

Day 3

Bin refinement

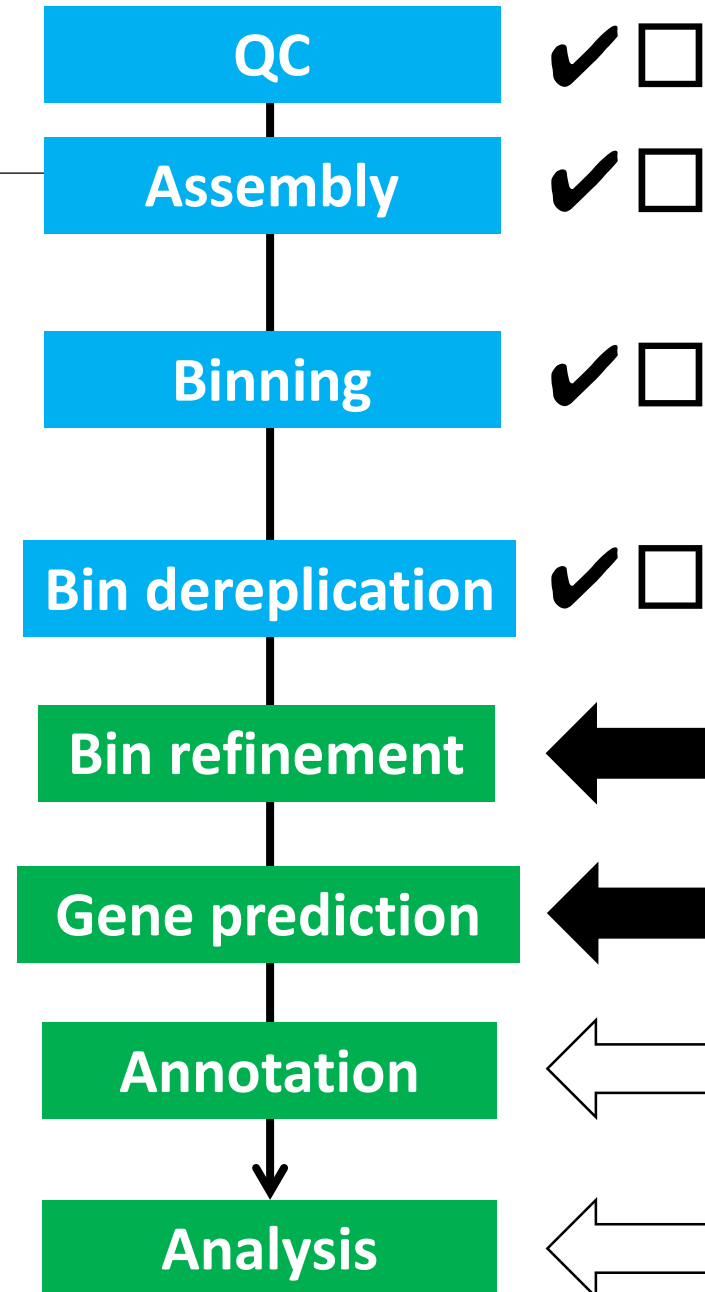
Gene prediction

Gene annotation



Day overview

- **Goals:**
 - **Bin refinement**
 - **Gene prediction**
 - **Annotation**
 - **Start analysis**

