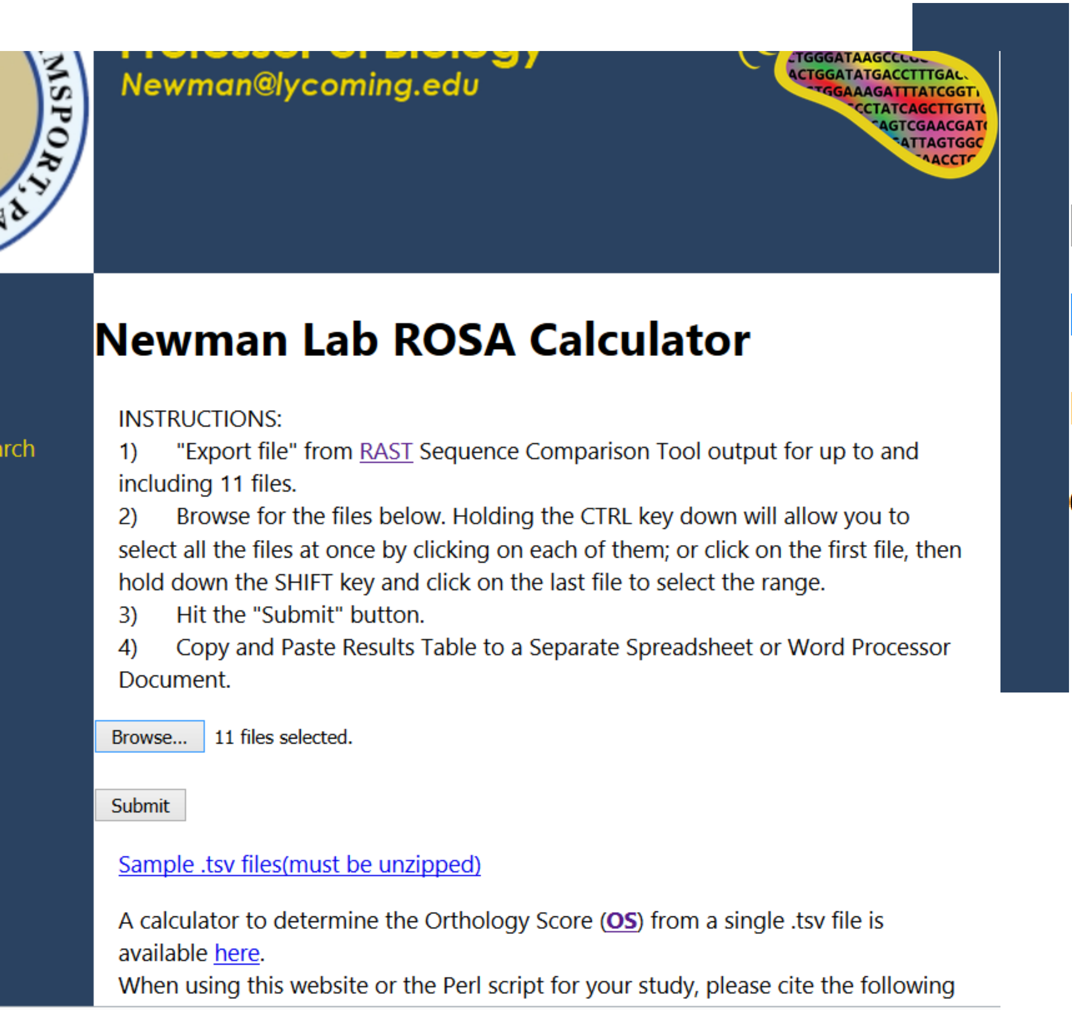
- ROSA takes into consideration both
  - similarity of orthologs at the protein level (AAI)
  - % of genome composed of orthologs (%BBH)
    - Differences in reference genome size create different %BBH values in reciprocal comparisons
- **Orthology Score (OS) = AAI^2 * %BBH**
- **ROSA = (OS$_{AB}$ + 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) calculated using tsv from up to 11 reciprocal sequence based comparisons





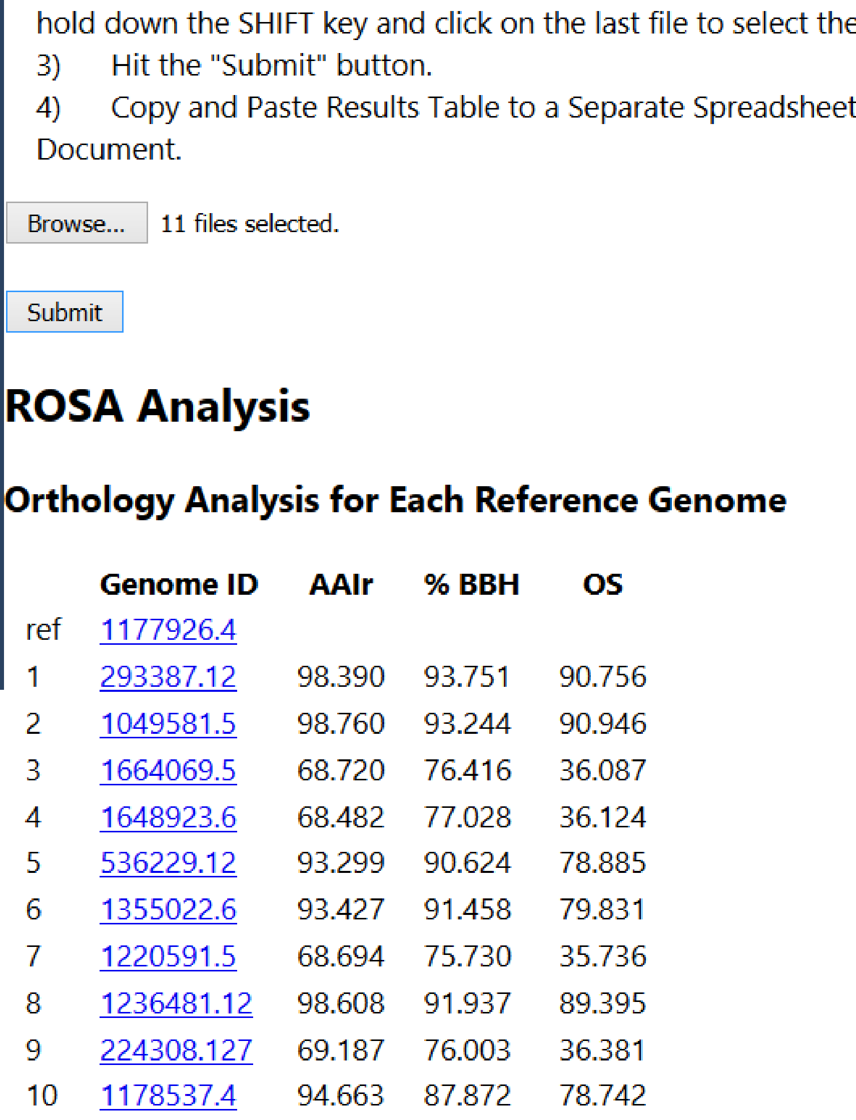
Newman@lycoming.edu

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.

Sample .tsv files(must be unzipped)

A calculator to determine the Orthology Score (OS) from a single .tsv file is available here.
When using this website or the Perl script for your study, please cite the following



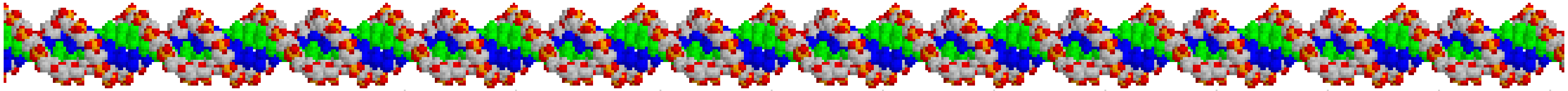
hold down the SHIFT key and click on the last file to select the
3) Hit the "Submit" button.
4) Copy and Paste Results Table to a Separate Spreadsheet
Document.

