Construction of environmental archaeal ammonia monooxygenase (*amoA*) & dissimilatory sulfite reductase (*dsrA*) clone libraries

Dusty, Maren, Hyunsoo, Abi, Bimo

Introduction

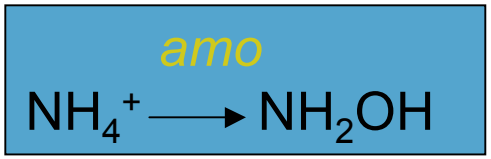
What?

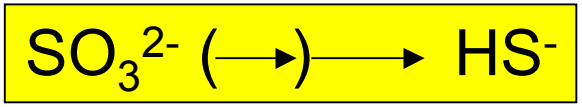
- A collection of environmental PCR fragments

Why?

- Phylogenetic and functional gene diversity
- Connection to the environment

The Genes





AmoA

- ~635 bp gene encodes α-subunit of ammonia monooxyenase (*amo*)
- Highly conserved across AOB
- AOA play a significant role in NH_4^+ - oxidation (Francis et al., 2005)

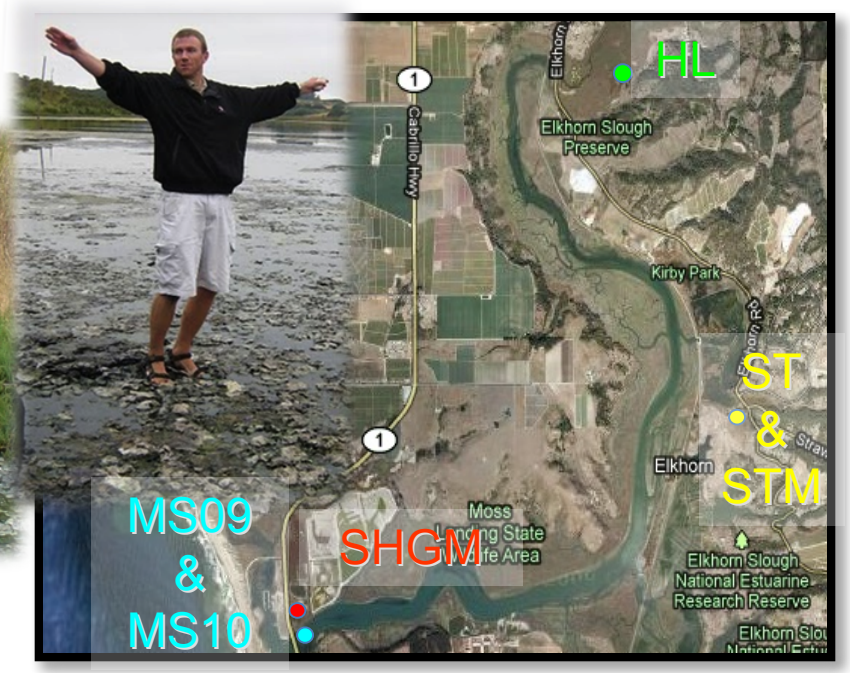
dsrA

- Gene for dissimilative sulfite reductase
- Catalyzes reduction of SO_3^{2-} to HS^- for energy conservation

Sampling- *amoA*

	Sample Location	Year	Tag	Reads	Usable Seq.
Liquid	Hopkins Aquarium	2009	aHA	Forward & Reverse	35
	Mouth of Slough	2009	aMS	Forward & Reverse	22
	Beach Sand	2009	aBS	Forward & Reverse	37
	Sea Water	2009	aSW	Forward & Reverse	35
	Cold Sandfilter	2009	aCS	Forward & Reverse	64
	Warm Sandfilter	2009	aWS	Forward & Reverse	44

Sampling- *dsrA*



	Sample Location	Year	Tag	Reads	Usable Seq.
Sediment	Hudson's Landing (natural)	2009	HL	Forward & Reverse	32
	Strawberry (natural)	2010	ST	Forward	25
	Strawberry (mat)	2010	STM	Forward	5
	Mouth of Slough (natural)	2009	MS09	Forward & Reverse	25
	Mouth of Slough (natural)	2010	MS10	Forward	10
	Shotgun (mat)	2010	SHGM	Forward	7

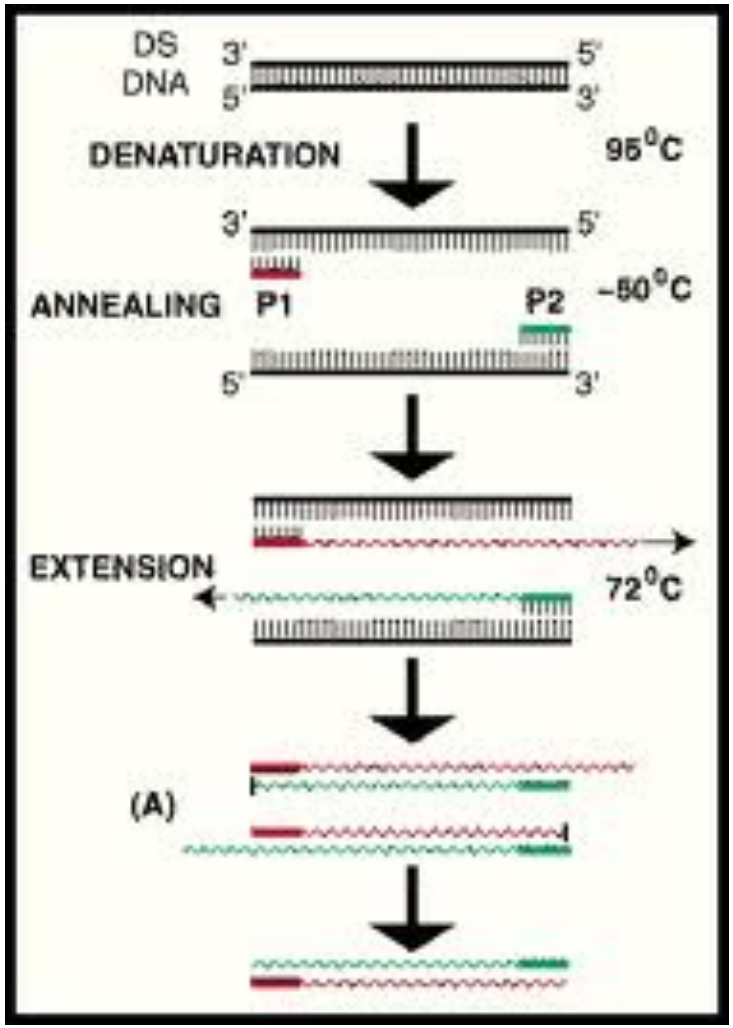
Experimental part

1. DNA extraction

- Soil or water
- Soil: "***MO*** Bio UltraTM Soil DNA Isolation Kit"

