The Power of Diversity I:
Prochlorococcus enters the genomic era

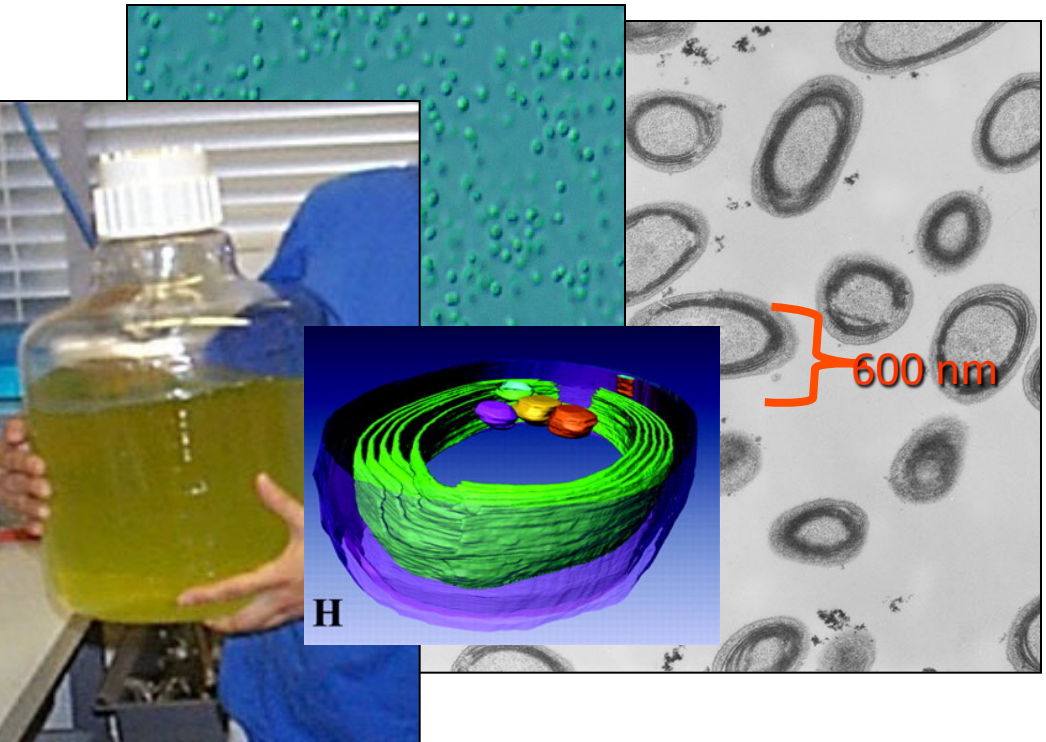
Chisholm

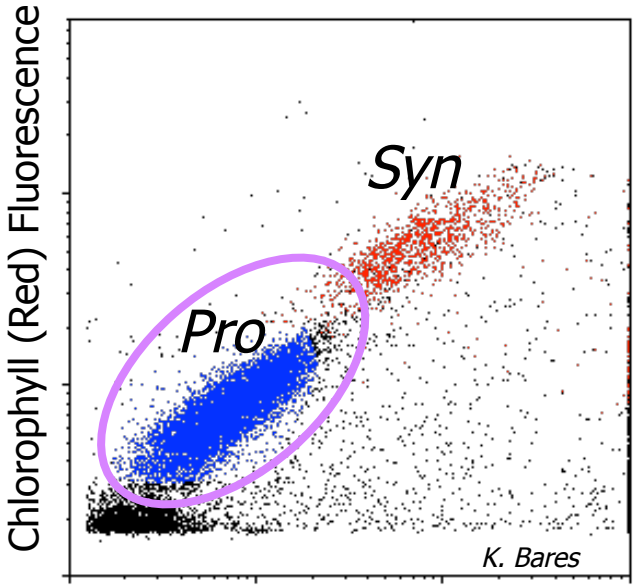
C-MORE Summer Course

June 6, 2012

What is *Prochlorococcus*?

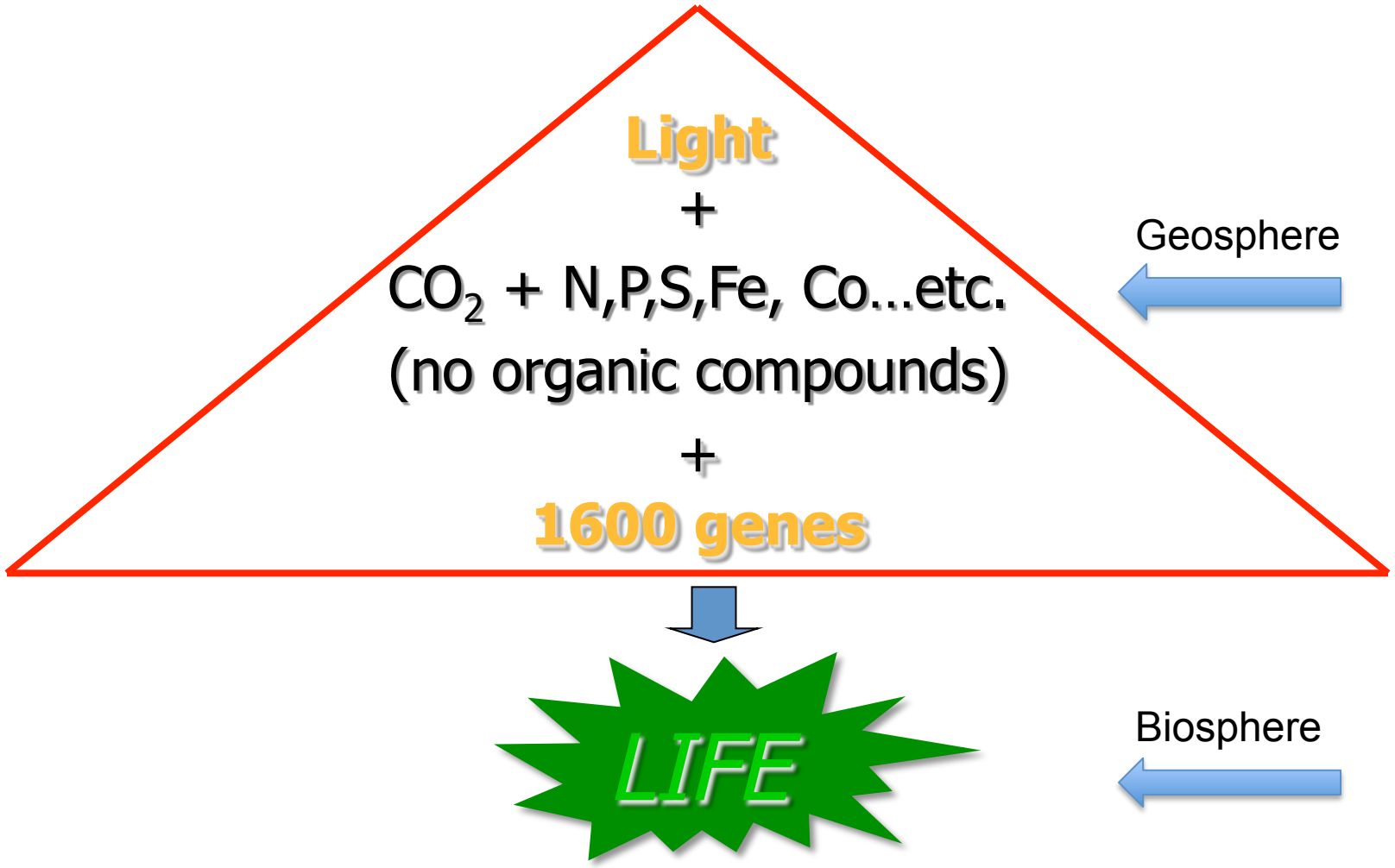
- Smallest cell in the oceans that fluoresces red
- Contains Divinyl Chl a and Chl b
- Oceanic cyanobacterium, 0.6 - 0.8 µm diameter
- Smallest (size and genome), and most abundant photosynthetic cell on Earth



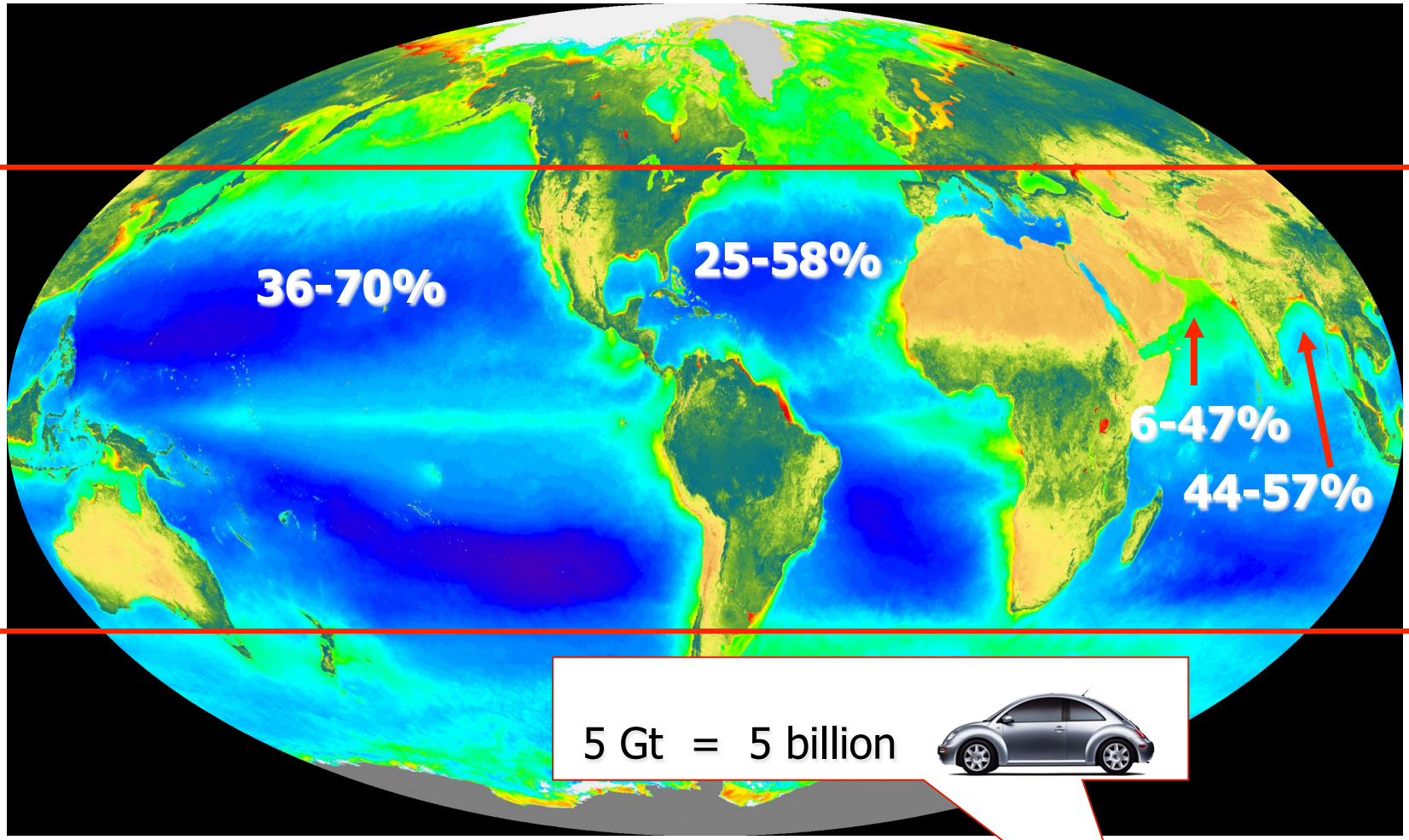


Prochlorococcus is the 'minimal life' form:

Smallest amount of information that can make life from scratch



A significant fraction of global chlorophyll

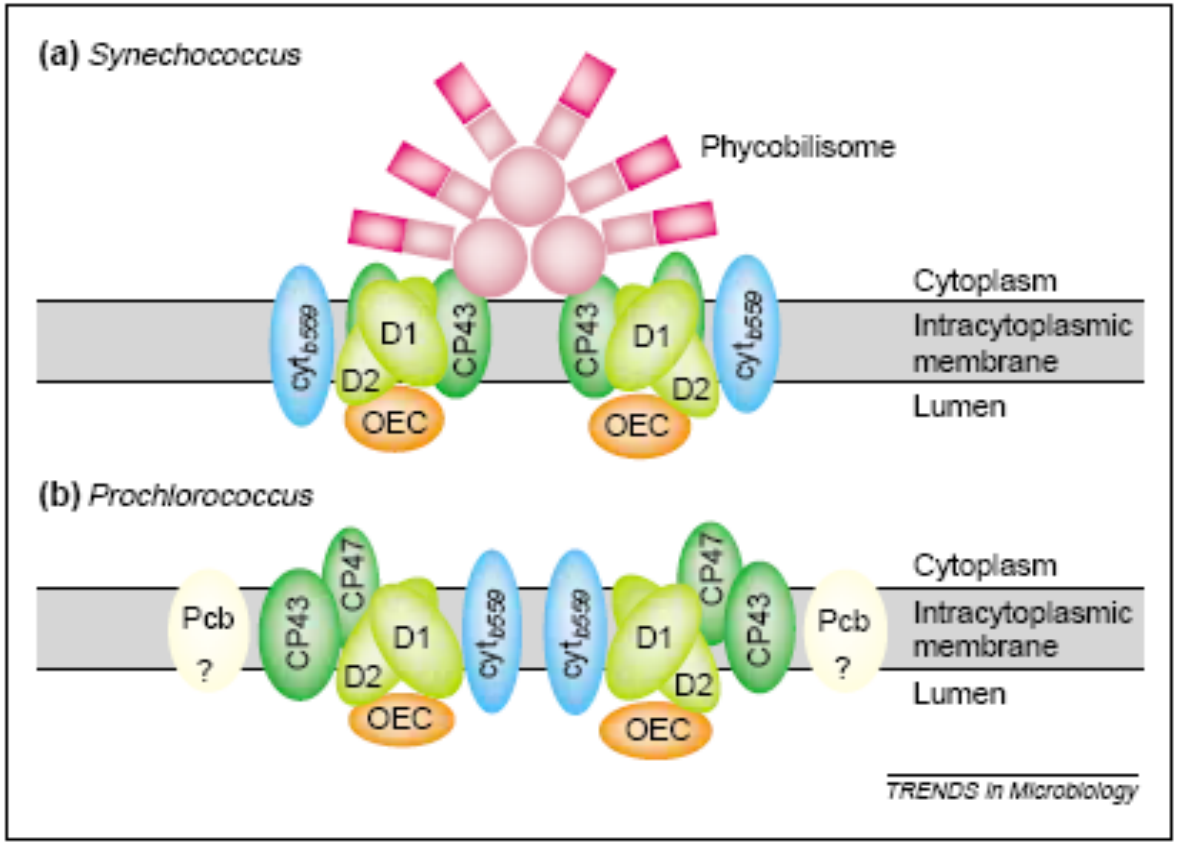


Record concentration:

700,000 cells ml^{-1}

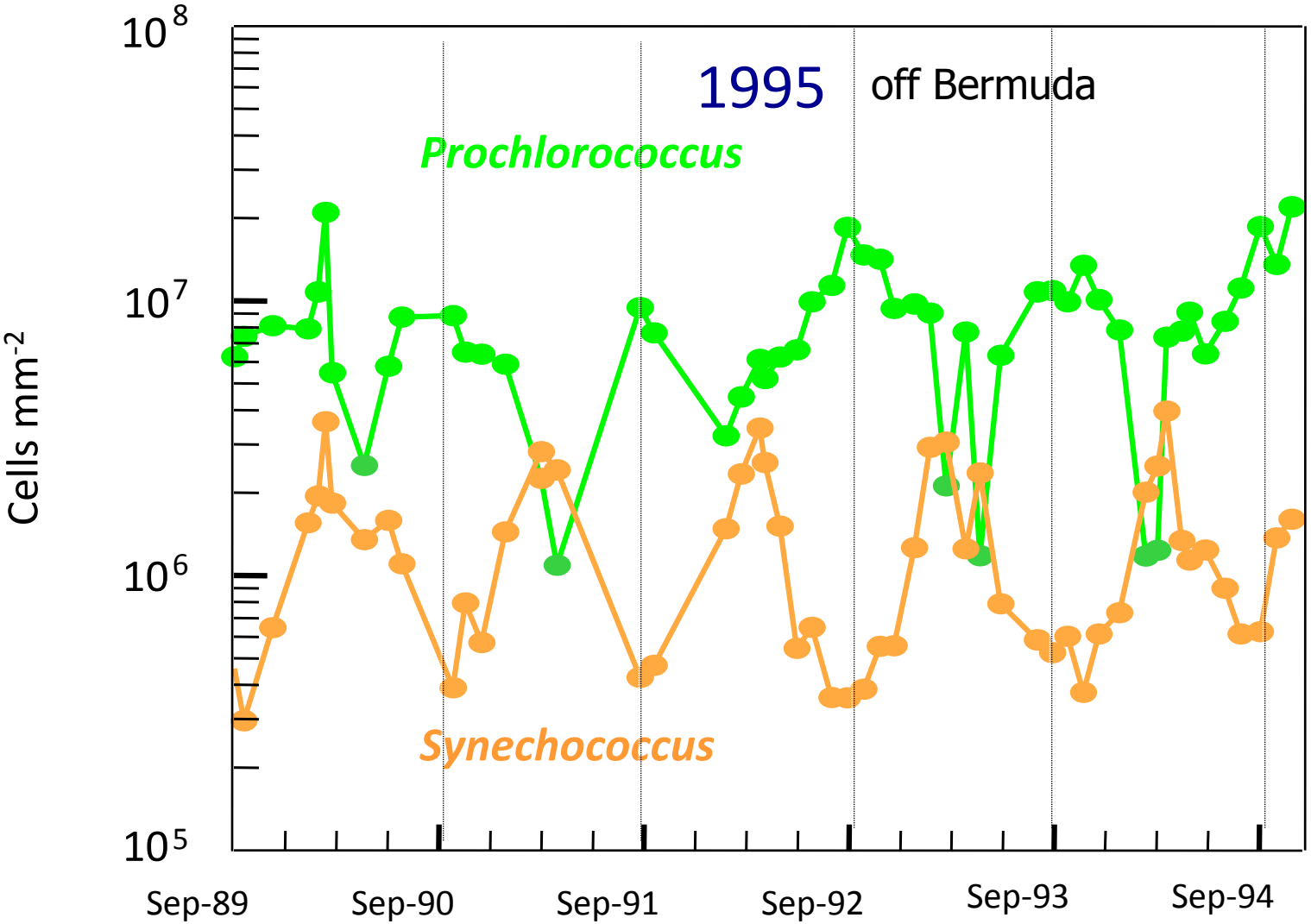
Global photosynthesis ≈ 5 Gt C yr^{-1}

Prochlorococcus is basically a slightly smaller *Synechococcus,* with a different light harvesting system

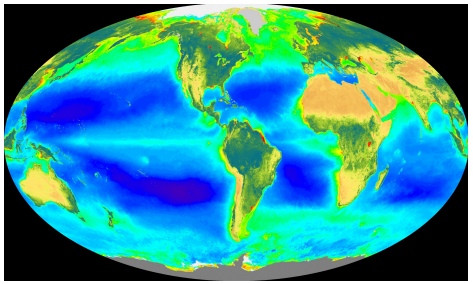


Ting et al 2006

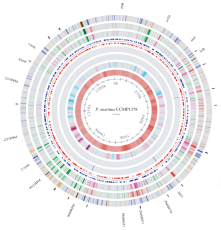
They share (and partition) the "small size bin"



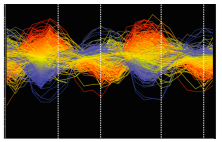
DuRand et al 2001



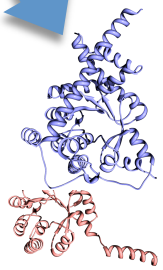
Biosphere



Genes



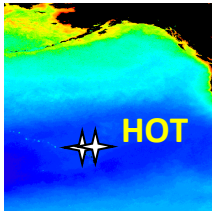
Transcripts



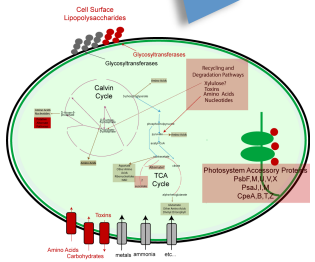
Proteins

Our Goal:

