



ANI-based analysis tools in IMG

Neha Varghese
njvarghese@lbl.gov
MGM-23

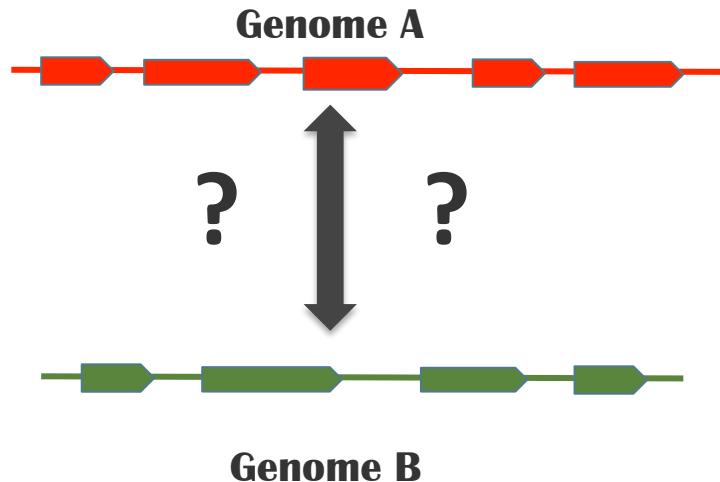




Introduction to ANI



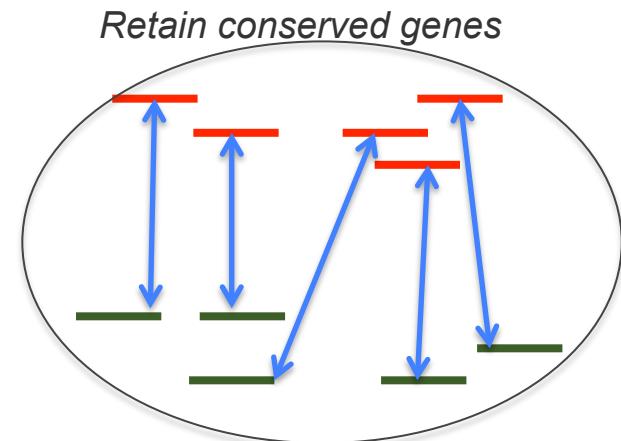
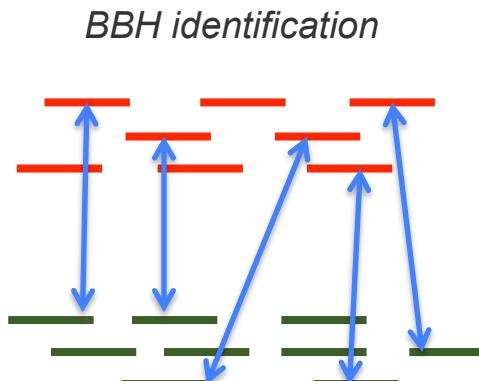
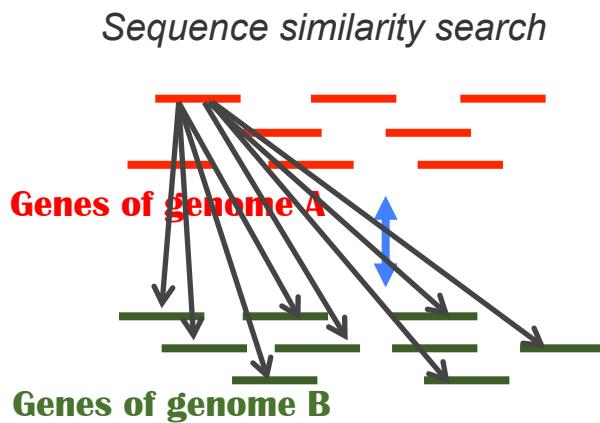
What is it?



**ANI provides a quantitative measure of
how related two genomes are on a genomic level.**

Thus, it is a means to compare the genetic relatedness among strains.

How is it computed?



$$AF = \frac{\text{lengths of BBH genes}}{\sum \text{length of genes in genome 1}}$$

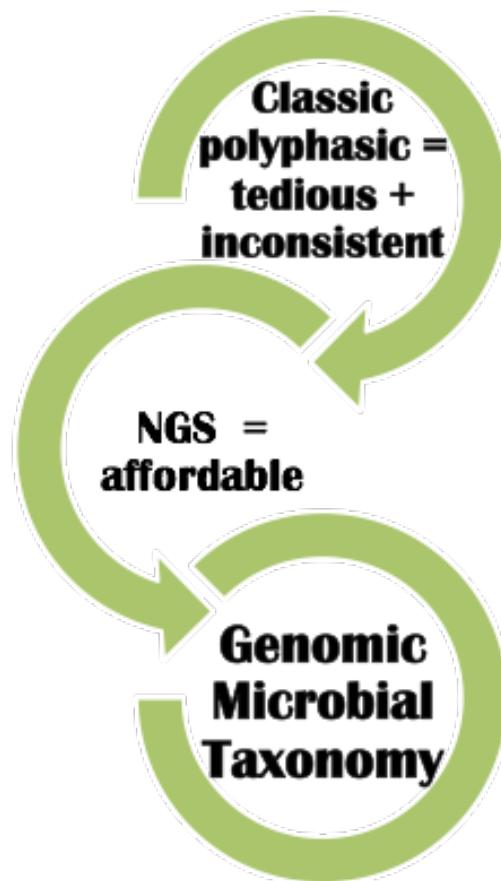
$$gANI = \frac{\sum_{bbh} (\text{Percent Identity} * \text{Alignment length})}{\text{lengths of BBH genes}}$$

Alignment fraction (AF) is the fraction of genes that are orthologous between two genomes.

Average Nucleotide Identity (ANI) is a measure of nucleotide-level genomic similarity between the conserved coding regions of two genomes

Why do we care?

- Microbial taxonomic assignments are inconsistent with emerging genetic, systematic, and phenotypic information for a large number of species



We believe,



The **AF and ANI** between two genomes,
accurately reflects
the degree of evolutionary distance
and should be used as the primary guide for
taxonomic species assignment,
supplementing the existing polyphasic approach .

Why use IMG for ANI based analysis?



The MiSI method (Microbial Species Identifier)

<https://ani.jgi-psf.org>

*Microbial species delineation using whole genome sequences.
Varghese et al. NAR 2015;doi: 10.1093/nar/gkv657*

Compute ANI,AF for all vs all.

genome1
genome2
genome3
genome4
genome5

		ANI(1>2)	ANI(2>1)	AF(1>2)	AF(2>1)
Genome1	Genome2	97.79	97.81	0.85	0.81
Genome1	Genome3	92.5	96.5	0.87	0.83
Genome1	Genome4	98.87	98.88	0.41	0.88
Genome1	Genome5	97.51	97.51	0.85	0.83
Genome2	Genome3	89.9	90.45	0.86	0.89
Genome2	Genome4	70.42	76.43	0.84	0.86
Genome2	Genome5	97.9	97.5	0.85	0.85
Genome3	Genome4	97.84	97.84	0.38	0.25
Genome3	Genome5	97.47	97.46	0.88	0.87

Cluster

*Identify pairs above predetermined thresholds
AF >= 0.6 and ANI >= 96.5*

Why use IMG for ANI based analysis?



