Microbial Genome Sequencing with GCAT-SEEK

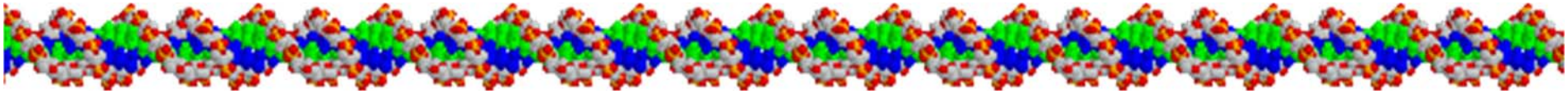
Jeff Newman – Lycoming College

ASMCUE May 15, 2014

http://www.lycoming.edu/~newman/ASMCUE2014-Newman.ppt

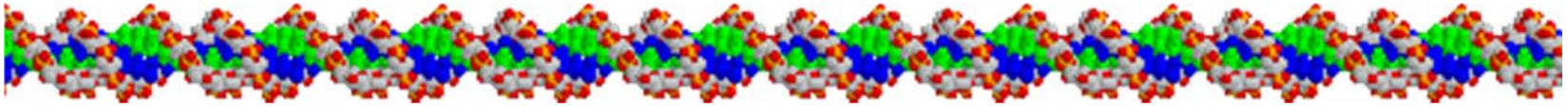
Human Genome 10th Anniversary

Overview



- GCAT→ GCAT-SEEK History, Workshops
- NextGen Sequencing Technologies
- Filtering and Assembly of Raw Data
- Automated Annotation with RAST
- Phenotypic Predictions
- Comparative Genomics
- Phylogenomic Analysis
- Use in the Classroom/Lab
- What's Next?
- **Take Home Message – Microbial Genome sequencing is surprisingly accessible, fast, easy, inexpensive and powerful.**

HHMI GCAT→ GCAT-SEEK



- Genome Consortium for Active Teaching (GCAT) founded by Malcolm Campbell (Davidson) in 2000 to bring Genomics (Microarrays) to the undergraduate curriculum.
  Multiple HHMI & NSF funded workshops
- ~2010 – Malcolm shifted to Synthetic Biology
  → GCAT-SynBio

Genome Consortium for Active Teaching

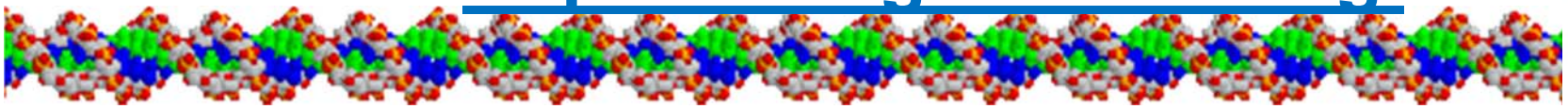
- Mike Boyle (Juniata) received NSF RCN-UBE pilot funding for UG NextGen seq network
  - Goal is to facilitate NextGen sequencing of UG faculty samples – students conduct course-based research.
  - Penn State Genomics Core facility.
  - Malcolm presented & approved use of GCAT "brand" → GCAT-SEEK
- Vince Buonaccorsi (Juniata) is PI of full NSF RCN-UBE for GCAT-SEEK.
  Also part of Juniata's HHMI Genomics Leadership Initiative

SEEKquence

HHMI

GCAT-SEEK Workshops

http://www.gcat-seek.org/



NSF

Summer Faculty Development Workshops – Designed for Beginners

- 2013 – Juniata College (Huntingdon, PA)
- 2014 – Lycoming College (Williamsport, PA) (June 2-6)
- 2015 – Juniata & Morgan State University (Baltimore, MD)
- 2016 – Juniata & California State Univ., Los Angeles
- 2017 – Hampton University (Hampton, VA)

Workshop Features Application-specific "Breakout Sessions" for sample prep & data analysis. Pedagogy & assessment sessions are combined.

- Pedagogy - Nancy Trun (Duquesne Univ.)
- Assessment – Tammy Tobin (Susquehanna Univ.)
- Prokaryotic Genomics – Jeff Newman (Lycoming)
- Metagenomics – Gina Lamendella (Juniata)
- Eukaryotic Genomics – Vince Buonaccorsi (Juniata)
- RNASeq – Mark Peterson (Juniata) & Arthur Hunt (Univ. of Kentucky)

How do we know microbes exist?

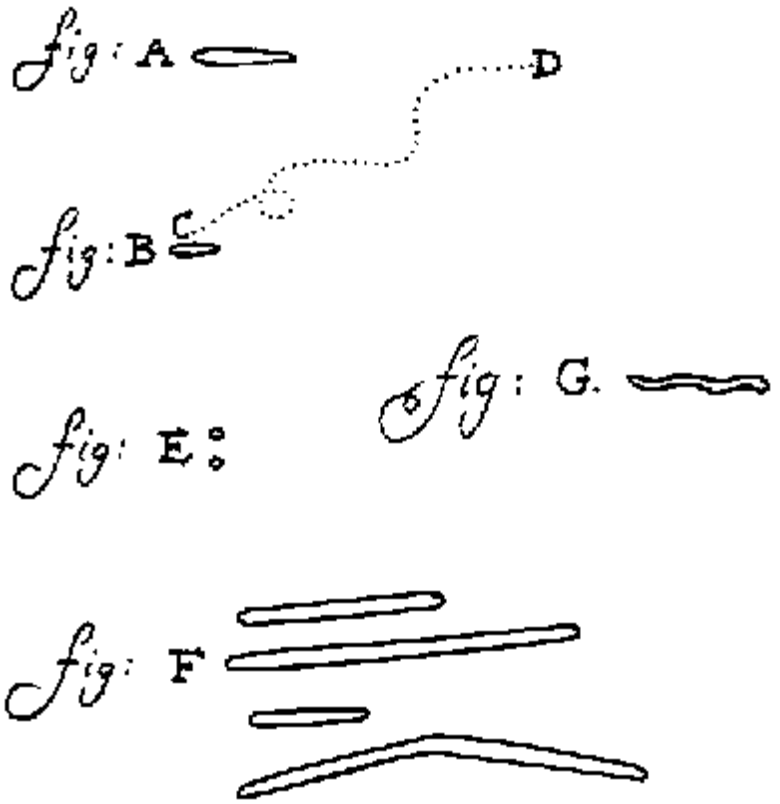
- What allowed them to be discovered?

Early microbiologists

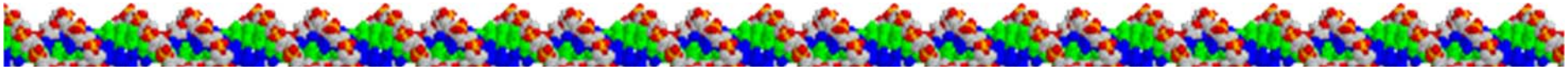
Far Side - Gary Larson

Who first observed Microbes?

1. Louis Pasteur
   0%
2. Galileo
   0%
3. Anton Von Leeuwenhoek
   0%
4. Robert Koch
   0%
5. Alexander Fleming
   0%



The 21st Century Microscope



Sequencing systems for every lab, application, and scale of study.

From the power of the HiSeq X to the speed of MiSeq, Illumina has the sequencer that's just right for you.

MiSeq
Focused power. Speed and simplicity for targeted and small genome sequencing.

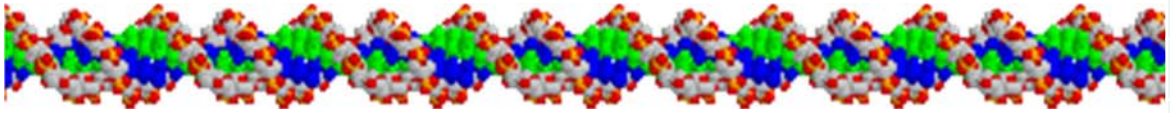
NextSeq 500
Flexible power. Speed and simplicity for everyday genomics.

HiSeq 2500
Production power. Power and efficiency for large-scale genomics.

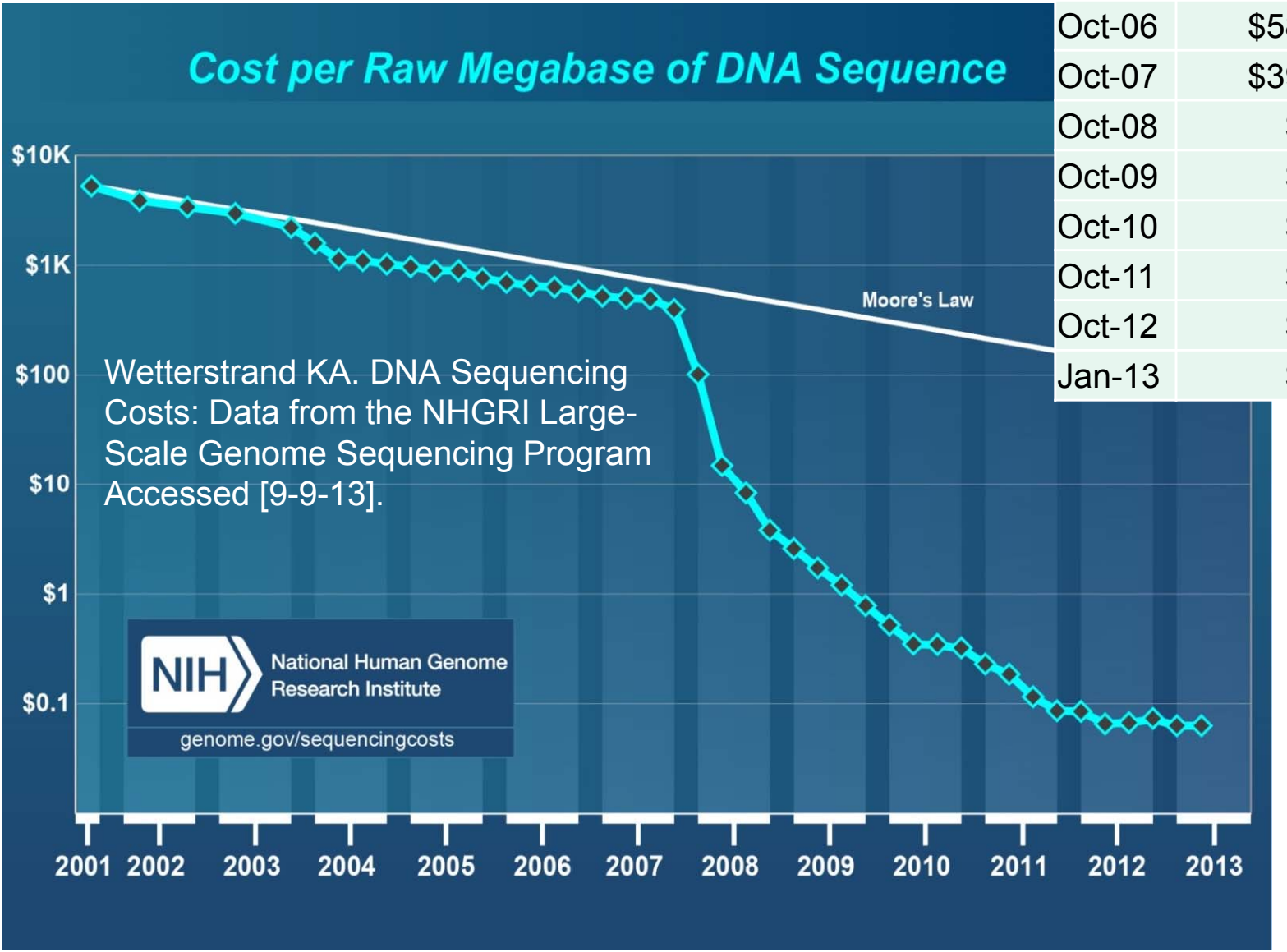
HiSeq X*
Population power. $1,000 human genome and extreme throughput for population-scale sequencing.

Key applications	Small genome, amplicon, and targeted gene panel sequencing.	Everyday genome, exome, transcriptome sequencing, and more.		Production-scale genome, exome, transcriptome sequencing, and more.		Population-scale human whole-genome sequencing.
Run mode	N/A	Mid-Output	High-Output	Rapid Run	High-Output	N/A
Flow cells processed per run	1	1	1	1 or 2	1 or 2	1 or 2
Output range	0.3-15 Gb	20-39 Gb	30-120 Gb	10-180 Gb	50-1000 Gb	1.6-1.8 Tb
Run time	5-65 hours	15-26 hours	12-30 hours	7-40 hours	< 1 day - 6 days	< 3 days
Reads per flow cell†	25 Million‡	130 Million	400 Million	300 Million	2 Billion	3 Billion
Maximum read length	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp	2 × 125 bp	2 × 150 bp

NextGen Sequencing Cost



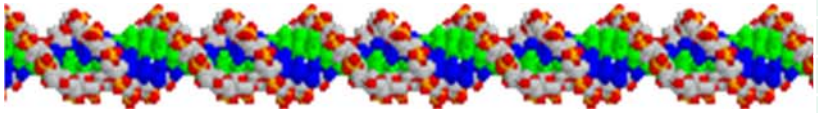
Date	Cost per Mb	Cost per Genome
Sep-01	$5,292.39	$95,263,072
Sep-02	$3,413.80	$61,448,422
Oct-03	$2,230.98	$40,157,554
Oct-04	$1,028.85	$18,519,312
Oct-05	$766.73	$13,801,124
Oct-06	$581.92	$10,474,556
Oct-07	$397.09	$7,147,571
Oct-08	$3.81	$342,502
Oct-09	$0.78	$70,333
Oct-10	$0.32	$29,092
Oct-11	$0.09	$7,743
Oct-12	$0.07	$6,618
Jan-13	$0.06	$5,671



From Spring 2001 → Spring 2013, cost/Mb decreased by factor of 88,000

Revolutionary!

