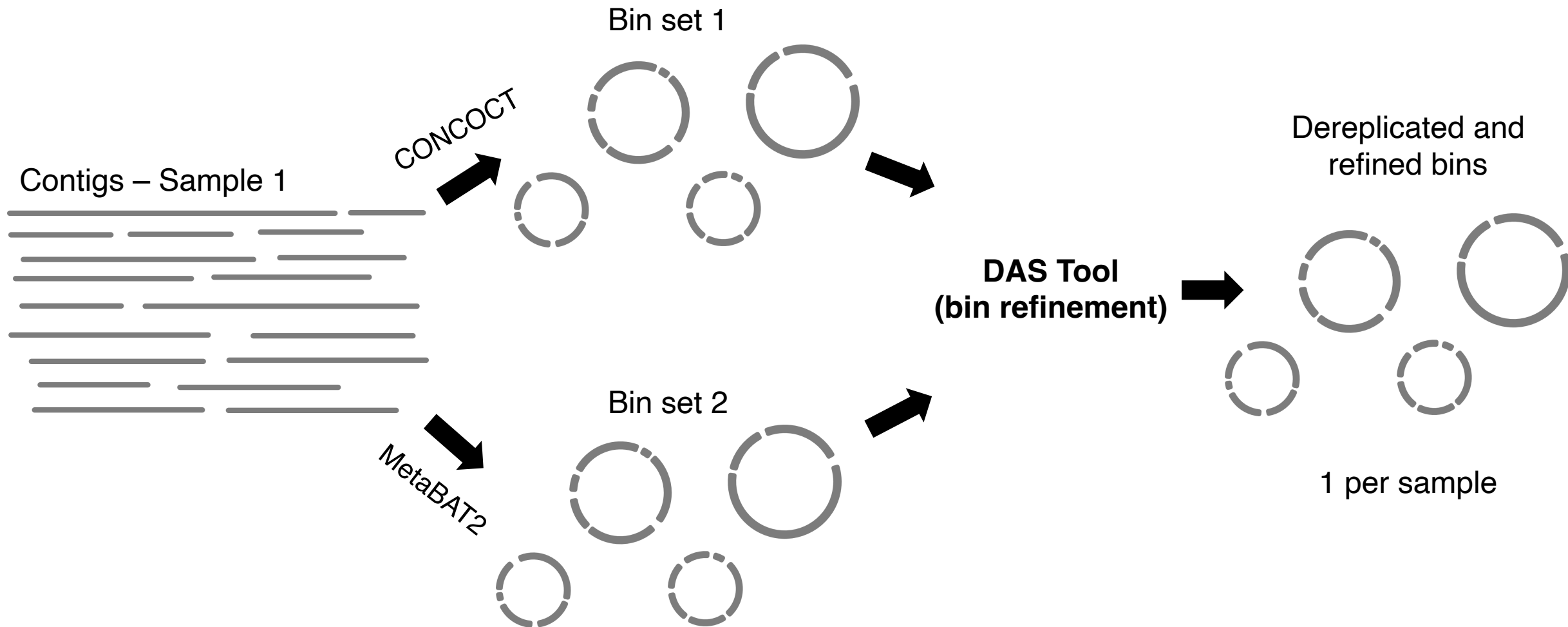
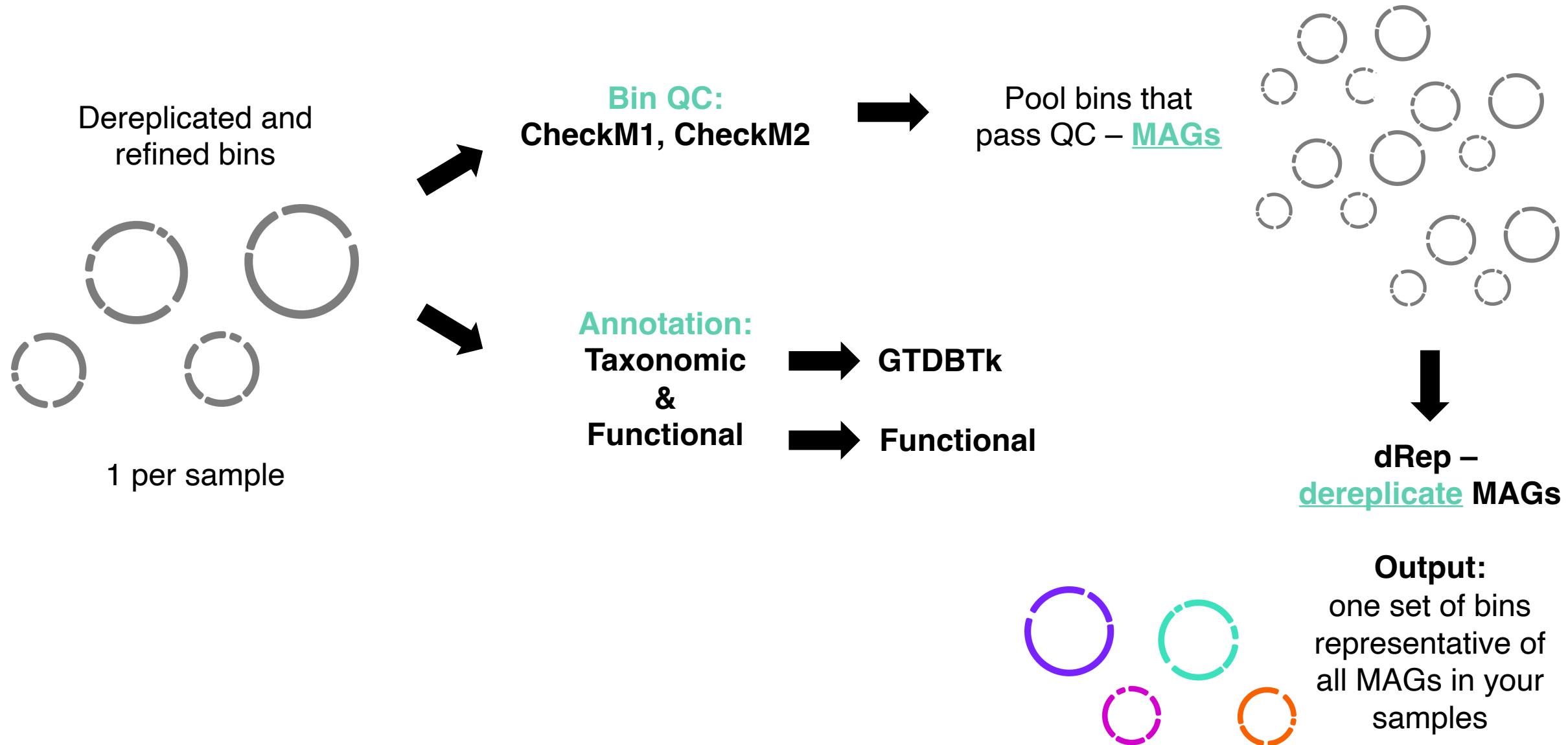


Workflow



Workflow



Taxonomic and functional annotation

Instructor: Dr. Emma Bell
(she/her)

December 15th, 2023

Hands-on bioinformatics for microbial meta-omics (ENV 621)

Outline

- Lecture
 - Introduction to taxonomic and functional annotation databases
 - Pairwise alignment vs. HMMs
 - Caveats of assigning function
- Case studies
- Practical
 - Evaluate bins
 - Taxonomically classify MAGs
 - Functionally annotate MAGs
 - Create a representative bin set

Annotation

```
>NG502_T1_k141_14455
TCGCTTCGTCTGTTGTAAAATGAAAGTGCTCTTGCATGTGAACTCTCCTT
TTGAATGTTTGGTAGCACTTTCATTTTAAAGGAGATCCTTCATGCAAGGG
CTATTTTATATACTTATCTCATTTTATCTAGCACACGGGTAAAGTTTGTT
TTTCAACCTGCGGAAGAAAGCACTGATGAATATGGATTGTCAGGATCACC
ATATATGATTCGGGCAGGCGTATATGACCATGTAGATGCCGCCATCGCCT
TGCATATGTGCCCATGGCTTCCGGTCGGAGCTGTTCAAATCAATGATGGC
TACAGTATGGCAAATGTGGATGTGTTTCGAGGGGAAAATTTATGGTACAGG
CGGCCATGGAGCTTACCCTGAACTTGGAACAGATCCGATTTGGATGCTTG
GTCCAGTGATGCAAGCTTTGTATGGAATCACTTCACGCAAAGTCTCTTCA
```

```
>NG502_T1_k141_14455_1 # 205 # 969 # 1 # ID=1_1;partial=00;st
tif=AGGA;rbs_spacer=5-10bp;gc_cont=0.437
MIRAGVYDHVDAALHMCPLVPGAVQINDGYSMANVDVFRGKIYGTGGHGAYPELGTD
PIWMLGPVMQALYGITSRKVSSLDAAVISIGRIQAGSASNIIPTEVFIEGTMRSYTPEVR
DLLELELKKAIFSIVGNLGGSYSLHVERGEPALNNHSVVNRWITKTVSEIYPDFQIVKKPF
GLGGEDFGYVSQKIPSSLFFLGCAIQDGVKRELHTPIFDIDEACLPITAILAQTVVHFL
NGKVSLSANEMLEV*
>NG502_T1_k141_14455_2 # 974 # 2038 # 1 # ID=1_2;partial=00;s
otif=GGAGG;rbs_spacer=5-10bp;gc_cont=0.412
MVHKLALIGFGVVGQGLVEILLNKQEKLQKEEGFEAVVVAISDLMKGSIIYHPDGLHLETV
LKVLIKETGNTENYPDVPGITKGWDSFKTTIRETNADTTIVEVSYTDVKTGQPATDHCKAAFE
```

.fna, .fa, .fasta – nucleotide



.faa – amino acid

Prodigal:

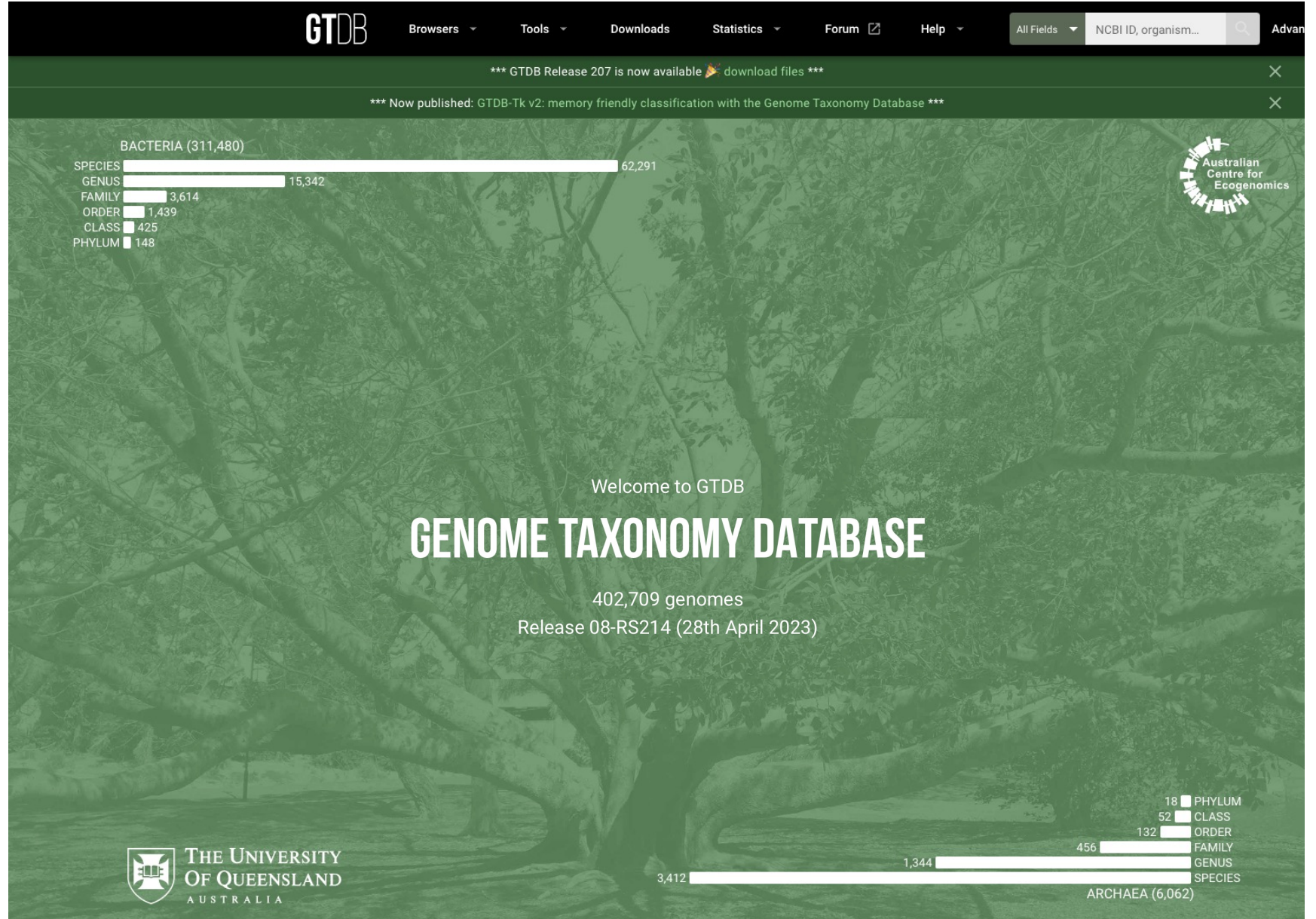
predict protein-coding genes

<https://github.com/hyattpro/Prodigal>

GTDB-Tk: Taxonomic annotation

GTDB-Tk: a toolkit for assigning
objective taxonomic
classifications to bacterial and
archaeal genomes.

 ecogenomics.github.io/GTDBTk/



The screenshot shows the GTDB (Genome Taxonomy Database) homepage. The background is a green-tinted image of a large, ancient tree. The GTDB logo is in the top left. Navigation links include Browsers, Tools, Downloads, Statistics, Forum, and Help. A search bar is in the top right. A banner at the top right says '*** GTDB Release 207 is now available download files ***'. Below this, another banner says '*** Now published: GTDB-Tk v2: memory friendly classification with the Genome Taxonomy Database ***'. On the left, a horizontal bar chart shows the taxonomic distribution for BACTERIA (311,480): SPECIES (62,291), GENUS (15,342), FAMILY (3,614), ORDER (1,439), CLASS (425), and PHYLUM (148). On the right, the Australian Centre for Ecogenomics logo is visible. In the center, the text reads 'Welcome to GTDB' followed by 'GENOME TAXONOMY DATABASE' in large letters. Below that, it says '402,709 genomes' and 'Release 08-RS214 (28th April 2023)'. At the bottom left is the University of Queensland Australia logo. At the bottom right, another horizontal bar chart shows the taxonomic distribution for ARCHAEA (6,062): PHYLUM (18), CLASS (52), ORDER (132), FAMILY (456), GENUS (1,344), and SPECIES (3,412).