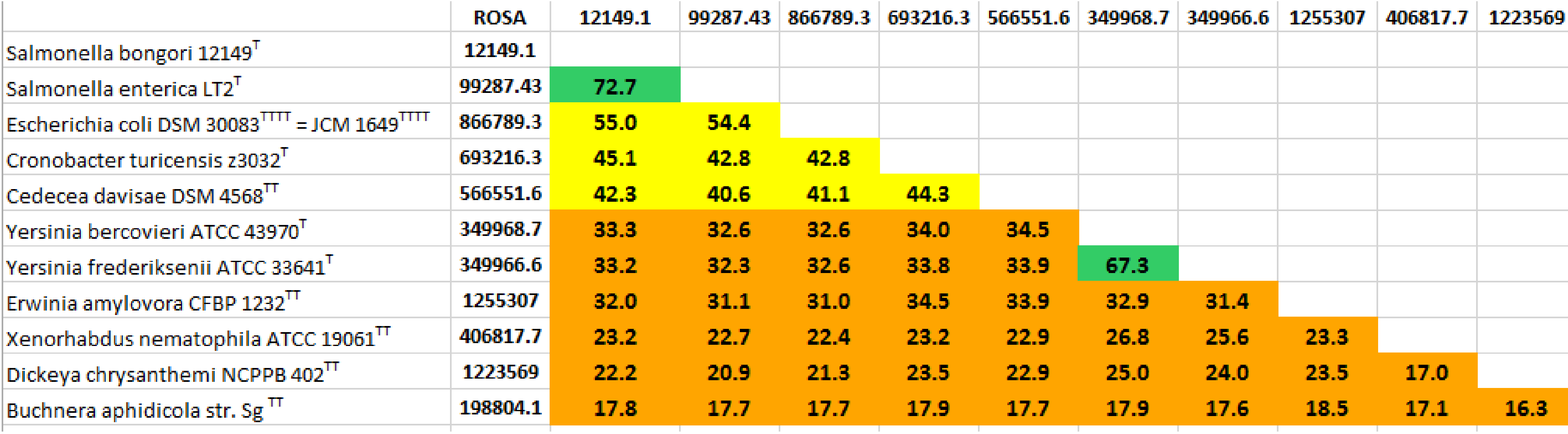
ROSA Analysis

Orthology Analysis for Each Reference Genome

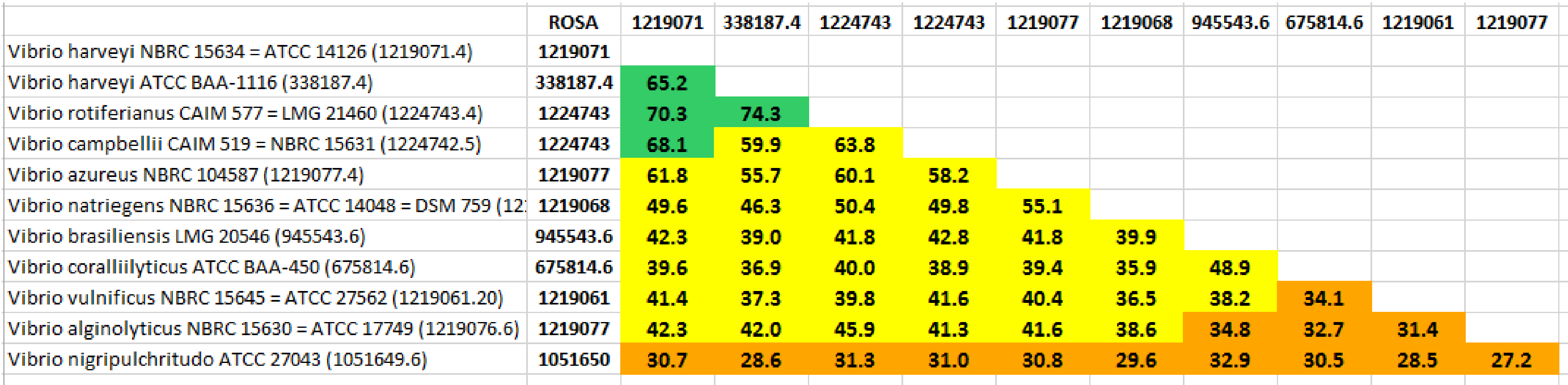
	Genome ID	AAIr	% BBH	OS
ref	1177926.4			
1	293387.12	98.390	93.751	90.756
2	1049581.5	98.760	93.244	90.946
3	1664069.5	68.720	76.416	36.087
4	1648923.6	68.482	77.028	36.124
5	536229.12	93.299	90.624	78.885
6	1355022.6	93.427	91.458	79.831
7	1220591.5	68.694	75.730	35.736
8	1236481.12	98.608	91.937	89.395
9	224308.127	69.187	76.003	36.381
10	1178537.4	94.663	87.872	78.742

Example ROSA tables



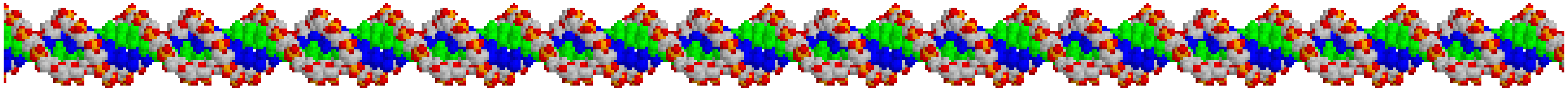


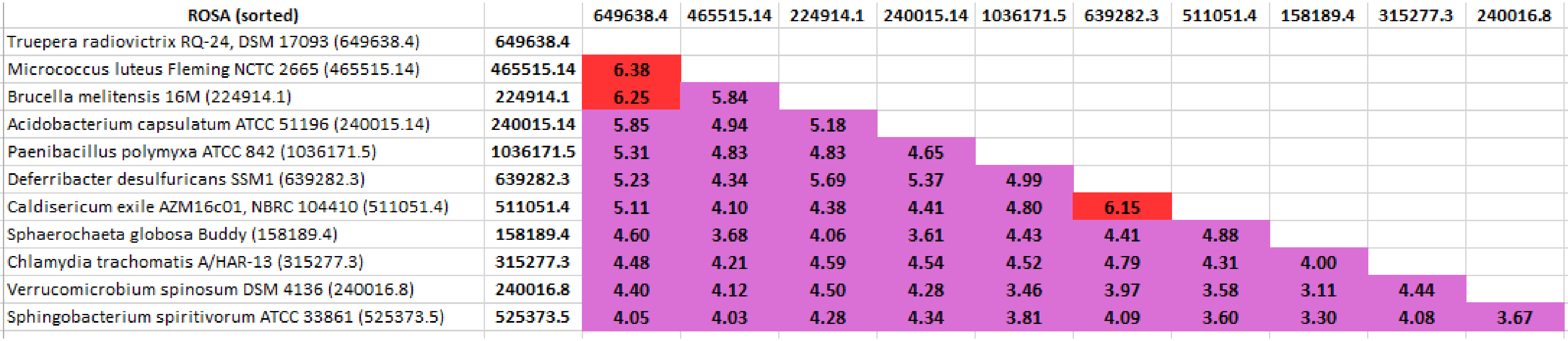
	ROSA	12149.1	99287.43	866789.3	693216.3	566551.6	349968.7	349966.6	1255307	406817.7	1223569
Salmonella bongori 12149^{T}	12149.1										
Salmonella enterica LT2^{T}	99287.43	72.7									
Escherichia coli DSM 30083^{TTTT} = JCM 1649^{TTTT}	866789.3	55.0	54.4								
Cronobacter turicensis z3032^{T}	693216.3	45.1	42.8	42.8							
Cedecea davisae DSM 4568^{TT}	566551.6	42.3	40.6	41.1	44.3						
Yersinia bercovieri ATCC 43970^{T}	349968.7	33.3	32.6	32.6	34.0	34.5					
Yersinia frederiksenii ATCC 33641^{T}	349966.6	33.2	32.3	32.6	33.8	33.9	67.3				
Erwinia amylovora CFBP 1232^{TT}	1255307	32.0	31.1	31.0	34.5	33.9	32.9	31.4			
Xenorhabdus nematophila ATCC 19061^{TT}	406817.7	23.2	22.7	22.4	23.2	22.9	26.8	25.6	23.3		
Dickeya chrysanthemi NCPPB 402^{TT}	1223569	22.2	20.9	21.3	23.5	22.9	25.0	24.0	23.5	17.0	
Buchnera aphidicola str. Sg ^{TT}	198804.1	17.8	17.7	17.7	17.9	17.7	17.9	17.6	18.5	17.1	16.3



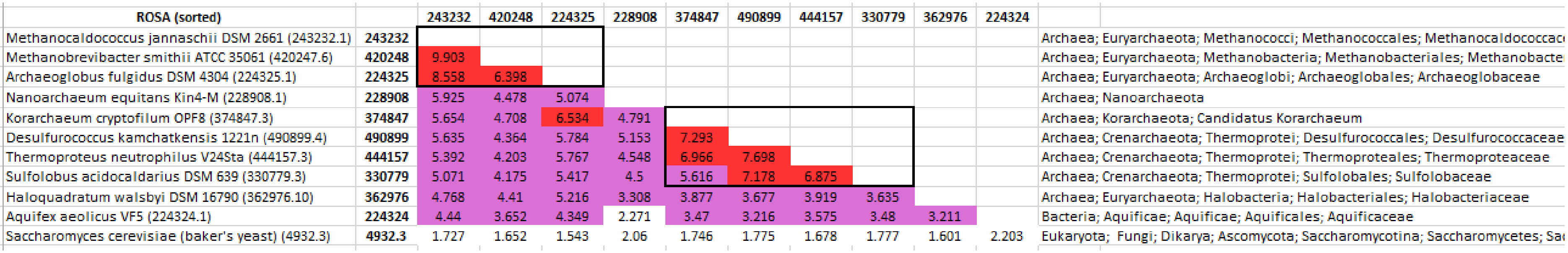
	ROSA	1219071	338187.4	1224743	1224743	1219077	1219068	945543.6	675814.6	1219061	1219077
Vibrio harveyi NBRC 15634 = ATCC 14126 (1219071.4)	1219071										
Vibrio harveyi ATCC BAA-1116 (338187.4)	338187.4	65.2									
Vibrio rotiferianus CAIM 577 = LMG 21460 (1224743.4)	1224743	70.3	74.3								
Vibrio campbellii CAIM 519 = NBRC 15631 (1224742.5)	1224743	68.1	59.9	63.8							
Vibrio azureus NBRC 104587 (1219077.4)	1219077	61.8	55.7	60.1	58.2						
Vibrio natriegens NBRC 15636 = ATCC 14048 = DSM 759 (12	1219068	49.6	46.3	50.4	49.8	55.1					
Vibrio brasiliensis LMG 20546 (945543.6)	945543.6	42.3	39.0	41.8	42.8	41.8	39.9				
Vibrio coralliilyticus ATCC BAA-450 (675814.6)	675814.6	39.6	36.9	40.0	38.9	39.4	35.9	48.9			
Vibrio vulnificus NBRC 15645 = ATCC 27562 (1219061.20)	1219061	41.4	37.3	39.8	41.6	40.4	36.5	38.2	34.1		
Vibrio alginolyticus NBRC 15630 = ATCC 17749 (1219076.6)	1219077	42.3	42.0	45.9	41.3	41.6	38.6	34.8	32.7	31.4	
Vibrio nigripulchritudo ATCC 27043 (1051649.6)	1051650	30.7	28.6	31.3	31.0	30.8	29.6	32.9	30.5	28.5	27.2

Example ROSA tables



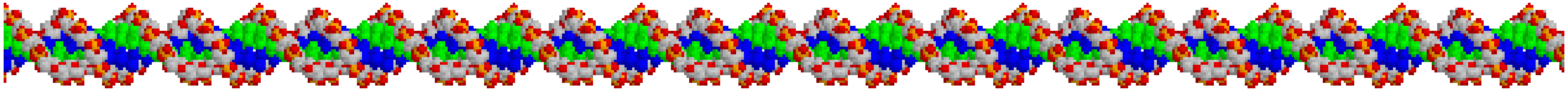


ROSA (sorted)		649638.4	465515.14	224914.1	240015.14	1036171.5	639282.3	511051.4	158189.4	315277.3	240016.8
Truepera radiovictrix RQ-24, DSM 17093 (649638.4)	**649638.4**										
Micrococcus luteus Fleming NCTC 2665 (465515.14)	**465515.14**	6.38									
Brucella melitensis 16M (224914.1)	**224914.1**	6.25	5.84								
Acidobacterium capsulatum ATCC 51196 (240015.14)	**240015.14**	5.85	4.94	5.18							
Paenibacillus polymyxa ATCC 842 (1036171.5)	**1036171.5**	5.31	4.83	4.83	4.65						
Deferribacter desulfuricans SSM1 (639282.3)	**639282.3**	5.23	4.34	5.69	5.37	4.99					
Caldisericum exile AZM16c01, NBRC 104410 (511051.4)	**511051.4**	5.11	4.10	4.38	4.41	4.80	6.15				
Sphaerochaeta globosa Buddy (158189.4)	**158189.4**	4.60	3.68	4.06	3.61	4.43	4.41	4.88			
Chlamydia trachomatis A/HAR-13 (315277.3)	**315277.3**	4.48	4.21	4.59	4.54	4.52	4.79	4.31	4.00		
Verrucomicrobium spinosum DSM 4136 (240016.8)	**240016.8**	4.40	4.12	4.50	4.28	3.46	3.97	3.58	3.11	4.44	
Sphingobacterium spiritivorum ATCC 33861 (525373.5)	**525373.5**	4.05	4.03	4.28	4.34	3.81	4.09	3.60	3.30	4.08	3.67



