



Using short-read SOLiD sequencing technology to characterize the microbial community in 500ml of seawater.

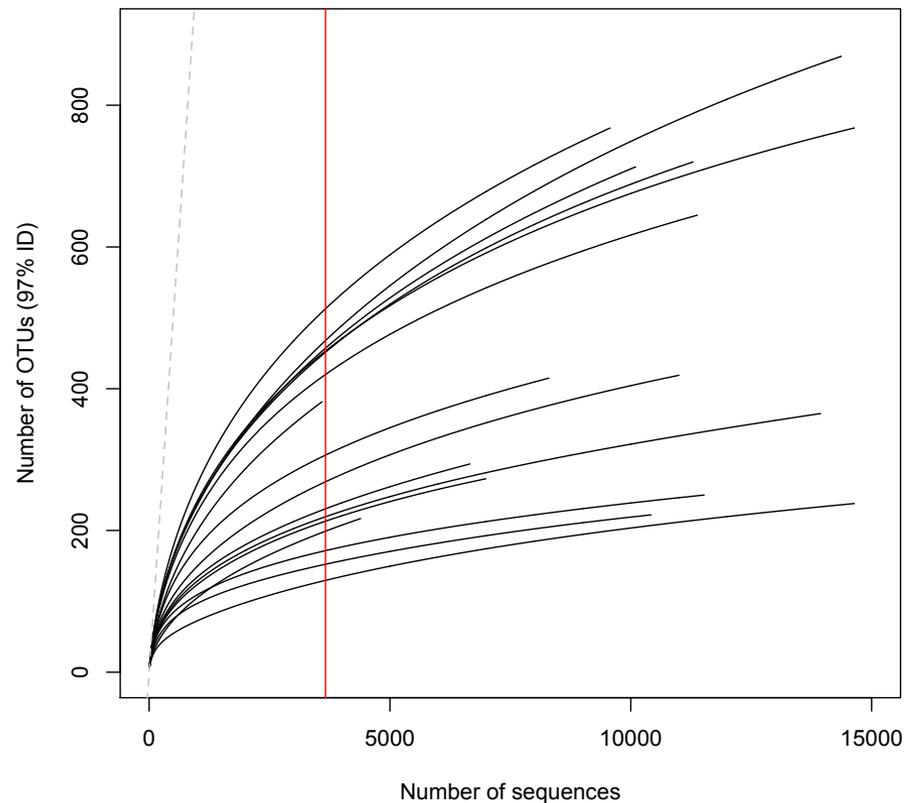
FISH 546: Bioinformatics

Winter 2013

Rachel Lange

Challenges of characterizing microbial communities

- Overwhelming abundance
 - Nearly impossible to detect each member in a sample



Challenges of characterizing microbial communities

- Majority of taxa are uncultured or unknown.
 - NCBI has 2601 complete prokaryotic genomes
 - Potentially >120,000 bacterial species in the ocean (Zinger et al 2011).

Challenges of characterizing microbial communities

- Majority of genes, even within complete genomes, have unknown functions.

Some omics words we would like to see

By Iddo on March 13th, 2013

Advertisomics: environmental sequencing aimed at obtaining popular press coverage with little or no scientific value. Samples obtained from an environment otherwise not of microbiological interest. *“Hey, did you hear they swabbed the car wheels in the building’s parking lot and found that the microbes all cluster by tire brand name?”*

Celebromics: sequencing the genome or microbiome of a celebrity. Generally the sequence is not even published, but just the act of sequencing it provides publicity for the lucky lab, the celeb, and maybe even a microbial species or two. *“They sequenced the genome of Keith Richards, and found a duplicated set of multiple drug resistance genes.”*

Contaminomics: sequencing results published prematurely, and later discovered that the major finding is the result of a contamination.

DuhOmics: unsurprising results from a genomic study. Usually confirming common knowledge that did not require a genomic study in the first place.

Lazarusomics: sequencing the genome of an extinct animal, including hominids, with the implicit or explicit promise that we will be able, very soon, to reverse the extinction.