Study *Prochlorococcus* at all scales of organization

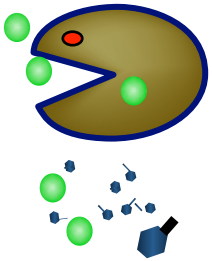


Ecosystem



Cell

Population

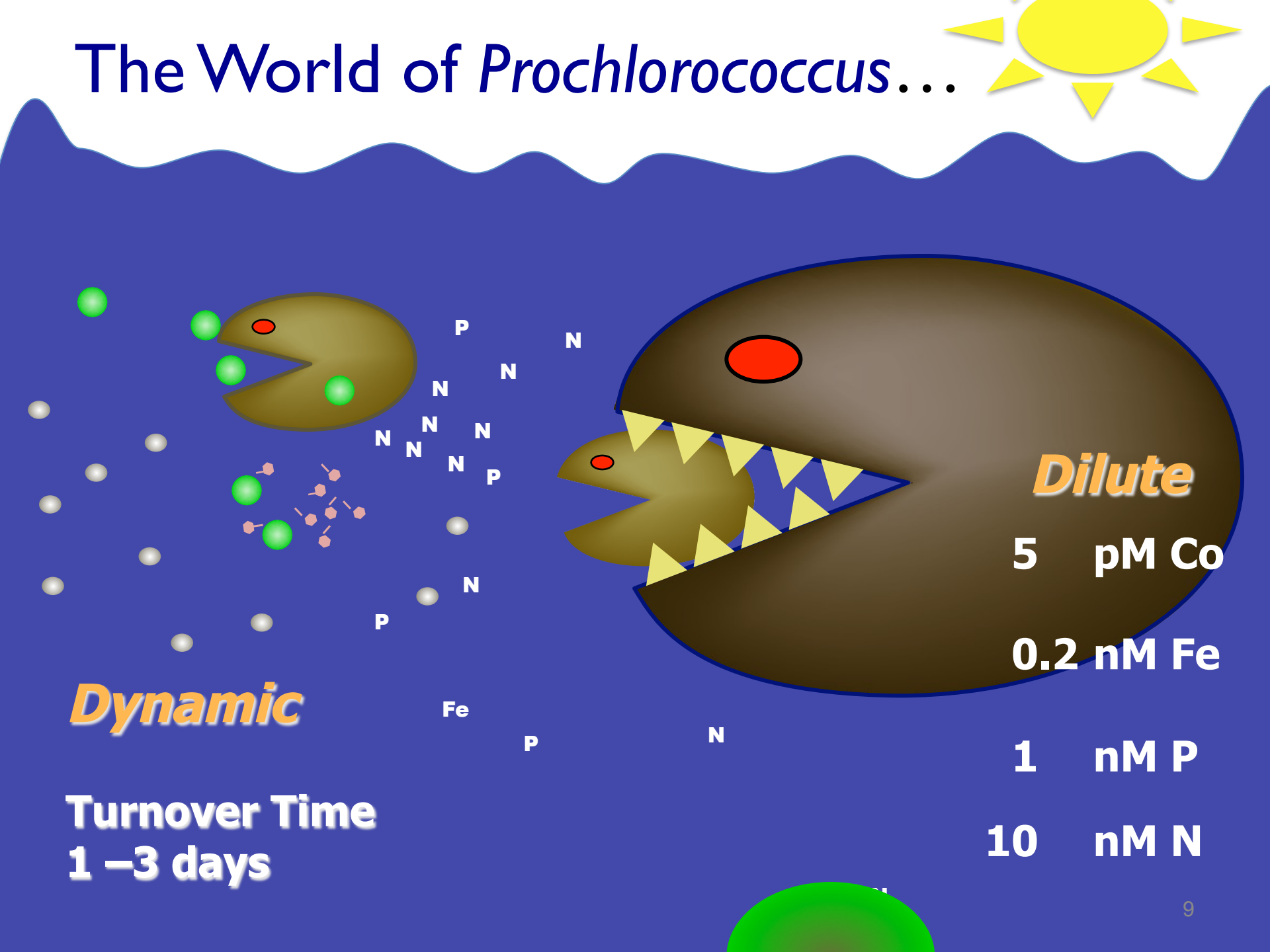


Food Web

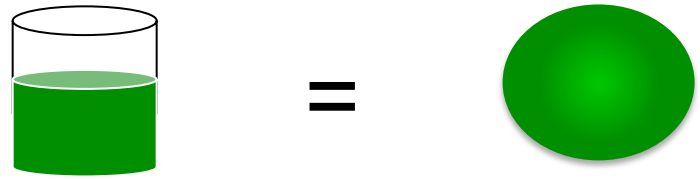
Components of two lectures

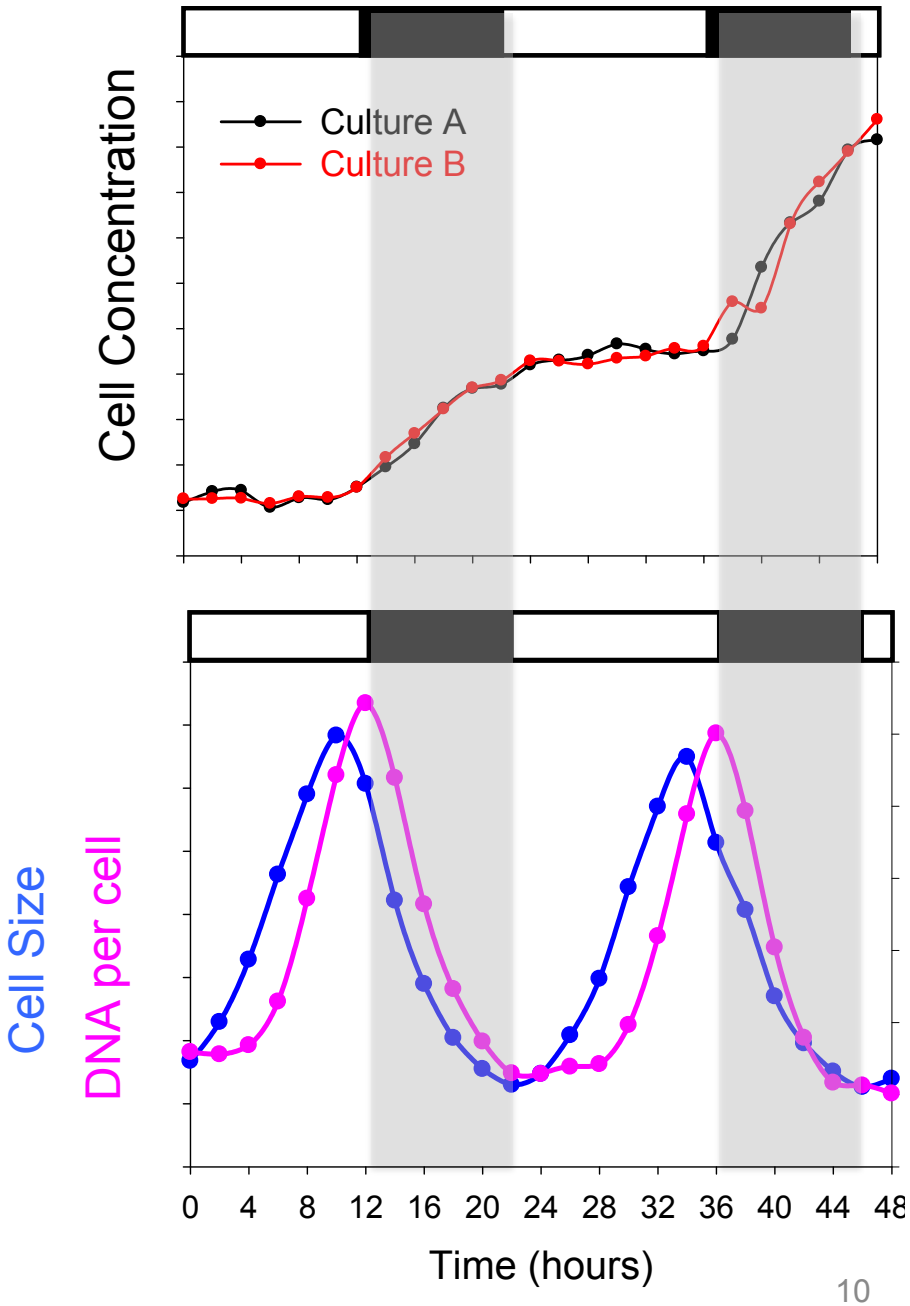
- The Cell
- Niche Dimensions of *Prochlorococcus*
- Single Cell Genomics
- Phage
- Interactions (and signaling?)

The World of *Prochlorococcus*…



Tight Cell Cycle Synchrony

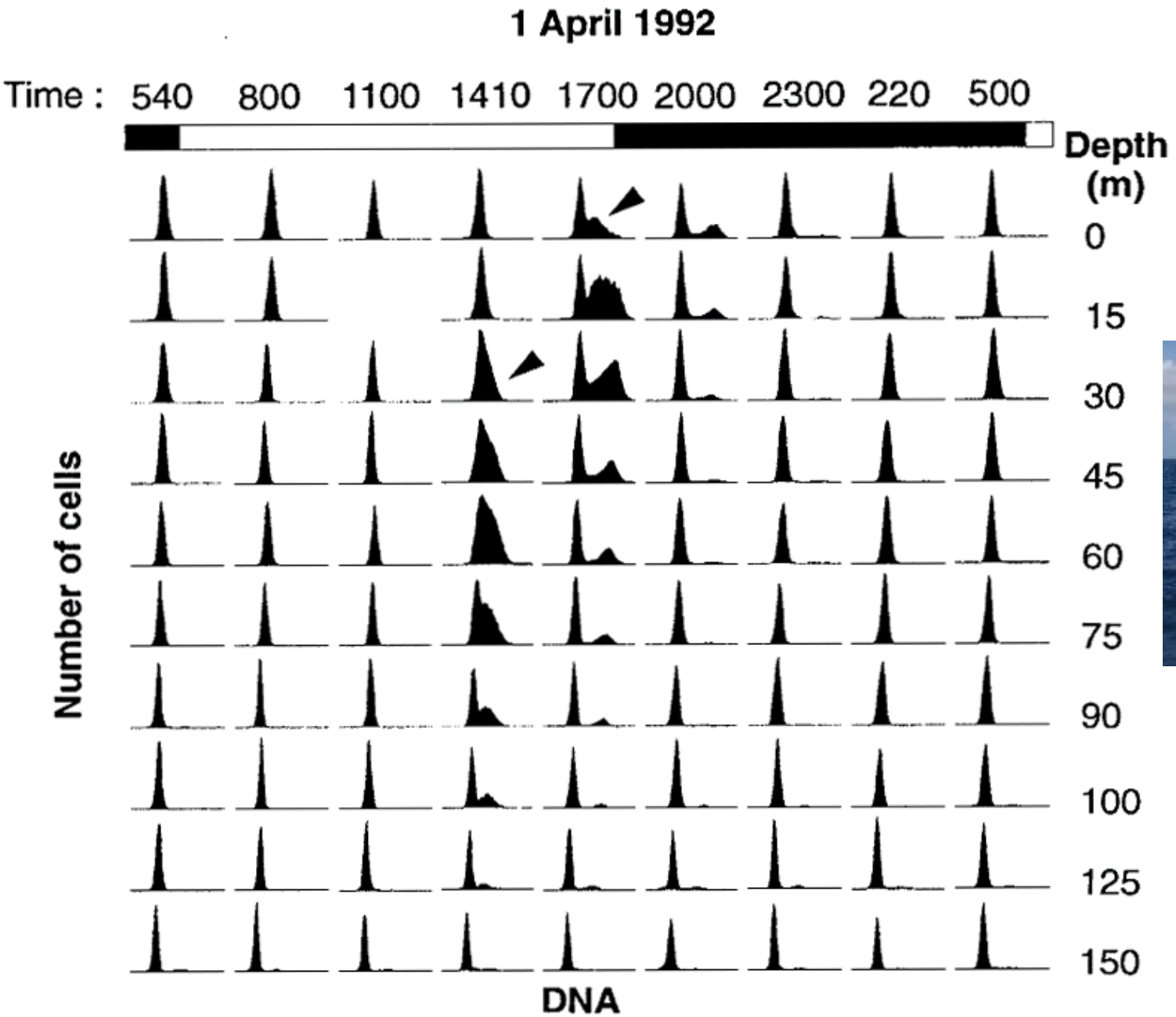




Zinser, Lindell et al. 2009

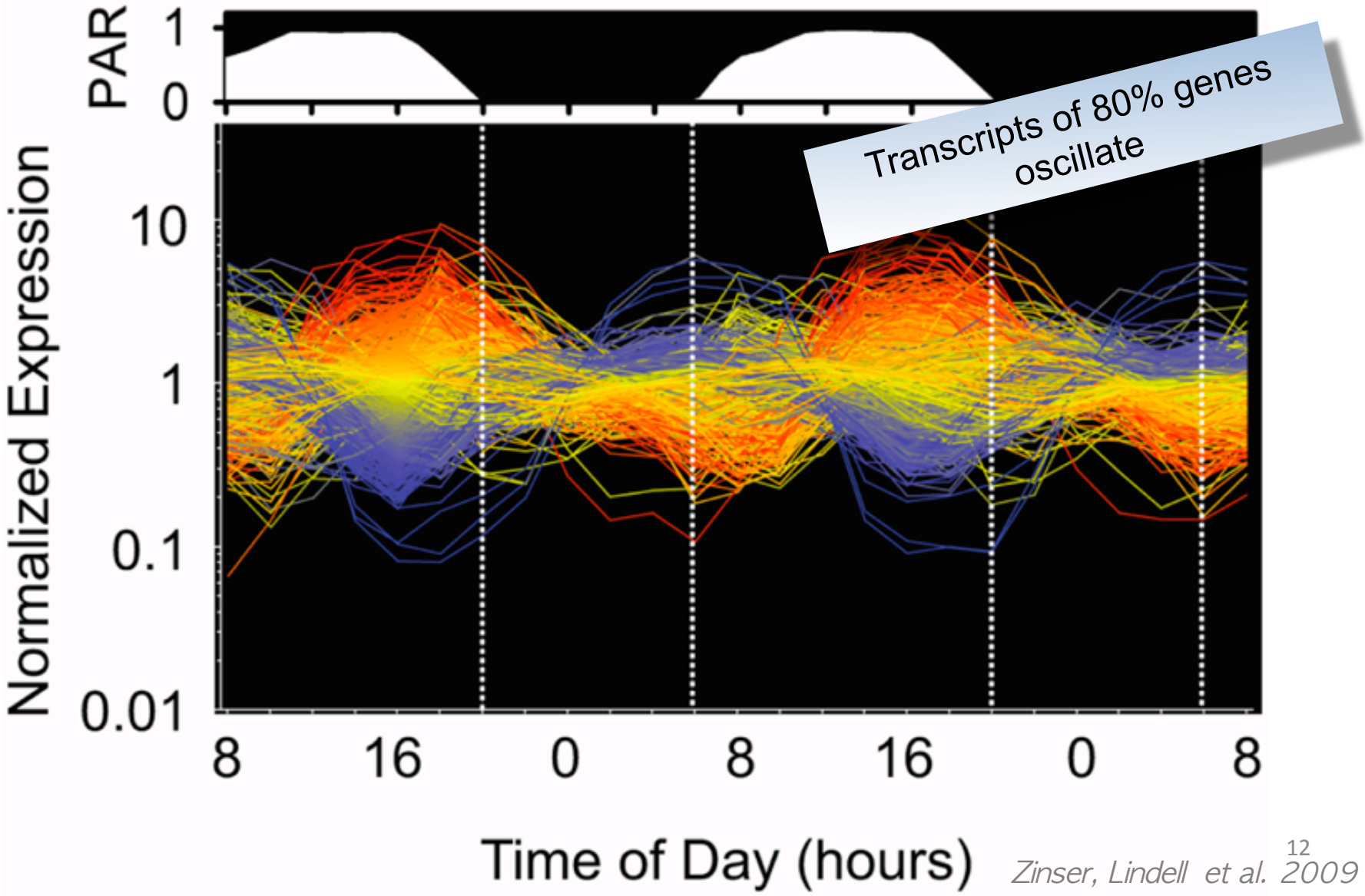
Can see it growing in the wild

Equatorial Pacific



Daniel Vaulot

Gene expression highly choreographed

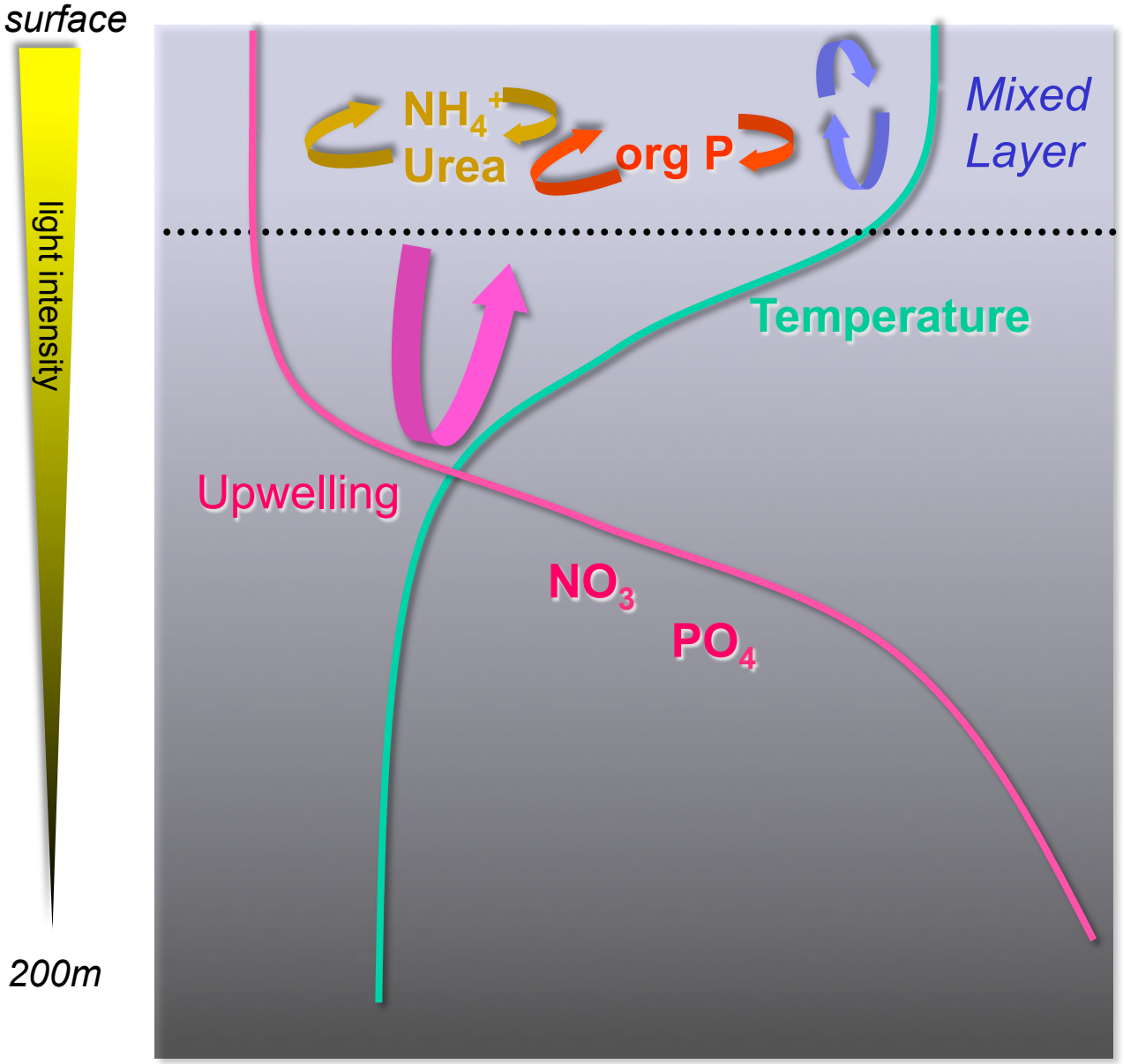


Zinser, Lindell et al. 2009

Components of two lectures

- The Cell
- Niche Dimensions of *Prochlorococcus*
- Single Cell Genomics
- Phage
- Interactions (and signaling?)

Structure of the ocean habitat

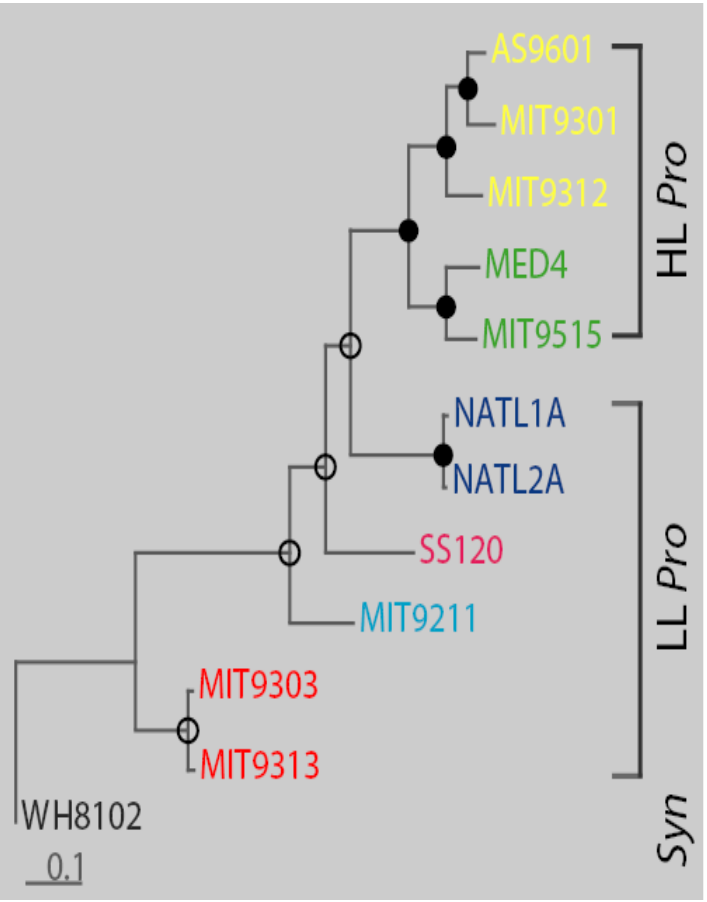


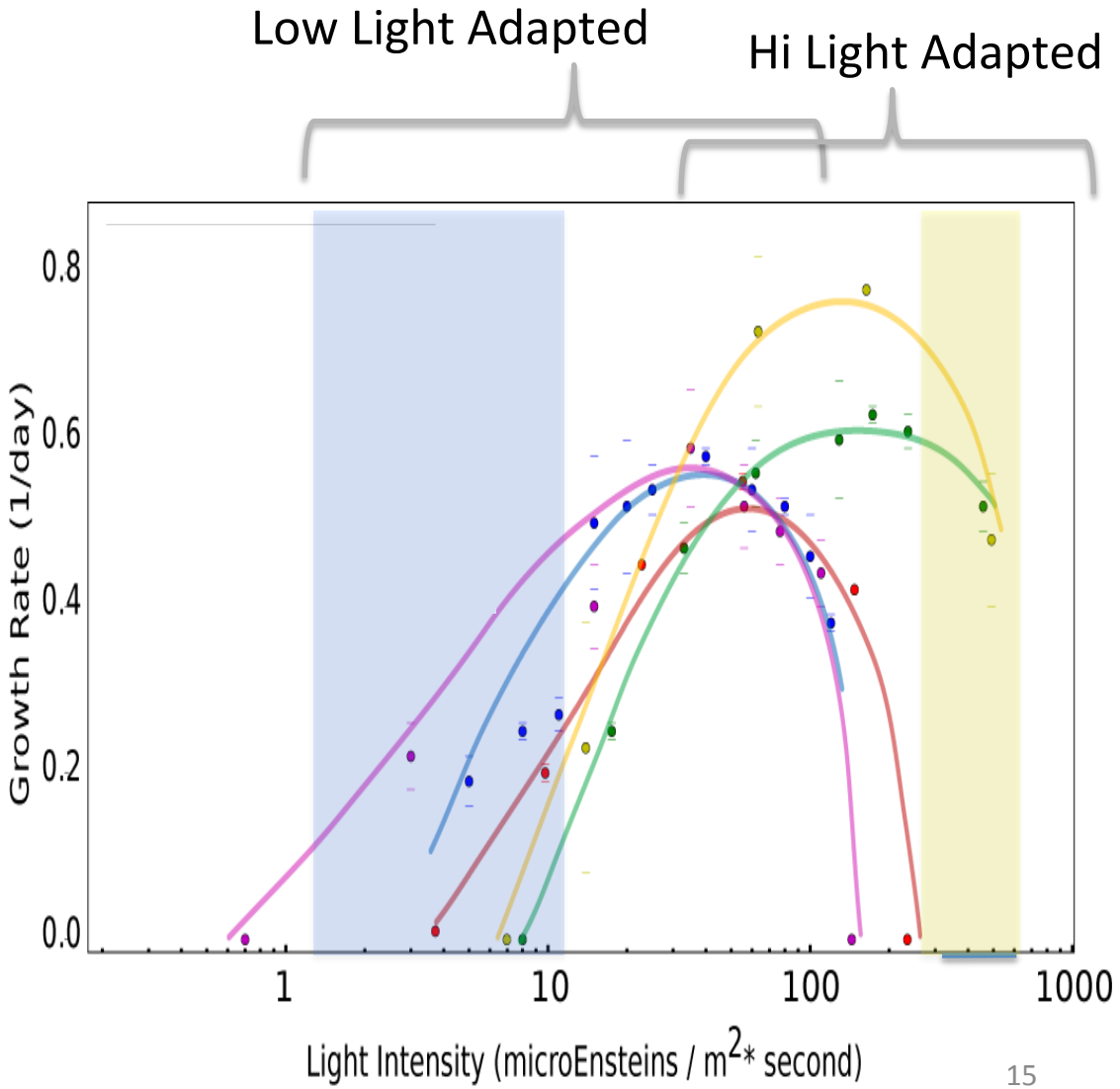
…physical and chemical gradients…

Niche dimensions!