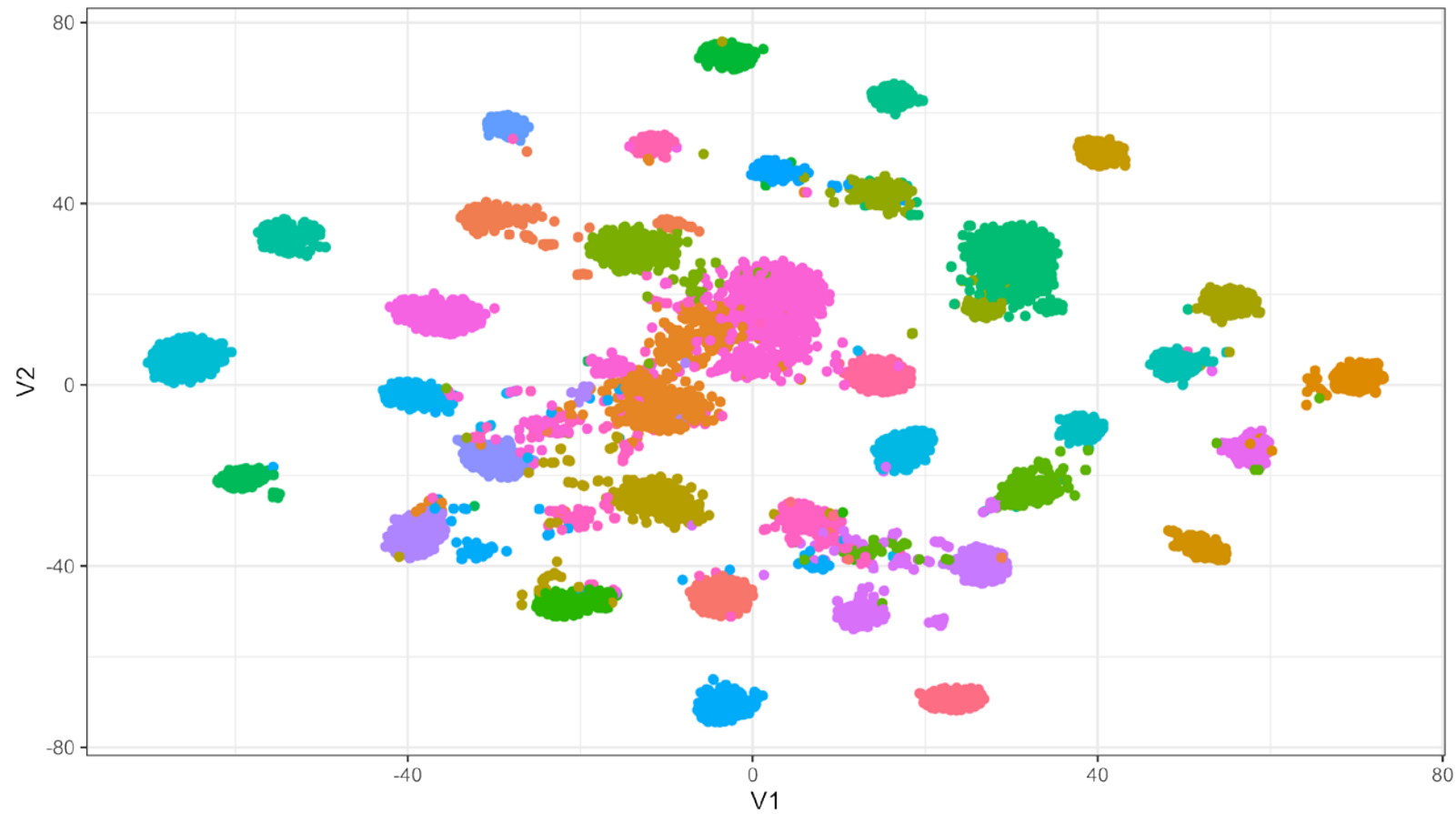


Bin refinement



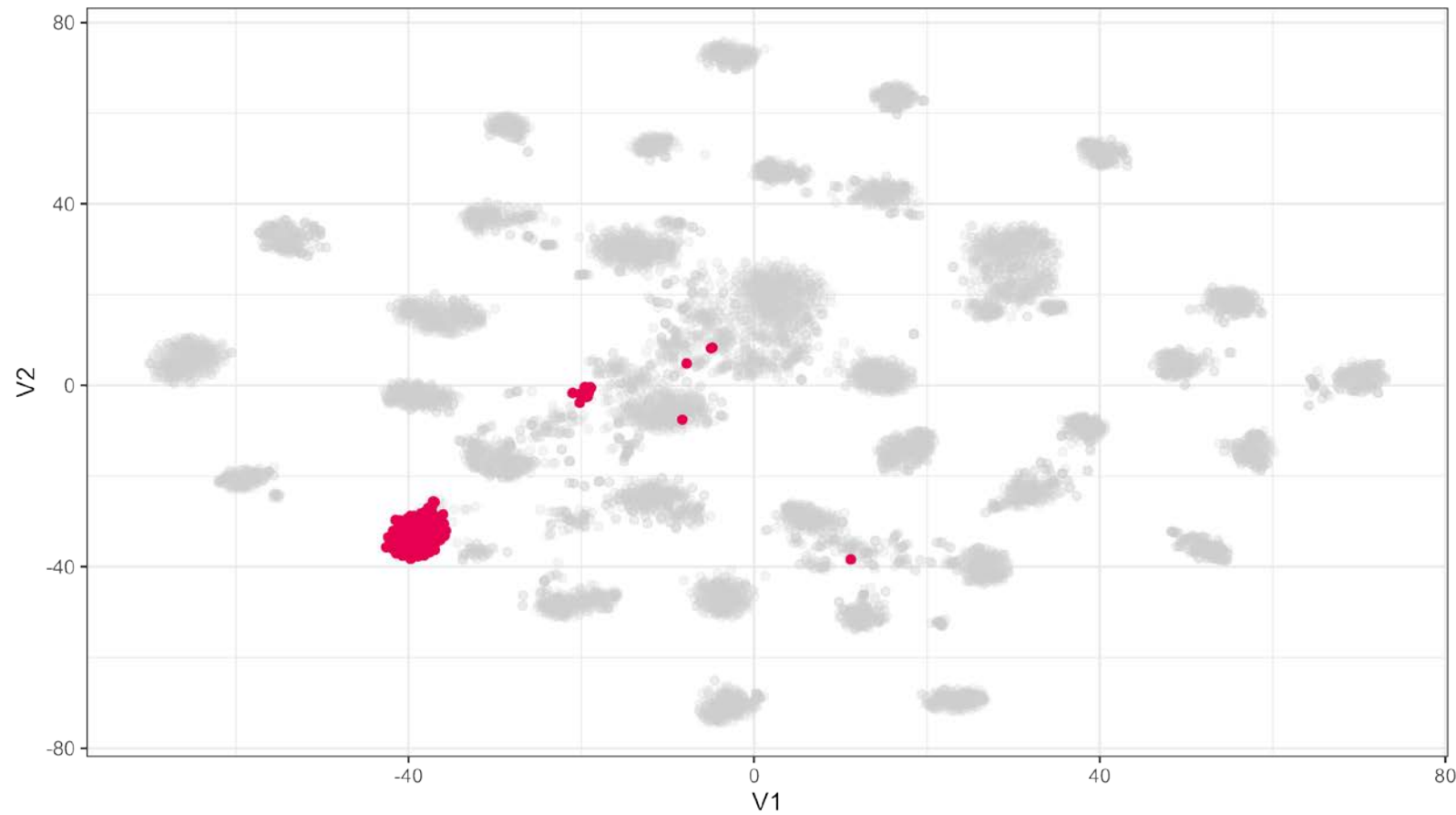
VizBin

- Inspect bin assignments



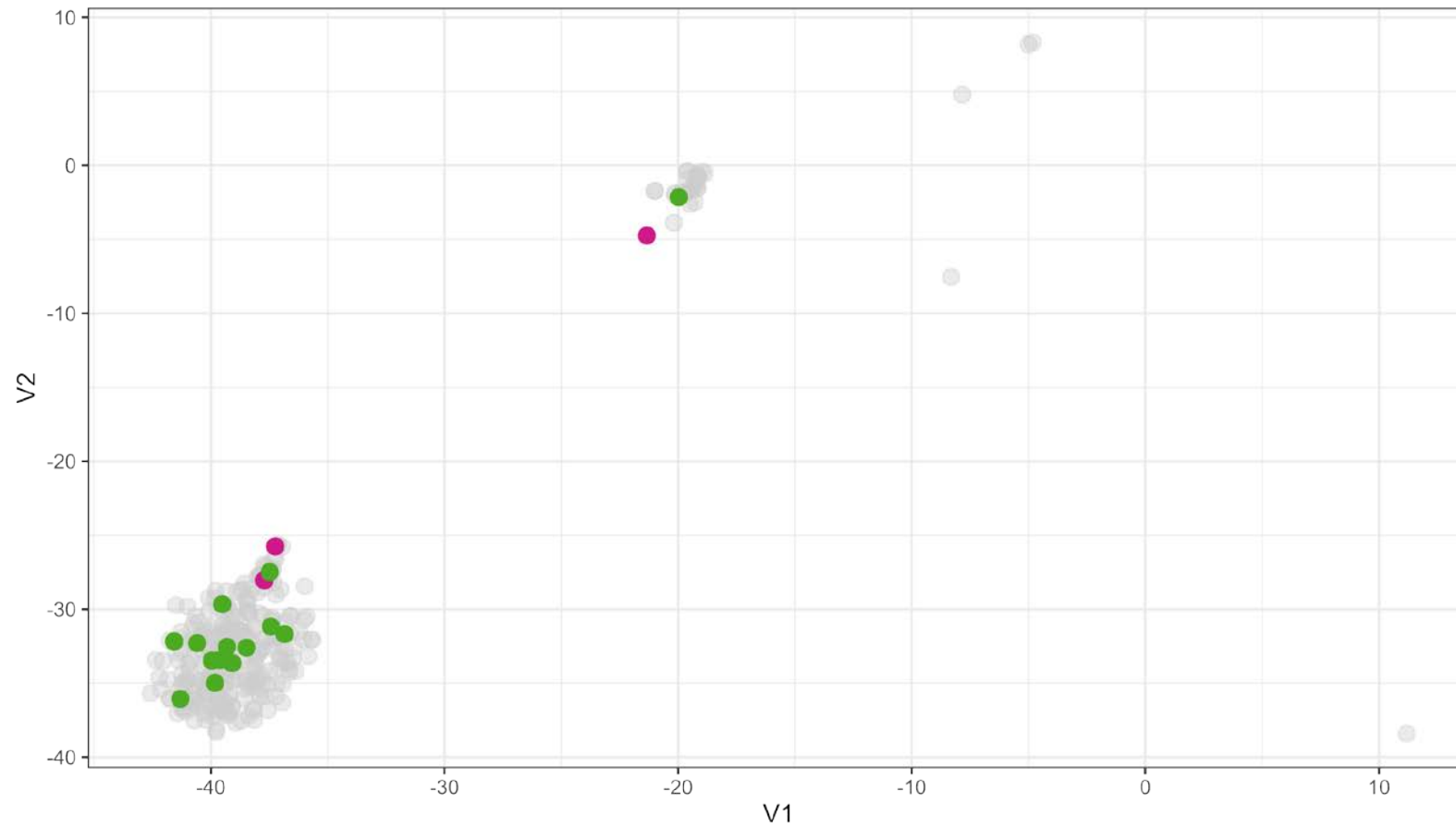
VizBin

- Use graphical user interface to assign bins or reassign contigs



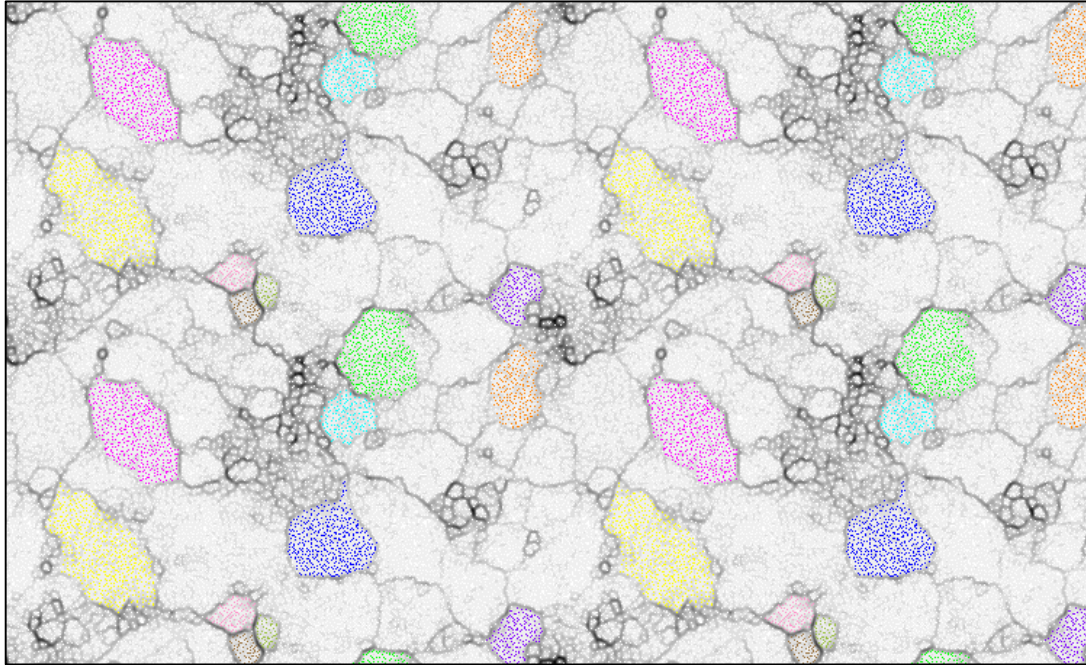
VizBin

- Identify contigs with unstable placement



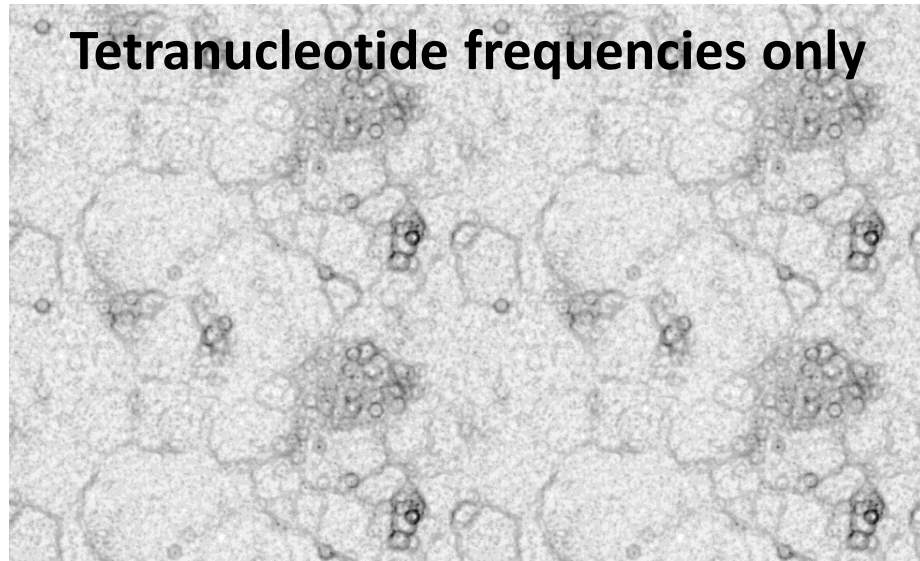
Alternative: ESOM

- Emergent Self Organizing Map (ESOM)
- esomana: <http://databionic-esom.sourceforge.net/user.html>

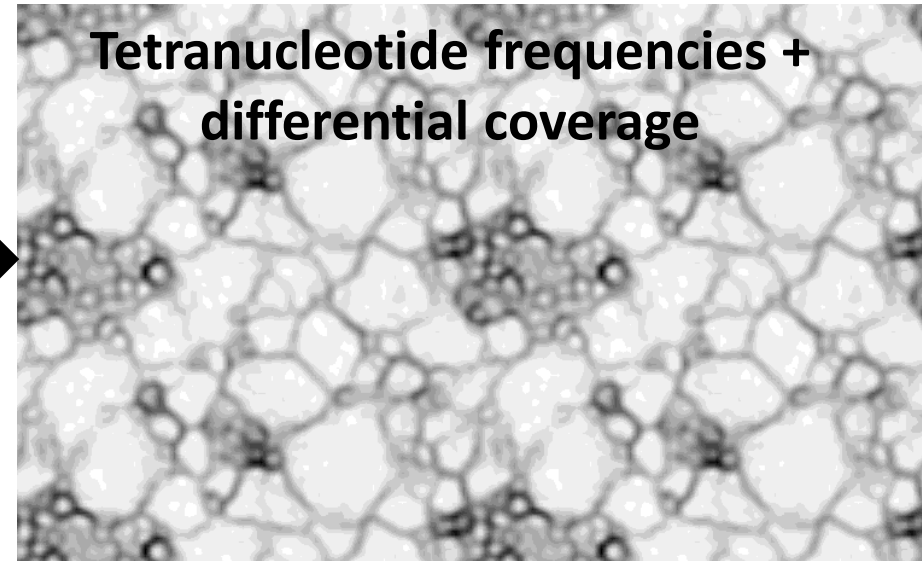


Alternative: ESOM

- **Flexible:** use whatever data you want, e.g.: tetranucleotide frequencies, coverage, both