Shockomics: related to advertisomics. Sequencing for shock value and pop publicity. Usually involving human or animal bodily secretions or parts you’d rather not have known about.

TooMuchInformationOmics: A result of the personal genomics and microbiome industry. No, I am not interested in that heel spur gene that you got from your grandmother, nor am I interested in the novelty of the chlamydia strain they found in your partner’s microbiome.

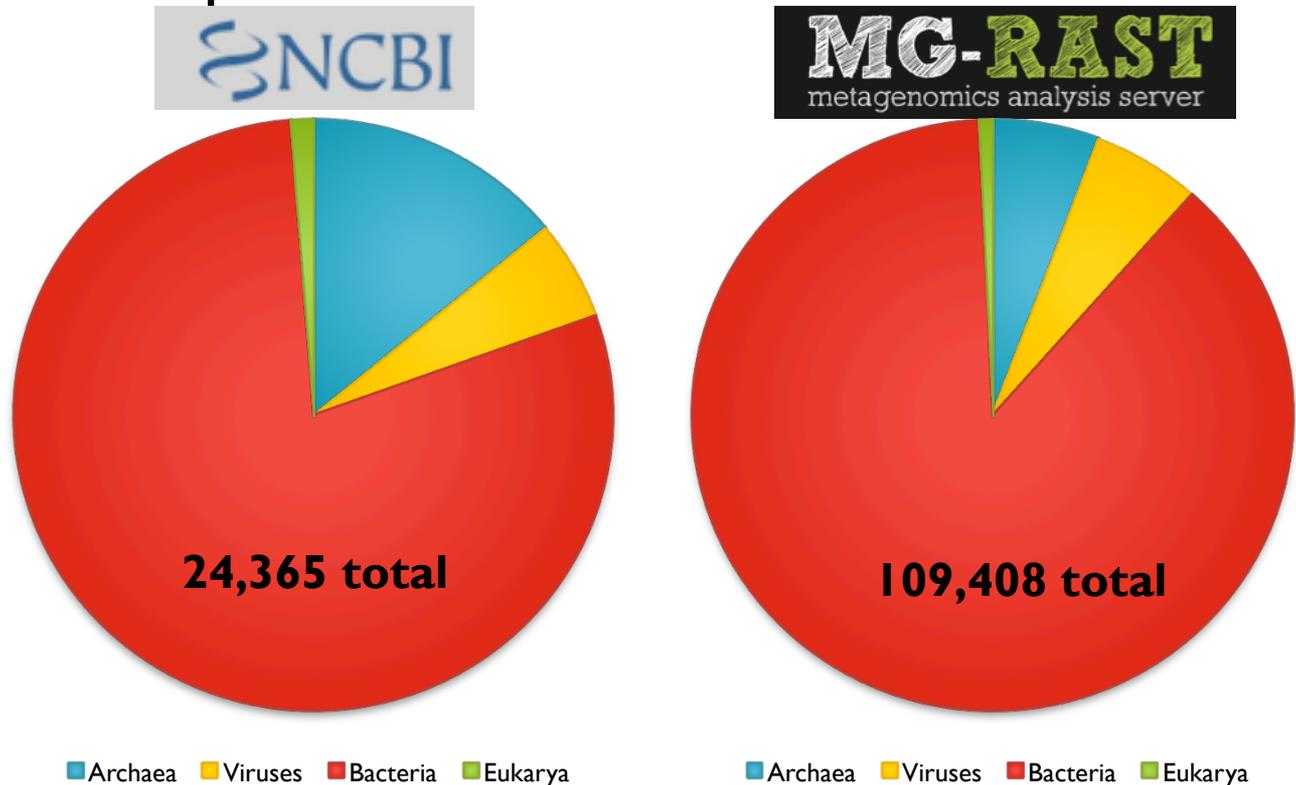
ZZZomics: an omics paper that makes you fall asleep half way through the introduction.

Main objective

- Can we fully characterize the microbial community in a small water sample from Golden Gardens, a human-disturbed environment?
- Approach:
 - Obtained an assembled metagenome from a single water sample.
 - See who's there and what they're doing.