Light adaptation defines ecotypes…

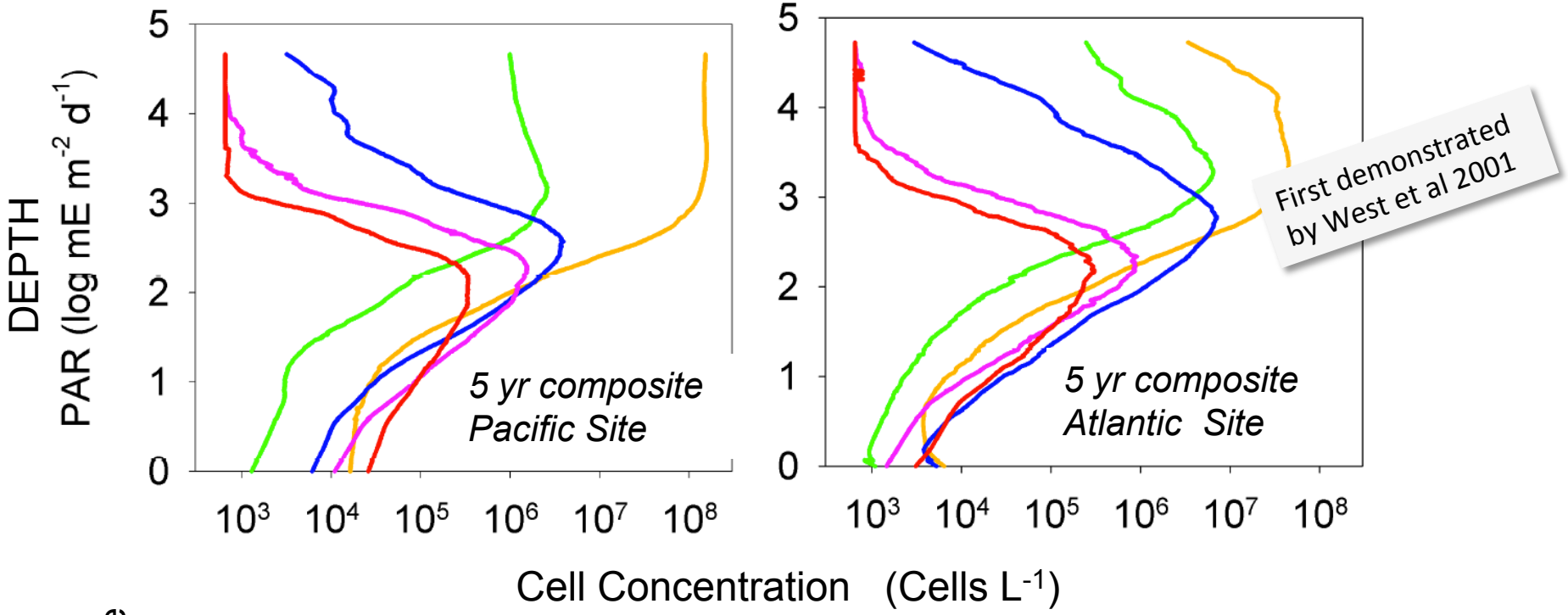
Whole Genome Phylogeny
(matches rRNA ITS phylogeny)

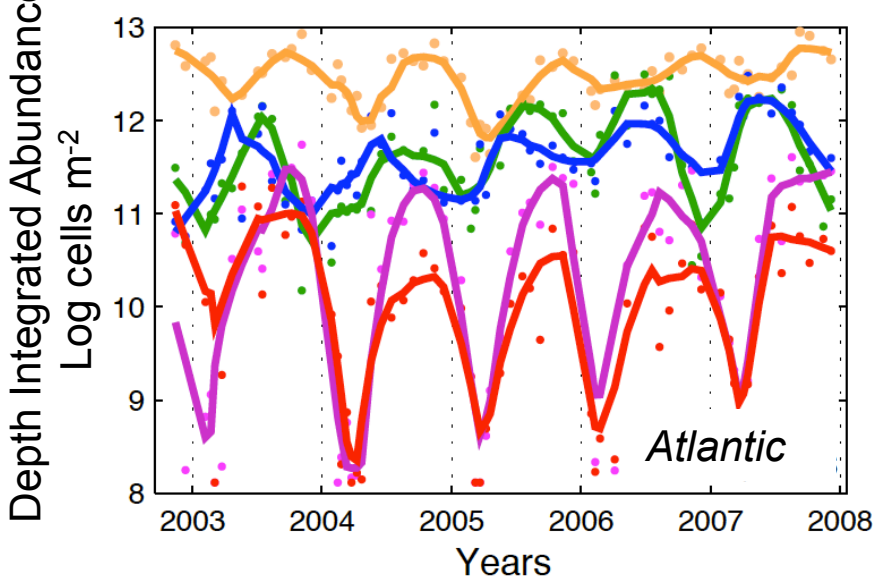


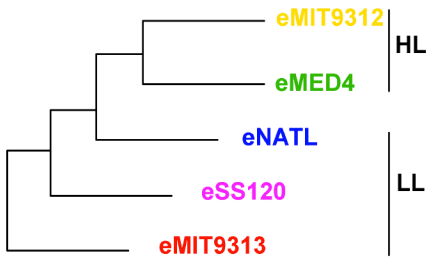


Moore, Rocap, Kettler and more, 1990 - 2006

…leading to niche differentiation

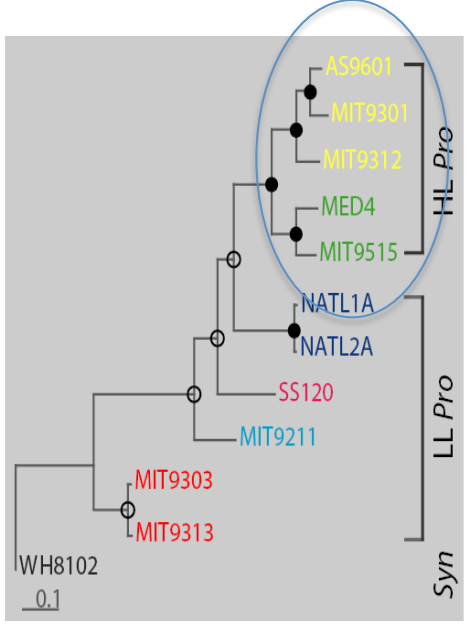


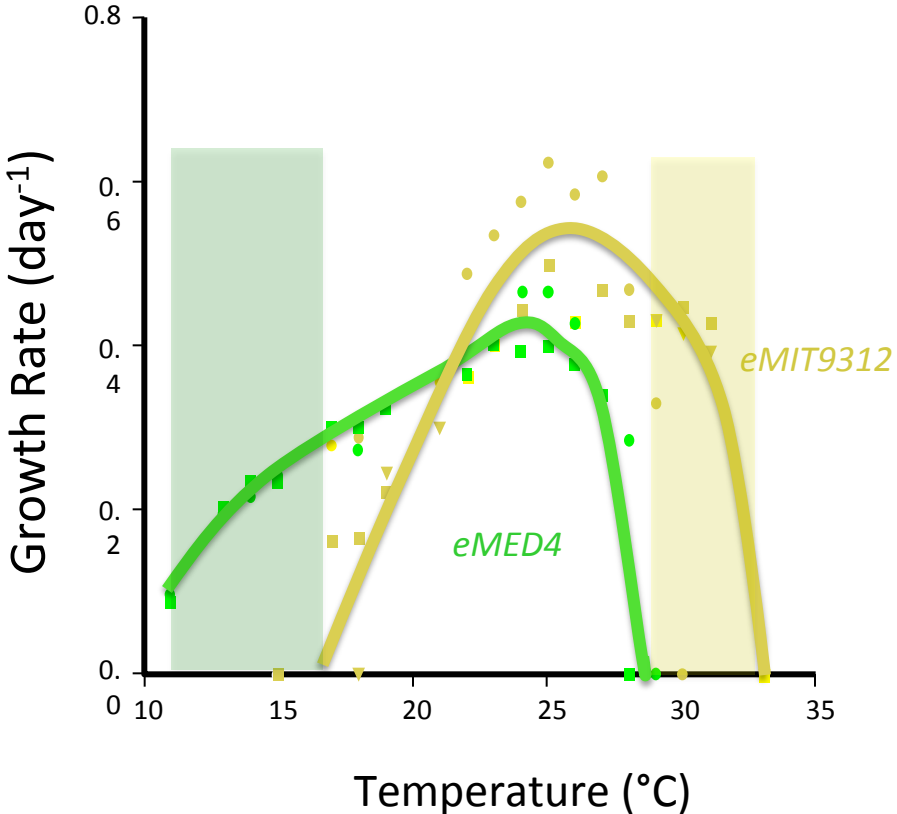


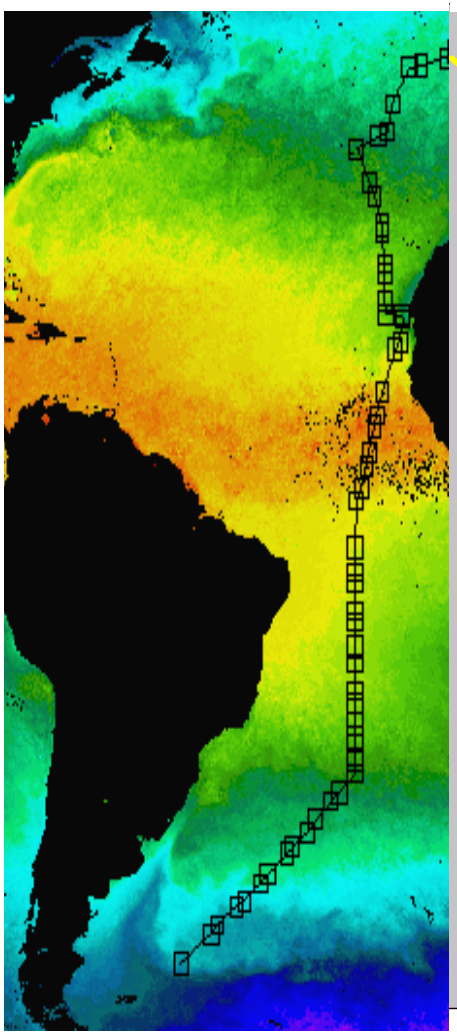


Malmstrom et al 2010

Temperature differentiates the two High Light ecotypes

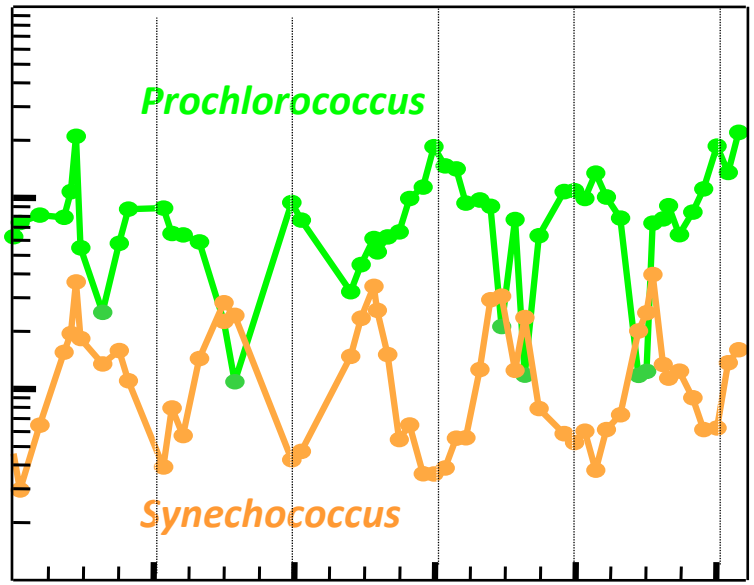




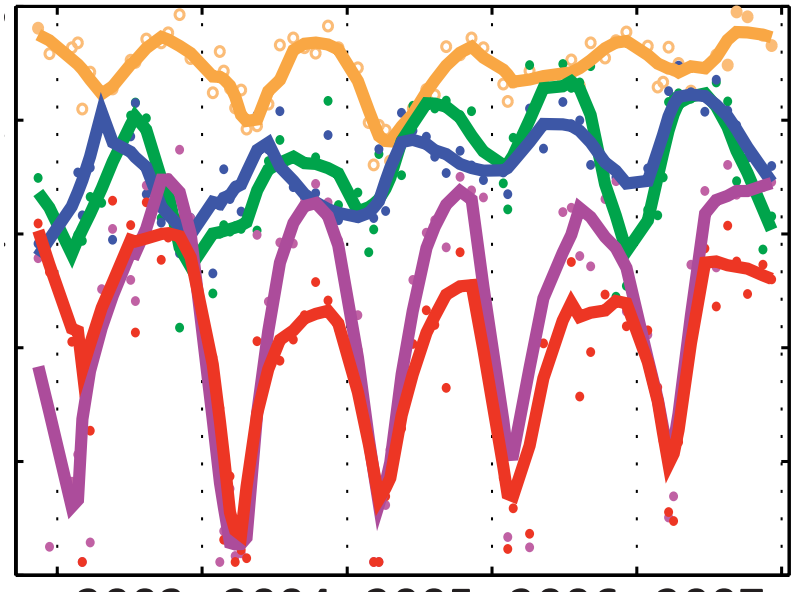


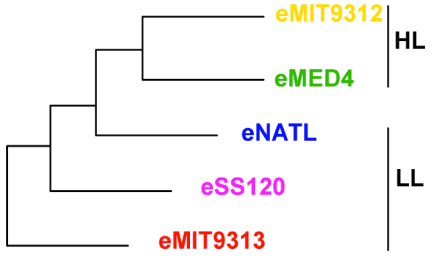
Johnson, Zinser et al, 2006

So there are meaningful layers, within layers, of diversity (leads to stability)

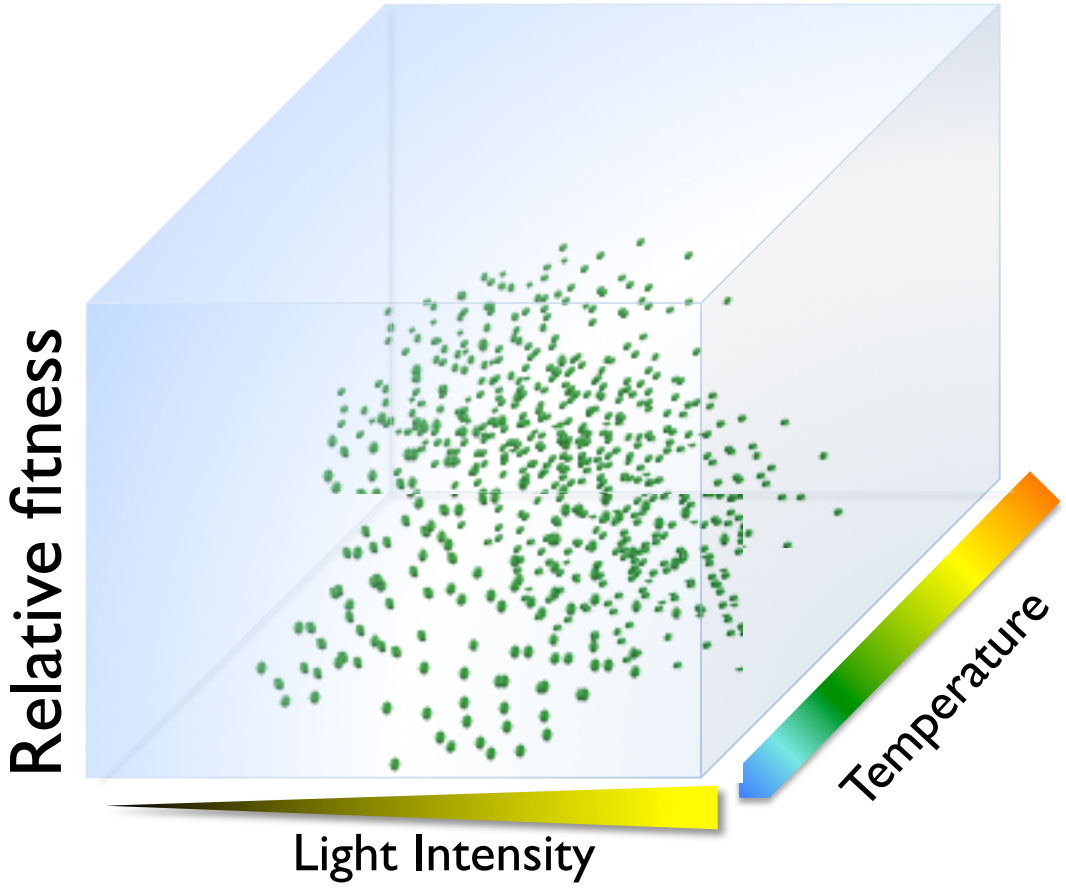


Prochlorococcus ecotypes





Beginning to understand *Prochlorococcus* niche differentiation in *two dimensions*



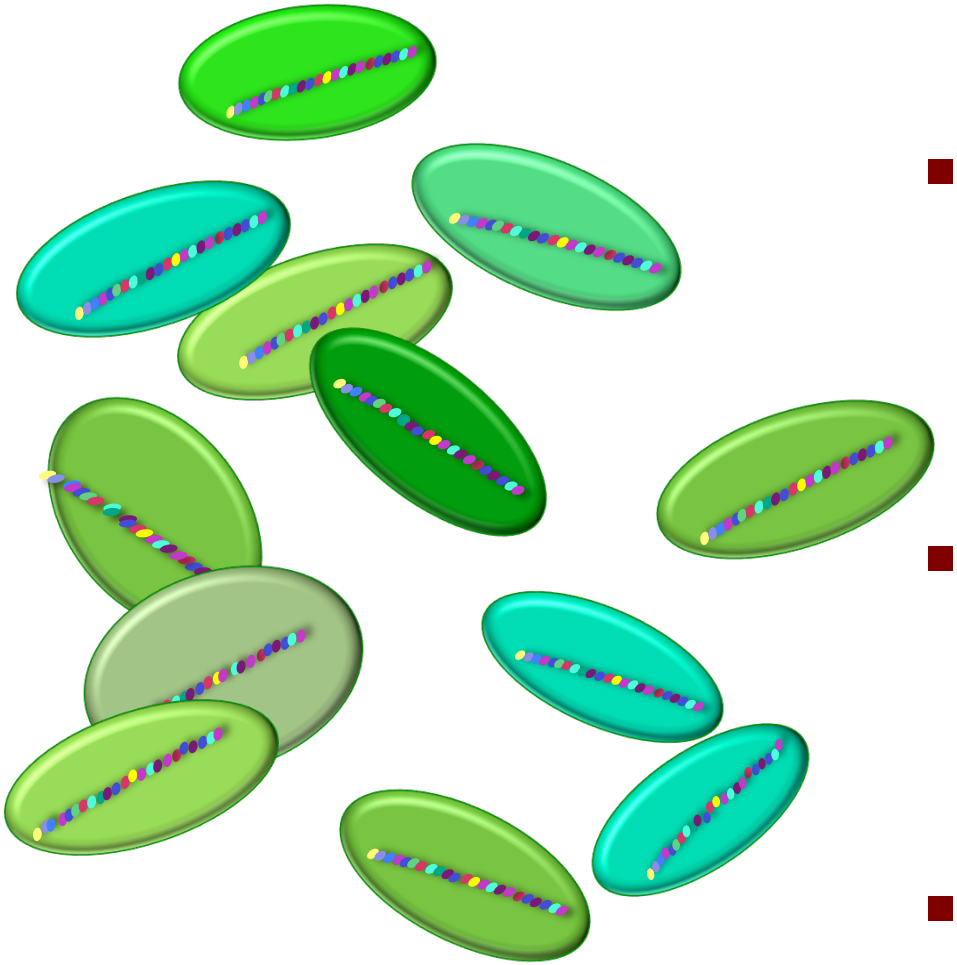
BUT Remember…

Ecological Niche:

***n**-dimensional*
hypervolume

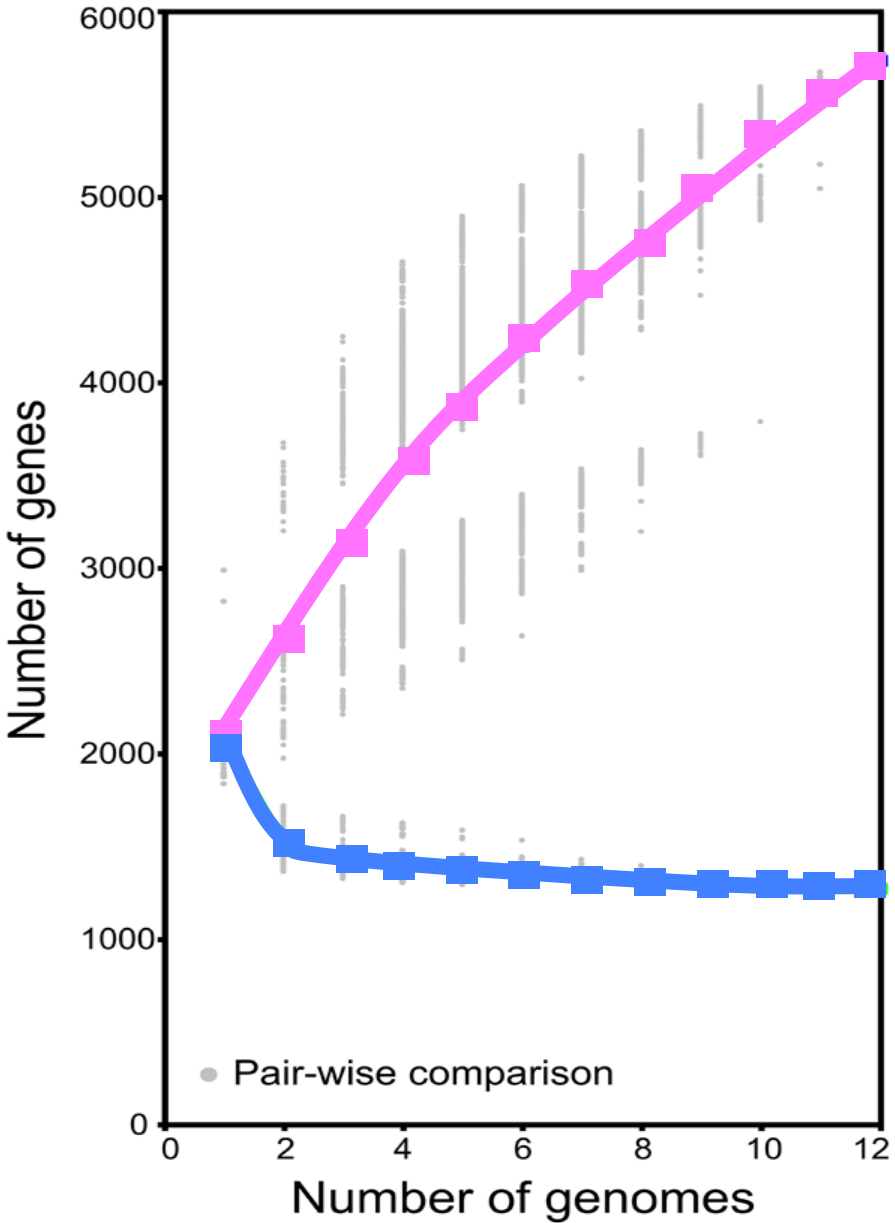
n-2 dimensions to go!