Indistinct bin boundaries between highly similar genomes

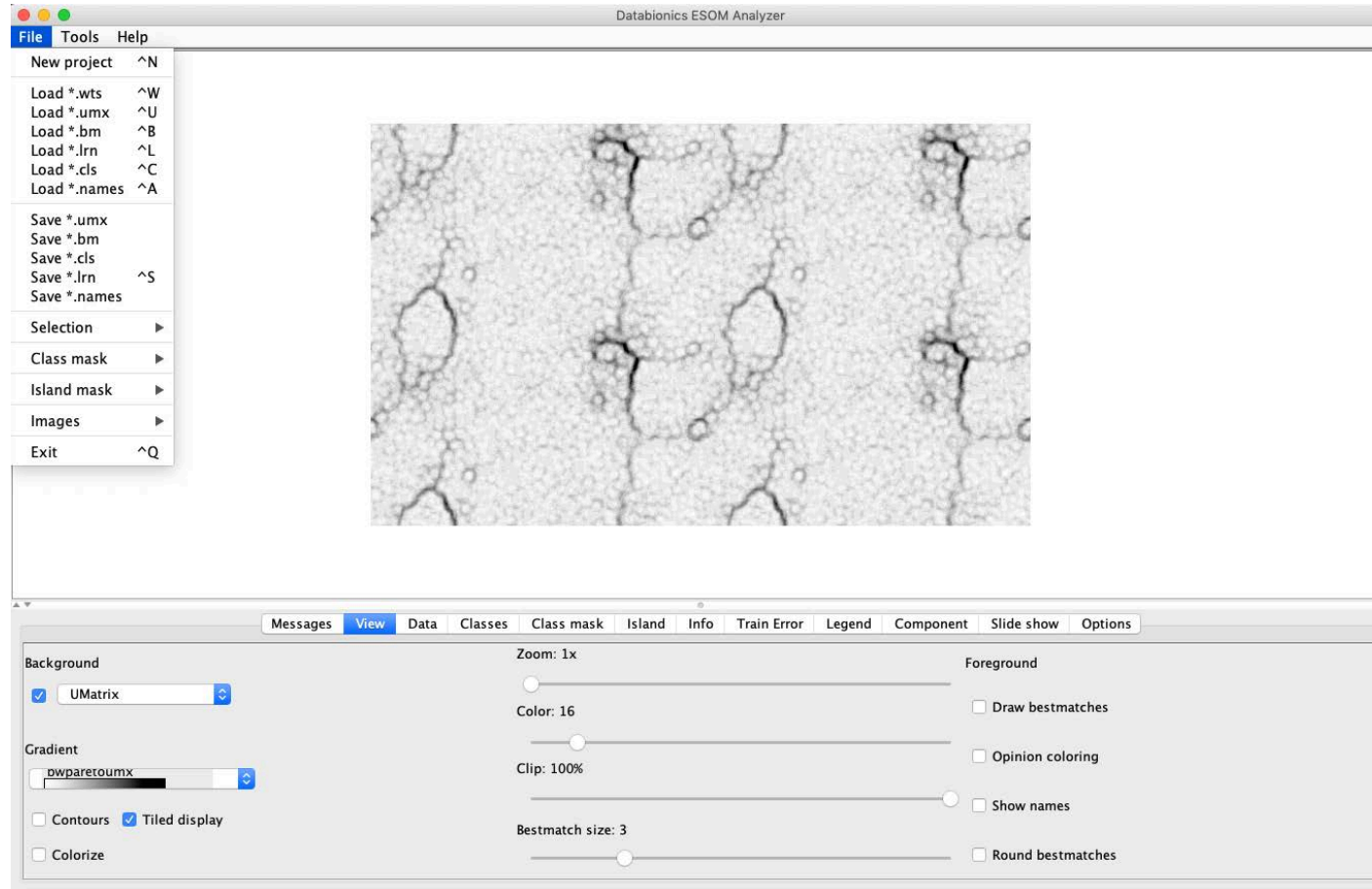


Clear bin boundaries between highly similar genomes using spatial gradient data



Alternative: ESOM

Topographic map of clusters

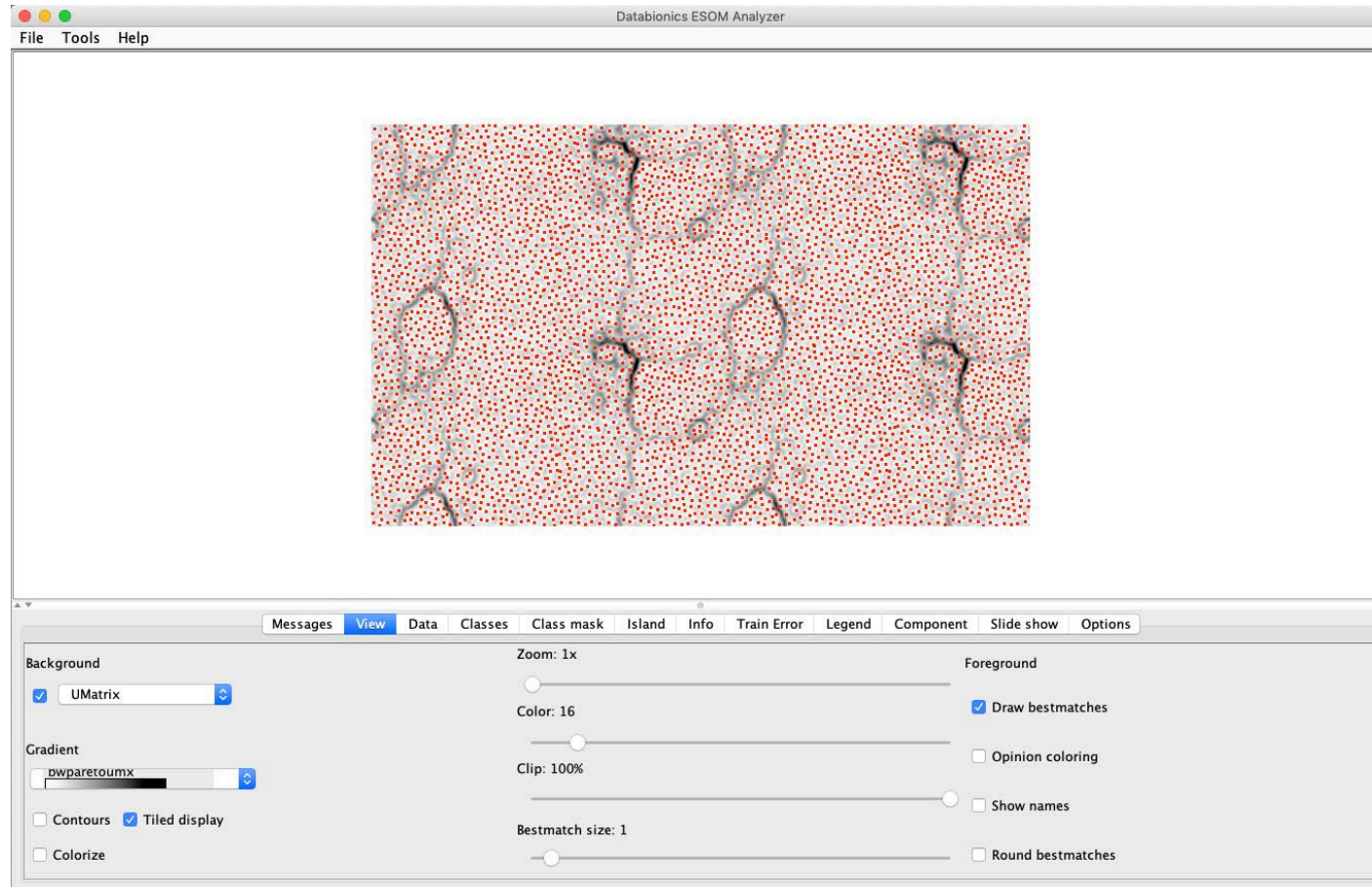


- Uses graphical user interface
- Must supply own prepared data (e.g. pre-calculate tetranucleotide frequencies)
- Dark lines = bin boundaries
- Strong lines = strong bin divisions



Alternative: ESOM

Map with contig fragments shown in red

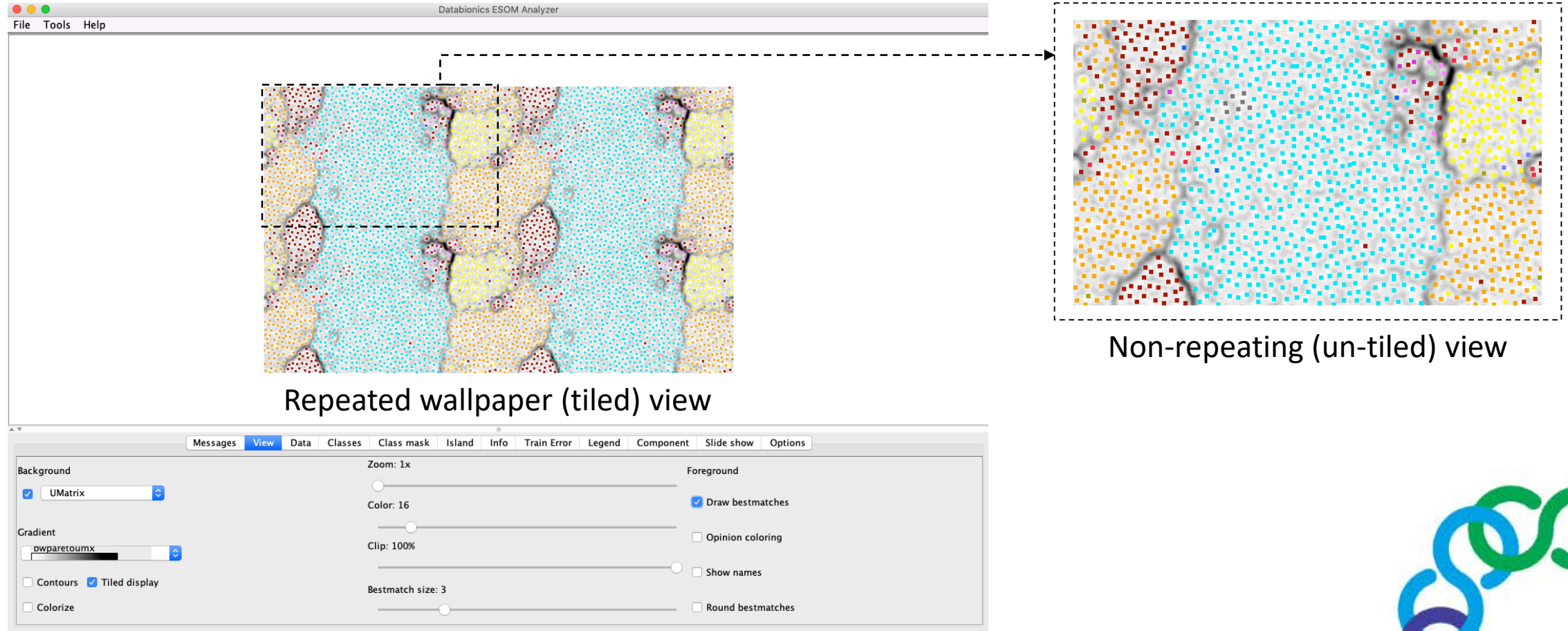


- Uses graphical user interface
- Must supply own prepared data (e.g. pre-calculate tetranucleotide frequencies)
- Dark lines = bin boundaries
- Strong lines = strong bin divisions



Alternative: ESOM

Map with contig fragments shown and coloured by bin assignment

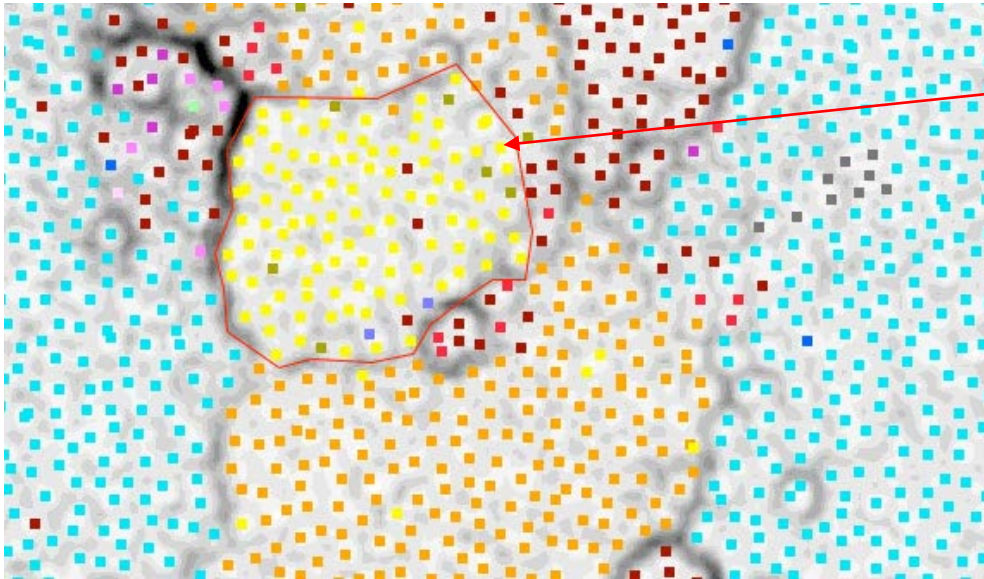


Colorado groundwater communities (Handley et al., 2013, ISME J)



Alternative: ESOM

Map with contig fragments shown and coloured by bin assignment

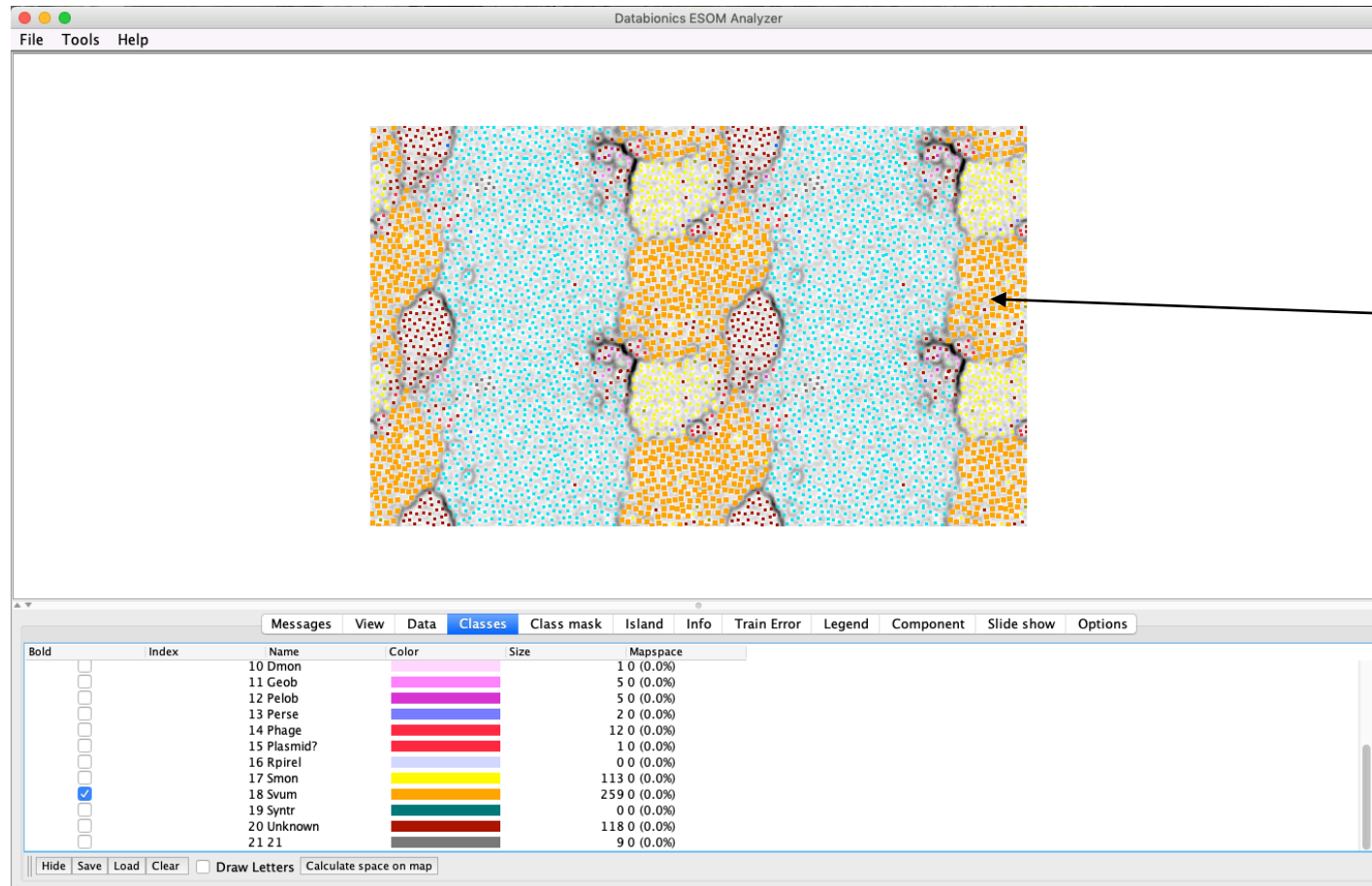


Like VizBin, bins are selected by manually drawing around boundary



Alternative: ESOM

Map with contig fragments shown and coloured by bin assignment



- Select pre-assigned bin to highlight contig fragments
- Choose/change bin colours
- Example: *Sulfurovum* bin highlighted



Task: Work with VizBin

Use VizBin to:

- Prepare input files for VizBin
- Project high-dimensional data down into a 2D plot
- Pick refined bins



Gene prediction



Gene prediction and annotation

- Genome prediction annotation is the process of attaching biological information to sequences
- It consists of three main steps:
 - Gene prediction
 - Prediction of protein sequences
 - Functional annotation: Attaching biological information to these elements



Gene prediction

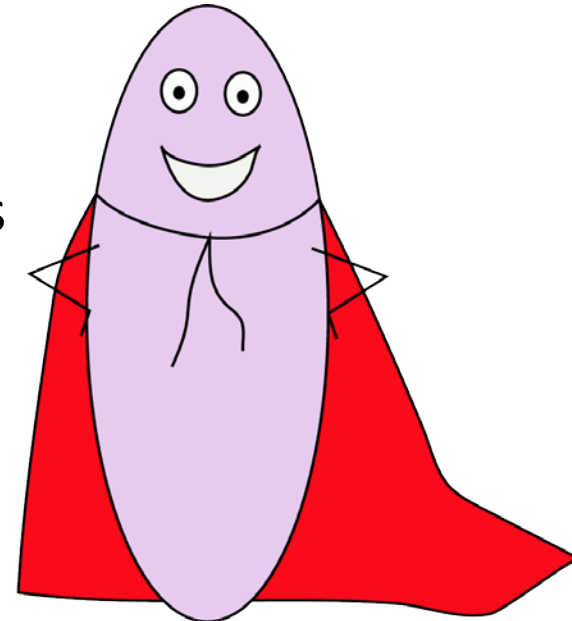
Aim:

- To identify regions of genomic DNA that encode putative genes present in high quality genomes

About 1/1000th of a human genome in size,
but with only 1/10th less coding DNA sequence
→ 100 x more power packed!!!