The MiSI method (Microbial Species Identifier)

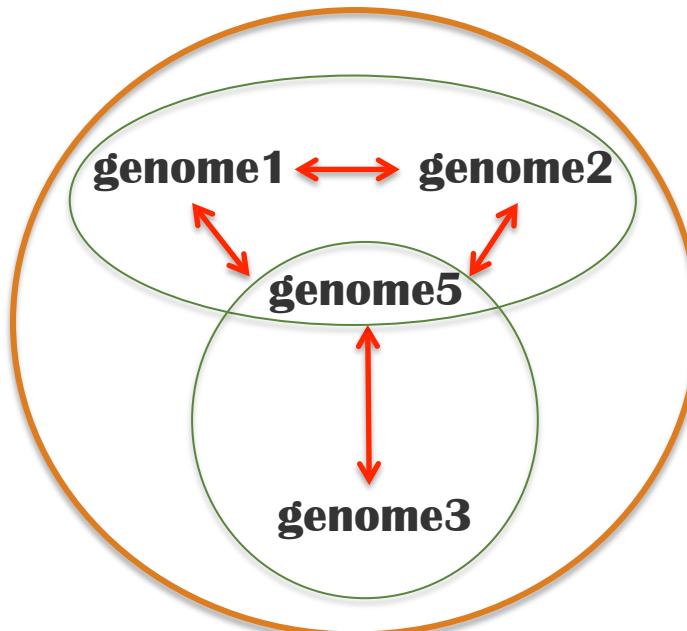
<https://ani.jgi-psf.org>

*Microbial species delineation using whole genome sequences.
Varghese et al. NAR 2015;doi: 10.1093/nar/gkv657*

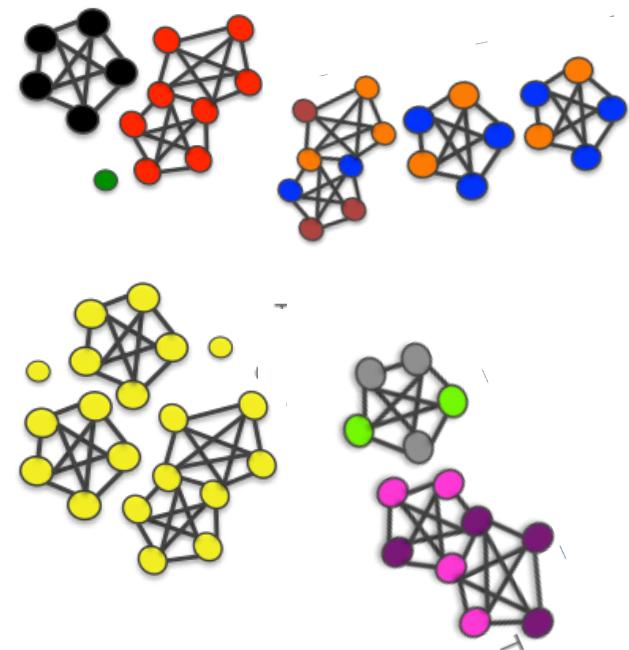
Genome pairs

Genome1	Genome2
Genome1	Genome5
Genome2	Genome5
Genome3	Genome5

Maximal Clique Enumeration

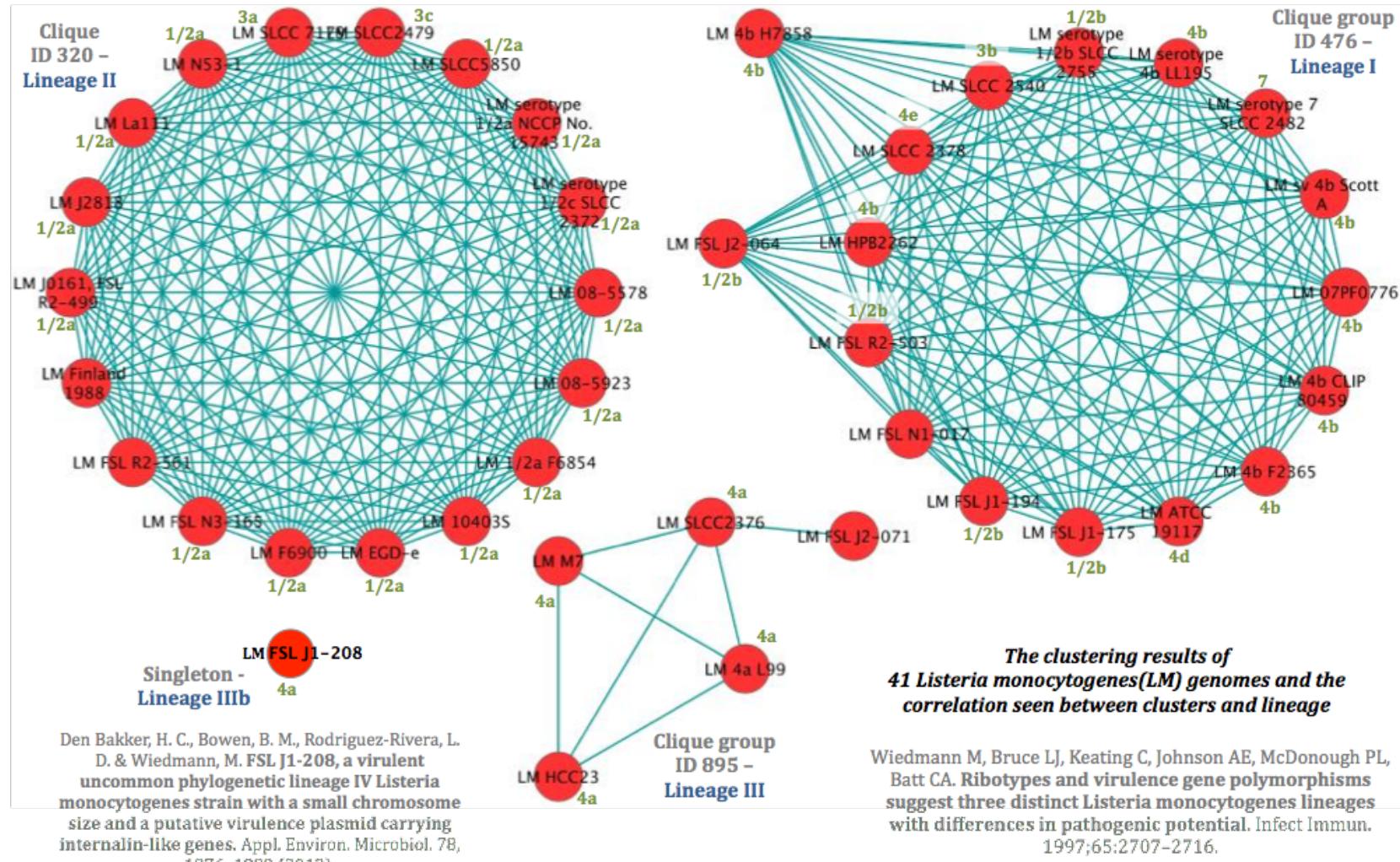


Cliques and clique-groups and singletons



Why use IMG for ANI based analysis?

The MiSI method (Microbial Species Identifier)



How can you use it?



- MiSI is fast, scalable, objective and an accurate reflection of the genetic similarity between a pair of genomes or a large set of genomes.
- This method can be applied to
 - Genome from metagenome
 - Draft genomes in up to 2500 scaffolds
 - Genomes reduced up to 25%
 - Genomes without a 16s
 - Previously annotated genomes without a species definition
- MCE implementation allows:
 - To determine if a genome of interest belongs to an existing species or is a novel candidate species
 - Provides the ability to sharpen our view of the evolutionary relationships between genomes

./ANICalculator -help

This tool will calculate the bidirectional average nucleotide identity (gANI) between two genomes.
Required input is the full path to the fna file (nucleotide sequence of genes in fasta format) of each query genome.

Either the rRNA and tRNA genes can be excluded, or provided in a list with the -ignoreList option.

Usage:./ANICalculator

- genome1fna <fna file of the first query genome> *REQUIRED*
- genome2fna <fna file of the second query genome> *REQUIRED*
- outfile <the output file> OR -stdout <output to screen> *Default: ANICalculator.out*
- outdir <output directory> *Default: Current directory*
- ignoreList <file containing list of genes to ignore (Should include TRNA and RNA genes)>
- logfile <log file> *Default: ANICalculator.log*
- help <prints this page>

Results are shown in tab-delimited format with following headers:

Genome1 <name of FNA file of genome1>

Genome2 <name of FNA file of genome2>

ANI(1->2) <Average nucleotide identity of the first genome to the second>

ANI(2->1) <Average nucleotide identity of the second genome to the first>

AF(1->2) <Alignment Fraction of the first genome to the second>

AF(2->1) <Alignment Fraction of the second genome to the first>



Tutorial

ANI analysis tools in IMG

MGM-23



How to get to ANI in IMG?