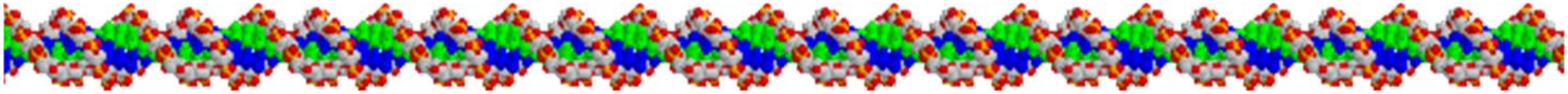
Shared MiSeq Runs



- NextGen Instruments generate more data than most UG faculty can use or afford.
- November 2013 – 27 bacteria @$200 each
- April, 2014 – Opened to Microedu Listserv →
  35 Bacteria and Phage from 16 institutions @$190/sample
- One run planned each semester and summer for the foreseeable future.

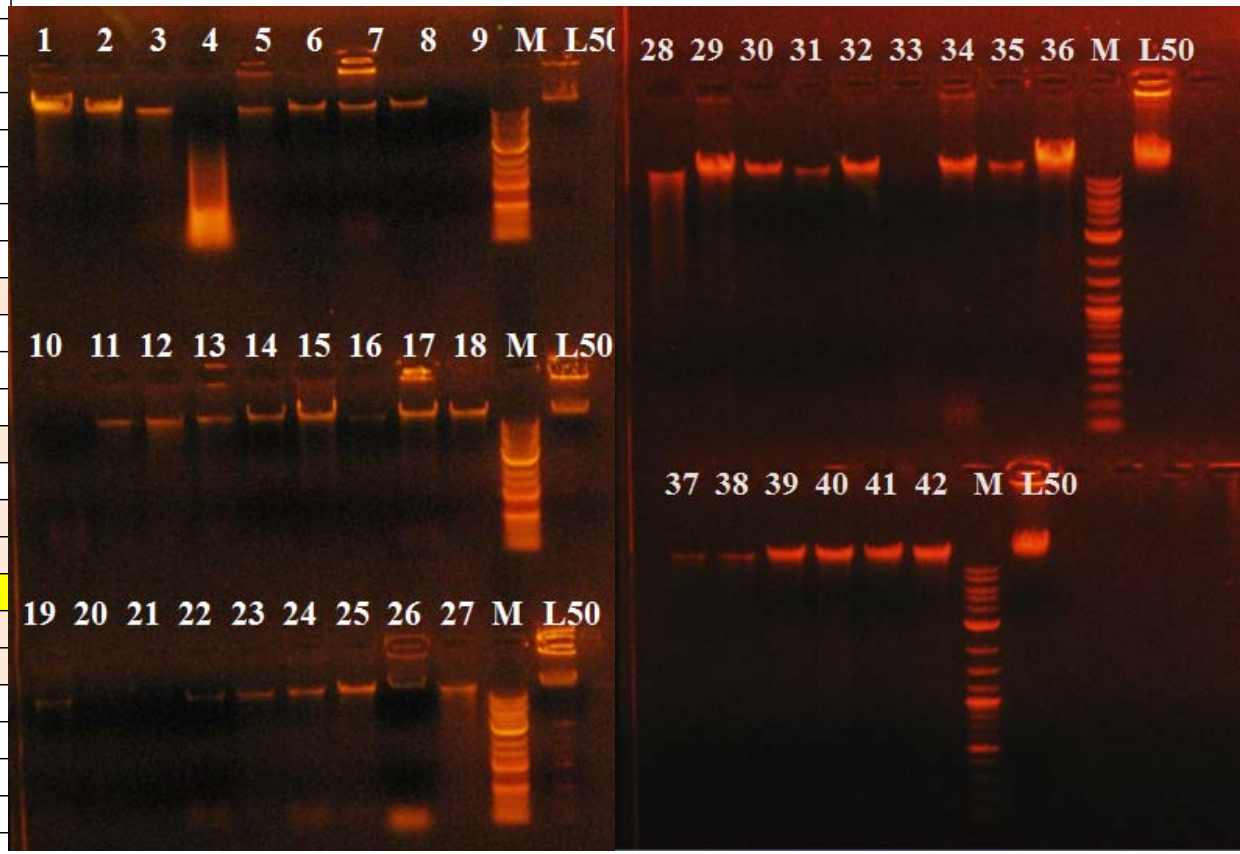
Sample	Reads est.	Bases est.
GSF665-1-E_coli-C06b	217,320	130,391,966
GSF665-2-Chryseobacterium-LO	1,317,872	790,723,170
GSF665-3-Linfield-KH	809,893	485,935,870
GSF665-4-Linfield-NH	301,171	180,702,758
GSF665-5-Exiguobacterium	794,482	476,689,384
GSF665-6-Plesiomonas_shigelloides	656,143	393,685,659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

gDNA from workshop or FedEx

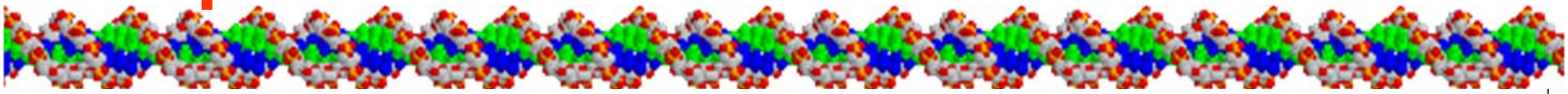


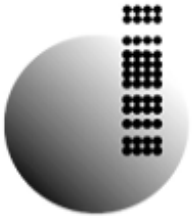
- QC via gel and Qubit to prep for shared run. Sequencing center requests 1 μg in 100 μL.

tube	Genomes to sequence	[DNA] (ng/uL) gel	[DNA] (ng/uL) Qubit	volume (uL)
1	Flavobacterium aquatile	200	104	100
2	Flavobacterium sp AED (franzi)	125	41.4	100
3	Flavobacterium reichenbachii	40	17.6	100
5	Flavobacterium sp. ABG (douthatii)	20	17.9	100
6	Flavobacterium chungangense	60	18.6	100
7	Flavobacterium chiliense	60	17	100
8	Flavobacterium sp KJJ	60	20.2	100
11	Bacillus indicus	15	13	100
12	Bacillus sp SJS (colbertis)	30	18.4	100
13	Pedobacter novum sp 20-19	15	18.6	100
14	Pedobacter borealis	40	22.4	100
15	Meiothermus	60	57.3	100
16	Chryseobacterium hispalense	5	3.2	200
18	Chryseobacterium sp FH2	40	20.4	100
19	Chryseobacterium antarcticum	10	11.9	100
22	Epilithonimonas Lactis	10	11.9	100
23	Epilithonimonas diehli FH1	20	6.75	150
24	Chryseobacterium piperi	25	15.5	100
25	Chryseobacterium angstadtii	40	9.76	150
26	Assia sp.	30	27.9	50
28	Flavobacterium succinicans	10	47.9	100
29	Chryseobacterium sp. JM1	40	59.5	70
31	Flavobacterium sp R30-53	10	9.15	125
34	Chryseobacterium sp BLS98	30	34.2	100
30	Chryseobacterium vrystaatense	20	16.8	100
36	Chryseobacterium luteum	75	74.7	100
17	Chryseobacterium formosense	50	24.3	100



Download data from web to your own computer or Juniata GCAT-SEEK server





THE CENTER FOR GENOMICS AND BIOINFORMATICS

Listing Directory: lims.cgb.indiana.edu/gs454/JeffNewman_Lycoming/GSF665-30Apr2014/

Name	Last modified	Size
Parent Directory		-
1-GSF-3May2014.tar.gz	03-May-2014 13:11	9.1G
GSF665-1-E-coli-C06b_S1_L001_R1_001.fastq.gz	03-May-2014 07:05	43M
GSF665-1-E-coli-C06b_S1_L001_R2_001.fastq.gz	03-May-2014 07:05	40M
GSF665-2-Chryseobacterium-LO_S2_L001_R1_001.fastq.gz	03-May-2014 07:05	184M
GSF665-2-Chryseobacterium-LO_S2_L001_R2_001.fastq.gz	03-May-2014 07:05	234M
GSF665-3-Linfield-KH_S3_L001_R1_001.fastq.gz	03-May-2014 07:05	129M
GSF665-3-Linfield-KH_S3_L001_R2_001.fastq.gz	03-May-2014 07:05	171M
GSF665-4-Linfield-NH_S4_L001_R1_001.fastq.gz	03-May-2014 07:05	51M
GSF665-4-Linfield-NH_S4_L001_R2_001.fastq.gz	03-May-2014 07:05	71M
GSF665-5-Exiguobacterium_S5_L001_R1_001.fastq.gz	03-May-2014 07:05	122M
GSF665-5-Exiguobacterium_S5_L001_R2_001.fastq.gz	03-May-2014 07:05	164M
GSF665-6-Plesiomonas-shigelloides_S6_L001_R1_001.fastq.gz	03-May-2014 07:05	105M
GSF665-6-Plesiomonas-shigelloides_S6_L001_R2_001.fastq.gz	03-May-2014 07:05	142M
GSF665-7-Halosimplex-carlsbadense_S7_L001_R1_001.fastq.gz	03-May-2014 07:05	101M
GSF665-7-Halosimplex-carlsbadense_S7_L001_R2_001.fastq.gz	03-May-2014 07:05	137M
GSF665-8-Phage-Eapen_S8_L001_R1_001.fastq.gz	03-May-2014 07:05	100M
GSF665-8-Phage-Eapen_S8_L001_R2_001.fastq.gz	03-May-2014 07:05	116M
GSF665-9-Phage-Aspire_S9_L001_R1_001.fastq.gz	03-May-2014 07:05	34M
GSF665-9-Phage-Aspire_S9_L001_R2_001.fastq.gz	03-May-2014 07:05	37M

Quality Filter Data

Flavobacterium-aquatile_S1_L001_R1_001.fastq
[File Converted]: D:\Data\Newman\MiSeq 2013-11-30\GSF634-1-Flavobacterium-aquatile_S1\GSF634-1-Flavobacterium-aquatile_S1_L001_R1_001_converted.fasta
[File Removed]: D:\Data\Newman\MiSeq 2013-11-30\GSF634-1-Flavobacterium-aquatile_S1\GSF634-1-Flavobacterium-aquatile_S1_L001_R1_001_removed.fasta

[Settings]
File Format Type: Illumina FASTQ
Quality Format: ASCII-33
"Median Score Threshold" Checked: TRUE Median Score Threshold: 20.00
"Max # of Uncalled Bases" Checked: TRUE Max # of Uncalled Bases: 3
"Called Base Number" Checked: TRUE Called Base Number: 25
"Trim or Reject Read" Checked: TRUE Trim or Reject Read While >= 3 Bases with Score <= 16.00
"Paired Reads Data" Checked: TRUE
"Remove 5' End and 3' End" Checked: FALSE
"Trim by Sequence" Checked: FALSE

[Filter Results]

- [Total Reads in the Input File]: 1028164
  - [Reads Converted Successfully]: 1026970
  - [Reads Failed to Convert]: 1194
    - [Reads Filtered by "Median Score"]: 538
    - [Reads Filtered by "Uncalled Bases"]: 507
    - [Reads Filtered by "Called Base Number in Each Read"]: 0
    - [Reads Rejected by "Base's Score"]: 149
  - [Reads Trimmed by "Base's Score"]: 46445
  - [Trimmed Bases by "Base's Score"]: 4372737
  - [Homopolymer Bases Trimmed]: 0

[Statistics]

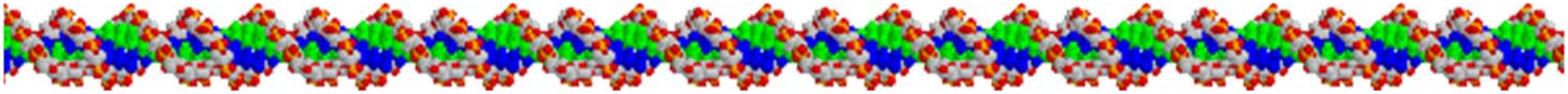
[Average Score of Each Base]: 33.85 33.88 33.88 33.91 33.90 37.77 37.78 37.69
37.67 37.67 37.68 37.65 37.66 37.66 37.66 37.67 37.63 37.66 37.66 37.67
37.60 37.62 37.65 37.64 37.63 37.60 37.60 37.60 37.59 37.55 37.57 .....

Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%
50	1 in 100000	99.999%

http://en.wikipedia.org/wiki/Phred_quality_score

Assembly statistics



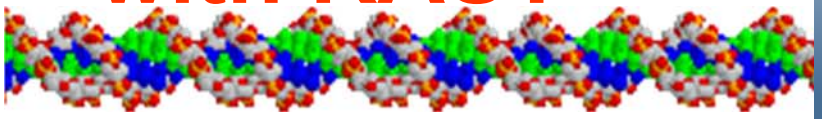
[SoftGenetics Assembler: Assembly Results Statistics Report]

- **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**

Annotation with RAST

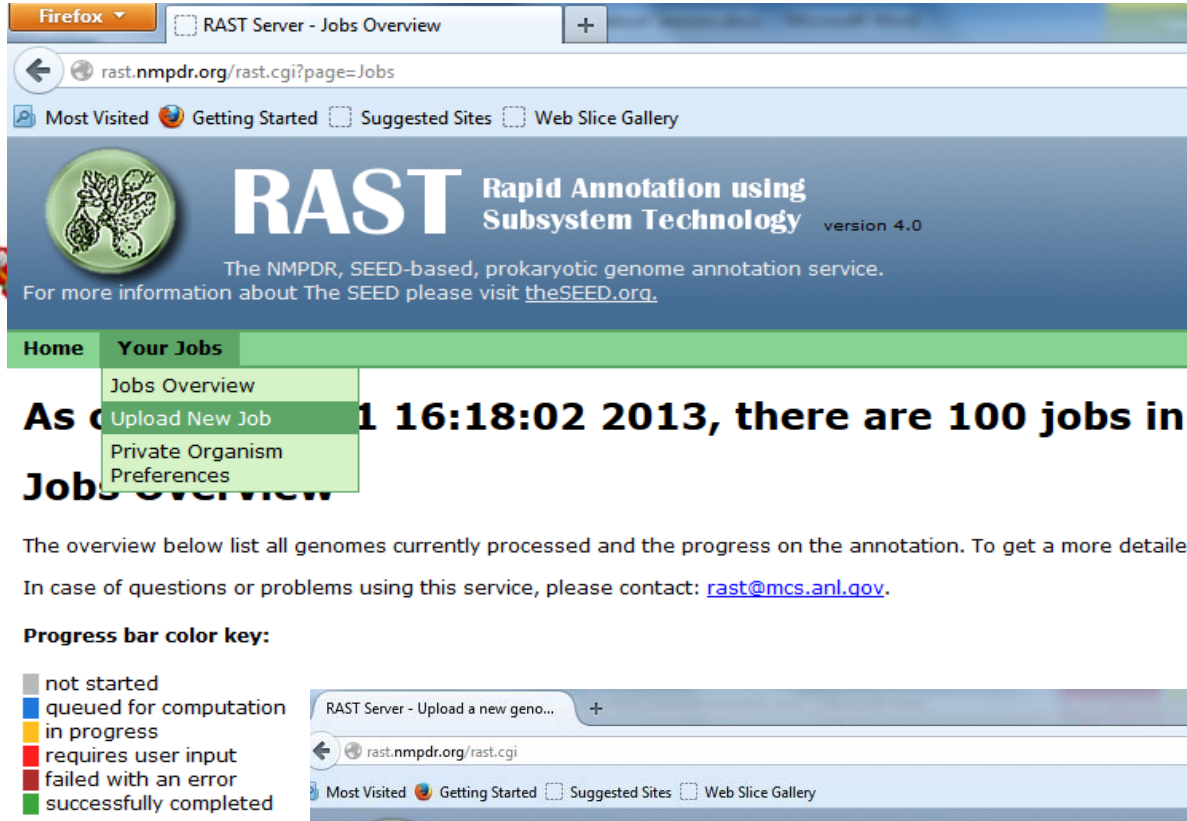


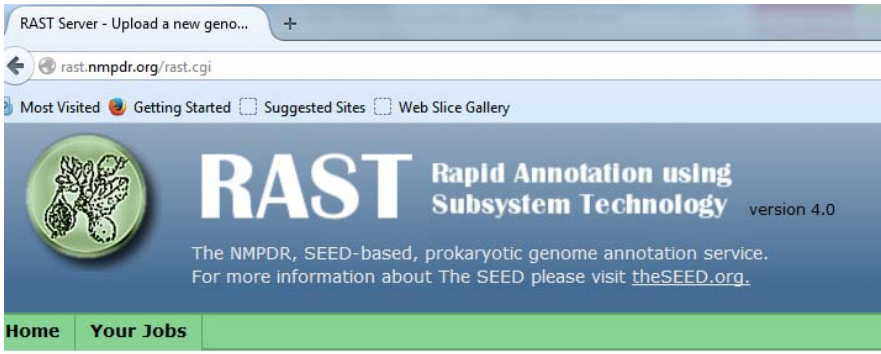
- http://rast.nmpdr.org/

- Login: newmanlab
- Pw: 16srrna1

- Upload genome as a .fasta file,
- input NCBI taxon ID,
- Select highest figfam#

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





Upload a Genome

Review genome data

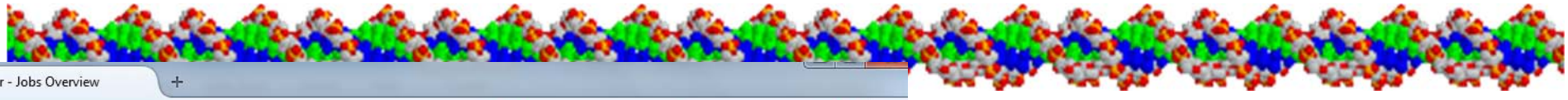
We have analyzed your upload and have computed the following information.

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 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

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.

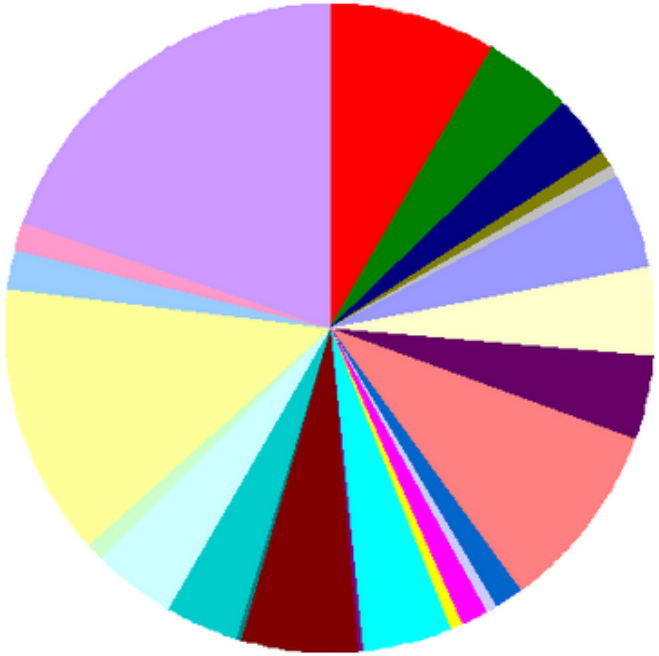
Available comparisons are function based, sequence based or via KEGG. You can also BLAST against this organism.

Subsystem Information

Subsystem Statistics | Features in Subsystems

Subsystem Coverage

Subsystem Category Distribution

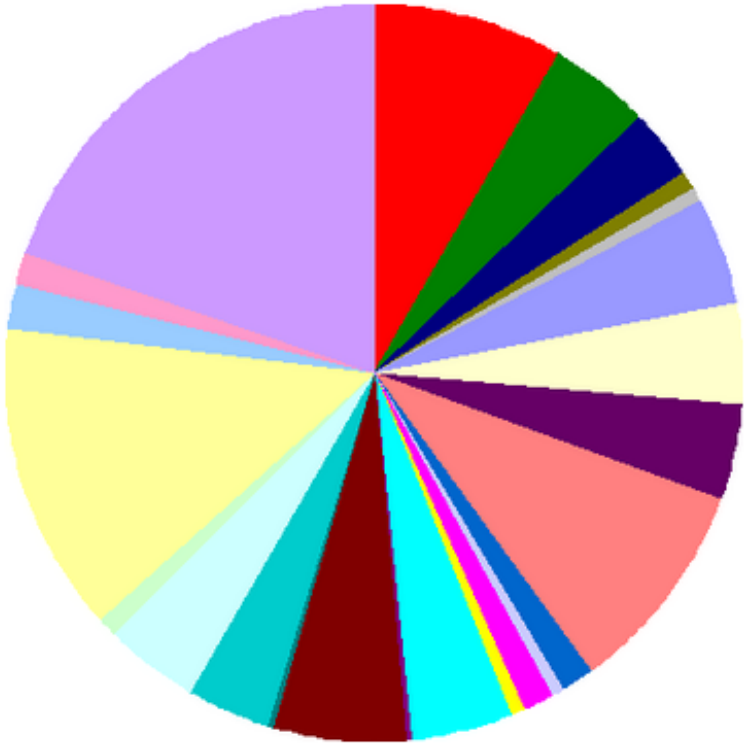


Subsystem Feature Counts

- 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)

Subsystem Coverage

Subsystem Category Distribution



Micro Course Assignment: Predict Phenotypes for Known Organism

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)

	F. douthatii	F. glaciei	F. succinicans
Negative Control	10	9	16
6 dextrin	98	19	56
D-maltose	98	98	99
1 D-trehalose	98	98	22
D-cellobiose	96	60	29
gentiobiose	100	77	99
sucrose	11	95	13
D-turanose	43	10	38
stachyose	18	45	24
pos control	97	100	100
pH 6	94	30	93
pH 5	11	7	11
D-raffinose	15	18	18
8 α-D-lactose	21	56	96
D-mellibiose	22	53	19
β-methyl-D-glucoside	18	56	17
D-salicin	15	51	20
N-acteyl-D-glucpsamine	100	44	27
N-acteyl-β-D-mannisamine	32	45	18
N-acetyl-D-galactosamine	99	57	24
N-acetyl neuraminic acid	10	20	15
1% NaCl	66	30	54
4% NaCl	14	13	12
8% NaCl	17	15	17
α-D-glucose	98	69	98
4 D-mannose	99	94	98
D-fructose	45	46	40
2 D-galactose	100	55	100
3-methyl glucose	15	14	7
D-fucose	11	54	25
L-fucose	32	51	31
L-rhamnose	30	51	25
inosine	11	26	38
1% Na-lactate	62	19	19
fusidic acid	20	37	19
D-serine	19	29	17
D-sorbitol	14	7	20
D-mannitol	21	50	29
D-arabitol	22	39	33
myo-inositol	16	38	12
3 glycerol	39	45	34
D-glucose-6-PO4	58	66	44
D-fructose-6-PO4	48	48	23
D-aspartic acid	14	5	24
D-serine	3	5	4
troleandomycin	17	38	19
rifamycin SV	96	43	70
minocycline	20	48	20

	F. douthatii	F. glaciei	F. succinicans
gelatin	54	13	97
glycyl-L-proline	85	90	92
L-alanine	56	31	10
L-arginine	85	52	100
L-aspartic acid	96	38	99
L-glutamic acid	95	56	100
L-histidine	86	10	35
L-pyroglutamic acid	29	39	32
7 L-serine	89	44	15
lincomycin	21	48	23
guanidine HCL	16	12	14
niaproof 4	17	11	17
pectin	49	43	23
5 D-galacturonic acid	97	64	34
L-galacturonic acid lactone	7	5	13
D-gluconic acid	34	49	36
5 D-glucuronic acid	82	69	40
glucuronamide	43	52	33
mucic acid	31	5	39
quinic acid	34	48	40
D-saccharic acid	33	25	35
vancomycin	97	44	82
tetrazolium violet	100	77	45
tetrazolium blue	99	61	99
p-hydroxy-phenylacetic acid	7	7	11
methyl pyruvate	37	5	24
D-lactic acid methyl ester	51	58	44
L-lactic acid	24	5	34
citric acid	47	5	56
α-keto-glutaric acid	37	33	46
D-malic acid	29	14	13
L-malic acid	40	41	98
bromo-succinic acid	11	6	30
nalidixic acid	40	49	25
LiCl	16	12	14
K-tellurite	23	25	24
tween-40	39	15	46
Y-amino-butyric acid	18	13	13
α-hydroxy-butyric aicd	20	8	49
β-hydroxy-D,L- butyric acid	22	10	22
α-keto-butyric acid	11	7	41
acetoacetic acid	77	52	67
propionic acid	23	8	27
acetic acid	69	28	94
formic acid	18	7	4
aztreonam	94	42	98
Na-butyrate	24	48	23
Na bromate	14	13	12

Reciprocal Orthology Score Average (ROSA) vs. *F. succ.*

Figure 6. Annotation of Genome with RAST

RAST Subsystems Carbohydrates (260)

Central carbohydrate metabolism (106)

Aminosugars (0); Organic acids (0); Glycoside hydrolases (0);

CO2 fixation (0)

Di- and oligosaccharides (23)

- Trehalose Biosynthesis (3)
- 1 • Trehalose Uptake and Utilization (4)
- 2 • Lactose and Galactose Uptake and Utilization (10)
- 8 • Lactose utilization (6) **(No Transporter)**

One-carbon Metabolism (34)

- 7 • Serine-glyoxylate cycle (29)
- One-carbon metabolism by tetrahydropterines (5)

Fermentation (23)

- Butanol Biosynthesis (9)
- Acetolactate synthase subunits (2)
- Fermentations: Lactate (3)
- Acetyl-CoA fermentation to Butyrate (9)