Prokaryote genomes:

- High gene density
- Genes = continuous stretches of coding DNA
- Absence of introns in the protein coding regions



Gene prediction

Gene finding algorithms for prokaryotes

- Homology:
 - Search by sequence similarity to homologous sequences
 - Based on the assumption that functional regions are more conserved evolutionarily than non-functional regions
- *Ab initio*:
 - Search by content: find genes by statistical properties that distinguish protein-coding DNA from non-coding DNA
 - Search by signals/sites, e.g. promoters, start and stop codons



Gene prediction

Homology: Sequence similarity searches

- Finding similarity in gene sequences between expressed sequence tags (ESTs), proteins, or other genomes to the input genome
- Local alignment:
 - BLAST family tools: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
 - Global alignment
 - GeneWise: <https://www.ebi.ac.uk/Tools/psa/genewise/>



Gene prediction

Ab initio search by content algorithms:

- Markov Models
- Dynamic Programming
- Linear discriminant analysis
- Linguist methods
- Neural Network



Gene prediction

Ab initio search by content: Markov Model Based Algorithms

- Most widespread algorithms for gene finding in prokaryotes are based on Markov Models
- Aim is to capture compositional differences among coding regions, “shadow” coding regions (coding on the opposite DNA strand) and non-coding DNA



Gene prediction

Markov Model Based Algorithms: Glimmer

- <https://ccb.jhu.edu/software/glimmer/>
- Interpolated Markov model (IMM) DNA discriminator
- Log-likelihood that a given interval on a DNA sequence was generated by a model of coding versus non-coding DNA



Gene prediction

Markov Model Based Algorithms: GeneMark/GeneMarkHMM/MetaGeneMark

- <http://exon.gatech.edu/GeneMark/>
- GeneMark is a family of gene prediction tools
- Genomic sequences can be analysed either by the self-training program GeneMarkS (sequences >50 kb) or using Heuristic Models by GeneMarkHMM
- Pre-trained model parameters are available for many species
- Metagenomics sequences can be analysed with MetaGeneMark



Gene prediction

Prodigal (PROkaryotic Dynamic Programming Genefinding ALgorithm)

- <http://compbio.ornl.gov/prodigal/>
- Based on Dynamic Programming, not Markov Models
- Gene-finding algorithm for prokaryote genomes developed to predict translation initiation sites more accurately.
- High accuracy in high GC content genomes
- Tends to predict longer genes rather than more genes (minimising number of false positives)



Gene prediction

Prodigal for metagenomics:

- Use anon (meta) mode with metagenomic data (or short sequence data)
 - Copes with diverse genomes
 - Unlike normal mode, it does not attempt to study the input sequence, and predict based on these assumptions
 - Uses pre-calculated training files, and predicts genes based on the best results
- Alternatively, use normal mode on each individual genome bin



Gene prediction

Prodigal for metagenomics:

- **Caveat:** unusual genetic codes
 - First uses genetic code 11 (stop codons TAA, TGA, TAG)
 - If genes are too short, uses alternative code 4 (TGA not a stop codon)
 - Will not try code 25, but will issue warning if genes are short
 - Must manually select code 25



Gene prediction

Prodigal for metagenomics:

- Important note:
 - Prodigal predicts coding DNA sequence ONLY
 - Provides nucleic acid (.fna) and amino acid (.faa) files
 - **DOES NOT** identify other features (e.g. rRNA, tRNA)
 - Combine with other prediction tools



Gene prediction

Predicting RNA features and non-coding regions:

- MeTaxa2: predicts ribosomal RNA sequences in a genome
- Aragorn: predicts tRNA and tmRNA sequences



Gene prediction

Predicting protein coding sequences in unassembled (short) reads

- FragGeneScan:
 - Tuning parameters for short sequences (and hence incomplete genes)
 - Model sequence error



Task: Gene prediction

Preparing data for gene prediction

1. Identify and prepare input files for each gene prediction tool (Prodigal, FragGeneScan, MeTaxa2 and Aragorn)
2. Configure parameters for gene prediction

Perform gene prediction

1. Run each job directly from the node (no slurm script required)

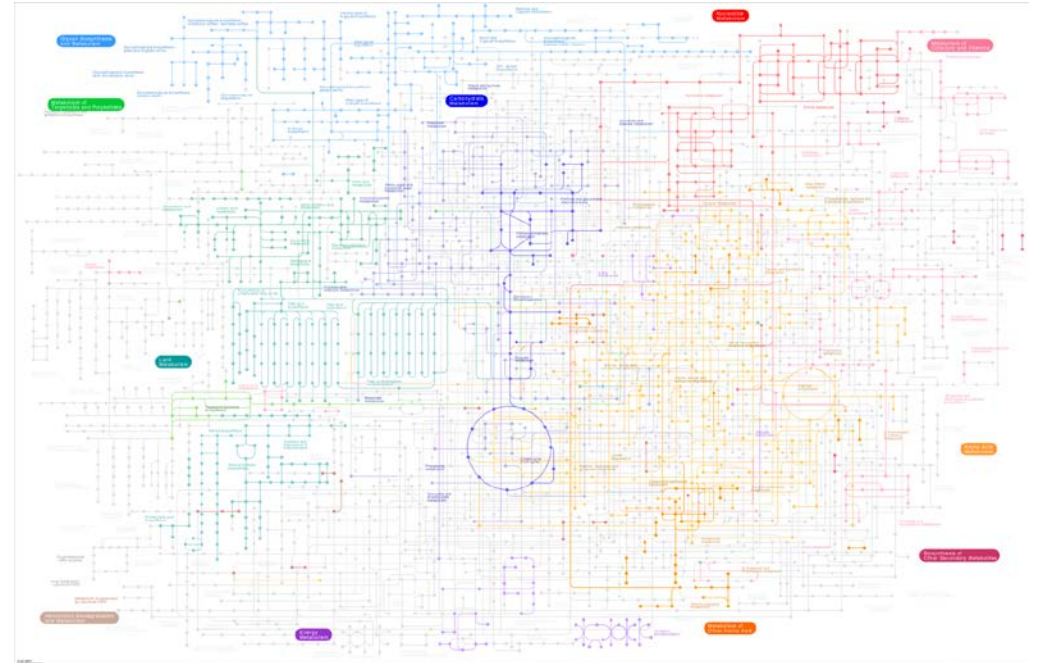


Gene annotation



Gene annotation

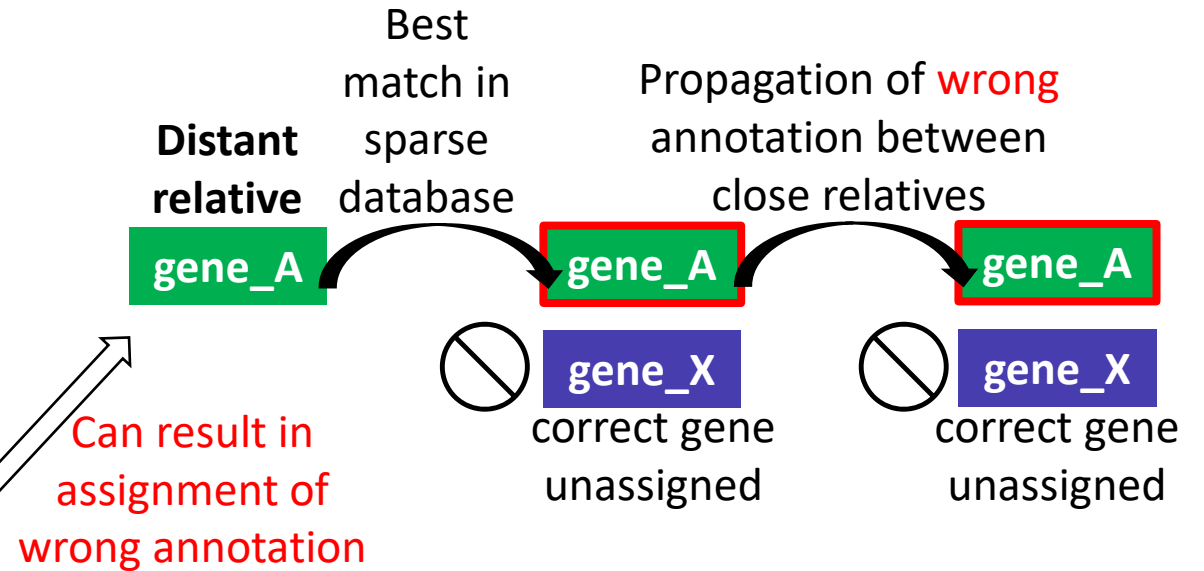
- Genome annotation attempts to predict gene function
- Predicted genes or protein sequences are compared against a curated set of reference sequences for which function is known, or is strongly suspected



Gene annotation

Caveat:

- Annotations are dependent on the reference database
- Environmental genomes can have:
 - Genes with distant homology matches to unrelated taxa
 - Large numbers of “hypothetical” gene annotations (= genes of unknown function)



Gene annotation

Caveat:

- Annotations are “advice”
- Automated annotations often need to be manually curated
- Interrogate if: expected functional gene is missing from annotations
- Gene synteny is a useful for missing gene discovery, e.g.:
 - check genes co-located in operons for putative functions
 - check for operon truncation (due to contig break)



Gene annotation

There are two main ways to perform gene annotation with protein sequences:

- BLAST-like gene annotation
- Domain annotation



Gene annotation

BLAST-like gene annotation

- Pairwise local alignment between the gene of interest (query sequence) and the sequences in the database (target sequence)
- Tools:
 - BLAST: web-based and stand alone
(usually too slow for metagenomics)
 - USEARCH (64-bit): fast (subscription needed)
 - Diamond: fast

Descriptions Graphic Summary **Alignments** Taxonomy

Alignment view Pairwise ? Download

100 sequences selected ?

