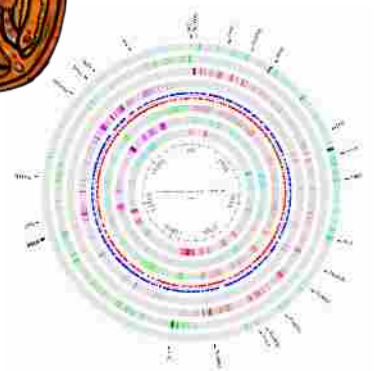
Croucher Course, Hongkong

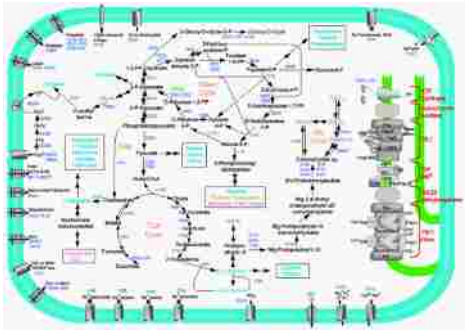
July 2025

Course 2: Using omics to decipher niche adaptation mechanisms in marine picocyanobacteria









Frédéric Partensky

Station Biologique de Roscoff

CNRS & Sorbonne University

Dept Adaptation & Diversity in the Marine Environment (UMR 7144)

Ecology of Marine Plankton (EcoMaP) Team

Roscoff, France

CNRS UPMC Station Biologique Roscoff

cnrs

SCIENCES SORBONNE UNIVERSITÉ

Croucher course 2 outline

I. Genetic diversity and evolution of picocyanobacteria

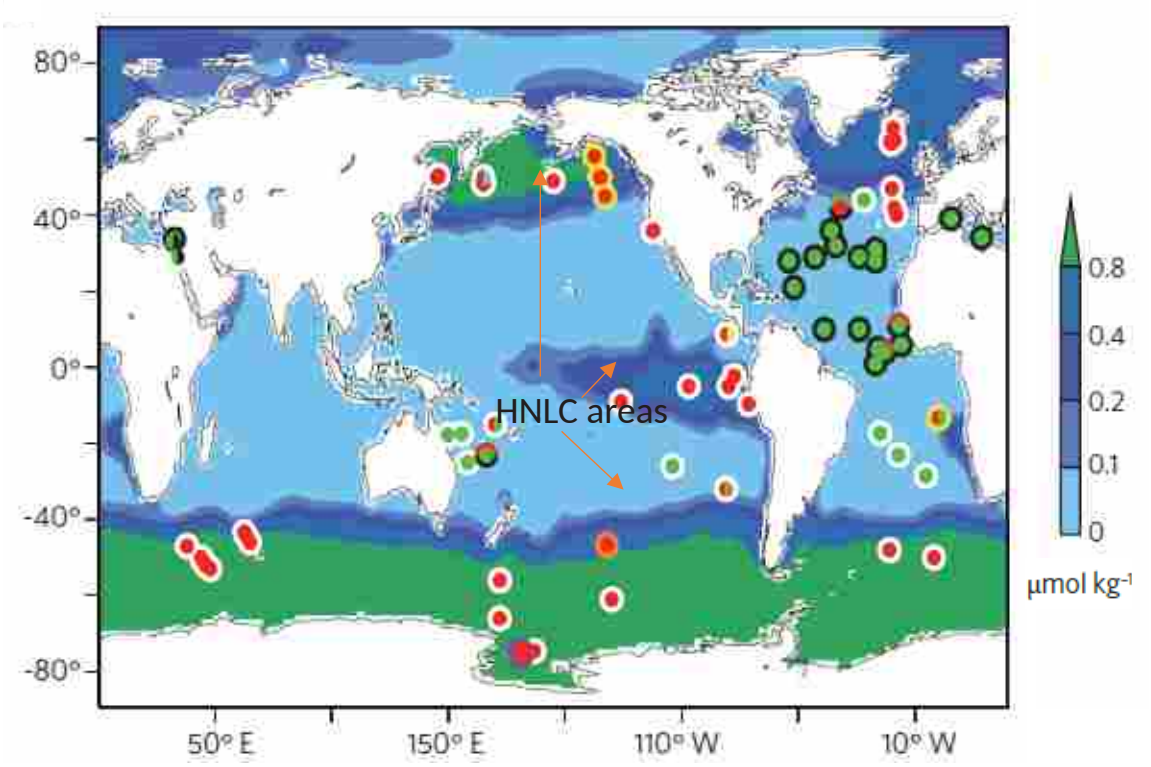
II. Ecological distribution of picocyanobacteria: are clades equivalent to ecotypes ?

III. Outcomes from comparative genomics

IV. Outcomes from metagenomics

Course 3: A multi-scale study of *Synechococcus* pigment types:
From genes to the global Ocean

Most of the Ocean is -N, -P or -Fe-limited or co-limited by several nutrients and this creates different nutrient niches for the phytoplankton, especially picocyanobacteria



Background : annual mean surface concentrations of **nitrates**

Primary (central circles) and secondary (outer circles) limiting nutrients:
N (green), P (black), Fe (red)

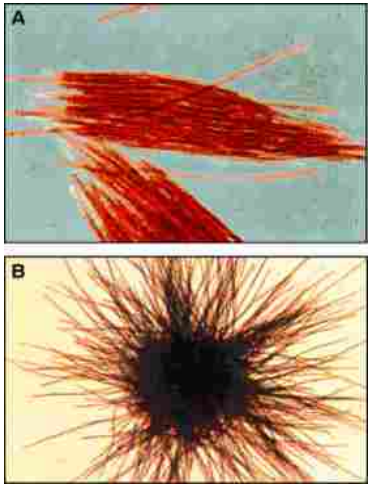
Limitations were inferred from chlorophyll and/or primary productivity increases after experimental amendment of N, P or Fe.