ROSA (sorted)		243232	420248	224325	228908	374847	490899	444157	330779	362976	224324	
Methanocaldococcus jannaschii DSM 2661 (243232.1)	**243232**											Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanocaldococcac
Methanobrevibacter smithii ATCC 35061 (420247.6)	**420248**	9.903										Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales; Methanobacte
Archaeoglobus fulgidus DSM 4304 (224325.1)	**224325**	8.558	6.398									Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales; Archaeoglobaceae
Nanoarchaeum equitans Kin4-M (228908.1)	**228908**	5.925	4.478	5.074								Archaea; Nanoarchaeota
Korarchaeum cryptofilum OPF8 (374847.3)	**374847**	5.654	4.708	6.534	4.791							Archaea; Korarchaeota; Candidatus Korarchaeum
Desulfurococcus kamchatkensis 1221n (490899.4)	**490899**	5.635	4.364	5.784	5.153	7.293						Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales; Desulfurococcaceae
Thermoproteus neutrophilus V24Sta (444157.3)	**444157**	5.392	4.203	5.767	4.548	6.966	7.698					Archaea; Crenarchaeota; Thermoprotei; Thermoproteales; Thermoproteaceae
Sulfolobus acidocaldarius DSM 639 (330779.3)	**330779**	5.071	4.175	5.417	4.5	5.616	7.178	6.875				Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae
Haloquadratum walsbyi DSM 16790 (362976.10)	**362976**	4.768	4.41	5.216	3.308	3.877	3.677	3.919	3.635			Archaea; Euryarchaeota; Halobacteria; Halobacteriales; Halobacteriaceae
Aquifex aeolicus VF5 (224324.1)	**224324**	4.44	3.652	4.349	2.271	3.47	3.216	3.575	3.48	3.211		Bacteria; Aquificae; Aquificae; Aquificales; Aquificaceae
Saccharomyces cerevisiae (baker's yeast) (4932.3)	**4932.3**	1.727	1.652	1.543	2.06	1.746	1.775	1.678	1.777	1.601	2.203	Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina; Saccharomycetes; Sa

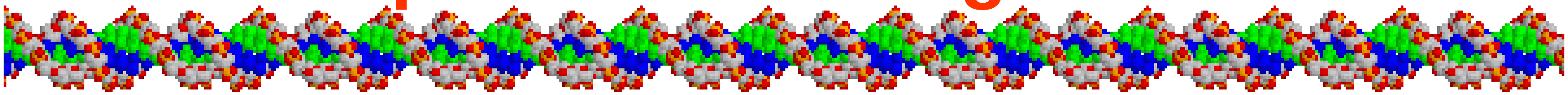
ROSA Thresholds

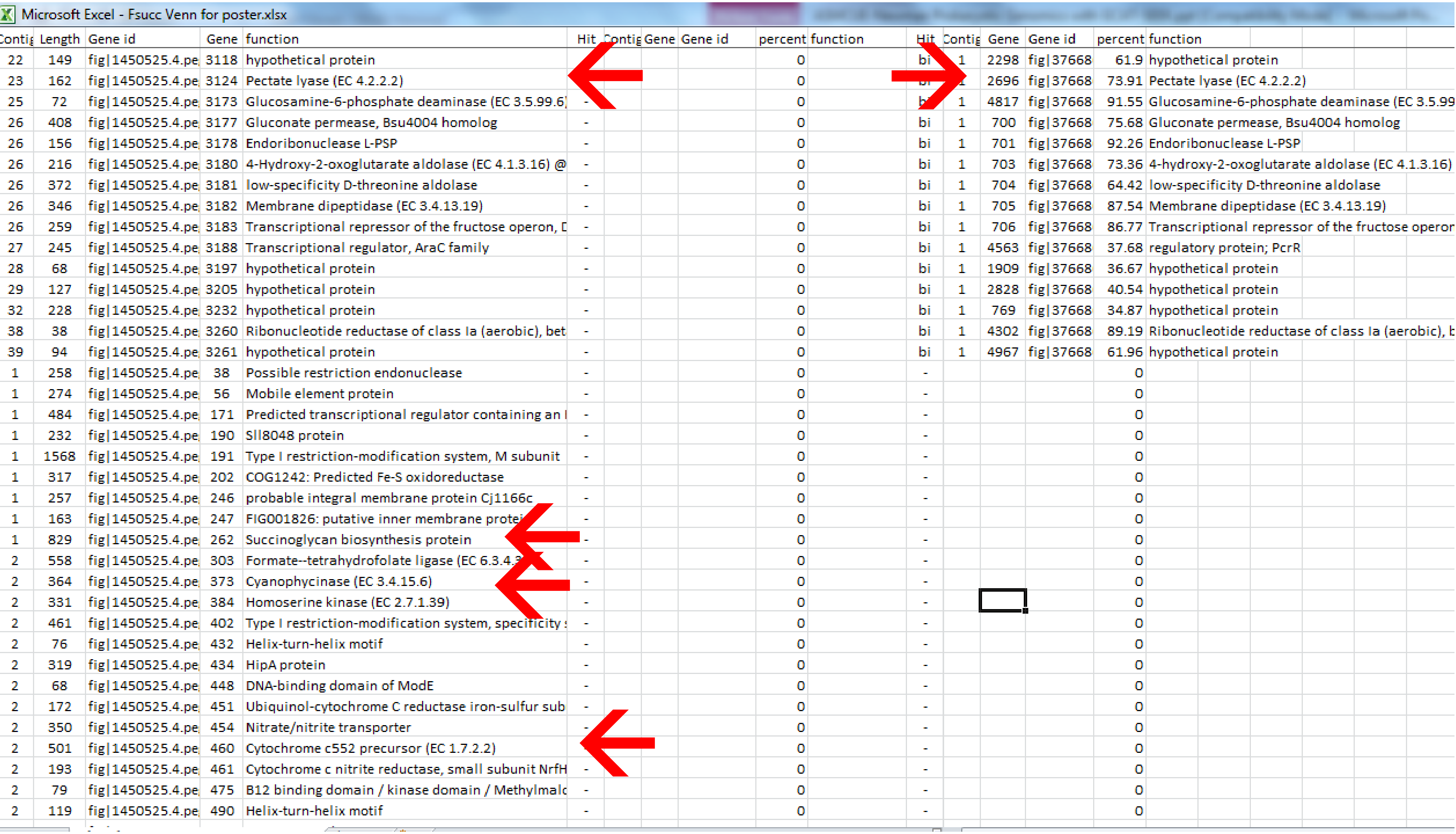


Level	Same	Different	expected range	min	max	mean	n=	below range	above range
8	species	strain	>65	49	99.59	85.98	312	3	N/A
7	genus	species	35-65	6.85	95.8	36.63	521	294	38
6	family	genus	15-35	5.8	54.99	18.86	524	125	19
5	order	family	10-15	4.75	23.19	11.58	314	79	22
4	class	order	8-10	4.15	14.99	8.22	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.