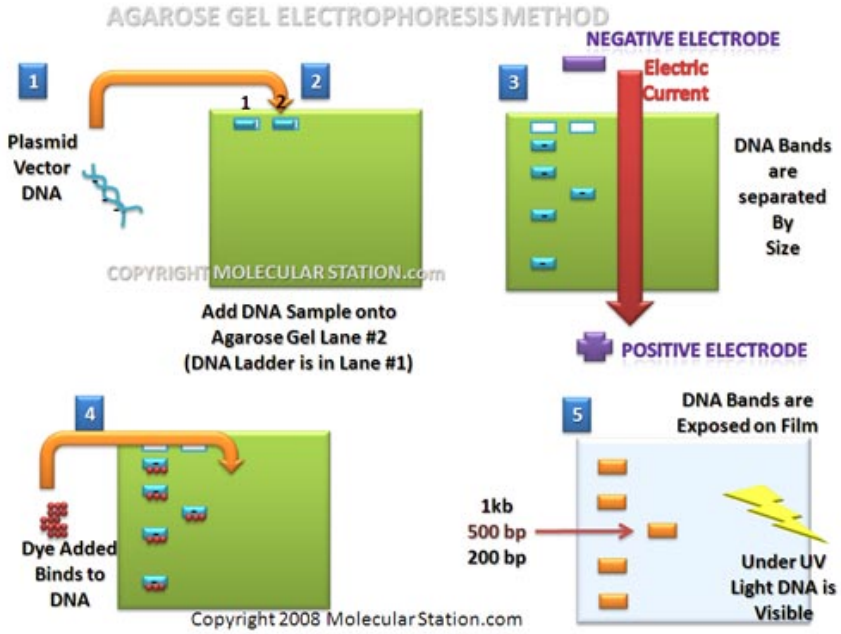
2. PCR

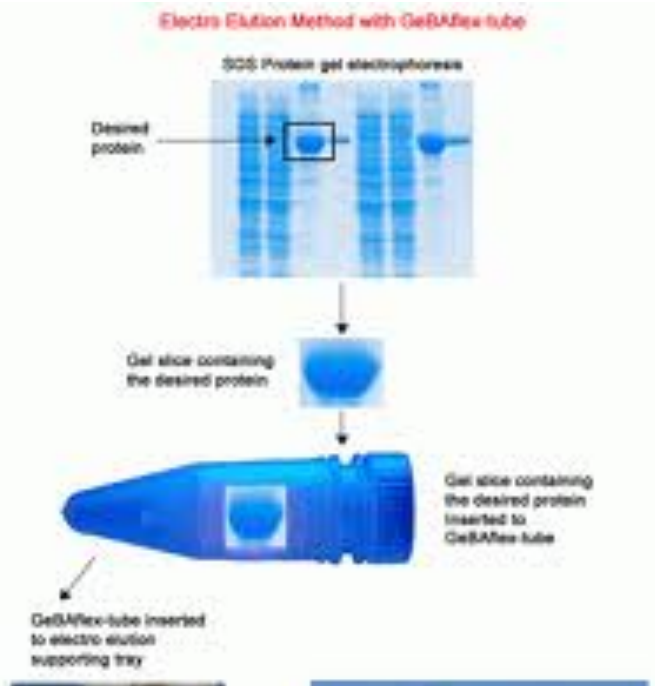
- With degenerate primers for *amoA* & *dsrA*



Experimental part

3. Agarose Gel electrophoresis



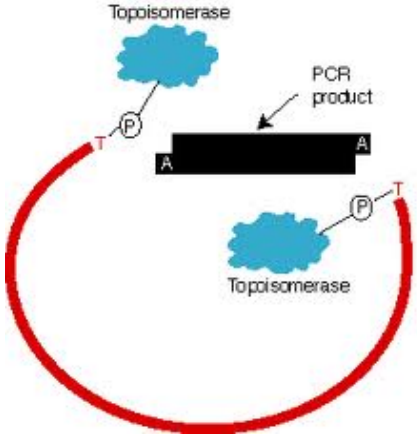


Purification of PCR product via Gel extraction

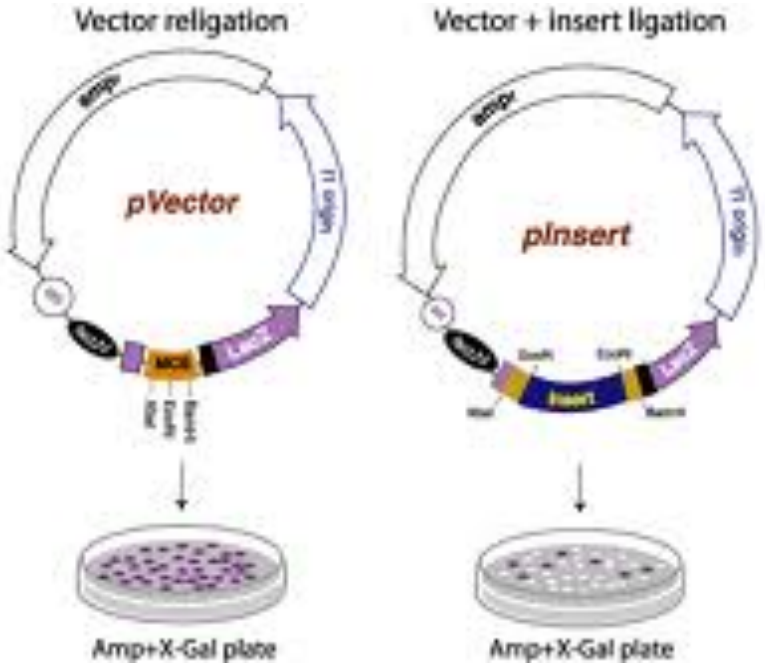
Experimental part

4. TA cloning

a. Ligation (TOPO)

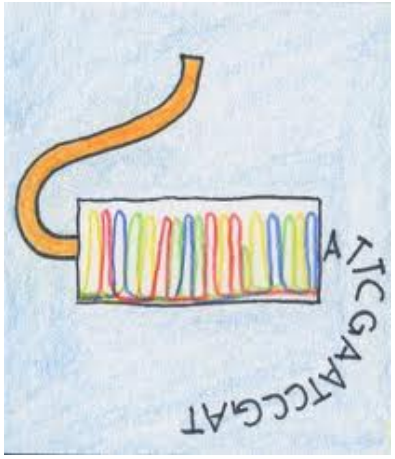


b. Transformation *E. coli*



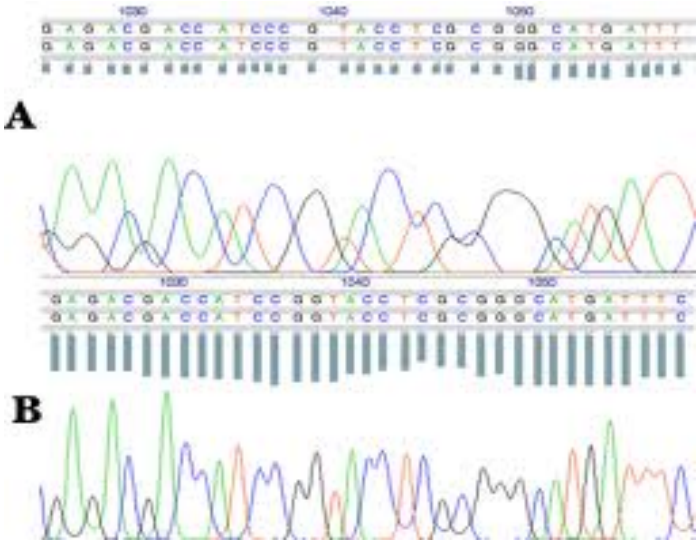
c. Blue & **white** colony selection

Sequencing & Analysis



- Sequencing of clone library
  - Sanger sequencing

- Comparative sequence analysis
  - Trimming sequences
  - Same size
  - Alignment
  - Create a FASTA file



Diversity of functional genes?