- **N** limitation in surface tropical ocean (Indian Ocean, most Atlantic and N Pacific oceans).
- **P** is a (co-) limiting factor in the Med Sea, N Atlantic and Gulf of Mexico.
- **Fe** limits productivity in the Southern Ocean as well as in the equatorial and subtropical S Pacific (HNLC areas).

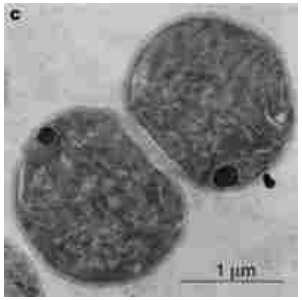
Moore et al. 2013, *Nature Geoscience*

Marine Cyanobacteria and diazotrophy

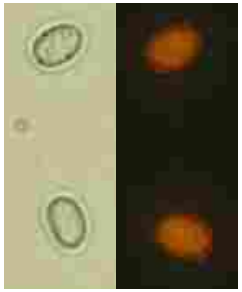
Free-living diazotrophs



Trichodesmium spp.



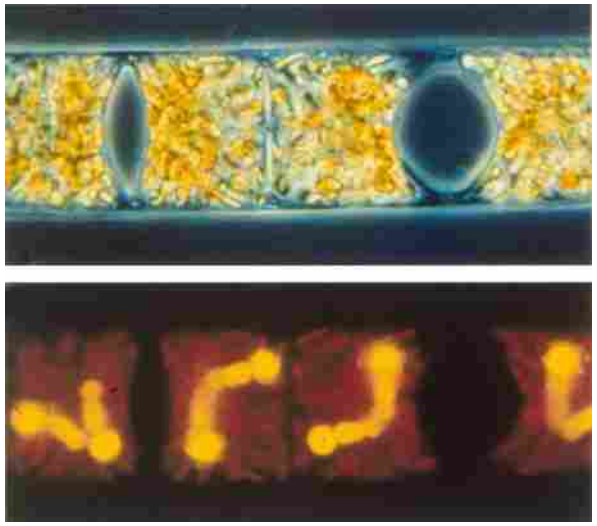
UCYN-B
Crocosphaera watsonii



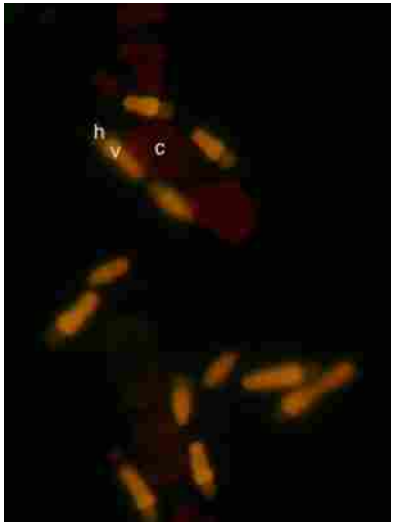
UCYN-C
Cyanothece sp.
(Crocosphaera subtropica)
Mares *et al.* J. Phycol. (2019)

Symbiotic diazotrophs

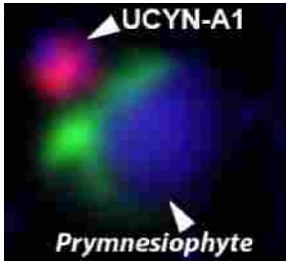




Calothrix
(Host: *Chaetoceros*)

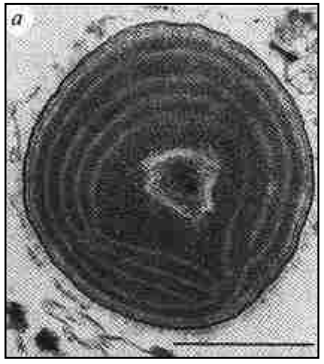


Richelia
(Host: *Hemiaulus*)

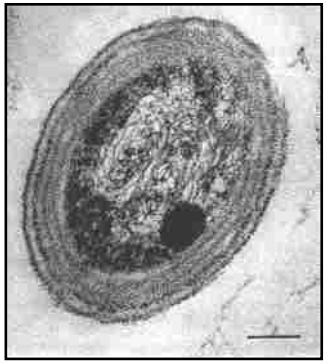


UCYN-A
Candidatus
Atelocyanobacterium
thalassa

Free-living non-diazotrophs



Synechococcus



Prochlorococcus

Photoheterotrophy in *Prochlorococcus*

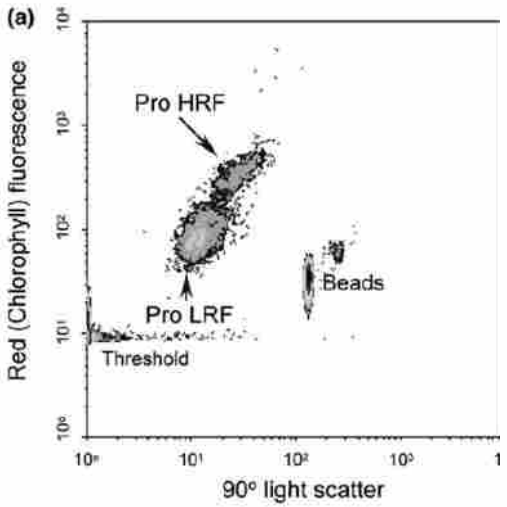
Picocyanobacteria compensate for the absence of diazotrophy by their small size (high surface/volume ratio and thus better nutrient absorption capacity) and specific abilities for photoheterotrophy.