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

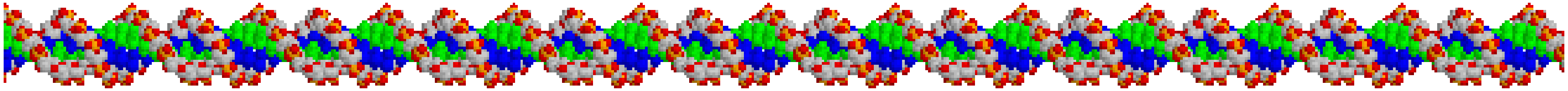


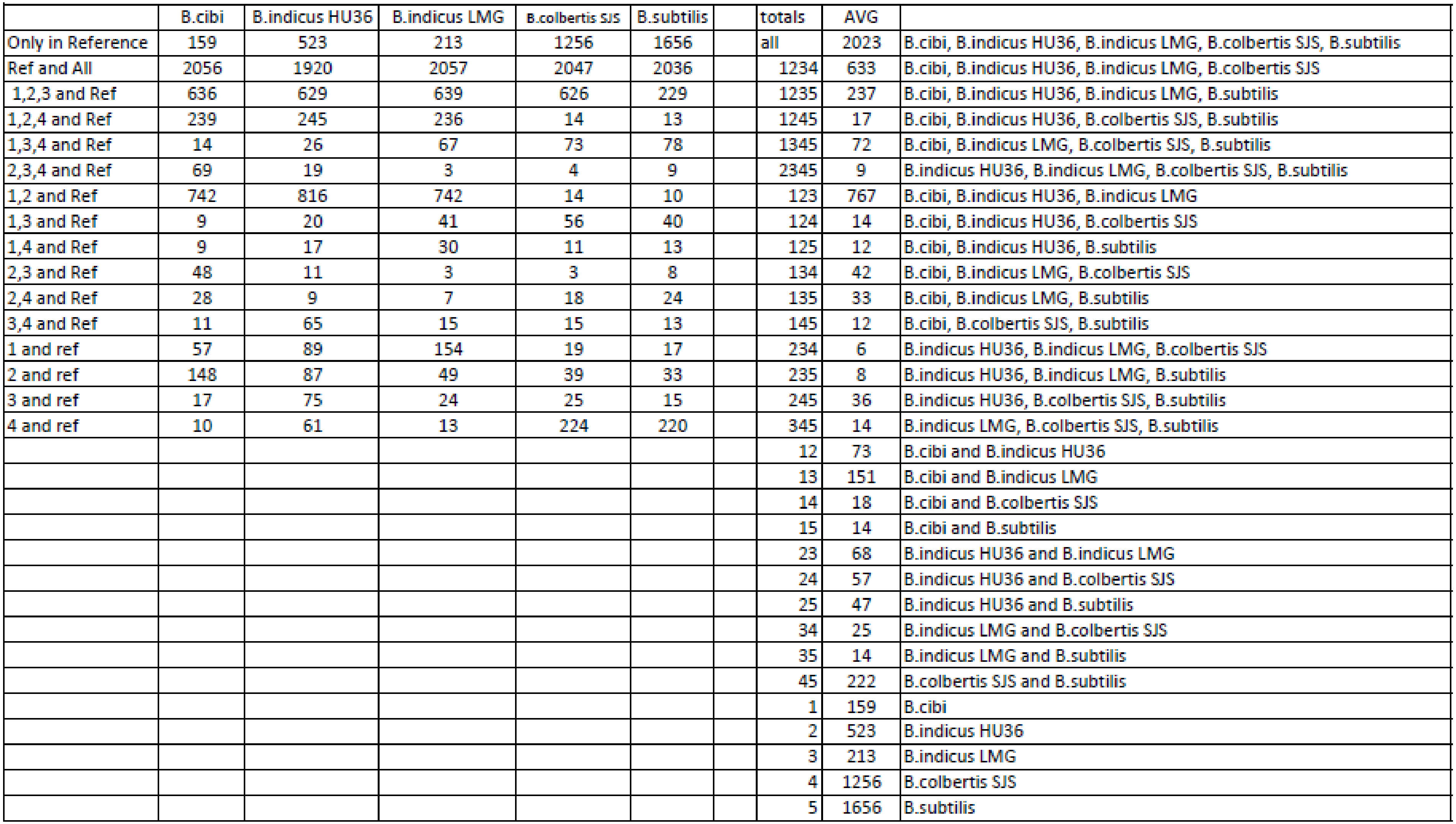


Microsoft Excel - Fsucc Venn for poster.xlsx

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	

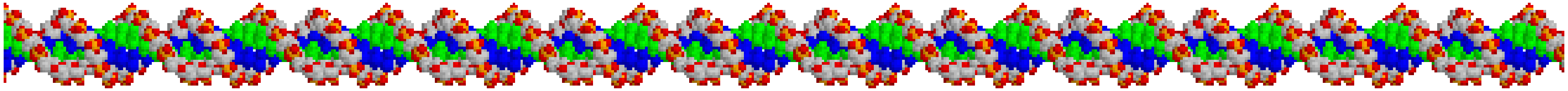
Venn Diagram Tool

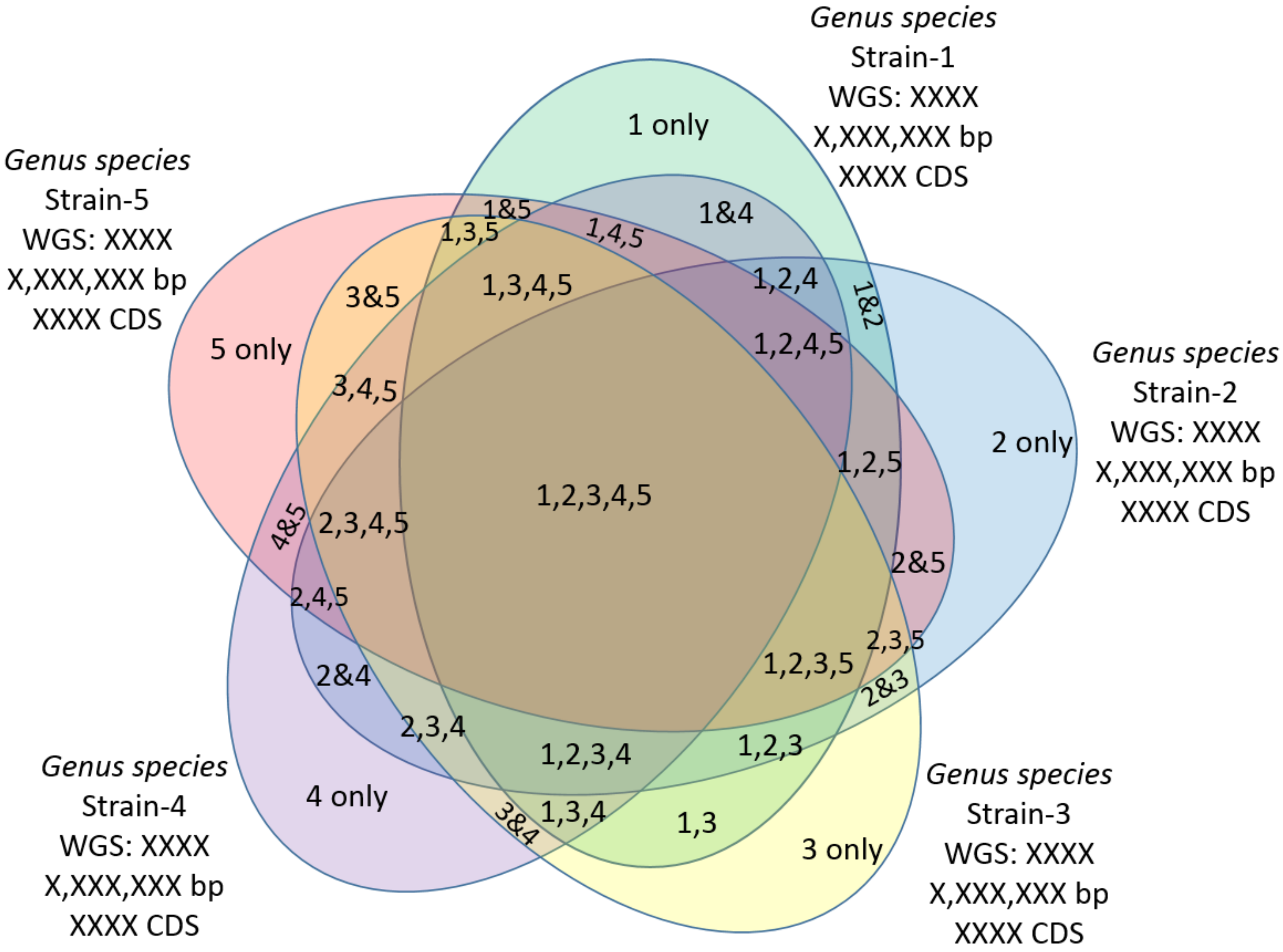




	B.cibi	B.indicus HU36	B.indicus LMG	B.colbertis SJS	B.subtilis		totals	AVG	
Only in Reference	159	523	213	1256	1656		all	2023	B.cibi, B.indicus HU36, B.indicus LMG, B.colbertis SJS, B.subtilis
Ref and All	2056	1920	2057	2047	2036		1234	633	B.cibi, B.indicus HU36, B.indicus LMG, B.colbertis SJS
1,2,3 and Ref	636	629	639	626	229		1235	237	B.cibi, B.indicus HU36, B.indicus LMG, B.subtilis
1,2,4 and Ref	239	245	236	14	13		1245	17	B.cibi, B.indicus HU36, B.colbertis SJS, B.subtilis
1,3,4 and Ref	14	26	67	73	78		1345	72	B.cibi, B.indicus LMG, B.colbertis SJS, B.subtilis
2,3,4 and Ref	69	19	3	4	9		2345	9	B.indicus HU36, B.indicus LMG, B.colbertis SJS, B.subtilis
1,2 and Ref	742	816	742	14	10		123	767	B.cibi, B.indicus HU36, B.indicus LMG
1,3 and Ref	9	20	41	56	40		124	14	B.cibi, B.indicus HU36, B.colbertis SJS
1,4 and Ref	9	17	30	11	13		125	12	B.cibi, B.indicus HU36, B.subtilis
2,3 and Ref	48	11	3	3	8		134	42	B.cibi, B.indicus LMG, B.colbertis SJS
2,4 and Ref	28	9	7	18	24		135	33	B.cibi, B.indicus LMG, B.subtilis
3,4 and Ref	11	65	15	15	13		145	12	B.cibi, B.colbertis SJS, B.subtilis
1 and ref	57	89	154	19	17		234	6	B.indicus HU36, B.indicus LMG, B.colbertis SJS
2 and ref	148	87	49	39	33		235	8	B.indicus HU36, B.indicus LMG, B.subtilis
3 and ref	17	75	24	25	15		245	36	B.indicus HU36, B.colbertis SJS, B.subtilis
4 and ref	10	61	13	224	220		345	14	B.indicus LMG, B.colbertis SJS, B.subtilis
							12	73	B.cibi and B.indicus HU36
							13	151	B.cibi and B.indicus LMG
							14	18	B.cibi and B.colbertis SJS
							15	14	B.cibi and B.subtilis
							23	68	B.indicus HU36 and B.indicus LMG
							24	57	B.indicus HU36 and B.colbertis SJS
							25	47	B.indicus HU36 and B.subtilis
							34	25	B.indicus LMG and B.colbertis SJS
							35	14	B.indicus LMG and B.subtilis
							45	222	B.colbertis SJS and B.subtilis
							1	159	B.cibi
							2	523	B.indicus HU36
							3	213	B.indicus LMG
							4	1256	B.colbertis SJS
							5	1656	B.subtilis

Venn Diagram Template





Genus species
Strain-1
WGS: XXXX
X,XXX,XXX bp
XXXX CDS

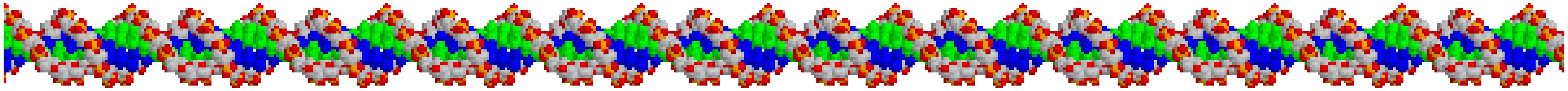
Genus species
Strain-5
WGS: XXXX
X,XXX,XXX bp
XXXX CDS