GTDB

Browsers Tools Downloads Statistics Forum Help

All Fields NCBI ID, organism...

*** GTDB Release 207 is now available download files ***

*** Now published: GTDB-Tk v2: memory friendly classification with the Genome Taxonomy Database ***

BACTERIA (311,480)

SPECIES 62,291

GENUS 15,342

FAMILY 3,614

ORDER 1,439

CLASS 425

PHYLUM 148

Australian Centre for Ecogenomics

Welcome to GTDB

GENOME TAXONOMY DATABASE

402,709 genomes

Release 08-RS214 (28th April 2023)

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

18 PHYLUM

52 CLASS

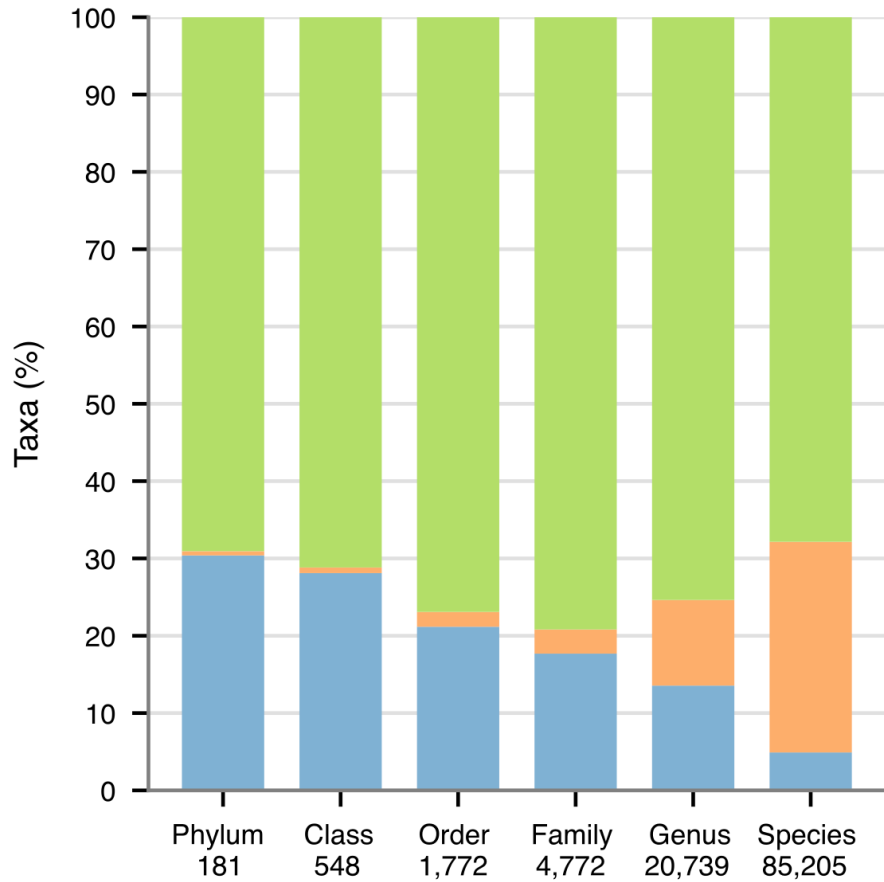
132 ORDER

456 FAMILY

1,344 GENUS

3,412 SPECIES

ARCHAEA (6,062)

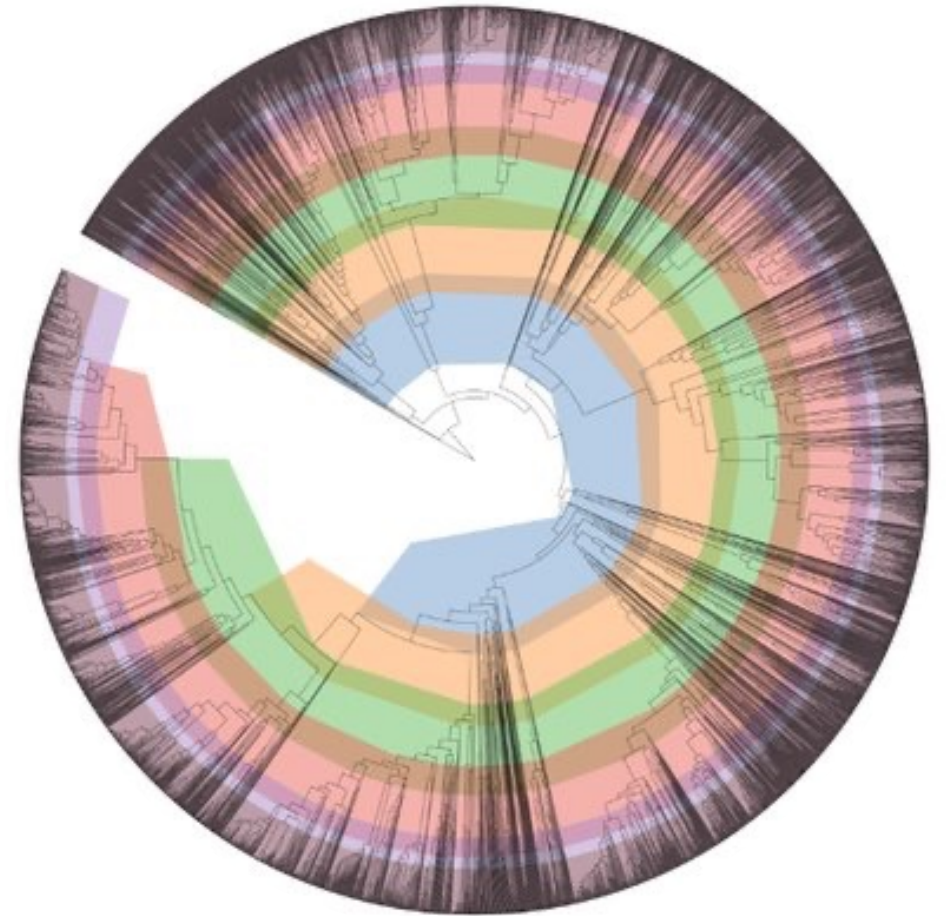


RefSeq



- Exclusively MAGs and/or SAGs
- Exclusively isolates
- Isolate and environmental genomes

GTDB



Colors indicate borders of:

- Phylum
- Class
- Order
- Family
- Genus
- Species

This is normal:

d__Bacteria;p__Verrucomicrobiota;c__Kiritimatiellia;o__SS1-B-03-39;f__UBA6053;g__UBA6053;s__

Taxa encompassing the largest number of genomes in the GTDB is given for each taxonomic ranks.

Phylum	Class	Order	Family	Genus	Species
Pseudomonadota 174,676	Gammaproteobacteria 150,865	Enterobacterales 80,364	Enterobacteriaceae 80,364	Escherichia 34,358	Escherichia coli 33,849
Bacillota 74,157	Bacilli 74,157	Lactobacillales 38,051	Staphylococcaceae 20,155	Staphylococcus 19,627	Klebsiella pneumoniae 14,975
Actinomycetota 33,994	Clostridia 33,352	Pseudomonadales 27,546	Streptococcaceae 19,088	Streptococcus 18,492	Staphylococcus aureus 14,959
Bacillota_A 33,495	Actinomycetia 28,469	Staphylococcales 20,270	Pseudomonadaceae 14,789	Klebsiella 18,145	Salmonella enterica 13,832



78.5% of sequenced
genomes belong to
four phyla



8.4% of sequenced
genomes are the
species E. coli

Accession	NCBI organism name	NCBI taxonomy	GTDB taxonomy
GCA_000008525.1	Helicobacter pylori 26695	d__Bacteria; p__Proteobacteria; c__Epsilonproteobacteria; o__Campylobacterales; f__Helicobacteraceae; g__Helicobacter; s__Helicobacter pylori	d__Bacteria; p__Campylobacterota; c__Campylobacteria; o__Campylobacterales; f__Helicobacteraceae; g__Helicobacter; s__Helicobacter pylori

Lost in literature – check GTDB for historic name

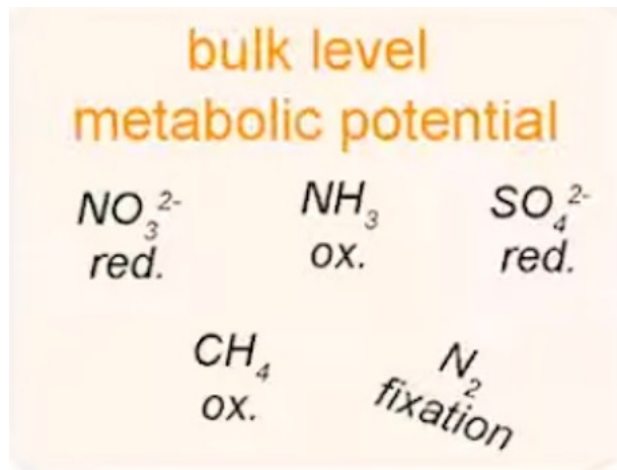
Taxonomy

The table below shows the parent taxa of the selected taxon as they appeared in specific releases.

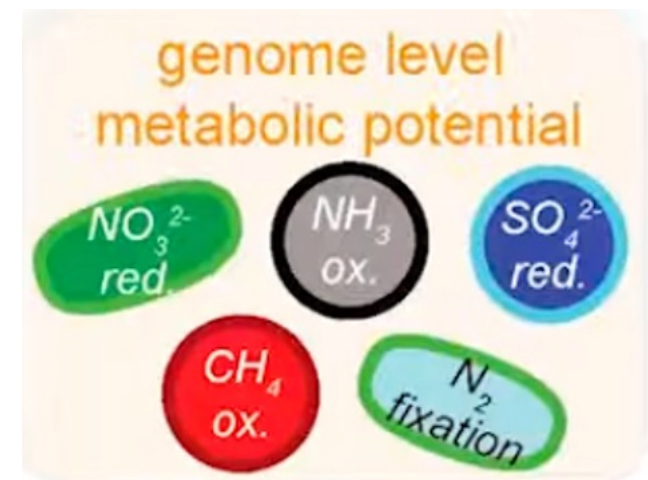
Release	Domain	Phylum	Class	Order	Family	Genus
R214	d__Bacteria	p__Bacillota_B	c__Desulfotomaculia	o__Desulfotomaculales	f__Desulfotomaculaceae	g__Desulfotomaculum
R207	d__Bacteria	p__Firmicutes_B	c__Desulfotomaculia	o__Desulfotomaculales	f__Desulfotomaculaceae	g__Desulfotomaculum
R202	d__Bacteria	p__Firmicutes_B	c__Desulfotomaculia	o__Desulfotomaculales	f__Desulfotomaculaceae	g__Desulfotomaculum
R95	d__Bacteria	p__Firmicutes_B	c__Desulfotomaculia	o__Desulfotomaculales	f__Desulfotomaculaceae	g__Desulfotomaculum
R89	d__Bacteria	p__Firmicutes_B	c__Desulfotomaculia	o__Desulfotomaculales	f__Desulfotomaculaceae	g__Desulfotomaculum
R86.2	d__Bacteria	p__Firmicutes_B	c__Desulfotomaculia	o__Desulfotomaculales	f__Desulfotomaculaceae	g__Desulfotomaculum
R83	d__Bacteria	p__Firmicutes_B	c__Desulfotomaculia	o__Desulfotomaculales	f__Desulfotomaculaceae	g__Desulfotomaculum
R80	d__Bacteria	p__Firmicutes	c__Desulfotomaculia	o__Desulfotomaculales	f__Desulfotomaculaceae	g__Desulfotomaculum

Functional annotation

Contigs



Bins



There are lots of databases to choose from...

- Kyoto Encyclopedia of Genes and Genomes (KEGG)
<https://www.kegg.jp/>
- EggNOG <http://eggno5.embl.de/>
- UniRef90
- RefSeq, RefSeq viral
- UniProt
- Tigrfam
- Pfam
- Specialised – HydDB (hydrogenase), CAZy (Carbohydrate-Active enZymes), MEROPs (peptidases),



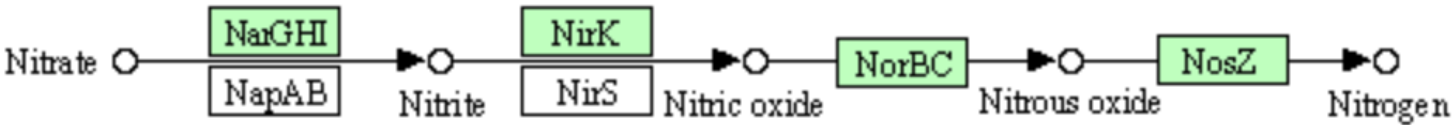
KEGG integrates functional annotation into biological pathway



Metabolism

- Global and overview maps
- [01100 Metabolic pathways \(404\)](#)
 - [01110 Biosynthesis of secondary metabolites \(170\)](#)
 - [01120 Microbial metabolism in diverse environments \(137\)](#)
 - [01200 Carbon metabolism \(54\)](#)
 - [01210 2-Oxocarboxylic acid metabolism \(22\)](#)
 - [01212 Fatty acid metabolism \(18\)](#)
 - [01230 Biosynthesis of amino acids \(66\)](#)
 - [01232 Nucleotide metabolism \(15\)](#)
 - [01250 Biosynthesis of nucleotide sugars \(19\)](#)
 - [01240 Biosynthesis of cofactors \(81\)](#)
 - [01220 Degradation of aromatic compounds \(31\)](#)
- Carbohydrate metabolism
- [00010 Glycolysis / Gluconeogenesis \(20\)](#)
 - [00020 Citrate cycle \(TCA cycle\) \(14\)](#)
 - [00030 Pentose phosphate pathway \(10\)](#)
 - [00040 Pentose and glucuronate interconversions \(2\)](#)
 - [00051 Fructose and mannose metabolism \(5\)](#)
 - [00052 Galactose metabolism \(1\)](#)
 - [00053 Ascorbate and aldarate metabolism \(3\)](#)
 - [00500 Starch and sucrose metabolism \(10\)](#)
 - [00520 Amino sugar and nucleotide sugar metabolism \(13\)](#)
 - [00620 Pyruvate metabolism \(25\)](#)
 - [00630 Glyoxylate and dicarboxylate metabolism \(20\)](#)
 - [00640 Propanoate metabolism \(11\)](#)
 - [00650 Butanoate metabolism \(14\)](#)
 - [00660 C5-Branched dibasic acid metabolism \(2\)](#)
 - [00562 Inositol phosphate metabolism \(3\)](#)
- Energy metabolism
- [00190 Oxidative phosphorylation \(32\)](#)
 - [00710 Carbon fixation in photosynthetic organisms \(8\)](#)
 - [00720 Carbon fixation pathways in prokaryotes \(13\)](#)
 - [00680 Methane metabolism \(13\)](#)
 - [00910 Nitrogen metabolism \(9\)](#)
 - [00920 Sulfur metabolism \(16\)](#)