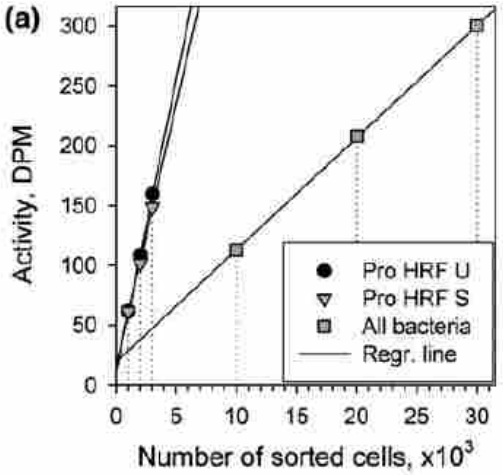
FEMS Microbiology Ecology 50 (2004) 153–161

Depth related amino acid uptake by *Prochlorococcus* cyanobacteria in the Southern Atlantic tropical gyre

Mikhail V. Zubkov [a,*], Glen A. Tarran [b], Bernhard M. Fuchs [c]

Method: Incubation of natural populations with radioactive ^{35}S-Methionine followed by flow cytometric cell sorting





Prochlorococcus type	Diameter[a] (μm)	Biomass (fg protein cell^{-1})	Methionine uptake (amol cell^{-1} d^{-1})
Pro LRF = HL phototype	0.58 ± 0.05 (14)	21	0.61 ± 0.22 (11)
Pro HRF = LL phototype	0.77 ± 0.08 (4)	48	2.5 ± 1.6 (4)

- ✓ About 25% of total bacterioplankton consumption of amino acids can be assigned to *Prochlorococcus*
- ✓ Pro LRF (= HL) have on average a comparable uptake rate than a marine bacterial cell, while the uptake rate of the Pro HRF (= LL) is 2.5–7 times higher.

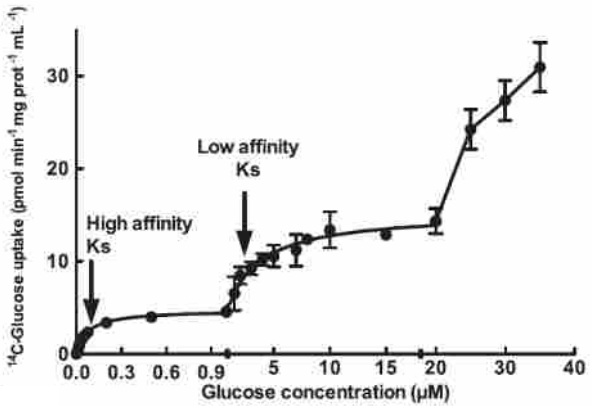
Photoheterotrophy in marine picocyanobacteria

PNAS | May 21, 2013 | vol. 110 | no. 21 | 8597–8602

Prochlorococcus can use the Pro1404 transporter to take up glucose at nanomolar concentrations in the Atlantic Ocean

María del Carmen Muñoz-Marín[a], Ignacio Luque[b], Mikhail V. Zubkov[c], Polly G. Hill[c], Jesús Díez[a], and José Manuel García-Fernández[a,1]

[a]Departamento de Bioquímica y Biología Molecular, Universidad de Córdoba, 14071 Córdoba, Spain; [b]Instituto de Bioquímica Vegetal y Fotosíntesis, Consejo Superior de Investigaciones Científicas and Universidad de Sevilla, 41092 Sevilla, Spain; and [c]National Oceanography Centre, Southampton SO14 3ZH, United Kingdom

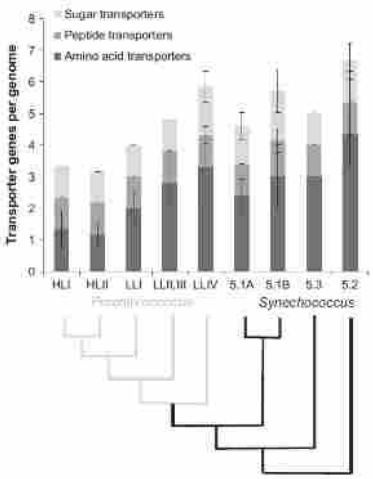


The ISME Journal (2016) 10, 2946–2957
© 2016 International Society for Microbial Ecology All rights reserved 1751-7362/16
www.nature.com/ismej