- Rarefaction analysis
- Total diversity coverage

MOTHUR

- Relatedness of our samples
- Relatedness of 2009 + 2010 samples
- Positioning of our samples in phylogenetic tree with all known dsrA/AmoA gene sequences

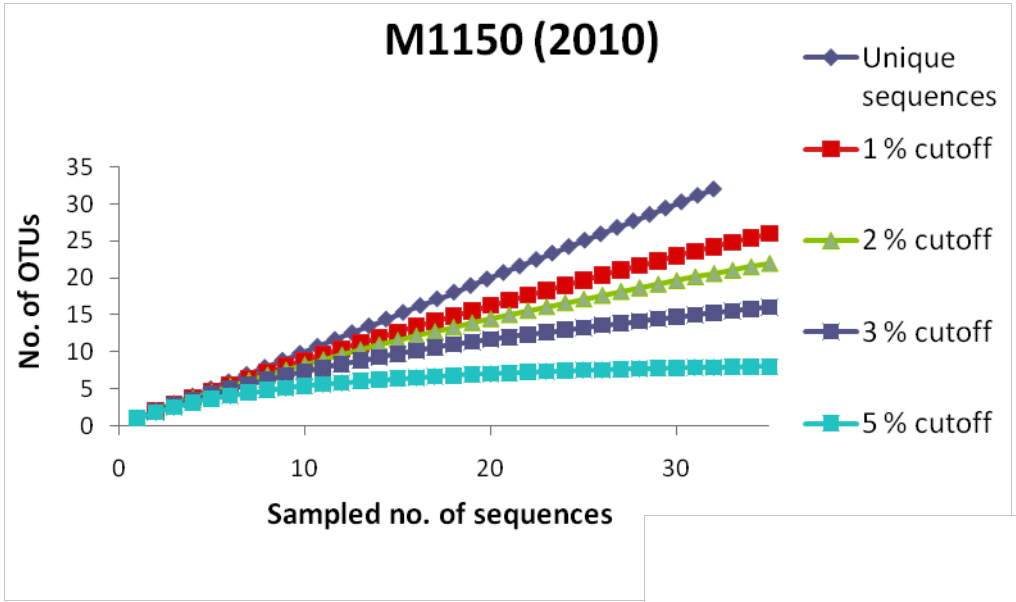
geneious

GENEIOUS
&
ARB

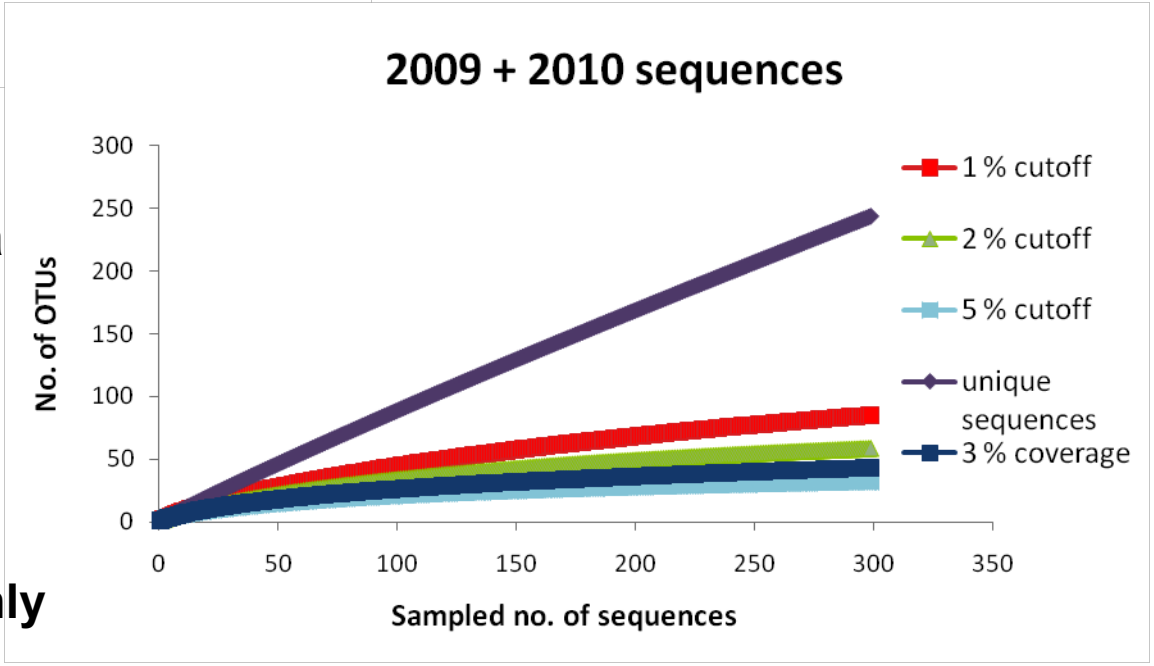
Our datasets

	AmoA	dsrA
Initial sequences 2010	2 * 96	4 * 48
Trimmed sequences 2010	50 + 35	7 + 25 + 8 + 5
Sequences from 2009	225 (6 sites)	67 (2 sites)
Sequences in database	1380 (18100 in Genbank)	4000 (in Genbank)

AmoA: rarefaction curves



- total diversity covered assuming a 5 % cutoff
- similar for AmoA Kelp and for all M1150 + Kelp



- already 80 % of total diversity covered assuming a 1 % cutoff
- total diversity covered for a 17 % cutoff

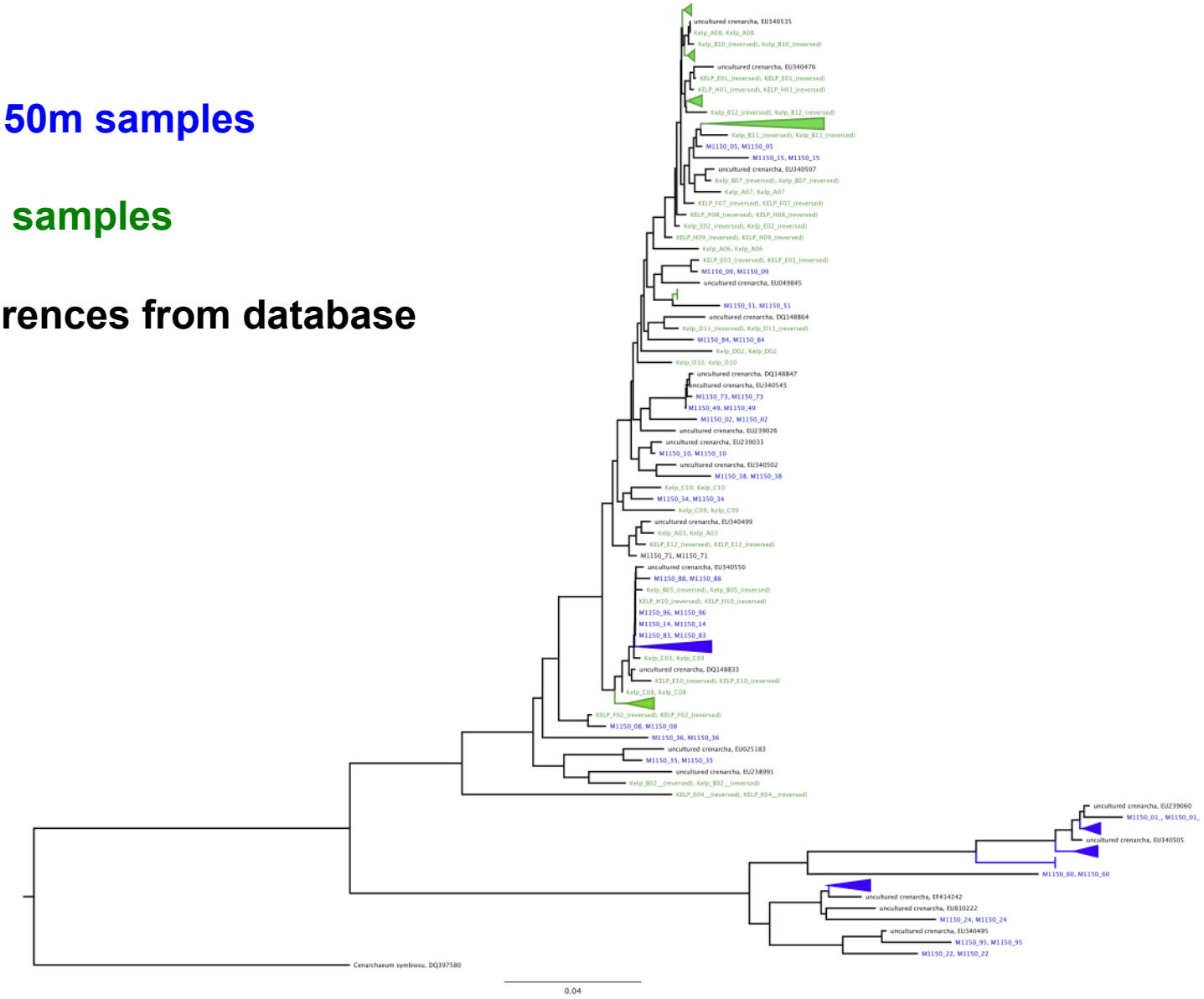
➡ **AmoA from ocean + aquarium samples are highly conserved**