Enter Genomics:
Genomes of 13 *strains (now 43)*



- What genes are "core" i.e. shared by all?
- How many unique genes in the gene pool of *ALL Prochlorococcus*?
- Their functions?

The core and flexible genomes



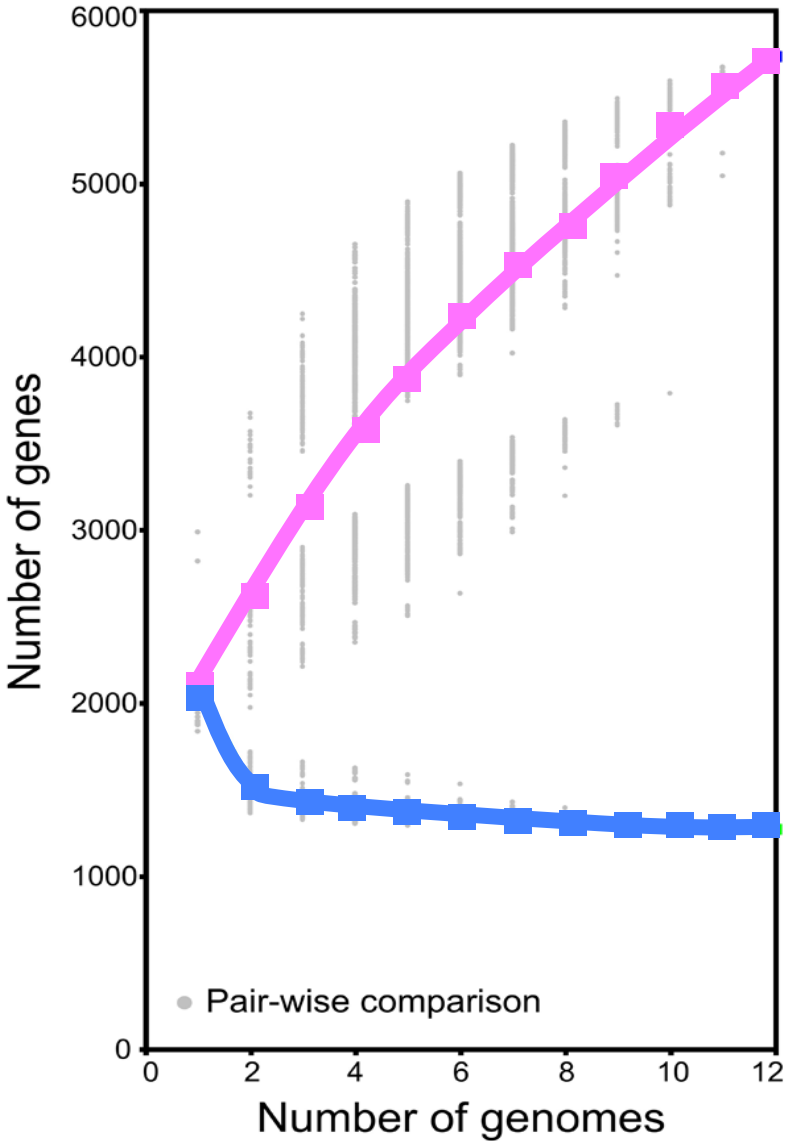
5736 'Flexible Genes' (and rising)

What is the global pan-genome?

1250 Shared Core Genes

Kettler, Martiny et al. 2007

Global Pan-genome is huge



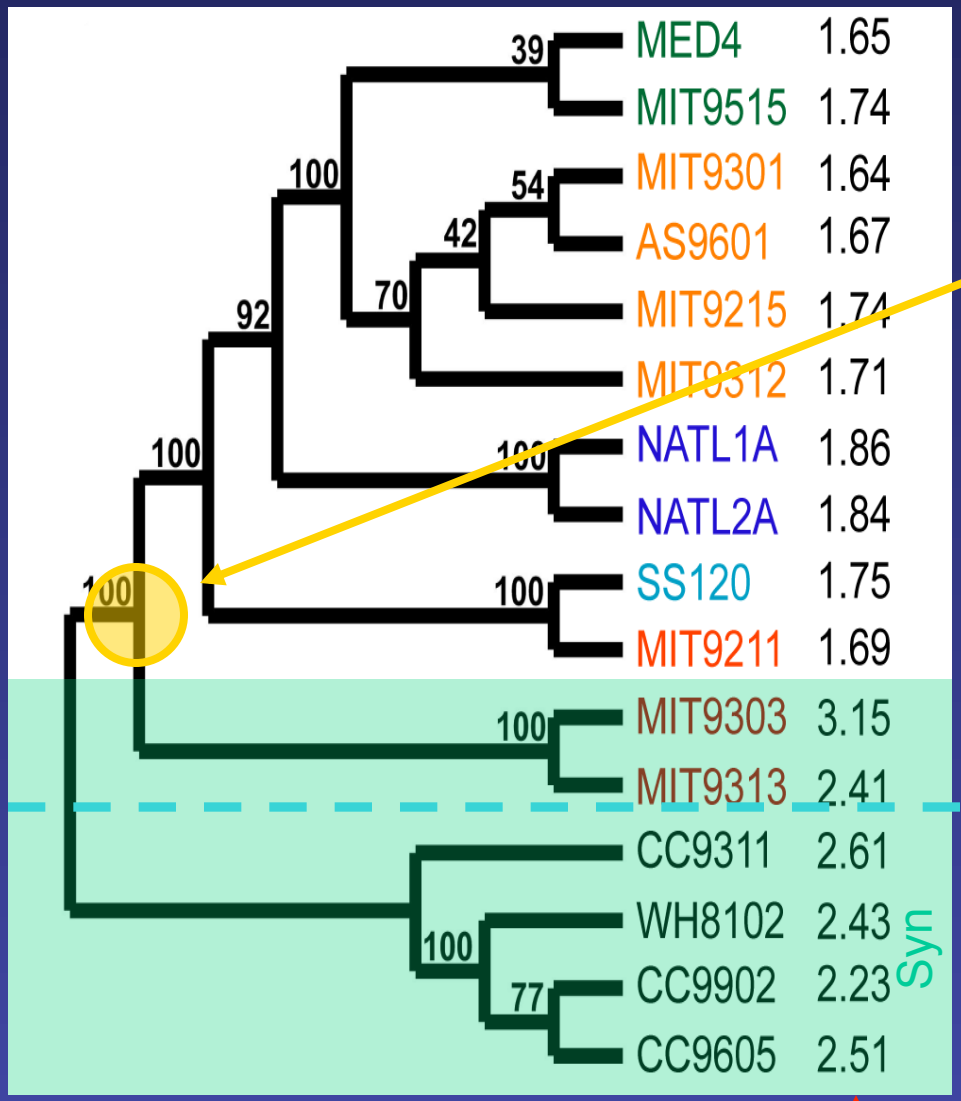
Total of 57,792 genes distributed among an *effective* population size of **10^{11} cells**

(Total number of cells in oceans is estimated at 10^{27})

1250 Core Genes

Baumdicker et al 2012. The infinitely many genes model for the distributed genome of bacteria. Genome Biol Evol. doi: 10.1093/gbe/evs016

Genome size in *Prochlorococcus* is variable



Systematic Genome Reduction? NO

small genomes

Genes are gained and lost – with most of the action in the "leaves of the tree"

large genomes

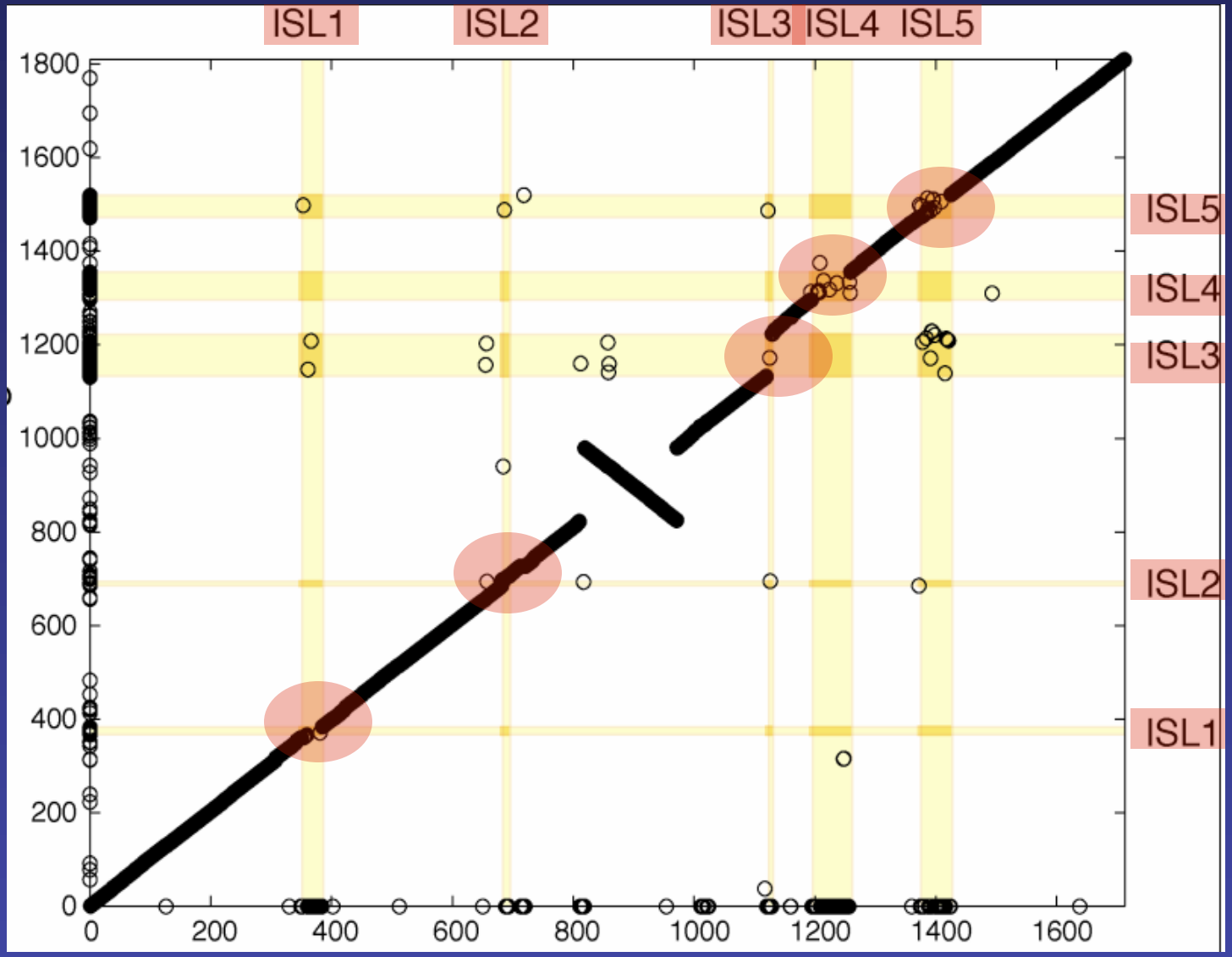
Genome Size (mb)

Kettler, Martiny et al. 2007

1574 Shared genes

1810 genes

1713 genes

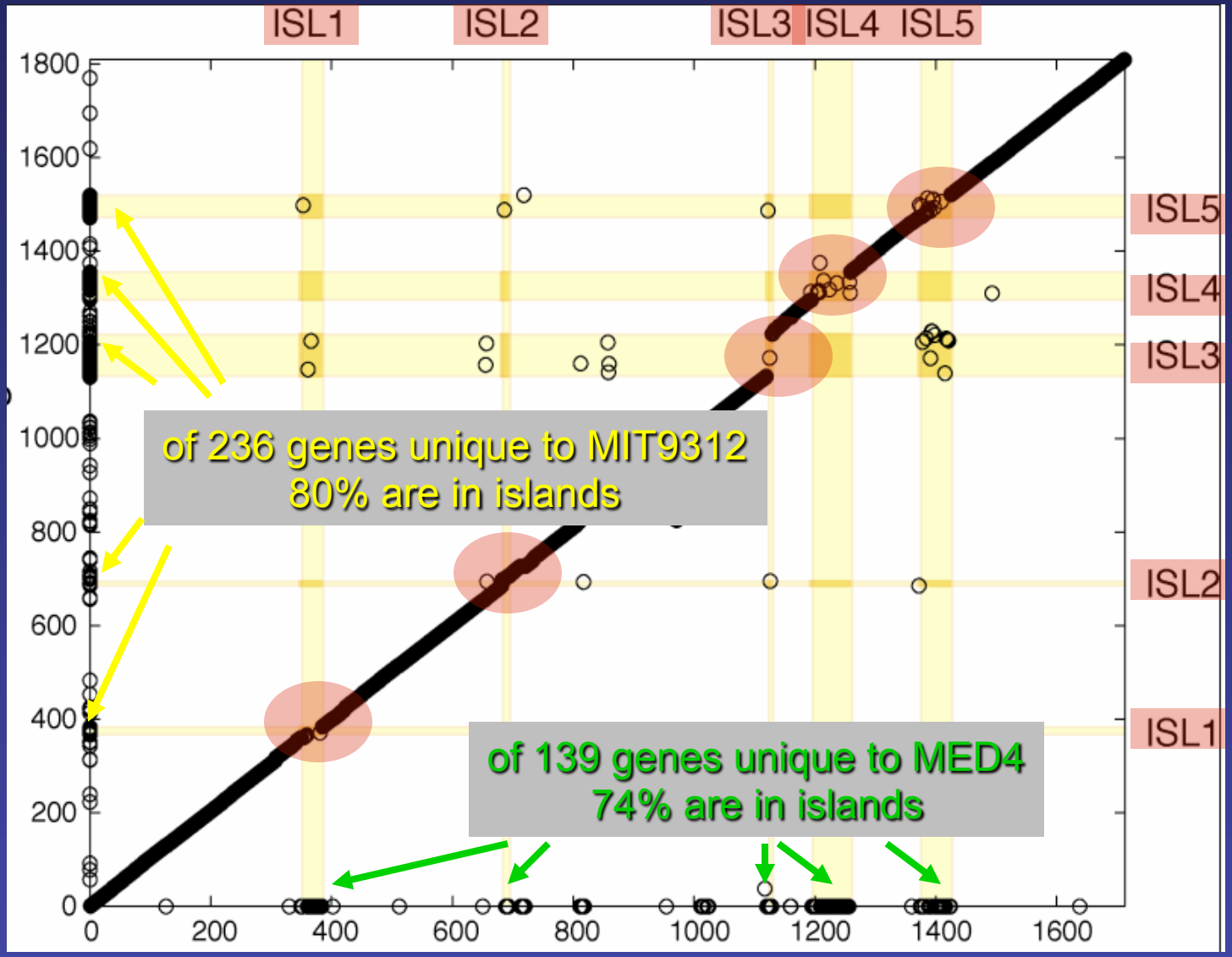


Coleman et al. 2006

1810 genes

1574 Shared genes

1713 genes



MIT9312 genome position

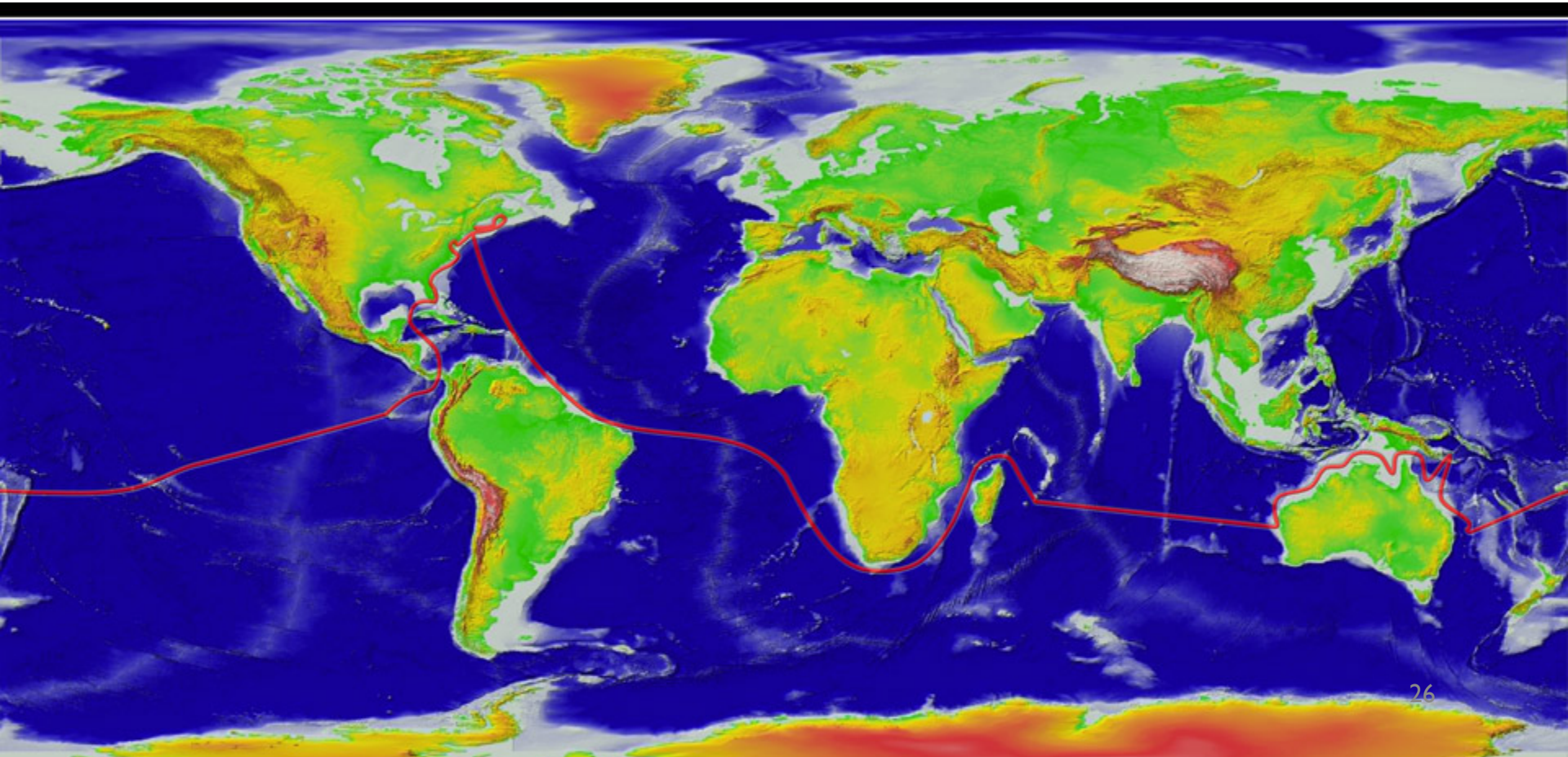
MED4 genome position

Phage are implicated as playing a role in island gene dynamics

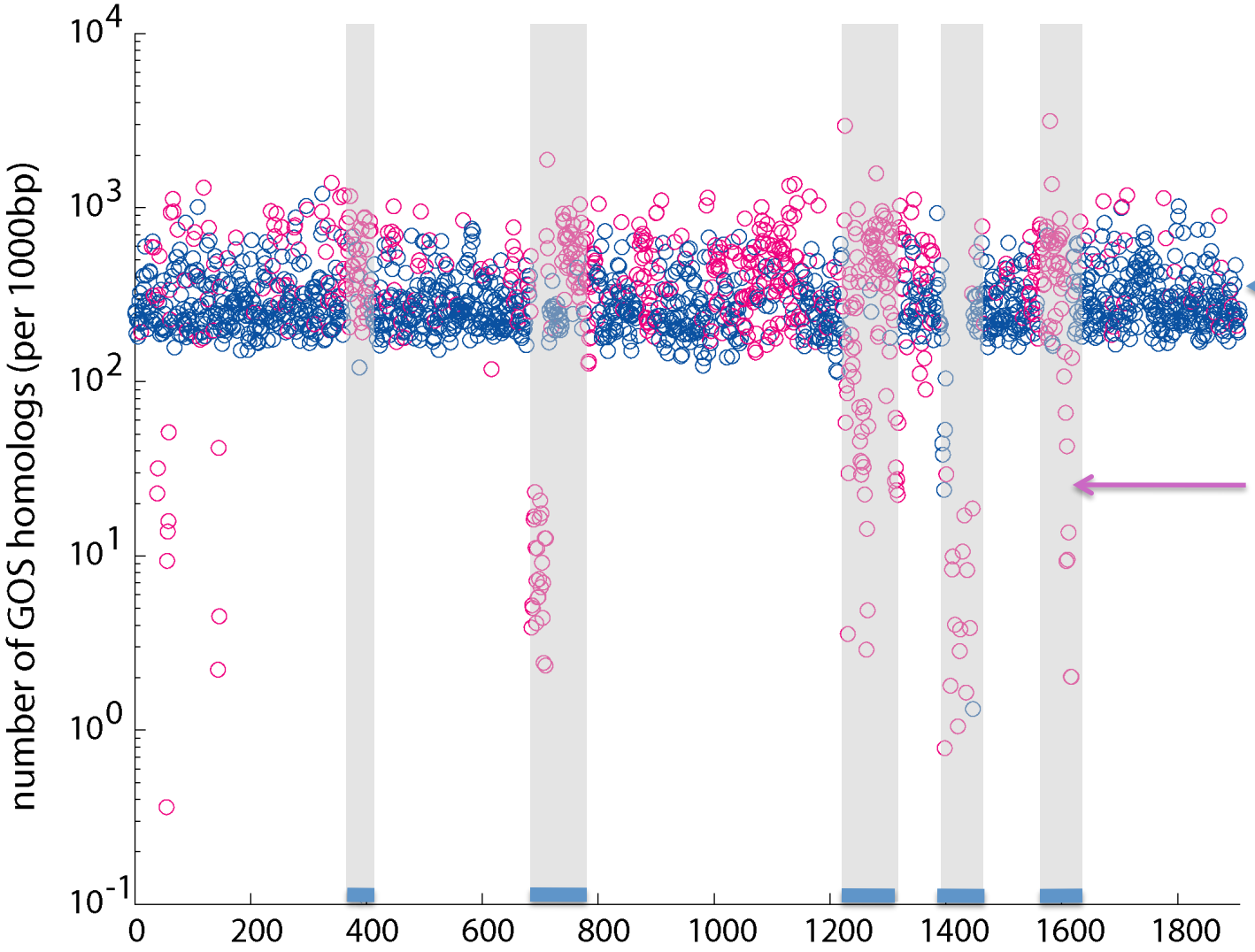
Coleman et al. 2006

Prochlorococcus genes are abundant in the Global Ocean Survey (GOS)