ORIGINAL ARTICLE

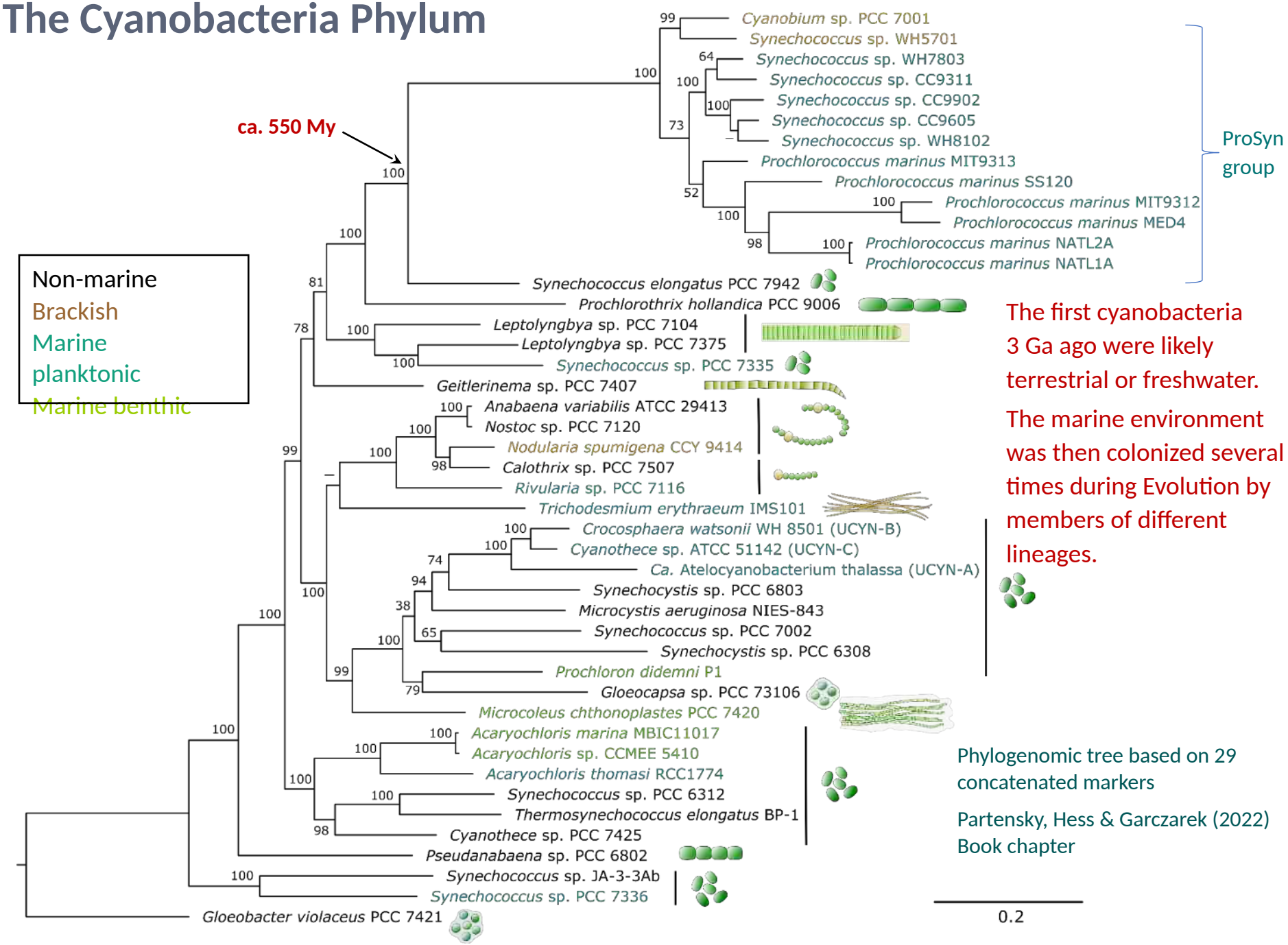
Global genetic capacity for mixotrophy in marine picocyanobacteria

Alexis P Yelton[1,5], Silvia G Acinas[2], Shinichi Sunagawa[3], Peer Bork[3], Carlos Pedrós-Alió[2] and Sallie W Chisholm[1,4]



More examples later on from metagenomics…

The Cyanobacteria Phylum



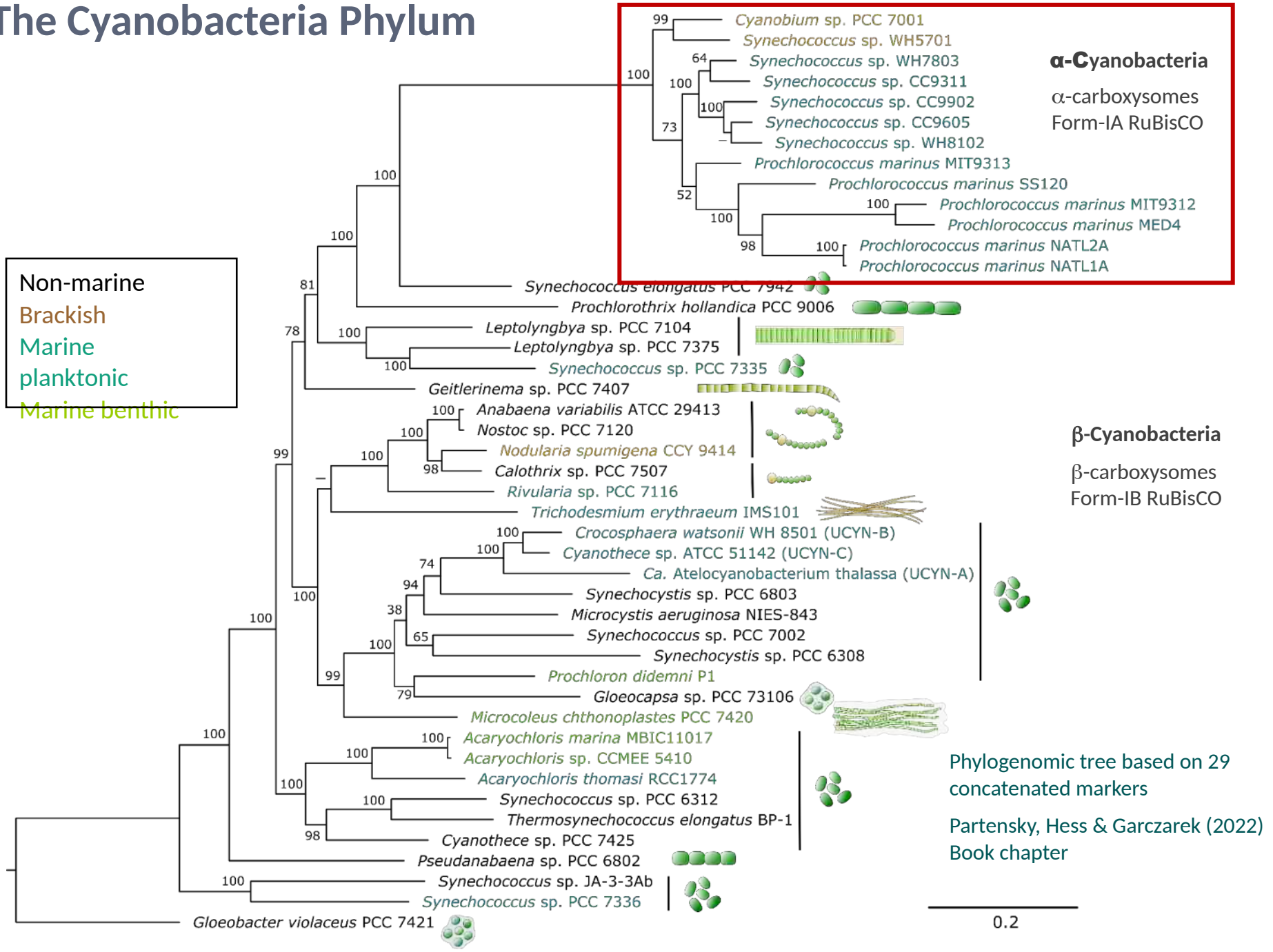
The first cyanobacteria 3 Ga ago were likely terrestrial or freshwater.