Denitrification:



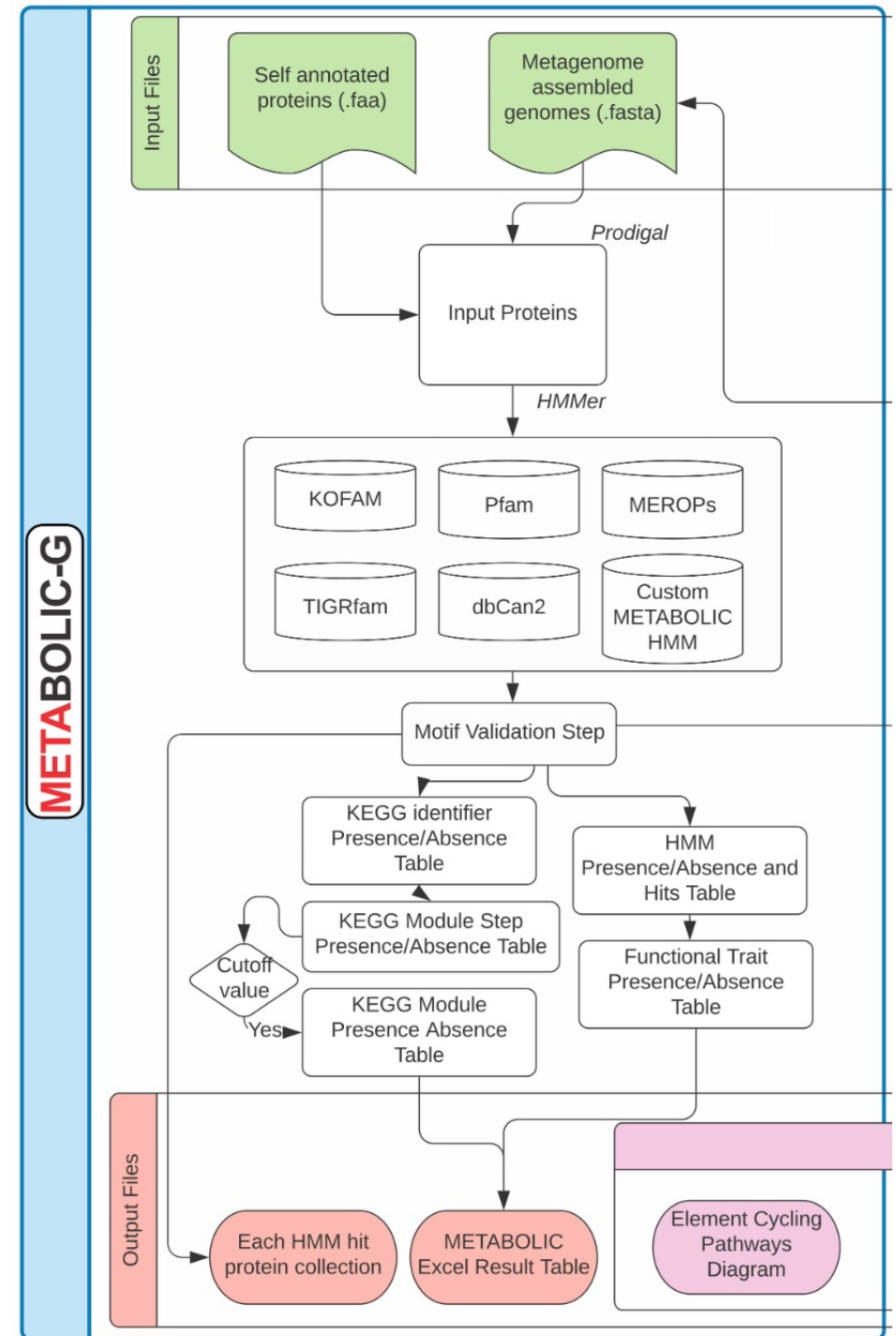
ORTHOLOGY: K00376

[Help](#)

Entry	K00376	KO
Symbol	nosZ	
Name	nitrous-oxide reductase [EC:1.7.2.4]	
Pathway	map00910 Nitrogen metabolism map01100 Metabolic pathways map01120 Microbial metabolism in diverse environments	
Module	M00529 Denitrification, nitrate => nitrogen	
Reaction	R02804 nitrogen:ferricytochrome-c oxidoreductase (N2O-forming)	
Brite	KEGG Orthology (KO) [BR:ko00001] 09100 Metabolism 09102 Energy metabolism 00910 Nitrogen metabolism K00376 nosZ; nitrous-oxide reductase Enzymes [BR:ko01000] 1. Oxidoreductases 1.7 Acting on other nitrogenous compounds as donors 1.7.2 With a cytochrome as acceptor 1.7.2.4 nitrous-oxide reductase K00376 nosZ; nitrous-oxide reductase	



<https://github.com/AnantharamanLab/METABOLIC>



Method: Pairwise

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

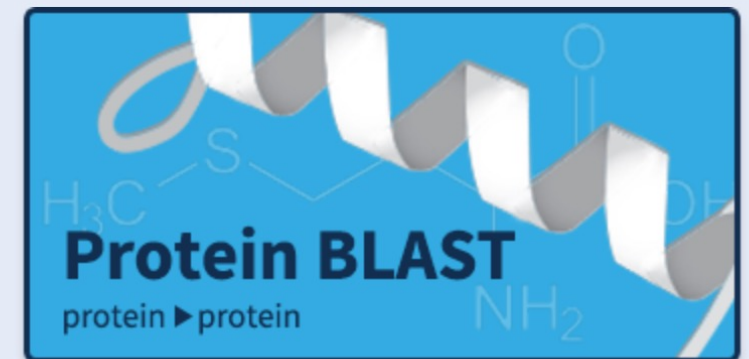
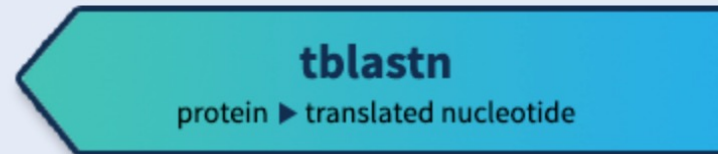
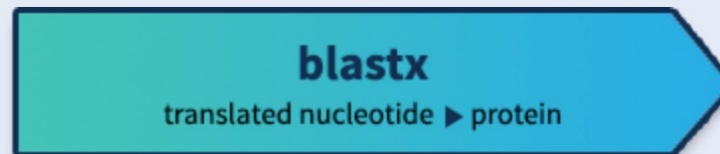
BLAST+ 2.15.0 is here!

We have included two exciting new features in the latest BLAST+ release

Tue, 28 Nov 2023

[More BLAST news...](#)

Web BLAST



Method: Pairwise

[Download](#) [Graphics](#)

P4201_109_bin90_k141_53671_1 1936,2029

Sequence ID: **Query_99571** Length: **94** Number of Matches: **1**

Range 1: 1 to 94 [Graphics](#)

[Next Match](#) [Previous Match](#)

Score		Expect	Identities	Gaps	Strand
174 bits(94)		5e-48	94/94(100%)	0/94(0%)	Plus/Plus
Query	1	AAATGTTGCGATGAAGCACCTGACTCCCGAAAAGAAGCACCTGACCCCAAAAATGATTC			60
Sbjct	1	AAATGTTGCGATGAAGCACCTGACTCCCGAAAAGAAGCACCTGACCCCAAAAATGATTC			60
Query	61	TCGATCAGAAAAATGTTGCGATGAAGCACCTGAC		94	
Sbjct	61	TCGATCAGAAAAATGTTGCGATGAAGCACCTGAC		94	

- Find related or similar sequences by comparing two sequences

Method: Pairwise

[Download](#) [GenBank](#) [Graphics](#)

Yponomeuta plumbellus genome assembly, chromosome: 10

Sequence ID: [OX371201.1](#) Length: 22918849 Number of Matches: 1

Range 1: 21903219 to 21903263 [GenBank](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score		Expect	Identities	Gaps	Strand	
58.4 bits(31)		1e-04	41/45(91%)	3/45(6%)	Plus/Plus	
Query	17	TGAGAAA-GG-C-AAATGAGAAAGGGACGAAATGAGAAAGGGACG				58
Sbjct	21903219	TGAGAAAGGGACGATATGAGAAAGGGACGAAATGAGAAAGGGACG				21903263

- Slow (which large volumes of data)
- Database errors / misannotated sequences – annotations are often given by the submitter
- Homology $\neq$ same function

Hidden Markov models (HMMs)

- Multiple sequence alignment – build a consensus and detect regions of similarity
- Use structural and mechanistic information (catalytic sites)
- Generate profiles using hidden Markov models (HMMs)
- More sensitive than pairwise
- Homology $\neq$ same function

	580	590	600	610
CP001820.1_1515 # 1721436 # 1	I	I	H	L
QTKX01000001.1_1880 # 183170	I	I	H	L
CP061474.1_2708 # 2708724 # 2	I	I	H	L
JADDIU010000002.1_90 # 77202	I	I	H	L
JAGGRA010000022.1_137 # 1273	I	I	H	L
FOJS01000005.1_105 # 89279 # 9	I	I	H	L
LT629731.1_1700 # 1828181 # 1	I	I	H	L
LT629731.1_1236 # 1375912 # 1	I	I	H	L
JABJUQ010000007.1_1912 # 187	I	I	H	L
RHNU01000019.1_40 # 40702 # 4	I	I	H	L
CP049703.1_1558 # 1596680 # 1	I	I	H	L
PVNE01000040.1_3 # 5649 # 746	I	I	H	L
RHNT01000030.1_6 # 5469 # 776	I	I	H	L
BJDN01000035.1_5 # 3283 # 532	I	I	H	L
RKLX01000004.1_37 # 44762 # 4	I	I	H	L
SLUL01000001.1_44 # 43903 # 4	I	I	H	L
CP012034.1_1205 # 1155810 # 1	I	I	H	L
CP012602.1_2085 # 2263430 # 2	I	I	H	L
CM002692.1_558 # 571722 # 574	I	I	H	L
JQMR01000001.1_991 # 1063335	I	I	H	L
PVNE01000006.1_92 # 97439 # 9	I	I	H	L
RSFW01000015.1_107 # 96368 #	I	I	H	L
LXMA01000038.1_163 # 131708	I	I	H	L
BDGB01000034.1_19 # 16914 # 1	I	I	H	L
FXTI01000004.1_86 # 80836 # 83	I	I	H	L
URPV01000016.1_22 # 22364 # 2	I	I	H	L
FOOK01000048.1_2 # 5254 # 754	I	I	H	L
FPBV01000001.1_76 # 86853 # 89	I	I	H	L
AUMM01000010.1_29 # 36455 #	I	I	H	L
BAMN01000039.1_2 # 1004 # 330	I	I	H	L
CAAFUD010000122.1_1 # 254 # 2	I	I	H	L
CAJLUB010000061.1_1 # 230 # 24	I	I	H	L
PGVE01000053.1_62 # 55921 # 5	I	I	H	L

Hidden Markov models (HMMs)

	353	357	361	805	548	552	599	600	621	
<i>Geobacillus stearothermophilus</i>	L A H Y Y T	E P D S F F G	V P D T I	I I H L W V	E G I	I G H H Y Y	A L E V I			NorZ
<i>Neisseria meningitis</i>	T A H Y T V	E G Q G F Y G	A A D L I	V V H L W V	E G F	T L H H L Y	A L E V V			
<i>Cupriavidus necator</i>	T A H Y T V	E G Q K F Y G	F G D V V	V V H L W V	E G F	T F H H L Y	A L E V V			
<i>Sphingomonas wittichii</i>	T A H Y T V	E G Q S F Y G	F G D V V	V V H L W V	E G I	T F H H L Y	A L E V V			
<i>Gammaproteobacterium HdN1</i>	T A H Y T V	E G Q T F Y G	F G D V V	I V H L W V	E G F	T F H H L Y	A L E V I			
<i>Candidatus Methylomirabilis oxyfera</i>	V A H Y R A	E P G K F Y G	P G D A V	I V H L W V	E G F	T G H H W Y	A M E V V			
<i>Chlorobi bacterium OLB7</i>	T V H D F V G F V N F F G	V G G A L	V I H M W A	E A F	I S H N F Y	T L Q V I			Nor-related	
<i>Bacteroidetes bacterium OLB12</i>	T V H D F V G F V N F F G	I G G S I	V I H M W A	E A F	I S H N F Y	T L Q V V				
<i>Flaviumibacter sp. ZG627</i>	T V H D F V G F V N F F G	I G G S L	V V H M W A	E A F	I S H N F Y	T L Q V I				
<i>Ansinibacterium sp. OR43</i>	T V H D F V G F V H F F G	A G G A V	V I H M W A	E A F	I S H N F Y	T L Q V I				
<i>Algoriphagus mannitolivorans</i>	T V H D F V G F T S F F G	V G G S L	V V H M W V	E A F	I S H N F Y	T L Q V V				
<i>Mariniradius saccharolyticus</i>	T V H D F V G F T N F F G	V G G S I	V V H M W V	E A F	I S H N F Y	T L Q V V				
<i>Indibacter alkaliphilus</i>	T V H D F V G F T K F F G	V G G A V	V V H M W V	E A F	I S H N F Y	T L Q V V				
<i>Cecembia lonarensis</i>	T V H D F V G F T K F F G	V G G A V	V V H M W V	E A F	I S H N F Y	T L Q V V				
<i>Muricauda ruestringensis</i>	T I N E F V D Y L G F F G	I G A C L	V V H M W V	E A F	I S H N F Y	T L Q F V				
<i>Arenibacter akgicola</i>	T I N E F I D Y L G Y F G	I G A C V	V V H M W V	E A F	I S H N F Y	T L Q F V				
<i>Candidatus Methylomirabilis oxyfera</i>	G A E D F V G G G P G E A	I G G V V	N I H M W V	E V T	I S H N F Y	T M Q V L			Nod-like	
<i>Candidatus Methylomirabilis sp.</i>	G A E D F V G G G P G E S	I G G A V	N I H M W V	E V T	I S H N F Y	T M Q V L				
<i>Polyangiaceae bacterium</i>	S A E D F I G G G P G E A	I G G M V	T V H M W V	E V T	I S H N F Y	T M Q V L				
<i>Acidobacteria bacterium</i>	S A E D F V G G G P A R G	I G G T V	T V H M W V	E V T	I S H N F Y	T L Q V L				
<i>Acidobacteria bacterium</i>	S A E D F V V A D G S T A	I G G V V	V V H M W V	E V T	I S H N F Y	T L Q V V				
<i>Magnetovibrio blakemorei</i>	G A T D F V - - - R P F G	L G G A V	V V H M W V	E V T	I A H N F Y	T L Q V L				
<i>Gammaproteobacterium HdN1</i>	A A W D F V - - - K P F G	I G I A V	V V H M W V	E V T	I S H N F Y	T L Q V L			Foram Nod-related	
<i>Globobulimina sp.</i>	C S V D V V - - - R P G G	V G G H I	V V H M W V	E V T	V G H N F Y	T T Q V M				
<i>Globobulimina sp.</i>	C A I D F V - - - R P L G	F G G H I	V V H M W V	E V T	V G H N F Y	T T Q V M				
<i>Nonionella stella Nor_I</i>	C A I D F V T I G G G M D	I G G H L	V V H M W V	E V T	V G H N F Y	T T Q V I				
<i>Nonionella stella Nor_II</i>	C A I D F V T L G G G M D	I G G H L	V V H M W V	E V T	V G H N F Y	T T Q V I				
<i>Nonionella stella Nor_III</i>	C A I D F V T L G G G M D	I G G H L	V V H M W V	E V T	V G H N F Y	T T Q V I				
<i>Nonionella stella/Bolivina argentea Nor_IV</i>	C A I D F V - Q V N G K I	L G G H L	V V H M W V	E V T	V G H N F Y	T T Q V I				
No species assignment Nor_V	C A S D F V - - - R P F G	A G G L V	V V H M W V	E V T	V A H N F Y	T W Q V L				
	Quinol binding site			Catalytic site						