Sugar alcohols (9)

- 3 • Glycerol and Glycerol-3-phosphate Uptake and Utilization (9

Polysaccharides (20)

- 6 • Glycogen metabolism (5)
- Cellulosome (15)

Monosaccharides (45)

- 4 • Mannose Metabolism (11)
- D-ribose utilization (2)
- Deoxyribose and Deoxynucleoside Catabolism (4)
- D-gluconate and ketogluconates metabolism (4)
- 5 • D-Galacturonate and D-Glucuronate Utilization (24)

Phenotype Comparisons

B

	Chryseo. populense	*Chryseo. hispalense*	*Chryseo. wanjuense*	*Chryseo. daecheongense*
neg control	19	34	25	38
dextrin	100	100	100	100
D-maltose	82	96	97	93
D-trehalose	95	19	99	98
D-cellobiose	8	21	27	35
gentiobiose	99	100	100	99
sucrose	11	19	22	99
D-turanose	8	14	8	16
stachyose	9	18	15	28
pos control	97	98	97	98
pH 6	96	98	96	97
pH 5	93	96	92	96
D-raffinose	14	16	19	26
α-D-lactose	10	11	20	18
D-melibiose	11	22	26	38
β-methyl-D-glucoside	8	14	12	18
D-salicin	7	8	6	11
N-acetyl-D-glucosamine	7	10	17	18
N-acetyl-β-D-mannosamine	7	10	15	21
N-acetyl-D-galactosamine	9	10	10	20
N-acetyl neuraminic acid	11	23	24	37
1% NaCl	94	95	93	94
4% NaCl	13	69	8	11
8% NaCl	17	19	10	15
α-D-glucose	82	93	99	91
D-mannose	98	84	95	88
D-fructose	99	98	99	21
D-galactose	13	12	14	23
3-methyl glucose	6	8	10	9
D-fucose	12	10	17	28
L-fucose	11	16	12	24
L-rhamnose	29	89	59	15

	Chryseo. populense	*Chryseo. hispalense*	*Chryseo. wanjuense*	*Chryseo. daecheongense*
inosine	9	11	8	9
1% Na-lactate	94	94	93	92
fusidic acid	11	10	9	7
D-serine	16	12	11	13
D-sorbitol	14	18	18	29
D-mannitol	9	10	15	14
D-arabitol	9	9	11	14
myo-inositol	11	11	22	26
glycerol	77	8	93	81
D-glucose-6-PO4	11	19	16	13
D-fructose-6-PO4	18	22	17	22
D-aspartic acid	7	8	4	4
D-serine	9	8	4	3
troleandomycin	9	9	38	11
rifamycin SV	95	90	95	95
minocycline	16	12	16	12
gelatin	100	100	100	100
glycyl-L-proline	91	42	81	88
L-alanine	8	8	4	3
L-arginine	12	22	18	59
L-aspartic acid	27	98	57	95
L-glutamic acid	97	99	90	99
L-histidine	12	13	9	14
L-pyroglutamic acid	10	11	12	8
L-serine	51	8	4	9
lincomycin	95	9	94	95
guanidine HCl	21	8	86	6
niaproof 4	12	11	9	8
pectin	22	97	44	90
D-galacturonic acid	95	97	78	18
L-galacturonic acid lactone	8	8	5	4
D-gluconic acid	9	11	10	9

	Chryseo. populense	*Chryseo. hispalense*	*Chryseo. wanjuense*	*Chryseo. daecheongense*
D-glucuronic acid	19	14	23	23
glucuronamide	17	16	10	13
mucic acid	10	10	10	7
quinic acid	11	11	13	7
D-saccharic acid	12	11	12	8
vancomycin	12	7	9	26
tetrazolium violet	98	43	91	85
tetrazolium blue	98	99	98	66
p-hydroxy-phenylacetic acid	13	12	8	10
methyl pyruvate	11	8	6	3
D-lactic acid methyl ester	11	8	8	5
L-lactic acid	12	8	16	7
citric acid	11	13	12	17
α-keto-glutaric acid	11	8	57	6
D-malic acid	10	8	11	6
L-malic acid	12	11	14	6
bromo-succinic acid	9	8	4	3
nalidixic acid	14	9	13	9
LiCl	14	73	11	6
K-tellurite	95	52	96	97
tween-40	94	95	95	98
γ-amino-butyric acid	14	8	16	9
α-hydroxy-butyric acid	17	10	13	6
β-hydroxy-D,L-butyric acid	15	9	20	7
α-keto-butyric acid	16	8	8	3
acetoacetic acid	75	86	72	82
propionic acid	14	8	8	8
acetic acid	97	94	99	99
formic acid	16	8	12	61
aztreonam	97	98	96	97
Na-butyrate	36	85	15	32
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

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.

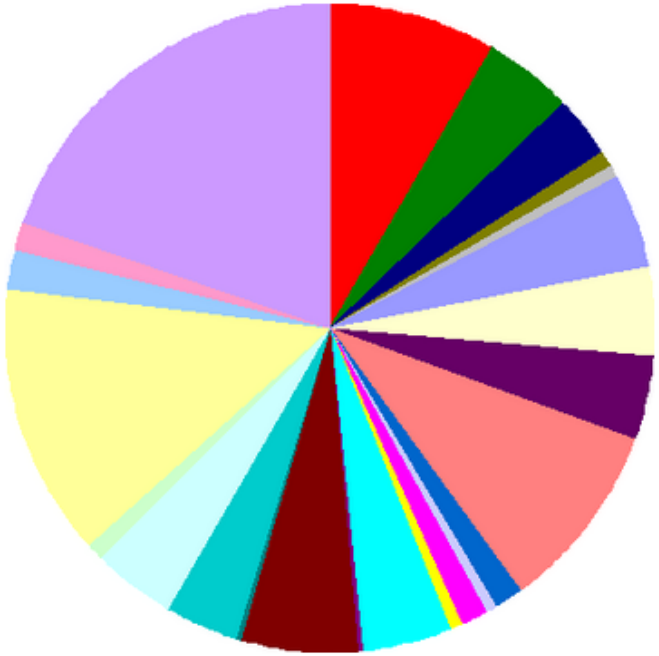
Available comparisons are function based, sequence based or via KEGG. You can also BLAST against this organism.

Subsystem Information

Subsystem Statistics | Features in Subsystems

Subsystem Coverage

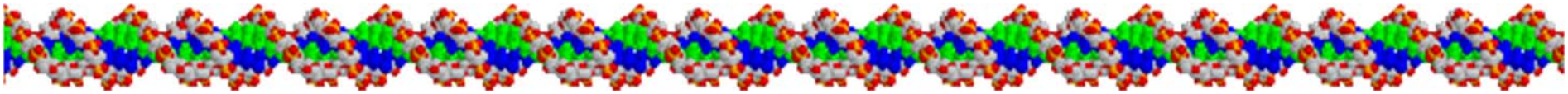
Subsystem Category Distribution



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)

Sequence-Based Comparison



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

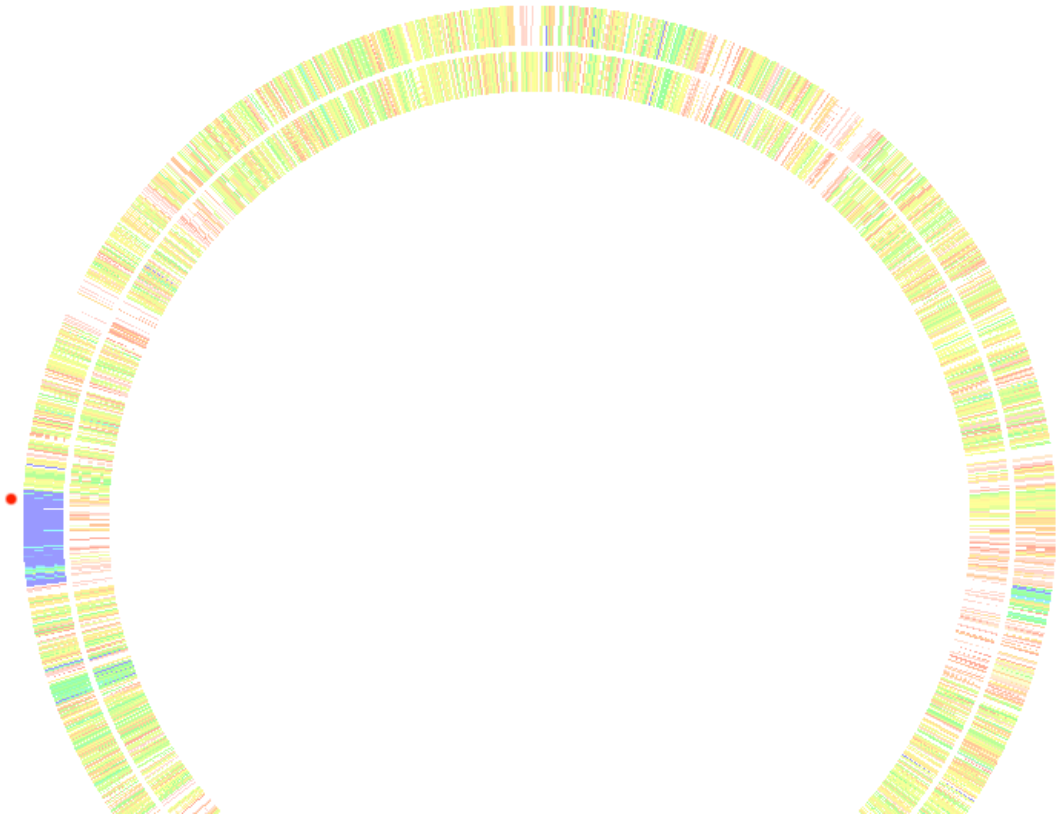
«first «prev displaying 3028 - 3057 of 4006 next» last»

percent identity 525257.7 < | percent identity 1144316.4 <

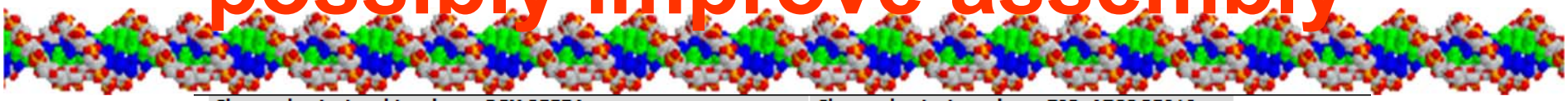
491205.4			525257.7			1144316.4		
Contig (all)	Gene	Length	Hit (all)	Contig (all)	Gene	Hit (all)	Contig (all)	Gene
21	3028	360	bi	1	241	bi	22	1525
21	3029	270	bi	1	240	uni	51	2777
21	3030	680	bi	1	239	uni	2	181
21	3031	92	bi	1	238	uni	14	1009
21	3032	580	bi	1	237	uni	49	2655
21	3033	298	bi	1	236	bi	2	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	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			

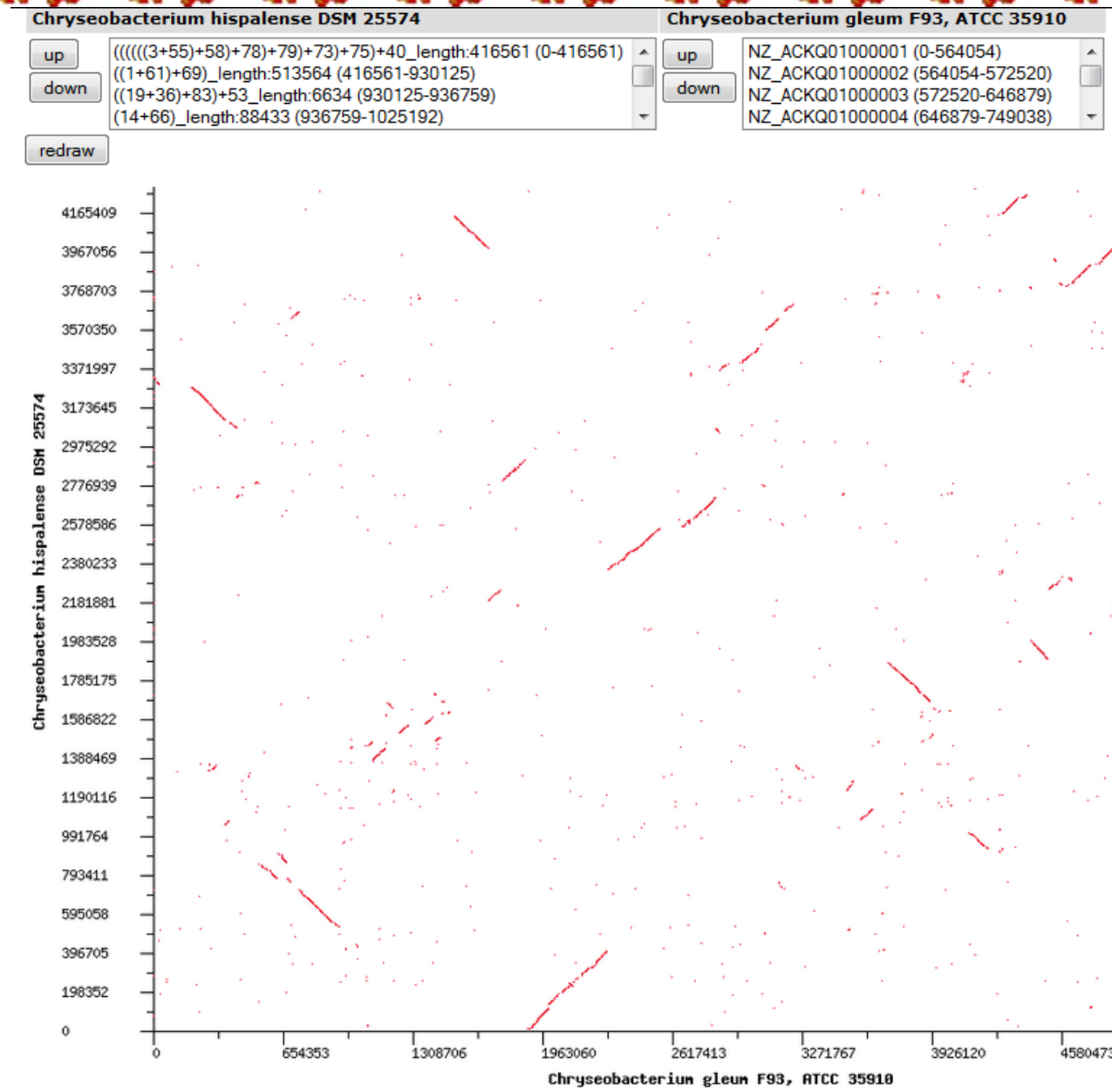
fig|525257.7.peg.224
location: NZ_ACKQ01000001 253632 254855
length: 407
identity: 1
function: Tyrosine type site-specific recombinase