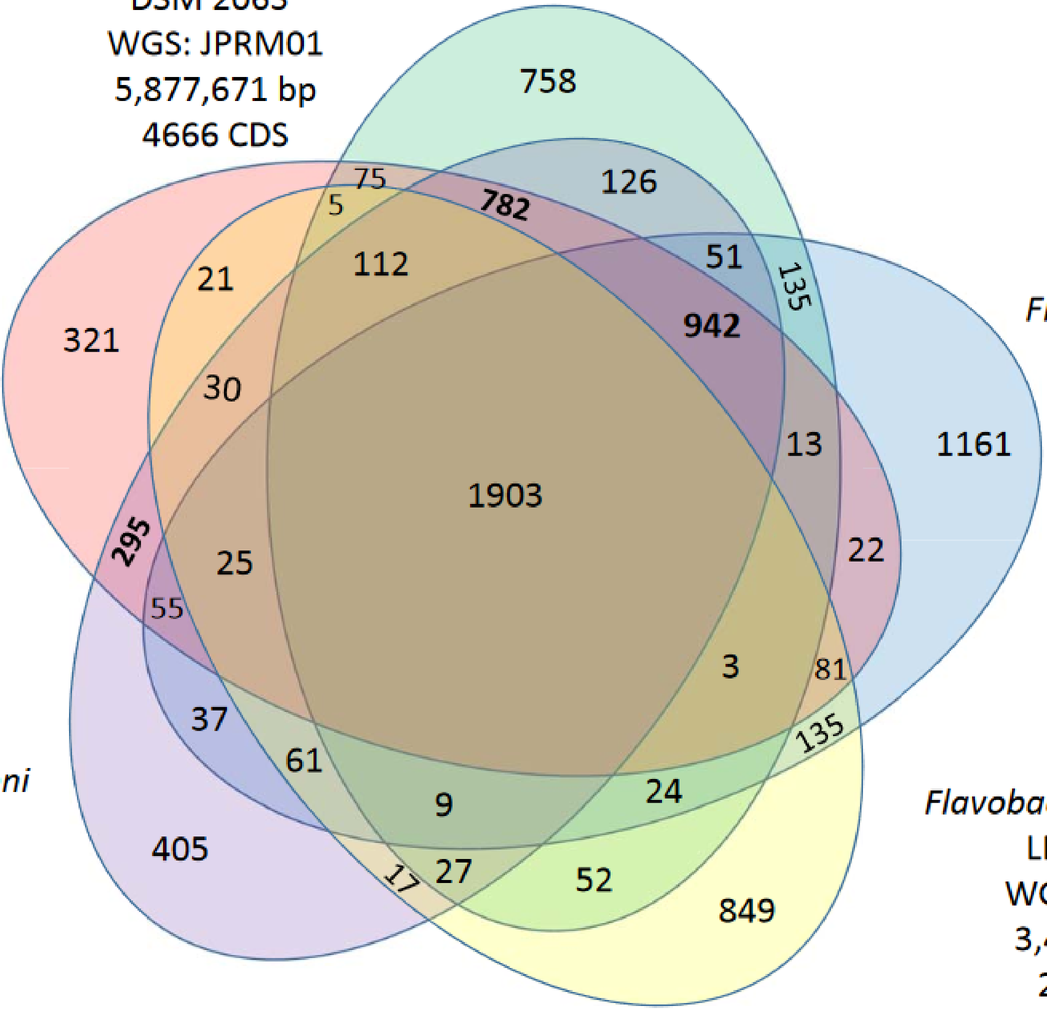
Genus species
Strain-2
WGS: XXXX
X,XXX,XXX bp
XXXX CDS

Genus species
Strain-4
WGS: XXXX
X,XXX,XXX bp
XXXX CDS

Genus species
Strain-3
WGS: XXXX
X,XXX,XXX bp
XXXX CDS

Identify Core, Genus or Family-Specific Genes





Flavobacterium hydatis
DSM 2063^T
WGS: JPRM01
5,877,671 bp
4666 CDS

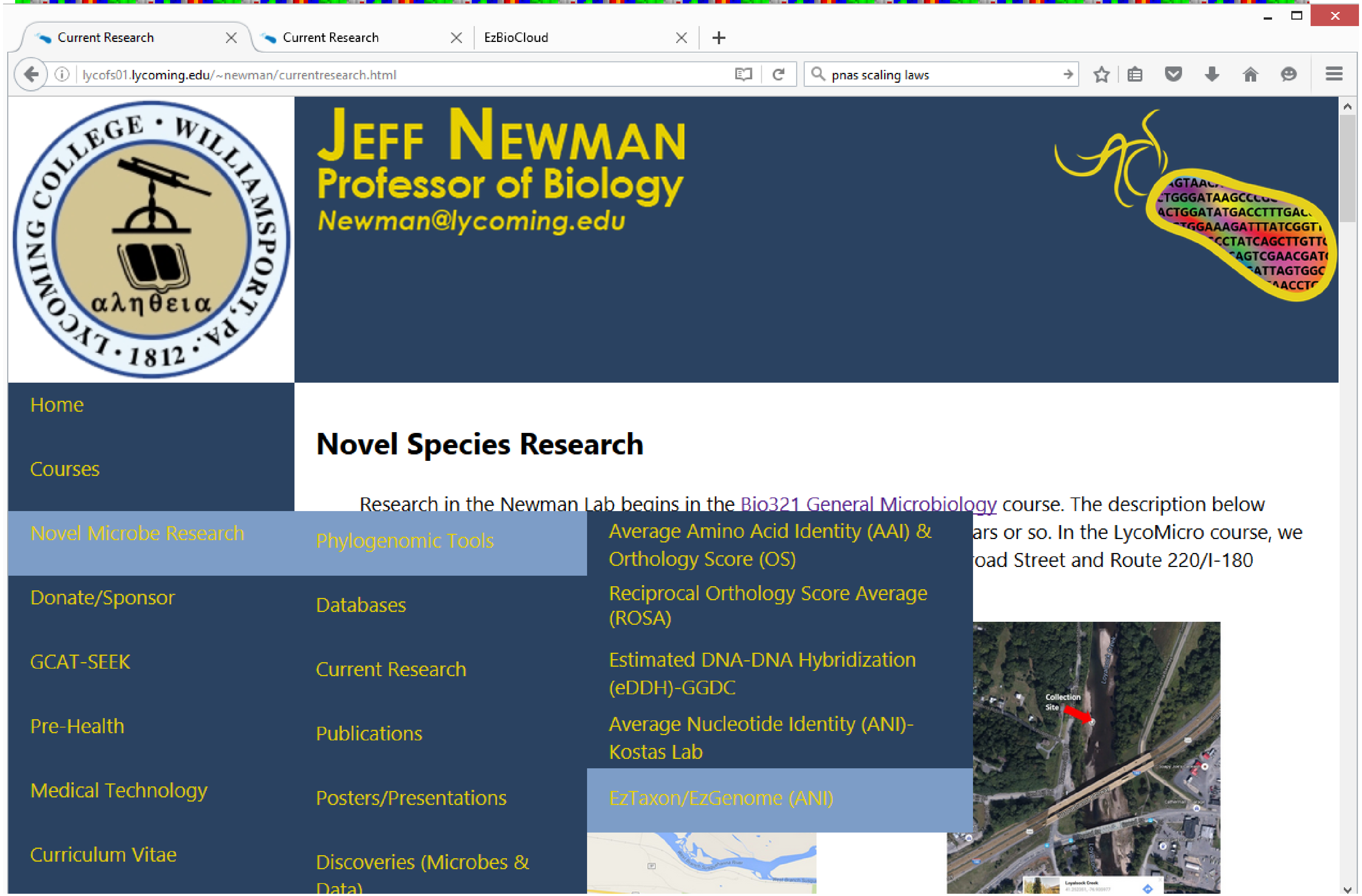
Flavobacterium falloni
JRM ^{T}
WGS: JSYO01
5,380,719 bp
4359 CDS

Flavobacterium hibernum
DSM 12611^T
WGS: JPRK01
5,283,672 bp
4288 CDS

Flavobacterium falloni
KMS
WGS: JSYP01
5,597,129 bp
4503 CDS

Flavobacterium aquatile
LMG 4008^T
WGS: JRHH01
3,490,856 bp
2996 CDS

Links/Tools available at novelmicrobe.com



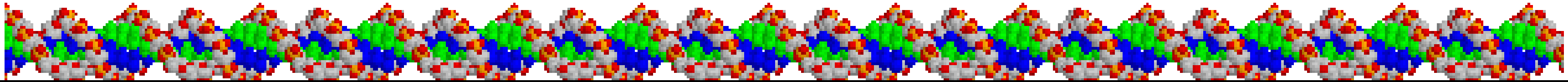
JEFF NEWMAN
Professor of Biology
Newman@lycoming.edu

- Home
- Courses
- Novel Microbe Research
  - Phylogenomic Tools
    - Average Amino Acid Identity (AAI) & Orthology Score (OS)
    - Reciprocal Orthology Score Average (ROSA)
    - Estimated DNA-DNA Hybridization (eDDH)-GGDC
    - Average Nucleotide Identity (ANI)-Kostas Lab
    - EzTaxon/EzGenome (ANI)
  - Databases
  - Current Research
  - Publications
  - Posters/Presentations
  - Discoveries (Microbes & Data)
- Donate/Sponsor
- GCAT-SEEK
- Pre-Health
- Medical Technology
- Curriculum Vitae

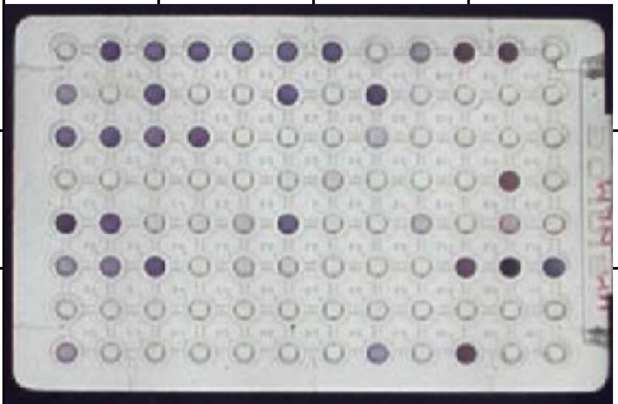
Novel Species Research

Research in the Newman Lab begins in the Bio321 General Microbiology course. The description below … ars or so. In the LycoMicro course, we … oad Street and Route 220/I-180