“Homology creep”



MICROBIAL
BIOTECHNOLOGY

Applied
Microbiology
International

Research article | Open Access |

Function-driven single-cell genomics

Tanja Woyke , Jessica Jarett

First published: 28 January 2015 | <https://doi.org/10.1111/1751-7915.12247> | Citations: 12

“While all these sequence-based methods have enabled tremendous progress in microbial ecology, they have also inadvertently given rise to an era of homology creep; one of today's key challenges is the slow pace of functional validation, which continues to lag further and further behind the rate of sequencing.”

- Case study 1: The genes I'm interested in aren't in a database
- Case study 2: I'm sure the microbes are doing it, but I can't find the genes

Geomicrobiology (Hubert Lab) Research



UNIVERSITY OF
CALGARY

Microbiology for [Oil Spill](#) Preparedness in Canada's Arctic Marine Environment

Marine Microbial Biodiversity and [Hydrocarbon Seep Detection](#)

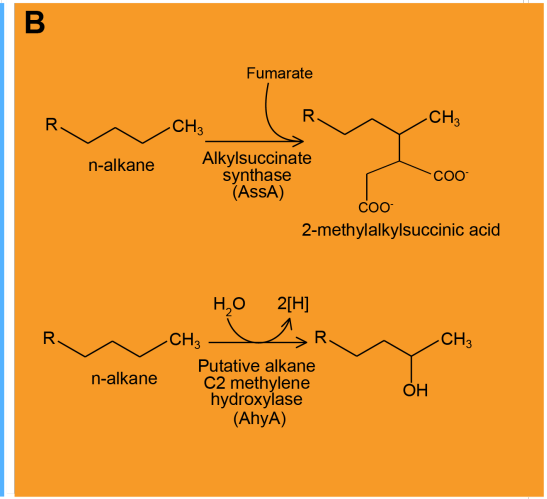
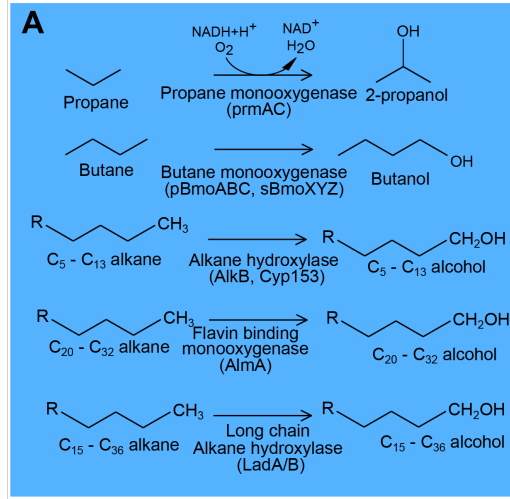
Challenges and Opportunities in [Petroleum Microbiology](#): the reservoir microbiome



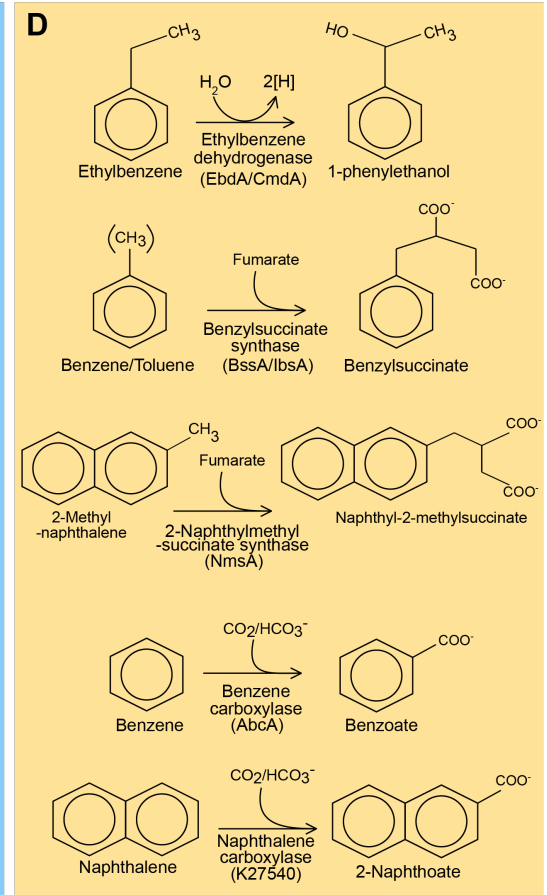
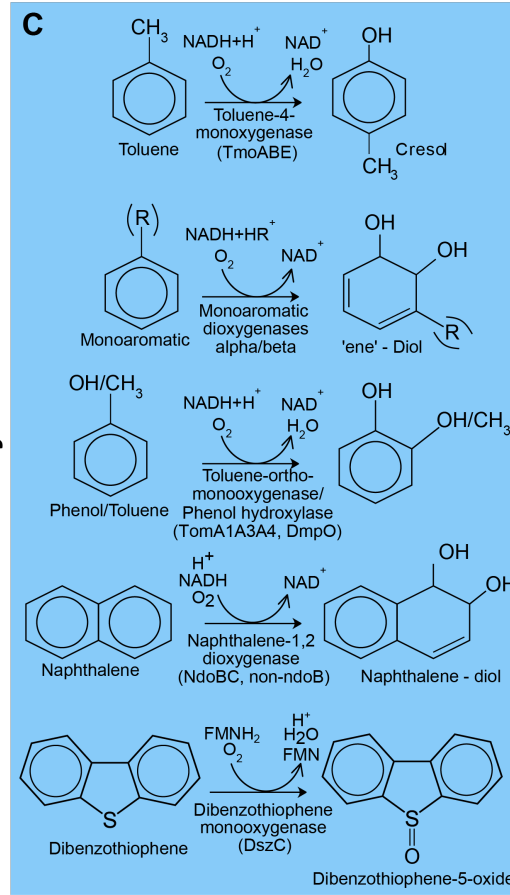
The genes I'm
interested in aren't
in a database:

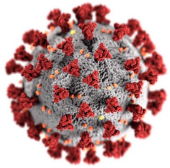
hydrocarbon
degradation
pathways

Alkanes



Aromatic Hydrocarbons





METHODS article

Front. Microbiol., 07 January 2022

Sec. Microbial Physiology and Metabolism

Volume 12 - 2021 | <https://doi.org/10.3389/fmicb.2021.764058>

CANT-HYD: A Curated Database of Phylogeny-Derived Hidden Markov Models for Annotation of Marker Genes Involved in Hydrocarbon Degradation



Varada Khot^{1†}



Jackie Zorz^{1†}



Daniel A. Gittins²



Anirban Chakraborty^{2‡}



Emma Bell²



María A. Bautista²



Alexandre J. Paquette¹



Alyse K. Hawley^{1‡}



Breda Novotnik¹



Casey R. J. Hubert²



Marc Strous¹



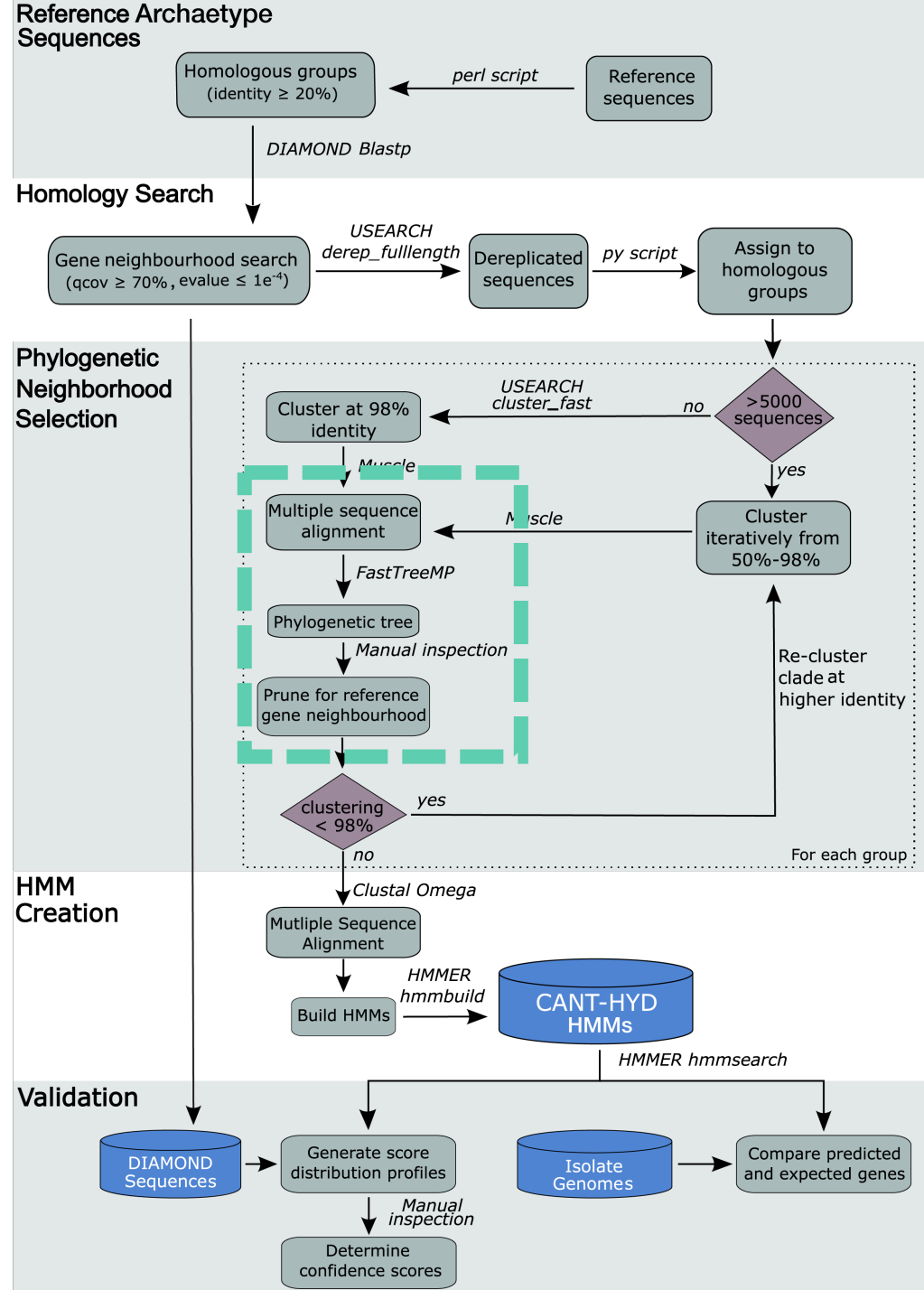
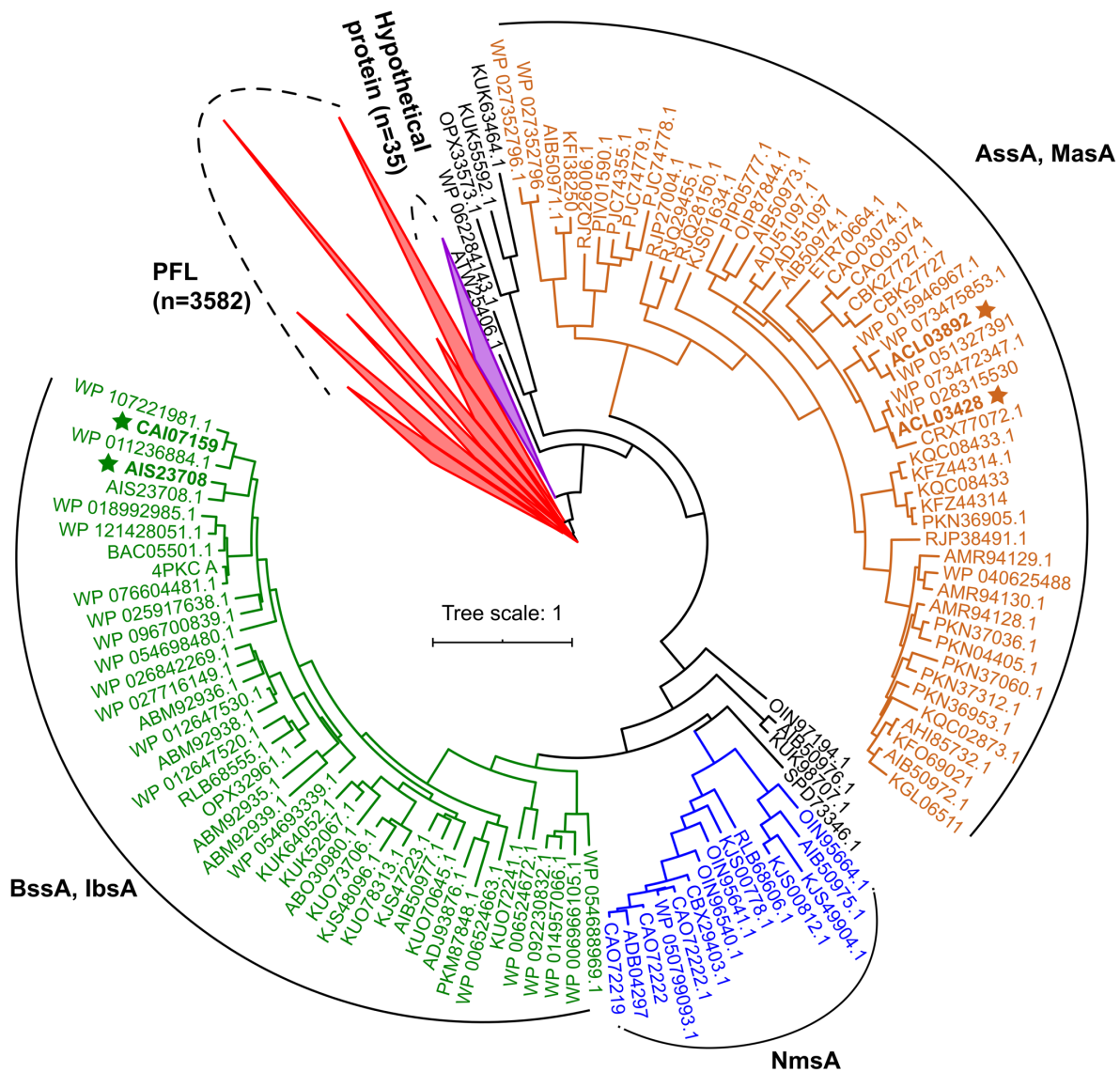
Srijak Bhatnagar^{2*‡}