The marine environment was then colonized several times during Evolution by members of different lineages.

Phylogenomic tree based on 29 concatenated markers

Partensky, Hess & Garczarek (2022) Book chapter

The Cyanobacteria Phylum

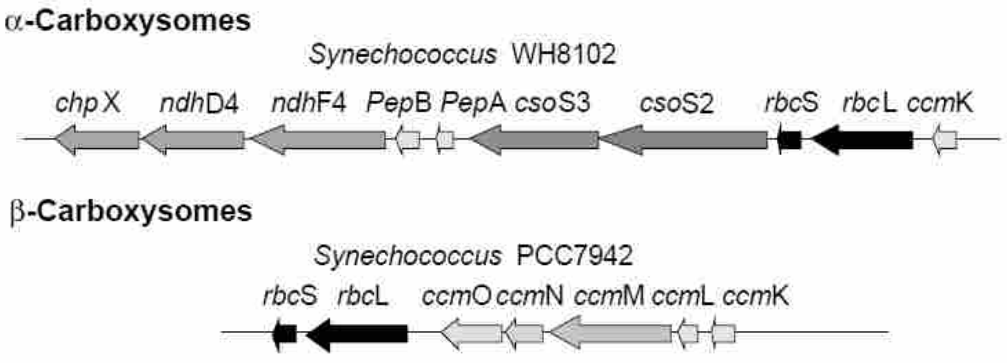


Phylogenomic tree based on 29 concatenated markers

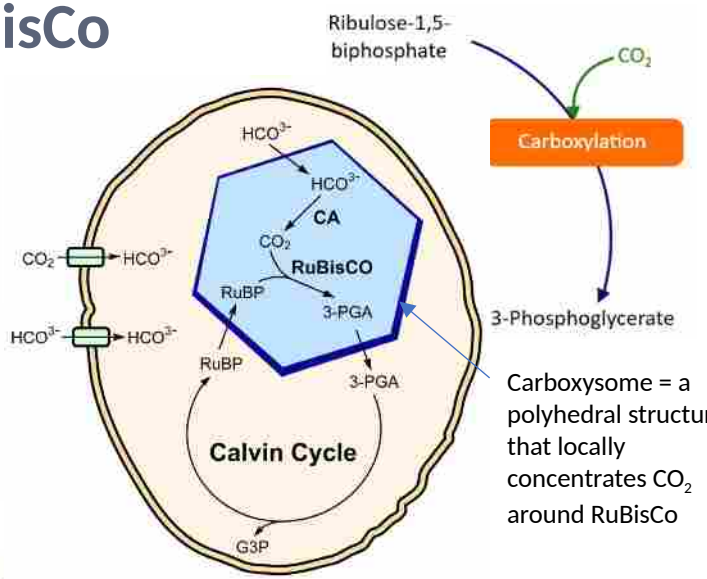
Partensky, Hess & Garczarek (2022) Book chapter

Differences between Forms IA and IB RuBisCo

Organization of the carboxysome gene cluster

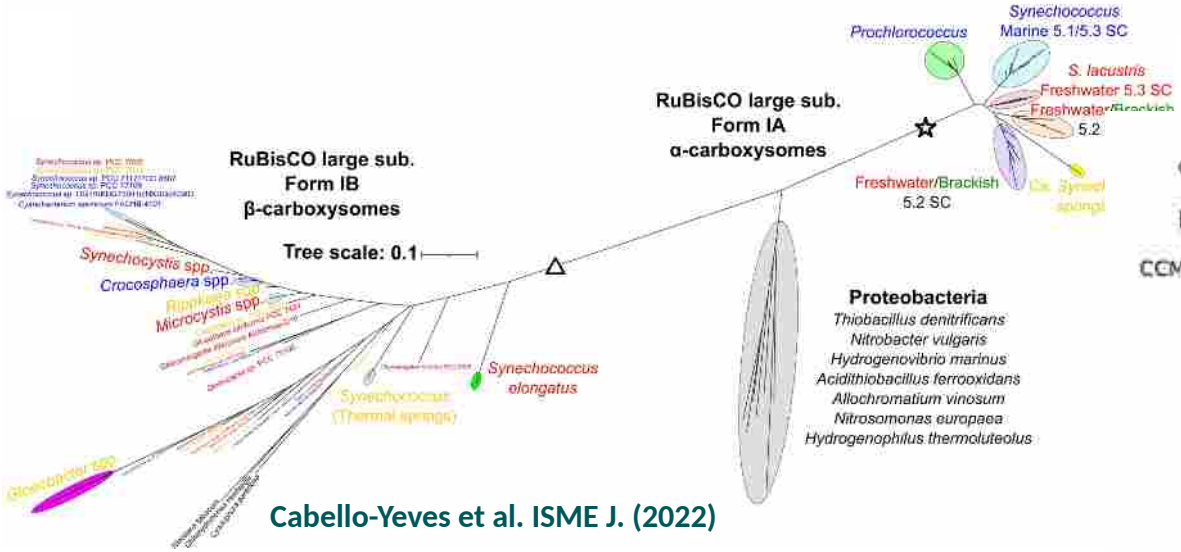


Badger et al. Funct. Plant Biol. (2002)