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

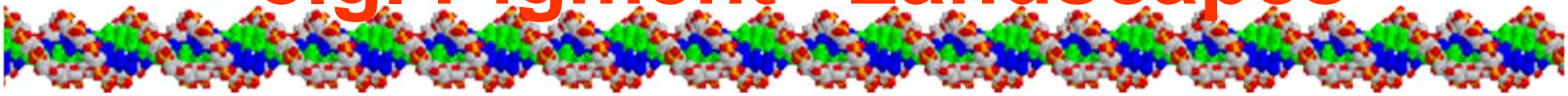


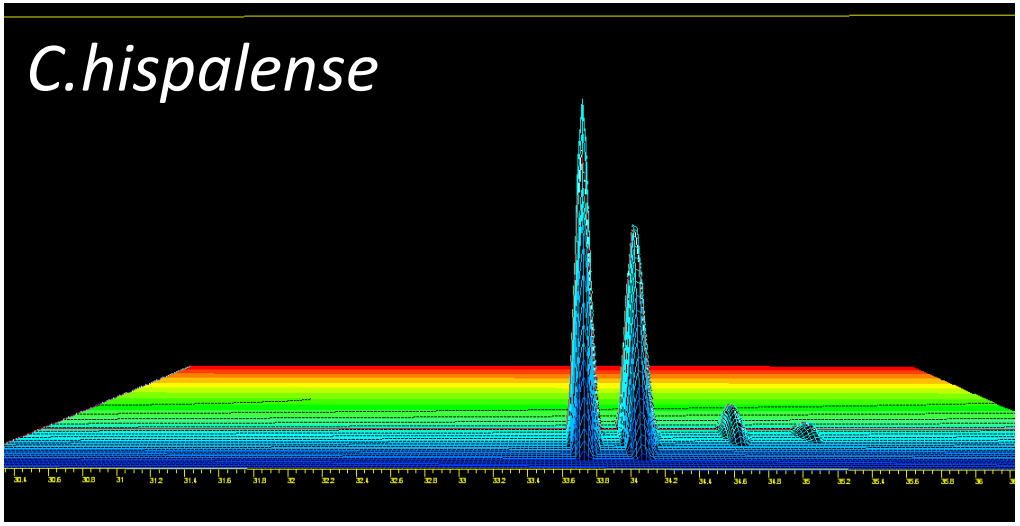
Dot Plots to Examine Synteny … possibly improve assembly

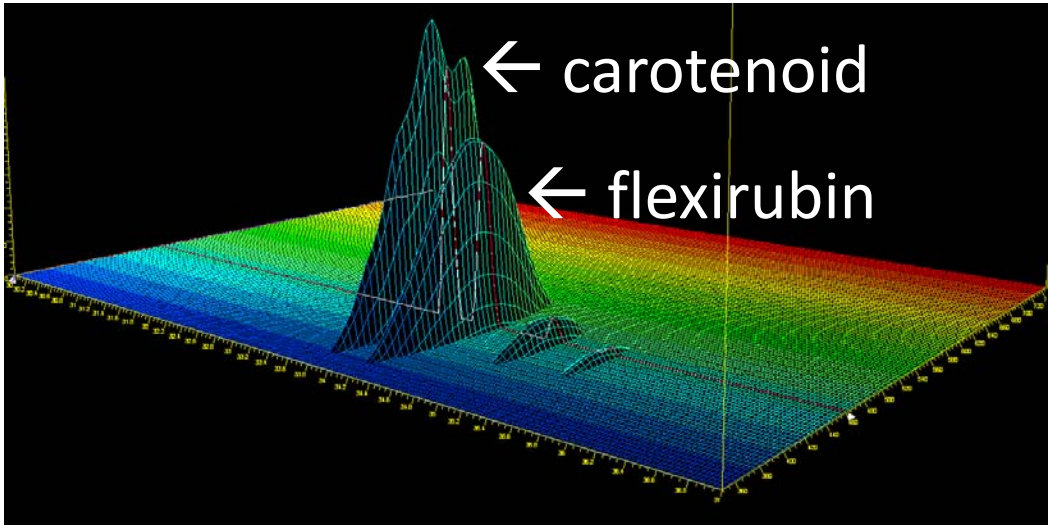


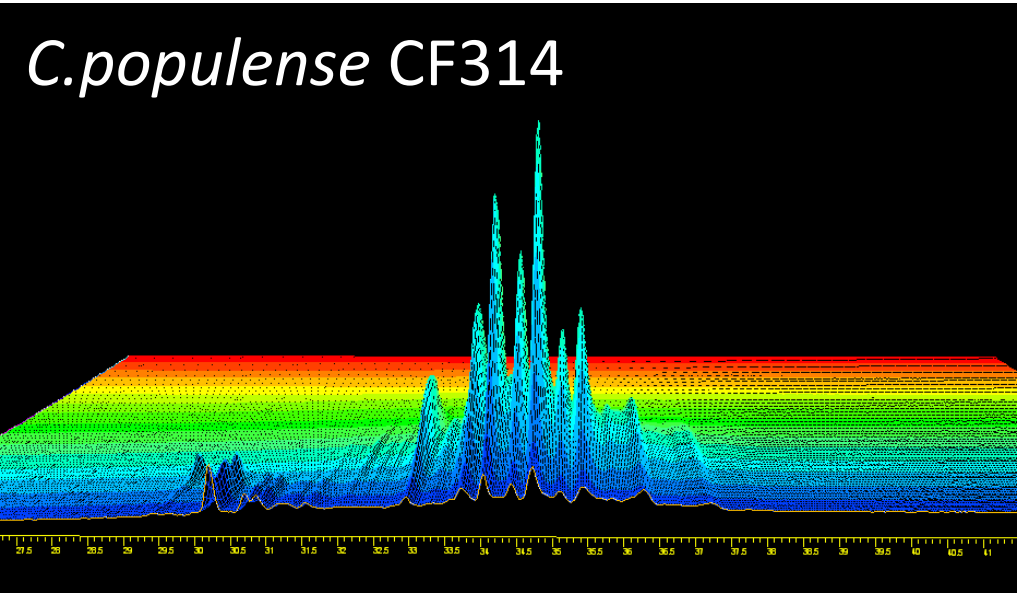


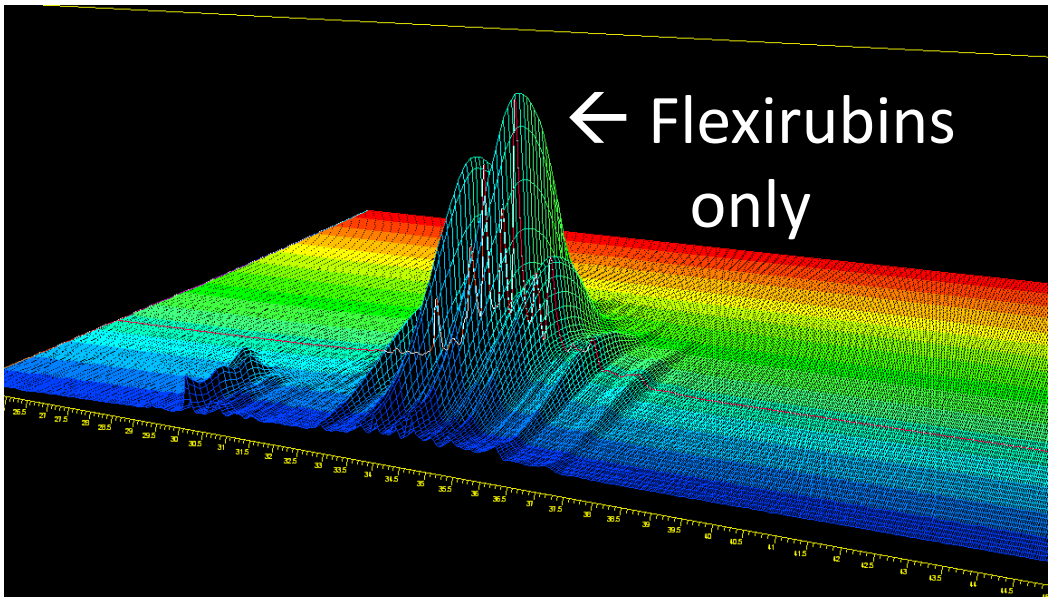
Explain phenotypic differences – e.g. Pigment "Landscapes"



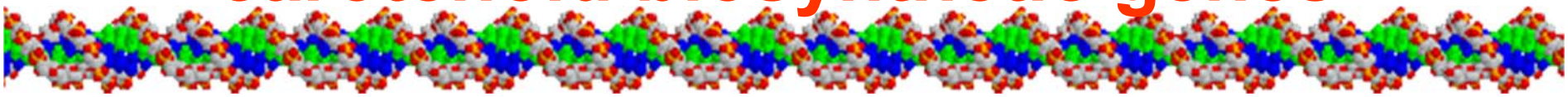








C.populense lacks carotenoid biosynthetic genes



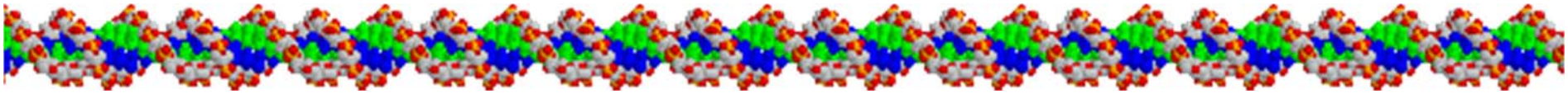
C.hispalense *C.populense*

491205.4			558151.4			1121286.3			236814.3			525257.7			1121287.3			1218103.4			558152.3			445961.3			59732.8			307480.3		
Contig all	Gene	Length	Hit all	Contig all	Gene	Hit all	Contig all	Gene	Hit all	Contig all	Gene	Hit all	Contig all	Gene	Hit all	Contig all	Gene	Hit all	Contig all	Gene	Hit all	Contig all	Gene	Hit all	Contig all	Gene	Hit all	Contig all	Gene	Hit all	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	-				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	-				19	4076
8	10											-			-			-			-			-								
8	10										2234	bi	1	511	bi	4	1403	bi	2	273	bi	75	3573	bi	5	753	bi	16	370	bi	2	1192
8	10										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