AmoA: phylogenetic tree of 2010 samples

M1 150m samples

Kelp samples

References from database



AmoA: comparison to known sequence data

- Most closely related sequences to 2010 samples (M1 150m & Kelp)
  - Polar Ocean (Kalanetra et al., unpublished)
  - Gulf of California (Beman et al., ISME 2008)
  - Monterey Bay M2_40m & MB_C130m (Francis et al., PNAS, 2005)

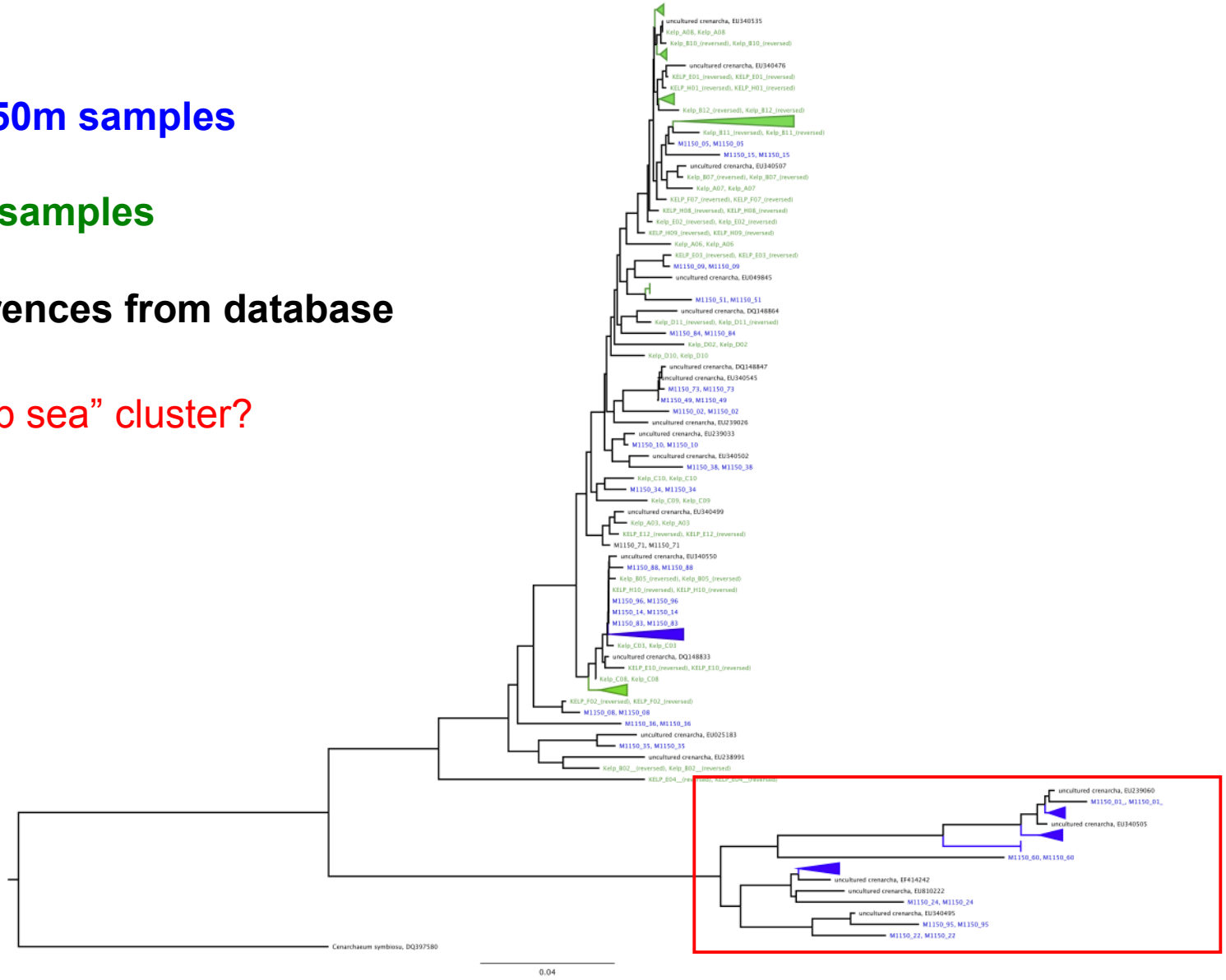
AmoA: phylogenetic tree of 2010 samples

M1 150m samples

Kelp samples

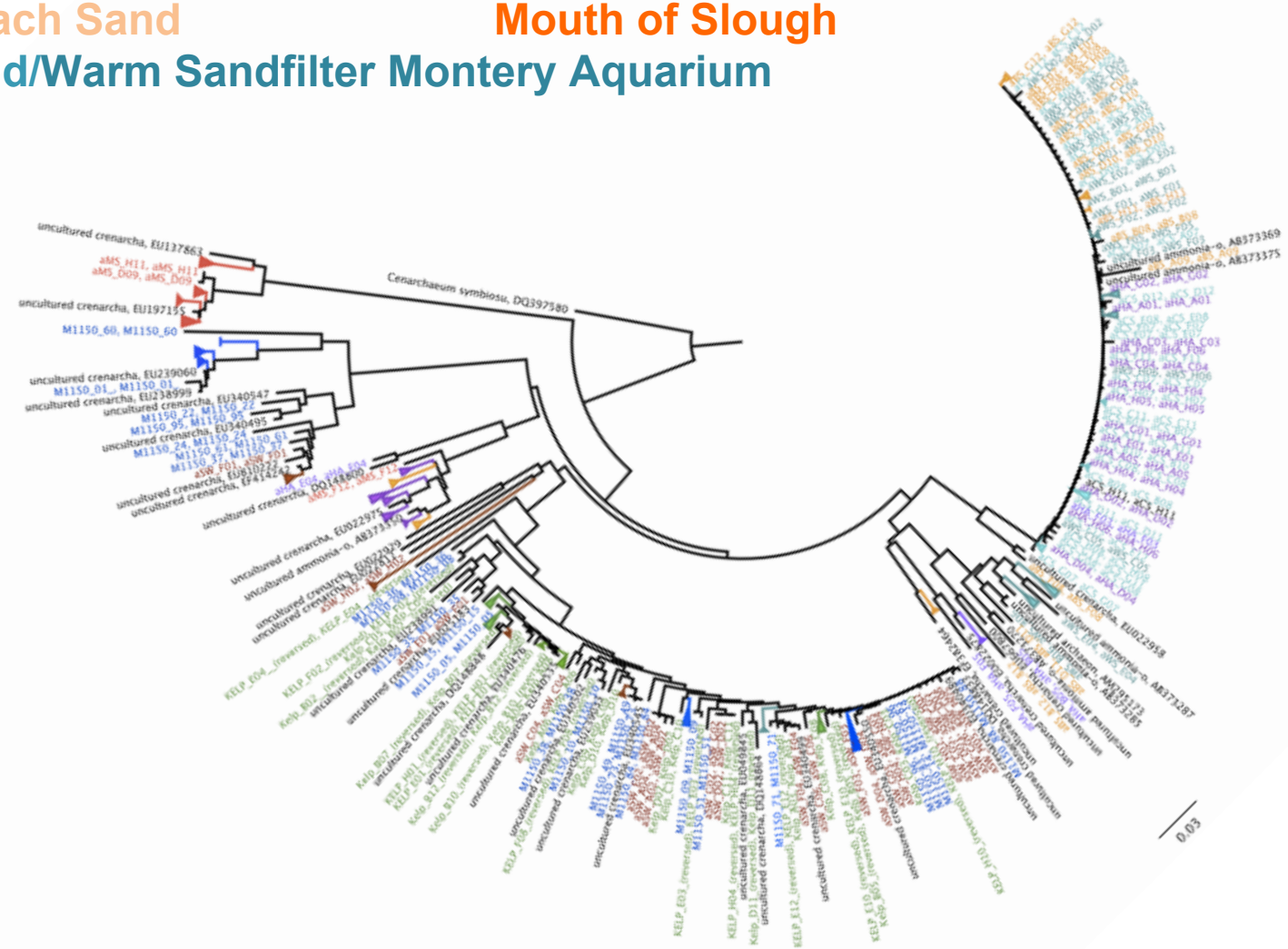
References from database

“deep sea” cluster?



AmoA: phylogenetic tree of 2009 & 2010 samples

Kelp samples (2010)
M1 150m samples (2010)
Beach Sand
Cold/Warm Sandfilter Montery Aquarium
Hopkins Aquarium
Seawater
Mouth of Slough



AmoA: comparison to known sequence data

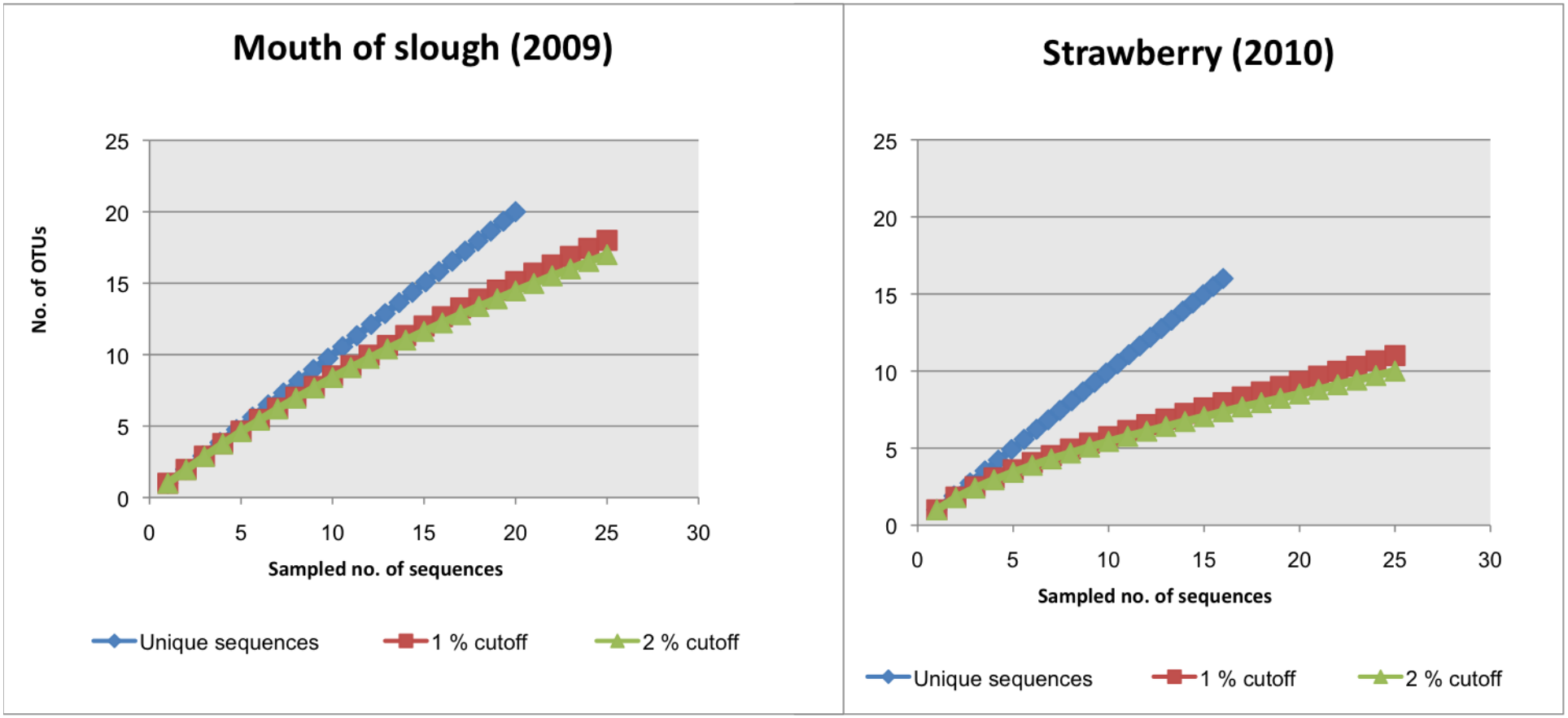
- Most closely related sequences to 2010 samples (M1 150m & Kelp)
  - Polar Ocean (Kalanetra et al., unpublished)
  - Gulf of California (Beman et al., ISME 2008)
  - Monterey Bay M2_40m & MB_C130m (Francis et al., PNAS, 2005)

- Distinct clusters:
  - Mouth of Slough samples ⇨ cluster with sandy loam & lake samples
  - Hopkins Aquarium samples ⇨ cluster with estuarine & Japanese aquarium samples
- Most disperse sequences: Beach sand (consistent with literature)

