

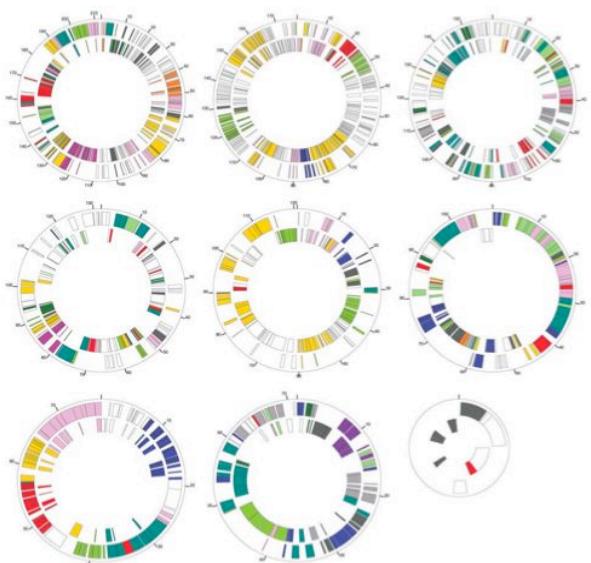
Metatranscriptomics: Measuring microbial gene expression in the ocean

Gene Tyson
DeLong Lab
gtyson@mit.edu

Agouron Summer Course 2008

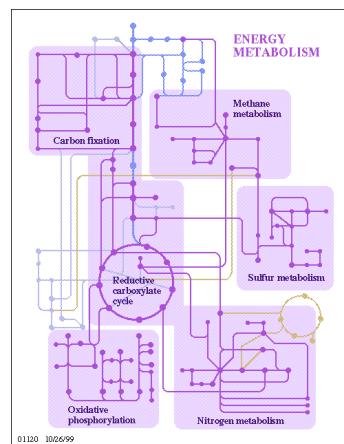
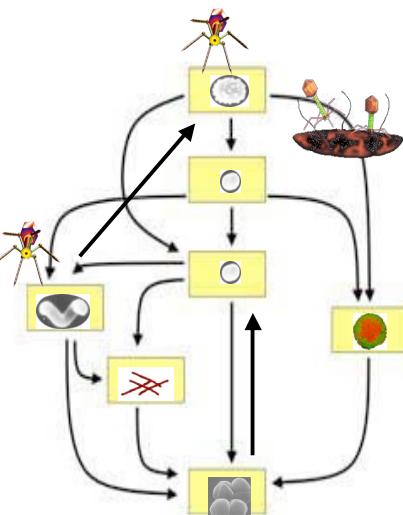
GENOMES

Community genomic (metagenomic) and transcriptomic data



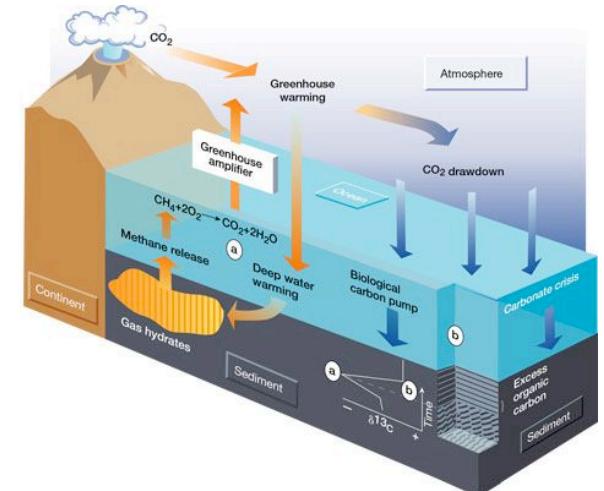
BIOMES

Community composition and interactions



Community metabolism

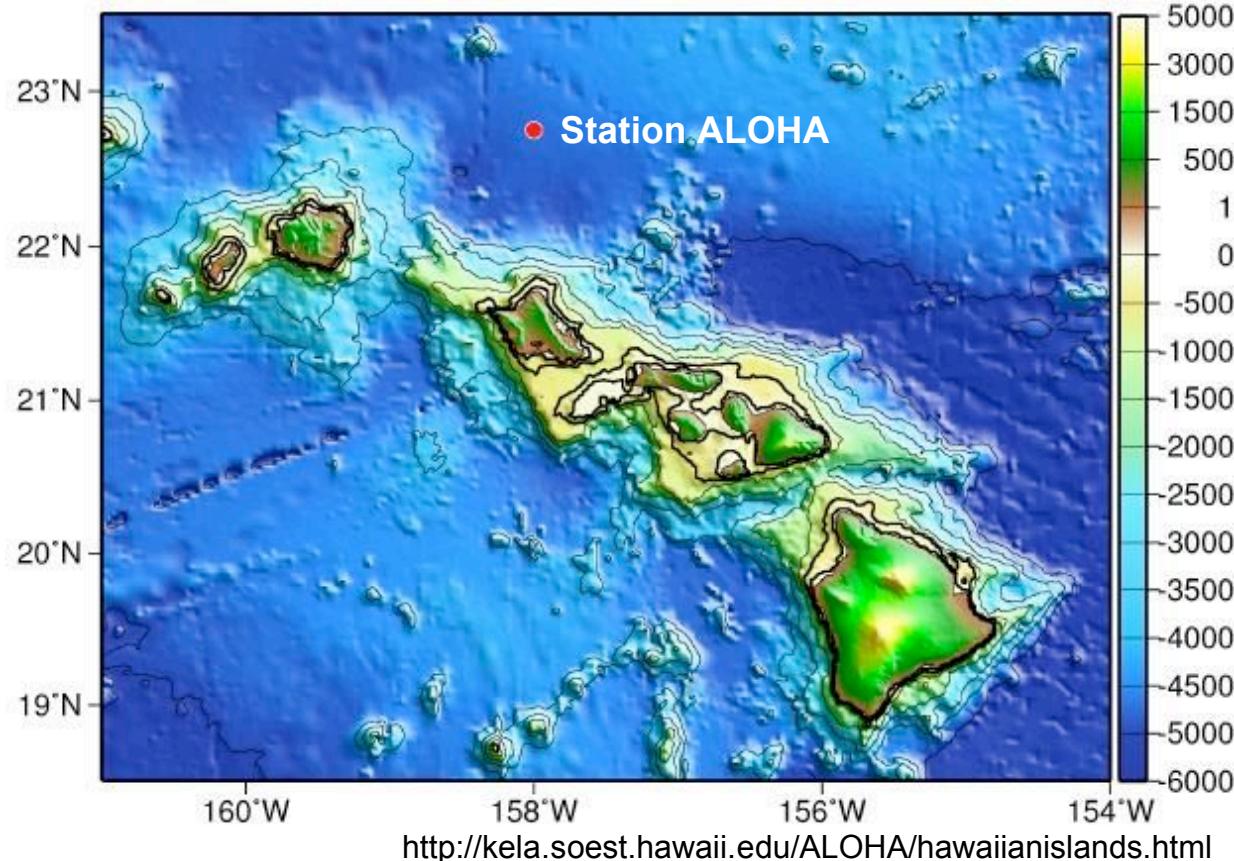
Ecosystem functions

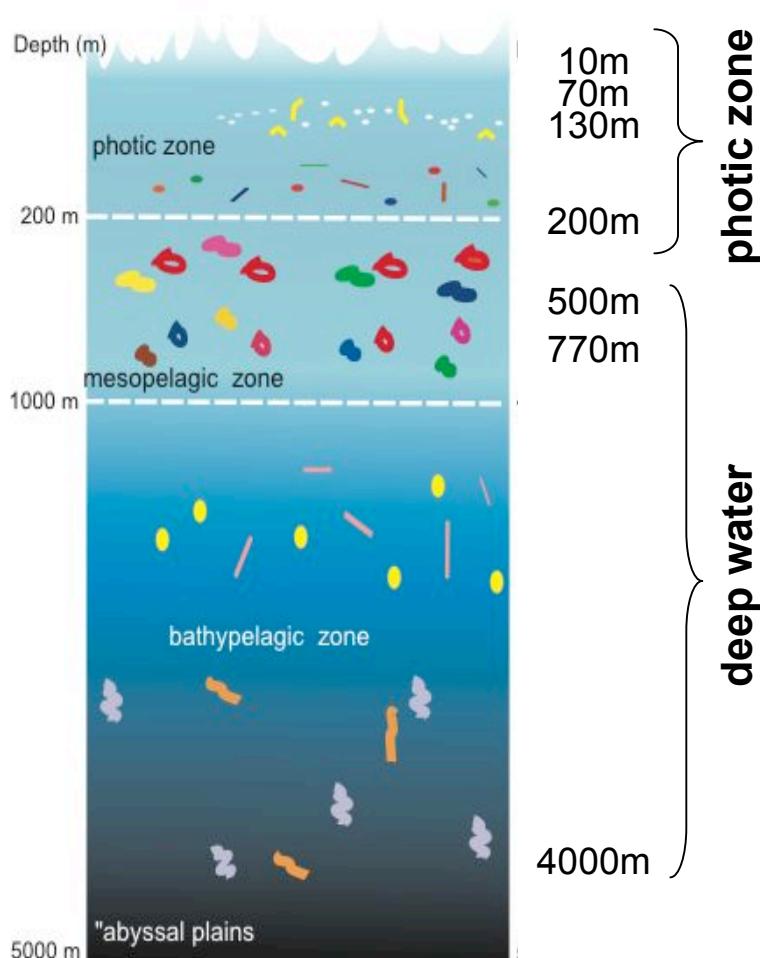


Station ALOHA - North Pacific Subtropical Gyre

Hawaii Ocean Time-series (HOT) program begun in October 1988

- monthly cruises to measure hydrography, chemistry and biology of the water column

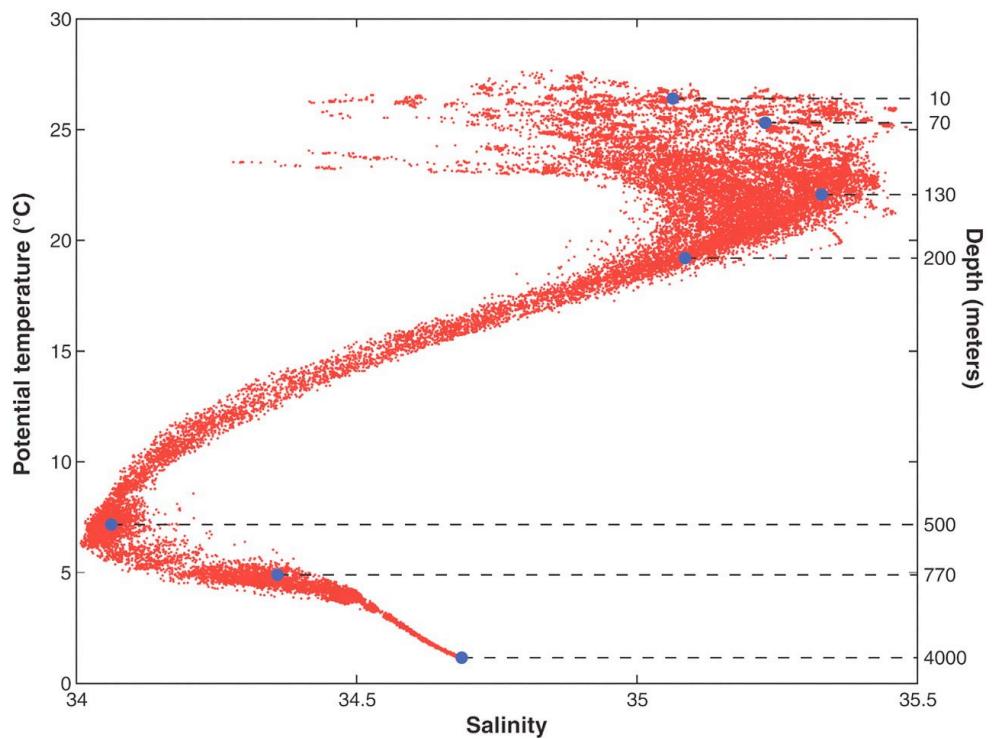


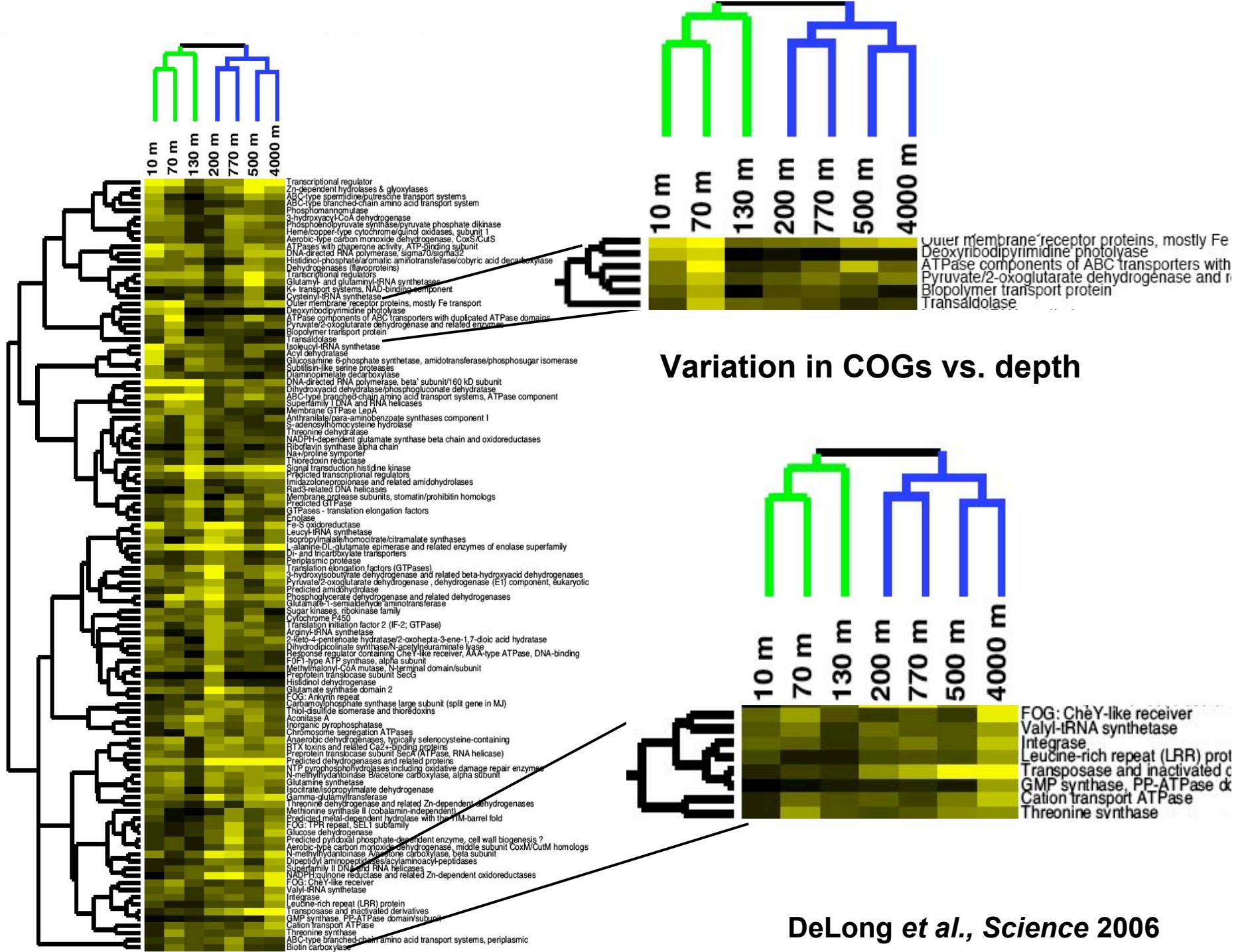


Community Genomics Among Stratified Microbial Assemblages in the Ocean's Interior

Edward F. DeLong,^{1*} Christina M. Preston,² Tracy Mincer,¹ Virginia Rich,¹ Steven J. Hallam,¹ Niels-Ulrik Frigaard,¹ Asuncion Martinez,¹ Matthew B. Sullivan,¹ Robert Edwards,³ Beltran Rodriguez Brito,³ Sallie W. Chisholm,¹ David M. Karl⁴

SCIENCE VOL 311 27 JANUARY 2006





Measuring gene expression in stratified marine microbial communities

Gene content
Metabolic potential
Genetic variation



Gene expression
Expression dynamics
Regulation

Metatranscriptomic Approach

Sample seawaters from biogeochemically different environments

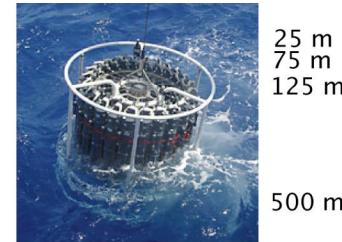
Collect microbial biomass ($\sim 0.2\mu\text{m} - 1.6\mu\text{m}$) from each depth

Extract community RNA and community genomic DNA

Amplify community RNA (preferentially mRNA), synthesis cDNA

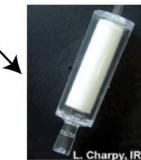
Pyrosequencing

HOT179



25 m
75 m
125 m

500 m



community genomic DNA

$\sim 1-5 \mu\text{g}$

community RNA

$\sim 50-100 \text{ ng}$

$\sim 2-5 \mu\text{g}$

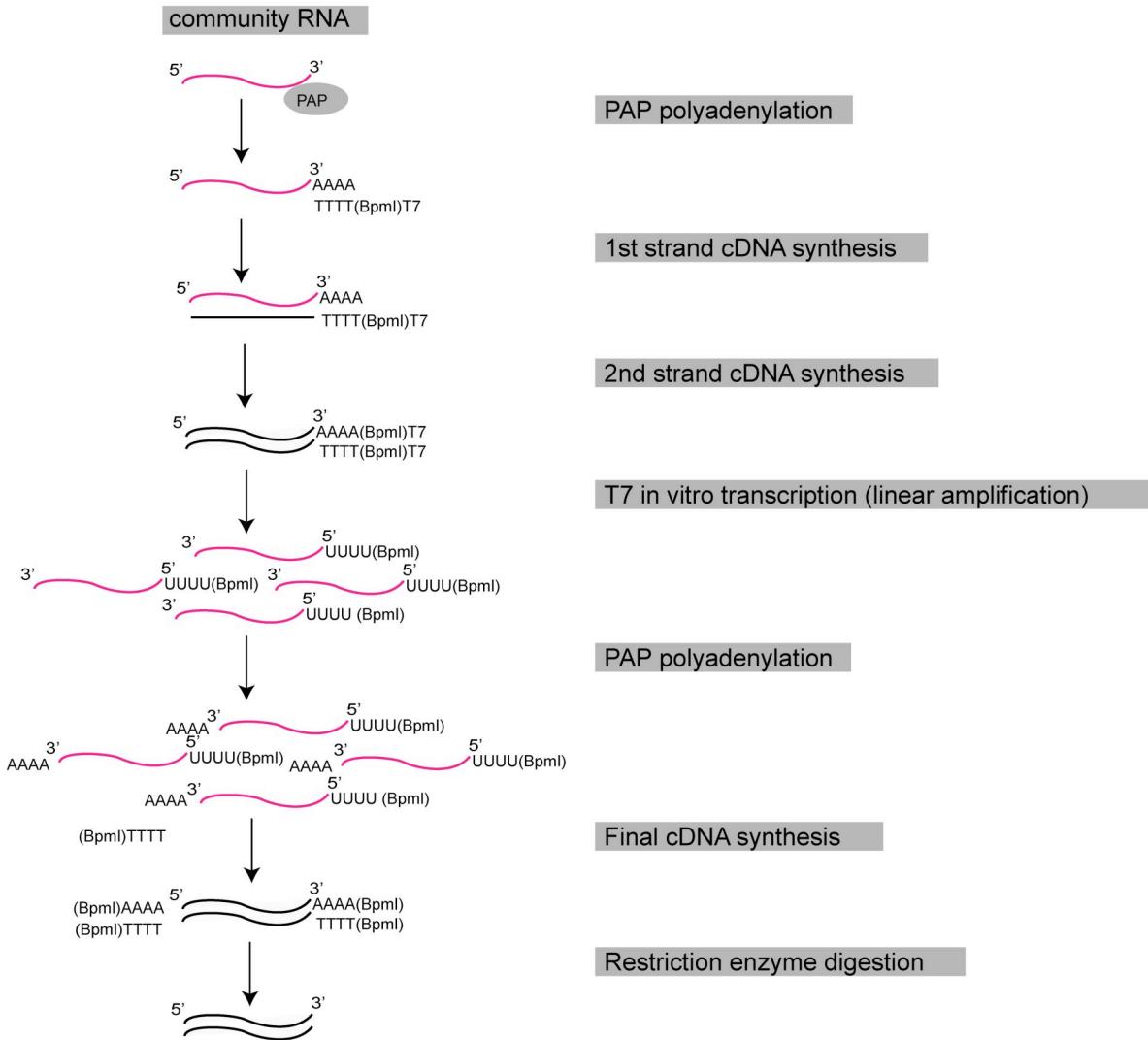
$\sim 12 \text{ Mb}$

$\sim 41 \text{ Mb}$

Coupled cDNA and DNA samples are necessary for generating a normalizing expression signal