I. Who's there?

- Approach
 - Used a web-server, MG-RAST, to assign taxonomy to each contig from the GG metagenome.
 - Compared to local blastn to NCBI database.

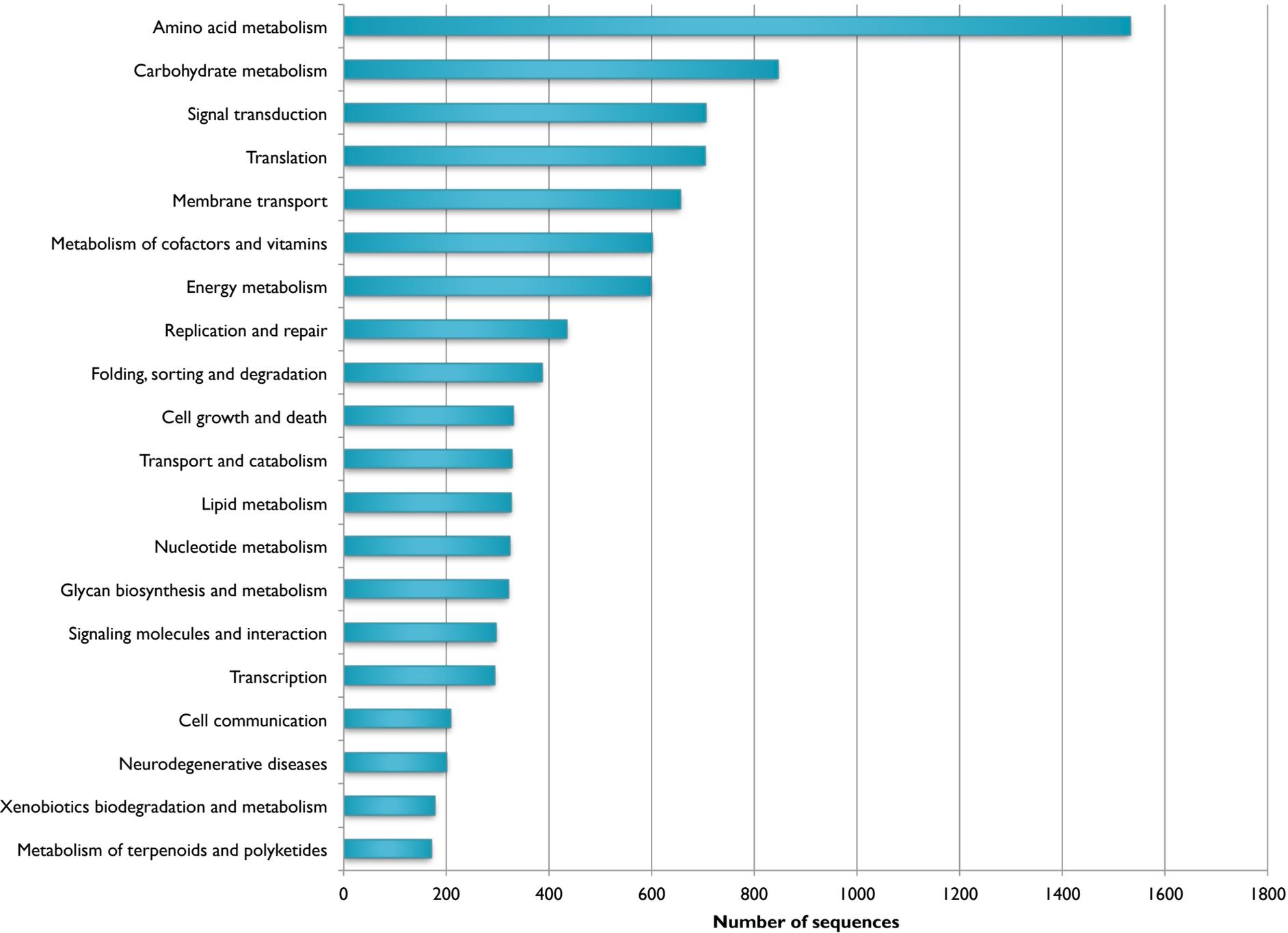


I. Who's there?

metagenome ▲▼	source ▲▼	domain ▲▼	phylum ▲▼	class ▲▼	order ▲▼	family ▲▼	genus ▲▼	species ▲▼	abundance ▲▼
4516213.3 ▼	M5NR ▼	all ▼							< ▼
4516213.3	M5NR	unassigned	unassigned	unassigned	unassigned	unassigned	unassigned	unassigned	31134
4516213.3	M5NR	Bacteria	Proteobacteria	Alphaproteobacteria	unclassified (derived from Alphaproteobacteria)	unclassified (derived from Alphaproteobacteria)	Candidatus Pelagibacter	Candidatus Pelagibacter ubique	8964
4516213.3	M5NR	Bacteria	Proteobacteria	Gammaproteobacteria	unclassified (derived from Gammaproteobacteria)	unclassified (derived from Gammaproteobacteria)	unclassified (derived from Gammaproteobacteria)	gamma proteobacterium HTCC2207	3522
4516213.3	M5NR	Archaea	Thaumarchaeota	unclassified (derived from Thaumarchaeota)	Nitrosopumilales	Nitrosopumilaceae	Nitrosopumilus	Nitrosopumilus maritimus	2968
4516213.3	M5NR	Bacteria	Proteobacteria	Alphaproteobacteria	unclassified (derived from Alphaproteobacteria)	unclassified (derived from Alphaproteobacteria)	Candidatus Pelagibacter	Candidatus Pelagibacter sp. HTCC7211	2742
4516213.3	M5NR	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	unclassified (derived from Rhodobacterales)	unclassified (derived from Rhodobacterales)	Rhodobacterales bacterium HTCC2255	2730
4516213.3	M5NR	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Gramella	Gramella forsetii	2149
4516213.3	M5NR	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Psychroflexus	Psychroflexus torquis	1531
4516213.3	M5NR	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Maribacter	Maribacter sp. HTCC2170	1351
4516213.3	M5NR	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	Flavobacterium johnsoniae	1347
4516213.3	M5NR	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Polaribacter	Polaribacter sp. MED152	1324
4516213.3	M5NR	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Roseobacter	Roseobacter denitrificans	1245
4516213.3	M5NR	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Croceibacter	Croceibacter atlanticus	1056
4516213.3	M5NR	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Polaribacter	Polaribacter irgensii	961
4516213.3	M5NR	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	Flavobacterium psychrophilum	949