> 10% of the sequenced fragments are from *Prochlorococcus* at many stations



Core and flexible genes, and islands, visible in metagenomic data



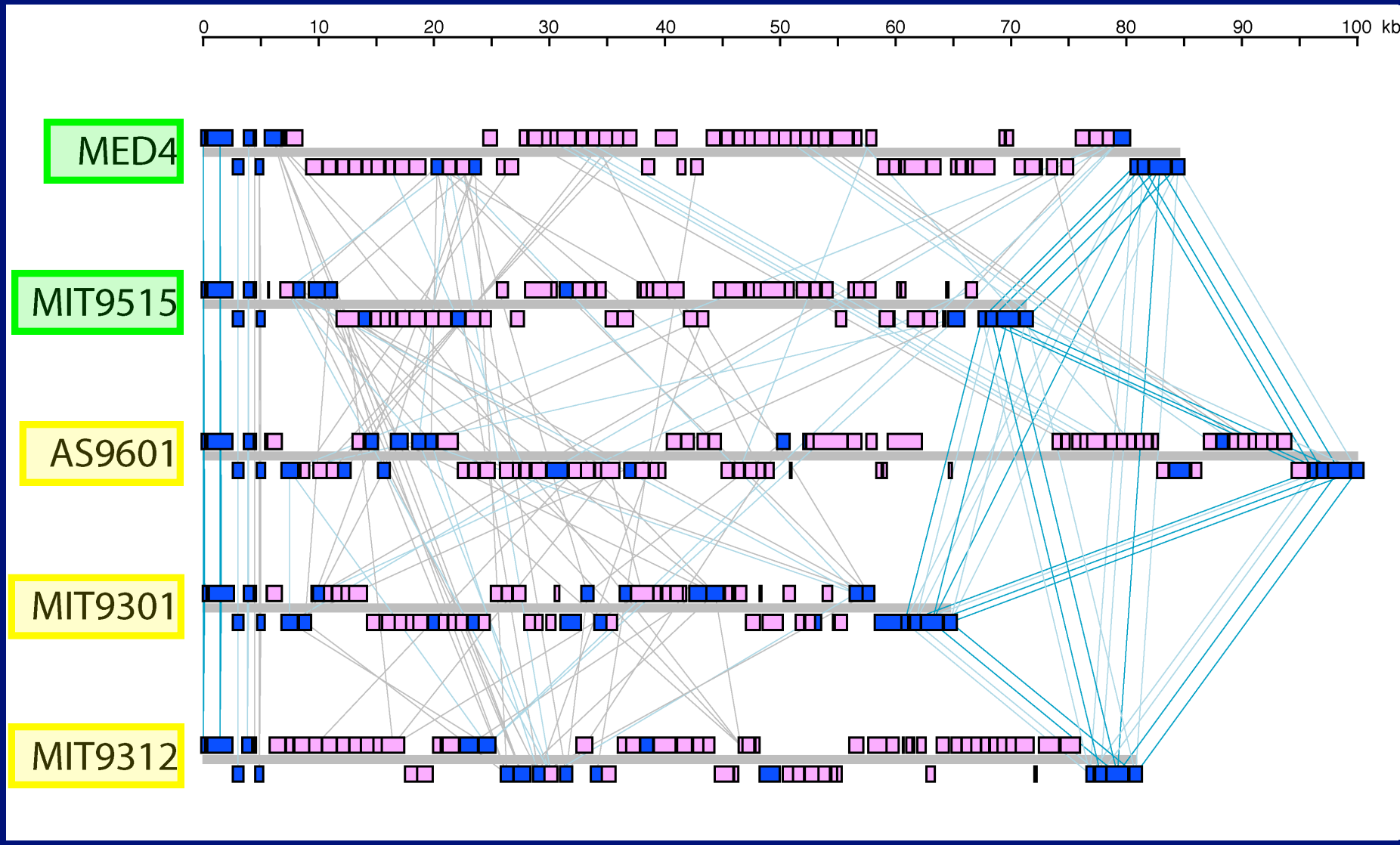
Core genes present in one copy per cell

flexible Genes more variable

Known Islands in cultured genome

Kettler, Martiny et al. 2007

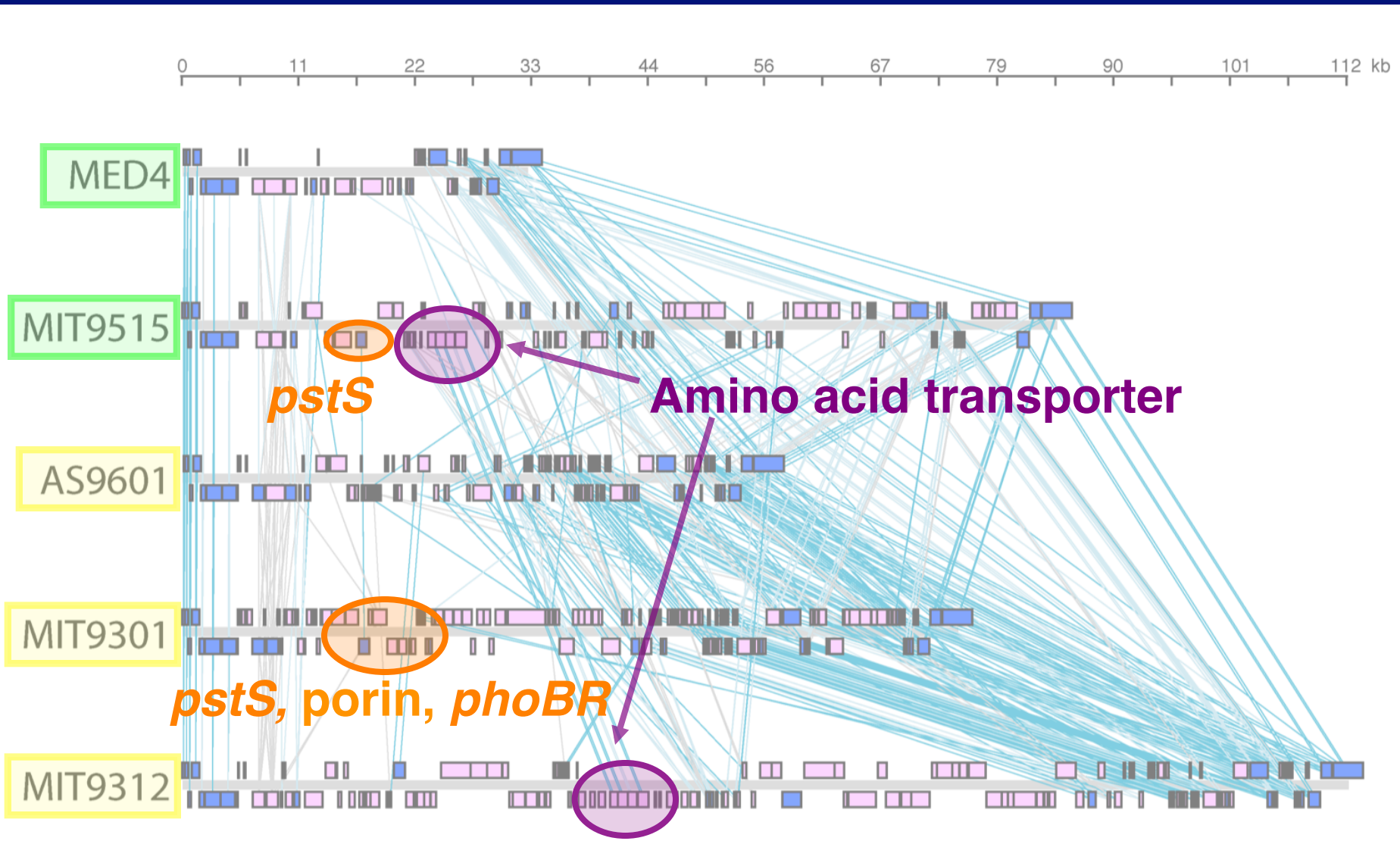
Island 4 - LPS Biosynthesis Genes



Found in all genomes

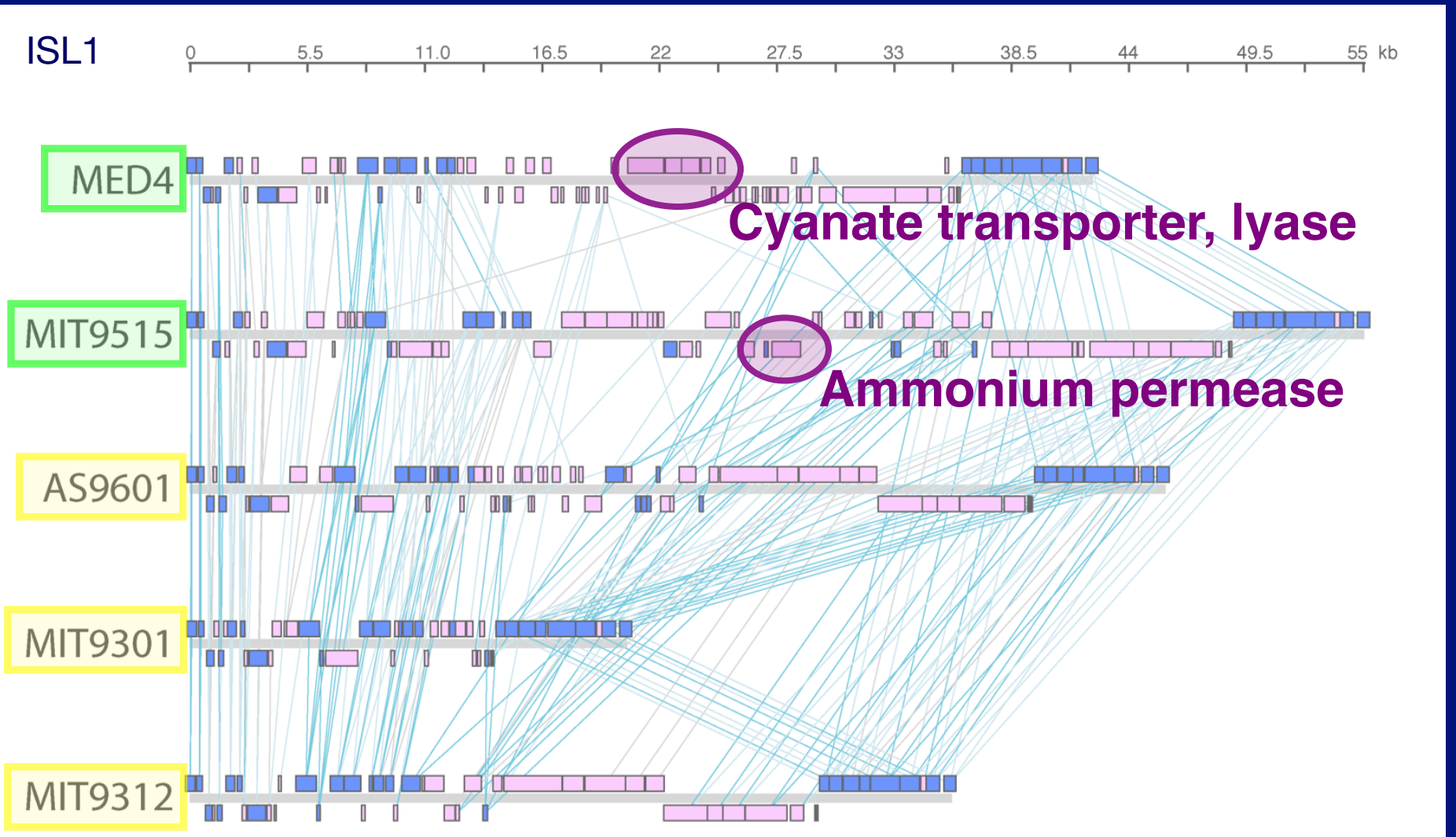
Coleman

Island 3 – Genes involved in N and P acquisition



Coleman

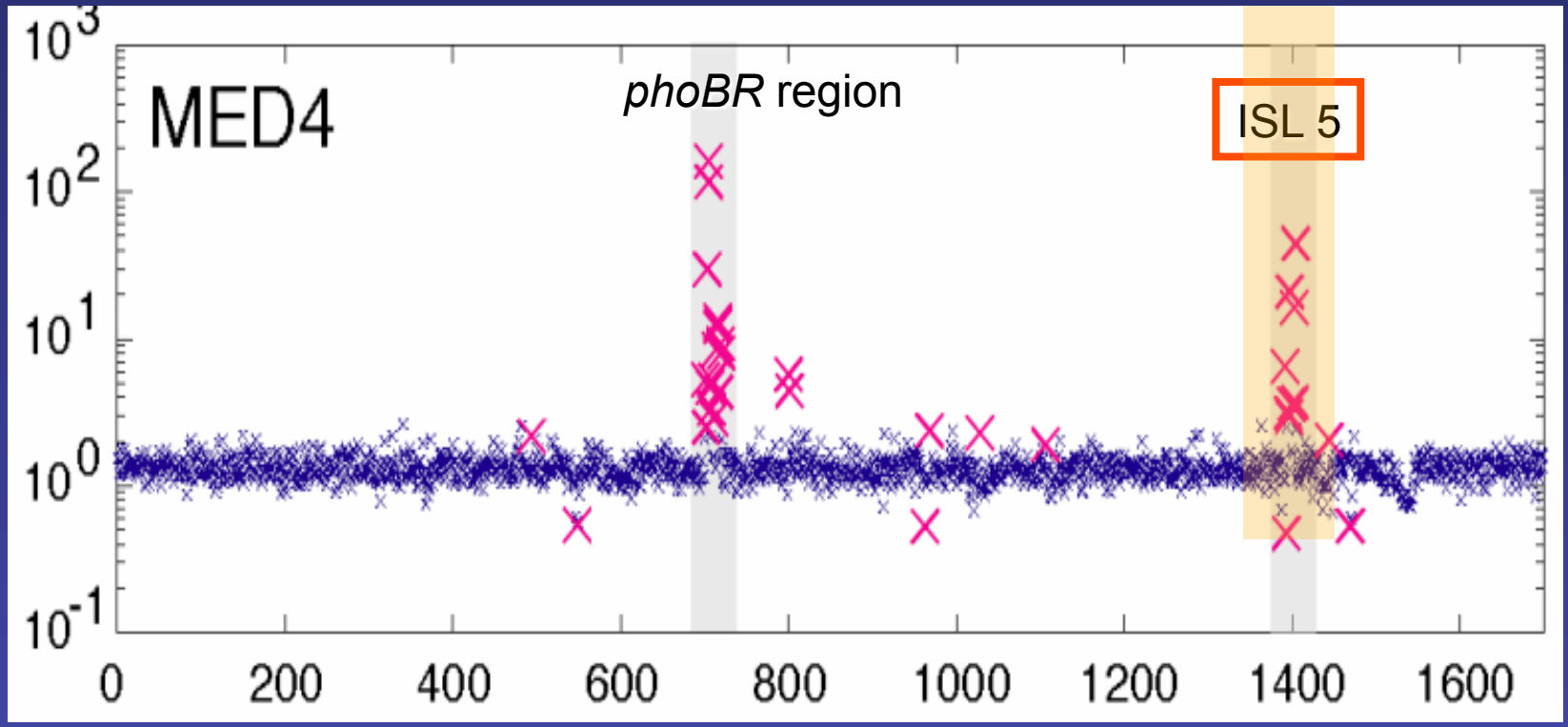
Island 1 – Genes involved in N acquisition



Coleman

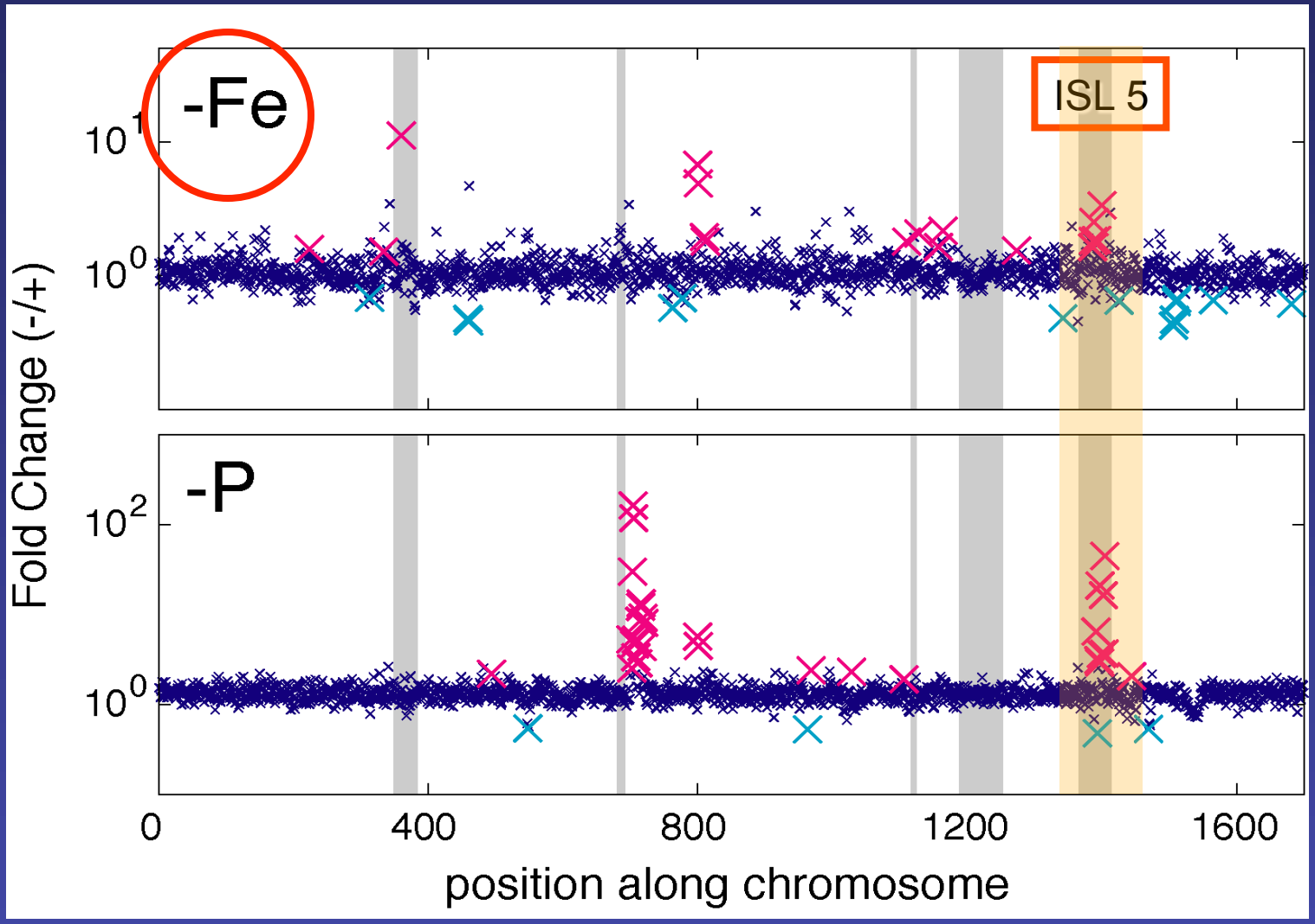
Island 5 genes (of unknown function) are upregulated under P-starvation in MED4

(but not in MIT9313)