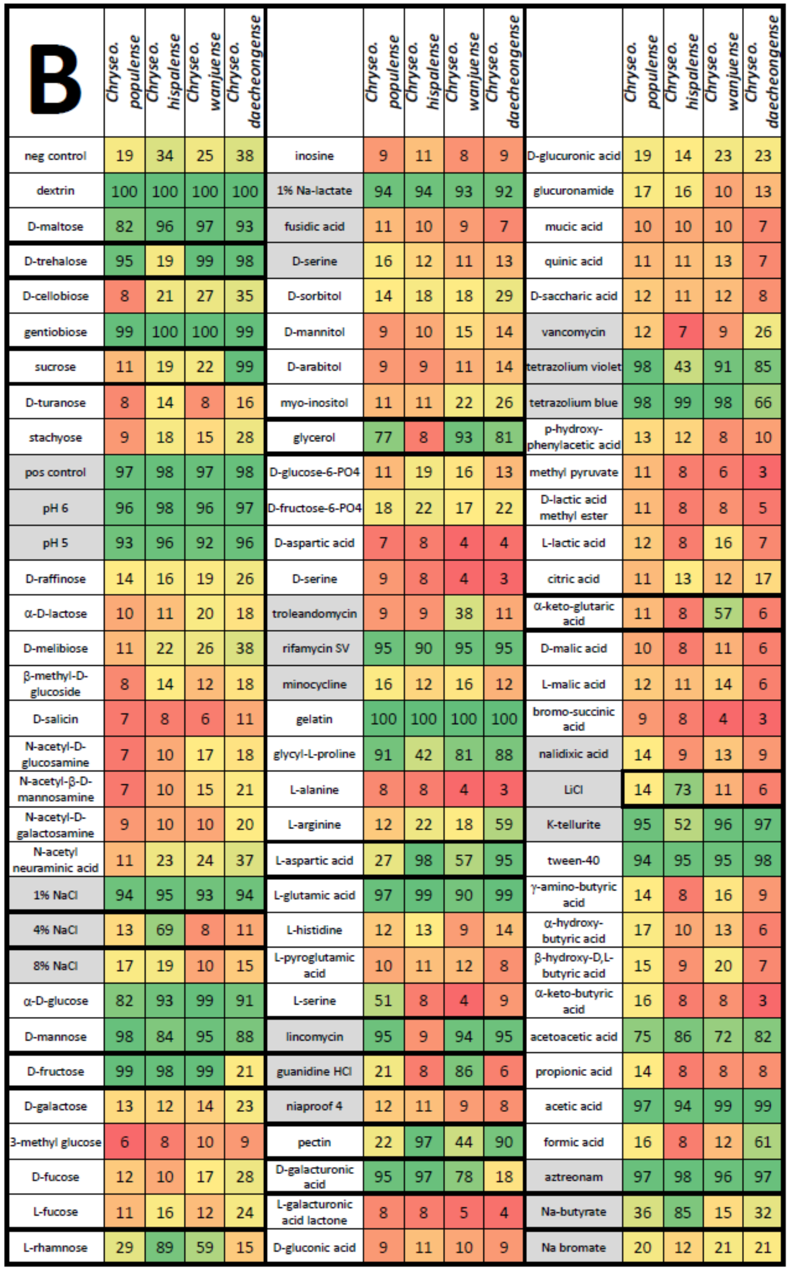
Biolog GenIII Plate



	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-glucosamine	N-acetyl-β-D-mannosamine	N-acetyl-D-galactosamine	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	Troleandomycin	rifamycin SV	minocycline
E	gelatin	glycyl-L-proline	L-alanine	L-arginine	L-aspartic acid	L-glutamic acid	L-histidine	L-pyroglutamic acid	L-serine	lincomycin	guanidine HCl	niaproof 4
F	pectin	D-galacturonic acid	L-galacturonic acid lactone					quinic acid	D-saccharic acid	vancomycin	tetrazolium violet	tetrazolium blue
G	p-hydroxyphenylacetic acid	methyl pyruvate	D-lactic acid methyl ester					L-malic acid	bromosuccinic acid	nalidixic acid	LiCl	K-tellurite
H	tween-40	γ-aminobutyric acid	α-hydroxybutyric acid					acetic acid	formic acid	aztreonam	Na-butyrate	Na bromate



Biolog Phenotype Comparisons



B	*Chryseo. populense*	*Chryseo. hispalense*	*Chryseo. wanjuense*	*Chryseo. daecheongense*		*Chryseo. populense*	*Chryseo. hispalense*	*Chryseo. wanjuense*	*Chryseo. daecheongense*		*Chryseo. populense*	*Chryseo. hispalense*	*Chryseo. wanjuense*	*Chryseo. daecheongense*
neg control	19	34	25	38	inosine	9	11	8	9	D-glucuronic acid	19	14	23	23
dextrin	100	100	100	100	1% Na-lactate	94	94	93	92	glucuronamide	17	16	10	13
D-maltose	82	96	97	93	fusidic acid	11	10	9	7	mucic acid	10	10	10	7
D-trehalose	95	19	99	98	D-serine	16	12	11	13	quinic acid	11	11	13	7
D-cellobiose	8	21	27	35	D-sorbitol	14	18	18	29	D-saccharic acid	12	11	12	8
gentiobiose	99	100	100	99	D-mannitol	9	10	15	14	vancomycin	12	7	9	26
sucrose	11	19	22	99	D-arabitol	9	9	11	14	tetrazolium violet	98	43	91	85
D-turanose	8	14	8	16	myo-inositol	11	11	22	26	tetrazolium blue	98	99	98	66
stachyose	9	18	15	28	glycerol	77	8	93	81	p-hydroxy-phenylacetic acid	13	12	8	10
pos control	97	98	97	98	D-glucose-6-PO4	11	19	16	13	methyl pyruvate	11	8	6	3
pH 6	96	98	96	97	D-fructose-6-PO4	18	22	17	22	D-lactic acid methyl ester	11	8	8	5
pH 5	93	96	92	96	D-aspartic acid	7	8	4	4	L-lactic acid	12	8	16	7
D-raffinose	14	16	19	26	D-serine	9	8	4	3	citric acid	11	13	12	17
α-D-lactose	10	11	20	18	troleandomycin	9	9	38	11	α-keto-glutaric acid	11	8	57	6
D-melibiose	11	22	26	38	rifamycin SV	95	90	95	95	D-malic acid	10	8	11	6
β-methyl-D-glucoside	8	14	12	18	minocycline	16	12	16	12	L-malic acid	12	11	14	6
D-salicin	7	8	6	11	gelatin	100	100	100	100	bromo-succinic acid	9	8	4	3
N-acetyl-D-glucosamine	7	10	17	18	glycyl-L-proline	91	42	81	88	nalidixic acid	14	9	13	9
N-acetyl-β-D-mannosamine	7	10	15	21	L-alanine	8	8	4	3	LiCl	14	73	11	6
N-acetyl-D-galactosamine	9	10	10	20	L-arginine	12	22	18	59	K-tellurite	95	52	96	97
N-acetyl neuraminic acid	11	23	24	37	L-aspartic acid	27	98	57	95	tween-40	94	95	95	98
1% NaCl	94	95	93	94	L-glutamic acid	97	99	90	99	γ-amino-butyric acid	14	8	16	9
4% NaCl	13	69	8	11	L-histidine	12	13	9	14	α-hydroxy-butyric acid	17	10	13	6
8% NaCl	17	19	10	15	L-pyroglutamic acid	10	11	12	8	β-hydroxy-D,L-butyric acid	15	9	20	7
α-D-glucose	82	93	99	91	L-serine	51	8	4	9	α-keto-butyric acid	16	8	8	3
D-mannose	98	84	95	88	lincomycin	95	9	94	95	acetoacetic acid	75	86	72	82
D-fructose	99	98	99	21	guanidine HCl	21	8	86	6	propionic acid	14	8	8	8
D-galactose	13	12	14	23	niaproof 4	12	11	9	8	acetic acid	97	94	99	99
3-methyl glucose	6	8	10	9	pectin	22	97	44	90	formic acid	16	8	12	61
D-fucose	12	10	17	28	D-galacturonic acid	95	97	78	18	aztreonam	97	98	96	97
L-fucose	11	16	12	24	L-galacturonic acid lactone	8	8	5	4	Na-butyrate	36	85	15	32
L-rhamnose	29	89	59	15	D-gluconic acid	9	11	10	9	Na bromate	20	12	21	21

C

Chryseobacterium populense **RAST subsystems**

Aminosugars (7)
- N-Acetyl-Galactosamine and Galactosamine Utilization (7) ←

Di- and oligosaccharides (29)
- Maltose and Maltodextrin Utilization (13)
- Trehalose Uptake and Utilization (6) ←
- Lactose and Galactose Uptake and Utilization (8)
- Lactose utilization (2)

Glycoside hydrolases (0)

One-carbon Metabolism (40)
- Serine-glyoxylate cycle (37)
- One-carbon metabolism by tetrahydropterines (3)

Organic acids (5)
- Glycerate metabolism (4)
- Lactate utilization (1)

Fermentation (34)
- Butanol Biosynthesis (14)
- Acetolactate synthase subunits (2)
- Acetyl-CoA fermentation to Butyrate (15)
- Acetoin, butanediol metabolism (3) ←

CO2 fixation (0)

Sugar alcohols (8)
- Glycerol and Glycerol-3-phosphate Uptake and Utilization (8) ←

Carbohydrates - no subcategory (0)

Polysaccharides (18)
- Glycogen metabolism (5)
- Cellulosome (13)

Monosaccharides (66)
- Mannose Metabolism (11)
- D-ribose utilization (3)
- Xylose utilization (12)
- Deoxyribose and Deoxynucleoside Catabolism (7)
- L-Arabinose utilization (4)
- D-Galacturonate and D-Glucuronate Utilization (29)

Chryseobacterium hispalense **RAST subsystems**

Aminosugars (0)

Di- and oligosaccharides (21)
- Maltose and Maltodextrin Utilization (8)
- Lactose and Galactose Uptake and Utilization (9)
- Lactose utilization (4)

Glycoside hydrolases (0)

One-carbon Metabolism (38)
- Serine-glyoxylate cycle (34)
- One-carbon metabolism by tetrahydropterines (4)

Organic acids (1)
- Lactate utilization (1)

Fermentation (31)
- Butanol Biosynthesis (16)
- Acetyl-CoA fermentation to Butyrate (15)

CO2 fixation (1)
- CO2 uptake, carboxysome (1)

Sugar alcohols (0)

Carbohydrates - no subcategory (0)

Polysaccharides (19)
- Glycogen metabolism (5)
- Cellulosome (14)

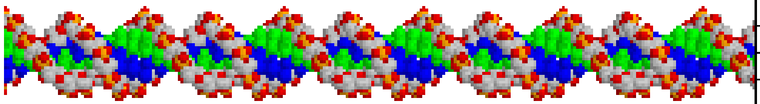
Monosaccharides (62)
- Mannose Metabolism (8)
- D-ribose utilization (3)
- Xylose utilization (10)
- Deoxyribose and Deoxynucleoside Catabolism (7)
- L-Arabinose utilization (4)
- D-Galacturonate and D-Glucuronate Utilization (30)