Home Find Genomes Find Genes Find Functions

Compare Genomes OMICS Workspace My IMG Data Marts Help

IMG/ER Content

Datasets	JGI	All
Bacteria	6112	42031
Archaea	361	1147
Eukarya	31	257
Plasmids	1	1220
Viruses	5182	
Genome Fragments	1201	
Metagenome	4701	10087
Total Datasets	61125	
My Private Datasets	12436	

Last Datasets Added On:

Genome	2016-04-13
Metagenome	2016-04-24

Project Map
[Metagenome Projects Map](#)
[System Requirements](#)

 Hands on training available at the [Microbial Genomics & Metagenomics Workshop](#)

The Integrated Microbial Genomes analysis and annotation of genome comparative context. The IMG contains datasets provided by IMG users and metagenome datasets.

IMG/ER provides users with two types of access: protected (password protected) genome data and/or metagenome datasets ([here](#)) and public (free access) genome and metagenome datasets ([here](#)).

IMG/ER Statistics

IMG/ER contains 244 public study datasets. These datasets follow:

Engineered	581
Bioreactor	16
Bioremediation	52
Biotransformation	28
Food production	3
Lab enrichment	102
Lab synthesis	3
Modeled	1
Solid waste	29
Unclassified	22
Wastewater	325

2016-04-25-08.00.01

Compare Genomes

Genome Statistics Synteny Viewers Abundance Profiles Phylogenetic Dist. Avg Nucleotide Ident. Pairwise ANI Distance Tree Same Species Plot Function Profile ANI Cliques Genome Clustering Genome Gene Best HmLgs Phylo. Marker COGs

community resource for a comprehensive genome and metagenome dataset available in their private (password protected) ([content/42/D1/D560](#)) or public ([content/42/D1/D568](#)) in the context of all public (free access) genome and metagenome datasets.

Read more...

News

- [Feb 2016 Strategy to Uncover More Microbial Lineages](#)
- [Feb 16 2016 Incorrect Metagenome KO/EC Assignment](#)
- [Feb to Mar 2016 NERSC New Facility](#)
- [Feb 1 2016 JGI Uncovers "Kryptonite", A Novel Bacterial Phylum](#)
- [Dec 24 2015 to Jan 3 2016 Happy Holidays](#)
- [Oct 26 2015 IMG UI](#)
- [Oct 13 2015 IMG Genome Annotation Pipeline SOP Publication](#)
- [Oct 8 2015 IMG UI Changes](#)
- [Oct 5 2015 After 10 Years, IMG Still Revolutionizing Genomics](#)
- [Sep 2015 IMG ABC Data Mart](#)
- [Sep 2015 MGM Workshops](#)

Landing page for ANI analysis in IMG



Home Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG Data Marts Help

[Home](#) > Compare Genomes

loaded.

Average Nucleotide Identity (ANI)

Average Nucleotide Identity (ANI) is a measure of nucleotide-level genomic similarity between the coding regions of two genomes. First, BBHs between a genome pair are computed as pairwise bidirectional best nSimScan hits of genes having 70% or more identity and at least 70% coverage of the shorter gene.

The Alignment Fraction (AF) is computed as:

$$AF = \frac{\text{lengths of BBH genes}}{\sum_{G1 \rightarrow G2} \text{length of genes in genome 1}}$$

The Average Nucleotide Identity (ANI) is computed as:

$$gANI = \frac{\sum_{bbh} (\text{Percent Identity} * \text{Alignment length})}{\text{lengths of BBH genes}}$$

Statistics

Alignment Fraction (AF), Average Nucleotide Identity (ANI) has been computed for 45,925 genomes from 7,771 unique species and 548,599,113 genome pairs.

20,234,198 pairs of these are between genomes assigned to the same taxonomic species.

Maximal clique enumeration generates 16,920 clusters, including 31,478 genomes from 2,510 taxonomic species.

14,447 genomes from 5,818 species are singletons.

Clique Type	Count	Count w/ Type Strain *
Clique	2,324	960
Clique-group	149	72
Singleton	14,447	2,731

* - where type strain is "Yes"

Pairwise ANI

Same Species Plot

ANI Cliques

Last updated Sun Sep 18 2016

Getting to “Pairwise ANI”



Home Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG Data Marts Help

[Home](#) > Compare Genomes

Average Nucleotide Identity (ANI)

Average Nucleotide Identity (ANI) is a measure of nucleotide-level genomic similarity between two genomes. It is calculated as the percentage of bidirectional best nSightScan hits of genes having 70% or more identity and at least 500 bp length.

The Alignment Fraction (AF) is computed as:

$$AF = \frac{\text{lengths of BBH genes}}{\sum \text{length of genes in genome 1}}$$

The Average Nucleotide Identity (gANI) is computed as:

$$gANI = \frac{\sum AF_{G1 \rightarrow G2}}{\sum AF_{G2 \rightarrow G1}}$$

Statistics

Alignment Fraction (AF), Average Nucleotide Identity (ANI) has been computed for 20,234,198 pairs of genomes. 20,234,198 pairs of these are between genomes assigned to the same taxonomic rank. Maximal clique enumeration generates 16,920 clusters, including 31,478 genomes. 14,447 genomes from 5,818 species are singletons.

Clique Type	Count	Count w/ Type Strain *
Clique	2,324	960
Clique-group	149	72
Singleton	14,447	2,731

* - where type strain is "Yes"

Compare Genomes

- Genome Statistics
- Synteny Viewers
- Abundance Profiles
- Phylogenetic Dist.
- Avg Nucleotide Ident.
 - Pairwise ANI
- Distance Tree
- Function Profile
- Genome Clustering
- Genome Gene Best HmLgs
- Phylo. Marker COGs

Same Species Plot

ANI Cliques

Same Species Plot

ANI Cliques

[Pairwise ANI](#) [Same Species Plot](#) [ANI Cliques](#)

Pairwise ANI analysis in IMG



Home Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG Data Marts Help

[Home](#) > Compare Genomes

Pairwise ANI

BBHs between a genome pair are computed as pairwise bidirectional best nSimScan hits of genes having 70% or more identity and at least 70% coverage of the shorter gene. You may either select genome(s) from IMG or you may upload a nucleotide sequence in FASTA format (using the [Upload File](#) button) to compute ANI to selected genome(s) in IMG.

Please select up to 100 genomes:

Quick Search: <enter a genome name to search>

Sequencing Status

Finished

Domain

Archaea

List Tree

Show

Search for: <click show, then enter a genome name to search>

Acidianus hospitalis W1 (A) [F]
Acidilobus saccharovorans 345-15 (A) [F]
Aciduliprofundum boonei T469 (A) [F]
Aciduliprofundum sp. MAR08-339 (A) [F]
Aeropyrum camini SY1, JCM 12091 (A) [F]
Archaeoglobus fulgidus 7324, DSM 8774 (A) [F]
Archaeoglobus fulgidus VC-16, DSM 4304 (A) [F]
Archaeoglobus fulgidus VC-16, DSM 4304 (A) [F]
Archaeoglobus profundus Av18, DSM 5631 (A) [F]
Archaeoglobus sulfaticallidus PM70-1, DSM 19444 (Asulf_version1) (A)
Archaeoglobus veneficus SNP6, DSM 11195 (A) [F]
Caldisphaera lagunensis IC-154, DSM 15908 (A) [F]

Selected Genomes

Pairwise 1:

loaded.
✓ Finished
Permanent Draft
Draft
All Finished, Permanent Draft and Draft

✓ Archaea
Bacteria
Eukaryota
GFragment
Plasmid
Viruses
Genome Cart

Pairwise 2:

Add Upload Sets Remove Upload File

Add Upload Sets Remove

ANI

Pairwise ANI analysis in IMG

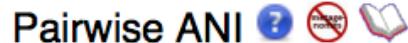
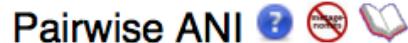


Home Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG Data Marts Help

[Home](#) > Compare Genomes

loaded.