RNA amplification

Modified Eberwine T7-based bacterial RNA amplification



Major challenges:

- 1) very short mRNA half life
- 2) low extraction yields
- 3) ribosomal RNAs
- 4) bacterial and archaeal mRNA is not polyadenylated

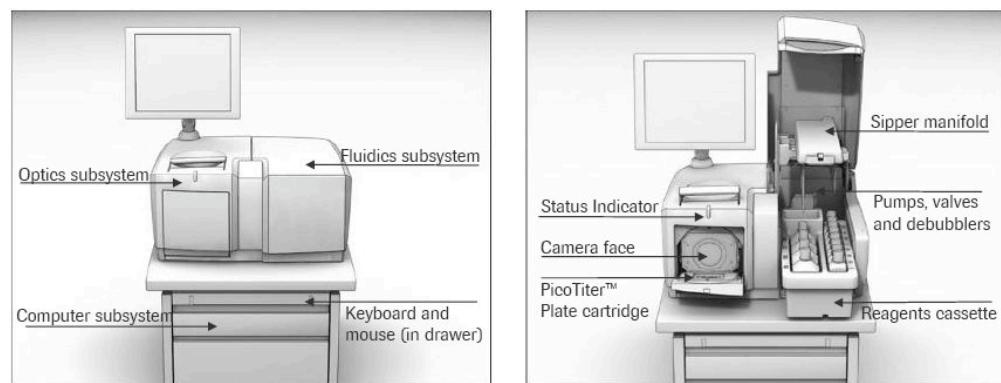
Modified method developed by Jorge
Frias-Lopez and Yanmei Shi

Pyrosequencing - 454 GS20 and FLX



OLD:

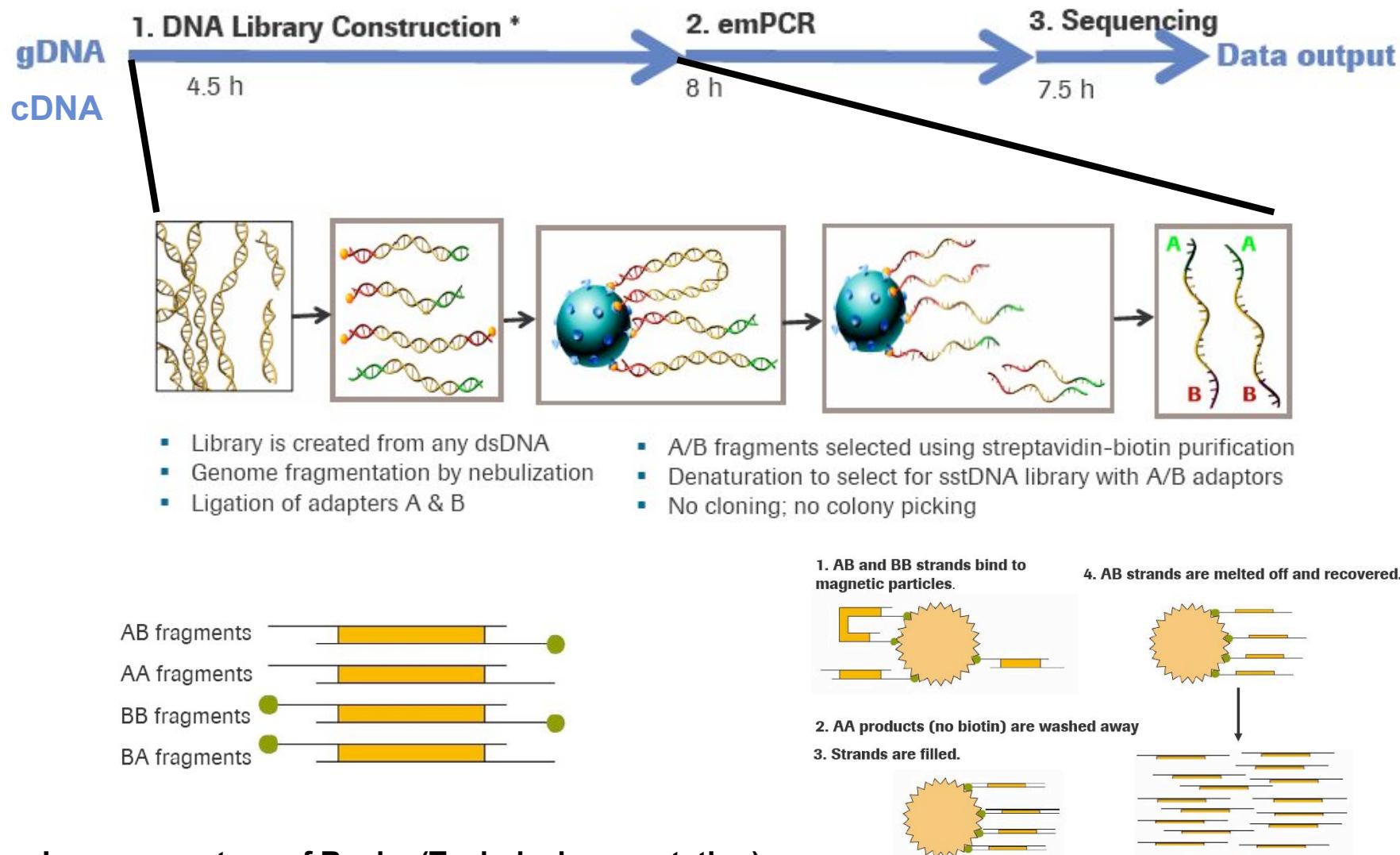
Genome Sequencer GS20
~ 40 million bases
~ >100bp
~ 400,000 reads



NEW:

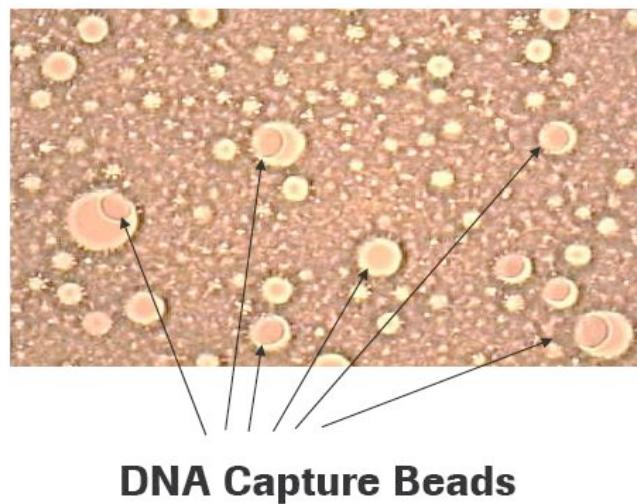
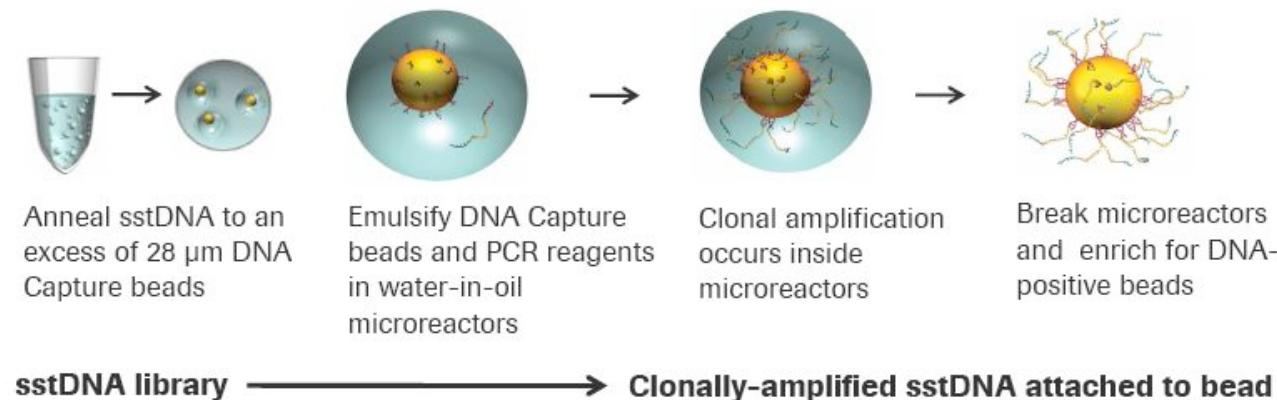
Genome Sequencer FLX
~ 100 million bases
~ >200bp
~ 400,000 reads

Pyrosequencing - Library construction

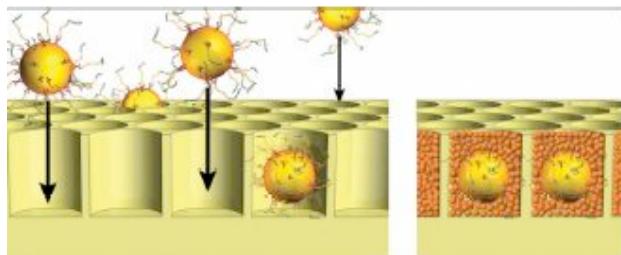


Images courtesy of Roche (Technical presentation)

Pyrosequencing - emPCR

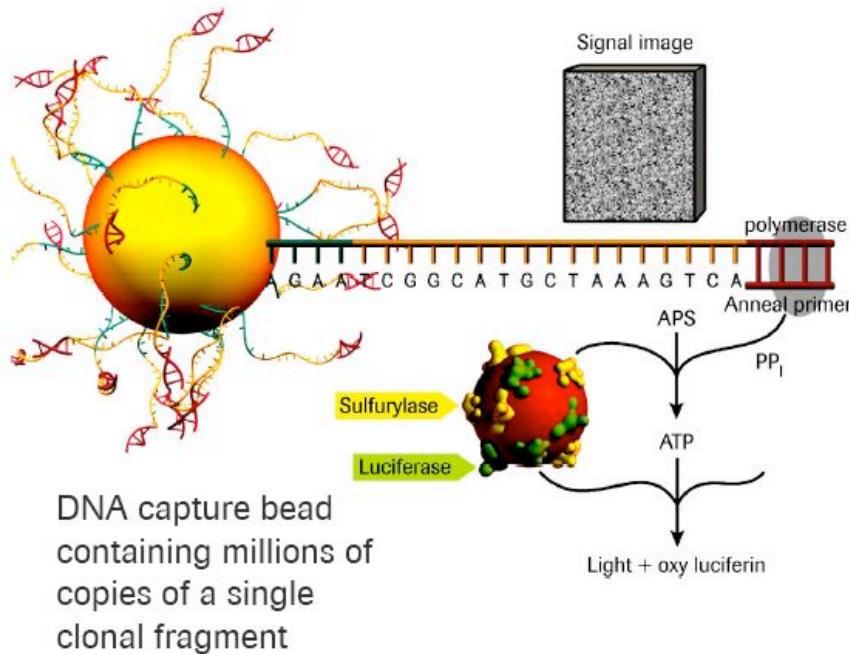


Pyrosequencing - sequencing

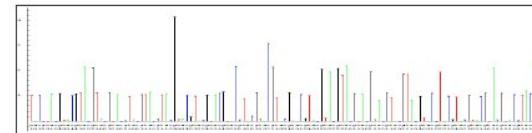
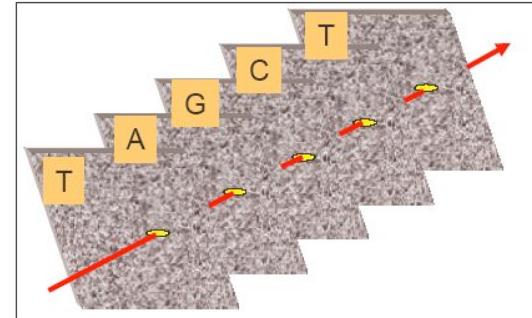
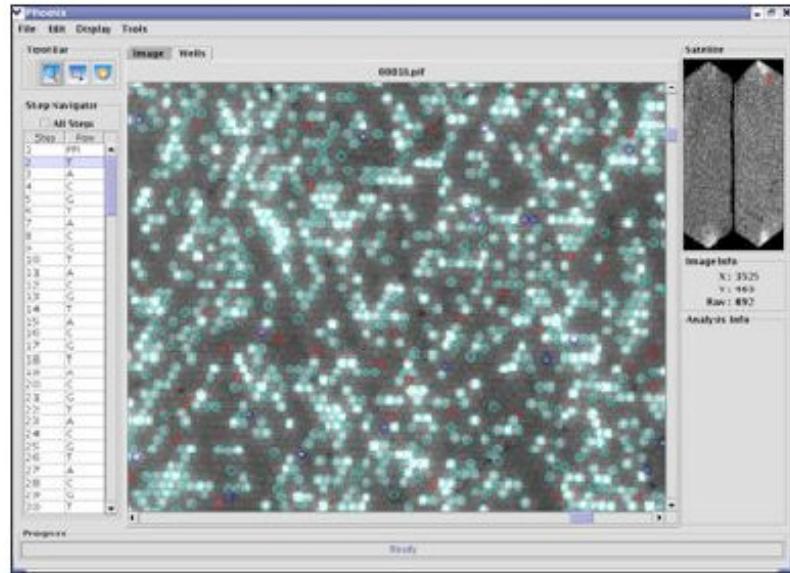


- Well diameter average for PicoTiterPlate is 44 µm
- A single clonally amplified sstDNA bead is deposited per well.
- A layer of packing and enzyme beads are deposited
- Plate is loaded into instrument for sequencing

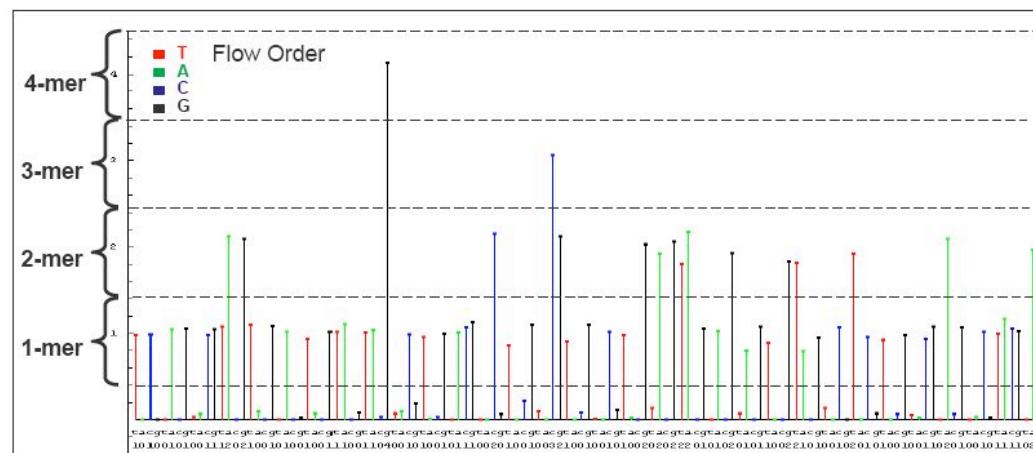
Amplified sstDNA library beads → **Packed PTP**



Pyrosequencing - imagine processing



Signal strength is determined by homopolymer length.



TCAG

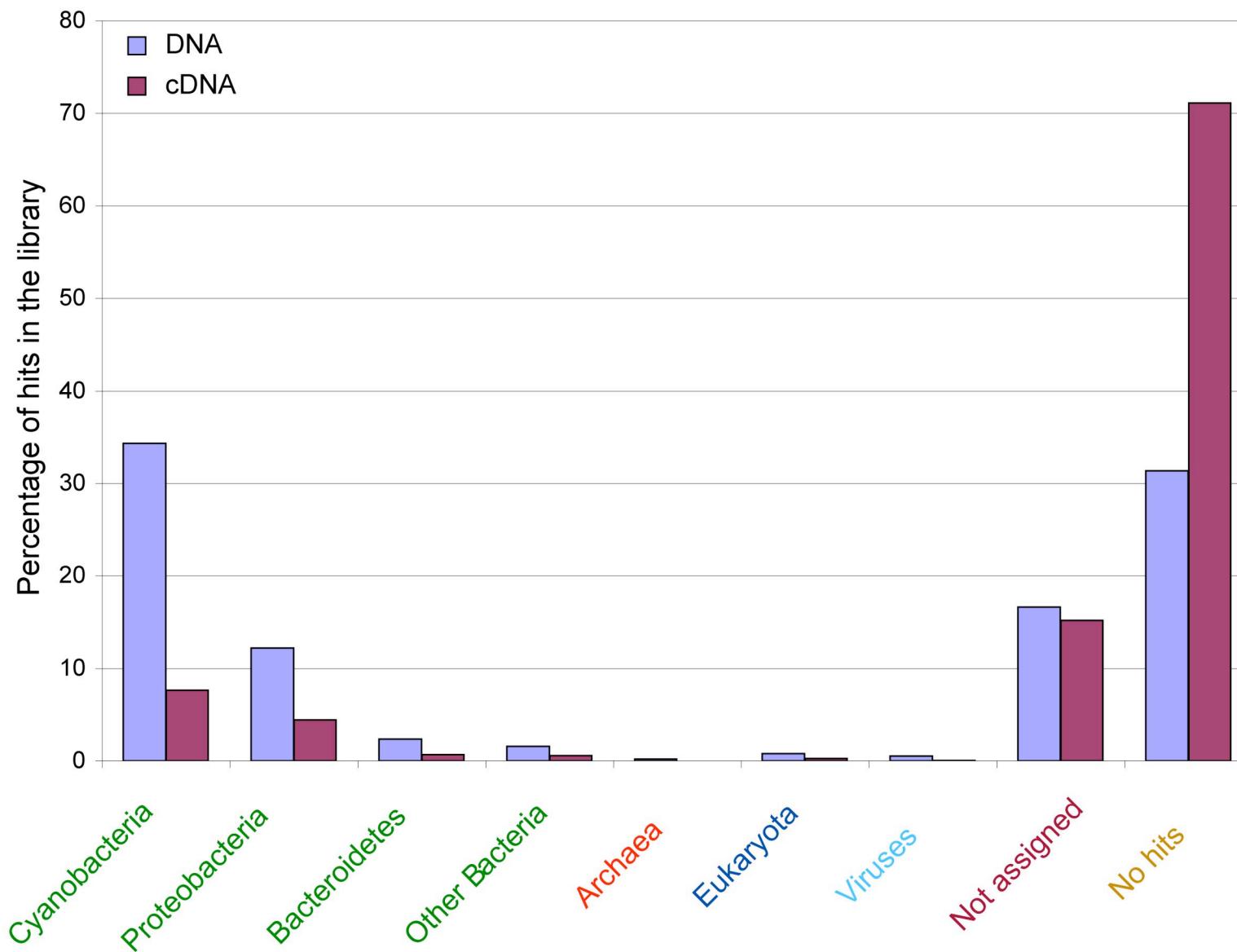
HOT179 - 75m

DNA and cDNA sample statistics

	DNA library	cDNA library
Number of reads	414,323	128,324
Average length (bp)	110	114
Number of rRNA reads	5,877	65,859
Total Base Pairs (Mb)	45.4	14.7

- Smaller than expected number of cDNA reads
 - inefficient removal of polyA tails
 - short cDNA fragments
- rRNA bearing reads constitutes ~50% of the cDNA library

HOT179 75m DNA and cDNA versus NCBI NR database



HOT179 - 75m DNA and cDNA sample stats

	DNA library	cDNA library
Number of reads	414,323	128,324
Average length (bp)	110	114
Number of rRNA reads	5,877	65,859
Total Base Pairs (Mb)	45.4	14.7
Number of NCBI-nr hits	205,747 (50%)	7,275 (13%)
Number of GOS peptide hits	290,741 (70%)	23,203 (43%)

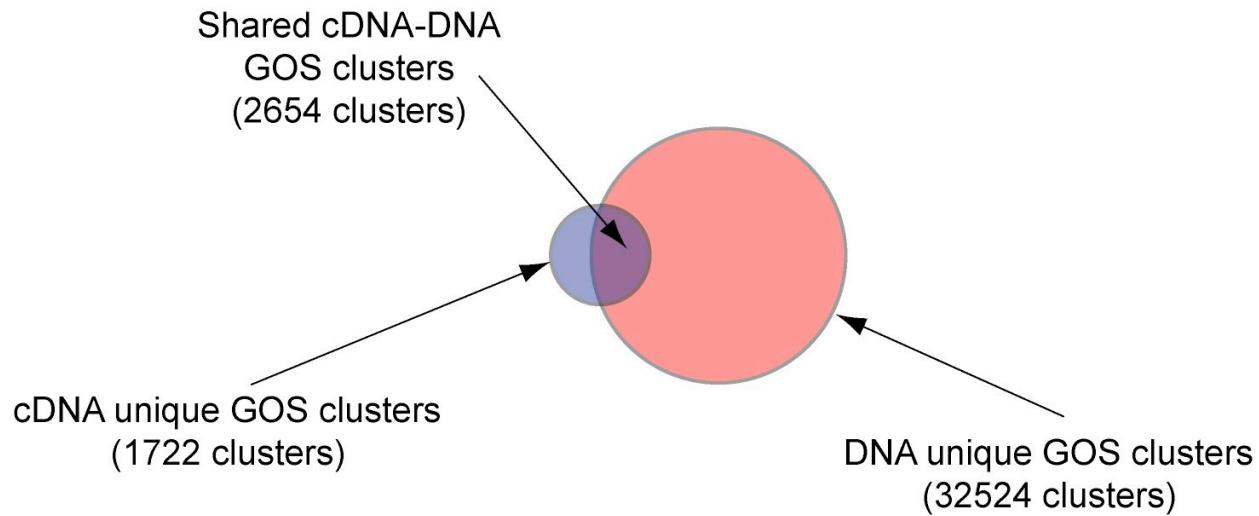
Significantly more hits to the GOS peptides for both DNA and cDNA reads
- increase is greater for the cDNA library

Large number of unknown cDNA reads

- possible presence of novel, rare genes that contribute to microbial expression profile and not detected in previous metagenomic surveys
- regulatory RNAs?