Martiny/Coleman et al 2006

ISLAND 5 Genes are also upregulated under iron starvation



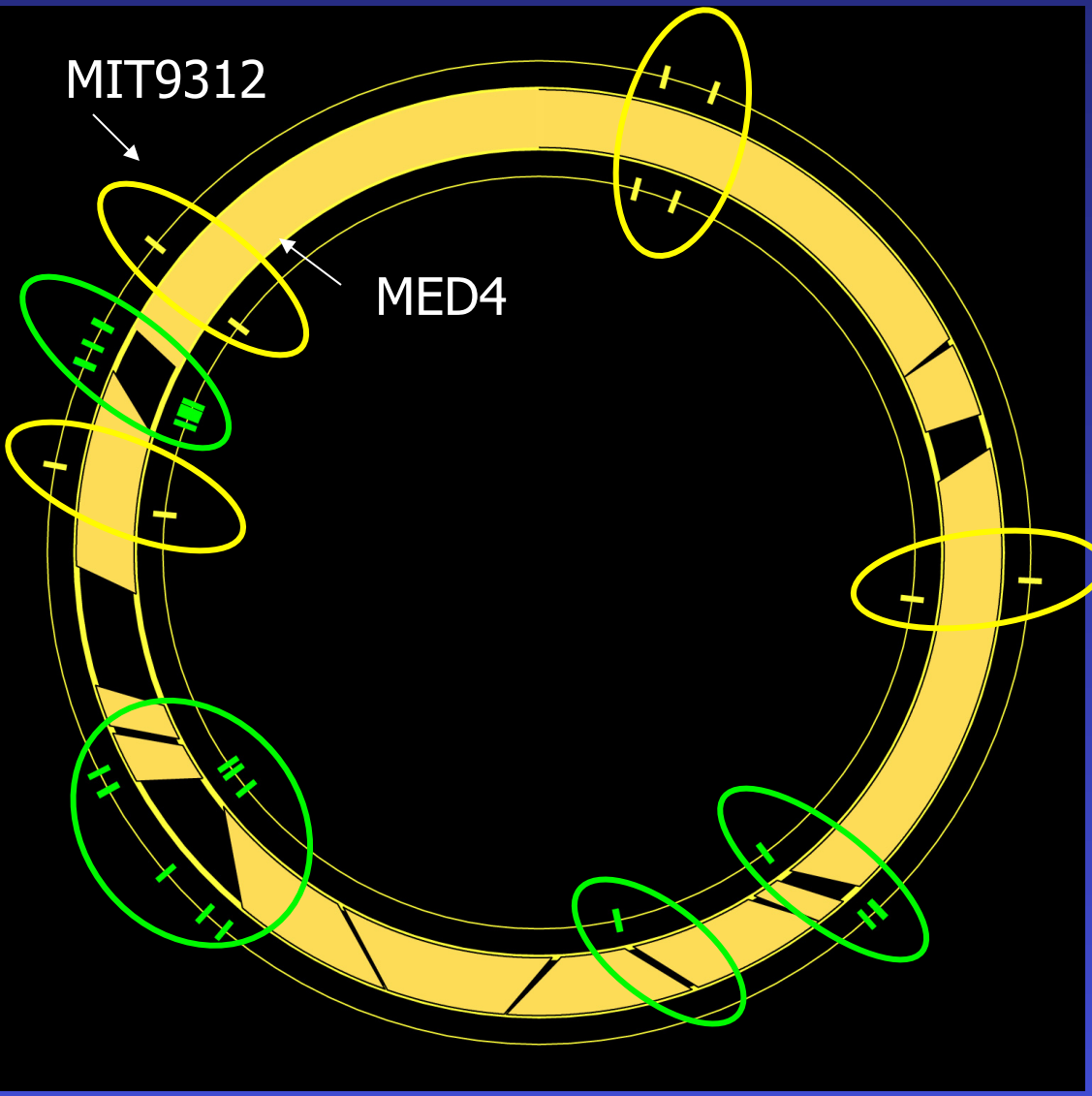
Thompson et al 2011

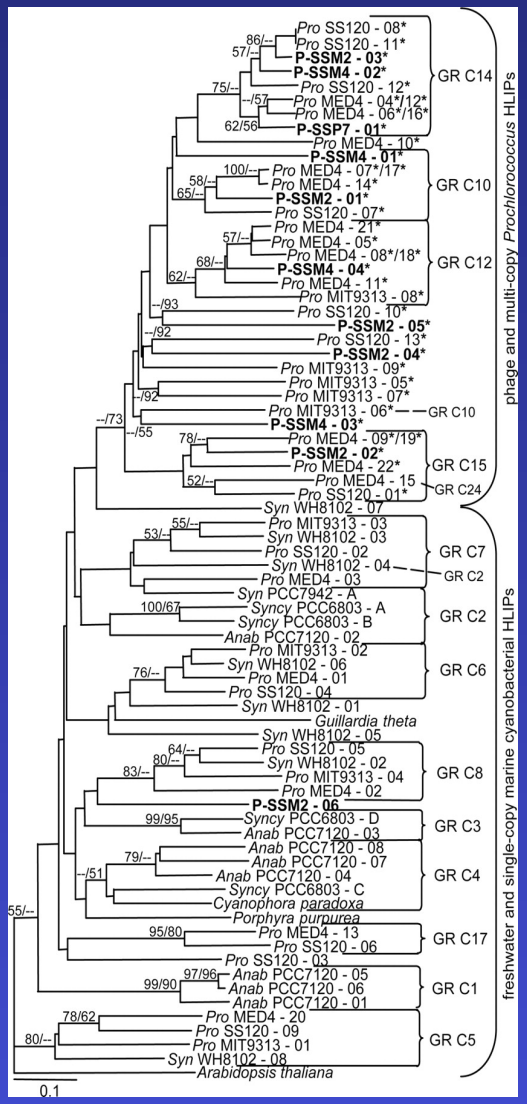
Components of two lectures

- The Cell
- Niche Dimensions of *Prochlorococcus*

  More on nutrients next time
- Single Cell Genomics
- Phage
- Interactions (and signaling?)

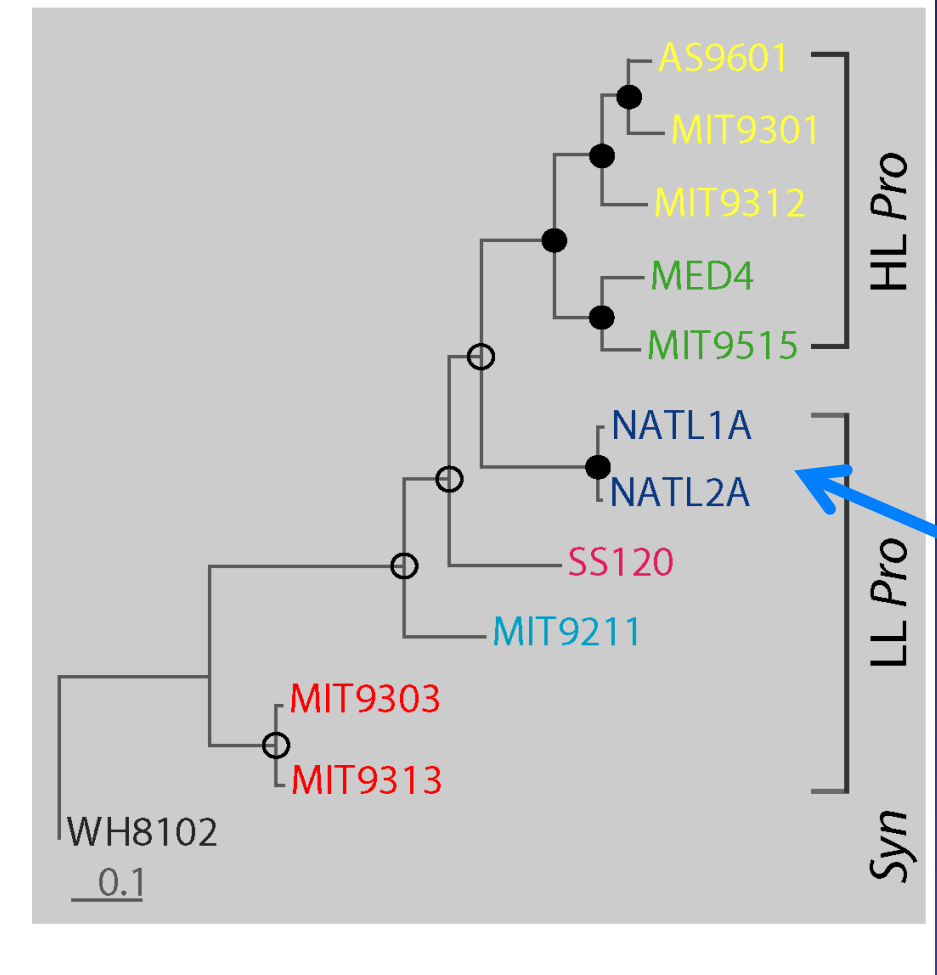
Phage-related HLIPs may play a role in island insertion





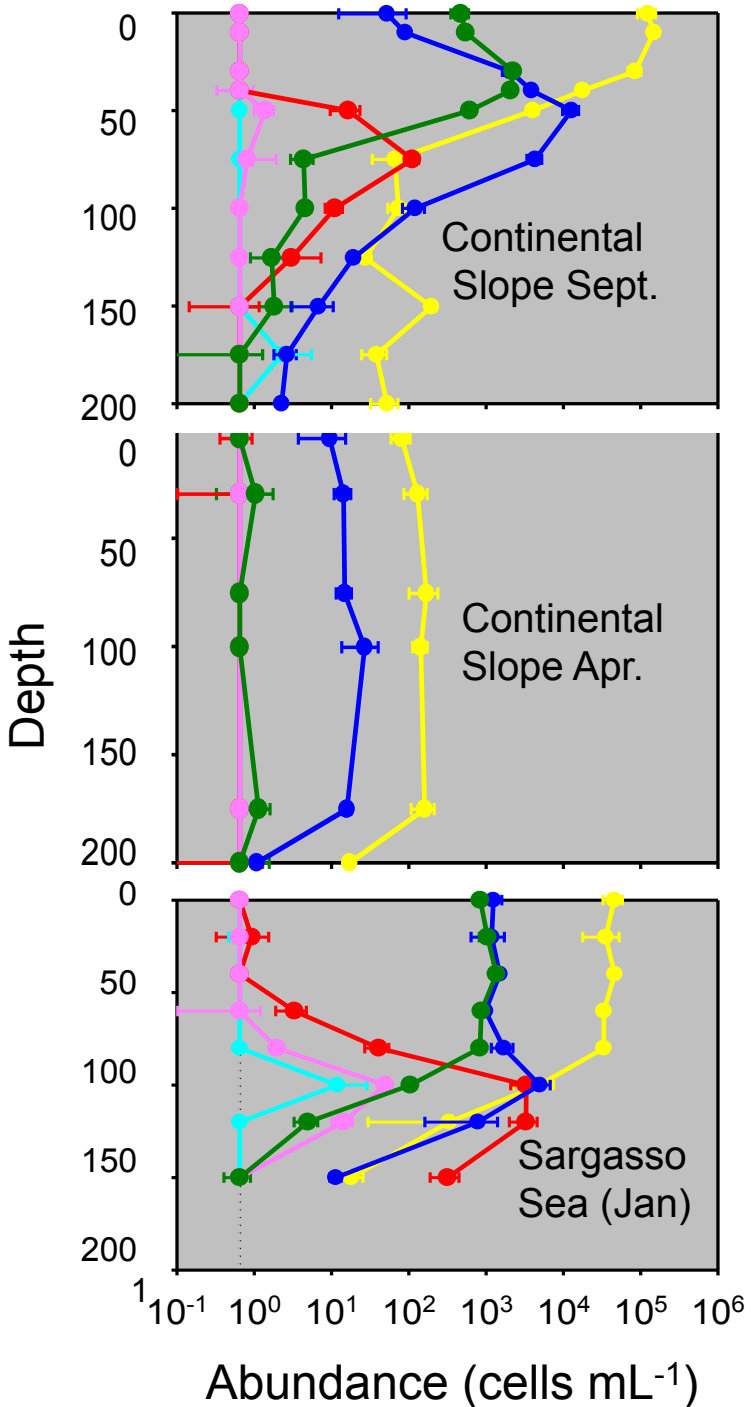
Lindell, Sullivan et al. PNAS 2005

Speaking of HLIPs…



"The NATLs" are very interesting in this regard

(note intermediate taxonomic position)

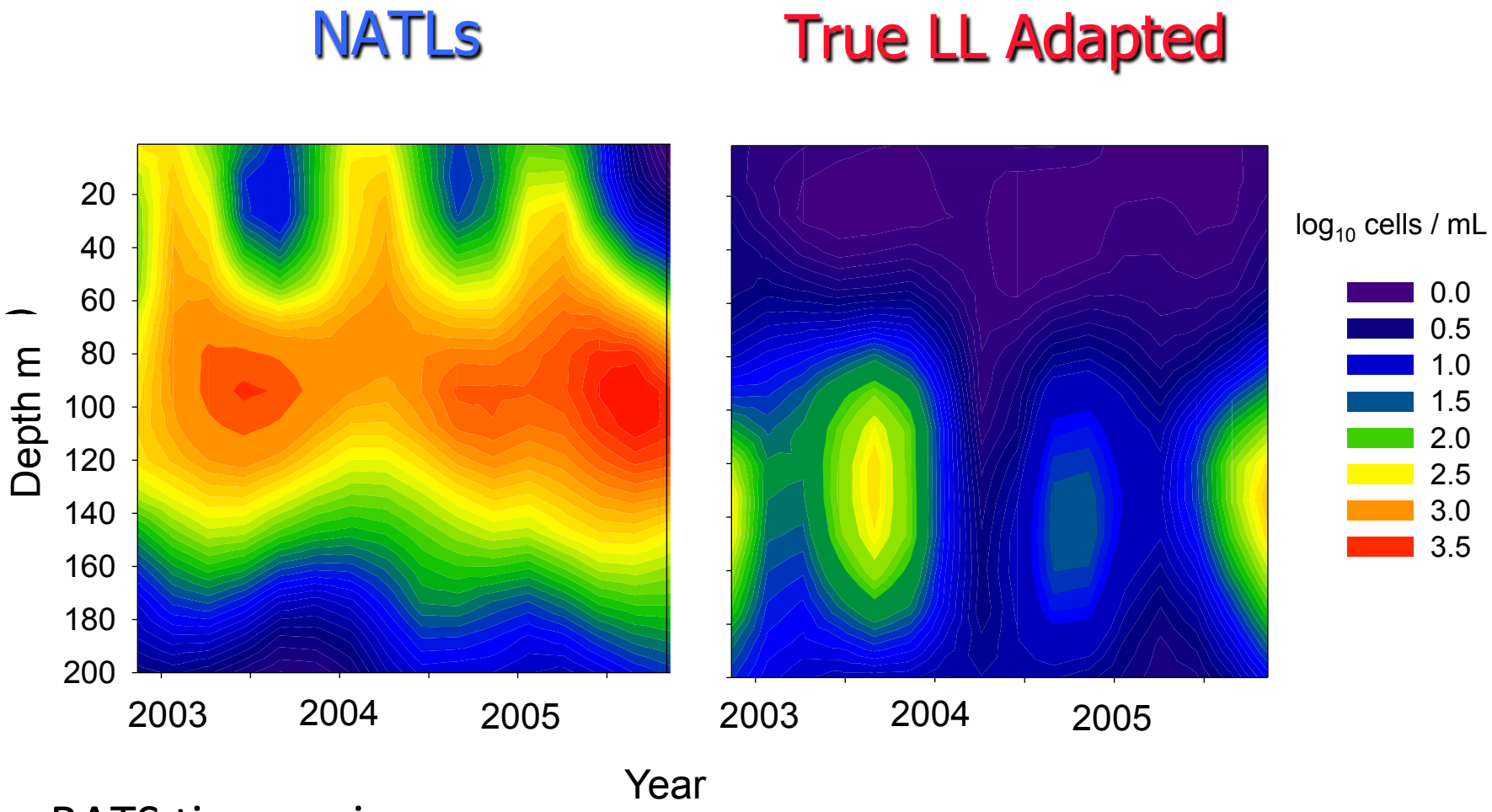


Handle high light better than other LL ecotypes…

…and deep mixing better than some HL ecotypes

Zinser et al 2007

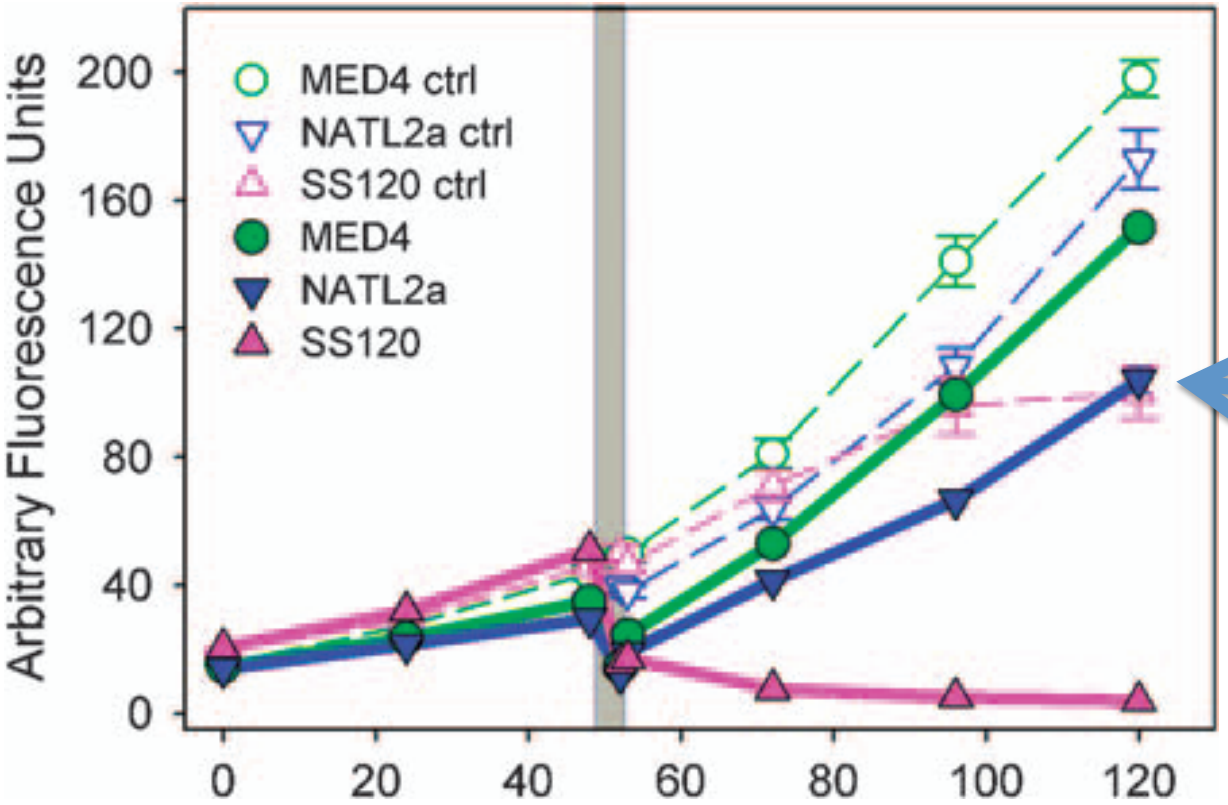
HYPOTHESIS: NATLs better adapted to fluctuating light



BATS time series

Malmstrom et al 2010

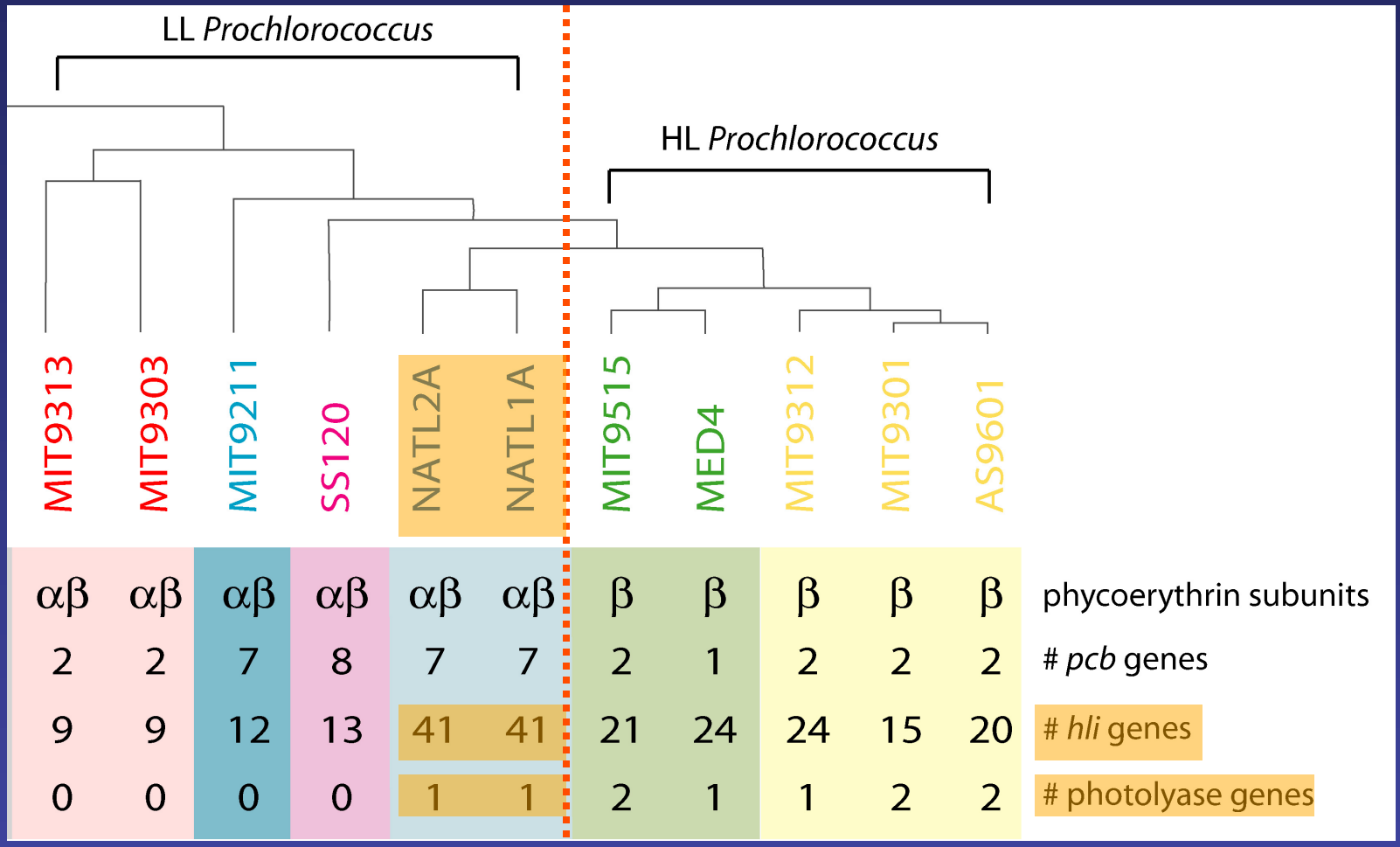
Back to the lab…



"The NATLs" are more resistant to light shock than other LL strains

Malmstrom et al 2010

What do the genomes tell us?