[Download](#) [GenPept](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

amidase [Gemmatimonadetes bacterium]
Sequence ID: [MBB27982.1](#) Length: 447 Number of Matches: 1
[See 1 more title\(s\)](#)

Range 1: 182 to 446 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
268 bits(686)	3e-84	Compositional matrix adjust.	139/265(52%)	175/265(66%)	0/265(0%)

Query 1 MGLKPTFGRI SLRGILPVSYELDHPCPFTRSVADA AVILQCLAGKDP LPLSADVPVDI
+GLKPT GR+S+ G++PVS+ LDHPCP T SV DAA ILQ +AG DP DPLSA
Sbjct 182 VGLKPTLGRVSVHGVPVPSFNLDHFGPLT SVGDAARILQVIAGYDPKDP LPSASETTTI

Query 61 RIEPLSRPPRVGIVRTYYPD NAETMRAATDDAIERLASEGA EFTDVHMPG SFAELHE
PL RPPR+G + Y+ + ADE M +AT AIE L GAE ++ MP SF LHE
Sbjct 242 TPRPLRPPRIGHLVGYFREQAEDMSSATQRAIECLQLAGAECIELEMPDSFGCLHE

Query 121 ALLLAVGAANVLDERYVAHRDAFP PSLCEIMERGRSAGAVDYARARRHQISFKSEVLA
+++ A DE++ HR+ +PP L +M+ G + AV YA AR+HQI F+ ++ +
Sbjct 302 RIIMVSEGAAYHDEQFGLHRNEYPPGLRSLMDEGLATS AVTYANARKHQIDFRLIQSI

Query 181 EGVDLLLTPATPTPAPSGLTSTGNPAFN SPWSYAGLPTIVLPAACSSDGLPAGIQLVAF
+D+LLTPAT TPAP L STGNPAFN SPWSY GLPTI LP GLPA IQLV
Sbjct 362 RDL DILLTPATLTPAPKTLESTGNPAFN SPWSYCGLPTISLPVEVGESGLPAAIQLVGE

Query 241 FAEIRLLTVSAWCETRLWNRTPSI 265
F+E RLL+++ WCE L WN P +
Sbjct 422 FSESRLLSIAQWCEQVLGWNHKPEL 446

Related Information
[Identical Proteins](#) -
Identical proteins to
MBB27982.1



Gene annotation

HMM-profiling of domains:

- Considers the query sequences as a collection of independently functioning protein folding domains
- Uses database of Hidden Markov models built from a collection of proteins that share a common domain
- Profiles build from statistical map of the
 - amino acid transitions (from position to position),
 - variations (differences at a position),
 - insertions/deletions between positions
- Tools: HMMer software (<http://hmmer.org/>)



Gene annotation

Common functional databases

- KEGG (Kyoto Encyclopedia of Genes and Genomes) (<https://www.kegg.jp>)

Very popular, each entry is well annotated, and often linked into “Modules” or “Pathways”
(Full access now requires a license fee)

- COGs (Clusters of Orthologous Groups of proteins) (<https://www.ncbi.nlm.nih.gov/COG/>)

Classify proteins from completely sequenced genomes on the basis of the orthology concept

- PFAM (<https://pfam.xfam.org>)

Focused more on protein domains based on hidden Markov models

- TIGRfam (<https://www.jcvi.org/tigrfams>)

Database of protein family definitions based on hidden Markov models



Gene annotation

Common functional databases (continued)

- The PANTHER (**P**rotein **A**nalysis **T**hrough **E**volutionary **R**elationships) Classification System (<http://pantherdb.org>)

Proteins are classified according to Family and subfamily, molecular function, biological process and pathway

- UniRef (UniProt Reference Clusters) (<https://www.uniprot.org/>)

Protein clustering at different levels (e.g. UniRef100, UniRef90, UniRef50)

- BioCyc Database Collection (<https://biocyc.org>)

14735 Pathway/Genome Databases (PGDBs), plus software tools

Subscriptions are required to access most of BioCyc

- MetaCyc Metabolic Pathway Database (<https://metacyc.org>)

2722 pathways from 3009 different organisms



Gene annotation

Graphical User Interface – MEGAN

- Toolbox for interactively analyzing microbiome data.
 - Taxonomic analysis using the NCBI taxonomy or a customized taxonomy such as SILVA
 - Functional analysis using InterPro2GO, SEED, eggNOG or KEGG
 - Bar charts, word clouds, Voronoi tree maps and many other charts
 - PCoA, clustering and networks
 - Supports metadata

<https://uni-tuebingen.de/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/informatik/lehrstuehle/algorithms-in-bioinformatics/software/megan6/>



Gene annotation

Some web-based annotation tools:

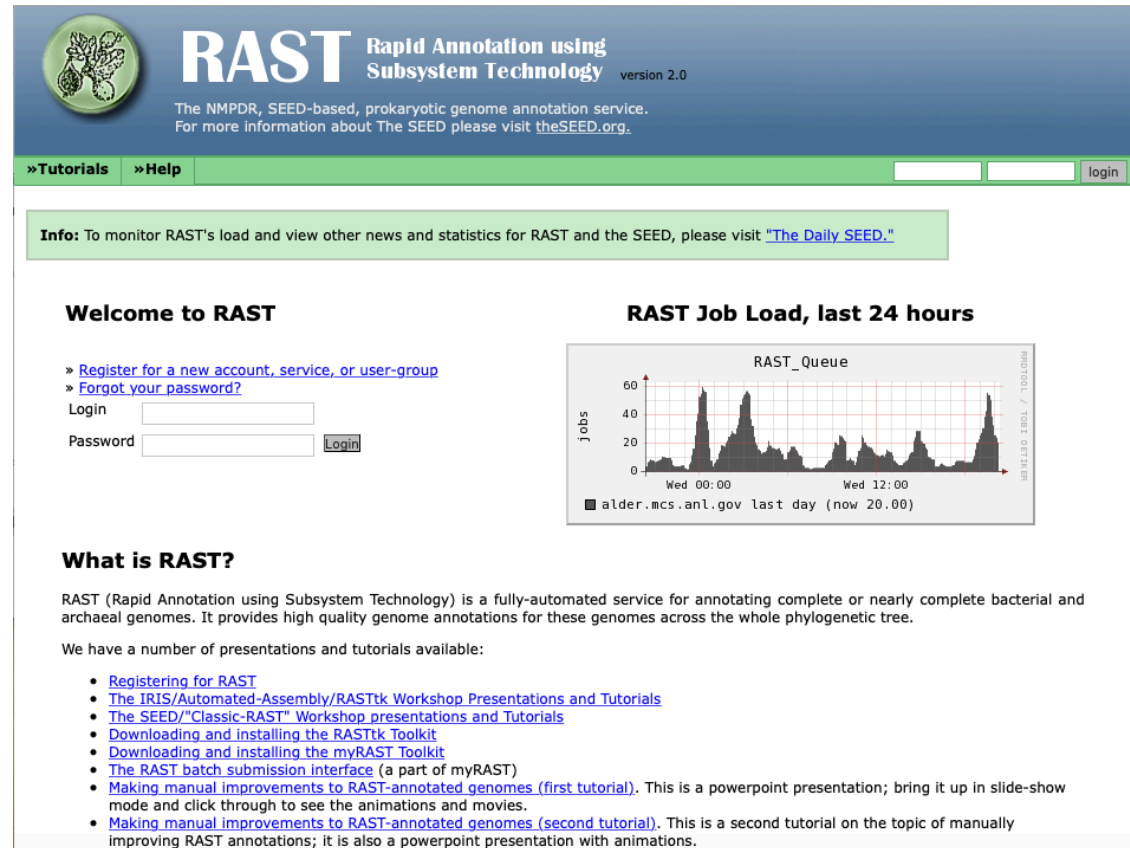
- Web BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)
- RAST/MG-RAST (Rapid Annotation using Subsystem Technology) Annotation Server
- KEGG Automatic annotation and KEGG mapping service
 - BLAST-Koala: BLAST search (<https://www.kegg.jp/blastkoala/>)
 - GHOST-Koala: GHOSTX search (<https://www.kegg.jp/ghostkoala/>)
 - KofamKOALA: HMM profile search (<https://www.genome.jp/tools/kofamkoala/>)
- IMG/M (The Integrated Microbial Genomes and Microbiomes)
(<https://img.jgi.doe.gov>)



Gene annotation

RAST Annotation Server (Rapid **Annotation** using Subsystem Technology):

- Fast annotation (~1 genome/day)
- Can use for individual genome bins
- It works well for genomes similar to large groups of reference genomes
- As usual: requires manual curation after initial annotation




The screenshot shows the RAST website interface. At the top, there is a logo for RAST (Rapid Annotation using Subsystem Technology) version 2.0. Below the logo, it states: "The NMPDR, SEED-based, prokaryotic genome annotation service. For more information about The SEED please visit theSEED.org." There are links for "»Tutorials" and "»Help", and a "login" button. A green box contains an "Info" message: "To monitor RAST's load and view other news and statistics for RAST and the SEED, please visit [The Daily SEED](http://TheDailySEED.org)." Below this, there is a "Welcome to RAST" section with links for "» Register for a new account, service, or user-group" and "» Forgot your password?". There are input fields for "Login" and "Password" with a "Login" button. To the right, there is a graph titled "RAST Job Load, last 24 hours" showing the "RAST_Queue" with a peak around 60 jobs. Below the graph, it says "alder.mcs.anl.gov last day (now 20.00)". The "What is RAST?" section describes the service as a fully-automated service for annotating complete or nearly complete bacterial and archaeal genomes. It provides high quality genome annotations for these genomes across the whole phylogenetic tree. It also lists a number of presentations and tutorials available:

- [Registering for RAST](#)
- [The IRIS/Automated-Assembly/RASTtk Workshop Presentations and Tutorials](#)
- [The SEED/"Classic-RAST" Workshop presentations and Tutorials](#)
- [Downloading and installing the RASTtk Toolkit](#)
- [Downloading and installing the myRAST Toolkit](#)
- [The RAST batch submission interface](#) (a part of myRAST)
- [Making manual improvements to RAST-annotated genomes \(first tutorial\)](#). This is a powerpoint presentation; bring it up in slide-show mode and click through to see the animations and movies.
- [Making manual improvements to RAST-annotated genomes \(second tutorial\)](#). This is a second tutorial on the topic of manually improving RAST annotations; it is also a powerpoint presentation with animations.

Gene annotation

RAST Annotation Server (Rapid Annotation using Subsystem Technology)



The SEED Viewer

SEED Viewer version 2.0

Welcome to the SEED Viewer - a read-only browser of the curated SEED data.
For more information about The SEED please visit theSEED.org.

[»Navigate](#) [»Organism](#) [»Comparative Tools](#) [»Help](#)

Organism Overview for *Candidatus Latescibacter anaerobius* SCGC AAA252-E07 (910047.4)

Genome	Candidatus Latescibacter anaerobius SCGC AAA252-E07 (Taxonomy ID: 910047)
Domain	Bacteria
Taxonomy	Bacteria; FCB group; Candidatus Latescibacteria; Candidatus Latescibacter; Candidatus Latescibacter anaerobius; Candidatus Latescibacter anaerobius SCGC AAA252-E07
Neighbors	View closest neighbors
Size	2,290,285
GC Content	42.1
N50	24013
L50	34
Number of Contigs (with PEGs)	164
Number of Subsystems	157
Number of Coding Sequences	2064
Number of RNAs	30

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

[Browse](#) [Compare](#) [Download](#) [Annotate](#)

Browse through the features of [Candidatus Latescibacter anaerobius SCGC AAA252-E07](#) both graphically and through a table. Both allow quick navigation and filtering for features of your interest. Each feature is linked to its own detail page.

Click [here](#) to get to the Genome Browser

Subsystem Information

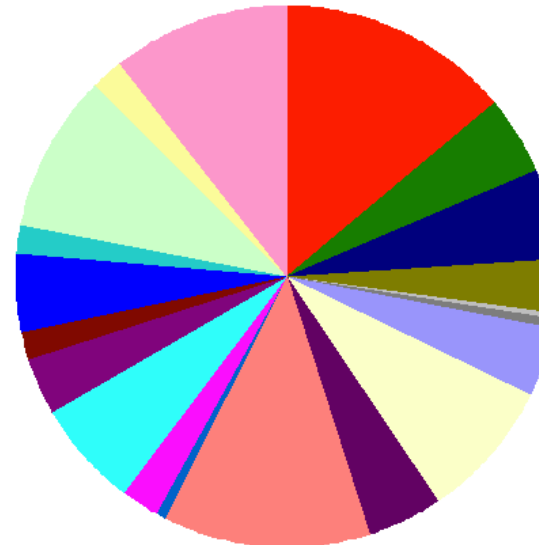
Subsystem Statistics

Features in Subsystems

Subsystem Coverage



Subsystem Category Distribution



Subsystem Feature Counts

- Cofactors, Vitamins, Prosthetic Groups, Pigments (67)
- Cell Wall and Capsule (23)
- Virulence, Disease and Defense (25)
- Potassium metabolism (15)
- Photosynthesis (0)
- Miscellaneous (2)
- Phages, Prophages, Transposable elements, Plasmids (2)
- Membrane Transport (21)
- Iron acquisition and metabolism (0)
- RNA Metabolism (39)
- Nucleosides and Nucleotides (21)
- Protein Metabolism (60)
- Cell Division and Cell Cycle (2)
- Motility and Chemotaxis (0)
- Regulation and Cell signaling (11)
- Secondary Metabolism (0)
- DNA Metabolism (31)
- Fatty Acids, Lipids, and Isoprenoids (15)
- Nitrogen Metabolism (9)
- Dormancy and Sporulation (1)
- Respiration (22)
- Stress Response (8)
- Metabolism of Aromatic Compounds (0)
- Amino Acids and Derivatives (45)
- Sulfur Metabolism (9)
- Phosphorus Metabolism (1)
- Carbohydrates (48)



Task: Gene annotation

Preparing data for gene annotation

1. Identify and prepare input files for gene annotation with Diamond
2. Configure parameters for gene annotation

Perform gene annotation

1. Prepare an annotation job to run under slurm
2. Use MEGAN to explore gene networks










Online resources and data analysis



Gene annotation

Identification of Biosynthetic Gene Clusters with antiSMASH

antiSMASH bacterial version  **Submit Bacterial Sequence**  **Submit Fungal Sequence**  **Submit Plant Sequence**  **Download**  **About**  **Help**  **Contact**

Server status: working

Running jobs: 0

Queued jobs: 0

Jobs processed: 595148

Nucleotide input

Results for existing job

Search a genome sequence for secondary metabolite biosynthetic gene clusters

Notification settings
 Email address (optional)

Data input
 NCBI accession number of desired sequence

Detection strictness: relaxed

strict relaxed loose

- Detects well-defined clusters containing all required parts.
- Detects partial clusters missing one or more functional parts.