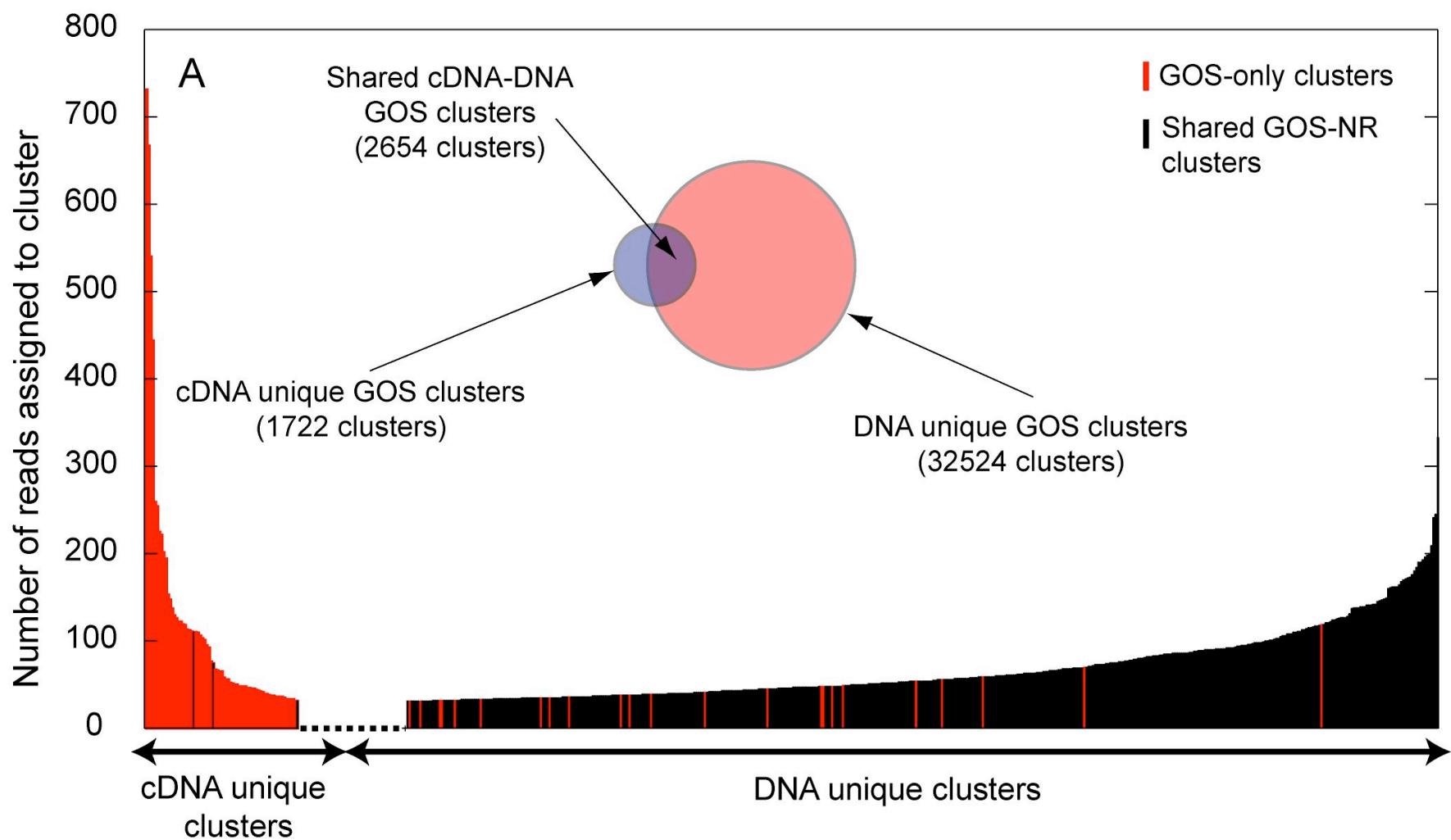
Station ALOHA - HOT179 - 75 m

Metatranscriptomic analysis



Station ALOHA - HOT179 - 75 m

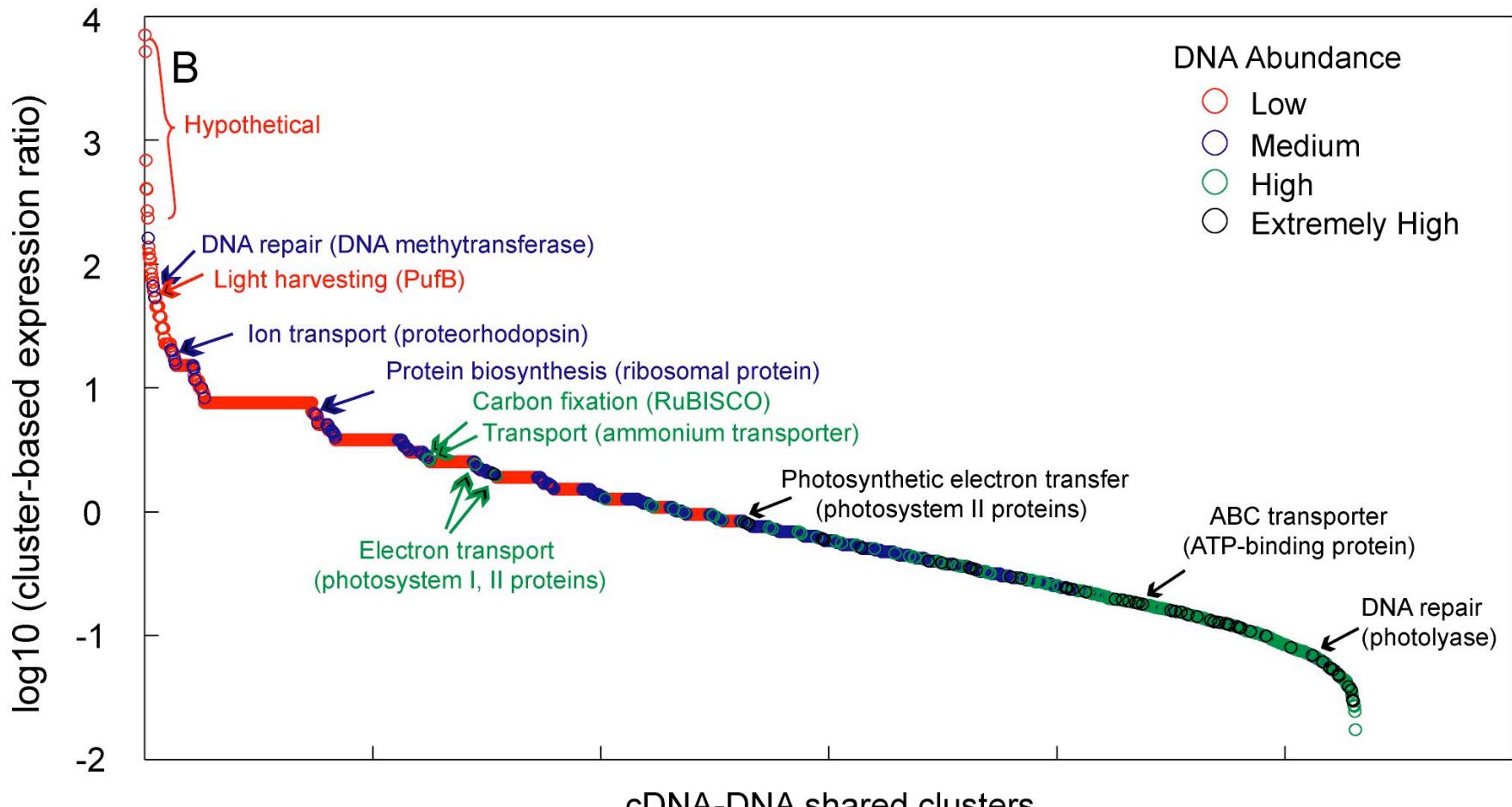
Metatranscriptomic analysis



Community Gene Expression Station ALOHA H179 - 75 m

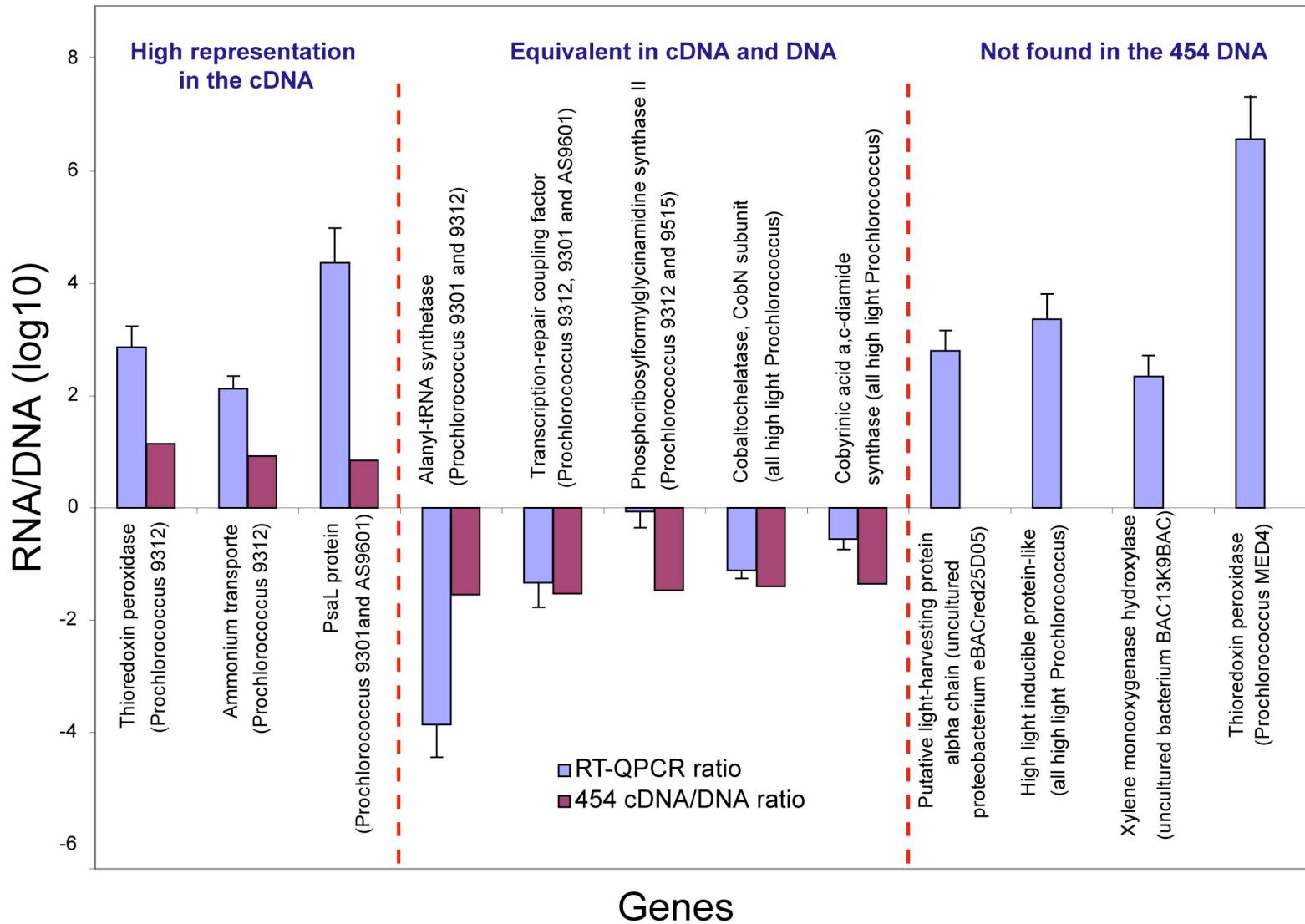
Cluster based expression ratio =

$$\frac{\text{number of cDNA reads mapped to each GOS protein cluster}}{\text{number of DNA reads mapped to each GOS protein cluster}}$$



Back to the Environment

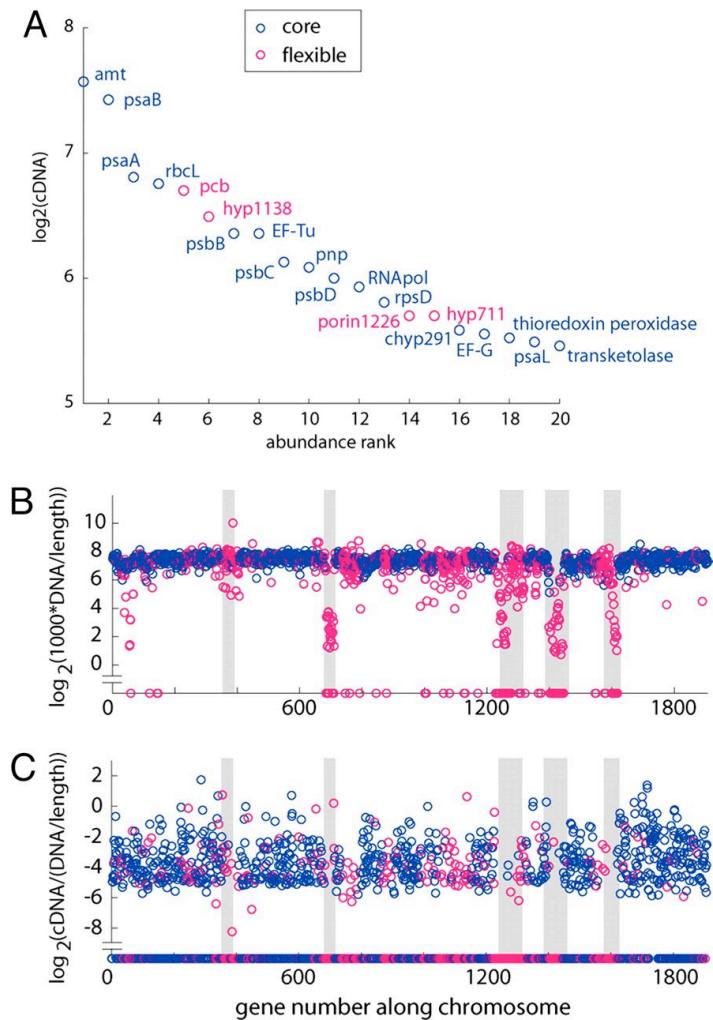
qRT-PCR/qPCR ratios



Jorge Frias-Lopez

Gene expression in *Prochlorococcus*

Station ALOHA - H179 - 75m



~ 90% of the putative *Prochlorococcus* reads mapped to MIT9301, AS9601 and MIT9312 - all high light species.

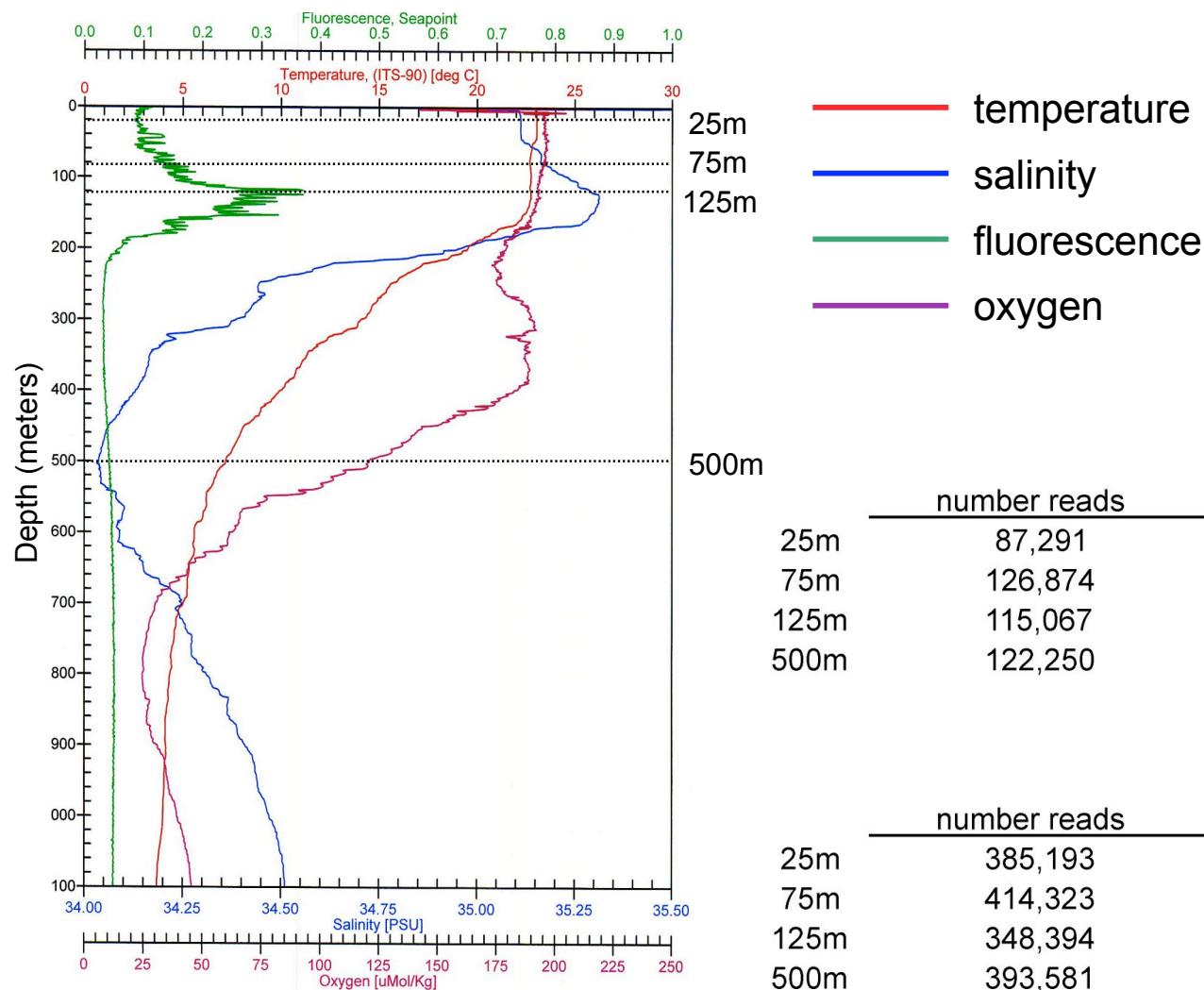
Highly expressed genes involved ammonium uptake (*amtB*), carbon fixation (*rbcL*) and photosynthesis (*psaA,B*).

DNA - all 'core' genes expression at relatively even abundance, 'flexible' genes variable.

cDNA - some 'flexible' genes highly expressed.