Coleman and Chisholm 2007

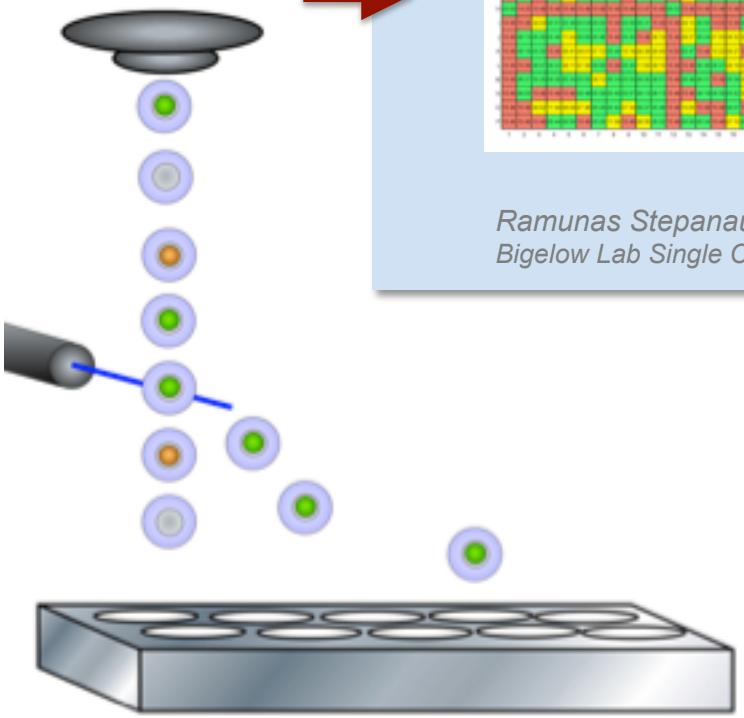
Components of two lectures

- The Cell
- Niche Dimensions of *Prochlorococcus*

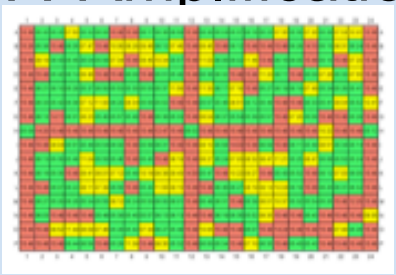
  More on nutrients next time
- Single Cell Genomics
- Phage
- Interactions (and signaling?)

Single Cell Genomics Pipeline

Flow Sort



MDA Amplification



Ramunas Stepanauskas
Bigelow Lab Single Cell Facility

Amplify and sequence identifier gene of interest:
eg. *rRNA ITS, narB, etc.*

Select candidates for sequencing

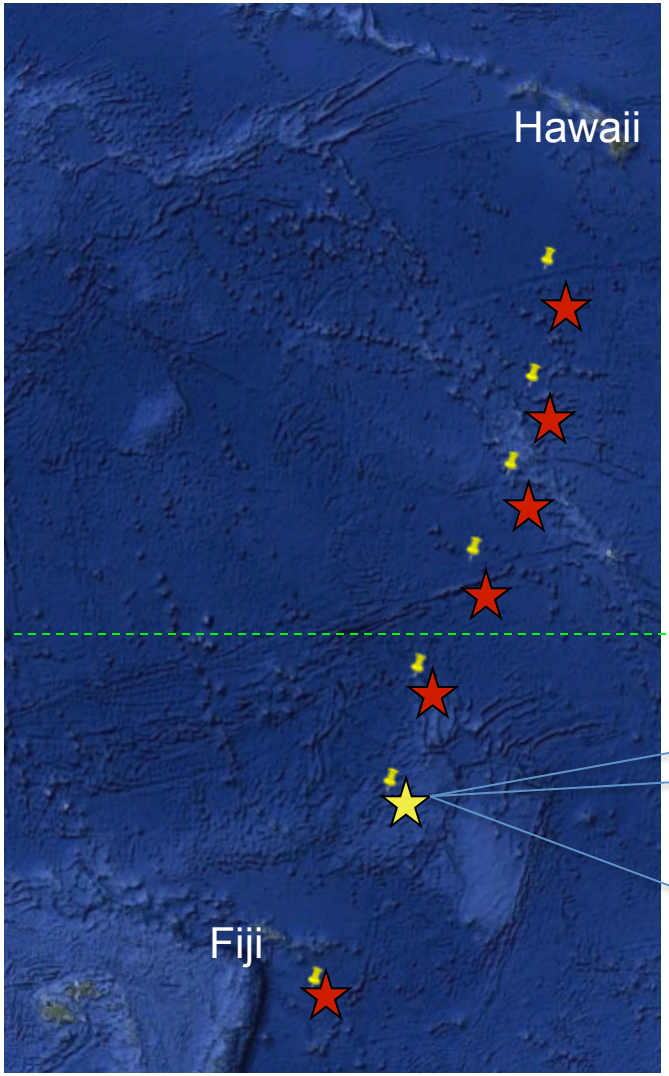
Re-amplify and make sequencing libraries

Illumina Sequencing

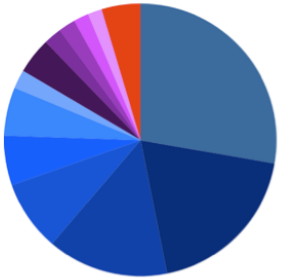
Assembly

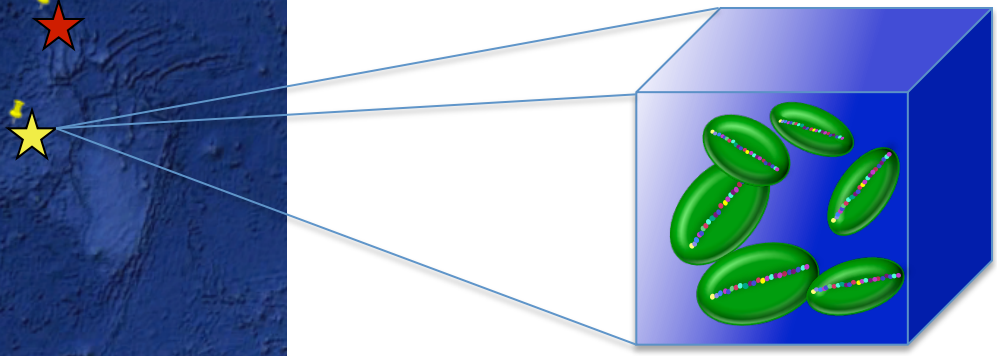
Rodrigue, Malmstrom, et al 2009

Genomes of 5 Wild *Prochlorococcus* Cells



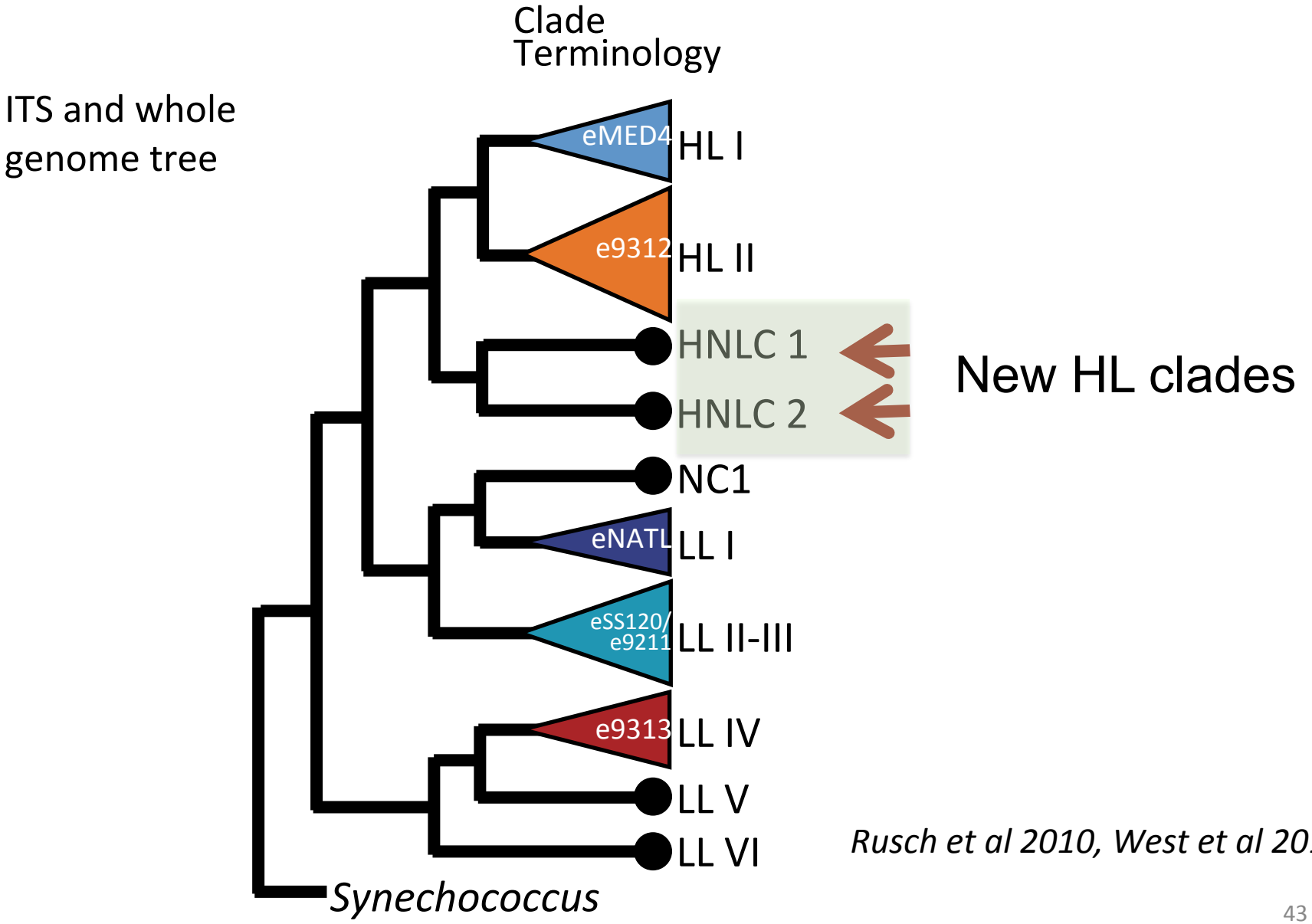
- Expanded recruited GOS reads by 15%
- Added to **641 new genes** to *Prochlorococcus* pan-genome
- New functions





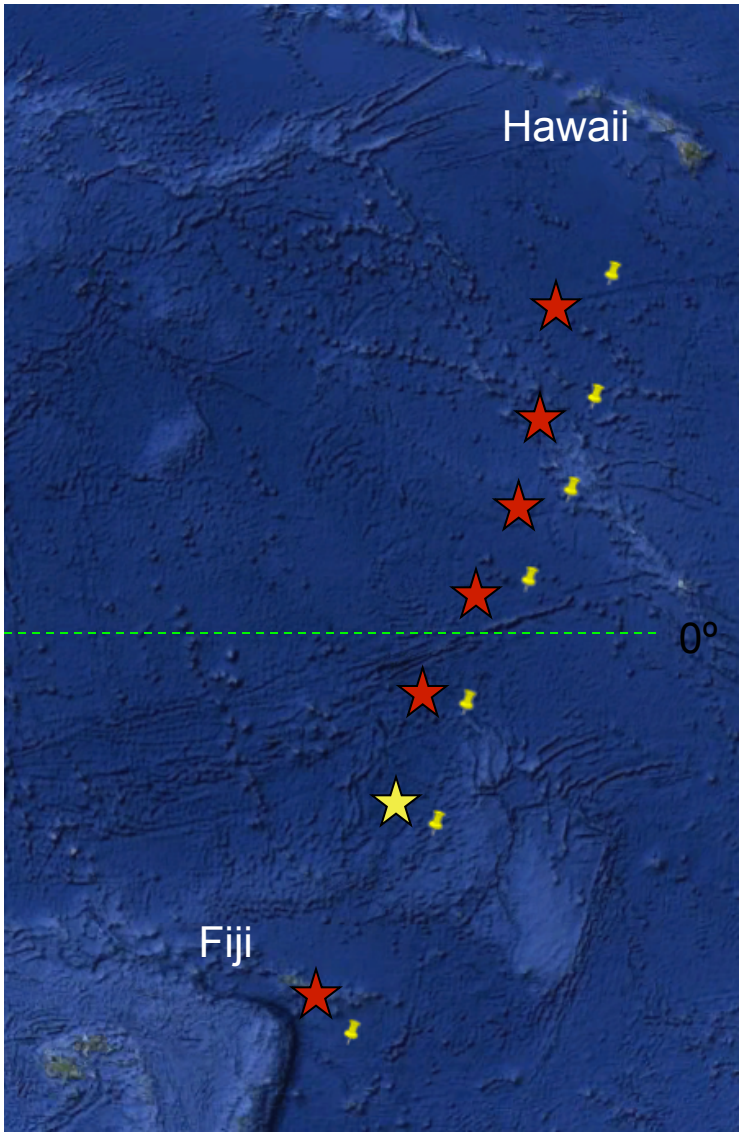
Malmstrom, Rodrigue, 2012 in revision

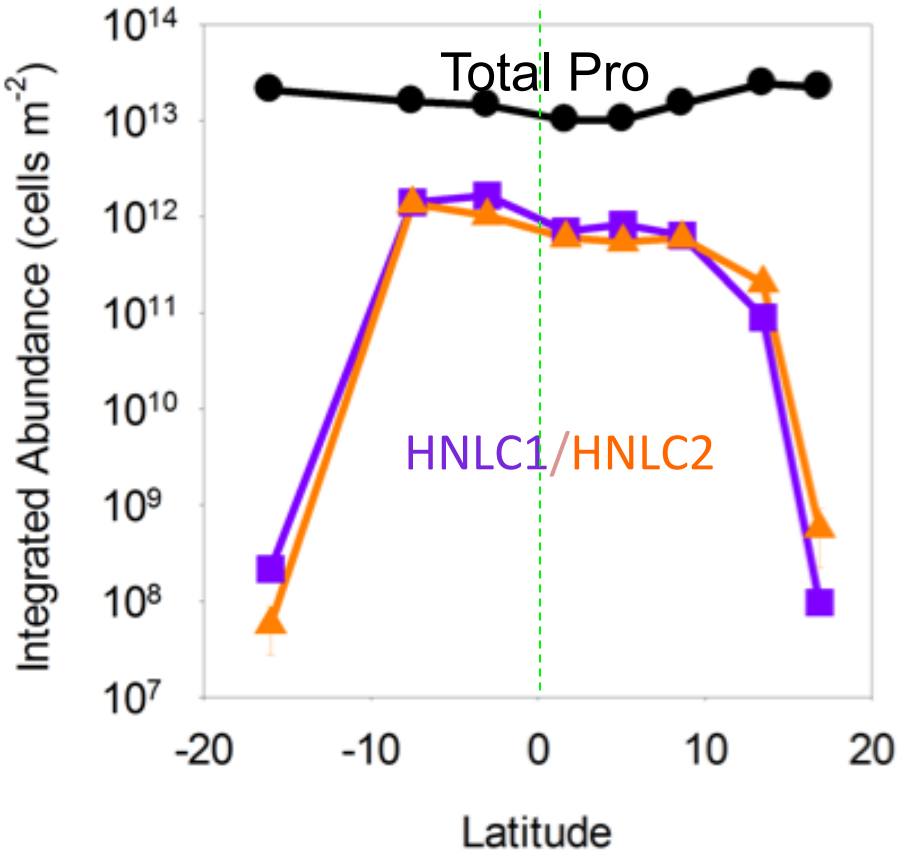
Two new clades



Rusch et al 2010, West et al 2011

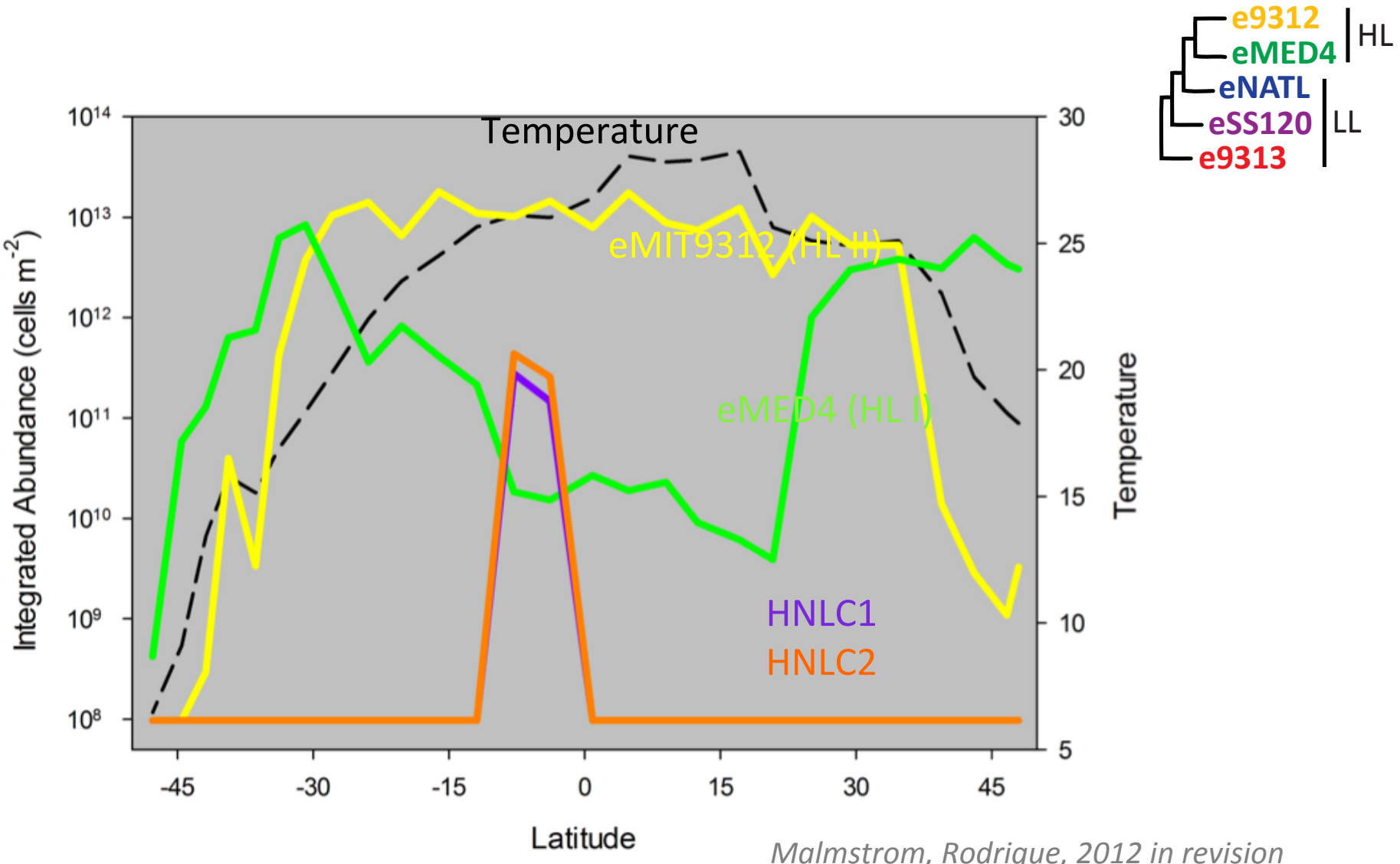
New HL clades restricted to equatorial waters





Malmstrom, Rodrigue, 2012 in revision

What is their distribution in the Atlantic?