Sequence-Based Comparison



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

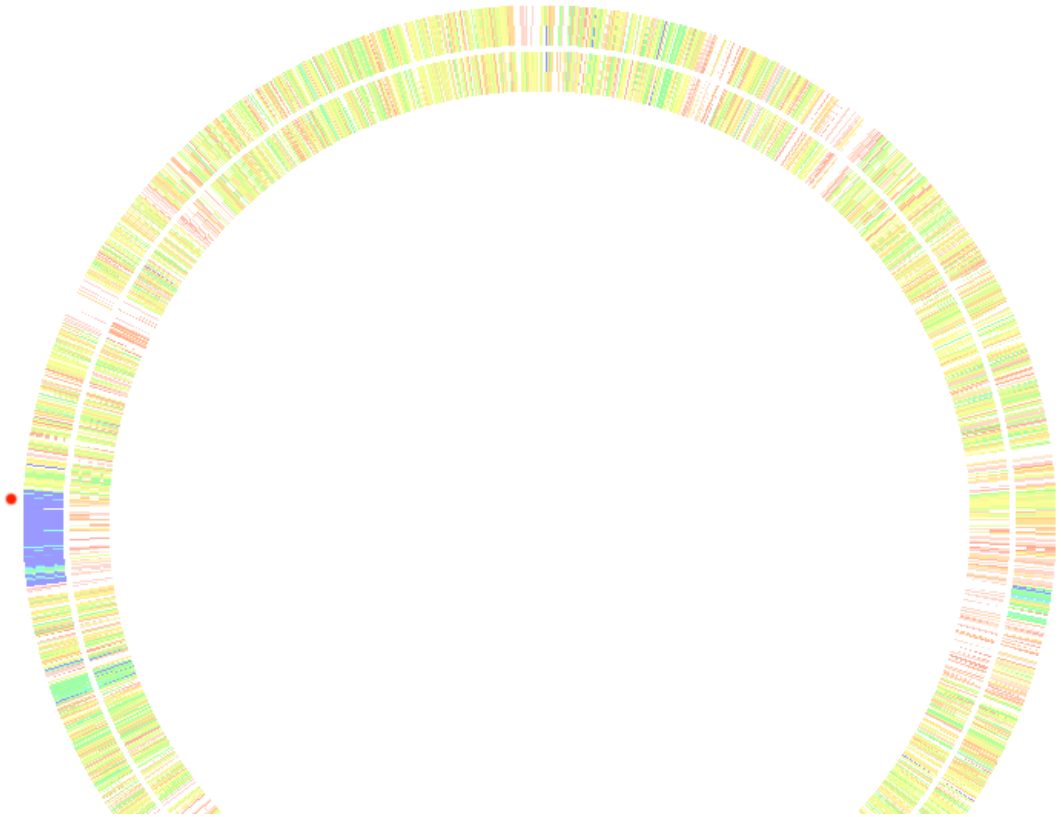
«first «prev displaying 3028 - 3057 of 4006 next» last»

percent identity 525257.7 < | percent identity 1144316.4 <

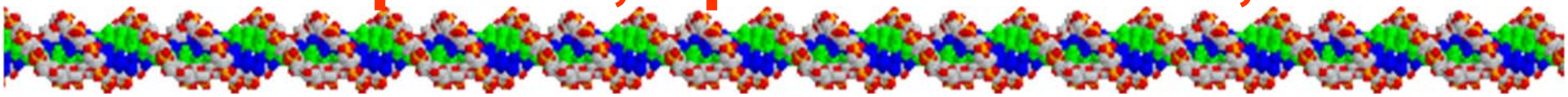
491205.4			525257.7			1144316.4		
Contig (all)	Gene	Length	Hit (all)	Contig (all)	Gene	Hit (all)	Contig (all)	Gene
21	3028	360	bi	1	241	bi	22	1525
21	3029	270	bi	1	240	uni	51	2777
21	3030	680	bi	1	239	uni	2	181
21	3031	92	bi	1	238	uni	14	1009
21	3032	580	bi	1	237	uni	49	2655
21	3033	298	bi	1	236	bi	2	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	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

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



Sequence-based comparison tables can be exported, opened in excel, sorted.



genome comparison.xlsx - Microsoft Excel

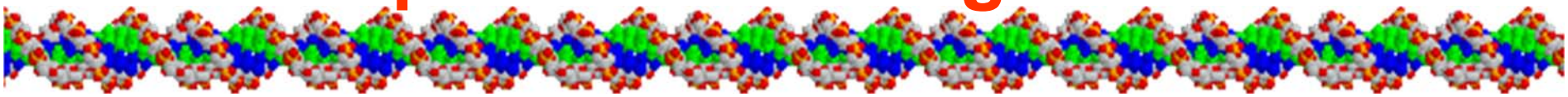
File | Home | Insert | Page Layout | Formulas | Data | Review | View | Acrobat

Paste | Cut | Copy | Format Painter | Clipboard | Calibri | 11 | Font | Wrap Text | Merge & Center | Alignment | General | Number | Conditional Formatting | Format as Table | Cell Styles | Styles | Insert | Delete | Format | Cells | AutoSum | Fill | Clear | Sort & Filter | Find & Select | Editing

R1 | Hit

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Contig	Gene	Length	Gene id	function	Hit	Contig	Gene	Gene id	percent identity	function	Hit	Contig	Gene	Gene id	percent identity	function	Hit
2	1	1	143	fig|6	Ribose 5-phosphate isomerase B (EC 5.3.1.6)	bi	725	1956	fig|66	72.54	Ribose 5-phosphate isomerase B (EC	bi	1019	2657	fig|6	72.54	Ribose 5-phosphate isomerase B (E	bi
3	1	2	719	fig|6	3'-to-5' exoribonuclease RNase R	bi	1268	3156	fig|66	87.63	3'-to-5' exoribonuclease R	bi	175	487	fig|6	89.82	3'-to-5' exoribonuclease	bi
4	1	3	226	fig|6	Lysine exporter protein (LYSE/YGGA)	bi	1036	2712	fig|66	78.87	hypothetical protein	bi	679	1913	fig|6	86.1	hypothetical protein	bi
5	1	4	310	fig|6	Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13)	bi	699	1882	fig|66	71.56	Muramoyltetrapeptide carboxypepti	bi	679	1912	fig|6	68.63	Muramoyltetrapeptide carboxypept	bi
6	1	5	120	fig|6	hypothetical protein	-				0		-				0		-
7	1	6	124	fig|6	hypothetical protein	bi	602	1670	fig|66	74.38	hypothetical protein	bi	456	1309	fig|6	71.54	Endonuclease (EC 3.1.-.-)	bi
8	1	7	251	fig|6	Oxidoreductase, short chain dehydrogenase/reductase fan	bi	602	1671	fig|66	77.6	Oxidoreductase, short chain dehydr	bi	456	1308	fig|6	77.6	Oxidoreductase, short chain dehyd	bi
9	1	8	240	fig|6	Inactive homolog of metal-dependent proteases, putative r	bi	485	1360	fig|66	74.34	Inactive homolog of metal-dependen	bi	519	1487	fig|6	73.51	Inactive homolog of metal-depende	bi
10	1	9	808	fig|6	TPR domain protein	bi	485	1361	fig|66	71.66	TPR domain protein	bi	1038	2694	fig|6	79.35	TPR domain protein	bi
11	1	10	42	fig|6	hypothetical protein	-				0		-				0		-
12	1	11	425	fig|6	Glutamyl-tRNA reductase (EC 1.2.1.70)	bi	459	1286	fig|66	87.01	Glutamyl-tRNA reductase (EC 1.2.1.7	bi	366	1060	fig|6	88.21	Glutamyl-tRNA reductase (EC 1.2.1.7	bi
13	1	12	304	fig|6	Porphobilinogen deaminase (EC 2.5.1.61)	bi	1710	3921	fig|66	81.03	Porphobilinogen deaminase (EC 2.5.	bi	366	1061	fig|6	79.54	Porphobilinogen deaminase (EC 2.5	bi
14	1	13	225	fig|6	uroporphyrinogen-III synthase(EC:4.2.1.75)	bi	678	1841	fig|66	59.09	hypothetical protein	bi	189	518	fig|6	79.43	Uroporphyrinogen III synthase HEM	bi
15	1	14	343	fig|6	Uroporphyrinogen III decarboxylase (EC 4.1.1.37)	bi	678	1840	fig|66	88.3	Uroporphyrinogen III decarboxylase	bi	189	517	fig|6	87.43	Uroporphyrinogen III decarboxylas	bi
16	1	15	181	fig|6	Acetyltransferase, including N-acetylases of ribosomal pro	bi	1175	2996	fig|66	58.62	GNAT family acetyltransferase	bi	189	516	fig|6	62.86	acetyltransferase, GNAT family prot	bi
17	1	16	90	fig|6	hypothetical protein	bi	61	105	fig|66	45.12	MtN3 and saliva related transmemb	bi	1326	3208	fig|6	75	hypothetical protein	-
18	1	17	404	fig|6	GTP-binding protein HflX	bi	61	104	fig|66	88.81	GTP-binding protein HflX	bi	1326	3209	fig|6	88.33	GTP-binding protein HflX	bi
19	1	18	238	fig|6	probable DNA alkylation repair enzyme	bi	175	505	fig|66	68.94	probable DNA alkylation repair enzy	bi	142	380	fig|6	69.23	predicted DNA alkylation repair enz	bi
20	1	19	112	fig|6	hypothetical protein	bi	175	504	fig|66	67.27	hypothetical protein	bi	142	379	fig|6	70.91	hypothetical protein	bi
21	1	20	359	fig|6	sodium/calcium exchanger membrane region	bi	981	2588	fig|66	88.55	Calcium/proton antiporter	bi	90	195	fig|6	83.29	Calcium/proton antiporter	bi
22	1	21	126	fig|6	SEC-independent protein translocase protein TATC	bi	981	2587	fig|66	80	SEC-independent protein translocase	bi	90	196	fig|6	86.25	SEC-independent protein translocas	bi
23	1	22	164	fig|6	hypothetical protein	bi	315	902	fig|66	82.17	hypothetical protein	bi	912	2425	fig|6	81.88	hypothetical protein	bi
24	1	23	103	fig|6	hypothetical protein	-				0		bi	912	2424	fig|6	36.46	hypothetical protein	-
25	1	24	453	fig|6	NADP-specific glutamate dehydrogenase (EC 1.4.1.4)	bi	187	548	fig|66	90.98	NADP-specific glutamate dehydroger	bi	912	2423	fig|6	94.12	NADP-specific glutamate dehydroge	bi
26	1	25	369	fig|6	Rossmann fold nucleotide-binding protein Smf possibly in	bi	560	1572	fig|66	76.55	Rossmann fold nucleotide-binding p	bi	899	2403	fig|6	74.4	Rossmann fold nucleotide-binding	bi
27	1	26	256	fig|6	conserved hypothetical protein, membrane	bi	560	1573	fig|66	68.57	hypothetical protein	bi	118	280	fig|6	79.13	putative transmembrane protein	bi

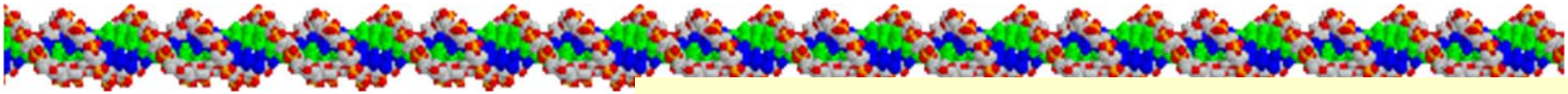
Sequence Based Comparison can ID unique and shared genes…..