RAST Rapid Annotation using Subsystem Technology

The SEED Viewer

SEEKquence

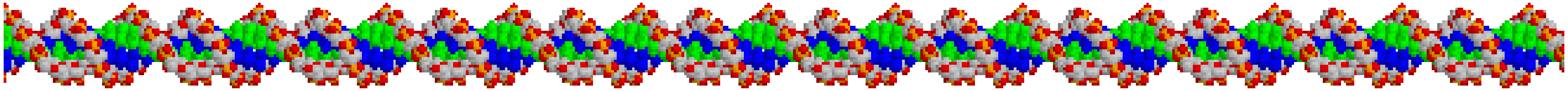
Biolog Phenotypes



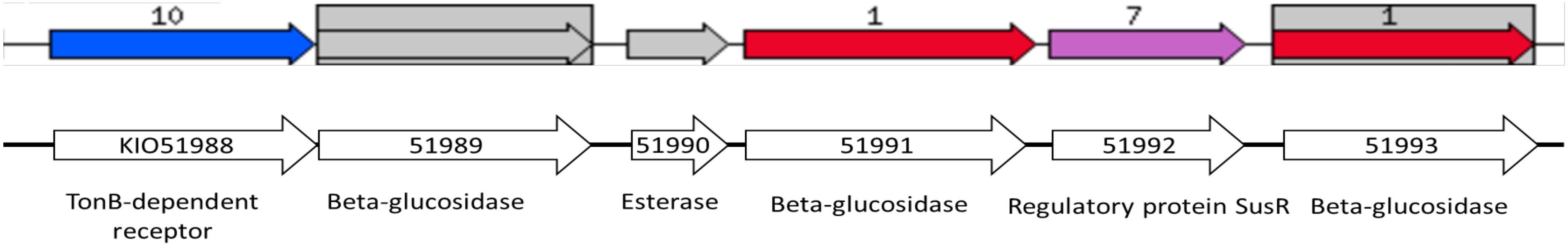
Biolog GenIII plates used to compare novel species candidates to closest relatives to identify phenotypic differences.

well	carbon source or condition	Flavobacterium gabrieli KJJ	Flavobacterium chiliense	Flavobacterium chungangense	Flavobacterium hibernum	well	carbon source or condition	Flavobacterium gabrieli KJJ	Flavobacterium chiliense	Flavobacterium chungangense	Flavobacterium hibernum
A01	neg control	21	24	24	23	E01	gelatin	92	99	93	98
A02	dextrin	99	98	99	98	E02	glycyl-L-proline	53	95	95	91
A03	D-maltose	90	98	99	98	E03	L-alanine	44	83	7	33
A04	D-trehalose	12	98	10	97	E04	L-arginine	44	87	22	64
A05	D-cellobiose	57	98	98	99	E05	L-aspartic acid	44	97	91	93
A06	gentiobiose	99	98	99	97	E06	L-glutamic acid	70	97	97	95
A07	sucrose	14	11	97	98	E07	L-histidine	30	71	23	23
A08	D-turanose	14	12	20	9	E08	L-pyroglutamic acid	10	9	13	14
A09	stachyose	14	13	11	32	E09	L-serine	44	89	74	73
A10	pos control	98	97	98	98	E10	lincomycin	44	11	10	10
A11	pH 6	97	96	96	98	E11	guanidine HCl	40	18	9	39
A12	pH 5	27	67	12	16	E12	niaproof 4	16	14	14	12
B01	D-raffinose	17	14	14	96	F01	pectin	28	45	97	71
B02	α-D-lactose	16	21	19	22	F02	D-galacturonic acid	56	98	98	96
B03	D-melibiose	17	13	14	10	F03	L-galacturonic acid lactone	12	8	9	53
B04	β-methyl-D-glucoside	14	17	98	97	F04	D-gluconic acid	10	14	27	8
B05	D-salicin	10	98	98	98	F05	D-glucuronic acid	18	47	97	51
B06	N-acetyl-D-glucosamine	99	97	6	96	F06	glucuronamide	17	21	33	18
B07	N-acetyl-β-D-mannosamine	77	16	10	12	F07	mucic acid	10	10	15	7
B08	N-acetyl-D-galactosamine	48	97	36	95	F08	quinic acid	10	12	21	16
B09	N-acetyl neuraminic acid	11	97	7	95	F09	D-saccharic acid	10	10	14	8
B10	1% NaCl	55	92	94	63	F10	vancomycin	63	94	95	94
B11	4% NaCl	14	13	13	12	F11	tetrazolium violet	90	99	45	70
B12	8% NaCl	19	16	16	11	F12	tetrazolium blue	99	99	95	96
C01	α-D-glucose	99	98	98	97	G01	p-hydroxy-phenylacetic acid	15	12	9	11
C02	D-mannose	99	97	98	97	G02	methyl pyruvate	14	71	12	85
C03	D-fructose	44	89	34	94	G03	D-lactic acid methyl ester	13	14	27	17
C04	D-galactose	99	97	51	97	G04	L-lactic acid	10	13	19	11
C05	3-methyl glucose	10	13	13	8	G05	citric acid	16	14	21	15
C06	D-fucose	12	16	6	8	G06	α-keto-glutaric acid	11	10	18	10
C07	L-fucose	16	20	68	87	G07	D-malic acid	13	11	16	12
C08	L-rhamnose	46	96	98	93	G08	L-malic acid	46	13	15	15
C09	inosine	11	8	7	8	G09	bromo-succinic acid	10	8	6	7
C10	1% Na-lactate	91	93	96	94	G10	nalidixic acid	16	13	12	13
C11	fusidic acid	11	10	9	9	G11	LiCl	11	11	9	11
C12	D-serine	16	15	10	12	G12	K-tellurite	21	18	16	18
D01	D-sorbitol	17	15	13	11	H01	tween-40	50	97	96	85
D02	D-mannitol	12	11	18	9	H02	γ-amino-butyric acid	16	12	12	16
D03	D-arabitol	10	12	11	7	H03	α-hydroxy-butyric acid	16	11	13	12
D04	myo-inositol	11	10	15	7	H04	β-hydroxy-D,L-butyric acid	13	11	19	13
D05	glycerol	44	8	8	8	H05	α-keto-butyric acid	12	8	7	7
D06	D-glucose-6-PO4	36	16	27	28	H06	acetoacetic acid	30	31	41	47
D07	D-fructose-6-PO4	27	21	90	34	H07	propionic acid	12	10	6	7
D08	D-aspartic acid	7	7	6	8	H08	acetic acid	51	91	89	60
D09	D-serine	7	7	6	6	H09	formic acid	16	11	8	9
D10	troleandomycin	12	10	9	9	H10	aztreonam	96	96	98	97
D11	rifamycin SV	96	94	96	96	H11	Na-butyrate	16	14	15	13
D12	minocycline	16	13	12	13	H12	Na bromate	15	14	16	12

β-glucosidase Paralogs



Beta-glucosidase cluster from *Flavobacterium hibernum*



Ortholog in
F. chungangense,
F. sp. CRH

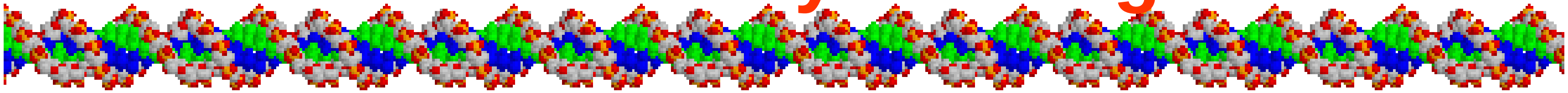
Correlates with
β-methyl-D-
glucoside
utilization

Ortholog in
F. chungangense,
F. chilense,
F. sp. AED,
F. sp. CRH,
F. aquidurense,
F. frigoris

Correlates with
Salicin utilization

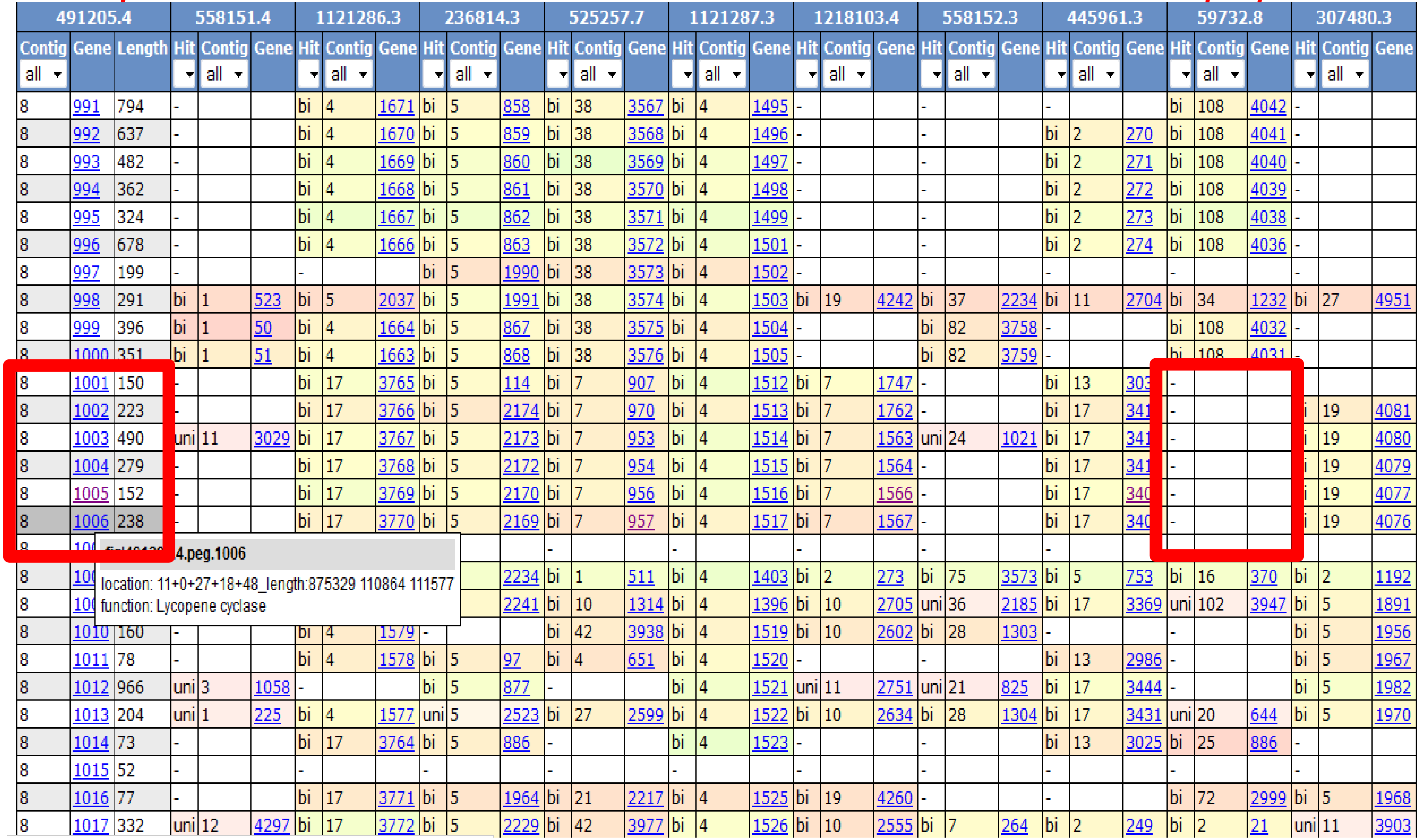
Ortholog in
F. chungangense,
F. chilense,
F. sp. CRH,
F. sp. KJJ,
F. hydatis
F. aquidurense,
F. johnsoniae,
F. reichenbachii,
F. denitrificans….