Extra features All off All on




☒ **KnownClusterBlast** ☐ **ClusterBlast** ☒ **SubClusterBlast**

☒ **ActiveSiteFinder** ☐ **Cluster Pfam analysis** ☐ **Pfam-based GO term annotation**






















Please be considerate in your use of antiSMASH. Help us keep antiSMASH available for everybody by limiting yourself to 5 concurrent jobs. Need to run more? See the [antiSMASH install guide](#) for instructions for getting your own antiSMASH installation.














anti
SMASH

antibiotics & Secondary Metabolite Analysis SShell
Version 4.2.0



Select Gene Cluster:

 Overview          10 11        19    

Identified secondary metabolite clusters

Cluster	Type	From	To	Most similar known
The following clusters are from record c00552_NODE_55.. (original name was: NODE_552_length_5782_cov_0.632764):				
Cluster 1	Other	1	5782	-
The following clusters are from record c00573_NODE_57.. (original name was: NODE_573_length_5701_cov_0.539986):				
Cluster 2	T3pks	1	5701	-
The following clusters are from record c00895_NODE_89.. (original name was: NODE_895_length_4918_cov_0.147126):				
Cluster 3	T1pks	1	4918	-
The following clusters are from record c02406_NODE_24.. (original name was: NODE_2406_length_3433_cov_0.091782):				
Cluster 4	Otherks	1	3433	-
The following clusters are from record c02525_NODE_25.. (original name was: NODE_2525_length_3365_cov_0.089712):				
Cluster 5	Nrps	1	3365	-
The following clusters are from record c02784_NODE_27.. (original name was: NODE_2784_length_3235_cov_0.116390):				
Cluster 6	Terpene	1	3235	-

<https://antismash.secondarymetabolites.org/>



Gene annotation

Identification of Carbohydrate-Active enZymes - CAZY Database



HOME ENZYME CLASSES ASSOCIATED MODULES GENOMES

Family Go

What's new
Definitions and Terminology
Help
Functional Data
Citing CAZY
PULDB
Enzyme & Glyco Resources
Commercial Providers
Scientific Meetings
About Us
Position(s) available

Suivre @CAZyDB

Welcome to the Carbohydrate-Active enZymes Database

The **CAZY** database describes the families of structurally-related catalytic and carbohydrate-binding modules (or functional domains) of enzymes that degrade, modify, or create glycosidic bonds.

Online since 1998, CAZY is a specialist database dedicated to the display and analysis of genomic, structural and biochemical information on Carbohydrate-Active Enzymes (CAZymes).

CAZY data are accessible either by browsing sequence-based families or by browsing the content of genomes in carbohydrate-active enzymes. New genomes are added regularly shortly after they appear in the daily releases of GenBank. New families are created based on published evidence for the activity of at least one member of the family and all families are regularly updated, both in content and in description.

An original aspect of the CAZY database is its attempt to cover all carbohydrate-active enzymes across organisms and across subfields of glycosciences. Please let us know if some families have escaped our attention, we will be happy to add them !

For a more extensive encyclopedic resource on the particular features of carbohydrate active enzymes, please visit [CAZypedia](#), a web site driven by the scientific community that studies these enzymes.

Reference for the CAZY database : In the 2014 database issue of Nucleic Acids Research, we summarized the many changes that have occurred in the CAZY database during the five previous years. Read the [Abstract](#) or the full [paper](#).

A new tool associated with the CAZY database ! PULDB is a database of Polysaccharide Utilization Loci (PULs) in Bacteroidetes. **PULDB** displays information on experimentally determined and predicted PULs for a number of Bacteroidetes genomes. Read the [Abstract](#) or the full [paper](#).

Enzyme Classes currently covered

Modules that catalyze the breakdown, biosynthesis or modification of carbohydrates and glycoconjugates :

- **Glycoside Hydrolases (GHs)** : hydrolysis and/or rearrangement of glycosidic bonds (see CAZypedia [definition](#))
- **GlycosylTransferases (GTs)** : formation of glycosidic bonds (see [definition](#))
- **Polysaccharide Lyases (PLs)** : non-hydrolytic cleavage of glycosidic bonds
- **Carbohydrate Esterases (CEs)** : hydrolysis of carbohydrate esters
- **Auxiliary Activities (AAs)** : redox enzymes that act in conjunction with CAZymes.

Associated Modules currently covered

Carbohydrate-active enzymes often display a modular structure with non-catalytic modules appended to the enzymes above

- **Carbohydrate-Binding Modules (CBMs)** : adhesion to carbohydrates

Query	Subject	E-value	Subject start	Subject end	Query start	Query end	Covered fraction
fig 6666666.197029.peg.1003	GH109.hmm	6.90E-12	1	117	2	112	0.920634921
fig 6666666.197029.peg.1015	GT4.hmm	3.00E-35	8	157	193	340	0.93125
fig 6666666.197029.peg.1034	PL17.hmm	0.00017	2	84	577	648	0.589928058
fig 6666666.197029.peg.1083	GH38.hmm	3.20E-28	25	182	33	183	0.583643123
fig 6666666.197029.peg.1124	GH117.hmm	2.60E-06	64	161	39	125	0.45971564
fig 6666666.197029.peg.1151	CE14.hmm	1.10E-16	1	124	9	127	0.991935484
fig 6666666.197029.peg.1183	GH109.hmm	2.70E-09	63	121	23	84	0.46031746
fig 6666666.197029.peg.1208	GH109.hmm	2.30E-12	3	112	6	106	0.865079365
fig 6666666.197029.peg.1232	GH109.hmm	1.70E-07	2	121	25	145	0.944444444
fig 6666666.197029.peg.1233	GH74.hmm	2.70E-11	30	117	164	254	0.373390558
fig 6666666.197029.peg.1233	GH74.hmm	4.70E-08	43	118	228	301	0.321888412
fig 6666666.197029.peg.1247	PL12.hmm	1.70E-27	1	138	388	516	0.992753623
fig 6666666.197029.peg.127	GH109.hmm	1.60E-12	2	122	5	119	0.952380952
fig 6666666.197029.peg.1289	GH109.hmm	1.60E-07	1	122	10	133	0.96031746
fig 6666666.197029.peg.1297	GH109.hmm	2.20E-08	4	121	5	111	0.928571429
fig 6666666.197029.peg.130	GH32.hmm	1.80E-08	43	171	93	232	0.436860068
fig 6666666.197029.peg.1325	GH109.hmm	1.20E-12	4	105	9	105	0.801587302
fig 6666666.197029.peg.1326	GH109.hmm	1.10E-09	4	118	10	115	0.904761905
fig 6666666.197029.peg.1327	GH109.hmm	2.30E-08	3	122	6	116	0.944444444
fig 6666666.197029.peg.1334	CE10.hmm	9.80E-16	99	316	3	221	0.636363636
fig 6666666.197029.peg.1340	GH28.hmm	7.60E-07	59	217	148	307	0.486153846
fig 6666666.197029.peg.1359	CBM9.hmm	1.60E-32	1	181	39	240	0.989010989

<http://www.cazy.org>



Gene annotation

Accurate classifier for hydrogenase sequences - HydDB



Classify

HydDB provides access to an accurate classifier for hydrogenase sequences and a curated database of hydrogenases by known type. The service is provided by the School of Biological Sciences, Monash University and the Bioinformatics Research Centre, Aarhus University.

Please cite! If you use HydDB for research, please cite the following paper: "Søndergaard D, Pedersen CNS, Greening C. **HydDB: A web tool for hydrogenase classification and analysis**. Sci Rep. 2016;6:34212. doi: 10.1038/srep34212.". The preprint is available on [bioRxiv](#). If you have any comments, corrections or questions contact [Chris Greening](#) or [Dan Søndergaard](#).

Classify

HydDB is unable to accurately check whether uploaded sequences correspond to hydrogenases or not. Instead, it is well-suited for functionally-predictive classification of known hydrogenases into different subgroups. Please ensure that all sequences that you upload correspond to catalytic subunits of hydrogenases (e.g. using conserved domain database and phylogenetic trees). Sequences that do not encode catalytic subunits of hydrogenases will still be classified, but the result may be wrong.

Sequences

Sequences File

Choose File no file selected

Instructions

To use the classifier to predict the type of one or more hydrogenases from sequence, either:

- paste your FASTA-formatted protein sequences into the text area, or
- upload a FASTA-formatted file with your protein sequences.

Press the "Submit" button to upload the sequences and begin the classification.

If you provided an e-mail address you will receive an e-mail when your job finishes or fails including a link to the results. You will also be able to download the results as a CSV file.

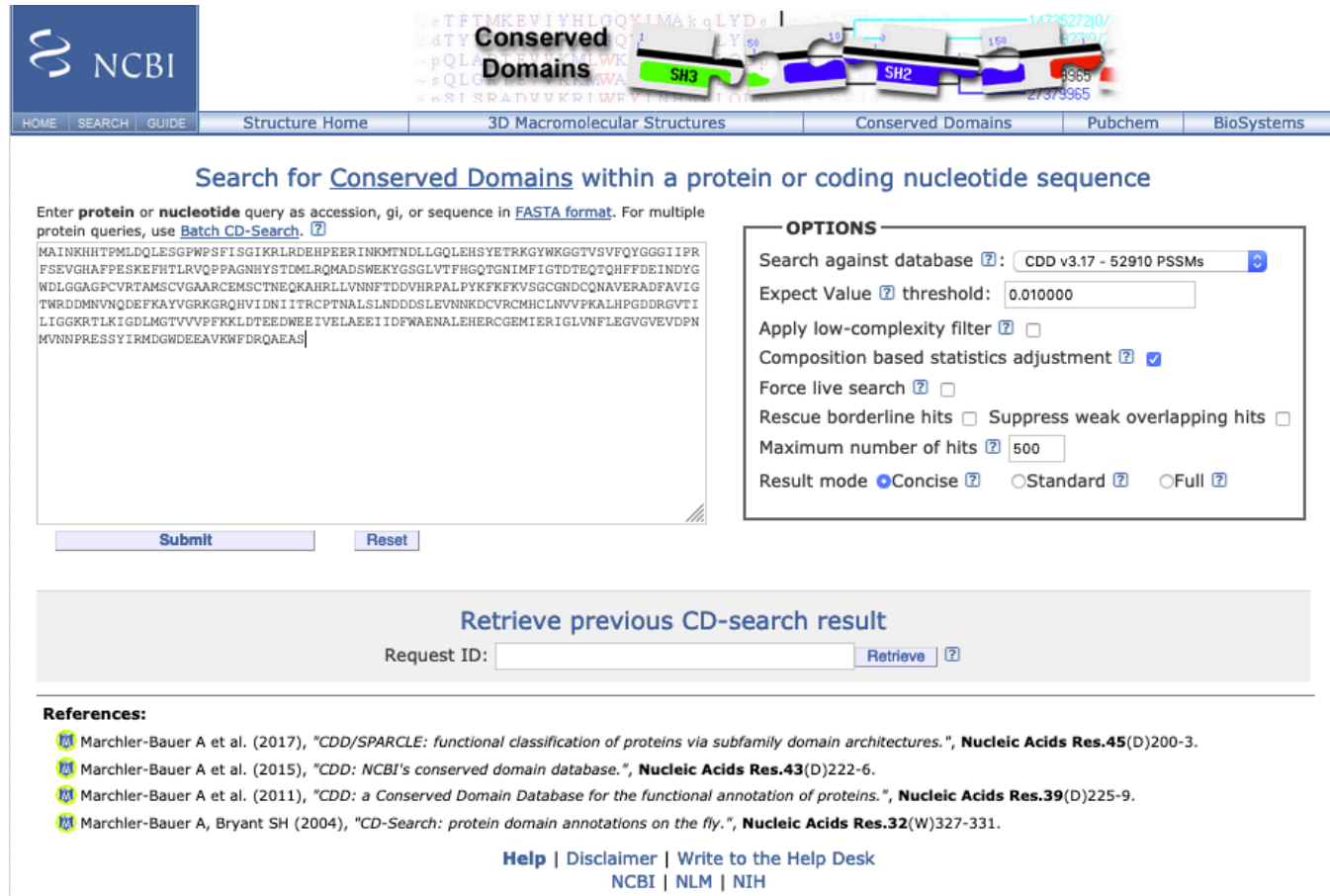
Only sequences encoding the catalytic subunits of hydrogenases will be classified, i.e. those binding the [NiFe]-centre (NiFe-hydrogenases), [FeFe]-centre (FeFe-hydrogenases), or [Fe]-centre (Fe-hydrogenases). Electron-transfer subunits, accessory proteins, and maturation factors cannot be classified by this service.