Prochlorococcus gene and transcript abundance using strain MIT9301 as a reference genome

HOT179 depth profile - cDNA and DNA read statistics

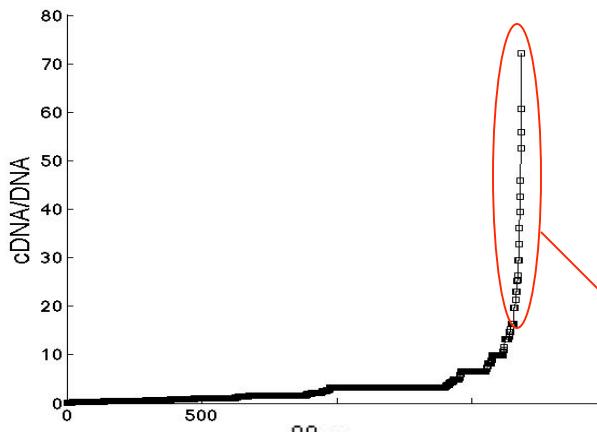


cDNA

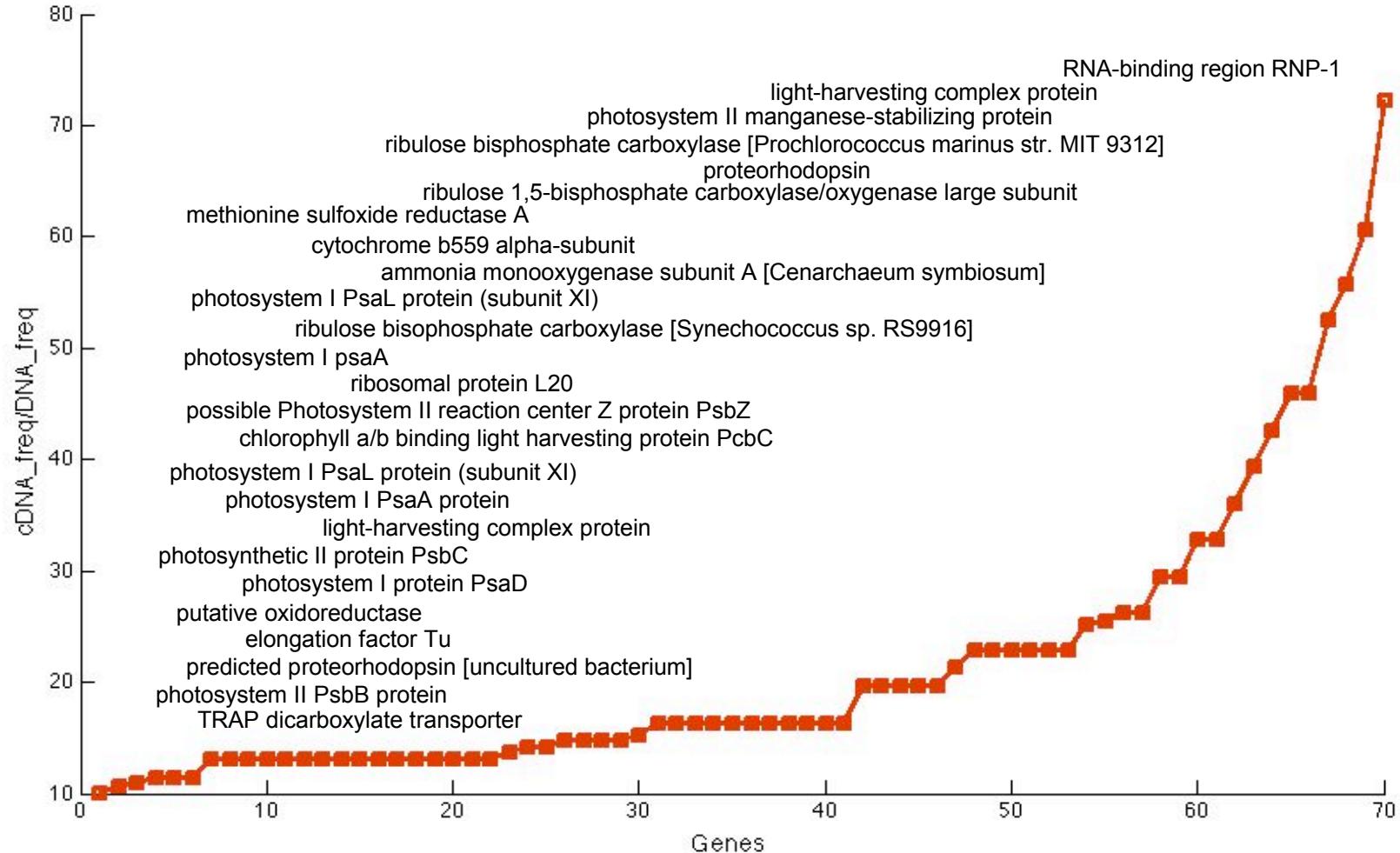
	number reads	average length	total
25m	87,291	105bp	9.2 Mb
75m	126,874	114bp	14.5Mb
125m	115,067	104bp	12.0Mb
500m	122,250	102bp	12.5Mb

DNA

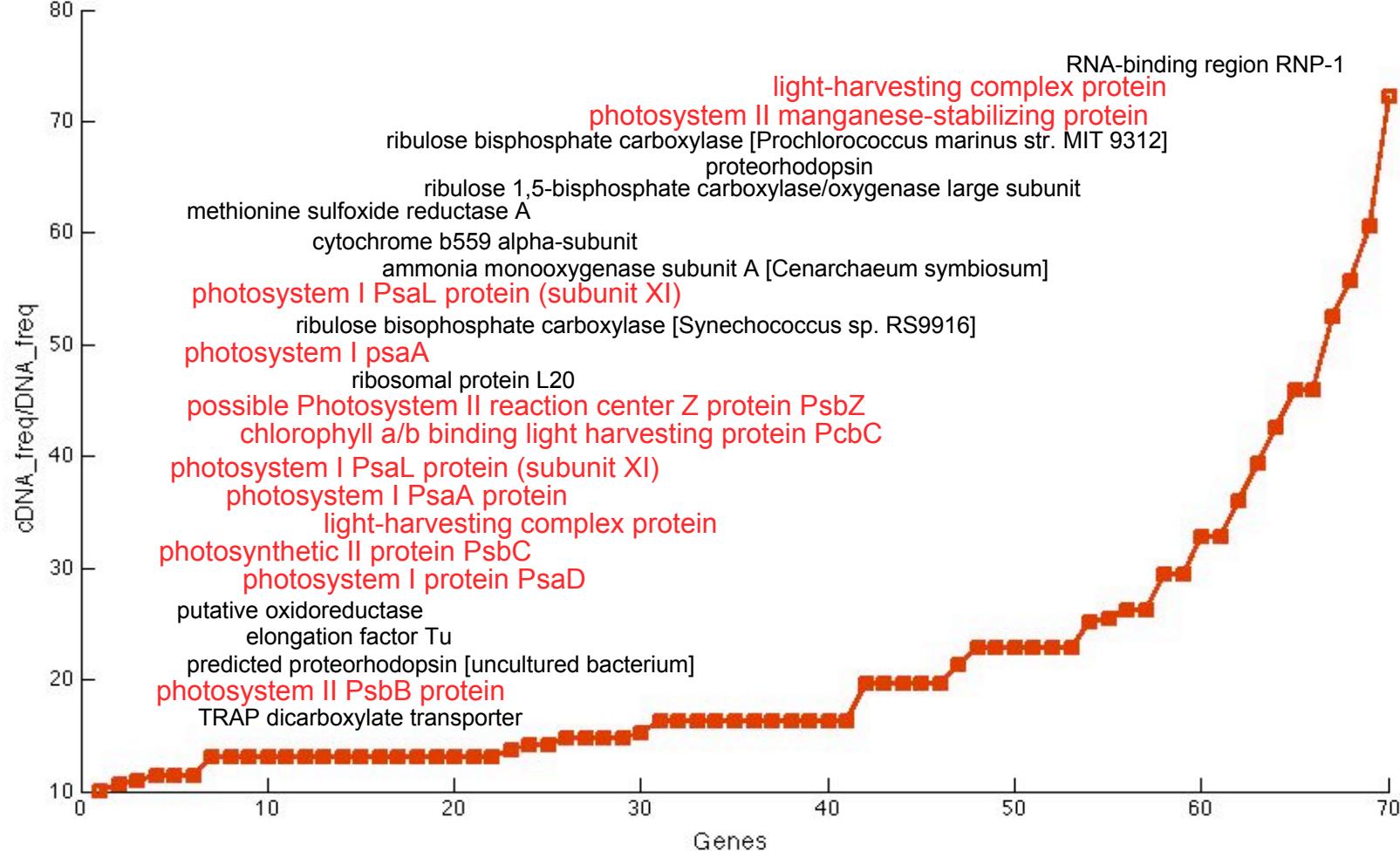
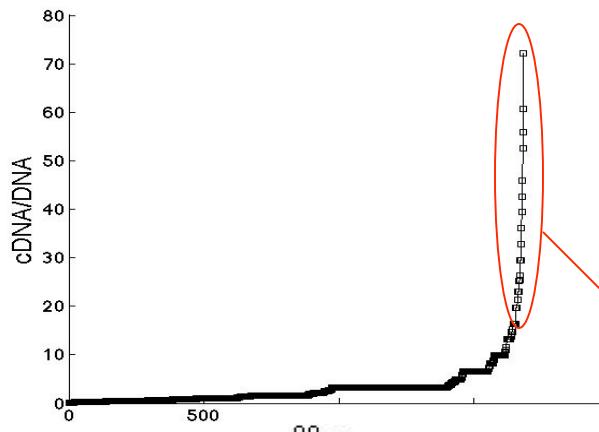
	number reads	average length	total
25m	385,193	108bp	41.6Mb
75m	414,323	110bp	45.6Mb
125m	348,394	108bp	37.6Mb
500m	393,581	106bp	41.7Mb



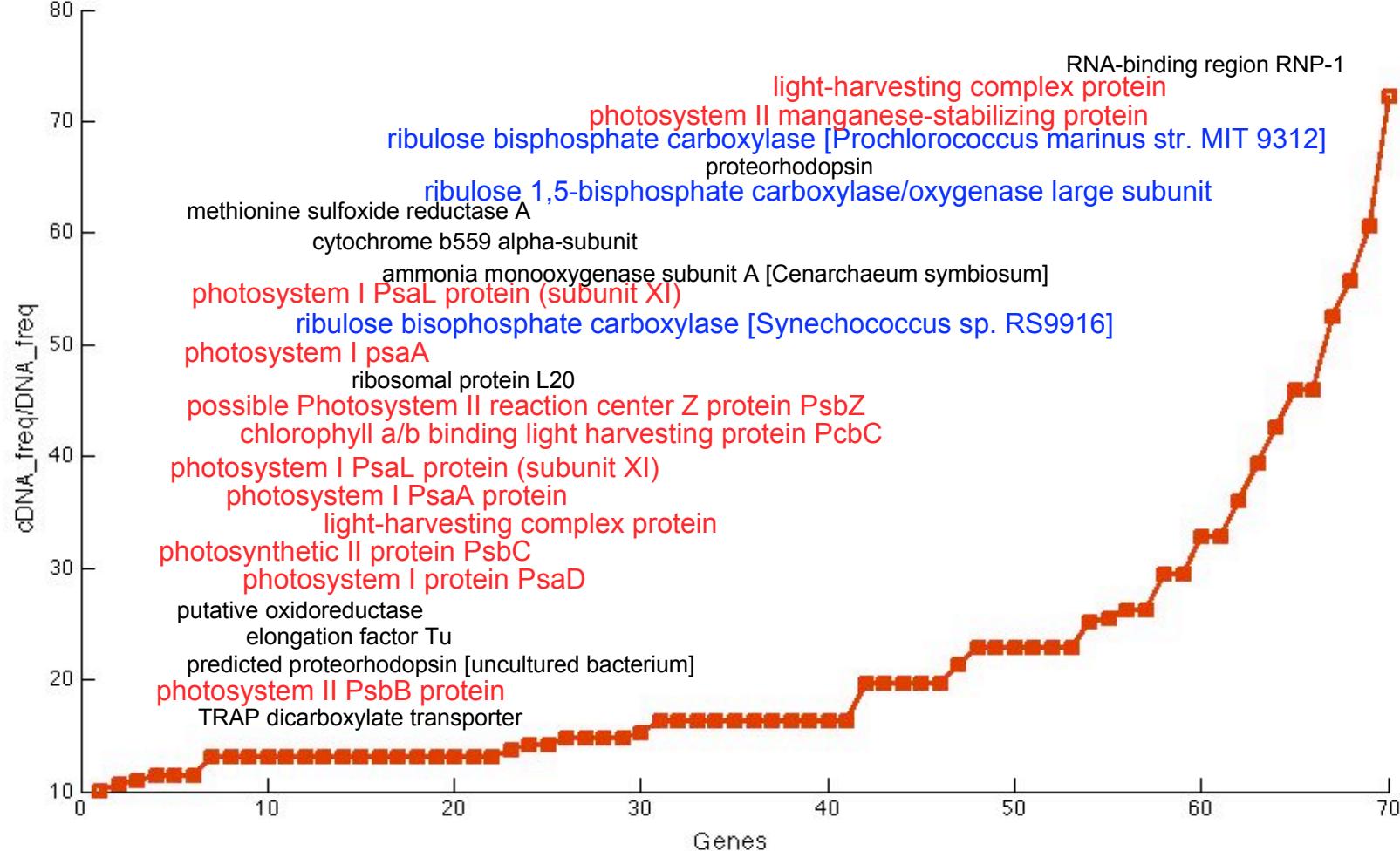
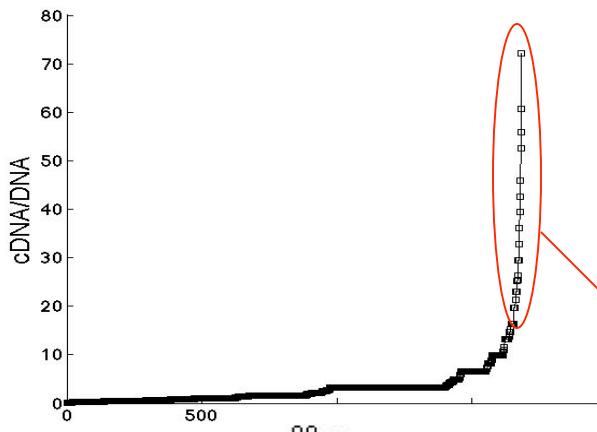
HOT179 125m highly expressed genes - based on NCBI gi number (hypotheticals removed)



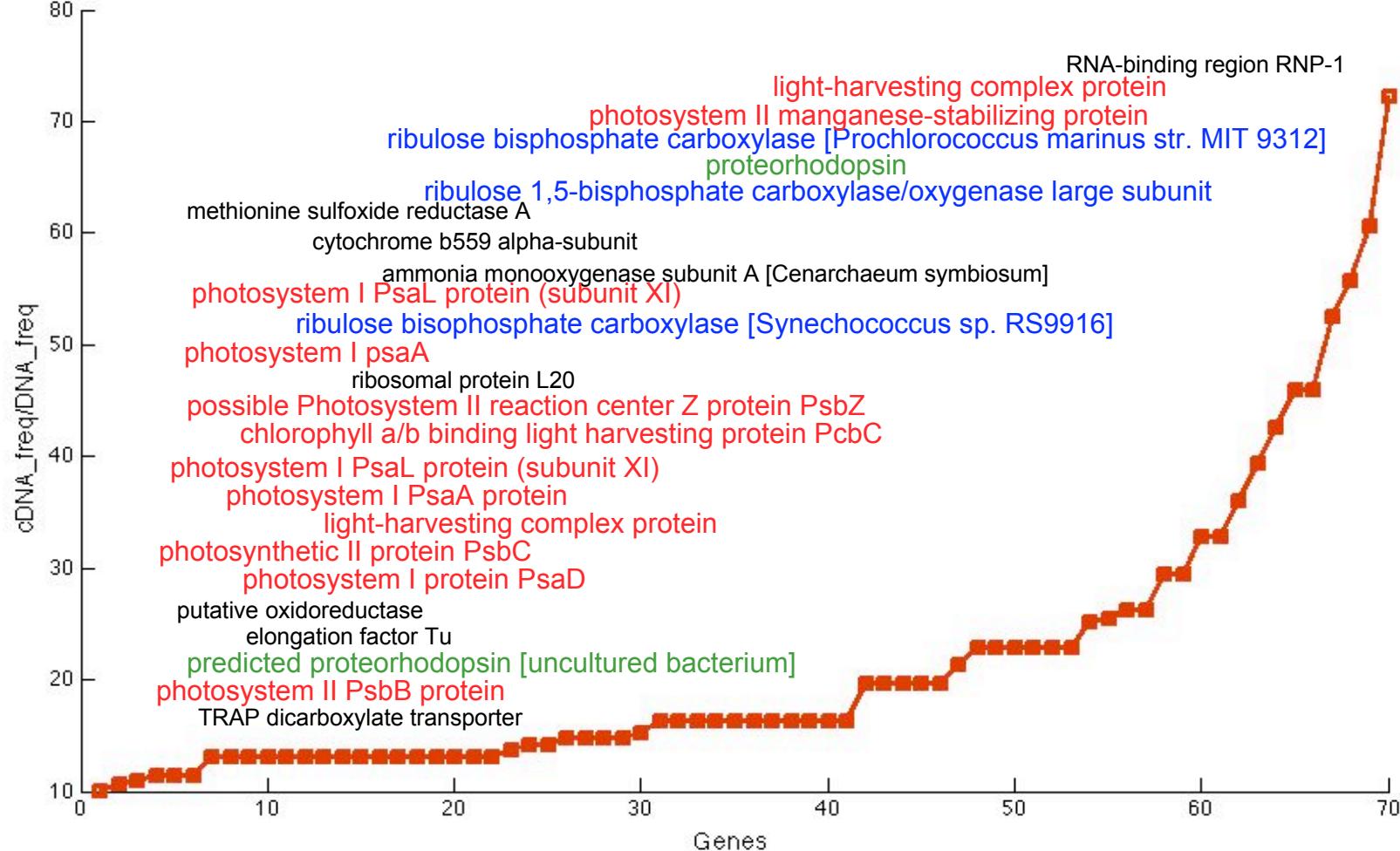
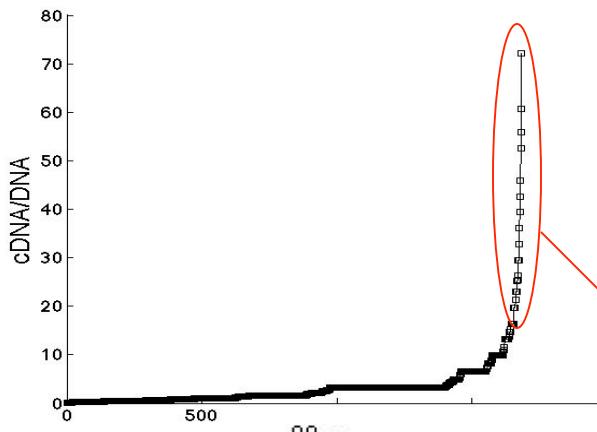
HOT179 125m highly expressed genes - based on NCBI gi number (hypotheticals removed)

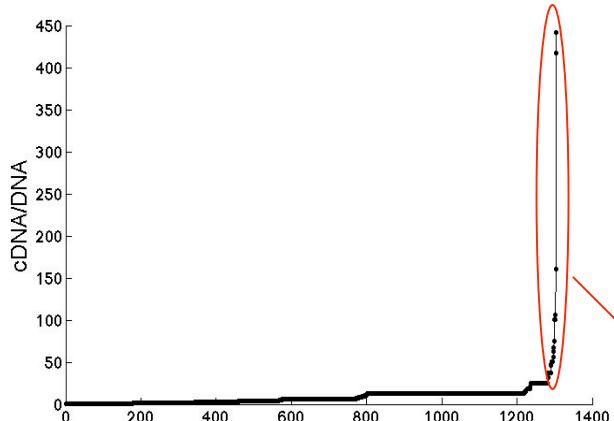


HOT179 125m highly expressed genes - based on NCBI gi number (hypotheticals removed)