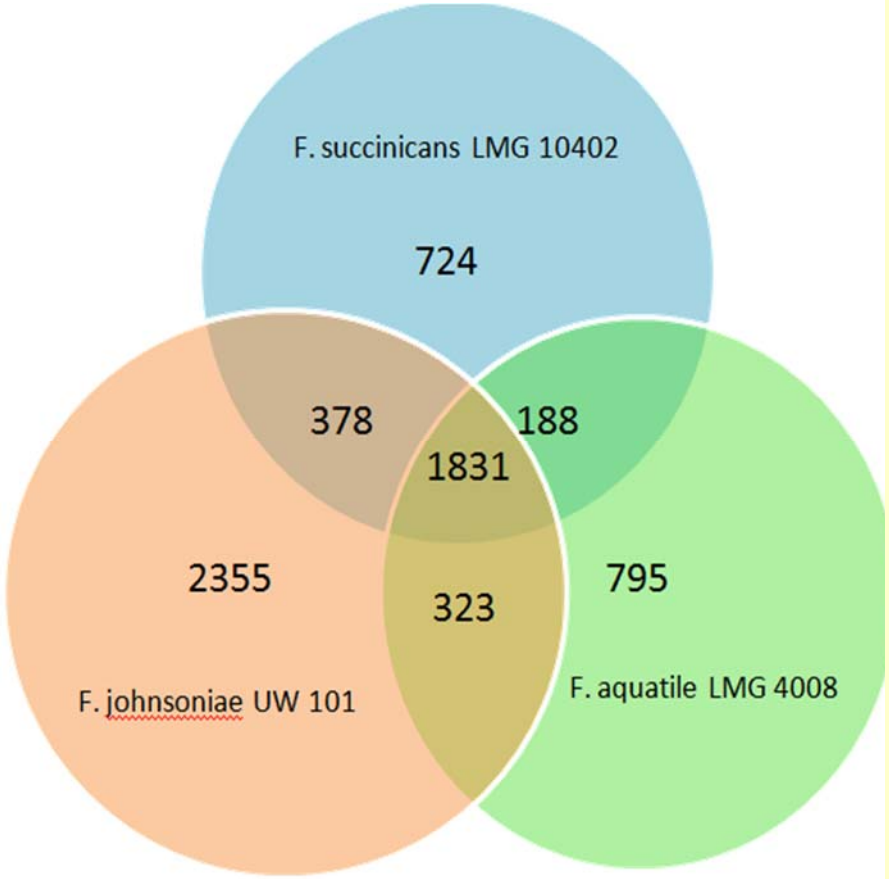


Microsoft Excel - Fsucc Venn for poster.xlsx

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 Diagrams

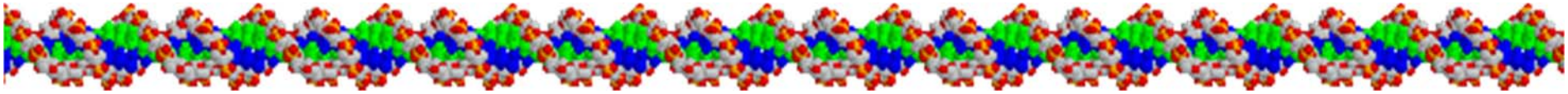


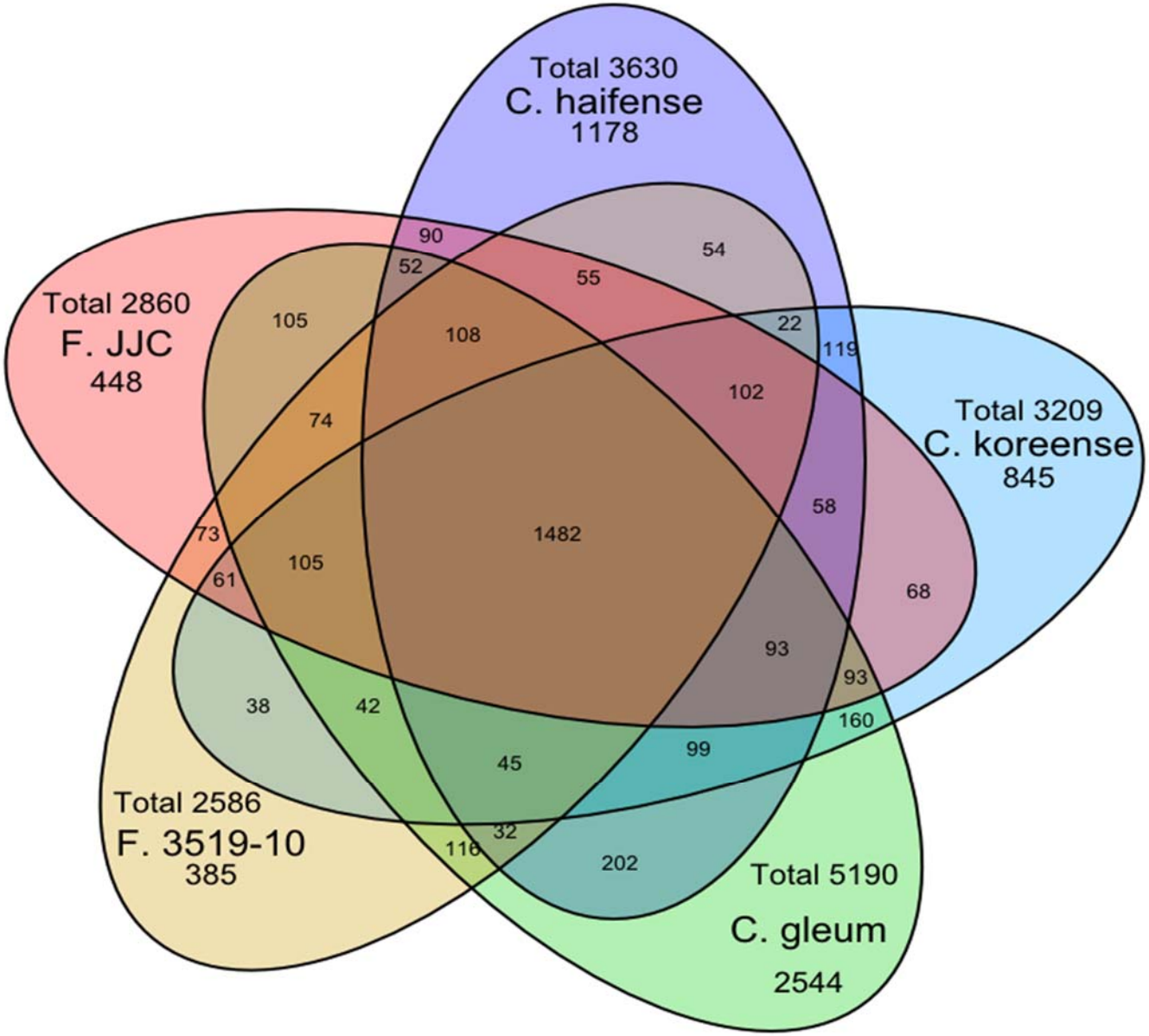


F. succinicans specific proteins

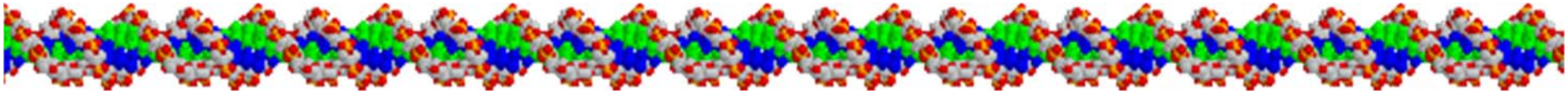
Gene#	Function	Gene#	Function
38	Possible restriction endonuclease	2026	4-hydroxyproline epimerase (EC 5.1.1.8)
191	Type I restriction-modification system, M subunit	2027	D-amino acid dehydrogenase small subunit (EC 1.4.99.1)
402	Type I restriction-modification system, specificity subunit S	2034	Ornithine cyclodeaminase (EC 4.3.1.12)
56	Mobile element protein	2035	putative secreted tripeptidyl aminopeptidase
202	COG1242: Predicted Fe-S oxidoreductase	2051	Phage protein
262	Succinoglycan biosynthesis protein	2073	Ubiquitin-protein ligase
303	Formate--tetrahydrofolate ligase (EC 6.3.4.3)	2117	Conjugative transposon protein TraB
373	Cyanophycinase (EC 3.4.15.6)	2133	Conjugative transposon protein TraO
384	Homoserine kinase (EC 2.7.1.39)	2144	reticulocyte binding protein 2 homolog a
451	Ubiquinol-cytochrome C reductase iron-sulfur subunit	2148	Phosphatidylserine/glycerophosphate/cardiolipin synthases
454	Nitrate/nitrite transporter	2321	Phosphate ABC transporter, periplasmic phosphate-BP PstS
460	Cytochrome c552 precursor (EC 1.7.2.2)	2322	Phosphate transport system permease protein PstC
461	Cytochrome c nitrite reductase, small subunit NrfH	2323	Phosphate transport system permease protein PstA
475	B12 binding/kinase domain Methylmalonyl-CoA mutase	2325	Phosphate transport system regulatory protein PhoU
695	High-affinity choline uptake protein BetT	2504	CRISPR-associated protein, Csn1 family
778	FIG022160: hypothetical toxin	2507	CRISPR-associated protein Cas2
779	FIG045511: hypothetical antitoxin (to FIG022160)	2508	CRISPR-associated protein Cas1
816	Thymidylate kinase (EC 2.7.4.9)	2556	Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)
833	S-adenosylmethionine decarboxylase related	2725	Metallo-beta-lactamase family protein, RNA-specific
837	Spermidine synthase (EC 2.5.1.16)	2736	Mobile element protein
841	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	2737	Mobile element protein
844	Prophage antirepressor	2818	putative sodium-dependent bicarbonate transporter
865	L-lactate permease	2914	Anaerobic C4-dicarboxylate transporter
1019	Deacetylases, acetoin utilization protein	2938	Predicted D-lactate dehydrogenase, Fe-S protein, FAD/FMN
1059	Put. metal chaperone, involved in Zn homeostasis, GTPase	2943	Isoaspartyl dipeptidase (EC 3.4.19.5) @ Asp-X dipeptidase
1084	Phosphodiesterase/alkaline phosphatase D	2944	Anaerobic C4-dicarboxylate transporter DcuC
1402	Methyltransferase (EC 2.1.1.-)	2958	Sialic acid-induced transmembrane protein/ mutarotase
1520	Transposase	3059	putative purple acid phosphatase precursor
1580	Protoporphyrinogen IX oxidase, oxygen-independent, HemG	3122	Metallo-beta-lactamase superfamily protein PA0057
1582	transcriptional regulator, XRE family	3134	probable polysaccharide biosynthesis transport protein
1584	DNA helicase II related protein	3136	Putative secreted polysaccharide polymerase
1838	Tyrosinase precursor (EC 1.14.18.1)	3137	Biotin carboxylase (EC 6.3.4.14)
1863	Glycosyltransferase (EC 2.4.1.-)	3140	Glycosyl transferase, group 1
1905	LSU ribosomal protein L36p	3145	Glucose-methanol-choline (GMC) oxidoreductase:NAD
1924	Nucleotidyltransferase (EC 2.7.7.-)	3154	Oxalate/formate antiporter
1925	DNA polymerase, beta-like region	3156	Pyruvate formate-lyase (EC 2.3.1.54)
1928	ATP-dependent DNA helicase	3224	Phage DNA replication protein
1929	ATP-dependent DNA helicase	3225	Phage external scaffolding protein #Protein D
1963	Methyltransferase FkbM	3226	Phage DNA binding protein
1964	Putative mannosyltransferase involved in polysacc. biosyn.	3227	Phage major capsid protein
2015	Endonuclease	3228	Phage major spike protein
		3229	Phage minor capsid protein - DNA pilot protein

Identify Core, Genus or Family-Specific Genes





Determine Phylogenomic Metrics - GGDC



GGDC

Genome-to-Genome Distance Calculator

DSMZ

About | GGDC 1.0 | GGDC 2.0 | FAQ | Contact | Legal Notice

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.

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:

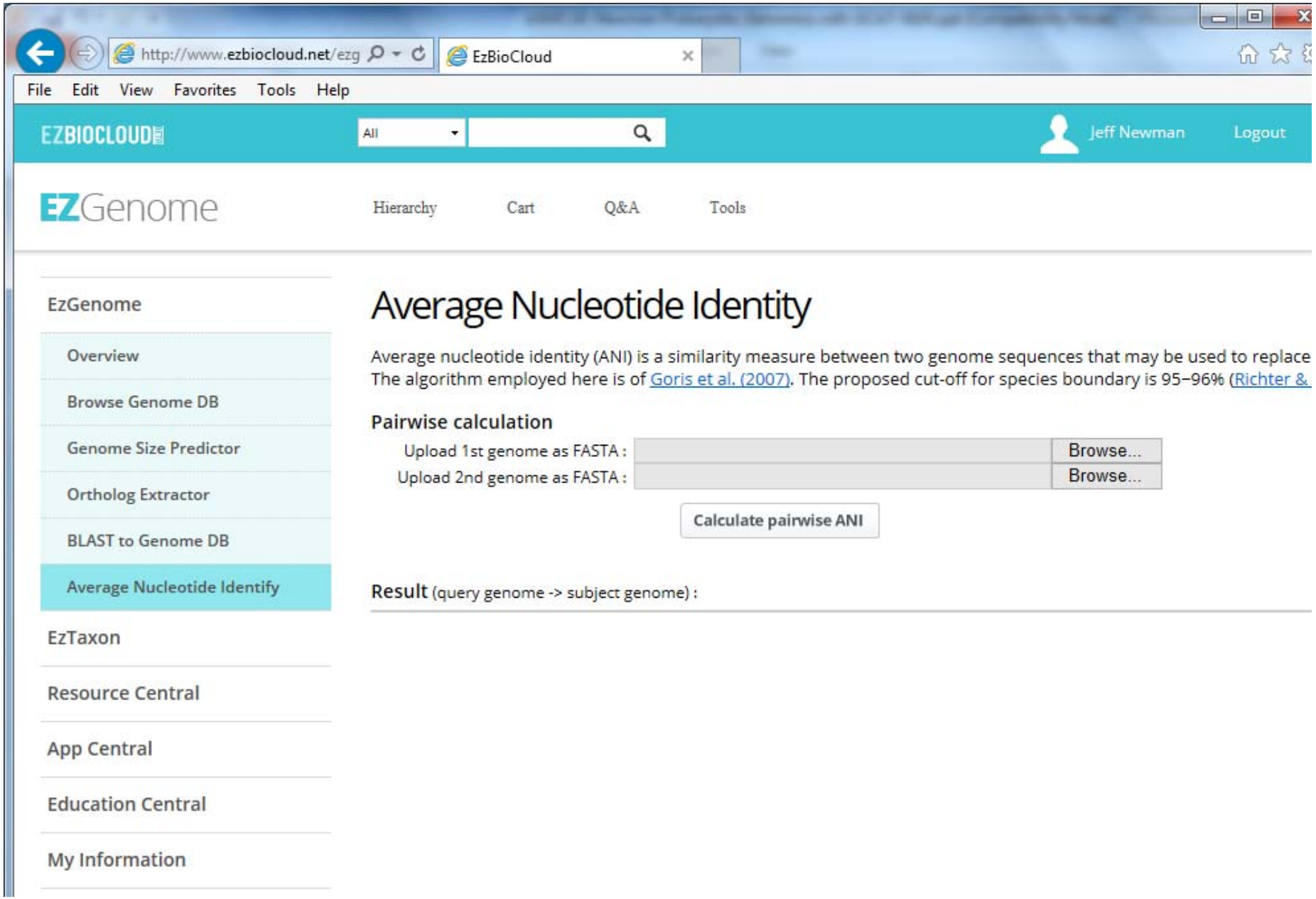
§ ANI options

Alignment options

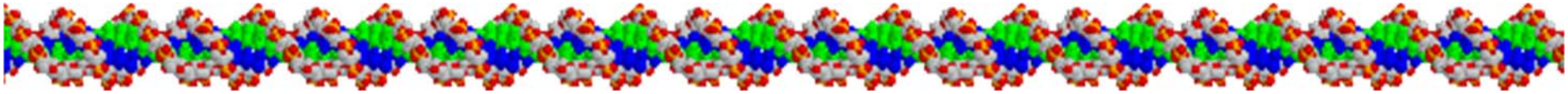
Fragment options

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

Determine Phylogenomic Metrics – Chun Lab EzGenome ANI Calculator



Determine Phylogenomic Metrics

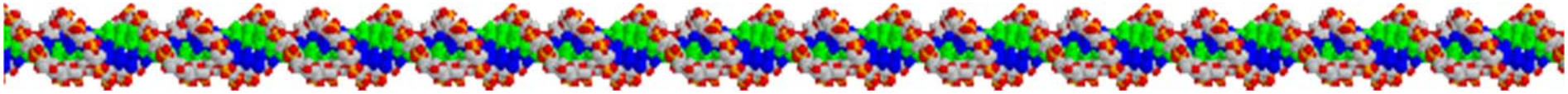


- Threshold for same vs different species is
  - 70% DDH (http://ggdc.dsmz.de/)
  - 95% ANI (http://enve-omics.ce.gatech.edu/ani/)
  - 95% AAI (http://lycofs01.lycoming.edu/~newman/ROSA/OrthologyScore.htm)
  - 65 ROSA (http://lycofs01.lycoming.edu/~newman/ROSA/index.htm)