Limits

<https://services.birc.au.dk/hyddb/>



NCBI Conserved Domain Search



The image shows the NCBI Conserved Domain Search web interface. At the top, there's a navigation bar with links to HOME, SEARCH, GUIDE, Structure Home, 3D Macromolecular Structures, Conserved Domains, Pubchem, and BioSystems. Below this is a search box with a placeholder text: "Search for Conserved Domains within a protein or coding nucleotide sequence". To the right of the search box is a diagram of a protein structure with domains labeled SH3, SH2, and SH1. Below the search box is a text area for entering a protein or nucleotide query. To the right of the text area is an "OPTIONS" box with various search parameters: Search against database (CDD v3.17 - 52910 PSSMs), Expect Value threshold (0.010000), Apply low-complexity filter (unchecked), Composition based statistics adjustment (checked), Force live search (unchecked), Rescue borderline hits (unchecked), Suppress weak overlapping hits (unchecked), Maximum number of hits (500), and Result mode (Concise selected). Below the search box are "Submit" and "Reset" buttons. Below the search box is a section for "Retrieve previous CD-search result" with a "Request ID:" field and a "Retrieve" button. At the bottom, there are "References:" and "Help | Disclaimer | Write to the Help Desk" links.

NCBI

HOME SEARCH GUIDE Structure Home 3D Macromolecular Structures Conserved Domains Pubchem BioSystems

Search for **Conserved Domains** within a protein or coding nucleotide sequence

Enter **protein** or **nucleotide** query as accession, gi, or sequence in [FASTA format](#). For multiple protein queries, use [Batch CD-Search](#).

MAINKHHHTPMLDQLESGPWPFSFISGIKRLRDEHPEERINKMTNDLLGQLEHSYETKRGYWKGGTVSVFQYGGGIIPR
FSEVGHAFPESEKFTLRVQPPAGNHYSTMLRQMADEMEKYGSLVTFHGQTGNIMFIGTDEQTHFFDEINDYG
NDLGGAGPCVRTAMSCVGAARCEMSTNEQKAHRLLVNFTDDVHRPALPYKFKFVSGCGNDQNAVERADFAVIG
TWRDDMNQVQDEFKAYVGRKGRQVIDNIITRCPTNALSLNDDSLVNNKDCVRCMHCLNVVPKALHPGDDRGVTI
LIGGKRTLKIGDLMTGVVVPFKKLDTEEDWEBIVELAEIIDFWAENALEHERCGEMIERIGLVNPLEGVGVEVDEN
MVNNPRESSYIRMDGWDEEAVKWFDRQAEAS

OPTIONS

Search against database

Expect Value threshold:

Apply low-complexity filter ☐

Composition based statistics adjustment ☒

Force live search ☐

Rescue borderline hits ☐ Suppress weak overlapping hits ☐

Maximum number of hits

Result mode ☒ Concise ☐ Standard ☐ Full

Retrieve previous CD-search result

Request ID: [Retrieve](#)

References:

- Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", **Nucleic Acids Res.**45(D)200-3.
- Marchler-Bauer A et al. (2015), "CDD: NCBI's conserved domain database.", **Nucleic Acids Res.**43(D)222-6.
- Marchler-Bauer A et al. (2011), "CDD: a Conserved Domain Database for the functional annotation of proteins.", **Nucleic Acids Res.**39(D)225-9.
- Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", **Nucleic Acids Res.**32(W)327-331.

[Help](#) | [Disclaimer](#) | [Write to the Help Desk](#)
NCBI | NLM | NIH

Search nucleotide/protein sequence(s) for conserved domains

Individual search: <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

Batch: <https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>



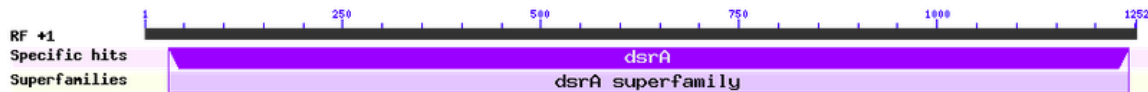
Conserved domains on [cl|1]

View Concise Results

NZ_JRAA01000001.1:c722683-721433 Solemya velum gill symbiont strain WH SV_sym_Scaffold_1, whole genome shotgun sequence

Graphical summary

☐ Zoom to residue level show extra options



Search for similar domain architectures

Refine search

List of domain hits

Name	Accession	Description	Interval	E-value
dsrA	TIGR02064	sulfite reductase, dissimilatory-type alpha subunit; Dissimilatory sulfite reductase catalyzes ...	31-1242	0e+00

sulfite reductase, dissimilatory-type alpha subunit; Dissimilatory sulfite reductase catalyzes the six-electron reduction of sulfite to sulfide, as the terminal reaction in dissimilatory sulfate reduction. It remains unclear however, whether trithionate and thiosulfate serve as intermediate compounds to sulfide, or as end products of sulfite reduction. Sulfite reductase is a multisubunit enzyme composed of dimers of either alpha/beta or alpha/beta/gamma subunits, each containing a siroheme and iron sulfur cluster prosthetic center. Found in sulfate-reducing bacteria, these genes are commonly located in an unidirectional gene cluster. This model describes the alpha subunit of sulfite reductase. [Central intermediary metabolism, Sulfur metabolism]

Pssm-ID: 273948 [Multi-domain] Cd Length: 402 Bit Score: 667.31 E-value: 0e+00

1	11	LDQLESGPWPSPFISG	12	IKRLRDEHPEERINKMTNDLLG	13	LEHSYETRGYWKGGTVSVFQYGGGIIPRFSEVGHAF	14	FPESKE	15	90
Cdd:TIGR02064	1	LDQLEKGPWPSPFVSEIKKTAAYRADYQVPVDPEDLLGVLELSYDERKTHWKGGIVSVFGYGGGVIGRYS	2	DQGEKFP	3	GVAE	4	80		
1	91	FHTLRVQPPAGNHYSTDLRQ	92	MADSWKEYSGGLVTFHGQTGNIMFIGTDEQTQHFDEINDYGWDLGGAGPCVRTAMSC	93	170				
Cdd:TIGR02064	81	FHTVRVQPPSKFYSTDYLRQLCDVWEKYSGLTNFHGQTDIVFLGTQTPQLQETFEELTNLGTDLGGSGSNLRT	82	PESC	83	160				
1	171	VGAARCEMSCTNEQKAHRLLVNNFTDDVHRPALPYKFKFVSGCNDQNAVERADFAVIGTWDRDMMVNQDE	172	FKAYVGR	173	250				
Cdd:TIGR02064	161	VGPARCFACVDTLKACYTELMEYQDELHRPAPFYKFKFVSGCPNDCAAIARSDFAVIGTWKDDIKVDQ	162	EAVKAYIAG	163	240				
1	251	KGRQHVINDITRCPTNALSLNDDDSL	252	EVNNKDCVRCMHCLNVVPKALHPGDDRGVTTILIGGKRTLKIGDLMGTVVVPFK	253	330				
Cdd:TIGR02064	241	WGFDEIENEVNRCPKKAISWDGSKELSIDNRECVRCMHCHINKMPKALHPGDE	242	RGVTTILIGGKAPILDGAQMGWVVPFV	243	320				
1	331	kLDTEDWEIEVLAETIIDFNAENALEHERCGEMIERIGLVNFLGVCVGVDPDMVNVPRESSYIRMDGWDEEAVKWF	332	410						
Cdd:TIGR02064	321	--EAEPPYDEIKELVEKIIDWDEEGKNRERIGETIKRLGLQKFLFVIGIEPDPQMVKEPRTNPYIFFKVEDEVPGG	322	WDA	323	398				
1	411	RQAE	412	414						
Cdd:TIGR02064	399	DIAE	400	402						

Conserved Domain Search results:
dsrA gene of *Solemya velum* gill symbiont strain WH

References:

Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", *Nucleic Acids Res.*45(D)200-3.





[Refine search](#)

?

sulfite reductase, dissimilatory-type alpha subunit; Dissimilatory sulfite reductase catalyzes the six-electron reduction of sulfite to sulfide, as the terminal reaction in dissimilatory sulfate reduction. It remains unclear however, whether trithionate and thiosulfate serve as intermediate compounds to sulfide, or as end products of sulfite reduction. Sulfite reductase is a multisubunit enzyme composed of dimers of either alpha/beta or alpha/beta/gamma subunits, each containing a siroheme and iron sulfur cluster prosthetic center. Found in sulfate-reducing bacteria, these genes are commonly located in an unidirectional gene cluster. This model describes the alpha subunit of sulfite reductase. [Central intermediary metabolism, Sulfur metabolism]

```

1          411  RQAE  414
Cdd:TIGR02064 399  DIAE  402

```


References:

Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", **Nucleic Acids Res.**45(D)200-3.



Gene annotation

KEGG: <https://www.genome.jp/kegg/pathway.html>



KEGG PATHWAY Database
Wiring diagrams of molecular interactions, reactions and relations

Menu PATHWAY BRITE MODULE KO GENES LIGAND NETWORK DISEASE DRUG DBGET

Select prefix: map Organism Enter keywords: Go Help

[New pathway maps | Update history]

Pathway Maps

KEGG PATHWAY is a collection of manually drawn [pathway maps](#) representing our knowledge on the molecular interaction, reaction and relation networks for:

- 1. Metabolism**
Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
- 2. Genetic Information Processing**
- 3. Environmental Information Processing**
- 4. Cellular Processes**
- 5. Organismal Systems**
- 6. Human Diseases**
- 7. Drug Development**

KEGG PATHWAY is the reference database for pathway mapping in **KEGG Mapper**.

Pathway Identifiers

Each pathway map is identified by the combination of 2-4 letter prefix code and 5 digit number (see **KEGG Identifier**). The prefix has the following meaning:

Prefix	Meaning
map	manually drawn reference pathway
ko	reference pathway highlighting KOs
ec	reference metabolic pathway highlighting EC numbers
rn	reference metabolic pathway highlighting reactions
<org>	organism-specific pathway generated by converting KOs to gene identifiers

and the numbers starting with the following:


Number	Meaning
011	global map (lines linked to KOs)
012	overview map (lines linked to KOs)
010	chemical structure map (no KO expansion)
07	drug structure map (no KO expansion)
other	regular map (boxes linked to KOs)

are used for different types of maps.

1. Metabolism

1.0 Global and overview maps

- 01100 Metabolic pathways
- 01110 Biosynthesis of secondary metabolites
- 01120 Microbial metabolism in diverse environments
- 01130 Biosynthesis of antibiotics
- 01200 Carbon metabolism
- 01210 2-Oxocarboxylic acid metabolism



KEGG Pathway Maps

[Brite menu | Download htext | Download json]

KEGG pathway maps Go

☐ One-click mode

Metabolism

- Global and overview maps
 - 01100 Metabolic pathways
 - 01110 Biosynthesis of secondary metabolites
 - 01120 Microbial metabolism in diverse environments
 - 01130 Biosynthesis of antibiotics
 - 01200 Carbon metabolism
 - 01210 2-Oxocarboxylic acid metabolism
 - 01212 Fatty acid metabolism
 - 01230 Biosynthesis of amino acids
 - 01220 Degradation of aromatic compounds
- Carbohydrate metabolism
- Energy metabolism
 - 00190 Oxidative phosphorylation
 - 00195 Photosynthesis
 - 00196 Photosynthesis - antenna proteins
 - 00710 Carbon fixation in photosynthetic organisms
 - 00720 Carbon fixation pathways in prokaryotes
 - 00680 Methane metabolism
 - 00910 Nitrogen metabolism
 - 00920 Sulfur metabolism
- Lipid metabolism
- Nucleotide metabolism
- Amino acid metabolism
- Metabolism of other amino acids
- Glycan biosynthesis and metabolism
- Metabolism of cofactors and vitamins
- Metabolism of terpenoids and polyketides
- Biosynthesis of other secondary metabolites
- Xenobiotics biodegradation and metabolism
- Chemical structure transformation maps

Gene annotation

KEGG: <https://www.genome.jp/kegg/pathway.html>



[[Brite menu](#) | [Download htext](#) | [Download json](#)]