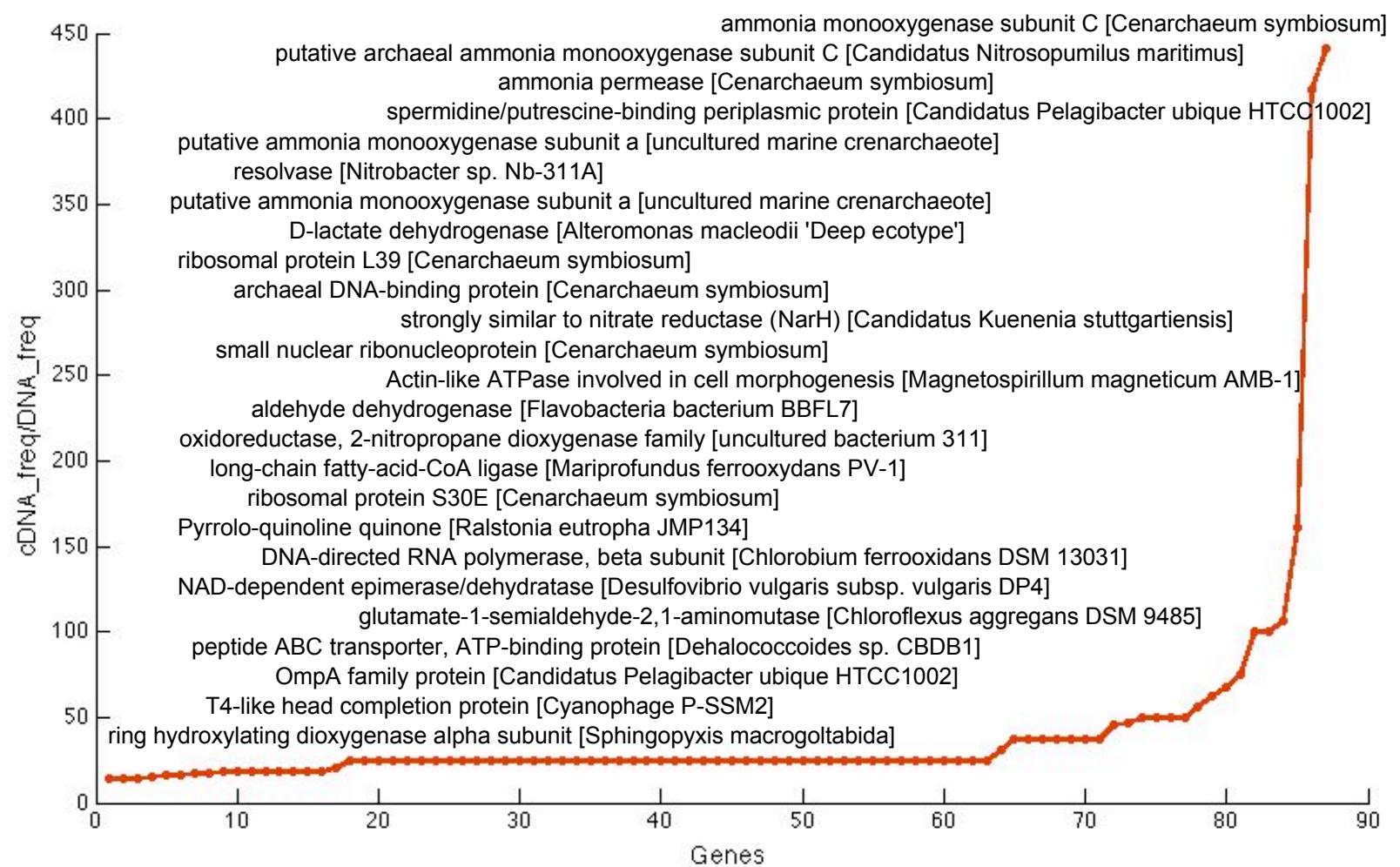


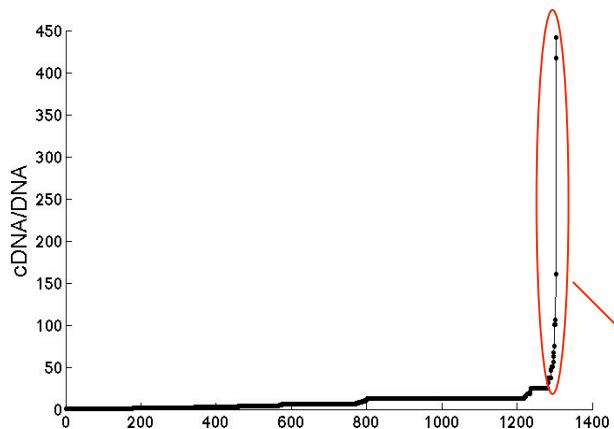
HOT179 125m highly expressed genes - based on NCBI gi number (hypotheticals removed)



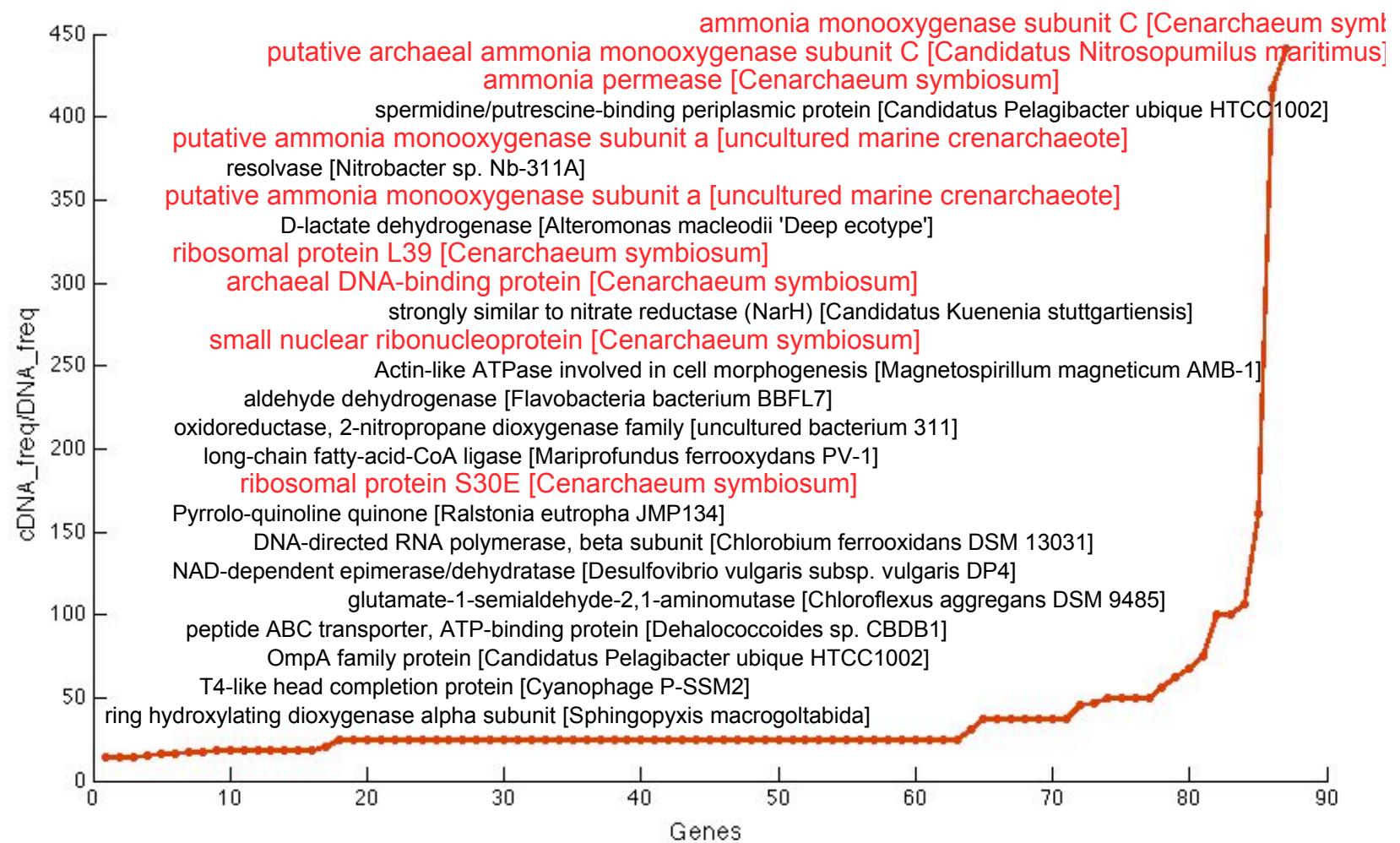


HOT179 500m highly expressed genes - based on NCBI gi number (hypotheticals removed)

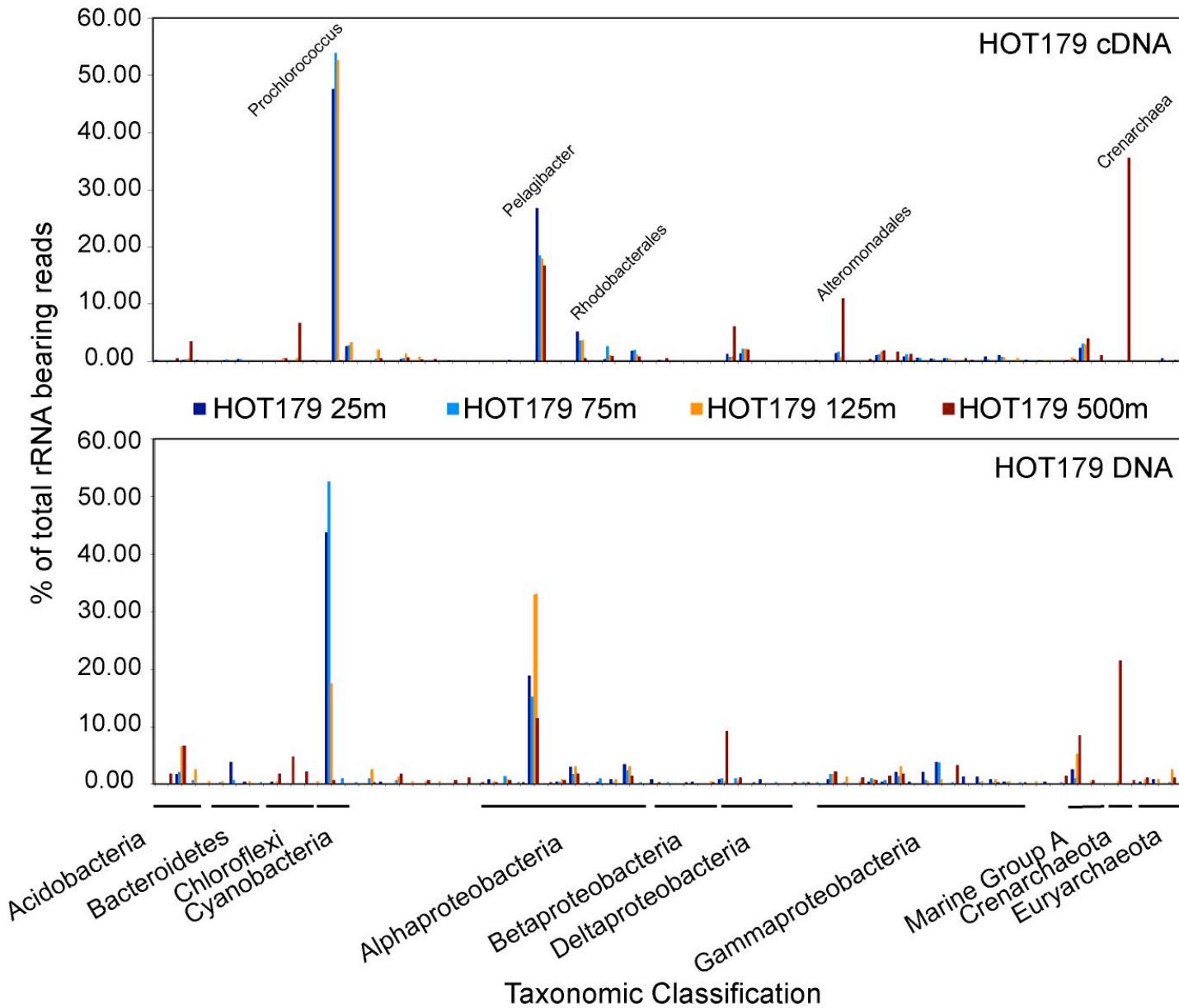




HOT179 500m highly expressed genes - based on NCBI gi number (hypotheticals removed)

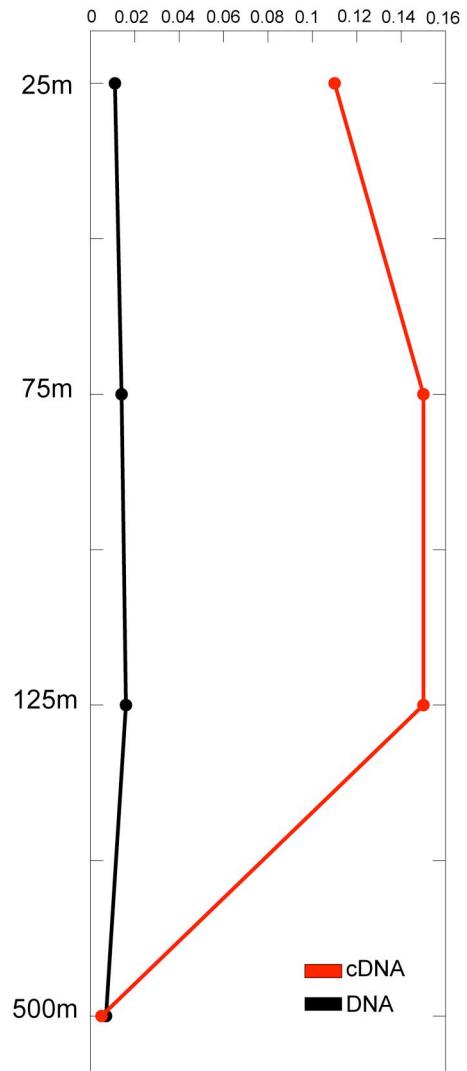


Taxonomic profile - DNA vs cDNA

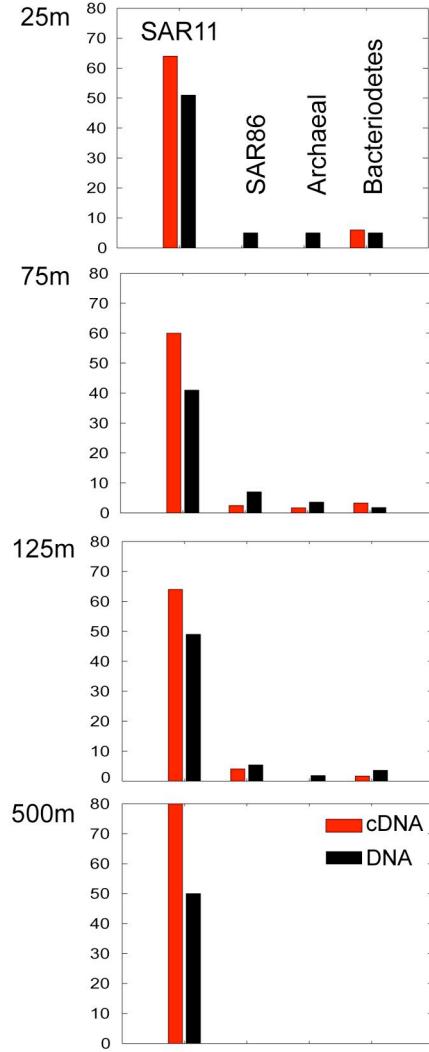


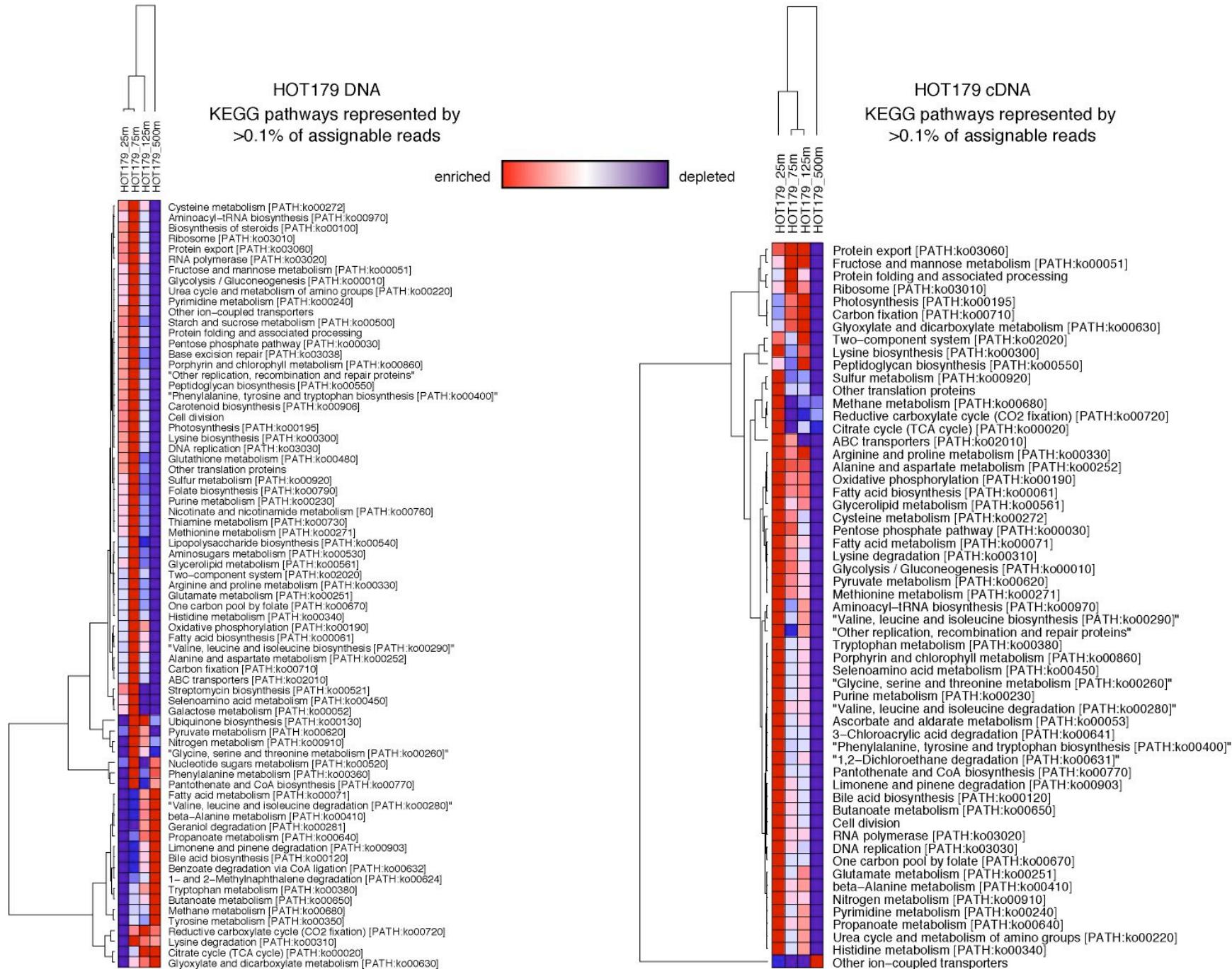
Proteorhodopsin expression through the depth profile

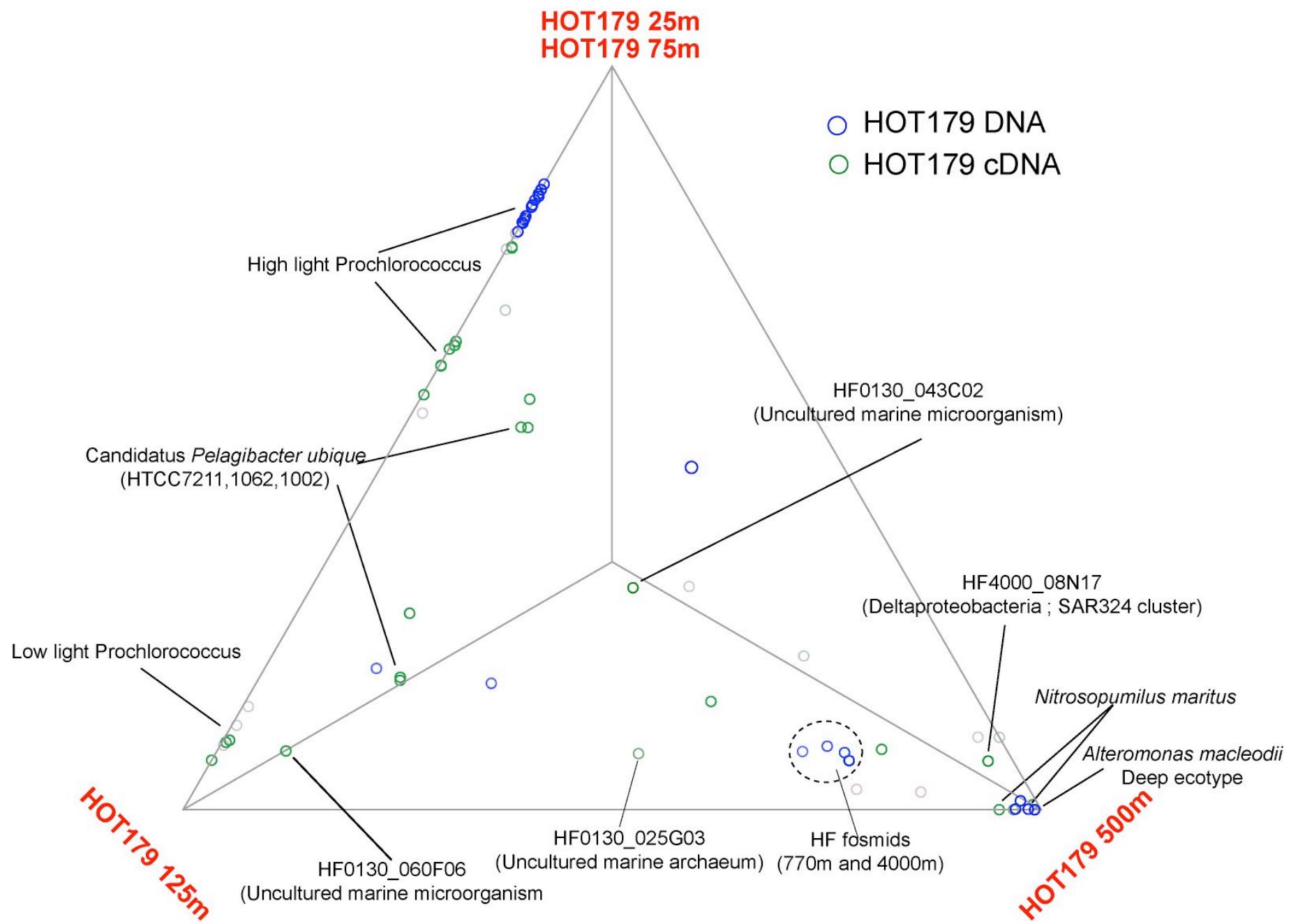
PR-coding reads relative to size of the library (%)