¹ Energy Bioengineering and Geomicrobiology Group, Department of Geoscience, University of Calgary, Calgary, AB, Canada

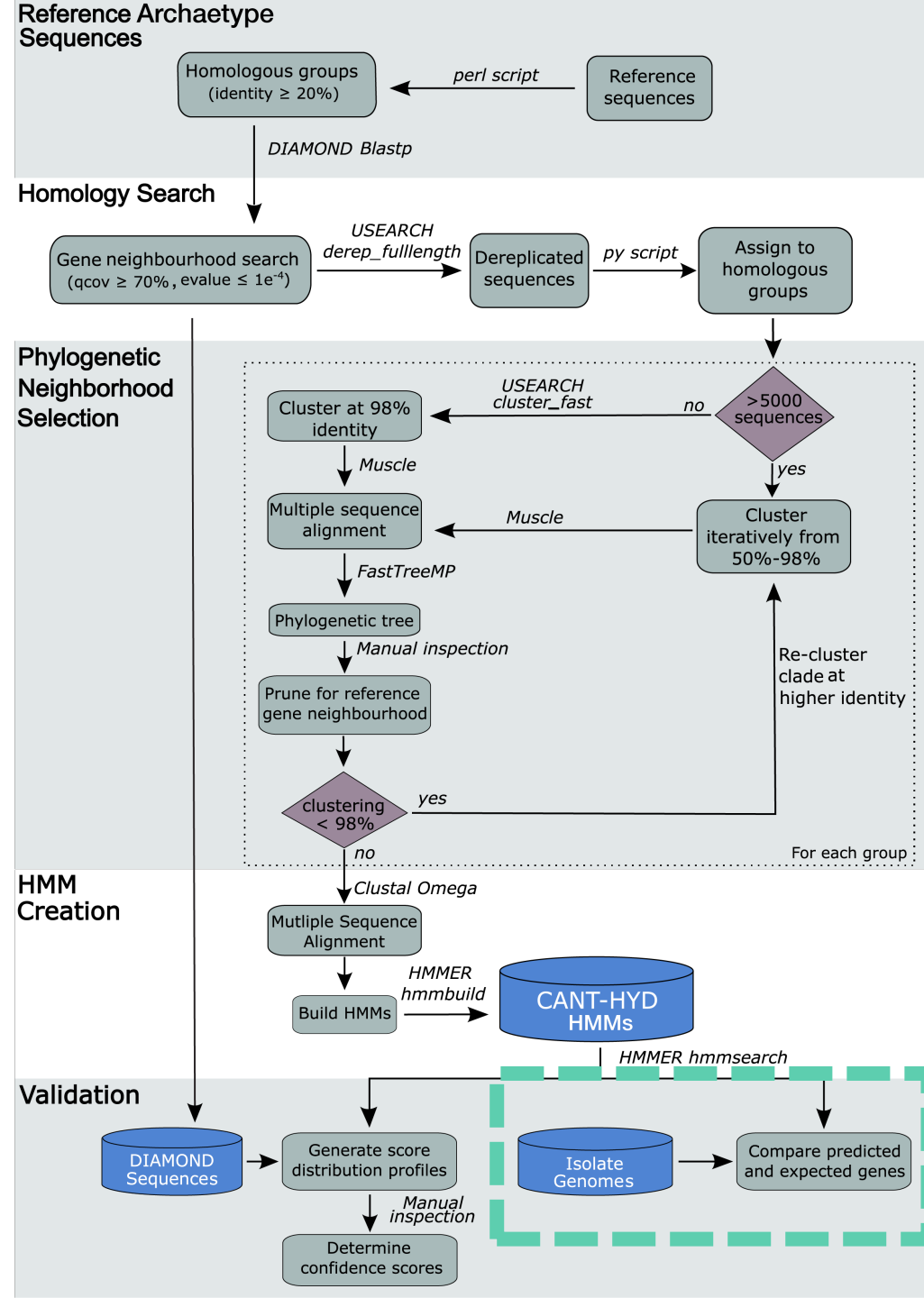
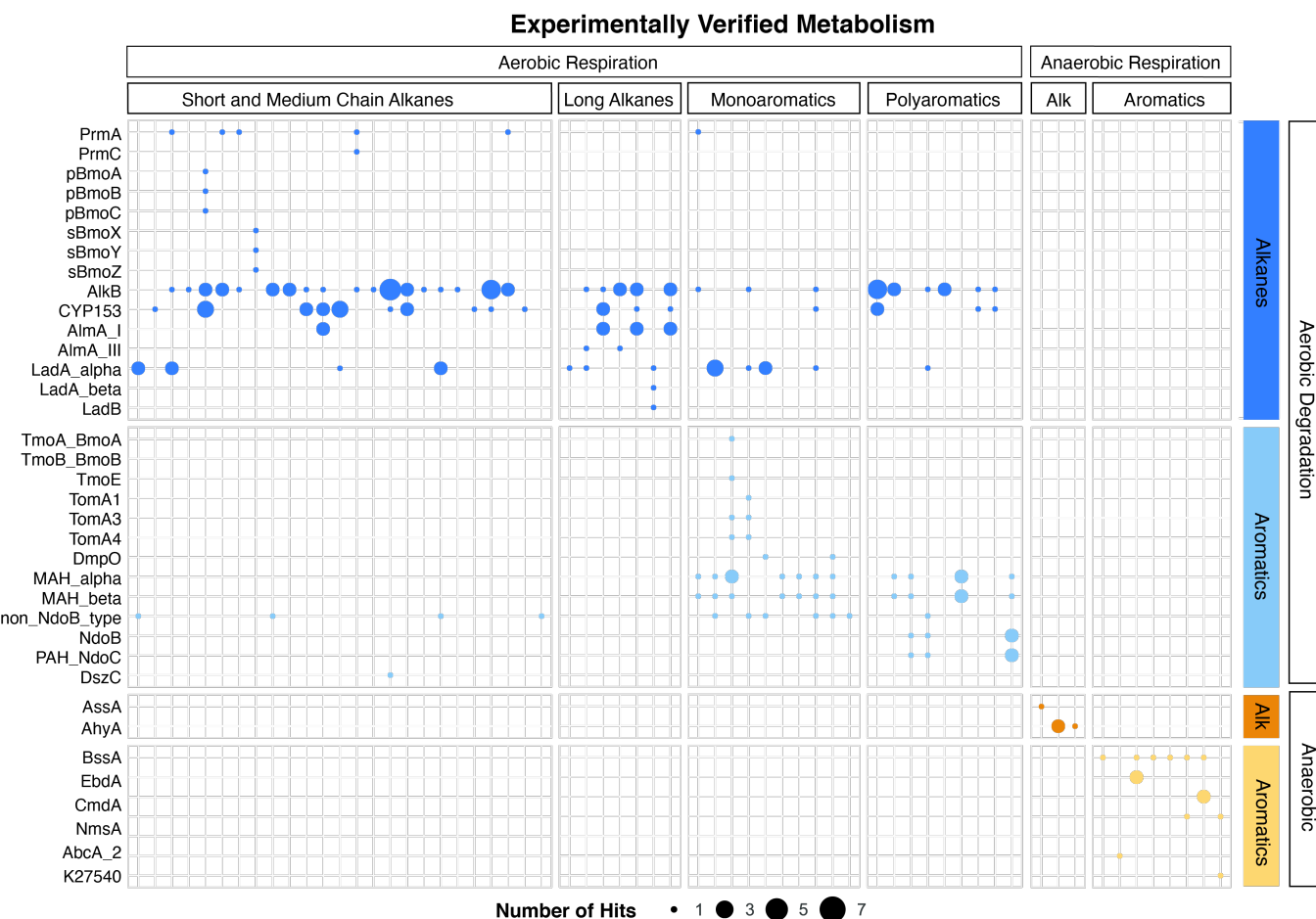
² Energy Bioengineering and Geomicrobiology Group, Department of Biological Sciences, University of Calgary, Calgary, AB, Canada

HOW IT ENDED

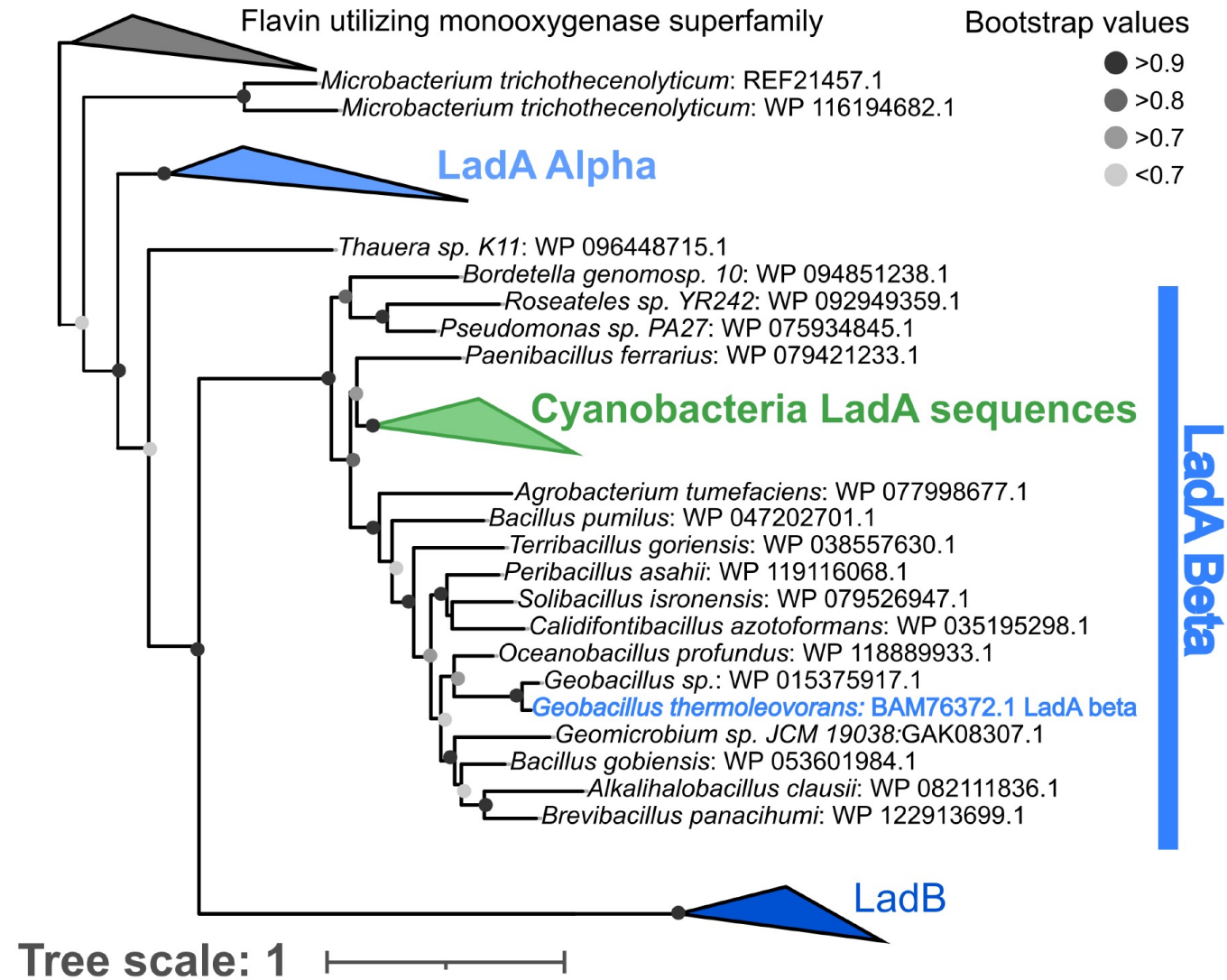
- Multiple sequence alignment and phylogenetic tree



- Multiple sequence alignment and phylogenetic tree (37 HMMs)