KEGG pathway maps

Go

▼ ▼ ▼ ☐ One-click mode

▼ Metabolism

▼ Global and overview maps

01100 Metabolic pathways

01110 Biosynthesis of sec
01120 Microbial metabolism

01120 Microbial metabolism in diverse enviro
01130 Biosynthesis of antibiotics

01130 Biosynthesis of amino acids
01200 Carbon metabolism

01210 2-Oxocarboxylic ac

01210 2-Oxocarboxylic acid metabolism
01212 Fatty acid metabolism

01230 Biosynthesis of am.

01220 Degradation of aromatic compounds

Carbohydrate metabolism

Energy metabolism

00190 Oxidative phosphorylation
00195 Photosynthesis

00195 Photosynthesis
00196 Photosynthesis

00196 Photosynthesis - antenna proteins
00710 Carbon fixation in photosynthetic

00710 Carbon fixation in photosynthetic
00720 Carbon fixation pathways in proka

00680 Methane metabolism

00910 Nitrogen metabolism

00920 Sulfur metabolism

Lipid metabolism

Nucleotide metabo

Amino acid metabolism

Metabolism of other amino acids
Glucose biosynthesis and metabol

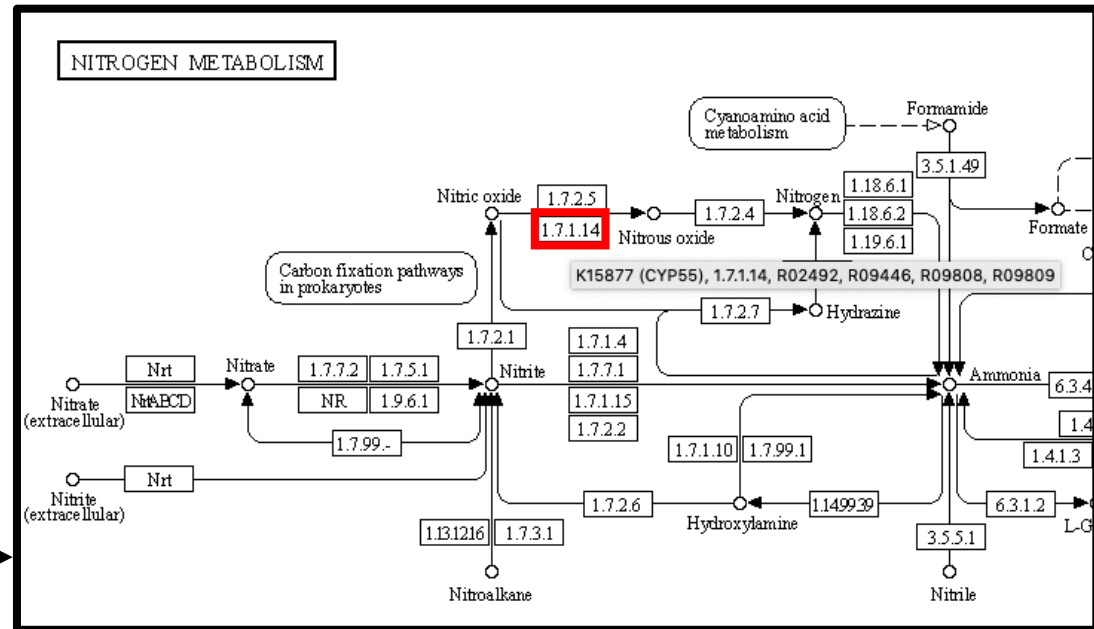
metabolism of cofactors and vit

metabolism of terpenoids and polyketides

biosynthesis of other secondary metabolites

Xenobiotics biodegradation and metabolism

Chemical structure transformation maps



Gene annotation

KEGG: <https://www.genome.jp/kegg/pathway.html>

KEGG Pathway Maps

[[Brite menu](#) | [Download htext](#) | [Download json](#)]

KEGG pathway maps

▼ ▼ ▼ ☐ One-click mode

▼ Metabolism


- ▼ Global and overview maps
 - 01100 Metabolic pathways
 - 01110 Biosynthesis of secondary metabolites
 - 01120 Microbial metabolism in diverse environments
 - 01130 Biosynthesis of antibiotics
 - 01200 Carbon metabolism
 - 01210 2-Oxocarboxylic acid metabolism
 - 01212 Fatty acid metabolism
 - 01230 Biosynthesis of amino acids
 - 01220 Degradation of aromatic compounds
- ▶ Carbohydrate metabolism
- ▼ Energy metabolism
 - 00190 Oxidative phosphorylation
 - 00195 Photosynthesis
 - 00196 Photosynthesis - antenna proteins
 - 00710 Carbon fixation in photosynthetic organisms
 - 00720 Carbon fixation pathways in prokaryotes
 - 00680 Methane metabolism
 - 00910 Nitrogen metabolism
 - 00920 Sulfur metabolism
- ▶ Lipid metabolism
- ▶ Nucleotide metabolism
- ▶ Amino acid metabolism
- ▶ Metabolism of other amino acids
- ▶ Glycan biosynthesis and metabolism
- ▶ Metabolism of cofactors and vitamins
- ▶ Metabolism of terpenoids and polyketides
- ▶ Biosynthesis of other secondary metabolites
- ▶ Xenobiotics biodegradation and metabolism
- ▶ Chemical structure transformation maps

KEGG ORTHOLOGY: K15877	
Entry	K15877 KO
Name	CYP55
Definition	fungal nitric oxide reductase [EC:1.7.1.14]
Pathway	ko00910 Nitrogen metabolism ko01100 Metabolic pathways ko01120 Microbial metabolism in diverse environments
Brite	KEGG Orthology (KO) [BR:ko00001] 09100 Metabolism 09102 Energy metabolism 00910 Nitrogen metabolism K15877 CYP55; fungal nitric oxide reductase Enzymes [BR:ko01000] 1. Oxidoreductases 1.7 Acting on other nitrogenous compounds as donors 1.7.1 With NAD+ or NADP+ as acceptor 1.7.1.14 nitric oxide reductase [NAD(P)+, nitrous oxide-fo K15877 CYP55; fungal nitric oxide reductase BRITE hierarchy
Other DBs	RN: R02492 R09446 R09808 R09809 GO: 0016966



Gene annotation

Metacyc: experimentally curated metabolic pathways



METACYC
A member of the BioCyc database collection

Two-day Introduction to
Pathway Tools Tutorial
Jan 16-17, 2020
Early Registration by Dec 12

Sites ▾ Search ▾ Genome ▾ Metabolism ▾ Analysis ▾ SmartTables ▾ Help ▾

Search Results for *dsrA*
using database *MetaCyc* what is this?

Genes (3) | Proteins (3) | EC Numbers (2)

Genes Gene/Gene Product pages contain: chromosomal location of gene; depiction of its operon; link to genome browser; detailed summaries of complexes; cofactors, activators, and inhibitors (for enzymes), depiction of regulon (for transcriptional regulators), protein features.

- *dsrA* - *Allochromatium vinosum*
- *dsrA* - *Desulfovibrio gigas*
- *dsrA* - *Archaeoglobus fulgidus*

Login to turn into a SmartTable.

Proteins Gene/Gene Product pages contain: chromosomal location of gene; depiction of its operon; link to genome browser; detailed summary of regulon (for transcriptional regulators), protein features.

- siroheme sulfite reductase, α subunit (*DsrA*) - *Allochromatium vinosum*
- sulfite reductase α subunit (*DsrA*) - *Desulfovibrio gigas*
- sulfite reductase, dissimilatory α subunit (*DsrA*) - *Archaeoglobus fulgidus*

Login to turn into a SmartTable.


EC Numbers EC Number pages contain: links to reaction and enzymes associated with the EC number in this database, names, description.

<https://metacyc.org/>



Gene annotation

Metacyc: experimentally curated metabolic pathways



Two-day Introduction to
Pathway Tools Tutorial
Jan 16-17, 2020
Early Registration by Dec 12

Sites Search Genome Metabolism Analysis SmartTables Help

Search Results for **dsrA**
using database *MetaCyc* what is this?

Genes (3) | Proteins (3) | EC Numbers (2)

Genes Gene/Gene Product pages contain: chromosomal location of gene; depiction of its open reading frame; cofactors, activators, and inhibitors (for enzymes), depiction of regulon (for transcriptional regulators).

Proteins Gene/Gene Product pages contain: chromosomal location of gene; depiction of its open reading frame; protein features.

- siroheme sulfite reductase, α subunit (**DsrA**) - *Allochromatium vinosum*
- sulfite reductase α subunit (**DsrA**) - *Desulfovibrio gigas*
- sulfite reductase, dissimilatory α subunit (**DsrA**) - *Archaeoglobus fulgidus*

Login to turn into a SmartTable.

EC Numbers EC Number pages contain: links to reaction and enzymes associated with the EC number.

Sites Search Genome Metabolism Analysis SmartTables Help

gene **dsrA** protein **sulfite reductase α complex**
Desulfovibrio gigas

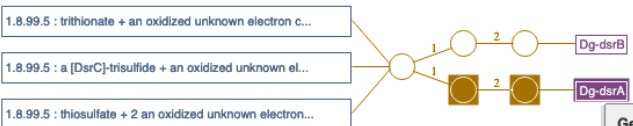
Accession ID	G-385 (MetaCyc)
Reactions	thiosulfate + 2 an oxidized unknown electron carrier + 3 H ₂ O \leftarrow 2 sulfite + 2 a reduced unknown electron carrier + 3 H ⁺ (catalyzed by complex a [DsrC]-trisulfide + an oxidized unknown electron carrier + 3 H ₂ O \leftarrow sulfite + a [DsrC protein]-dithiol + a reduced unknown electron carrier + 3 H ⁺ (catalyzed by complex a) + 3 H ₂ O \leftrightarrow 3 sulfite + a reduced unknown electron carrier + 4 H ⁺ (catalyzed by complex a))
Pathways	dissimilatory sulfate reduction I (to hydrogen sulfide) dissimilatory sulfate reduction II (to thiosulfate)

Summary GO Terms (1) Reactions (3) References Show All

Subunit Composition [DsrA]₂

Component of dissimilatory sulfite reductase (extended summary available):
[(DsrA)₂][(DsrB)₂]

Gene-Reaction Schematic



Gene: dsrA G-385
Product: sulfite reductase α subunit, subunit of sulfite reductase α complex, dissimilatory sulfite reductase
Species: *Desulfovibrio gigas*

GO Terms:
Cellular Component: GO:0005829 - cytosol []

Enzymatic activity: sulfite reductase (thiosulfate-forming) (dissimilatory sulfite reductase)

2 sulfite + 2 a reduced unknown electron carrier + 3 H⁺ \rightarrow thiosulfate + 2 an oxidized unknown electron carrier + 3 H₂O

<https://metacyc.org/>



Gene annotation

- The **PSORT** family - prediction of protein localization sites in cells.
- Useful for making cell schematics!



[Updates](#) | [Documentation](#) | [Resources](#) | [Contact](#)

Submit a Sequence to PSORTb version 3.0.2

Based on a study last performed in 2010, PSORTb v3.0.2 is the most precise bacterial localization prediction tool available. PSORTb v3.0.2 has a number of [improvements](#) over PSORTb v2.0.4. Version 2 of PSORTb is maintained [here](#).

You can currently submit one or more Gram-positive or Gram-negative bacterial sequences or archaeal sequences in FASTA format (?). Copy and paste your FASTA-formatted sequences into the textbox below or select a file containing your sequences to upload from your computer. Web display mode is limited to the analysis of approximately 100 proteins. For larger analyses, either enter your email address in the form below (results of up to 5000 per submission returned by email) or for even larger analyses we can help you or you can download the standalone version.

See also:

- [Updates](#)
- [Precomputed genome results](#)
- [Limitations of PSORTb v.3.0](#)
- [PSORTb User's Guide](#)
- [Docker PSORTb web service](#) (what is [docker](#)?)
- [Download standalone PSORTb](#)
- [Docker standalone PSORTb](#) (what is [docker](#)?)

<http://psort.org/>

Choose an organism type (?):
 Required

Choose Gram stain (?):
 Required

Advanced Gram stain options (?):
 Required

Output format (?):

Show results (?):

Email address:

Copy and paste your FASTA sequences below

or upload from file:
(uploads limited to 50KB, approximately 100 proteins, in Web display mode, enter an email address to use email mode if you need to analyze more proteins)

no file selected



Task: Analyze data for group work

Determine which genome(s) have the following attributes, and the genetic mechanisms used for these attributes:

1. Denitrification (Nitrate or nitrite to nitrogen)
2. Ammonia oxidation (Ammonia to nitrite or nitrate)
3. Anammox (Ammonia and nitrite to nitrogen)
4. Sulfur oxidation (SOX pathway, thiosulfate to sulfate)
5. Sulfur reduction (DSR pathway, sulfate to sulfide)
6. Photosynthetic carbon fixation
7. Non-photosynthetic carbon fixation (Reverse TCA or Wood-Ljungdahl)
8. Non-polar flagella expression due to a chromosomal deletion
9. Plasmid-encoded antibiotic resistance
10. Aerobic (versus anaerobic) metabolism



Summary of online resources

Resources to help interpret your data:

- KEGG: <https://www.genome.jp/kegg/pathway.html>
- BioCyc: <https://biocyc.org/>
- MetaCyc: <https://metacyc.org/>
- HydDB: <https://services.birc.au.dk/hyddb/>
- PSORT: <https://psort.hgc.jp/>



**Optional: work with own data, or
continue group task**