Malmstrom, Rodrigue, 2012 in revision

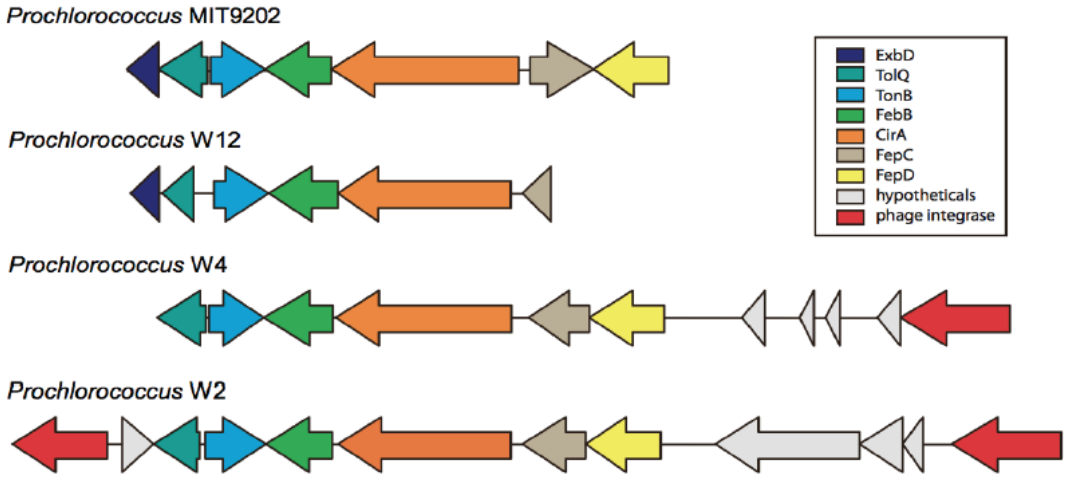
Siderophore transport genes found in wild cells, and in one of our cultures (9202)

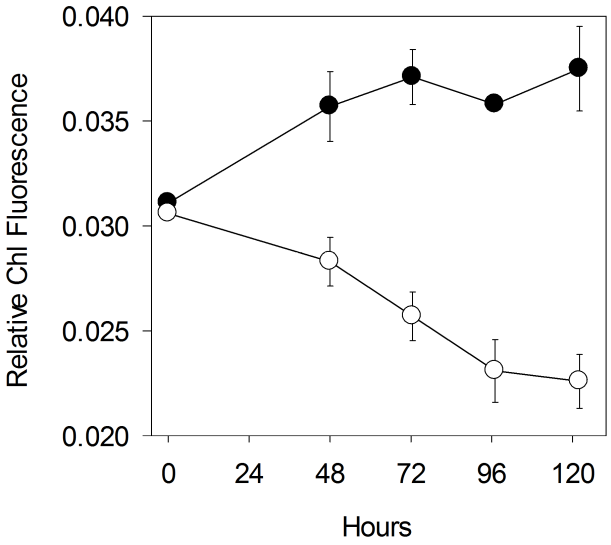
Culture

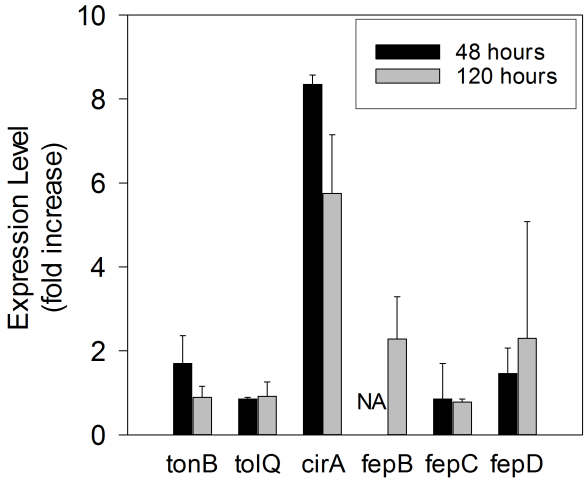
Wild single cell

Wild single cell

Wild single cell



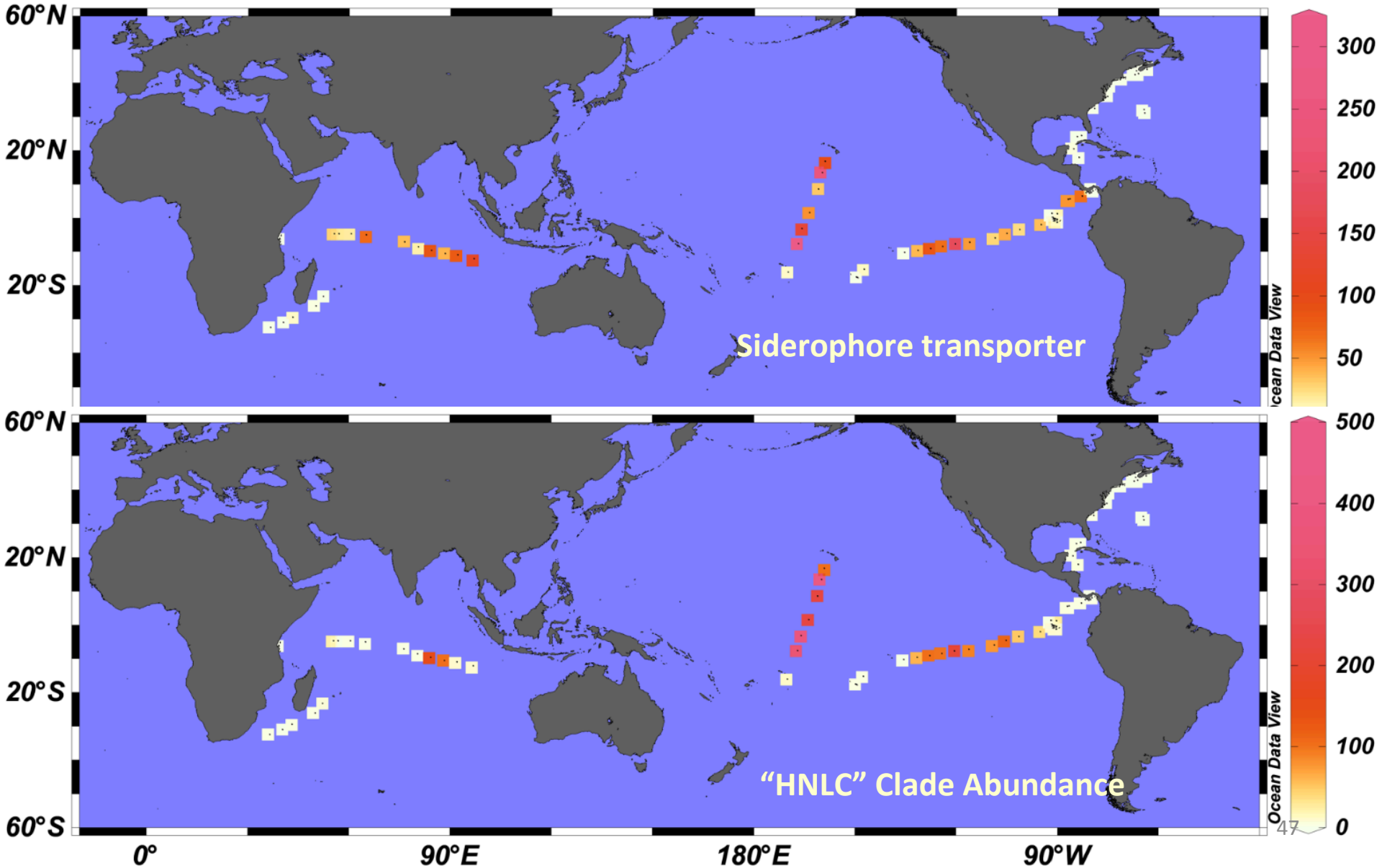




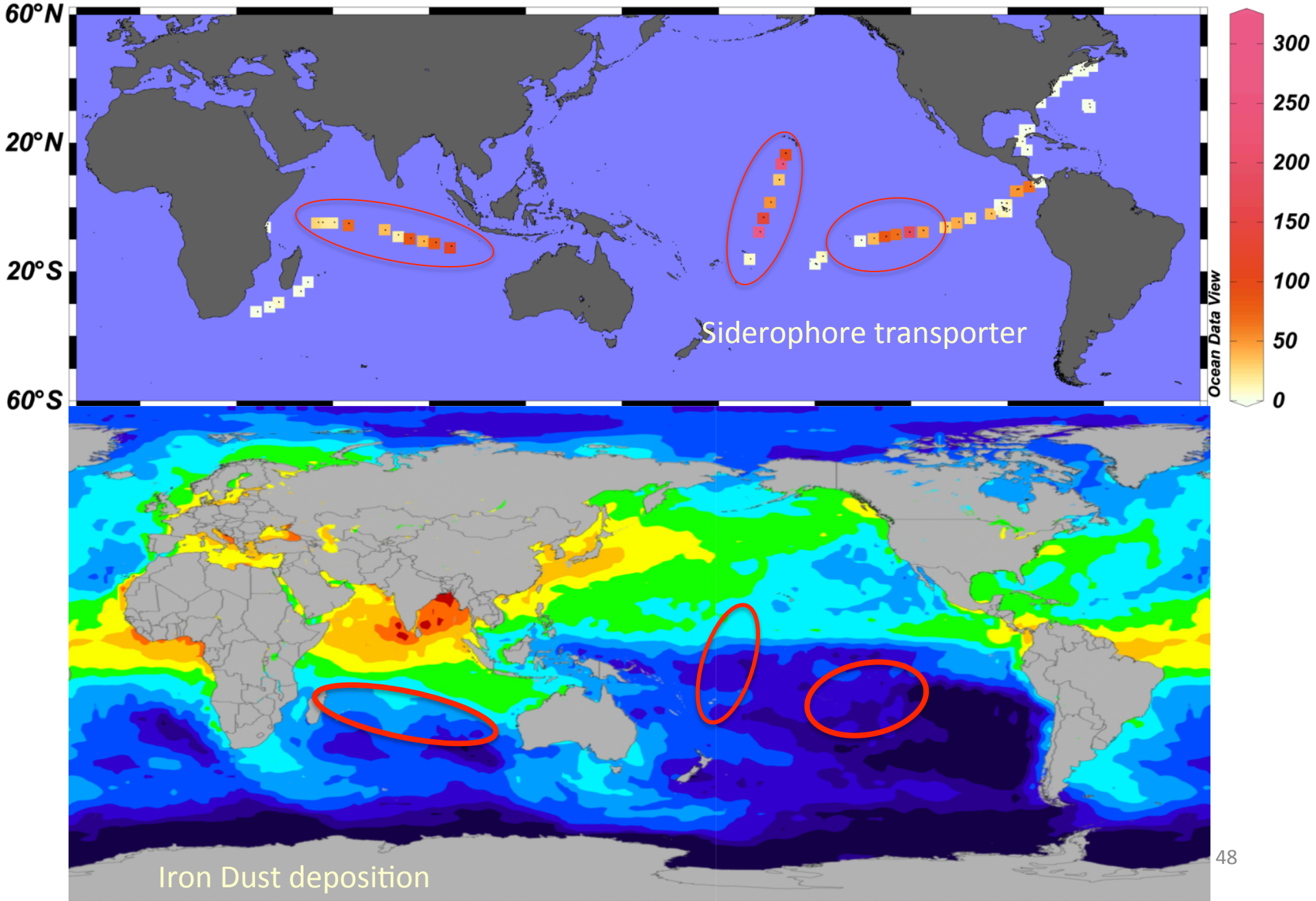
Genes are expressed under conditions of iron starvation in cultures

Malmstrom, Rodrigue, 2012 in revision

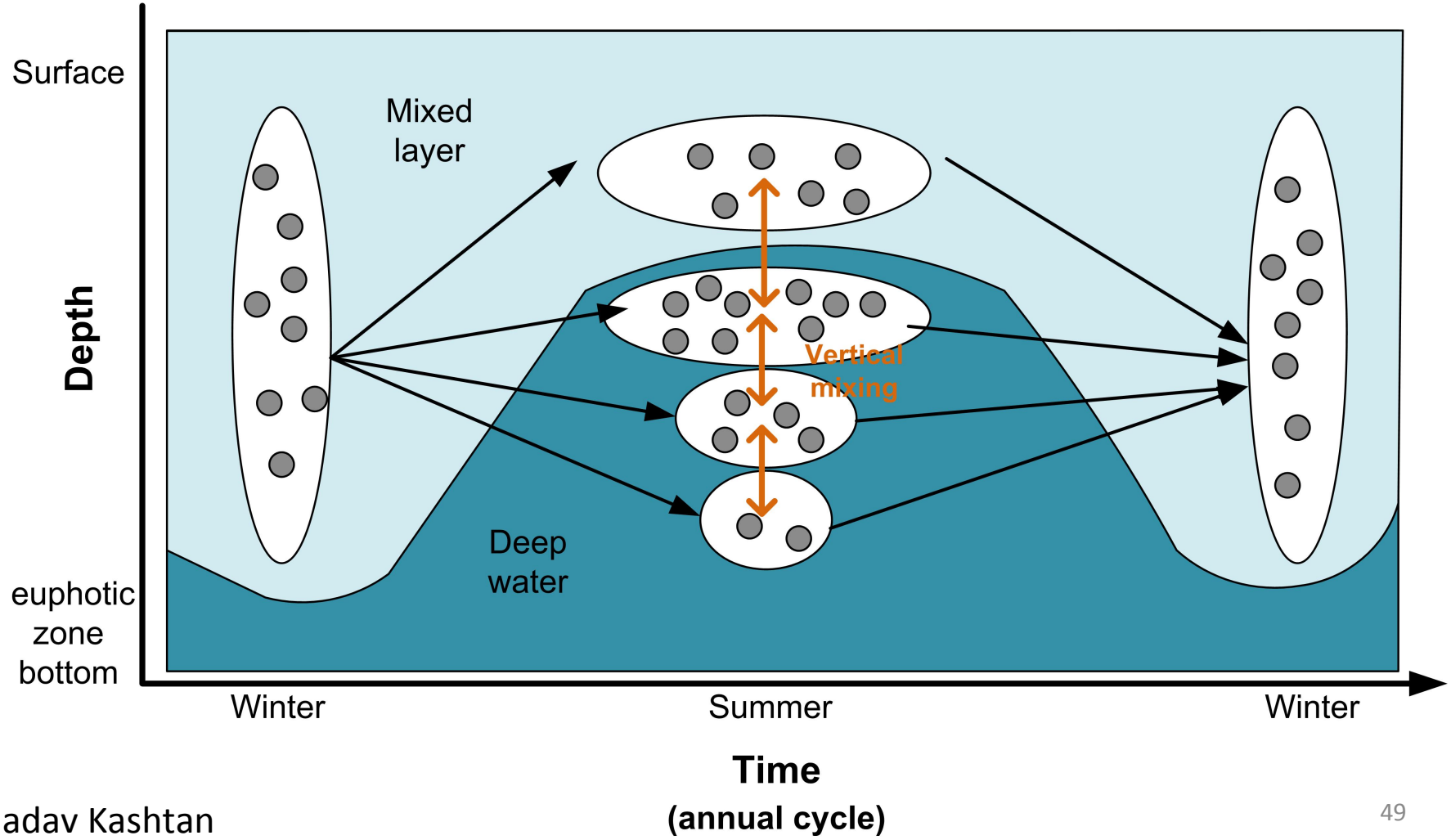
Global distributions through GOS metagenomics



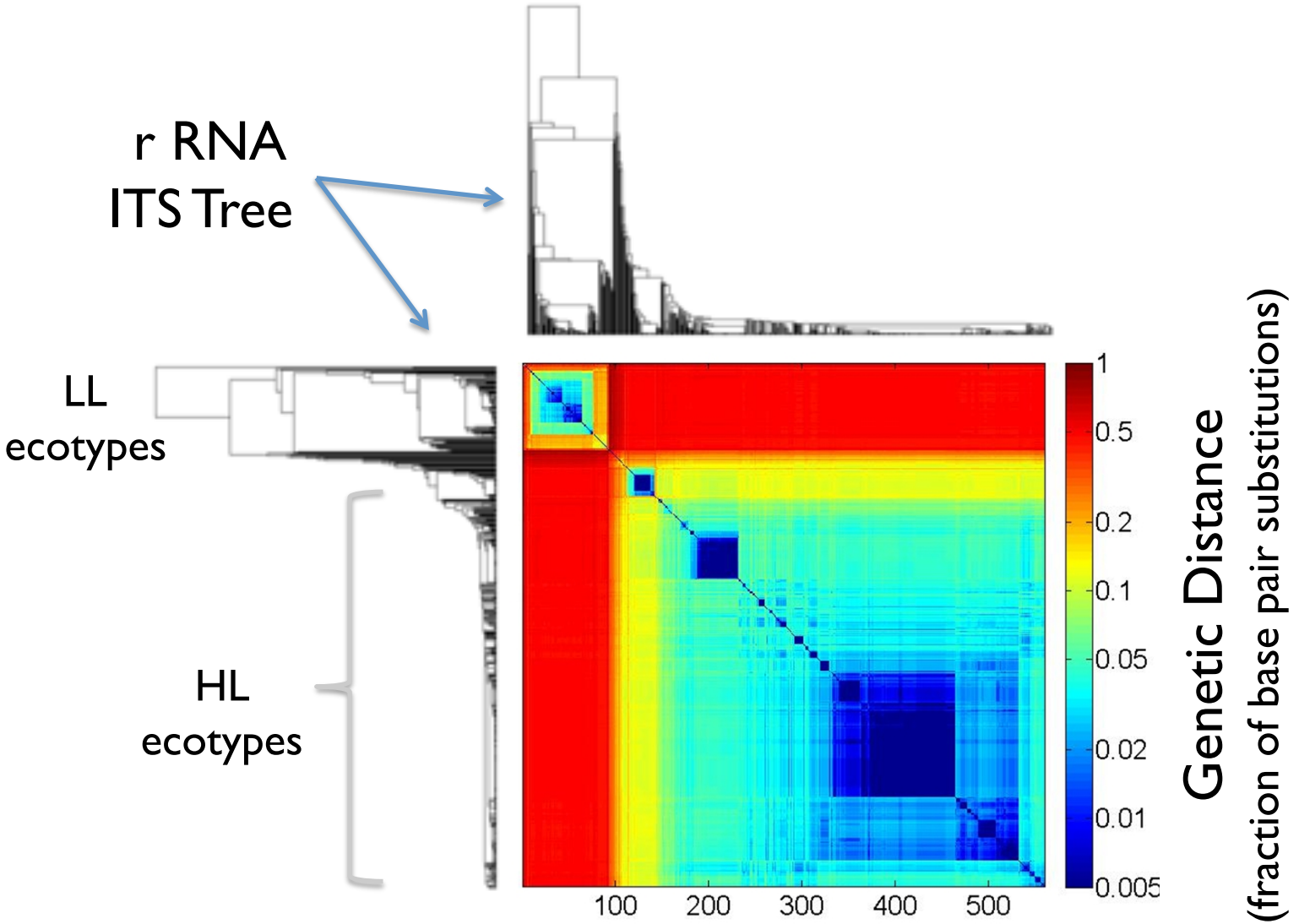
Makes sense, so far…



Population genetics of wild *Prochlorococcus* cells through single cell genomics



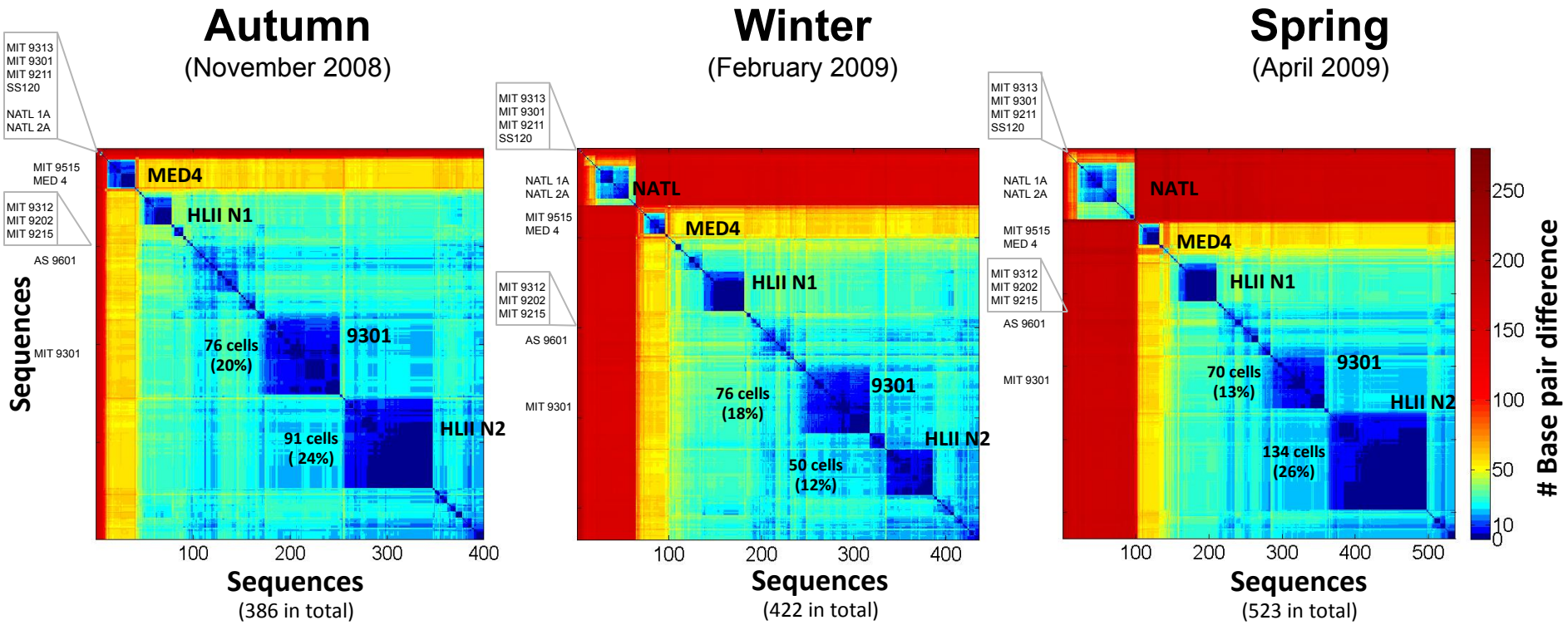
Representing the data

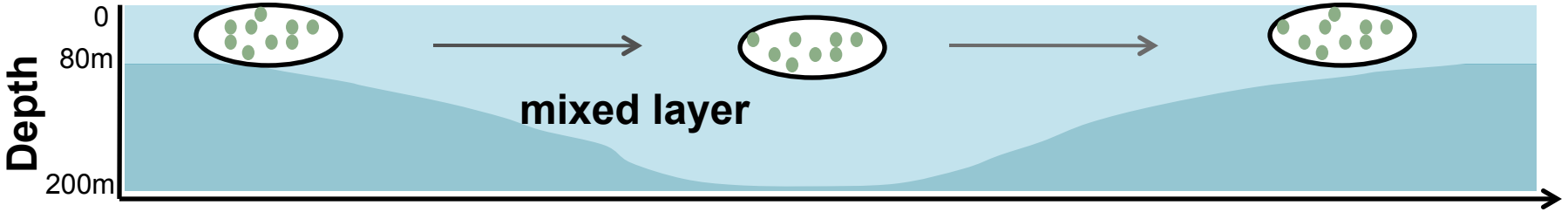


Nadav Kashtan

Prochlorococcus seasonal population dynamics in the Atlantic Ocean

(based on ribosomal ITS sequences analysis of samples from Bermuda)

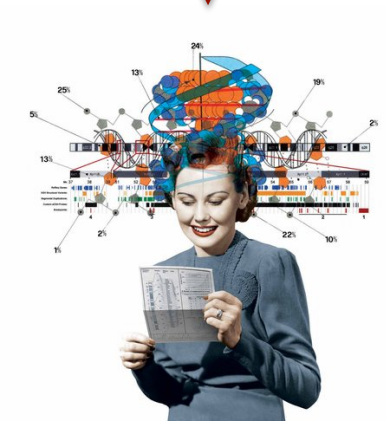




Knee-deep in Genomes

- 50 cultured *Prochlorococcus* strains
- 100's of wild single cells
- 100's of virus strains
- Megagenomic data
- Endless Ocean

Genome sequencing
Fast & Cheap!



Staggering amounts of data