	Average Nucleotide Identity (ANI)			GGDC DDHest		
	1	2	3	1	2	3
1. *Bacillus indicus* LMG 22858^T						
2. *Bacillus cibi* DSM 16189^T	98.27			80.40		
3. *Lycobacillus colbertis* SJS sp. nov.	80.41	80.40		19.30	19.30	

Average Nucleotide Identity (ANI) and Estimated DNA-DNA Hybridization

Determine Phylogenomic Metrics – NewmanLab AAI, %BBH, OS, ROSA Calculator



RAST Server - Job De... | Seed Viewer - Multi-... | Seed Viewer - Blast ... | Seed Viewer - Blast ... | Newman Lab ROSA ...

lycofs01.lycoming.edu/~newman/rosa/

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

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... No files selected.

Once the files have loaded, the Submit button will appear.

This .zip file contains set of sample .tsv files (must be unzipped) to test the ROSA calculator.

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 article:
Krebs, J.E., Gale, A.N., Anspach, T.J., Sontag, T.C., Keyser, V.K., Kirk, K.E., Peluso, E.M., and Newman J.D (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.

Average Amino Acid Identity (AAI) & 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.

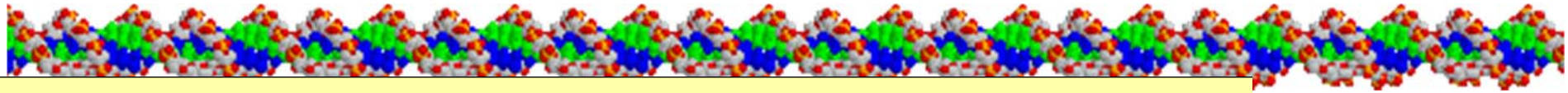
Browse... Flavobacterium chilense LMG 26360.tsv
Submit Reset

Orthology Analysis

	Genome ID	AAIr	% BBH	OS
ref	946677.5			
1	1453505.3	80.52320313223849	64.73281224797111	41.972666406555646
2	1453498.5	66.97895740021758	43.34503960547669	19.445368161076843
3	1341154.4	66.25092324987445	40.195293461355035	17.642457235669315
4	1107311.4	65.6102782103684	40.04610691236834	17.23868210961846
5	37752.3	85.6201182922618	70.15191755958291	51.42700038998396
6	991.3	77.68355325494804	68.21575333093035	41.16639564074482
7	376686.16	83.14456074619576	75.62674946267803	52.2809078931338
8	1341181.5	65.03299105160328	41.04196497668159	17.357836898292163
9	362418.3	82.38634513064508	68.5219875468246	46.50936663739819
10	1450525.4	71.50145565729076	47.735852462538865	24.404754849972214

A Perl script for this analysis that takes the filename on the command line and produces the results above (tab-delimited on standard o

Average Amino Acid Identity (AAI) & Reciprocal Orthology Score Average (ROSA)



Matrices

Average Amino Acid Identity (AAIr)	1453498.5	1453505.3	1341154.4	946677.5	1107311.4	37752.3	991.3	376686.16	1341181.5	362418.3	1450525.4
1453498.5		67.023	68.804	66.979	67.824	67.462	66.821	66.396	67.994	66.722	67.463
1453505.3	66.818		66.055	80.523	65.398	81.337	76.264	79.933	65.194	80.729	70.325
1341154.4	68.921	66.244		66.251	82.497	65.970	66.303	65.347	84.250	65.424	66.953
946677.5	67.052	80.710	66.604		65.654	85.689	77.687	83.126	64.782	82.159	71.768
1107311.4	68.043	65.220	82.507	65.610		65.368	66.117	64.829	85.135	64.809	66.427
37752.3	67.337	81.477	66.241	85.620	65.734		79.388	82.535	66.202	82.181	71.071
991.3	66.707	76.513	66.539	77.684	66.469	79.446		76.224	65.760	76.126	72.126
376686.16	66.641	80.394	66.075	83.145	65.732	82.802	76.616		65.186	82.100	70.978
1341181.5	68.324	65.470	84.330	65.033	85.266	65.988	65.897	64.855		64.414	66.575
362418.3	66.951	80.994	65.951	82.386	65.552	82.440	76.179	82.338	64.997		71.121
1450525.4	67.409	70.500	66.965	71.501	66.469	71.045	72.025	70.608	66.553	70.923	

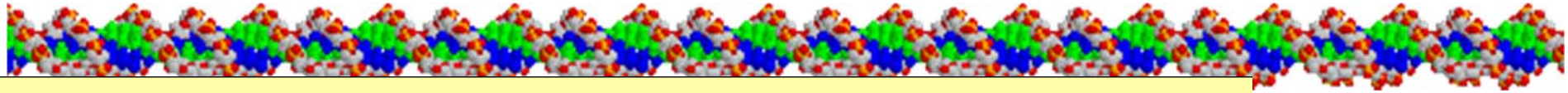
Percent Bidirectional Best Hit (% BBH)	1453498.5	1453505.3	1341154.4	946677.5	1107311.4	37752.3	991.3	376686.16	1341181.5	362418.3	1450525.4
1453498.5		44.466	74.491	43.345	67.545	47.710	44.411	42.109	68.719	45.599	66.023
1453505.3	69.513		73.368	64.733	68.304	70.040	59.482	65.507	67.765	71.106	75.041
1341154.4	66.304	41.827		40.195	73.919	44.792	41.510	39.864	73.876	42.870	60.647
946677.5	71.462	68.345	73.579		69.131	78.070	69.530	73.291	70.140	73.238	76.776
1107311.4	64.099	42.056	79.224	40.046		45.182	41.340	39.426	78.258	42.306	60.214
37752.3	70.909	66.317	73.987	70.152	69.322		63.918	67.181	67.535	69.125	77.141
991.3	72.199	61.449	74.858	68.216	69.342	69.831		64.251	69.677	64.352	76.579
376686.16	71.792	70.970	75.000	75.627	68.885	77.025	67.397		69.024	77.199	75.712
1341181.5	66.384	42.007	80.794	41.042	79.553	44.617	41.716	39.535		42.731	62.417
362418.3	70.195	70.293	73.436	68.522	67.222	71.993	61.663	69.829	66.911		72.264
1450525.4	67.455	48.944	69.261	47.736	64.047	53.095	48.360	45.408	65.234	47.749	

37752.3	24.513	24.630	25.979	26.933	32.882						
991.3	24.123	24.354	25.696	25.979	32.463	42.180					
1453505.3	23.404	23.551	25.183	25.505	30.719	45.181	35.285				
946677.5	23.397	23.519	25.141	25.787	31.975	54.375	41.565	43.247			
362418.3	22.999	23.328	25.145	25.882	30.286	47.807	36.539	46.227	47.973		
376686.16	22.980	23.167	24.884	25.223	30.390	49.287	38.446	43.862	51.462	49.688	

	376686.16	1341181.5	362418.3	1450525.4
	22.980			
	49.688	22.999		
	30.390	28.280	30.286	

	2418.3	376686.16

Average Amino Acid Identity (AAI) & Reciprocal Orthology Score Average (ROSA)



Matrices

Average Amino Acid Identity (AAIr)	1453498.5	1453505.3	1341154.4	946677.5	1107311.4	37752.3	991.3	376686.16	1341181.5	362418.3	1450525.4
1453498.5		67.023	68.804	66.979	67.824	67.462	66.821	66.396	67.994	66.722	67.463
1453505.3	66.818		66.055	80.523	65.398	81.337	76.264	79.933	65.194	80.729	70.325
1341154.4	68.921	66.244		66.251	82.497	65.970	66.303	65.347	84.250	65.424	66.953
946677.5	67.052	80.710	66.604		65.654	85.689	77.687	83.126	64.782	82.159	71.768
1107311.4	68.043										
37752.3	67.337										
991.3	66.707										
376686.16	66.641										
1341181.5	68.324										
362418.3	66.951										
1450525.4	67.409										

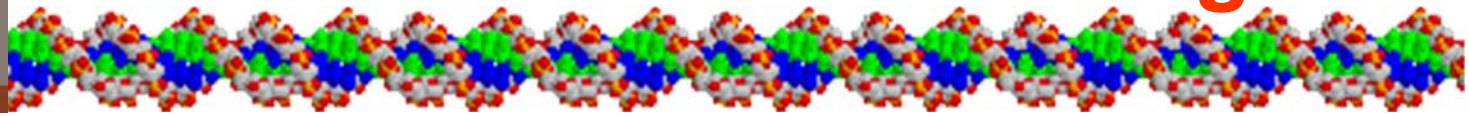
Percent Bidirectional Best Hit (% BBH)	14534
1453498.5	
1453505.3	69.5
1341154.4	66.3
946677.5	71.4
1107311.4	64.0
37752.3	70.9
991.3	72.1
376686.16	71.7
1341181.5	66.3
362418.3	70.1
1450525.4	67.4

Reciprocal Orthology Score Average (ROSA)	1453498.5	1453505.3	1341154.4	946677.5	1107311.4	37752.3	991.3	376686.16	1341181.5	362418.3	1450525.4
1453498.5											
1453505.3	25.505										
1341154.4	33.379	25.183									
946677.5	25.787	43.247	25.141								
1107311.4	30.374	23.551	52.119	23.519							
37752.3	26.933	45.181	25.979	54.375	24.630						
991.3	25.979	35.285	25.696	41.565	24.354	42.180					
376686.16	25.223	43.862	24.884	51.462	23.167	49.287	38.446				
1341181.5	31.380	23.404	54.948	23.397	57.279	24.513	24.123	22.980			
362418.3	25.882	46.227	25.145	47.973	23.328	47.807	36.539	49.688	22.999		
1450525.4	30.350	30.719	29.122	31.975	27.433	32.882	32.463	30.390	28.280	30.286	

ROSA (sorted)	1341181.5	1107311.4	1341154.4	1453498.5	1450525.4	37752.3	991.3	1453505.3	946677.5	362418.3	376686.16
1341181.5											
1107311.4	57.279										
1341154.4	54.948	52.119									
1453498.5	31.380	30.374	33.379								
1450525.4	28.280	27.433	29.122	30.350							
37752.3	24.513	24.630	25.979	26.933	32.882						
991.3	24.123	24.354	25.696	25.979	32.463	42.180					
1453505.3	23.404	23.551	25.183	25.505	30.719	45.181	35.285				
946677.5	23.397	23.519	25.141	25.787	31.975	54.375	41.565	43.247			
362418.3	22.999	23.328	25.145	25.882	30.286	47.807	36.539	46.227	47.973		
376686.16	22.980	23.167	24.884	25.223	30.390	49.287	38.446	43.862	51.462	49.688	

asm2014
114th General Meeting | American Society for Microbiology
May 17–20, 2014 | Boston, Massachusetts

Come see our posters at the ASM General Meeting!



Sunday May 18, 2014 10:45 AM - 12:00 PM

- #436 Andrew Gale, Jordan Krebs, Eileen Peluso, Jeff Newman - Integration of Average Amino Acid Identity (AAI) and Percentage of Orthologous Genes in a Single Phylogenomic Metric, the Reciprocal Orthology Score Average (ROSA).

Tuesday May 20, 2014 12:30 PM - 1:45 PM

- #2850 Samantha Stropko & Jeff Newman, Identification and Characterization of *Lycobacillus* colbertis SJS gen. nov, sp. nov.
- #2852 Tori Bortniak & Jeff Newman, Characterization of *Chryseobacterium populense* sp. nov.
- #2853 Tom Sontag & Jeff Newman, *Chryseobacterium haifense*: A Genomic Report
- #2854 Ashley Gimbel & Jeff Newman, Classification and Identification of *Flavobacterium douthatii* ABG sp. nov.
- #2855 Kyle Swovick & Jeff Newman, *Flavobacterium succinicans*: A Genomic Report.

Friday & Saturday, November 7 & 8, 2014

- Allegheny Branch of the ASM meeting at Lycoming

Thank You!

HHMI

NSF

SEEKquence