Pairwise ANI



BBHs between a genome pair are computed as pairwise bidirectional best nSimScan hits of genes having 70% or more identity and at least 70% coverage of the shorter gene. You may either select genome(s) from IMG or you may upload a nucleotide sequence in FASTA format (using the Upload File button) to compute ANI to selected genome(s) in IMG

Please select up to 100 genomes:

Quick Search: <enter a genome name to search>

Sequencing Status Domain
Finished Archaea
 List Tree Show ?

Search for: <click show, then enter a genome name to search>

Selected Genomes

Pairwise 1:

Add Upload Sets Remove Upload File

Pairwise 2:

Add Upload Sets Remove

My Workspace - Genome Sets

Please select some genome sets:

No workspace genome sets.

Submit Cancel

ANI

Getting to “Same species plot”

Home Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG Data Marts Help

Home > Compare Genomes

Average Nucleotide Identity (ANI)

Average Nucleotide Identity (ANI) is a measure of nucleotide-level genomic similarity between two genomes. It is computed as the bidirectional best nSimScan hits of genes having 70% or more identity and at least 50% length.

The Alignment Fraction (AF) is computed as:

$$AF = \frac{\text{lengths of BBH genes}}{\sum \text{length of genes in genome 1}}$$

The Average Nucleotide Identity (gANI) is computed as:

$$gANI = \frac{\sum \text{length of genes in genome 1}}{\sum \text{length of genes in genome 2}}$$

Statistics

Alignment Fraction (AF), Average Nucleotide Identity (ANI) has been computed for 20,234,198 pairs of genomes. 20,234,198 pairs of these are between genomes assigned to the same taxonomic rank. Maximal clique enumeration generates 16,920 clusters, including 31,478 genome pairs. 14,447 genomes from 5,818 species are singletons.

Clique Type	Count	Count w/ Type Strain *
Clique	2,324	960
Clique-group	149	72
Singleton	14,447	2,731

* - where type strain is "Yes"

Compare Genomes

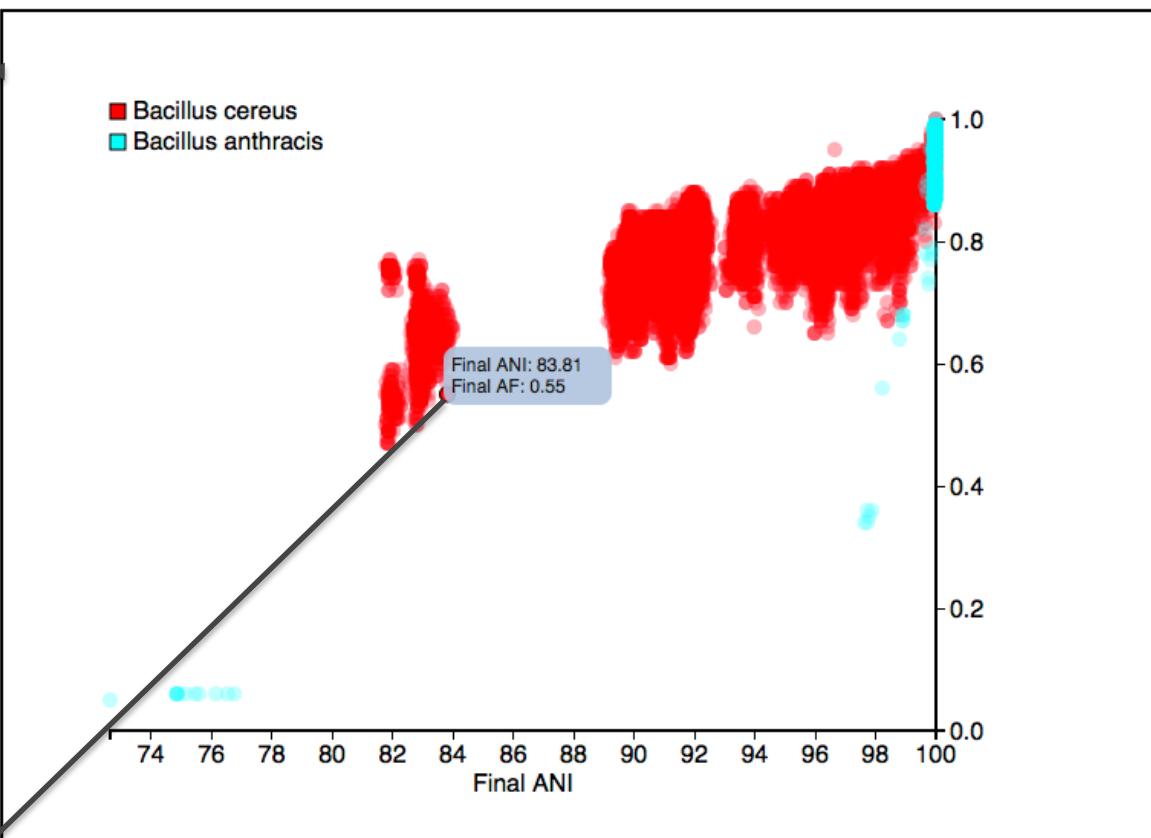
- Genome Statistics
- Synteny Viewers
- Abundance Profiles
- Phylogenetic Dist.
- Avg Nucleotide Ident. ▾
- Distance Tree
- Function Profile
- Genome Clustering
- Genome Gene Best HmLgs
- Phylo. Marker COGs

- Pairwise ANI
- Same Species Plot
- ANI Cliques

ANI Same Species Plot

Please select no more than 10 species to analyze the genomes in each species using pairwise ani:
Domain: Bacteria Species: *Bacillus anthracis* Add Species

Bacillus cereus *Bacillus anthracis*



Set of Selected Genomes from Plot

- Species: *Bacillus cereus* | Taxon: 640753006
- Species: *Bacillus cereus* | Taxon: 2537561969

Getting to “ANI Cliques”

The diagram illustrates the navigation path from the JGI homepage to the ANI Cliques feature. A large black arrow points from the main title "Getting to ‘ANI Cliques’" down to the "Compare Genomes" section of the JGI interface. From there, another black arrow points to the "Avg Nucleotide Ident." link, which is highlighted with a dark blue background. Finally, a third black arrow points to the "ANI Cliques" button at the bottom of the page.

Home Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG Data Marts Help

[Home](#) > Compare Genomes

Average Nucleotide Identity (ANI)

Average Nucleotide Identity (ANI) is a measure of nucleotide-level genomic similarity between two genomes. It is calculated as the alignment fraction (AF) of bidirectional best nSimScan hits of genes having 70% or more identity and at least 500 bp length.

The Alignment Fraction (AF) is computed as:

$$AF = \frac{\text{lengths of BBH genes}}{\sum \text{length of genes in genome 1}}$$

The Average Nucleotide Identity (gANI) is computed as:

$$gANI = \frac{\sum AF_{G1 \rightarrow G2}}{\sum AF_{G2 \rightarrow G1}}$$

Statistics

Alignment Fraction (AF), Average Nucleotide Identity (ANI) has been computed for 20,234,198 pairs of genomes. 20,234,198 pairs of these are between genomes assigned to the same taxonomic rank. Maximal clique enumeration generates 16,920 clusters, including 31,478 genome pairs. 14,447 genomes from 5,818 species are singletons.