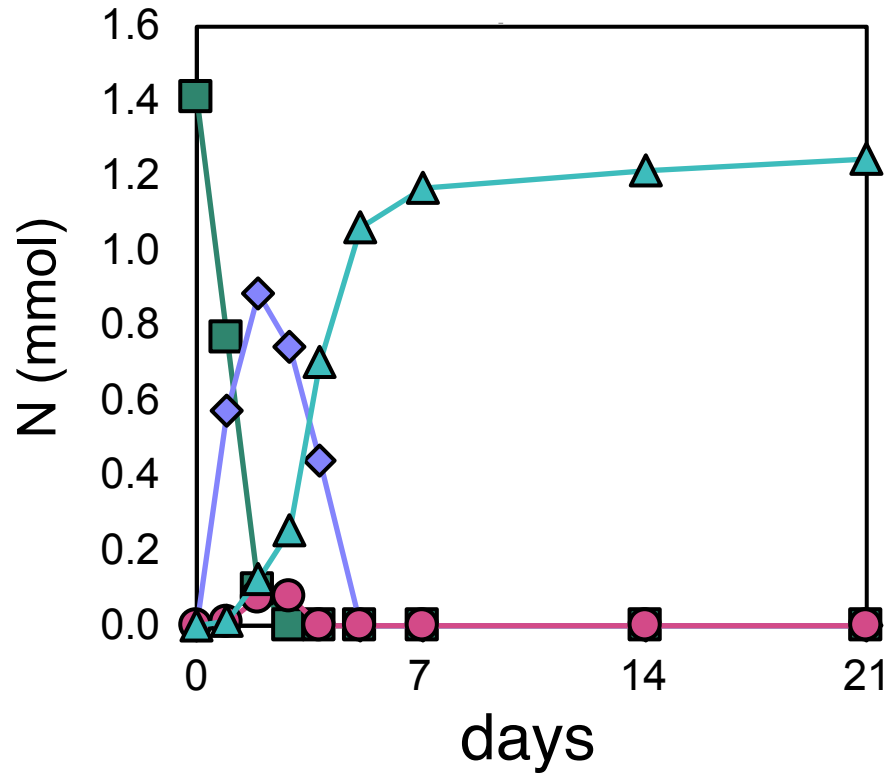
A



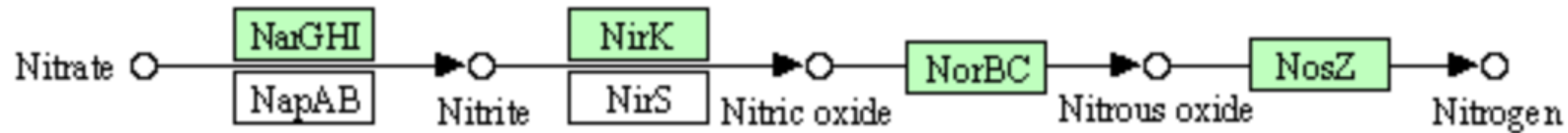
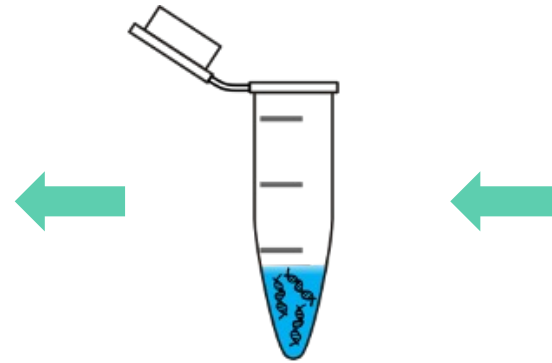
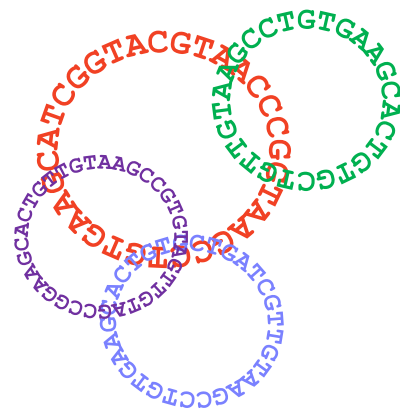
- LadA in cyanobacterial genomes show potential for aerobic hydrocarbon degradation in global ocean surface waters

- Case study 1: The genes I'm interested in aren't in a database
- Case study 2: I'm sure the microbes are doing it, but I can't find the genes

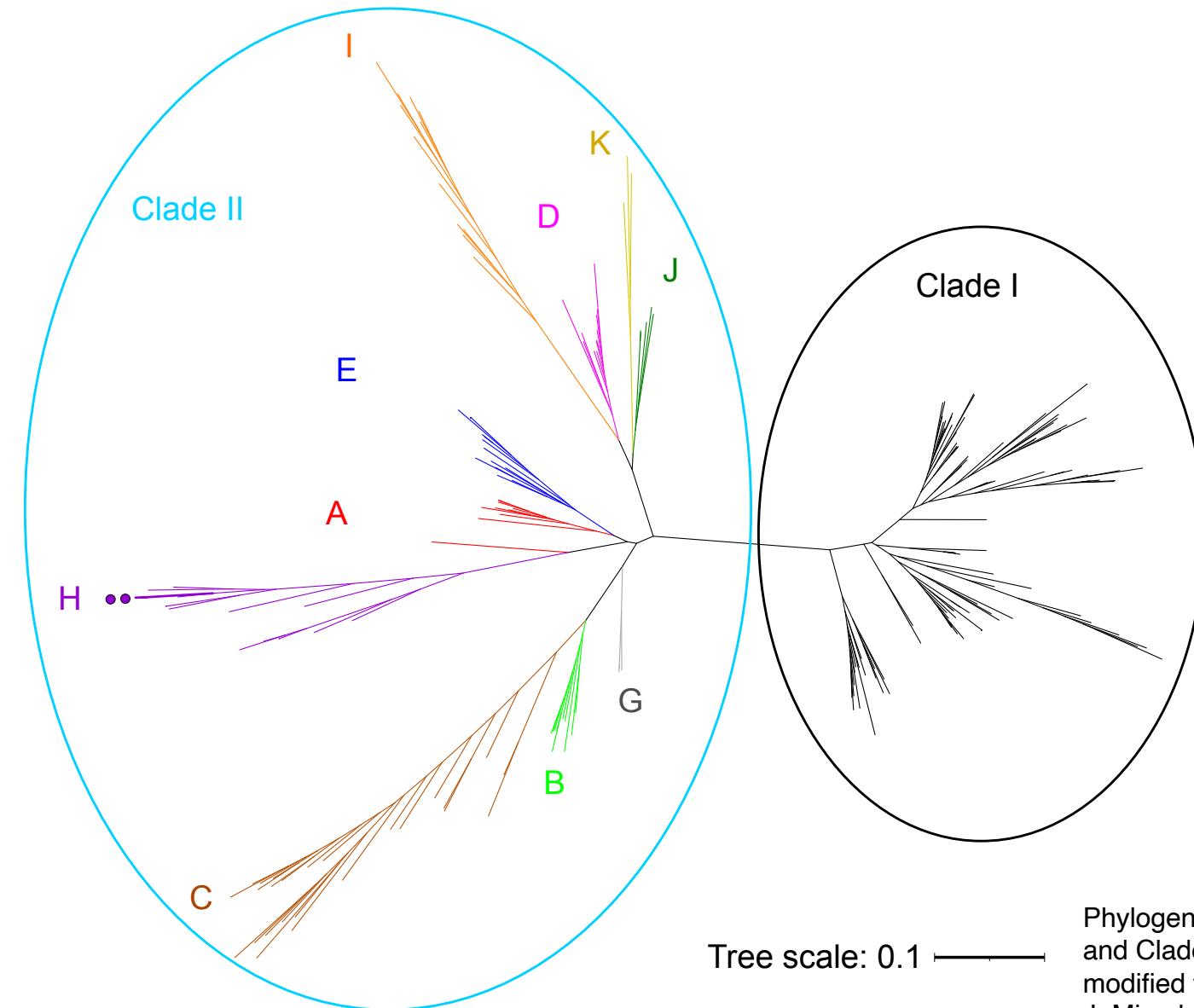
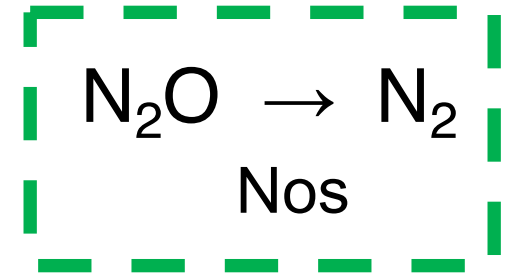
I'm sure my microbes are doing it...



I'm sure my microbes are doing it...



N₂OR microorganisms are taxonomically diverse

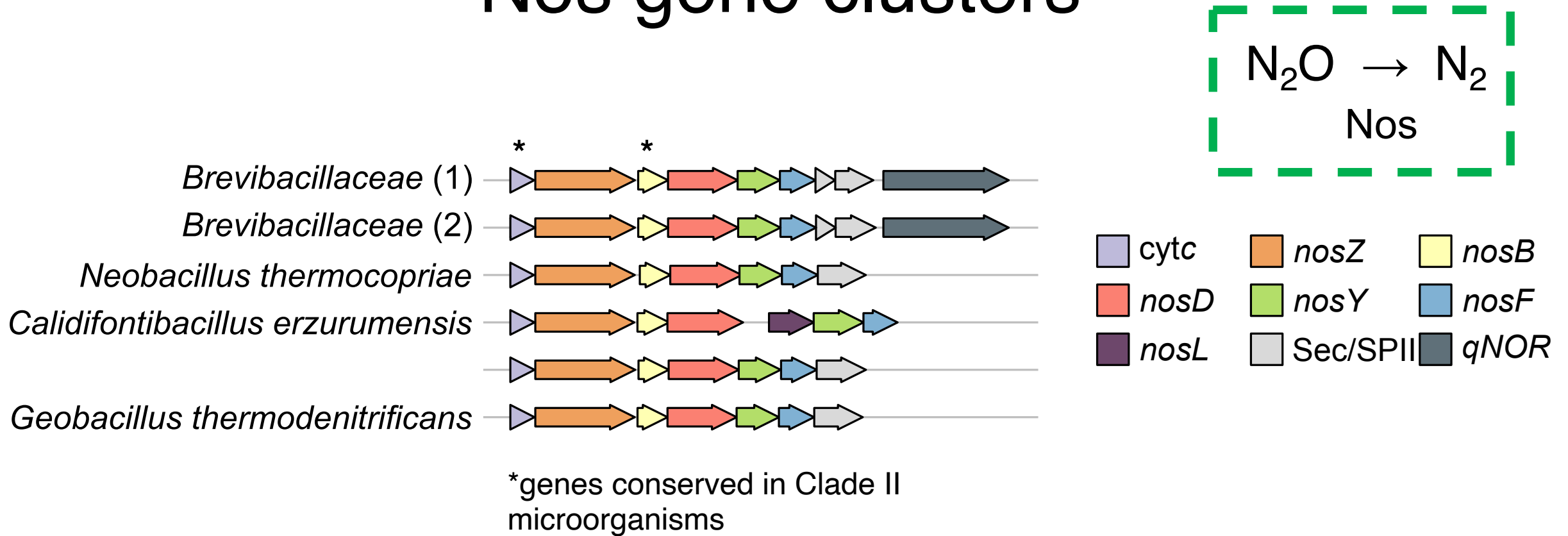


- **Clade I:** Well-studied Proteobacteria
- **Clade II:** Phylogenetically diverse
Twelve bacterial and
archaeal phyla (to date)
- **Clade II H:** *Firmicutes*

Tree scale: 0.1 ———

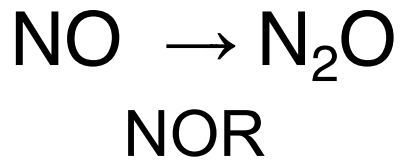
Phylogenetic tree of Clade I (typical)
and Clade II (atypical) NosZ:
modified from Chee-Sanford *et al.*,
J. Microbiol. Methods (2020) 172

Clade II microorganisms have distinct Nos gene clusters

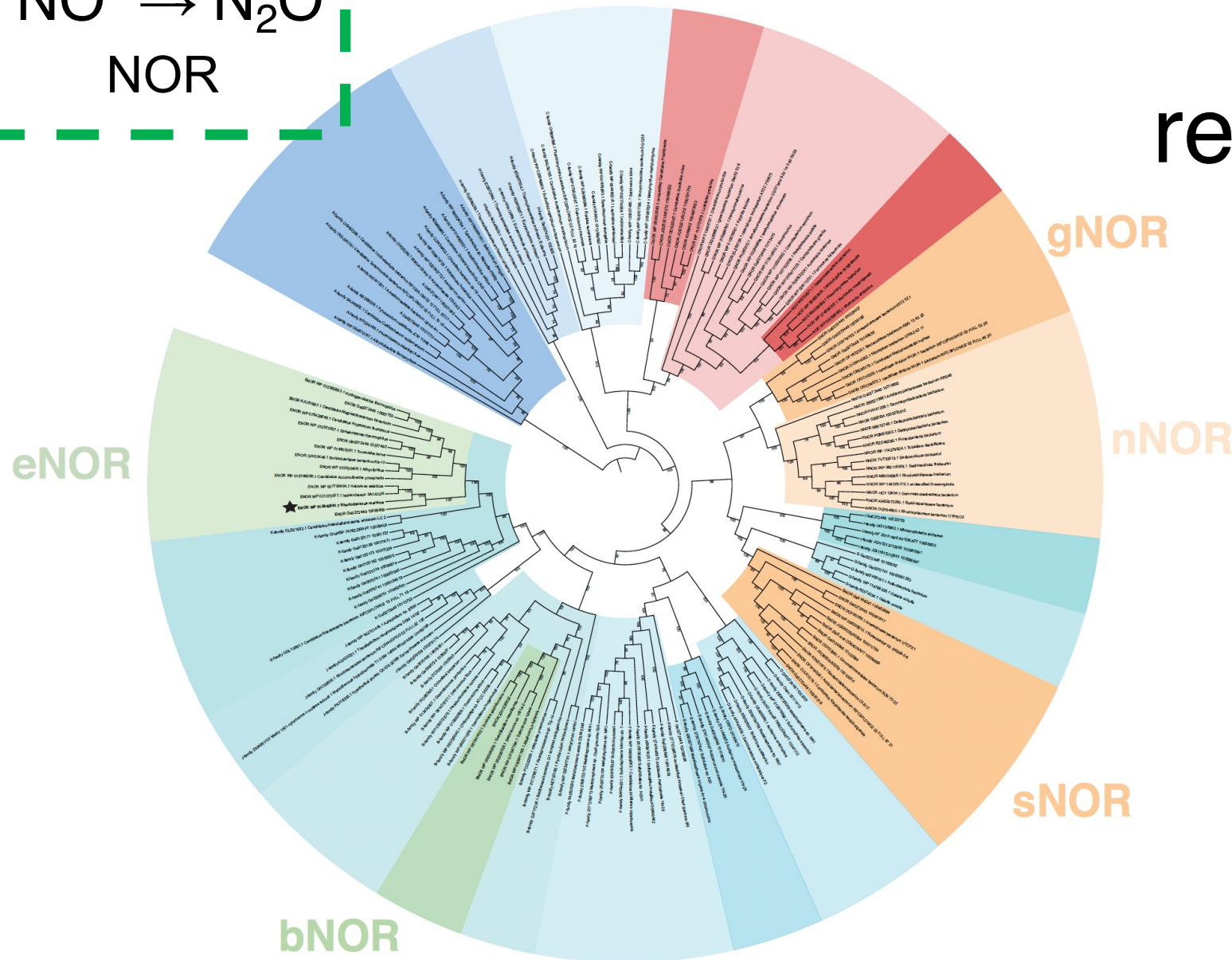


Cellular localisation: PSORTb

Transmembrane helices: DeepTMHMM



Nitric oxide reductases evolved from oxygen reductases



cNOR = in database (mostly Proteobacteria)
qNOR = annotated as cNOR
bNOR = experimentally verified (not in database)
eNOR = newly experimentally verified in this paper
sNOR, nNOR, gNOR = proposed