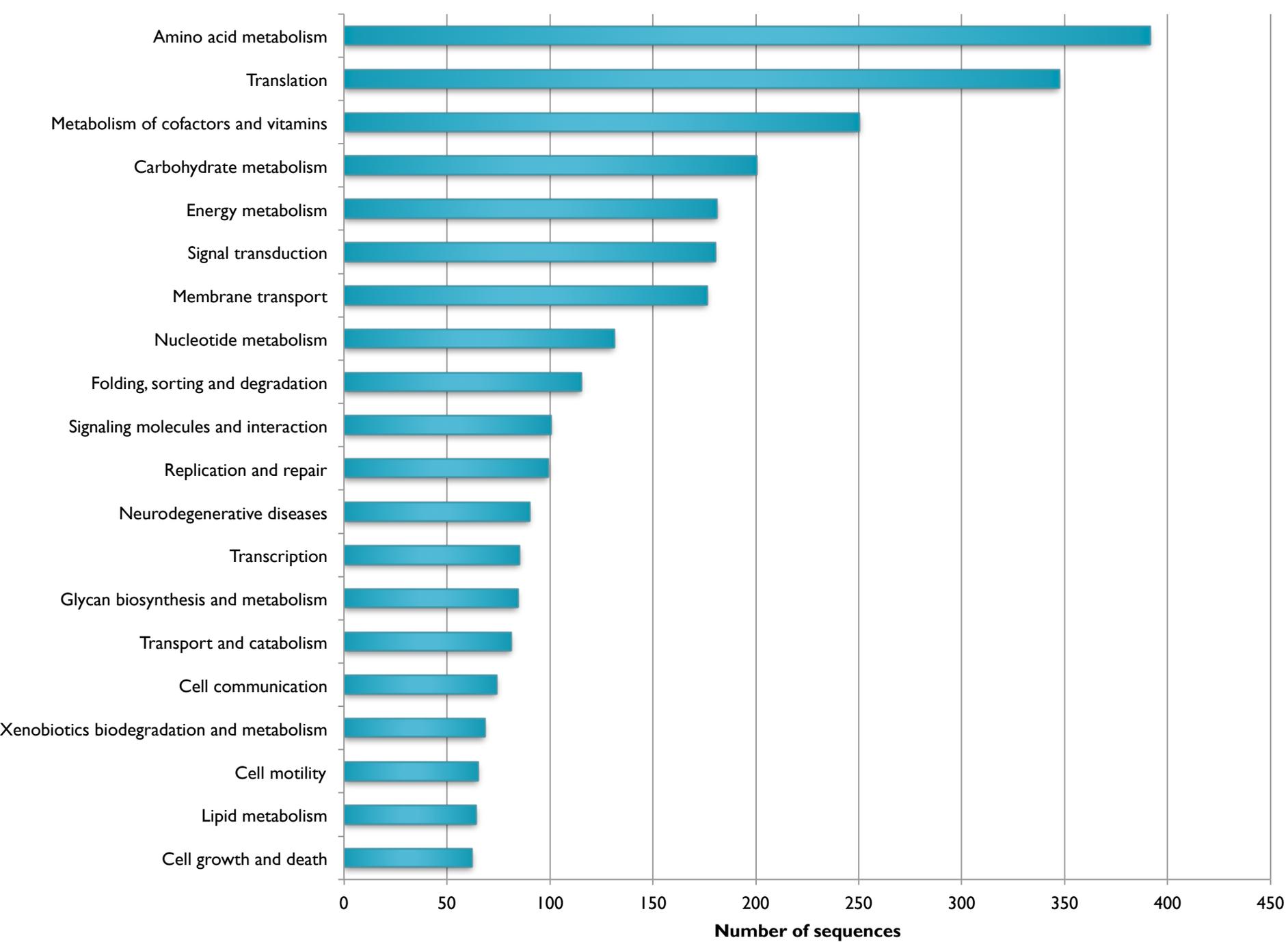
What are the abundant types doing?

- Candidatus Pelagibacter ubique (SAR11)
 - One of the most common marine bacteria
 - Accounted for 6.095% of sequences in GG



What are the abundant types doing?

- **Nitrosopumilus maritimus**
 - One of the most common marine archaea
 - Accounted for 2.081% of sequences in GG
 - Ammonia oxidizing capability in mesophilic, aerobic environments like GG.



Do bacterial or archaeal ammonia oxidizers dominate the community?

- Approach:
 1. Assemble fasta file of reference amoA gene sequences from published bacterial or archaeal studies.
 2. Make a blast db using GG metagenome.
 3. BlastN to identify all bacterial or archaeal amoA genes in the GG metagenome
 4. Compare abundance, diversity, and phylogeny of each domain.
- Tools used: NCBI, blastn, Galaxy, Clustal, MrBayes, iPlant.