Clique Type	Count	Count w/ Type Strain *
Clique	2,324	960
Clique-group	149	72
Singleton	14,447	2,731

* - where type strain is "Yes"

[Pairwise ANI](#) [Same Species Plot](#) **ANI Cliques**

ANI cliques in IMG – All cliques



Home Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG Data Marts Help

[Home](#) > Compare Genomes

16937 cliques and 11596 species loaded

ANI Cliques

All Cliques [by Species](#) [by Taxonomy](#) [Clique Groups](#)

Filter column: Clique ID [Filter](#) text [Apply](#) [?](#)

[Export](#) Page 1 of 170 << first < prev 1 [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) next > last >> 100

Column Selector

✓ Clique ID
Clique Type
Intra-Clique ANI
Intra-Clique AF
Contributing Species
Genome Count
All Columns

Clique ID	Clique Type	Intra-Clique ANI	Intra-Clique AF	Contributing Species	Genome Count
1	clique-group	97.84	0.818	Escherichia coli, Escherichia sp. 3_2_53FAA, Escherichia sp. 4_1_40B, Escherichia sp. TW10509, Escherichia sp. TW15838, Shigella boydii, Shigella dysenteriae, Shigella flexneri, Shigella sonnei, Shigella sp. 28760	2841
2	clique-group	98.88	0.907	Salmonella enterica	1490
3	clique-group	98.91	0.929	Staphylococcus aureus	4641
4	clique-group	98.55	0.861	Streptococcus pneumoniae	518
5	clique-group	98.39	0.866	Acinetobacter baumannii, Acinetobacter isolate CA1, Acinetobacter isolate CA2, Acinetobacter isolate CA3, Acinetobacter isolate CA4, Acinetobacter isolate CA5, Acinetobacter nosocomialis, Acinetobacter sp. NCTC 10304	973

ANI cliques in IMG – All cliques



Home Find Genomes Find Genes Find Functions Compare Genomes OMICS View

Home > Compare Genomes 16937 cliques

ANI Cliques

All Cliques by Species by Taxonomy Clique Groups

Filter column: Clique ID Filter text Apply ?

Export Page 1 of 170 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector

Clique ID	Clique Type	Intra-Clique ANI	Intra-Clique AF	Contributing Species	Genome Count
1	clique-group	97.84	0.818	Escherichia coli, Escherichia sp. 3_2_53FAA, Escherichia sp. 4_1_40B, Escherichia sp. TW10509, Escherichia sp. TW15838, Shigella boydii, Shigella dysenteriae, Shigella flexneri, Shigella sonnei, Shigella sp. 28760	2841
2	clique-group	98.88	0.907	Salmonella enterica	1490
3	clique-group	98.91	0.929	Staphylococcus aureus	4641
4	clique-group	98.55	0.861	Streptococcus pneumoniae	518
5	clique-group	98.39	0.866	Acinetobacter baumannii, Acinetobacter isolate CA1, Acinetobacter isolate CA2, Acinetobacter isolate CA3, Acinetobacter isolate CA4, Acinetobacter isolate CA5, Acinetobacter nosocomialis, Acinetobacter sp. NCTC 10304	973

Clique Details

Clique ID: 5

Overview Genomes in Clique Similar Cliques

Clique Information	
Clique ID	5
Clique Type	clique-group
Intra-Clique ANI	98.39
Intra-Clique AF	0.866
Total Genomes	973

ANI cliques in IMG - All cliques



Clique Details

Clique ID: 5

[Overview](#) [Genomes in Clique](#) [Similar Cliques](#)

Clique Information	
Clique ID	5
Clique Type	clique-group
Intra-Clique ANI	98.39
Intra-Clique AF	0.866
Genomes	973

Clique Details

Clique ID: 5

[Overview](#) [Genomes in Clique](#) [Similar Cliques](#)

Note: Similar cliques have gANI >= 90 with the current clique.

Filter column: Filter ?

Page 1 of 1 << first < prev 1 next > last >>

Column Selector

Clique ID	Clique Type	Average Identity with Clique 5	Average Gene Content Similarity with Clique 5	Contributing Species
159	clique	91.59	0.796	Acinetobacter baumannii, Acinetobacter genomospecies 13TU, Acinetobacter nosocomialis, Acinetobacter sp. RUH2624

Page 1 of 1 << first < prev 1 next > last >>

ANI cliques in IMG – All cliques



Clique Details

Clique ID: 5

Overview Genomes in Clique Similar Cliques

Add Selected to Genome Cart Select All Clear All Pairwise ANI

Filter column: Genome ID Filter text: Apply ?

Export Page 1 of 10 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector Select Page Deselect Page

Select	Genome ID	Genome Name	Species	Domain
<input type="checkbox"/>	640069301	Acinetobacter baumannii ATCC 17978	Acinetobacter baumannii	B
<input type="checkbox"/>	641522602	Acinetobacter baumannii AYE	Acinetobacter baumannii	B
<input type="checkbox"/>	641522603	Acinetobacter baumannii SDF	Acinetobacter baumannii	B
<input type="checkbox"/>	642555102	Acinetobacter baumannii ACICU	Acinetobacter baumannii	B
<input type="checkbox"/>	643348502	Acinetobacter baumannii AB0057	Acinetobacter baumannii	B
<input type="checkbox"/>	643348503	Acinetobacter baumannii AB307-0294	Acinetobacter baumannii	B
<input type="checkbox"/>	645058757	Acinetobacter baumannii AB900	Acinetobacter baumannii	B
<input type="checkbox"/>	647000204	Acinetobacter baumannii 6013150	Acinetobacter baumannii	B
<input type="checkbox"/>	647533101	Acinetobacter baumannii ATCC 19606, CIP 70.34	Acinetobacter baumannii	B
<input type="checkbox"/>	648276604	Acinetobacter baumannii AB056	Acinetobacter baumannii	B
<input type="checkbox"/>	650377901	Acinetobacter baumannii 1656-2	Acinetobacter baumannii	B
<input type="checkbox"/>	650377902	Acinetobacter baumannii TCDC-AB0715	Acinetobacter baumannii	B
<input type="checkbox"/>	2503754004	Acinetobacter isolate CA1	Acinetobacter isolate CA1	B

ANI cliques in IMG - All cliques



Clique Details

Clique ID: 5

Overview Genomes in Clique Similar Cliques

Add Selected to Genome Cart Select All Clear All Pairwise ANI

Filter column: Genome ID Filter text: Apply ?

Export Page 1 of 10 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector Select Page Deselect Page



Pairwise ANI ? 📖

Clique ID: 5

Filter column: Genome1 ID Filter text: Apply ?

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector

Genome1 ID ▲	Genome1 Name	Genome2 ID	Genome2 Name	ANI1->2	ANI2->1	AF1->2	AF2->1	Total BBH	Precomputed ?
640069301	Acinetobacter baumannii ATCC 17978	641522603	Acinetobacter baumannii SDF	97.68	97.69	0.610	0.640	2101	Yes
640069301	Acinetobacter baumannii ATCC 17978	641522602	Acinetobacter baumannii AYE	97.54	97.55	0.840	0.690	2847	Yes
641522602	Acinetobacter baumannii AYE	641522603	Acinetobacter baumannii SDF	97.79	97.8	0.590	0.760	2213	Yes

Export Page 1 of 1 << first < prev 1 next > last >> All

ANI cliques - Species



Home Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG Data Marts Help

[Home](#) > [Compare Genomes](#)

16937 cliques and 11596 species loaded

ANI Cliques

All Cliques **by Species** by Taxonomy Clique Groups

Filter column: Domain Filter text Apply ?

Export Page 1 of 116 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

✓ Domain
Species
Number of Genomes
Number of Cliques
Represented Clique Type(s)
All Columns

Domain	Species	Number of Genomes	Number of Cliques	Represented Clique Type(s)
A	Acidianus hospitalis	1	1	singleton
A	Acidilobus saccharovorans	1	1	singleton
A	Acidilobus sulfurireducens sp. nov.	1	1	singleton
A	Acidiplasma aeolicum	3	2	clique, singleton
A	Acidiplasma cupricumulans	1	1	clique
A	Acidiplasma sp. MBA-1	1	1	clique
A	Aciduliprofundum boonei	1	1	singleton
A	Aciduliprofundum sp. MAR08-339	1	1	singleton
A	Aeropyrum camini	1	1	singleton
A	Archaeoglobus archaean SCGC AC-334-A09	1	1	singleton
A	Archaeoglobus fulgidus	3	1	clique
A	Archaeoglobus sulfaticallidus	1	1	singleton
A	Archaeoglobus veneficus	1	1	singleton
A	Caldisphaera lagunensis	1	1	singleton
A	Caldivirga maquilingensis	1	1	singleton
A	Candidatus Acidianus copahuensis	1	1	singleton

ANI cliques - Sp

Home Find Genomes Find Genes Find Functions

[Home](#) > Compare Genomes

ANI Cliques

All Cliques [by Species](#) [by Taxonomy](#) [Clique Groups](#)

Filter column: Domain Filter text

Export Page 1 of 116 << first < prev 1 2 3 4 5 > next > last >> All

Column Selector

Domain	Species	N
A	Acidianus hospitalis	
A	Acidilobus saccharovorans	
A	Acidilobus sulfurireducens sp. nov.	
A	Acidiplasma aeolicum	3
A	Acidiplasma cupricumulans	1
A	Acidiplasma sp. MBA-1	1
A	Aciduliprofundum boonei	1
A	Aciduliprofundum sp. MAR08-339	1
A	Aeropyrum camini	1
A	Archaeoglobus archaean SCGC AC-334-A09	1
A	Archaeoglobus fulgidus	3
A	Archaeoglobus sulfaticallidus	1
A	Archaeoglobus veneficus	1
A	Caldisphaera lagunensis	1
A	Caldivirga maquilingensis	1
A	Candidatus Acidianus copahuensis	1

Genomes for Species

Genus Species: Acidiplasma aeolicum

Taxonomy: Archaea;Euryarchaeota; Thermoplasmata; Thermoplasmatales; Ferroplasmaceae; Acidiplasma

Filter column: Clique ID Filter text Apply ?