heme-copper oxidoreductase (HCO) superfamily

Murali et al., (2021)
 doi: <https://doi.org/10.1101/2021.10.15.464467>

References sequences

		H255 E259	Y270	Y276	Y284	H307	T338 T341 T345
nNOR							
<i>Thermomicrobium</i> sp002898255	GBD20489.1	FWLFAHNLMEAMGIMTLGAIYAIVPRYTR-SGQLY-S	PRAAVVAMILYTMAAIPAFG	HHLYTWVTGNPEVLQNVSR-STSWATGFIAATLTA	FNVLTV-----	WRNGL	
<i>Sedimenticola selenatireducens</i>	WP_084609916.1	FYIFAHNLMEAMAIMVISAVYATLPLYLADGTRKLYS	DKLANLALWILLVTSVTSFF	HHFYTTNPLGPSALAY-HGNFMSWATGVGAAL-STFT	ILATI-----	WKHGI	
<i>Sedimenticola selenatireducens</i>	PLX61751.1	FYIFAHNLMEAMAIMVI SAVYATLPLYLADGTRKLYS	DKLANLALWILLVTSVTSFF	HHFYTTNPLGPSALAY-HGNFMSWATGVGAAL-STFT	ILATI-----	WKHGI	
<i>Rhodocyclaceae bacterium</i> UTPRO2	OQY64980.1	FFFFAHNLMEAMAIMVASAIYATLPLYLADGSRKLF	SKMANLALWILLVTSITSGL	HHFITFPNQPAALSYWGS-IMSWG	TGIGAAL-SIFTVFATI-----	WQHGL	
gNOR							
<i>Sulfurimonas autotrophica</i>	WP_013326534.1	WWGLDLVADGLVLIYVAGSWYLLATLIT--GQKLF-MENVARAALMLELLVSWMVWS	HHLLA-DQGQPEMMKLISGEMVTA	FELLTQGI-ALFITL	TVTL-----	WKARP	
<i>Ignavibacteria bacterium</i> GWA2_36_19	OGU38604.1	WWGLDLIADGLVLIFVAGTWYLLATLIT--GKKLF-MENWARAALFVEMVSVTWVWS	HHLMS-DQAQPGILKILSGEMVTA	FELITQGL-AFFITL	ATL-----	WSARP	
<i>Ignavibacteria bacterium</i> CG2_30_36_16	OIP63421.1	WWGLDLIADGLVLIFVAGTWYLLATLIT--GKKLF-MENWARAALFVEMVSVTWVWS	HHLMS-DQAQPGILKILSGEMVTA	FELITQGL-AFFITL	ATL-----	WSARP	
candidate division WOR-1 bacterium	OGW14262.1	WWGLDLIADGLVLIYVAGTWYLLATLIS--GKELY-MRNVARAALLVELIVSWNVWA	HHLLS-DQAQPNIMKIIISGEMVTA	FELVTMGI-AIFITL	KTLL-----	WEARP	
candidate division WOR-1 bacterium	OGC04573.1	WWGLDLIADGLVLIYVAGTWYLLAMIIT--GRQIF-MQNFARAALFVELVSVWFVWS	HHLLS-DQTQPVMMRIFSGEMITAFELVTSGI-AVFLT	LATL-----		WQARP	
sNOR							
<i>Bacteroidetes bacterium</i> 37-13	OJV27025.1	TFFFGHTIANEALYLGLATLYELLPEVS--GRP	KFKTTWYVALGWNC	AIIFILGAFF	HHLYM-DFVQPKGFQIFGQ-IASYFATIPSVVV	TIISIVTLL-----	YNNKI
<i>Geobacillus</i> sp.	WP_023634191.1	IYAFGHIFANSVIYMGVIAVYELPKYT--NRP-WKSYKIFLIAWNMSTLFTII	IYP	HHLLM-DFVMPK-WMLIIGQVFSYLNGLPVLV	TAFGALMIV-----	YRSGI	
<i>Rhodanobacter denitrificans</i>	WP_015448124.1	IYWFGHVMVINATIYMGVIAVYELLPRYT--GRP-YGISRPFLWSWA	ASTVFV	IIVFP	HHLLM-DYAEPR-WMLVMGQIIISYAAGFPVFLV	TAYGVLTNI-----	HRSGL
<i>Nitrosococcus halophilus</i>	WP_013031504.1	IYFFGHVFINATIYASVTAVYELLPRYT--GRP-WKTSKVFYAAWLAIVF	VMMAVYP	HHLLM-DFAMPP-WALIVGQVL	SYGSGVPVMVVTGYGALMIV-----	YRSGI	
<i>Thioalkalivibrio</i> sp. ALRh	WP_019592254.1	TYFFGHVFINATIYMAVIGVYELPKYT--GRP-WKVS	RVFLAAWA	ASTVMVLLVYP	HHLLM-DFSQPTSLHVLGQ-VISYTSGLPVL	VTAWGALTNI-----	YRSGI
bNOR							
<i>Salinicoccus qingdaonensis</i>	WP_092985759.1	FWAFGHTLVN	VWYLVAVSAWYLVVPKVI--GGKLF-SDSLARAVVILIVVLNVP	GGFHHQIV-DPGFTEGLKFMHL-FMSLAI	AVPSLL-TAFALFATLERTGRRRG-G-KGLLGWFWKL		
<i>Salinicoccus</i> sp. YB14-2	WP_092985759.1	FWAFGHTLVN	VWYLVAVSAWYLVVPKVI--GGKLF-SDSLARAVVILIVVLNVP	GGFHHQIV-DPGFTEGLKFMHL-FMSLAI	AVPSLL-TAFALFATLERTGRRRG-G-KGLLGWFWKL		
<i>Jeotgalicoccus psychrophilus</i>	WP_026860023.1	FWSFGHTLVN	VWYLVAVSAWYIVLPKVI--GGKIF-SDSLARLVVILIVLNVP	GGFHHQIV-DPGFTEGLKFMHL-FMSLAI	GFPSLM-TAFALFATLERAGRNGK-G-KGLFGWFFKL		
<i>Virgibacillus dakarensis</i>	WP_088049698.1	FWSFGHTLVN	IWYIVATSAWYVVVPKII--GGRVF-SDKLARLVVLLVILNIP	GGFHHQII-DPGISESVKFLHV-FMSISIA	FPSLM-TAFAMFAVFERAGRKLK-G-KGLLGWFFKL		
<i>Sporosarcina</i> sp. HYO08	KXH87068.1	FWSFGHTLVN	IWYLTAVSAWYTIVPKII--GGRRF-SDTLTRVVVILLVITNIP	GGFHHQIV-DPGMGEALKYMHV-FMSLAI	GFPSLM-TAYAMFSVFERTGRLKG-G-KGLLGWFFKL		
<i>Sporosarcina ureilytica</i>	WP_075529285.1	FWSFGHTLVN	IWYLTAVSAWYVIVPKII--GGKRF-SDTLTRVVVAMLVITNIP	GGFHHQIV-DPAMGPALKYMHV-FMSLSIA	FPSLM-TAFAMFYVFERTGRAKG-G-KGLFGWLKKL		
<i>Bacillus</i> sp. FJAT-27445	WP_059173536.1	FWAFGHTLVN	IWYLTAVSAWYVVIIPKLM--GGRRW-SDTLTRVVI	IALLVVMNITGGFHHQIV-DPGITESVKFMHV-FMSLAI	GFPSLM-TAYALFAVFEKTARRNG-G-KGILGWYKKM		
<i>Bacillus</i> sp. EB01	WP_043932172.1	FWAFGHTLVN	IWYLTATSAWYVVIPKLM--GGKRW-SDTLTRVVI	ISLVVMNITGGFHHQIV-DPGISEAVKYMHV-FMSLAI	GFPSLM-TAYALFATFEKTARRNG-G-KGLVGWYKKM		
<i>Thalassobacillus cyri</i>	WP_093046123.1	FWAFGHTLVN	IWYMTAISAWYVFVPKII--DGRRW-SDMLTRIVVIALVIMNITGGFHHQII-DPGISESVKYMHV-FMSLAI	GFPSLM-TAYAMFAVFEKTARKKG-G-RGLVGWYKKM			
<i>Neobacillus bataviensis</i>	WP_007083088.1	FWAFGHTLVN	IWYLTAVSAWYVIVPKII--SGKRW-SDTLTRVVI	IALLVVMNITGGFHHQIV-DPGISPSVKFMHV-FMSLAI	GFPSLM-TAWAMFAVFERTARKQG-G-KGLIGWYKKM		
eNOR							
<i>Rhodothermus marinus</i>	WP_012842681.1	YWIIGHSSQQINLAAMITVWYFLTHVVG--GA	EVV-SEKLSRTAFILYLF	FINMGAAHLLA-DPGVHMSWKHWNTSYAVYGA	VLASMI-HAFAIPAGLEAGRRKRGLGSQGLFGWLWSA		
<i>Alicyclophilus denitrificans</i>	WP_013520406.1	WWAFGHSSQQINVAAHISIWYAVAAIAF--GAKPM-SERVS	RGAFLLYILFLQLASAHHLLA-DPGVSTEWKIVNTSYFYFAVLGSMI-HALSIPGAMEVAQRAKYN-KGLFEWL	LRKA			
<i>Candidatus Kryptonium thompsoni</i>	WP_075426648.1	WFGFGHPAQQINLAAMSVWYLLSYFTV--GGVTP-SEKVS	RTAFLLYVLFINAGSAHLLLV-DPGFGAPWKIFNTSYVIYMAVLASLI-HCFAIPASAEIGQRKRGN-KGLFTW	LIIKA			
<i>Bacteroidetes bacterium</i> OLB11	KXK43645.1	WWGLGHSSQQINVAAMVSIWYVMSFLAV--GGTSI-NEKVS	RSFAVLYILFICLASAHHLLT-DPGVSSAWKVWNTSYAMYLA	VLASMI-HAFAVPSSSFESAQRWGYN-KGLFEWLSKA			
<i>Gemmatimonadetes bacterium</i> SCN 70-22	ODT01569.1	WWGLGHSSQQINVAAMVAIWYMLAALTV--GGVVL-NEKVS	RSFAVLYILFISMASAHHLLV-DPGFGPAWKIVNTSYFYMAVLASMI-HGFTVPAGMELGMRLRGYT-DGLFGWL	LRRA			

Black: conserved proton channels
Red: active site

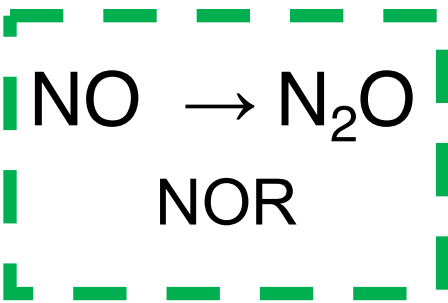
Murali et al., (2021)
doi: <https://doi.org/10.1101/2021.10.15.464467>