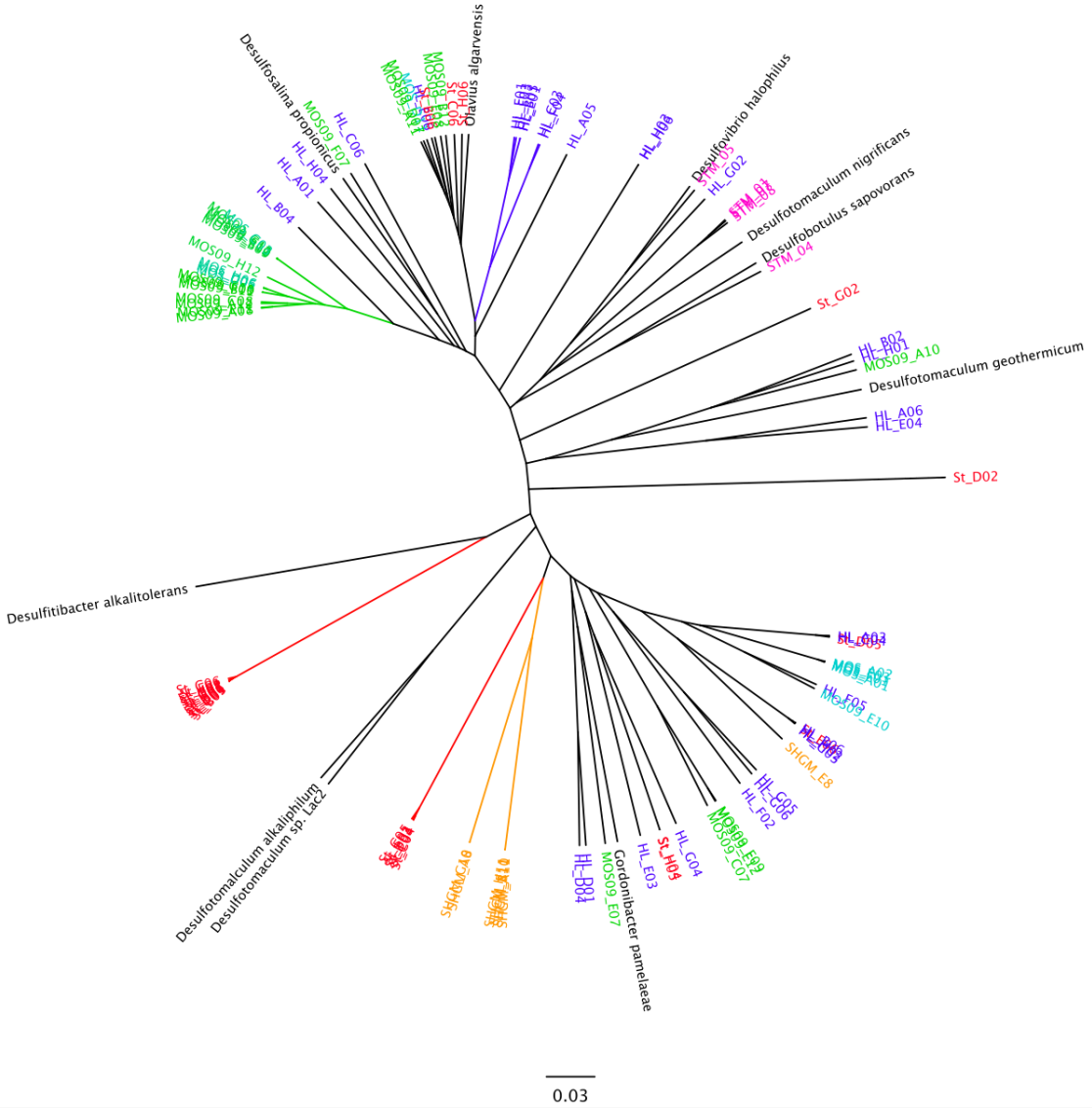
Our datasets

	AmoA	dsrA
Initial sequences 2010	2 * 96	4 * 48
Trimmed sequences 2010	50 + 35	7 + 25 + 8 + 5
Sequences from 2009	225 (6 sites)	67 (2 sites)
Sequences in database	1380 (18100 in Genbank)	4000 (in Genbank)