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector Select Page Deselect Page

Select	Genome ID	Genome Name	Clique ID
<input type="checkbox"/>	2681813050	Acidiplasma aeolicum V	1268
<input type="checkbox"/>	2651869795	Acidiplasma aeolicum VT	1268
<input type="checkbox"/>	2596583664	Acidiplasma aeolicum V	9985

Export Page 1 of 1 << first < prev 1 next > last >> All

Add Selected to Genome Cart Select All Clear All

1 1 singleton

3 2 clique, singleton

1 1 clique

1 1 clique

1 1 singleton

ANI cliques - Species



Home

Home > C

ANI C

All Cliques

Filter colu

Export

Column

Domain

A

A

A

Cliques for Species

Genus Species: Acidiplasma aeolicum

Taxonomy: Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales; Ferroplasmaceae; Acidiplasma

Filter column: Clique ID Filter text: Apply ?

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector

	Clique ID	Clique Type	Genome Count
			Acidiplasma aeolicum only
	1268	clique	2
	9985	singleton	1

Export Page 1 of 1 << first < prev 1 next > last >> All

	Acidiplasma saundersii sp. nov.				
A	Acidiplasma aeolicum		3	2	singleton, clique, singleton
A	Acidiplasma cupricumulans		1	1	clique
A	Acidiplasma sp. MBA-1		1	1	clique
A	Aciduliprofundum boonei		1	1	singleton
A	Aciduliprofundum sp. MAR08-339		1	1	singleton
A	Aeropyrum camini		1	1	singleton
A	Archaeoglobus archaean SCGC AC-334-A09		1	1	singleton
A	Archaeoglobus fulgidus		3	1	clique
A	Archaeoglobus sulfaticallidus		1	1	singleton
A	Archaeoglobus veneficus		1	1	singleton
A	Caldisphaera lagunensis		1	1	singleton
A	Caldivirga maquilingensis		1	1	singleton
A	Candidatus Acidianus copahuensis		1	1	singleton

My IMG

Data Marts

Help

pecies loaded

ANI cliques - Taxonomy



Home Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG Data Marts Help

[Home](#) > Compare Genomes

ANI Cliques

All Cliques by Species by Taxonomy Clique Groups

hint: The leaf nodes display Genus Species followed by the count of cliques for that genus-species

Sequencing Status

Finished

Domain

Bacteria

Show

[Expand All](#) [Collapse All](#)

+ Bacteria

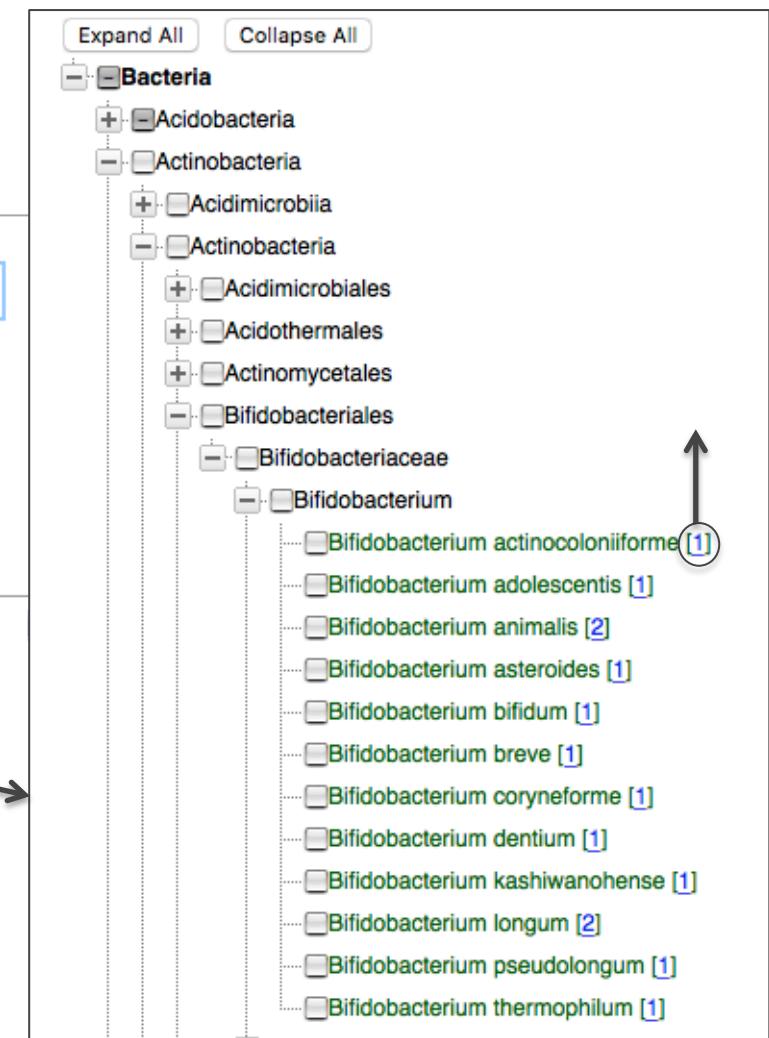
Contact us
[Accessibility/Section 508](#)
[Disclaimer](#)

JGI is a DOE Office of Science User Facility managed by [Lawrence Berkeley National Laboratory](#)
© 1997-2015 The Regents of the University of California

Version 4.540 Oct. 2015
gpweb05 gemini2_shared 5.016000 2016-04-26-12.22.52 128.3.89.208