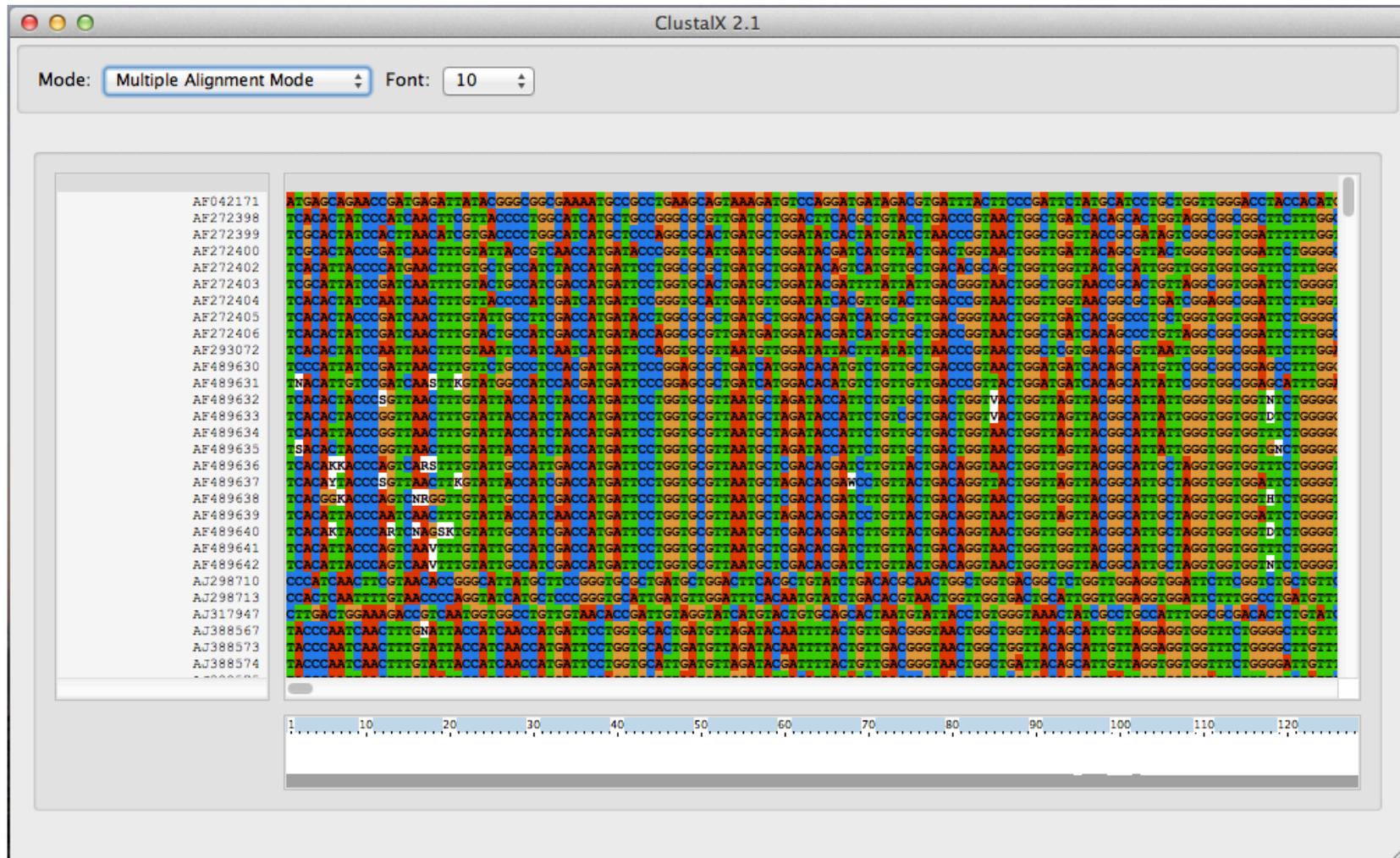
Abundance of amoA genes

	Bacterial	Archaeal
Reference amoA	46	59
GG amoA	309	855
Total sequences	95994	6287
Ratio amoA/Total	0.003	0.136

- Archaeal amoA are more than 2x as abundant as Bacterial.
- The proportion of amoA/Total is considerably larger!

Diversity and phylogeny of amoA?

The amoA sequences obtained do not align with the reference amoA sequences. They are very divergent and when I blast them individually they do not match any reference or environmental amoA sequences.



Conclusions

- Short-read sequences from SOLiD sequencing technology can be used to characterize complex microbial communities.
- MG-RAST is an efficient web-server for analyzing the taxonomic and functional diversity within a metagenome.
- We need more cultured representatives in order to correctly annotate metagenomes
 - Or use metagenomes to assemble representatives (Iverson, V. et al. *Science*. 335, 587-590. (2012)).