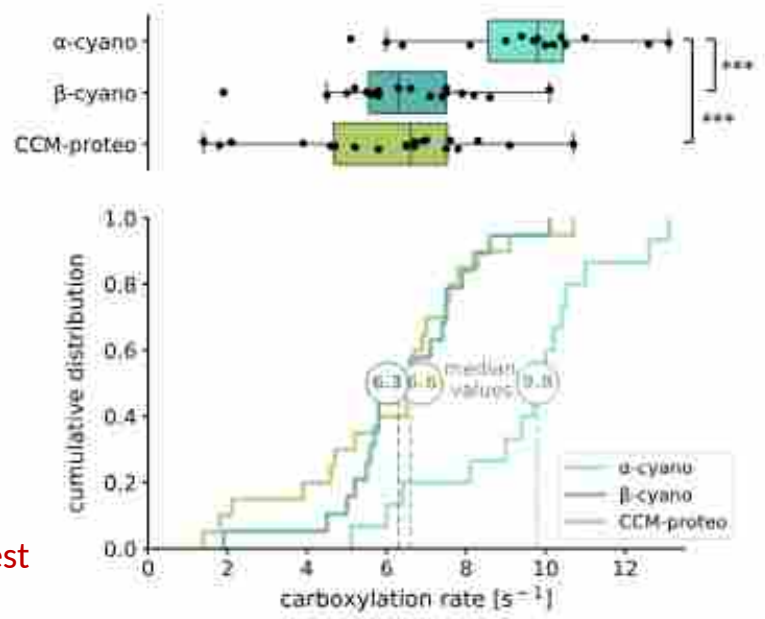


Carboxysome = a polyhedral structure that locally concentrates CO_2 around RuBisCo

Kerfeld Microbe magazine (2010)



Cabello-Yeves et al. ISME J. (2022)



De Pins et al. EMBO J (2024)

A high-throughput kinetic characterization of >100 bacterial Form-I RuBisCos has revealed that α-cyanobacteria's RuBisCos exhibit the highest carboxylation rates (≈10 s^{-1} median vs. <7 s^{-1} in other groups).

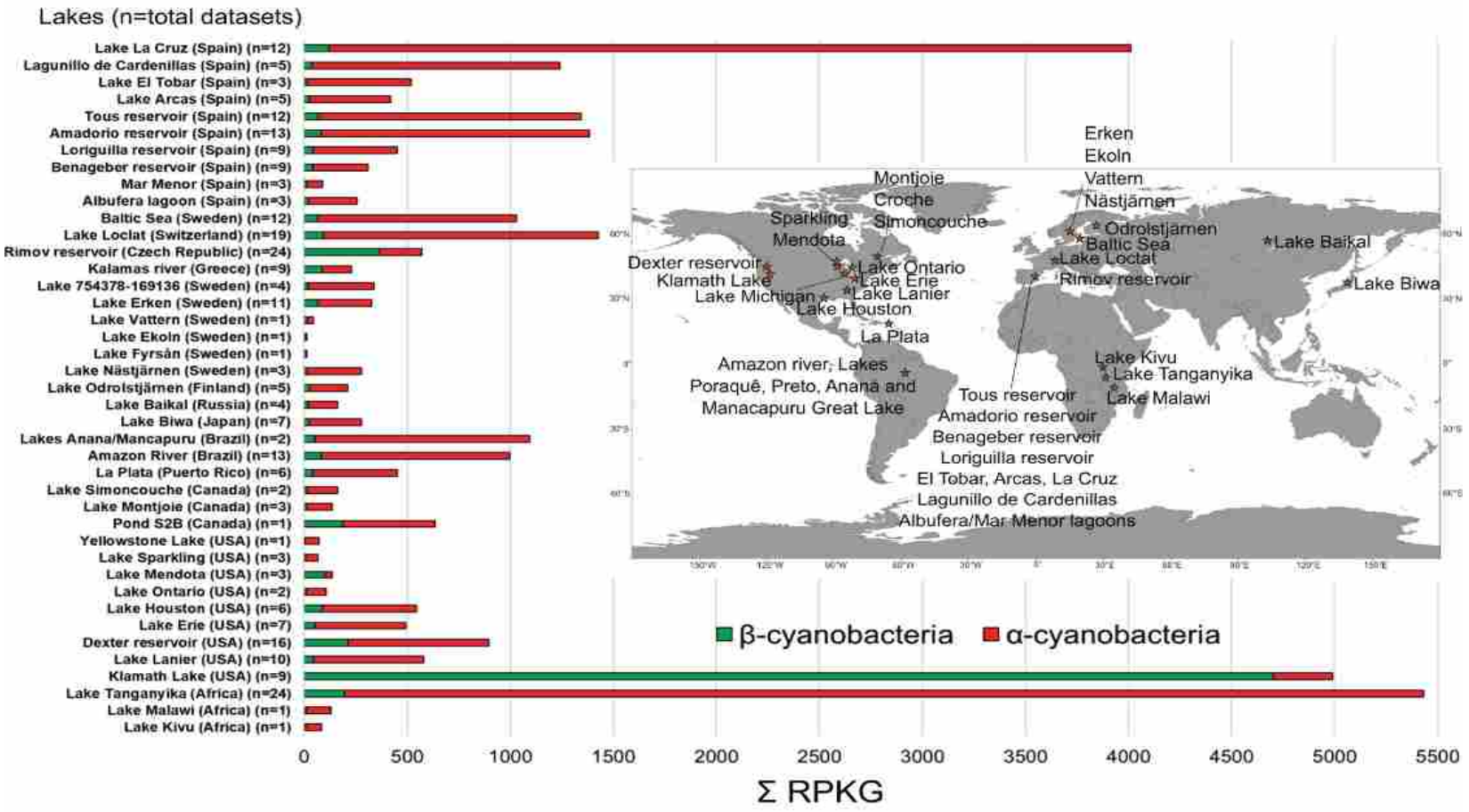
=> Potential competitive advantage over β-cyanobacteria.