✓ Finished
Permanent Draft
Draft
All Finished, Permanent Draft and Draft

✓ Archaea
Bacteria
Eukaryota



ANI cliques – Clique groups



Home Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG Data Marts Help

[Home](#) > Compare Genomes

16937 cliques and 11596 species loaded

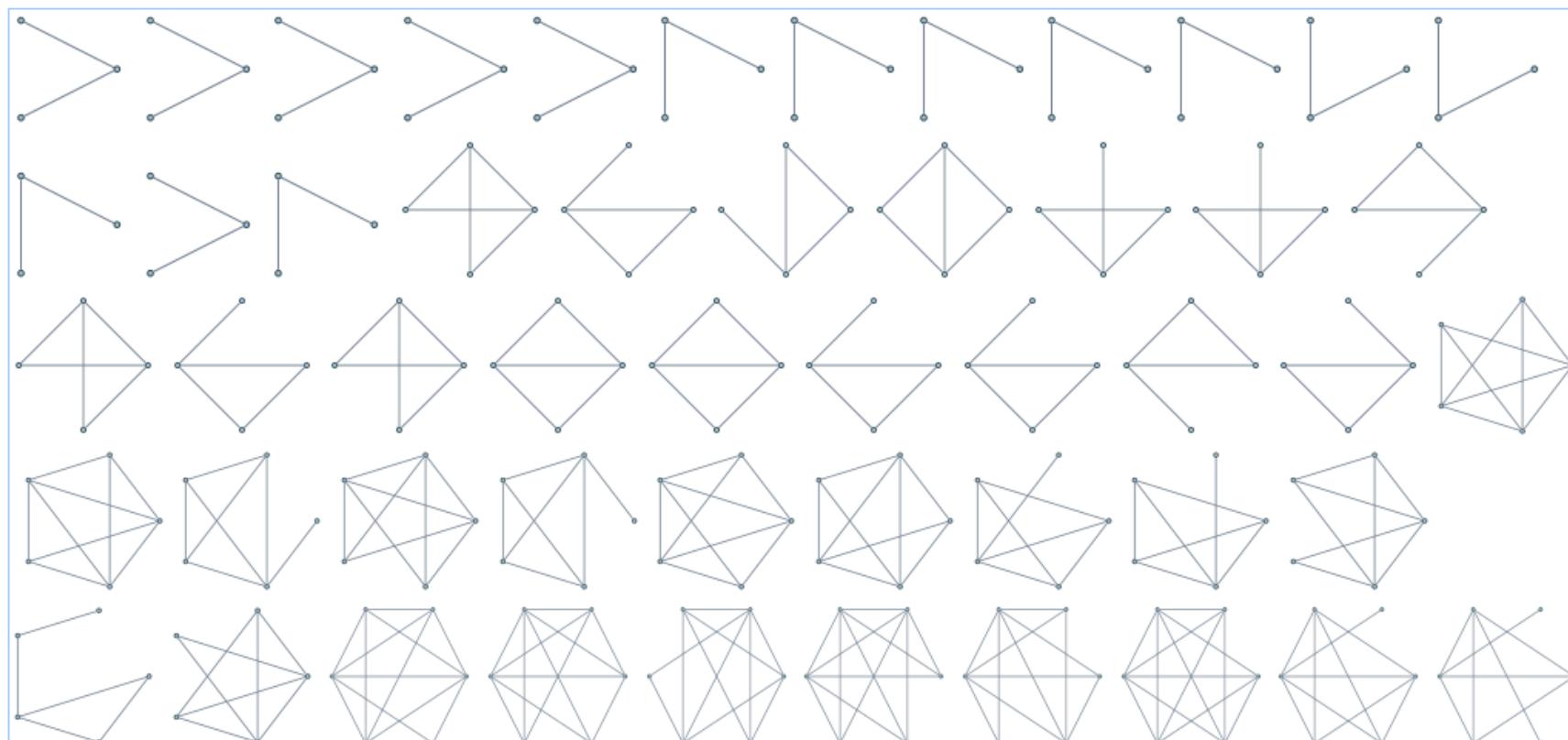
ANI Cliques

All Cliques by Species by Taxonomy **Clique Groups**



Click on a clique group to get detailed information about that group.

Any two genomes in a clique group will be connected by an edge in the graph if they have an ANI ≥ 96.5 and an AF ≥ 0.6 .



ANI cliques – Clique groups

Home > Compare G

ANI Cliques

All Cliques by

hint: Click
Any
an ANI ≥ 96.5 .

Clique Details

Clique ID: 1425

Overview

Genomes in Clique

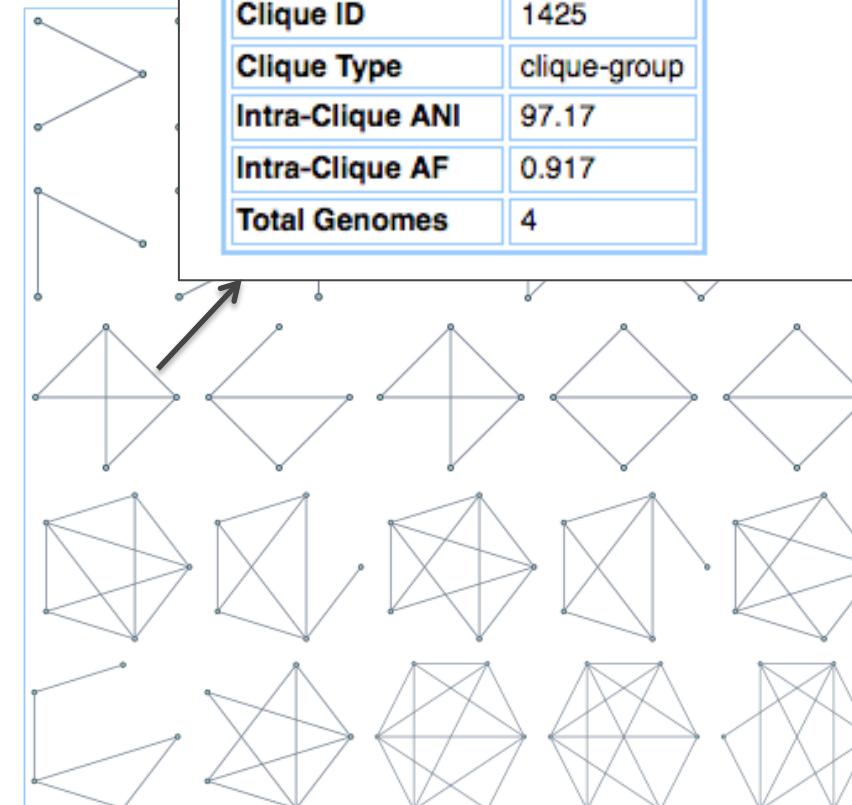
Similar Cliques

Group

Clique Information

Clique ID	1425
Clique Type	clique-group
Intra-Clique ANI	97.17
Intra-Clique AF	0.917
Total Genomes	4

cliques and 11596 species loaded



Clique Details

Clique ID: 1425

Overview

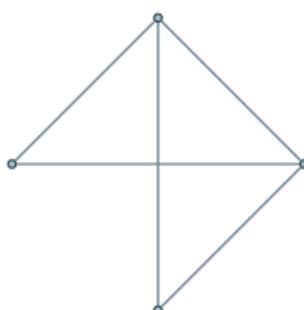
Genomes in Clique

Similar Cliques

Group

hint:

Any two genomes in a clique group will be connected by an edge in the graph if they have an ANI ≥ 96.5 and an AF ≥ 0.6 .





Exercises



Exercises:



- ① Use “Pairwise ANI” tool to compute the ANI,AF of *Nonomuraea sp. NBRC 110462* genome to all other genomes of the same genus and determine to which species this genome could be assigned to.

- ② Use “ANI cliques” tool to identify the cluster that the species *Clostridium carboxidivorans* is assigned to. Explore the cluster to determine type of cluster and identify the missing links if the type is a clique group.



Exercise solutions



Solution 1



Pairwise ANI



BBHs between a genome pair are computed as pairwise bidirectional best nSimScan hits of genes having 70% or more identity and at least 70% coverage of the shorter gene. You may either select genome(s) from IMG or you may upload a nucleotide sequence in FASTA format (using the [Upload File](#) button) to compute ANI to selected genome(s) in IMG

Please select up to 100 genomes:

Quick Search: <enter a genome name to search>

Sequencing Status

List Tree

Search for: <enter a genome name to search>

- 464_Arbrobacter_TAA_AAG (unscreened) (B) [P]
- 464_Chloroherpeton_CGT_GTA (unscreened) (B) [P]
- 464_Chloroherpeton_TAA_GTA (unscreened) (B) [P]
- 464_Dongia_CGT_CTC (unscreened) (B) [P]
- 464_Euzebya_TAA_CTC (unscreened) (B) [P]
- 464_Euzebyaceae_AGG_AGA (unscreened) (B) [P]
- 464_Euzebyaceae_AGG_GTA (unscreened) (B) [P]
- 464_Euzebyaceae_CGT_CTA (unscreened) (B) [P]
- 464_Fabibacter_CGT_TAG (unscreened) (B) [P]
- 464_Martelella_CGT_TAT (unscreened) (B) [P]
- 464_Nitratirfractor_CGT_AGA (unscreened) (B) [P]
- 464_OD1_AGG_ACT (unscreened) (B) [P]
- 464_OD1_AGG_TAG (unscreened) (B) [P]

Selected Genomes

Pairwise 1:

Add Upload Sets Remove Upload File

Pairwise 2:

Add Upload Sets Remove

ANI

Solution 1

Pairwise ANI



metagenomes



BBHs between a genome pair are computed as pairwise bidirectional best nSimScan hits of genes having 70% or more identity and at least 70% coverage of the shorter gene. You may either select genome(s) from IMG or you may upload a nucleotide sequence in FASTA format (using the [Upload File](#) button) to compute ANI to selected genome(s) in IMG

Please select up to 100 genomes:

Quick Search: <enter a genome name to search>

Sequencing Status Domain

All Finished, Permanent Draft and Draft Bacteria

List Tree Show ? Selected: 7

Search for: Nomonuraeal

- Nonlabens ulvanivorans JCM 19298 (B) [P]
- Nonlabens ulvanivorans PLR (B) [P]
- Nonlabens xylophagum DSM 16809 (B) [D]
- Nonomuraea coxensis DSM 45129 (B) [P]
- Nonomuraea jiangxiensis CGMCC 4.6533 (B) [D]
- Nonomuraea maritima CGMCC 4.5681 (B) [D]
- Nonomuraea pusilla DSM 43357 (B) [D]
- Nonomuraea solani CGMCC 4.7037 (B) [D]
- Nonomuraea sp. NBRC 110462 (B) [P]
- Nonomuraea wenchangensis CGMCC 4.5598 (B) [D]
- Nosocomiicoccus massiliensis NP2 (B) [P]
- Nostocales genome binned from the metagenome of a non-axenic Nostoc sp.
- Nostoc azollae 0708 (B) [F]
- Nostoc linckiae EC1-08 (B) [D]

Selected Genomes

Pairwise 1:

Add Upload Sets Remove Upload File

Pairwise 2:

Add Upload Sets Remove

ANI

Solution 1



Pairwise ANI



Filter column: **Genome1 ID** Filter text: Apply ?

Export

Page 1 of 1 << first < prev 1 next > last >>

All

Column Selector	Genome1 ID	Genome1 Name	Genome2 ID	Genome2 Name	ANI1->2	ANI2->1	AF1->2	AF2->1
	2671180301	<u>Nonomuraea sp.</u> <u>NBRC 110462</u>	2675903060	<u>Nonomuraea wenchangensis</u> <u>CGMCC 4.5598</u>	83.67	83.69	0.580	0.500
	2671180301	<u>Nonomuraea sp.</u> <u>NBRC 110462</u>	2675903066	<u>Nonomuraea maritima CGMCC 4.5681</u>	83.38	83.36	0.510	0.550
	2671180301	<u>Nonomuraea sp.</u> <u>NBRC 110462</u>	2675903140	<u>Nonomuraea solani CGMCC 4.7037</u>	83.23	83.25	0.600	0.400
	2671180301	<u>Nonomuraea sp.</u> <u>NBRC 110462</u>	2515154141	<u>Nonomuraea coxensis DSM 45129</u>	83.87	83.91	0.560	0.540
	2671180301	<u>Nonomuraea sp.</u> <u>NBRC 110462</u>	2599185218	<u>Nonomuraea pusilla DSM 43357</u>	99.56	99.56	0.960	0.940
	2671180301	<u>Nonomuraea sp.</u> <u>NBRC 110462</u>	2675903065	<u>Nonomuraea jiangxiensis CGMCC 4.6533</u>	83.31	83.32	0.570	0.440

Solution 2

ANI Cliques

Red arrow pointing to the "by Species" tab.

Red arrow pointing to the "Filter column: Species" dropdown and the "ostridium carboxidivorans" input field.

Red arrow pointing to the "Number of Cliques" value (1) and the "Represented Clique Type(s)" value ("clique-group").

Domain	Species	Number of Genomes	Number of Cliques	Represented Clique Type(s)
B	Clostridium carboxidivorans	4	1	clique-group

Clique Details

Red arrow pointing to the "Clique ID: 1896".

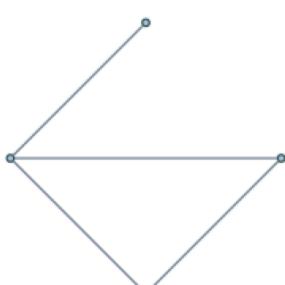
Red arrow pointing to the "Group" tab.

Clique Information	
Clique ID	1896
Clique Type	clique-group
Intra-Clique ANI	99.99
Intra-Clique AF	0.932
Total Genomes	4

Clique Details

Red arrow pointing to the "Group" tab.

Hint: Any two genomes in a clique group will be connected by an edge in the graph if they have an ANI ≥ 96.5 and an AF ≥ 0.6 .



Solution 2



Clique Details

Clique ID: 1896



Overview Genomes in Clique Similar Cliques Group

Filter column: Genome ID Filter text: Apply ?

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector		Select Page	Deselect Page	
Select	Genome ID	Genome Name	Species	Domain
<input type="checkbox"/>	645058736	Clostridium carboxidivorans P7	Clostridium carboxidivorans	B
<input type="checkbox"/>	647533125	Clostridium carboxidivorans P7	Clostridium carboxidivorans	B
<input type="checkbox"/>	2636415692	Clostridium carboxidivorans P7	Clostridium carboxidivorans	B
<input type="checkbox"/>	2654588030	Clostridium carboxidivorans P7 null replaces 81740	Clostridium carboxidivorans	B

Export Page 1 of 1 << first < prev 1 next > last >> All

Add Selected to Genome Cart Select All Clear All Pairwise ANI

Solution 2



Clique Details

Clique ID: 1896

Overview Genomes in Clique Similar Cliques Group

4 of 4 rows selected

Filter column: Genome ID Filter text : Apply ?

Export Page 1 of 1 < first < prev 1 next > last >> All

Column Selector Select Page Deselect Page

Select	Genome ID	Genome Name	Species	Domain
<input checked="" type="checkbox"/>	645058736	Clostridium carboxidivorans P7	Clostridium carboxidivorans	B
<input checked="" type="checkbox"/>	647533125	Clostridium carboxidivorans P7	Clostridium carboxidivorans	B
<input checked="" type="checkbox"/>	2636415692	Clostridium carboxidivorans P7	Clostridium carboxidivorans	B
<input checked="" type="checkbox"/>	2654588030	Clostridium carboxidivorans P7 null replaces 81740	Clostridium carboxidivorans	B

Export Page 1 of 1 << first < prev 1 next > last >> All

4 of 4 rows selected

Add Selected to Genome Cart Select All Clear All Pairwise ANI

Two red arrows highlight specific interaction points on the interface. One arrow points to the 'Page 1 of 1' link under the first set of pagination controls. A second, larger red arrow points diagonally from the 'Page 1 of 1' link in the first section towards the 'Pairwise ANI' button in the second section.

Solution 2



Pairwise ANI



Clique ID: 1896

Filter column: Filter ?

Page 1 of 1 << first < prev 1 next > last >> All ?

Column Selector	Genome1 ID	Genome1 Name	Genome2 ID	Genome2 Name	ANI1->2	ANI2->1	AF1->2	AF2->1	Total BBH	Precomputed ?
	645058736	Clostridium carboxidivorans P7	2636415692	Clostridium carboxidivorans P7	99.99	99.99	0.960	0.470	5025	Yes
	645058736	Clostridium carboxidivorans P7	2654588030	Clostridium carboxidivorans P7 null replaces 81740	99.99	99.99	0.960	0.930	5048	Yes
	645058736	Clostridium carboxidivorans P7	647533125	Clostridium carboxidivorans P7	99.99	99.99	0.740	0.950	3965	Yes
	647533125	Clostridium carboxidivorans P7	2654588030	Clostridium carboxidivorans P7 null replaces 81740	99.99	99.99	0.970	0.730	3892	Yes
	647533125	Clostridium carboxidivorans P7	2636415692	Clostridium carboxidivorans P7	99.99	99.99	0.970	0.370	3868	Yes
	2636415692	Clostridium carboxidivorans P7	2654588030	Clostridium carboxidivorans P7 null replaces 81740	100	100	0.990	0.980	10346	Yes

Page 1 of 1 << first < prev 1 next > last >> All ?