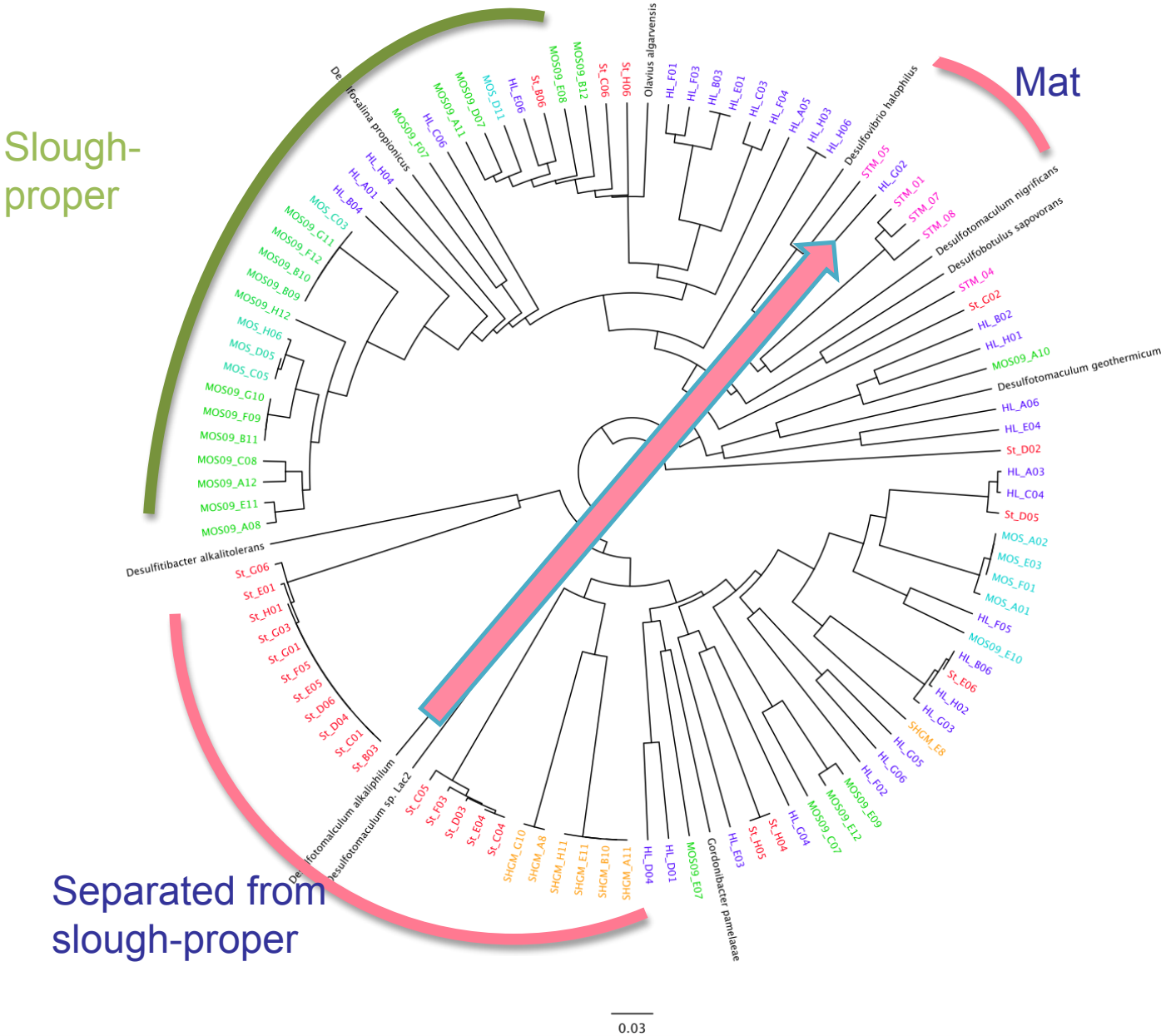
Comparison of diversity for *dsr*A gene

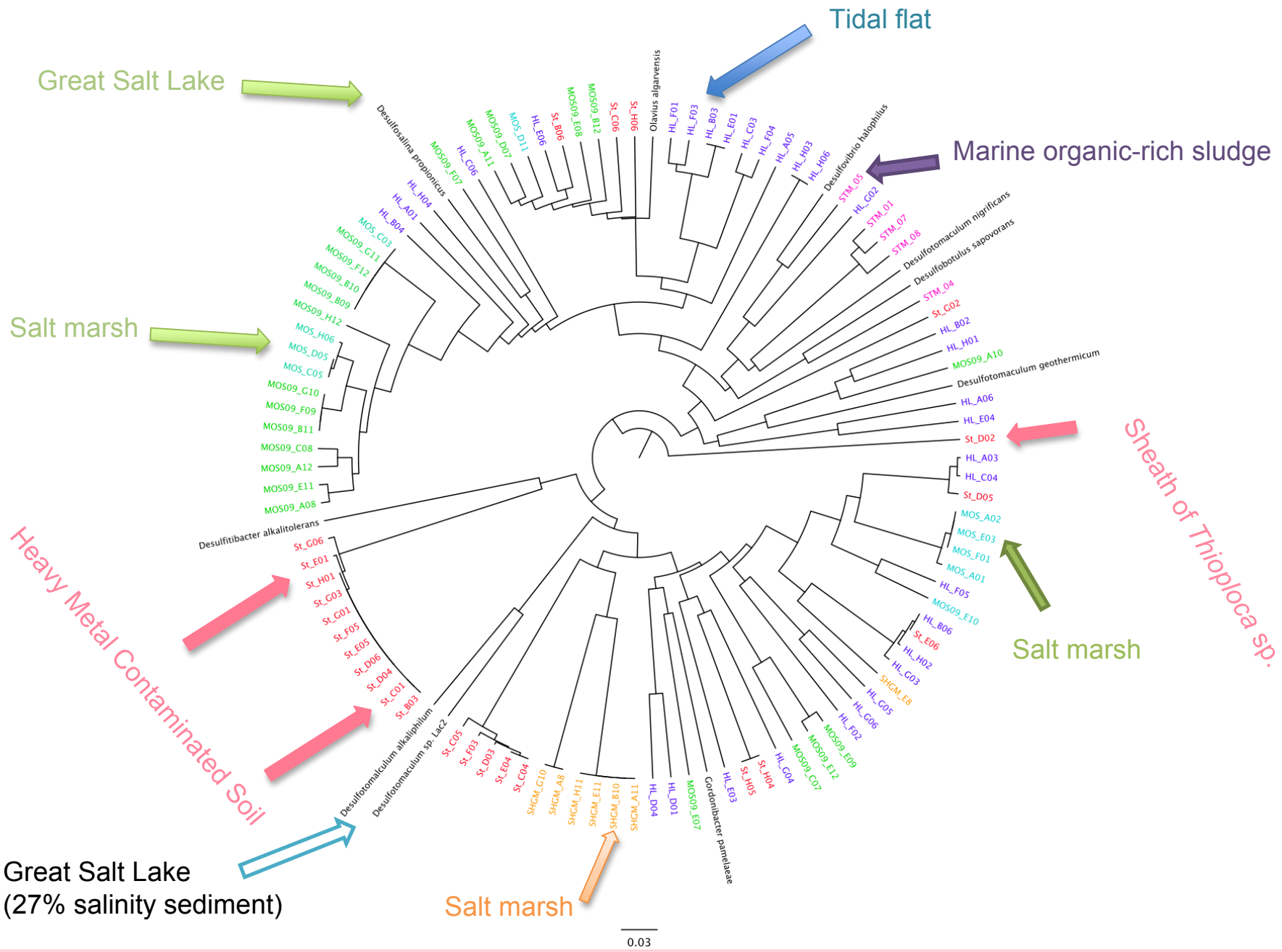


Neighbor-joining tree of dsrA gene sequences sampled in 2009 and 2010

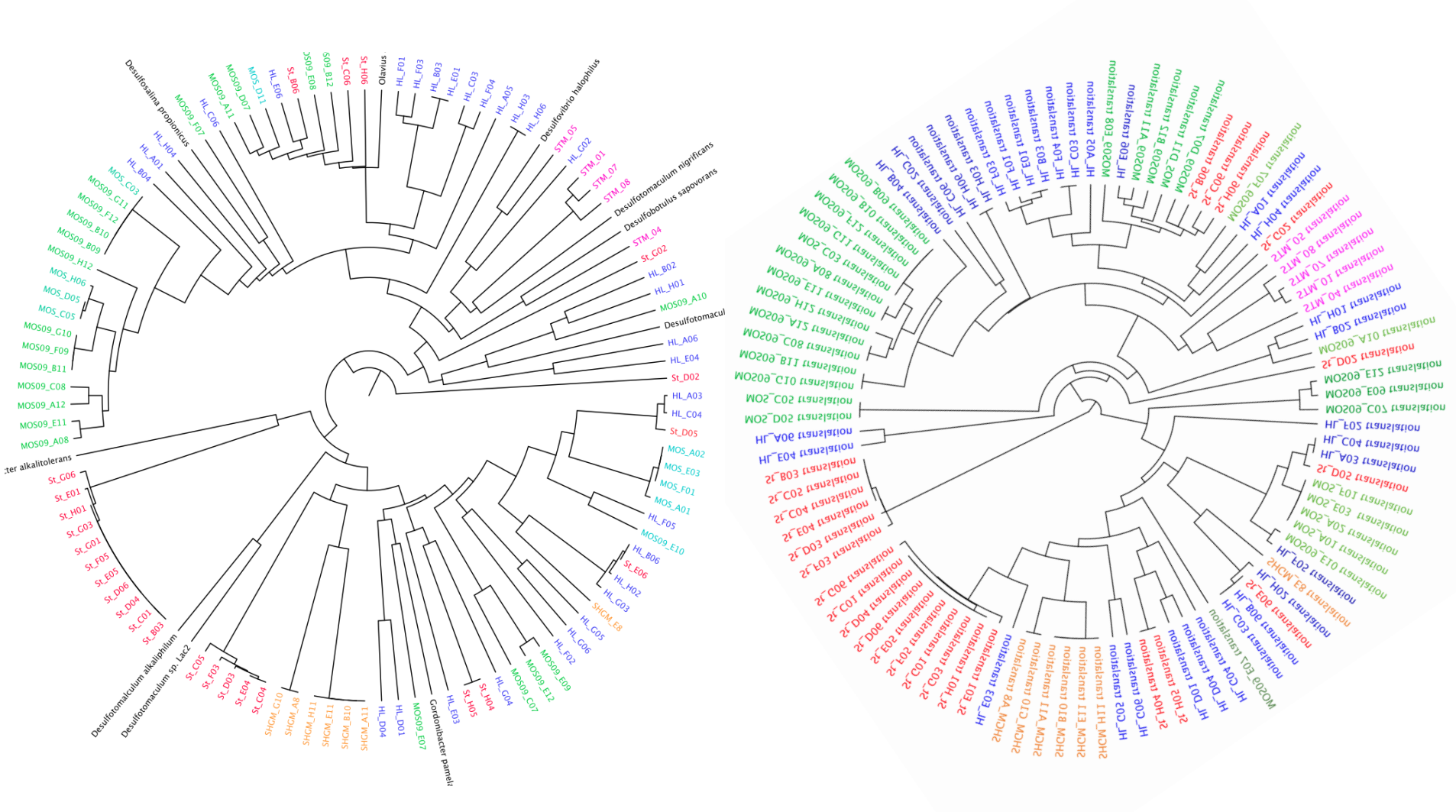



HL
Elkhorn
1
Cabrillo Hwy
Elkhorn Slough
Preserve
Zmudowski
State Beach
Kirby Park
ST
&
STM
1
Elkhorn
Moss
Landing State
Wildlife Area
SHGM
MS09
&
MS10
Elkhorn Slough
National Estuarine
Research Reserve


Slough-proper
Mat
Separated from slough-proper
Olavius algarvensis
Desulfovibrio halophilus
Desulfotomaculum nigrificans
Desulfobotulus sapovorans
Desulfotomaculum geothermicum
Desulfitibacter alkalitolerans
Desulfotomaculum alkaliphilum
Desulfotomaculum sp. Lac2
Gordonibacter pamelaeae
0.03


Tidal flat
Great Salt Lake
Marine organic-rich sludge
Salt marsh
Sheath of Thioploca sp.
Salt marsh
Heavy Metal Contaminated Soil
Great Salt Lake
(27% salinity sediment)
Salt marsh
0.03
Olavius algarvensis
Desulfovibrio halophilus
Desulfotomaculum nigrificans
Desulfobotulus sapovorans
Desulfotomaculum geothermicum
Desulfosalina propionicus
Desulfitibacter alkalitolerans
Desulfotomaculum alkaliphilum
Desulfotomaculum sp. Lac2
Gordoniibacter pamelaeae

Nucleotide vs. Amino Acid



Reduced estimated OTUs for 5% cutoff binning from 65 to 49 for total dsrA

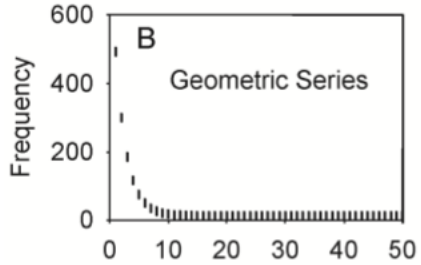
Diversity still high when considering primary structure

What Might We Say?