Taxonomic affiliation of PRs (%)

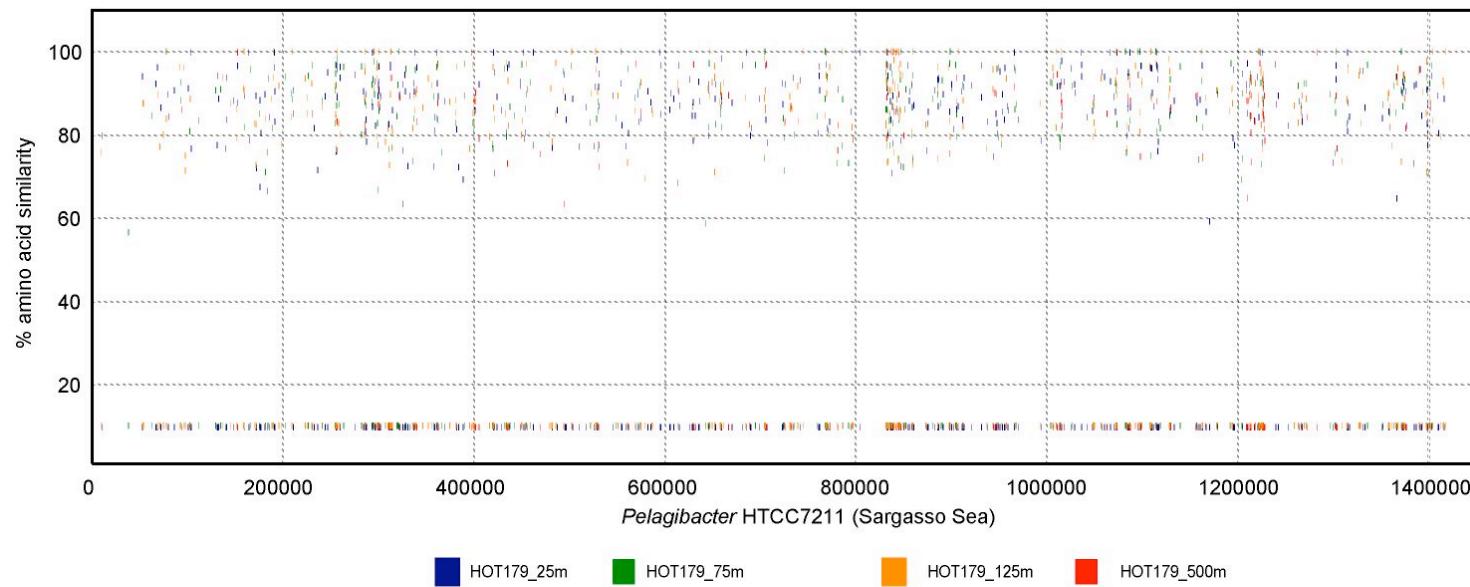




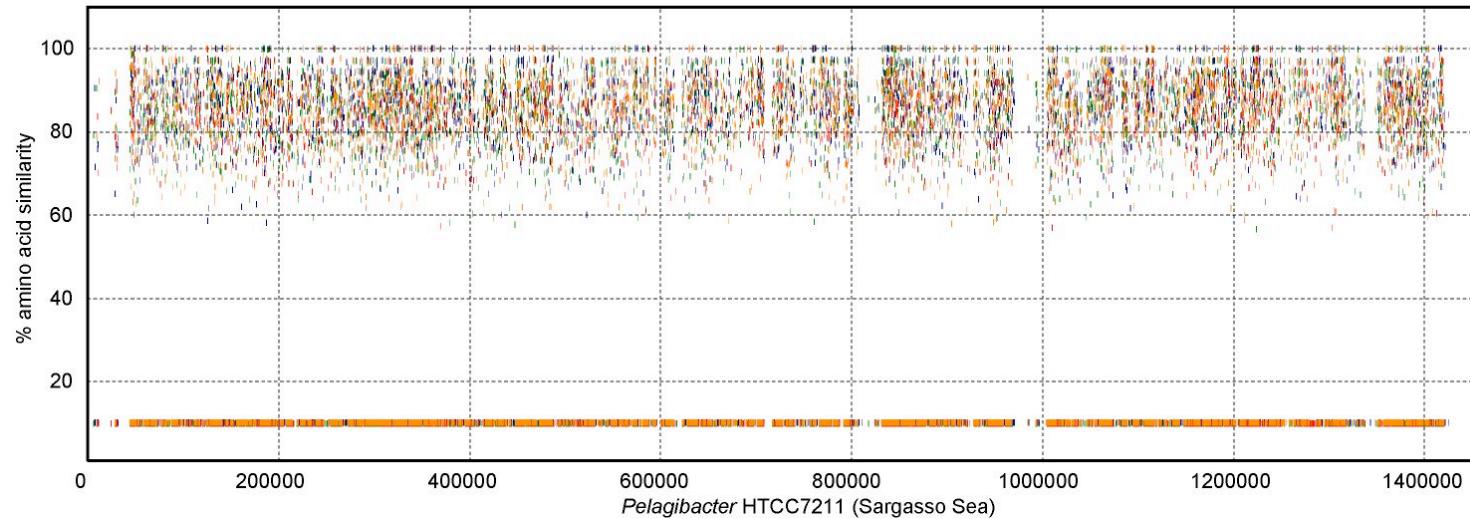


Pelagibacter HTCC7211 Coverage and Gene Expression

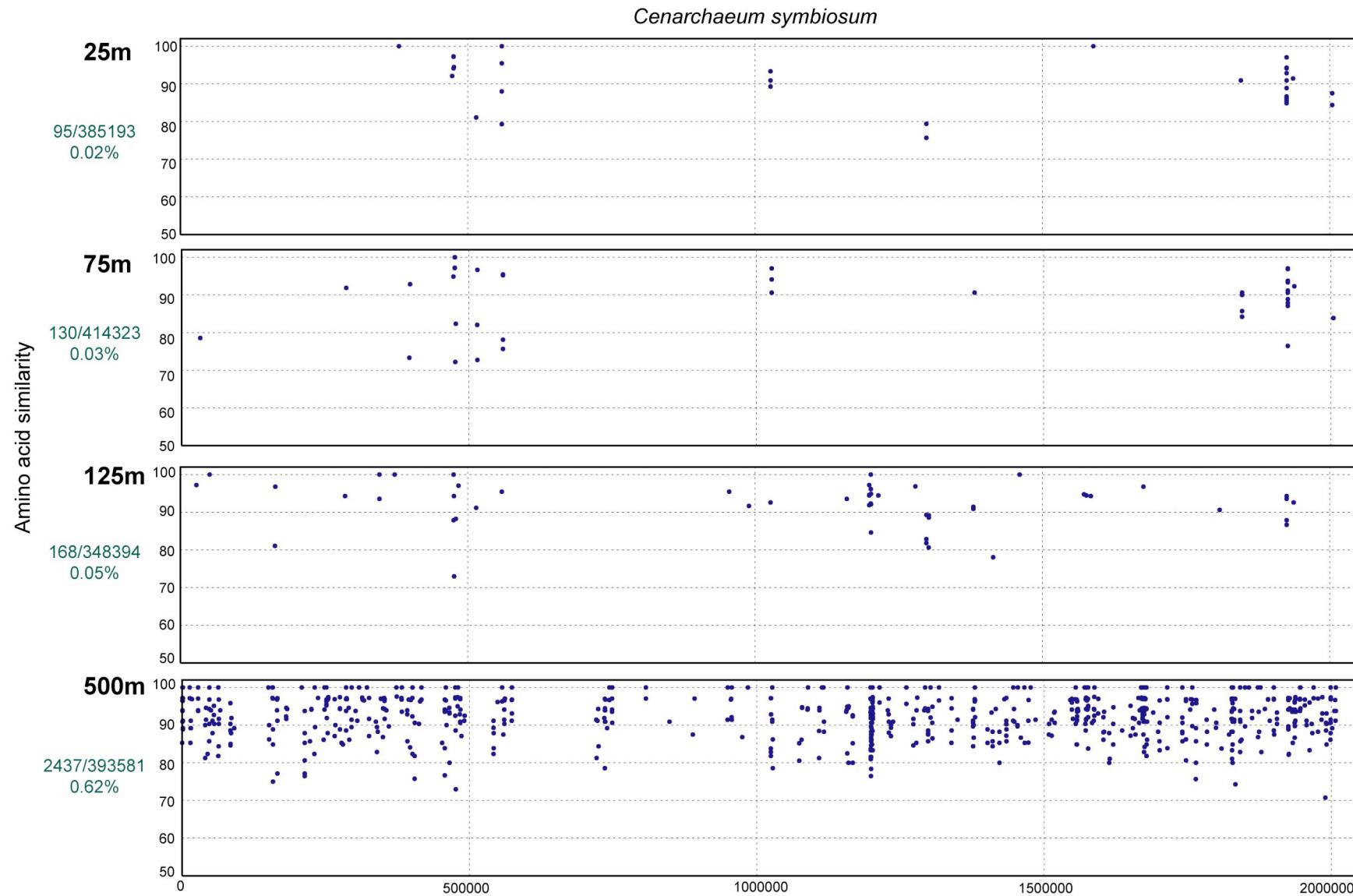
HOT179 cDNA



HOT179 DNA

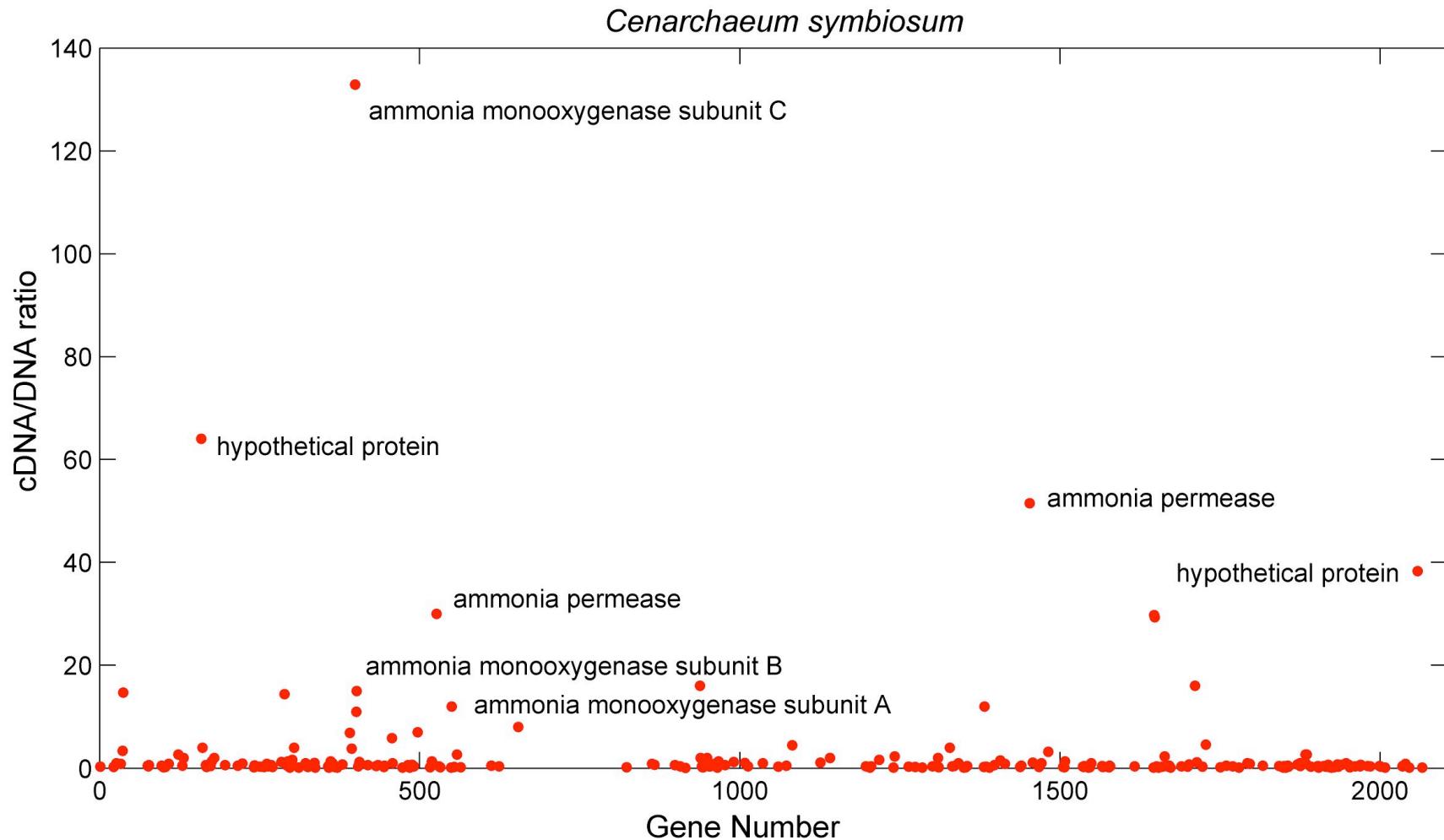


Cenarchaeum symbiosum coverage - depth profile

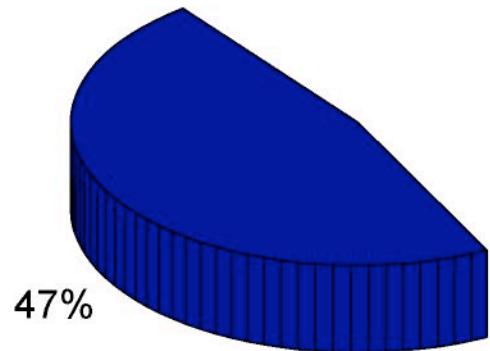


Cenarchaeum symbiosum cDNA/DNA ratio

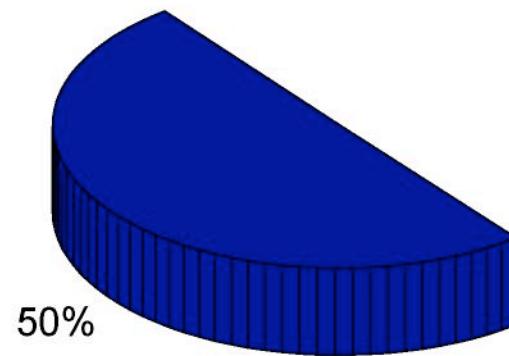
Station ALOHA - H179 - 500m



HOT179 25m

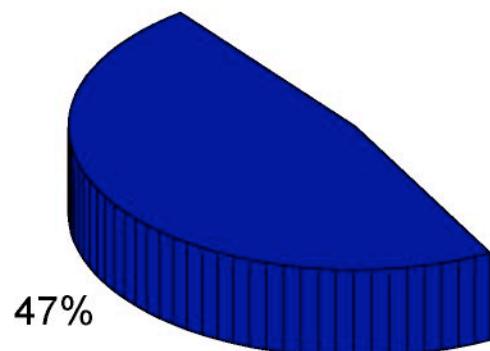


HOT179 75m

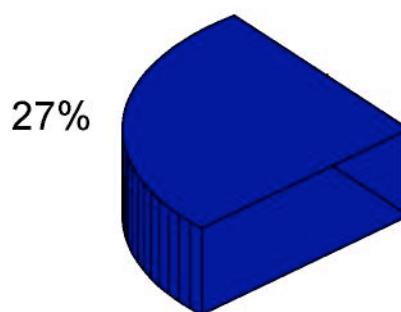


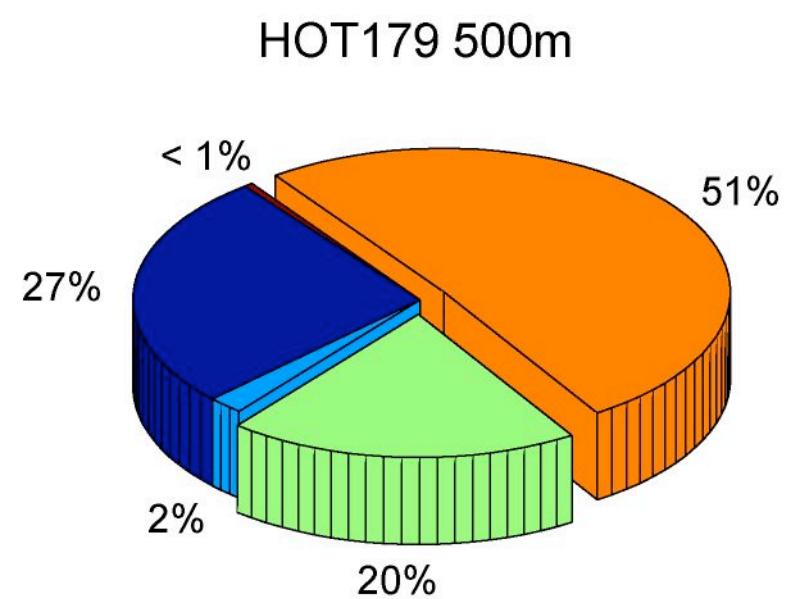
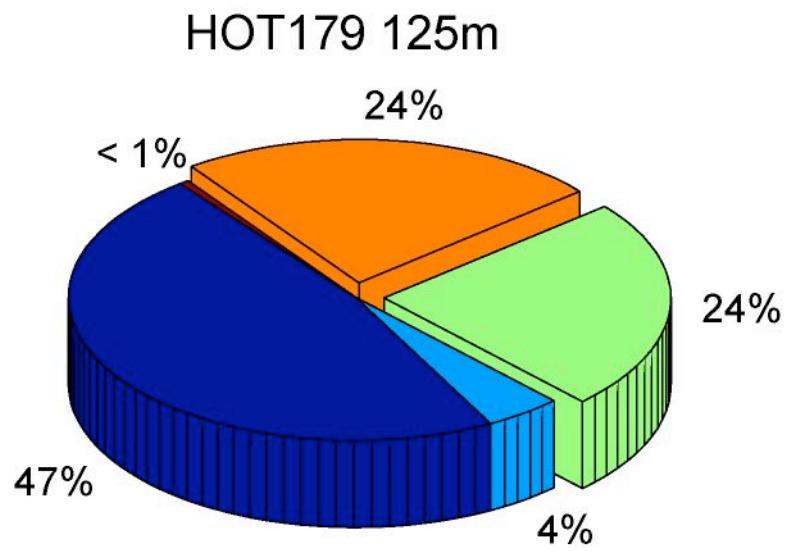
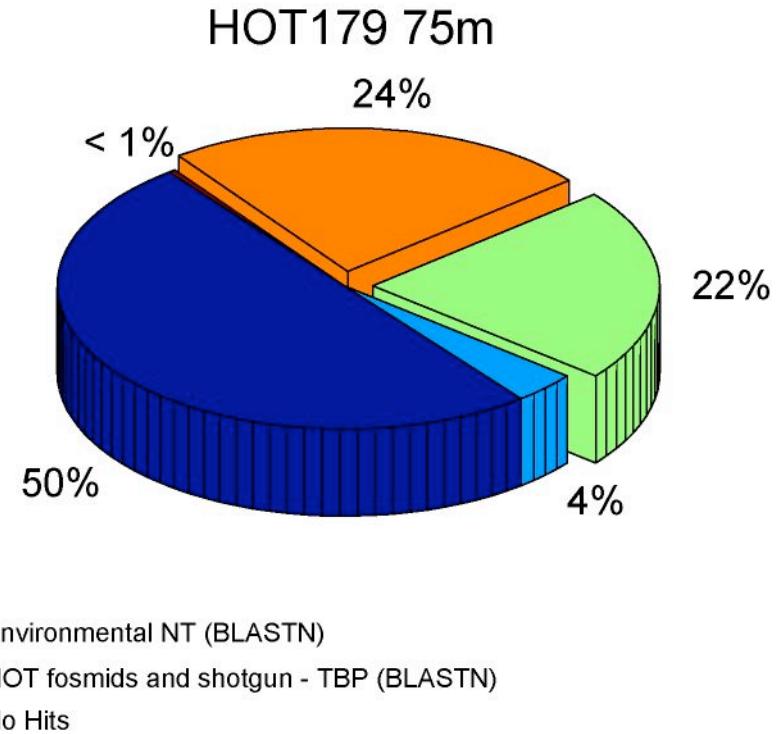
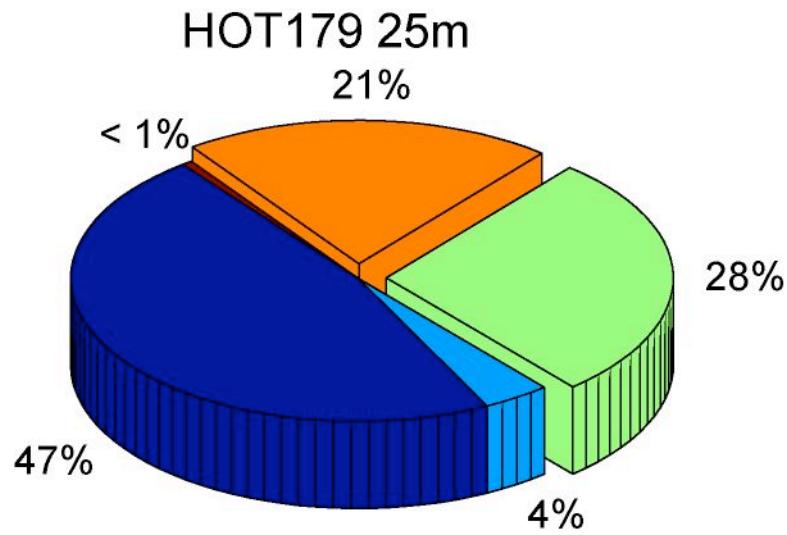
■ NR and GOS (BLASTX)