α-cyanobacteria dominate not only in oceans but also great lakes

The ISME Journal (2022) 16:2421–2432

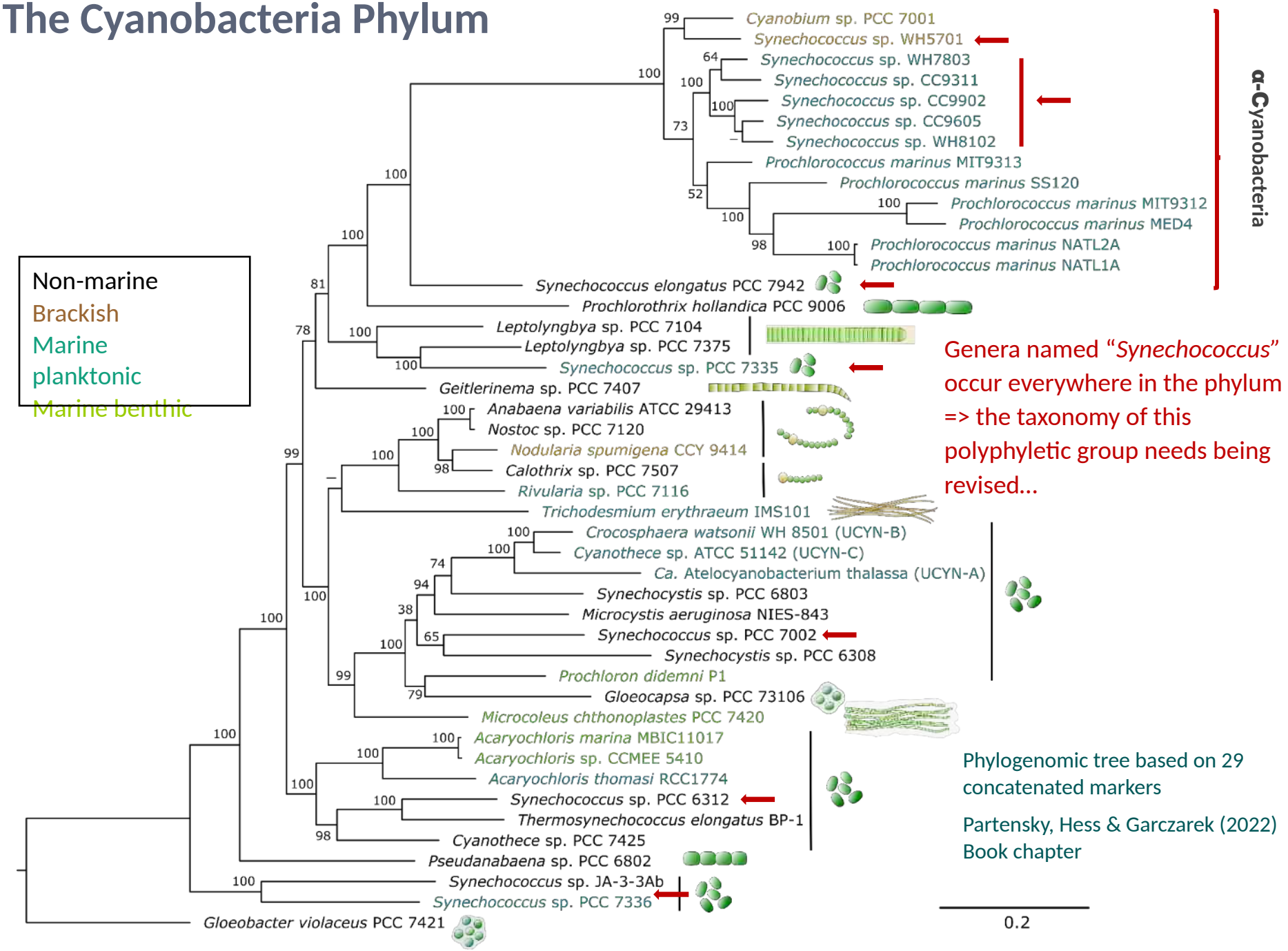
α-cyanobacteria possessing form IA RuBisCO globally dominate aquatic habitats

Pedro J. Cabello-Yeves[1,12], David J. Scanlan[2,12], Cristiana Callieri[3], Antonio Picazo[4], Lena Schallenberg[5], Paula Huber[6,7], Juan J. Roda-Garcia[1], Maciej Bartosiewicz[8], Olga I. Belykh[9], Irina V. Tikhonova[9], Alberto Torcello-Requena[2], Paula Martin De Prado[2], Andrew D. Millard[10], Antonio Camacho[4], Francisco Rodriguez-Valera[1,11] and Richard J. Puxty[2,12]



The paradigm of a separation in carboxysome type across the salinity divide does not hold true!

The Cyanobacteria Phylum



Genera named "*Synechococcus*" occur everywhere in the phylum => the taxonomy of this polyphyletic group needs being revised…

Phylogenomic tree based on 29 concatenated markers

Partensky, Hess & Garczarek (2022) Book chapter

The Cyanobacteria Phylum

Genus

Cyanobacteria/Subsection I/Incertae Sedis/

Form- *Synechococcus*

Michael Herdman and Rosmarie Rippka

Syn.e.cho.coc' cus. Gr. n. *synechia* continuity; Gr. n. *kokkos* grain or kernel; M.L. masc. n. *Synechococcus* contiguous cells.