Chryseo. populense sp. nov lacks carotenoid biosynthetic genes



C.hispalense

C.populense



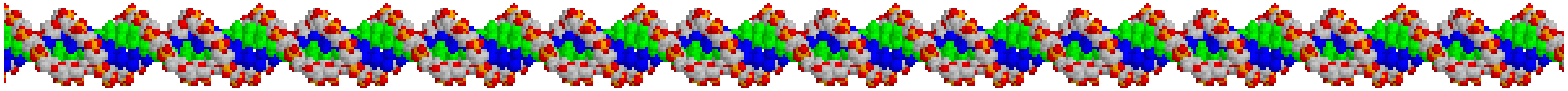
491205.4			558151.4			1121286.3			236814.3			525257.7			1121287.3			1218103.4			558152.3			445961.3			59732.8			307480.3		
Contig all ▾	Gene	Length	Hit ▾	Contig all ▾	Gene	Hit ▾	Contig all ▾	Gene	Hit ▾	Contig all ▾	Gene	Hit ▾	Contig all ▾	Gene	Hit ▾	Contig all ▾	Gene	Hit ▾	Contig all ▾	Gene	Hit ▾	Contig all ▾	Gene	Hit ▾	Contig all ▾	Gene	Hit ▾	Contig all ▾	Gene	Hit ▾	Contig all ▾	Gene
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	-			i	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	-			i	19	4080
8	1004	279	-			bi	17	3768	bi	5	2172	bi	7	954	bi	4	1515	bi	7	1564	-			bi	17	341	-			i	19	4079
8	1005	152	-			bi	17	3769	bi	5	2170	bi	7	956	bi	4	1516	bi	7	1566	-			bi	17	340	-			i	19	4077
8	1006	238	-			bi	17	3770	bi	5	2169	bi	7	957	bi	4	1517	bi	7	1567	-			bi	17	340	-			i	19	4076
8	10											-			-			-			-			-			-			-		
8	100										2234	bi	1	511	bi	4	1403	bi	2	273	bi	75	3573	bi	5	753	bi	16	370	bi	2	1192
8	100										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

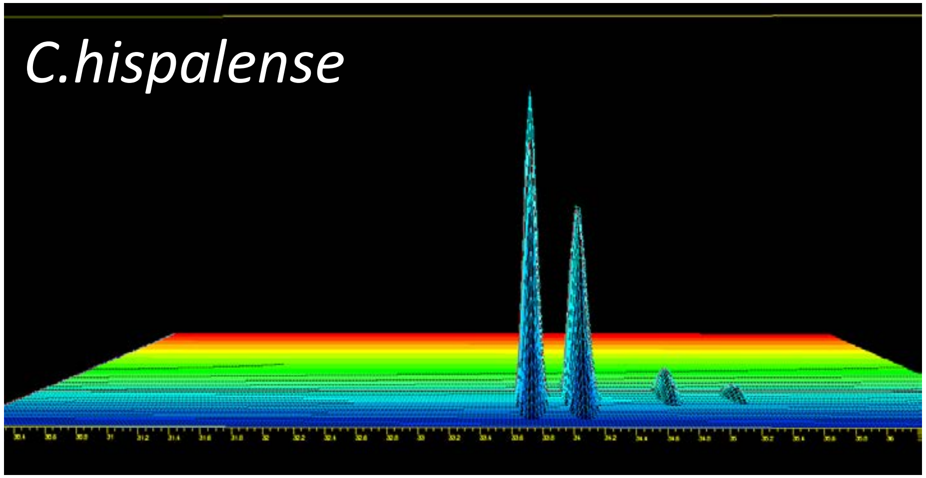
4.peg.1006

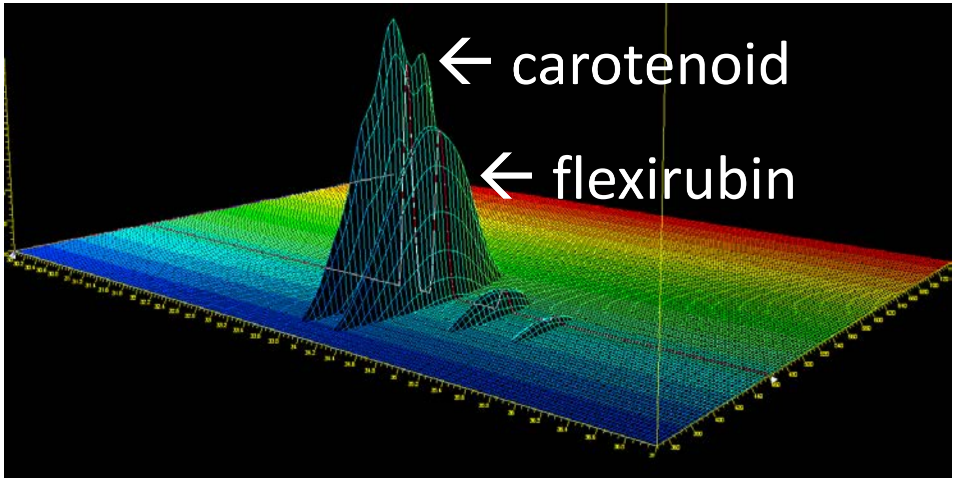
location: 11+0+27+18+48_length:875329 110864 111577

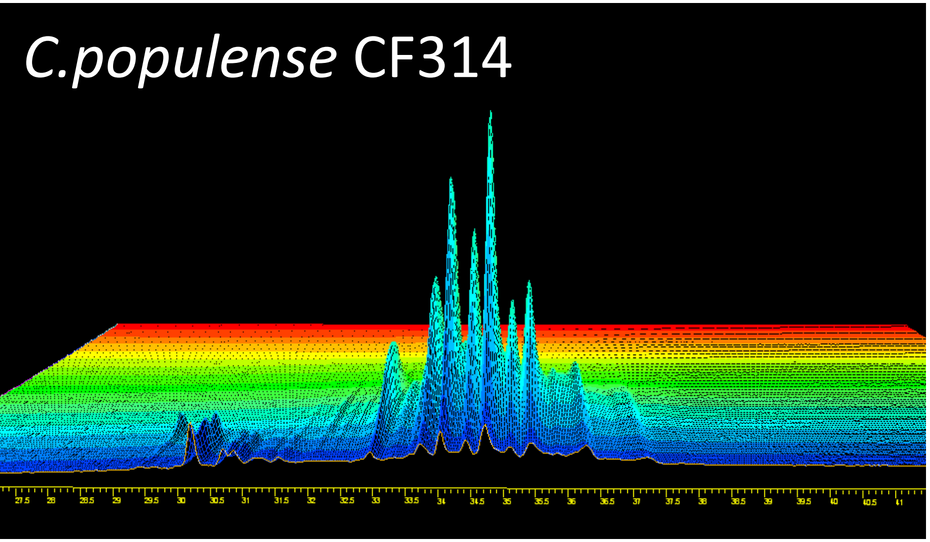
function: Lycopene cyclase

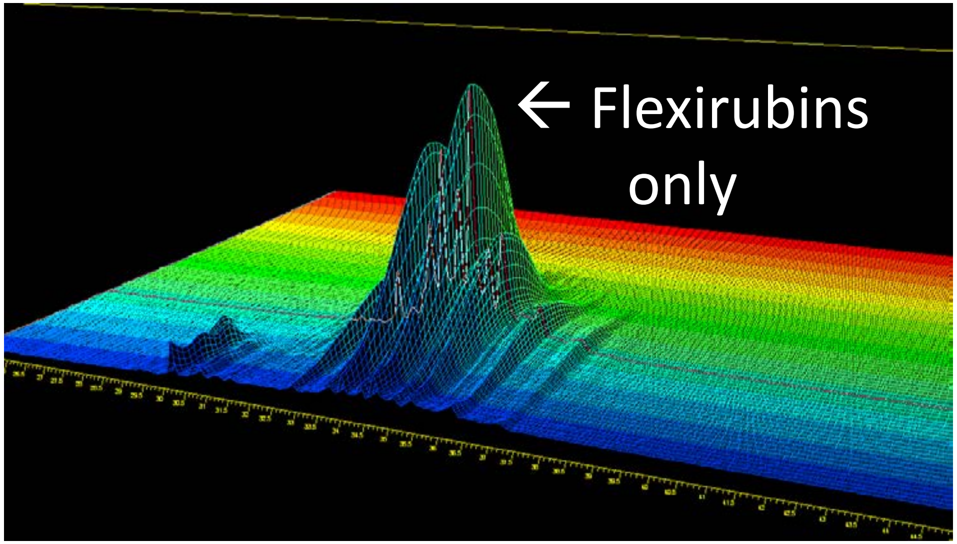
Explain phenotypic differences – e.g. Pigment "Landscapes"



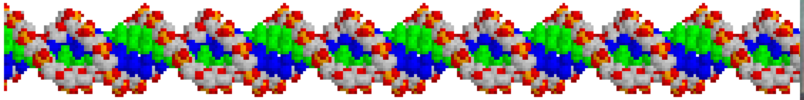








Thank You!



HHMI

NSF

SEEKquence