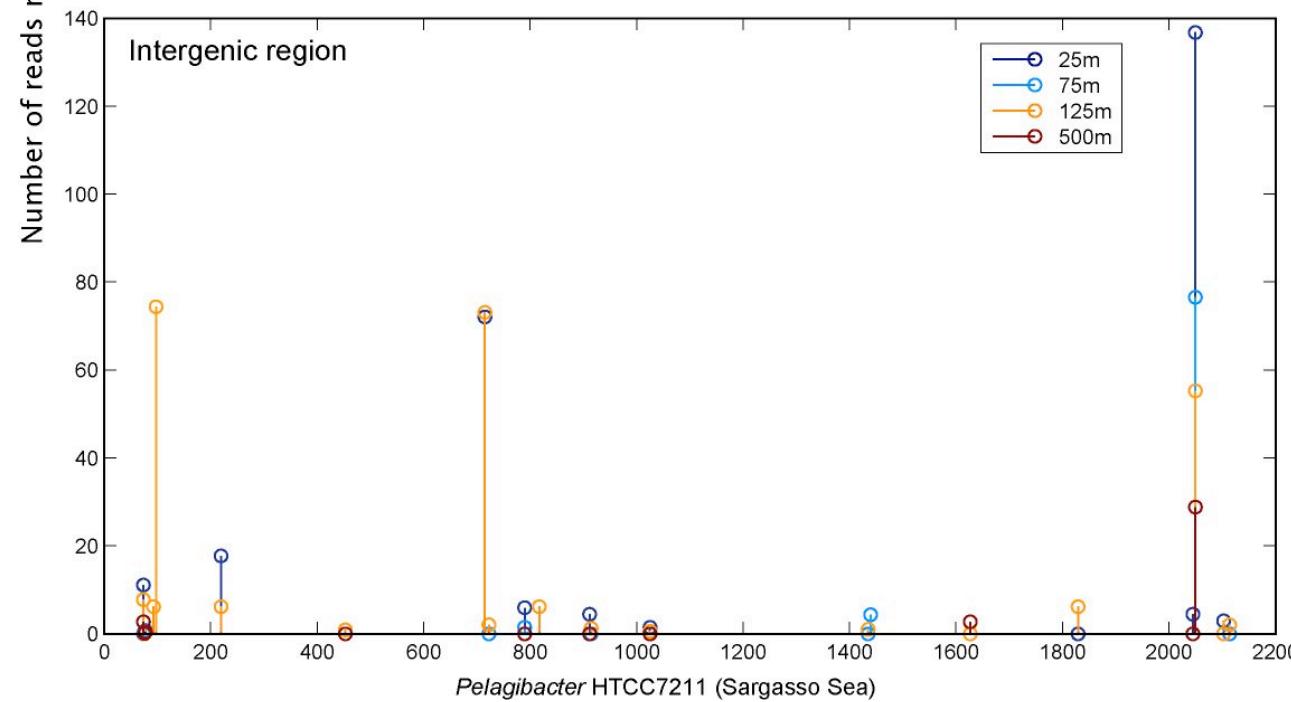
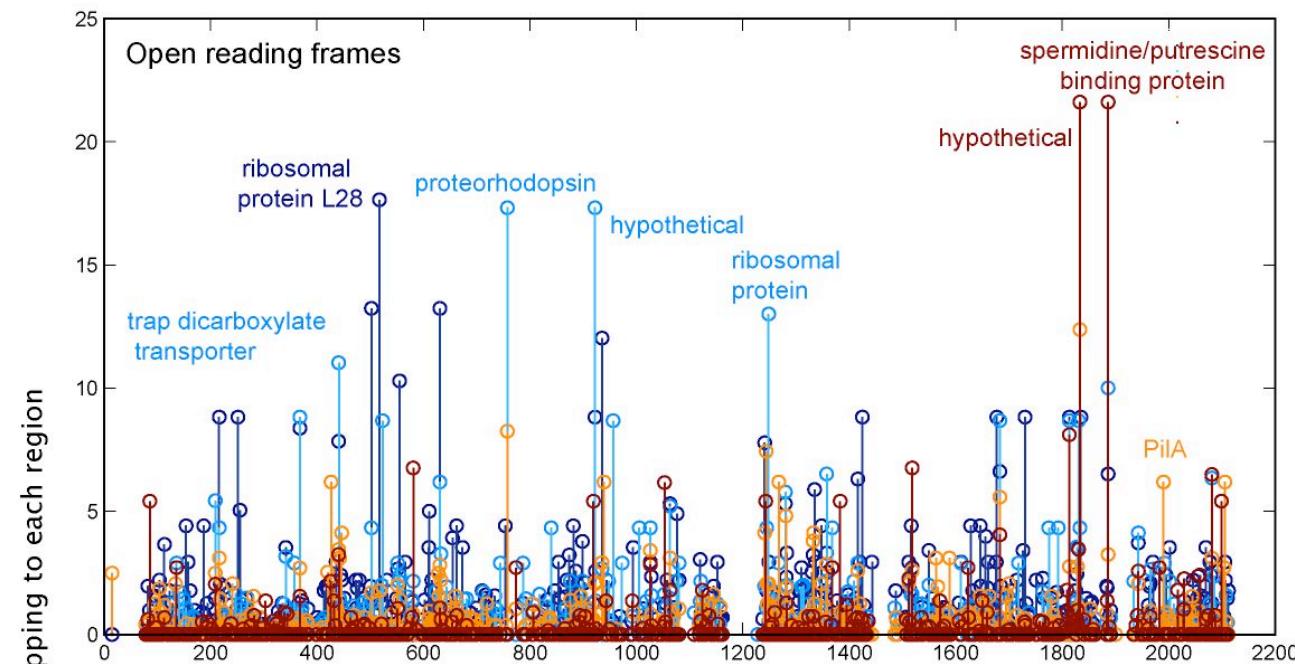
HOT179 125m



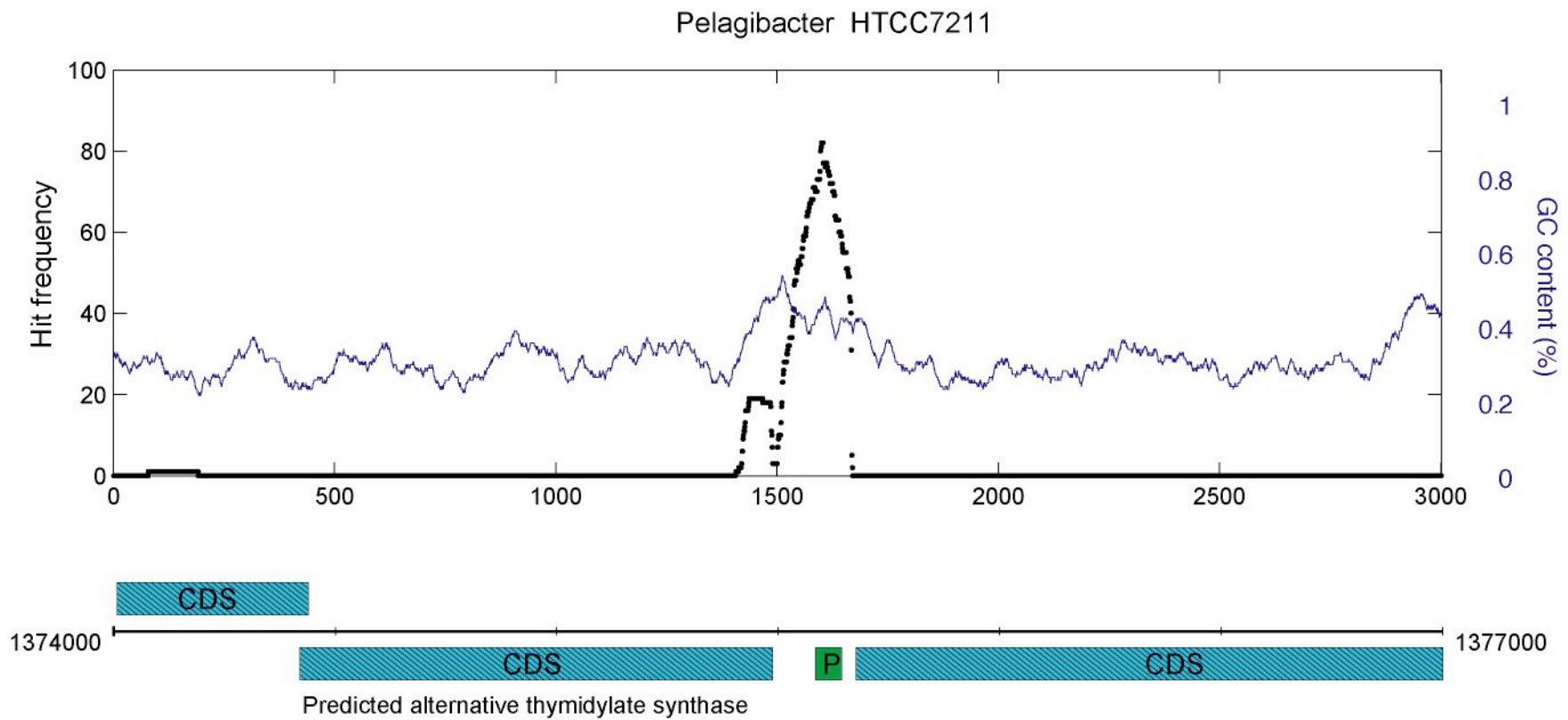
HOT179 500m



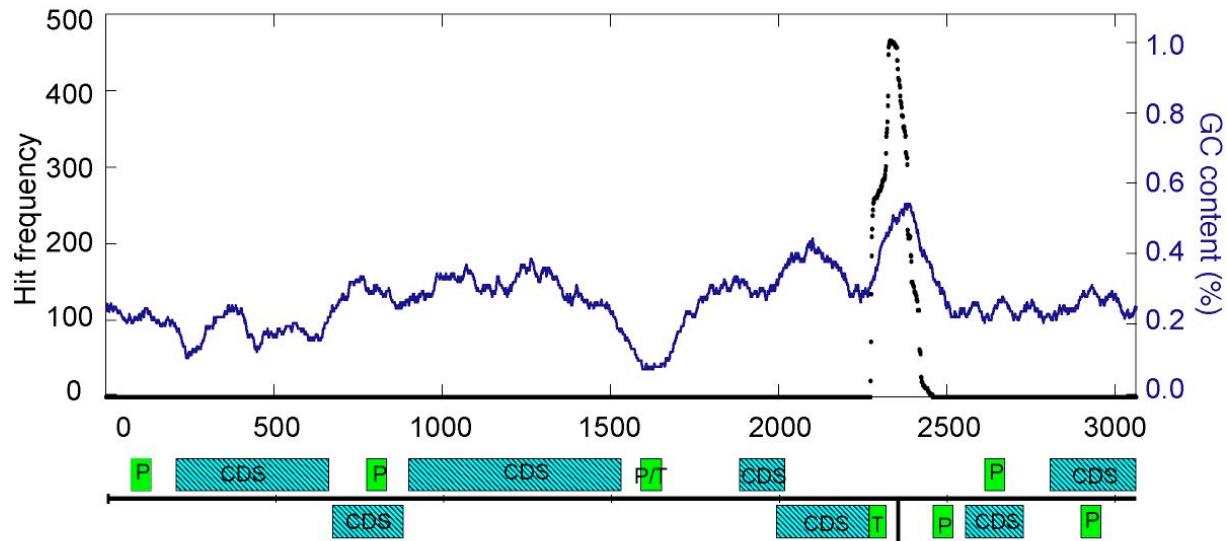




Possible small non-coding RNA in Pelagibacter?



Detecting small non-coding RNA in the environment

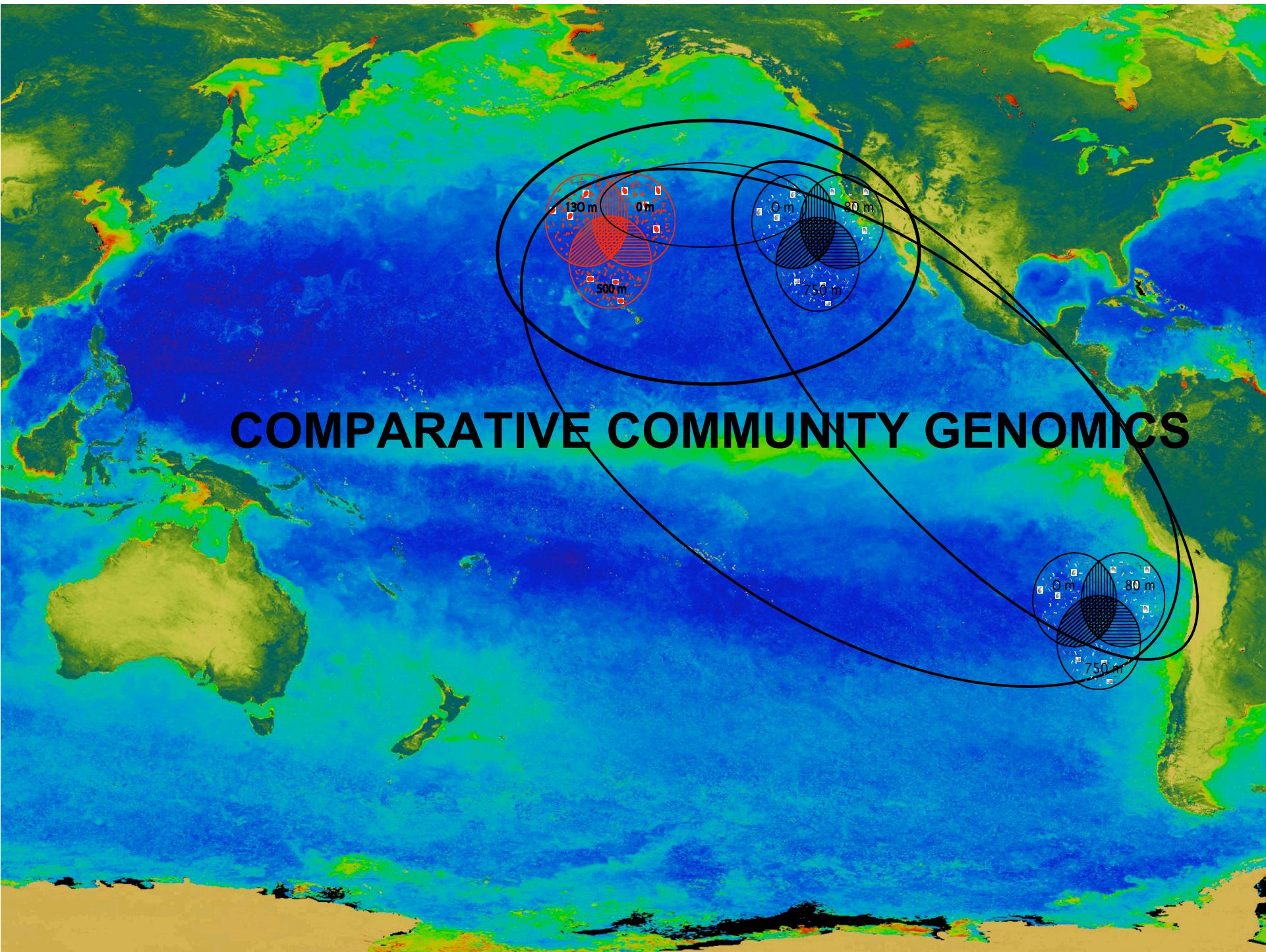


**Detected a large number
of cDNA reads mapped to
intergenic regions on
environmental sequences**

Transcriptional Terminator: Size=36 nt.
 $\Delta G=-13.90$ kcal/mol

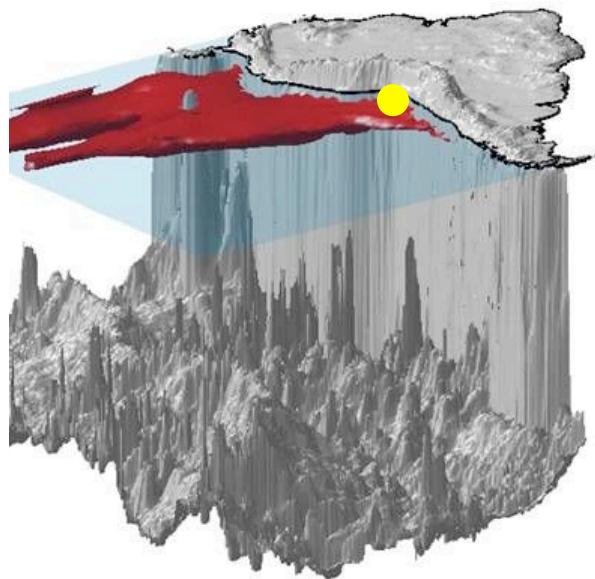
Antiterminator: Size=61 nt.
 $\Delta G=-16.90$ kcal/mol