- Multiple sequence alignment and phylogenetic tree

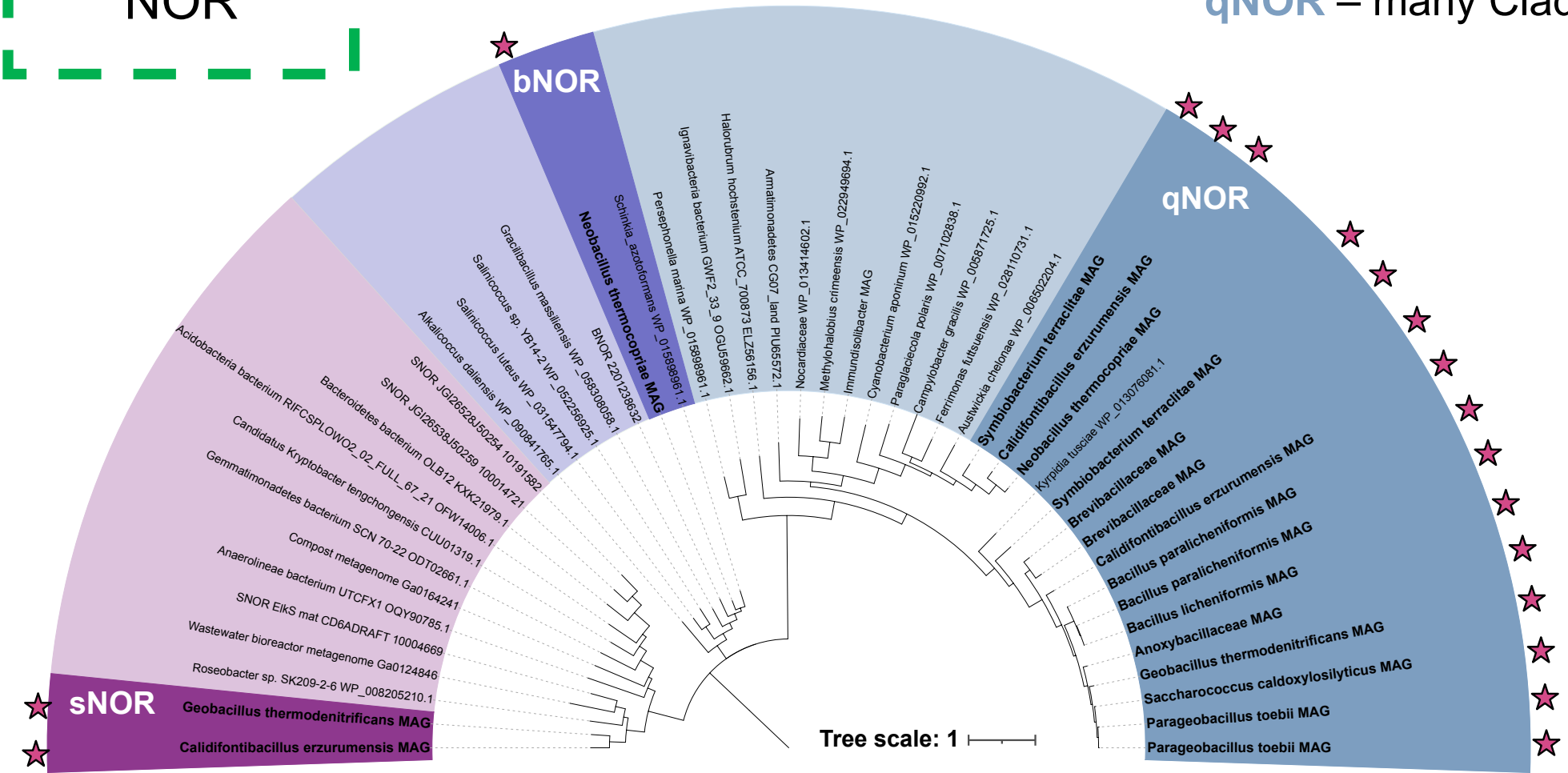
	580	590	600	610	620	630	640	650	660	670																																																																																										
CP001820.1_1515 # 1721436 # 1	I	V	I	H	L	W	V	E	G	F	F	E	V	F	S	T	T	L	M	A	F	I	F	V	T	L	G	L	V	S	R	A	G	T	V	A	A	I	S	S	G	A	I	Y	L	I	G	G	I	P	G	T	F	H	H	L	Y	F	T	G	V	T	S	T	I	V	A	T	G	A	S	F	S	A	L	E	V	V	P	L	V	L	L	G	Y	E	A	F	E	N	Y	T	R	L	H	S	A	P	W	M
QTKX01000001.1_1880 # 183170	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	M	V	Q	M	K	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	M	G	A	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	E	A	I	W	I	A	L	G	S	V	F	S	A	L	E	V	I	P	L	T	L	L	L	E	A	Y	E	Q	Y	K	M	M	R	D	G	G	V	D	
CP061474.1_2708 # 2708724 # 2	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	L	V	Q	L	G	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	E	A	I	W	I	A	L	G	S	V	F	S	A	L	E	V	I	P	L	T	L	L	L	E	A	Y	E	Q	Y	K	M	M	R	D	G	G	V	N		
JADDIU010000002.1_90 # 77202	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	L	V	Q	L	G	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	E	A	I	W	I	A	L	G	S	V	F	S	A	L	E	V	I	P	L	T	L	L	L	E	A	Y	E	Q	Y	K	M	M	R	D	G	G	V	I		
JAGGRA010000022.1_137 # 1273	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	L	V	Q	M	K	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	E	A	I	W	I	A	L	G	S	V	F	S	A	L	E	V	I	P	L	T	L	L	L	E	A	Y	E	Q	Y	K	M	M	R	D	G	G	V	N		
FQJS01000005.1_105 # 89279 # 9	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	L	V	Q	M	K	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	E	A	I	W	I	A	L	G	S	V	F	S	A	L	E	V	I	P	L	T	L	L	L	E	A	Y	E	Q	Y	K	M	M	R	D	G	G	V	N		
LT629731.1_1700 # 1828181 # 1	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	M	V	Q	M	K	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	E	A	I	W	I	A	L	G	S	V	F	S	A	L	E	V	V	P	L	T	L	L	L	E	A	Y	E	H	Y	K	T	M	K	A	G	G	I	D		
LT629731.1_1236 # 1375912 # 1	I	V	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	L	V	R	M	K	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	L	G	I	G	H	H	Y	Y	N	G	S	S	E	V	W	I	A	I	G	S	V	F	S	A	L	E	I	P	L	T	L	L	L	E	A	Y	S	Q	Y	K	M	L	R	D	G	G	I	D				
IABJUQ010000007.1_1912 # 187	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	L	V	Q	M	N	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	E	I	W	I	A	L	G	S	V	F	S	C	L	E	V	I	P	L	T	L	L	V	L	E	A	Y	E	Q	Y	K	M	M	K	D	G	G	V	D		
RHNU01000019.1_40 # 40702 # 4	I	V	H	L	W	V	E	G	I	F	E	S	F	A	V	I	L	I	G	Y	L	M	V	D	M	K	L	T	T	I	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	T	G	V	M	G	H	H	Y	F	W	E	G	D	H	S	I	W	L	A	L	G	S	C	F	S	A	L	E	V	I	P	L	C	L	L	V	W	E	A	Y	T	H	Y	R	V	K	D	S	S	M	E				
CP049703.1_1558 # 1596680 # 1	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	L	V	Q	L	R	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	A	E	I	W	I	A	L	G	A	V	F	S	A	L	E	V	I	P	L	T	L	L	L	E	A	Y	E	Q	Y	K	M	M	R	D	G	G	V	N		
PVNE01000040.1_3 # 5649 # 746	V	V	H	L	W	V	E	G	F	F	E	V	F	A	T	V	V	I	A	F	L	F	T	R	M	G	L	L	R	I	S	T	A	T	T	S	V	L	F	S	T	I	V	F	L	F	G	G	I	I	G	T	F	H	H	L	Y	F	S	G	T	P	I	G	V	L	A	F	G	A	T	F	S	A	L	E	V	V	P	L	V	L	I	G	F	E	A	Y	E	N	L	T	L	S	R	A	R	P	W	V
RHNT01000030.1_6 # 5469 # 776	I	V	H	L	W	V	E	G	I	F	E	S	F	A	V	I	L	I	G	Y	L	L	V	D	M	K	L	T	T	I	R	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	T	G	V	V	G	M	G	H	H	Y	F	W	E	G	D	H	S	I	W	L	A	L	G	A	C	F	S	A	L	E	V	I	P	L	C	L	L	V	W	E	A	Y	T	H	Y	R	V	Y	R	D	S	M	L	D	
BJDN01000035.1_5 # 3283 # 532	I	V	H	L	W	V	E	G	I	F	E	S	F	A	V	I	L	I	G	Y	L	M	V	D	L	K	L	T	T	T	R	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	M	G	M	G	H	H	Y	F	W	E	G	D	H	S	I	W	L	A	L	G	A	C	F	S	A	L	E	V	I	P	L	C	L	L	W	E	A	Y	T	H	Y	R	V	F	K	D	S	K	V	D		
RKLX01000004.1_37 # 44762 # 4	I	V	H	L	W	V	E	G	I	F	E	S	F	A	V	I	L	I	G	Y	L	L	V	D	M	K	L	T	T	I	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	T	G	V	M	G	M	G	H	H	Y	F	W	E	G	D	H	S	I	W	L	A	L	G	S	C	F	S	A	L	E	V	I	P	L	C	L	L	W	E	A	Y	T	H	Y	K	V	R	D	S	Q	L	E			
SLUL01000001.1_44 # 43903 # 4	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	L	V	Q	M	K	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	A	E	V	W	I	A	L	G	A	V	F	S	A	L	E	V	I	P	L	T	L	L	V	L	E	A	Y	E	Q	Y	K	M	M	R	D	G	G	V	N	
CP012034.1_1205 # 1155810 # 1	I	V	H	L	W	V	E	G	I	F	E	S	F	A	V	I	L	I	G	Y	L	L	V	D	M	K	L	T	T	I	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	T	G	V	V	G	M	G	H	H	Y	F	W	E	G	D	H	S	I	W	L	A	L	G	A	C	F	S	A	L	E	V	I	P	L	C	L	L	I	W	E	A	Y	T	H	Y	R	V	Y	H	D	S	Q	K	N	
CM012602.1_2085 # 2263430 # 2	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	M	V	Q	M	K	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	A	E	I	W	I	A	L	G	A	V	F	S	A	L	E	V	I	P	L	T	L	L	L	E	A	Y	E	Q	Y	K	M	M	R	D	G	G	V	N		
CP002692.1_558 # 571722 # 574	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	L	V	Q	M	K	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	A	E	I	W	I	A	L	G	A	V	F	S	A	L	E	V	I	P	L	T	L	L	L	E	A	Y	E	Q	Y	K	M	M	R	D	G	G	V	N		
IQMR01000001.1_991 # 1063335	I	V	H	L	W	V	E	G	I	F	E	S	F	A	V	I	L	I	G	Y	L	M	V	N	M	K	L	T	T	V	R	S	T	V	R	M	L	Y	F	Q	L	I	L	L	G	A	G	V	V	G	M	G	H	H	Y	F	W	E	G	D	H	S	I	W	L	A	L	G	A	C	F	S	A	L	E	V	V	P	L	C	L	L	I	W	D	A	Y	T	H	Y	K	I	M	R	D	S	G	K	E	
PVNE01000006.1_92 # 97439 # 9	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	L	V	Q	M	G	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	T	G	H	H	Y	Y	W	I	G	A	P	E	I	W	I	A	L	G	A	V	F	S	A	L	E	V	I	P	L	T	L	L	L	E	A	Y	G	H	Y	R	I	M	K	Q	G	G	V	S		
RSFW01000015.1_107 # 96368 # 9	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	M	V	Q	M	K	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	M	G	A	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	E	V	W	I	A	L	G	S	V	F	S	A	L	E	V	I	P	L	T	L	L	L	E	A	Y	E	Q	Y	K	M	M	R	D	G	G	V	N		
LXMA01000038.1_163 # 131708	I	I	H	L	W	V	E	G	I	F	E	V	F	A	V	V	V	I	G	F	L	L	V	Q	M	K	L	V	T	K	K	S	T	V	R	A	L	Y	F	Q	L	I	L	L	G	S	G	V	I	G	I	G	H	H	Y	Y	N	G	S	S	A	E	I	W	I	A	L	G	A	V	F	S	A	L	E	V	I	P	L	T	L	L	L	E	A	Y	E	Q	Y	K	M	M	R	D	G	G	V			

* Amino acids essential for function

• Multiple sequence alignment and phylogenetic tree

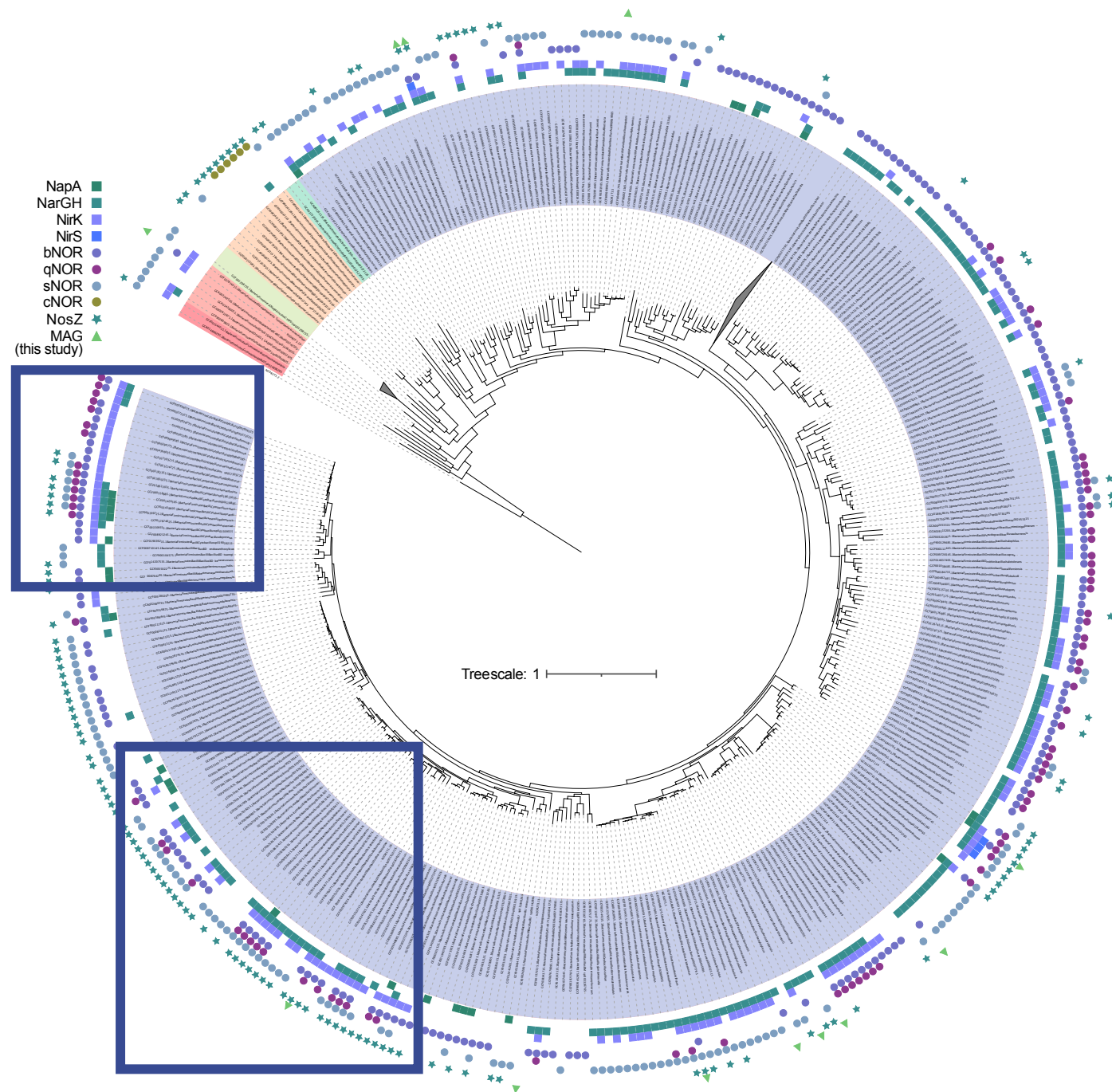


sNOR – found in both Bacteria and Archaea
bNOR – only found in the *Bacilli*
qNOR – many Clade II (lacks C subunit)



N₂O production and consumption by *Firmicutes* underestimated

- Complete denitrification (clade II) in diverse environments
- Cultivation needed to demonstrate the relevance of different nitric-oxide reductase



Outline

- Lecture
 - Introduction to taxonomic and functional annotation databases
 - Pairwise alignment vs. HMMs
 - Caveats of assigning function
- Case studies
- Practical
 - Evaluate bins
 - Taxonomically classify MAGs
 - Functionally annotate MAGs
 - Create a representative bin set