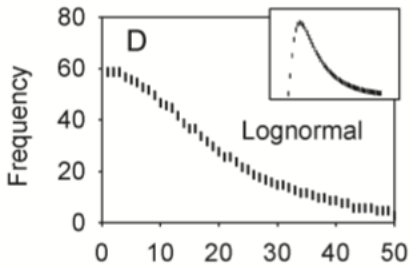
- Clone Libraries:
  - A collection of gene sequences and related environmental or functional information
  - Allows possible relationships between environment and function to be drawn
  - Allows comparison of relative community richness between environments
- Diversity of the sample must be well-represented for the questions asked to be appropriately answered

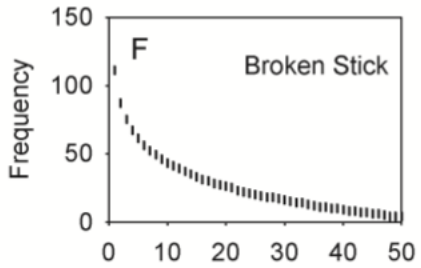
What is “Well-Represented”?

- Yield stable Chao estimates
  - Too low: “highly uneven phylotype abundance distribution”; a lot of single-frequency bins



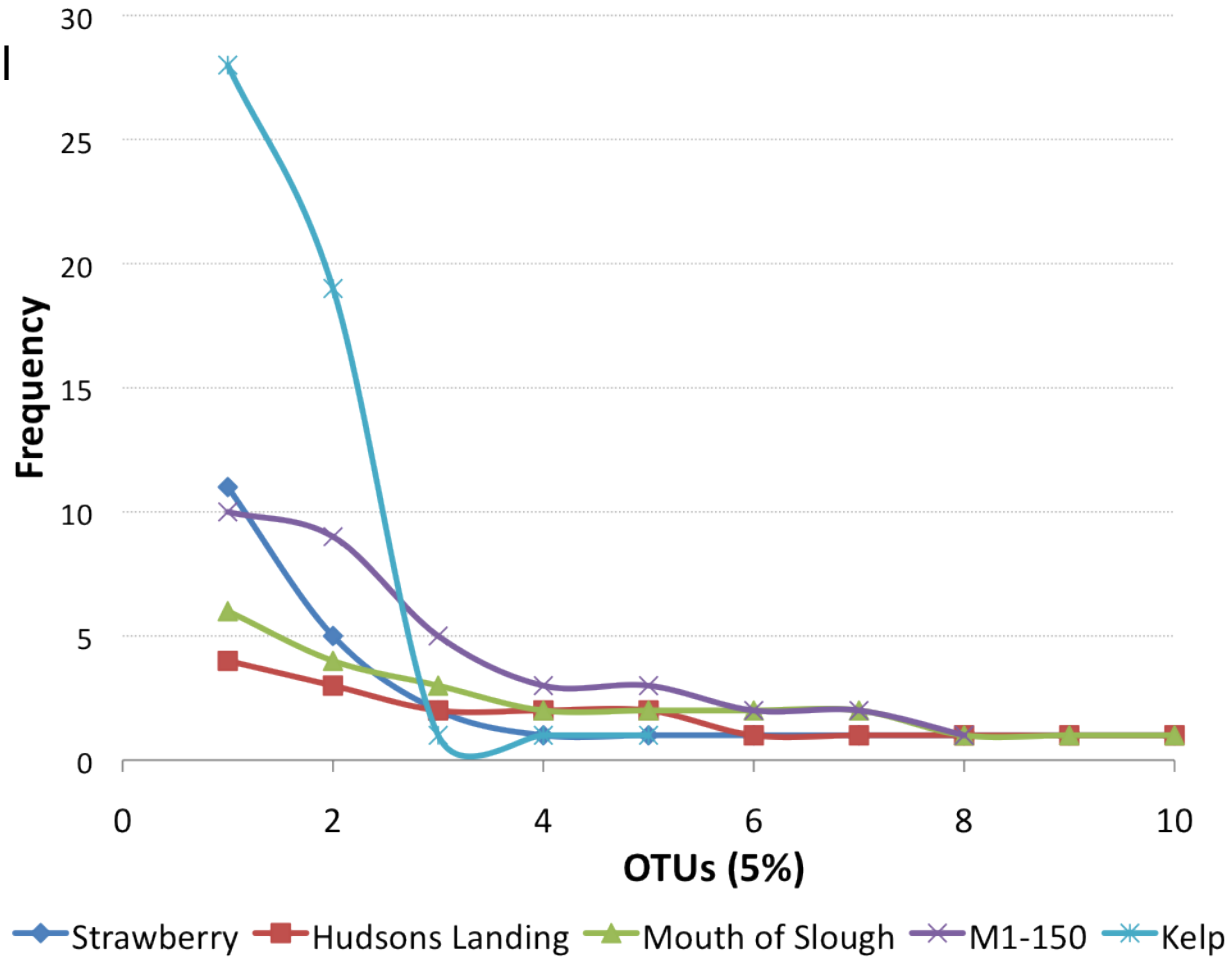
  - Satisfactory: “more even phylotype abundance”





What is “Well-Represented”?

- Too low: single-bin tail
  - Strawberry
  - Hudson’s Landing
  - Mouth of Slough
  - All of the others with <10 samples
- Satisfactory: “more even phylotype abundance”
  - M1-150
  - Kelp
- *dsrA*: Needs more samples!
- *amoA*: diversity well captured



Improving Analysis

- Need more sequences!
  - DNA extraction bias
  - Forward and Reverse reads would increase amount of usable sequences from given purified colonies

Thanks, everyone!

Backup

- **Degenerate primers**
- *Main article: Degenerate bases*
- Sometimes *degenerate primers* are used. These are actually mixtures of similar, but not identical primers. They may be convenient if the same gene is to be amplified from different organisms, as the genes themselves are probably similar but not identical. The other use for degenerate primers is when primer design is based on protein sequence. As several different codons can code for one amino acid, it is often difficult to deduce which codon is used in a particular case. Therefore primer sequence corresponding to the amino acid isoleucine might be "ATH", where A stands for adenine, T for thymine, and H for adenine, thymine, or cytosine, according to the genetic code for each codon, using the IUPAC symbols for degenerate bases. Use of degenerate primers can greatly reduce the specificity of the PCR amplification. The problem can be partly solved by using touchdown PCR.
- *Degenerate primers* are widely used and extremely useful in the field of microbial ecology. They allow for the amplification of genes from thus far uncultivated microorganisms or allow the recovery of genes from organisms where genomic information is not available. Usually, degenerate primers are designed by aligning gene sequencing found in GenBank. Differences among sequences are accounted for by using IUPAC degeneracies for individual bases. PCR primers are then synthesized as a mixture of primers corresponding to all permutations. TOPO Cloning
- From Wikipedia, the free encyclopedia

- **TOPO Cloning** is a molecular biology technique in which DNA fragments amplified by either Taq or Pfu polymerases are cloned into specificvectors without the requirement for DNA ligases.
- [edit]Principle
- The technique utilises the inherent biological activity of DNA topoisomerase I. The biological role of topoisomerase is to cleave and rejoinsupercoiled DNA ends to facilitate replication. Vaccinia virus topoisomerase I specifically recognises DNA sequence 5´-(C/T)CCTT-3'. Duringreplication, the enzyme digests DNA specifically at this sequence, unwinds the DNA and re-ligates it again at the 3' phosphate group of thethymidine base.