Anti-antiterminator: Size=45 nt.
 $\Delta G=-13.90$ kcal/mol



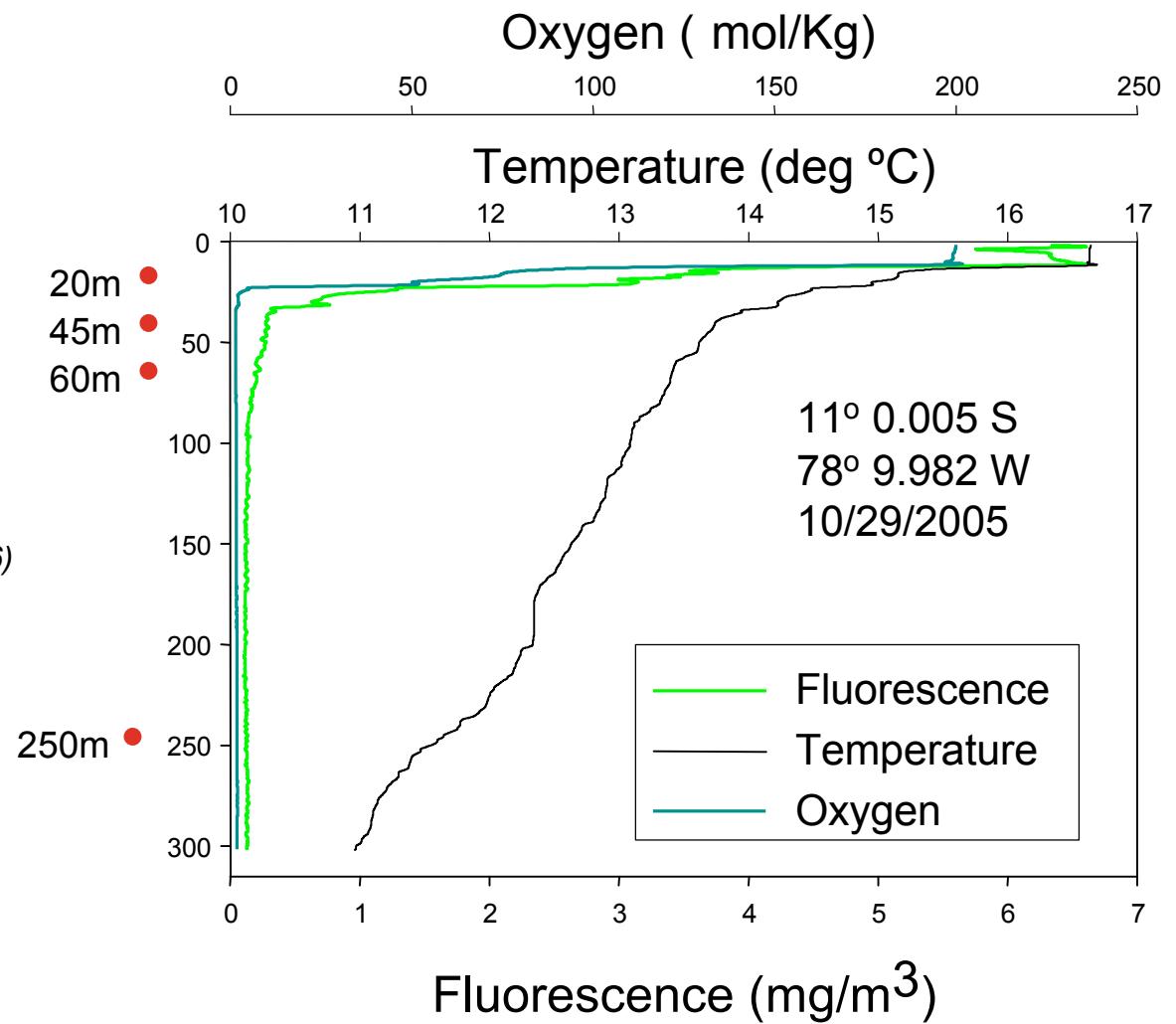
COMPARATIVE COMMUNITY GENOMICS

OMZ metagenomic and metatranscriptomic libraries



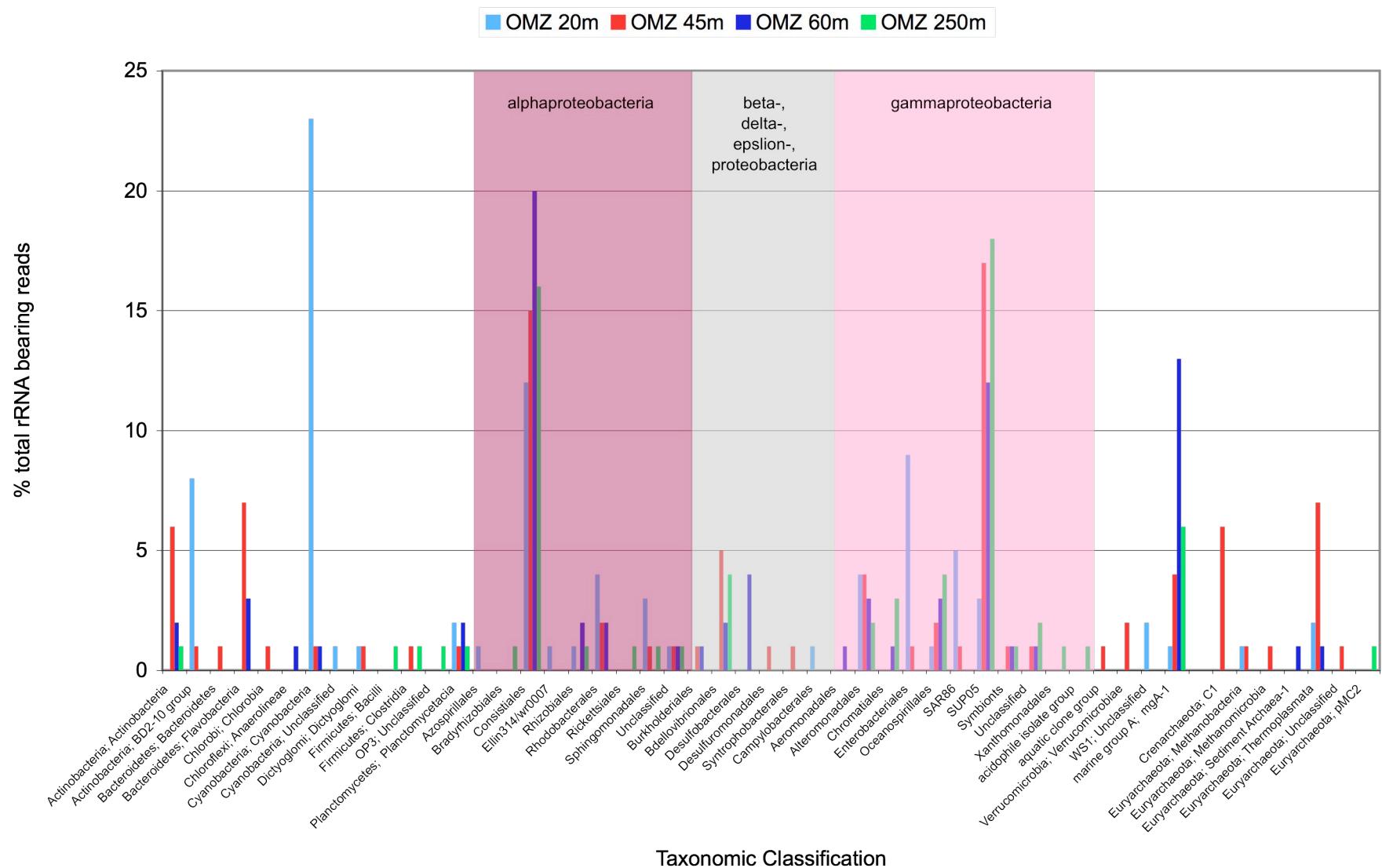
De Pol & Ulloa, Zonas de minimo oxígeno (2006)

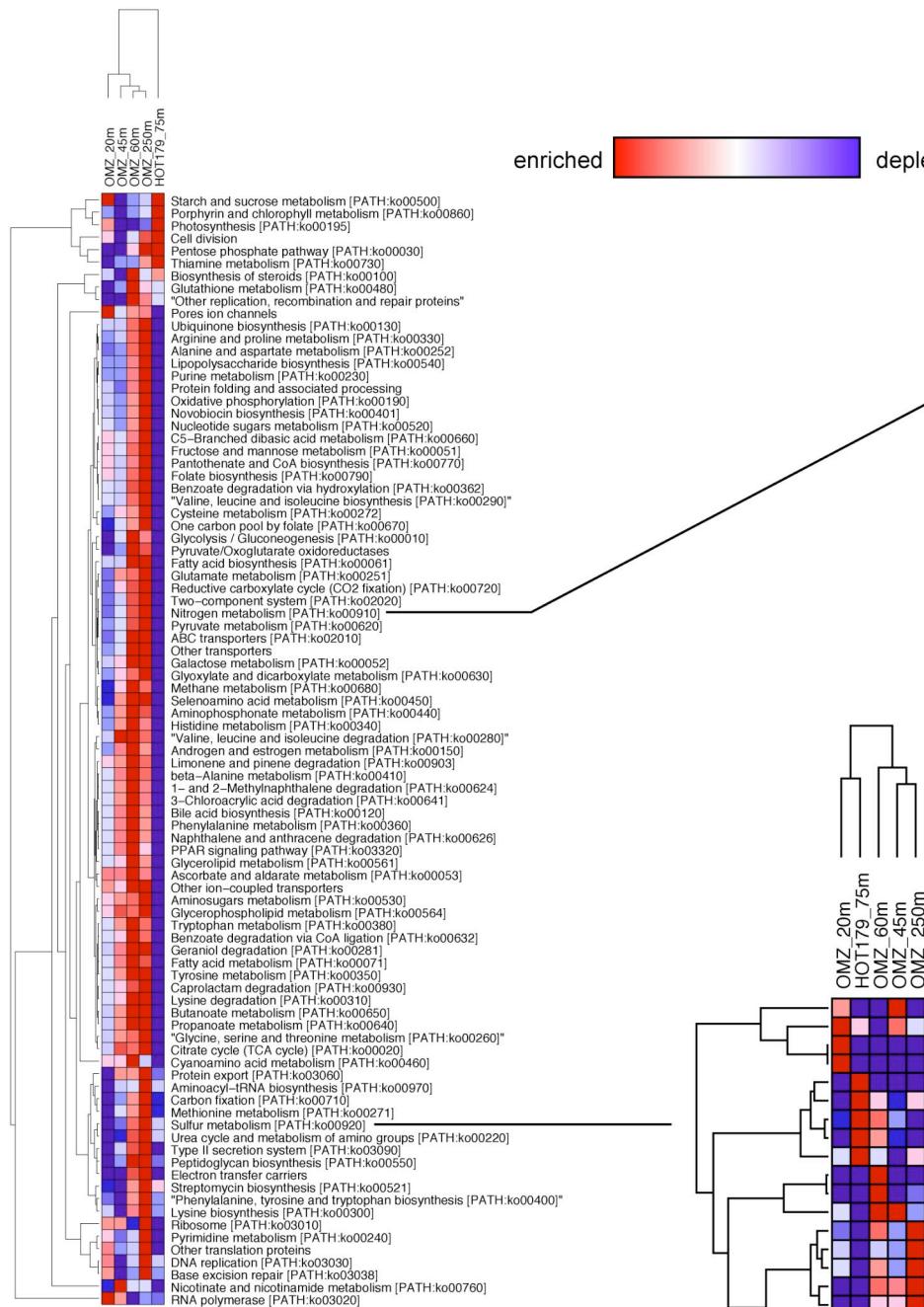
- fosmid library (40,000 clones/per depth, archive ~1.6Gb)
- 454 FLX DNA (1/4 chip/per depth)
- 454 FLX cDNA (in progress)



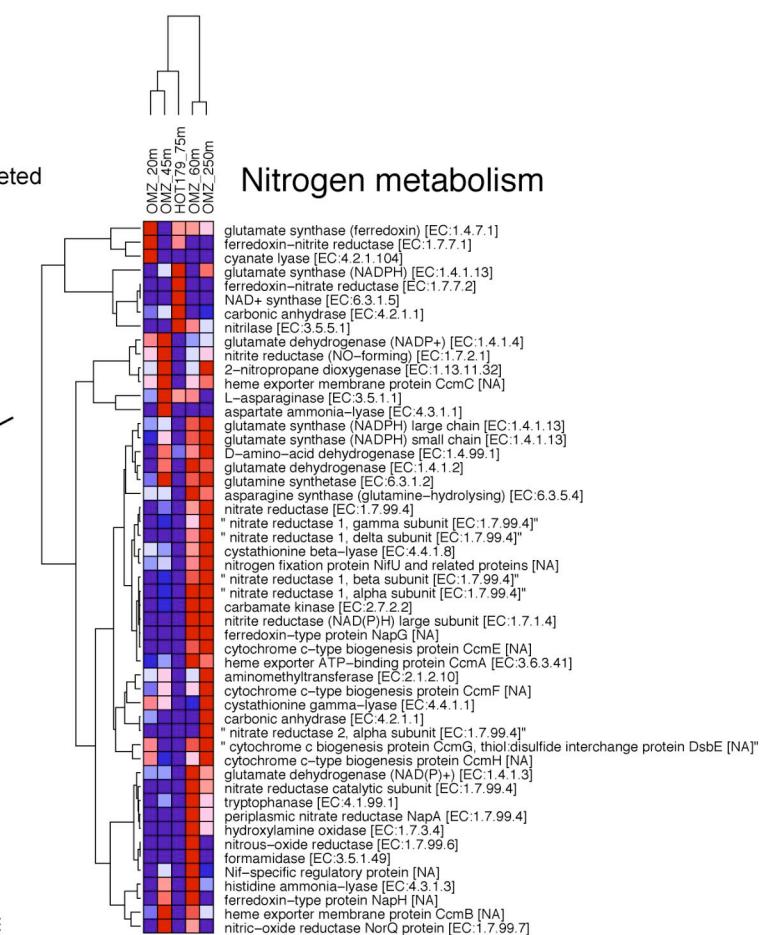
Samples collected by:
Osvaldo Ulloa
Ann Thompson (Chisholm Lab)

OMZ depth profile 16S rRNA taxonomic classification

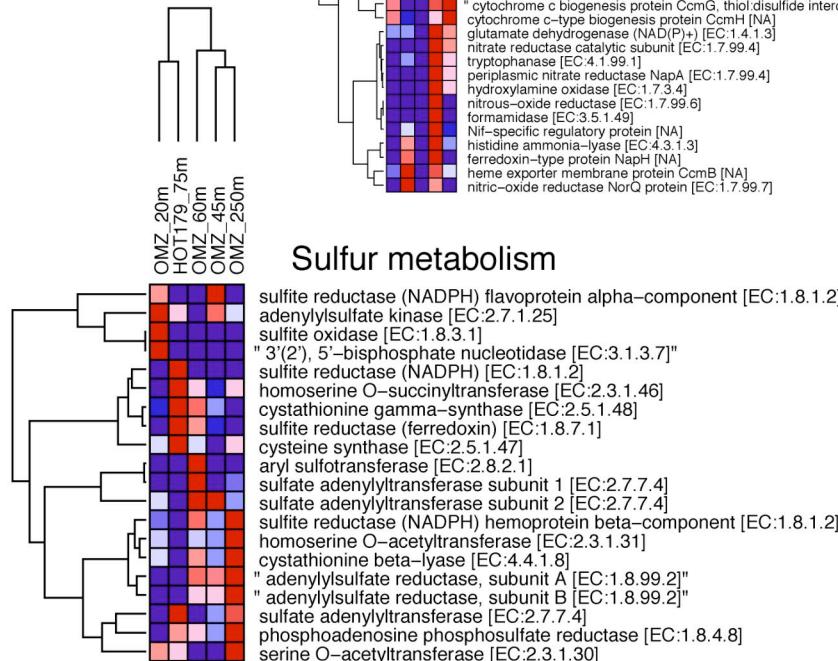




Nitrogen metabolism



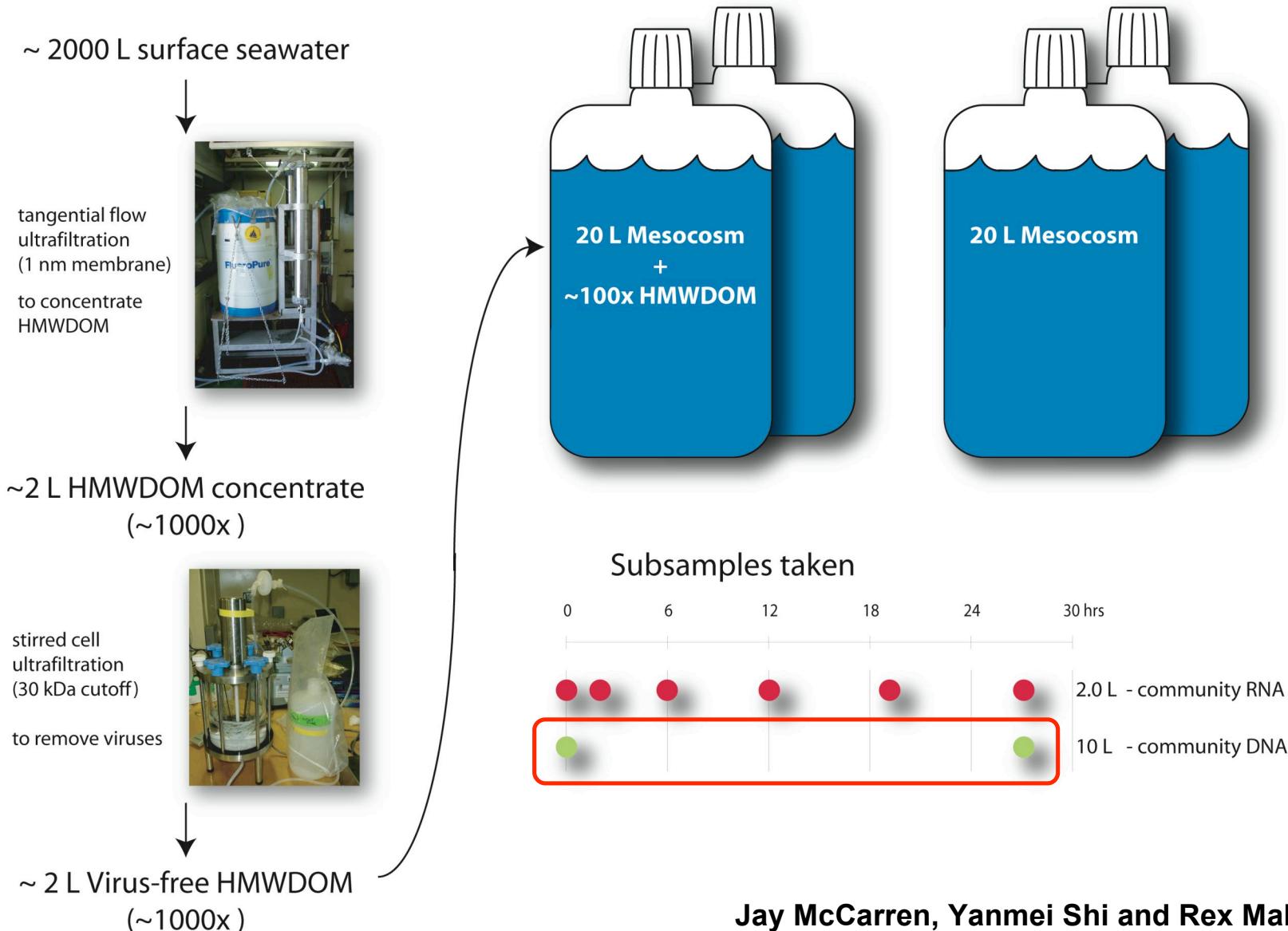
Sulfur metabolism



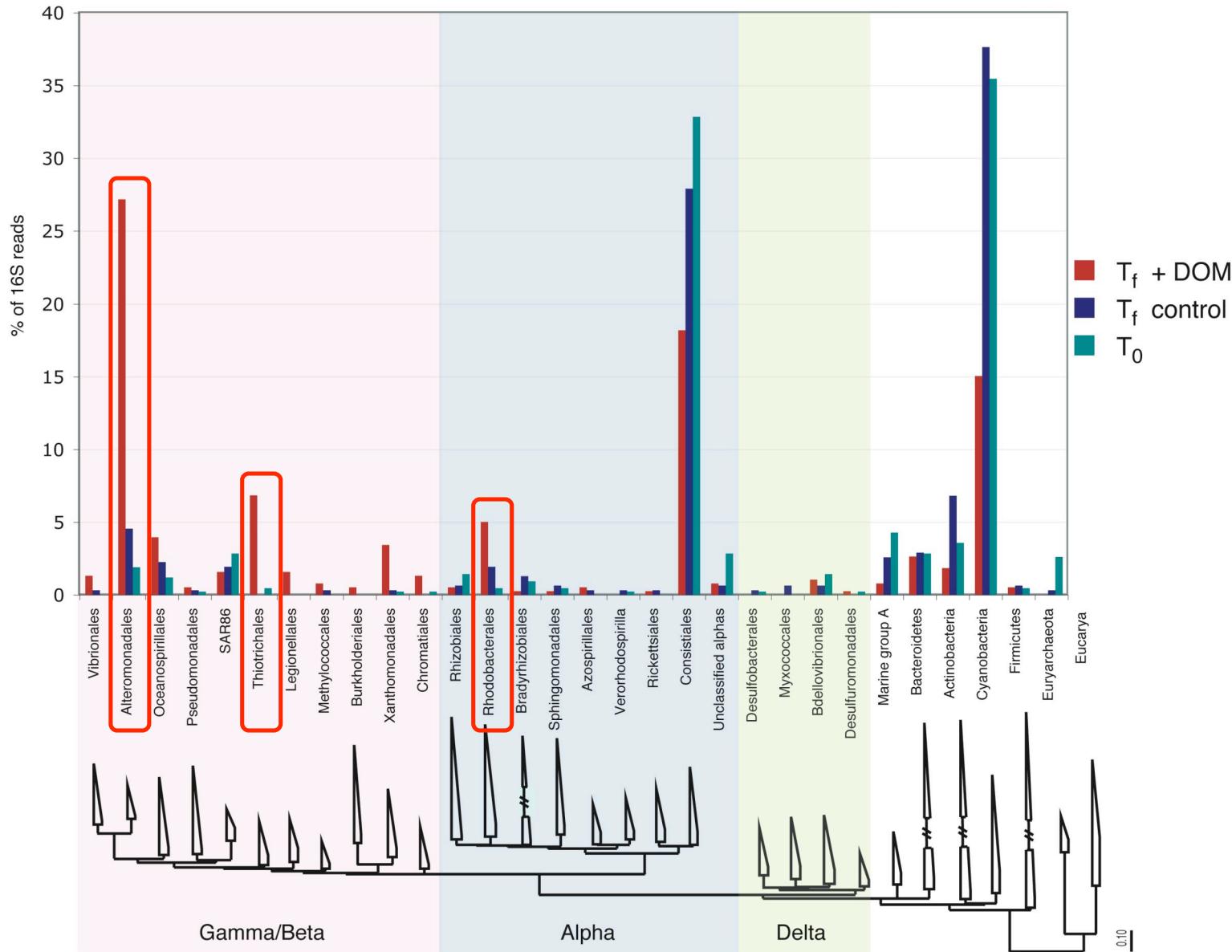
Conclusions

- Metatranscriptomic analysis of marine microbial assemblages reveals discernible patterns of gene expression
- Gene discovery
 - ecological importance of hypothetical proteins in the environment
- New era in experimental microbial ecology
 - microbial assemblage dynamics
 - perturbation experiments
- Comparative genomic and transcriptomic analyses from spatially and geochemical distinct marine environments

Experimental community transcriptomics



Shift in community composition following 27hrs incubation



DeLong Lab

- Justin Buck
- Rachel Barry
- John Eppley
- Sarah Lincoln
- Jay McCarren
- Julia Maresca
- Chon Martinez
- Tsultrim Palden
- Vinh Pham
- Virginia Rich
- Yanmei Shi
- Laure-Anne Ventouras



Collaborators

MIT - Chisholm Lab

Penn State - Stephan Schuster

Universidad de Concepcion

CMORE & HOT Teams

Yanmei Shi, Gene Tyson and Ed DeLong:
Metatranscriptome analyses

Jorge Frias-Lopez, Maureen Coleman and
Penny Chisholm: *Prochlorococcus*
transcriptomics