Bergey's Manual of Systematics of Archaea and Bacteria

This article is © 2001
Online © 2015

The α-Cyanobacteria radiation is called "Cluster 5" in the Bergey's manual and is subdivided in two different Syn. "subclusters" and two *Cyanobium* groups; while the type species *Synechococcus elongatus* belongs to "Cluster 1".

Synechococcus Nägeli, 1849

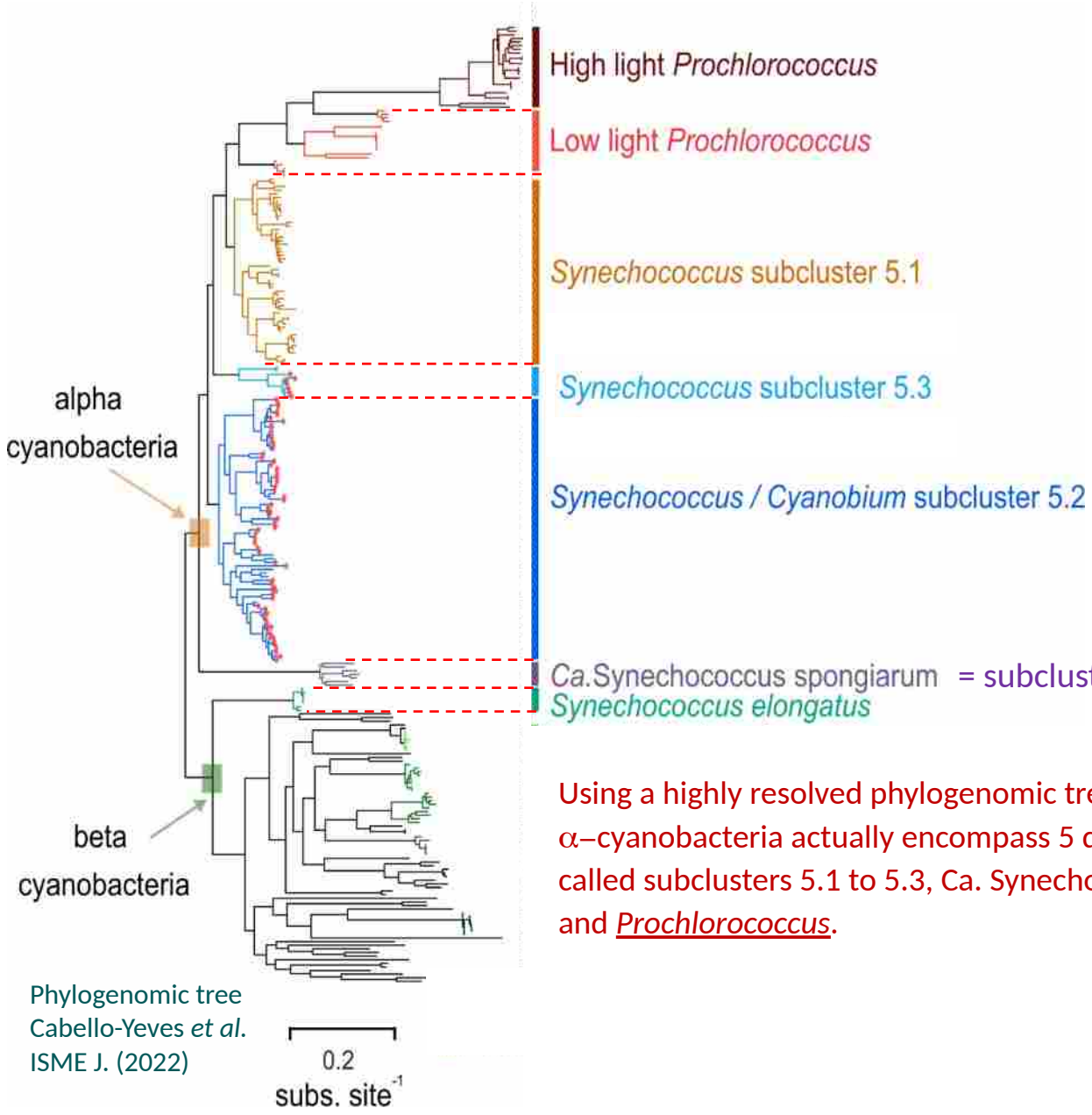
Holotype species: *Synechococcus elongatus* (Nägeli) Nägeli





16S rDNA tree, made by A. Wilmotte

Genetic diversity within α-Cyanobacteria



Phylogenomic tree
Cabello-Yeves *et al.*
ISME J. (2022)

Using a highly resolved phylogenomic tree and many more sequences, α–cyanobacteria actually encompass 5 deep monophyletic lineages called subclusters 5.1 to 5.3, Ca. Synechococcus spongarium (= SC 5.4?), and *Prochlorococcus*.

Genetic diversity within *Prochlorococcus*



Johnson and Sieburth, 1979

High Light (HL) Phototypes
Low DV-Chl *b/a*

Low Light (LL) phototypes
High DV-Chl *b/a*

Tree from Biller et al. (2015) Nature Rev. Microbiol.

Prior to 2010, the *Proc.* genus was thought to account 4 HL-adapted clades (HLI-IV) and 4 LL-adapted clades (LLI-IV) occupying different niches in the Ocean.

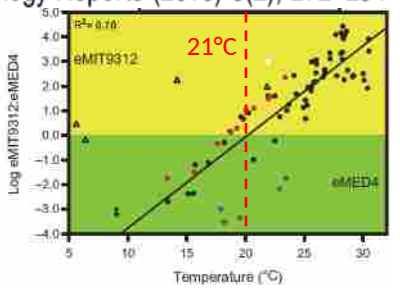
By 2015, several studies had unveiled novel clades in previously under-sampled environments (arrows).



Environmental Microbiology Reports (2016) **8**(2), 272–284

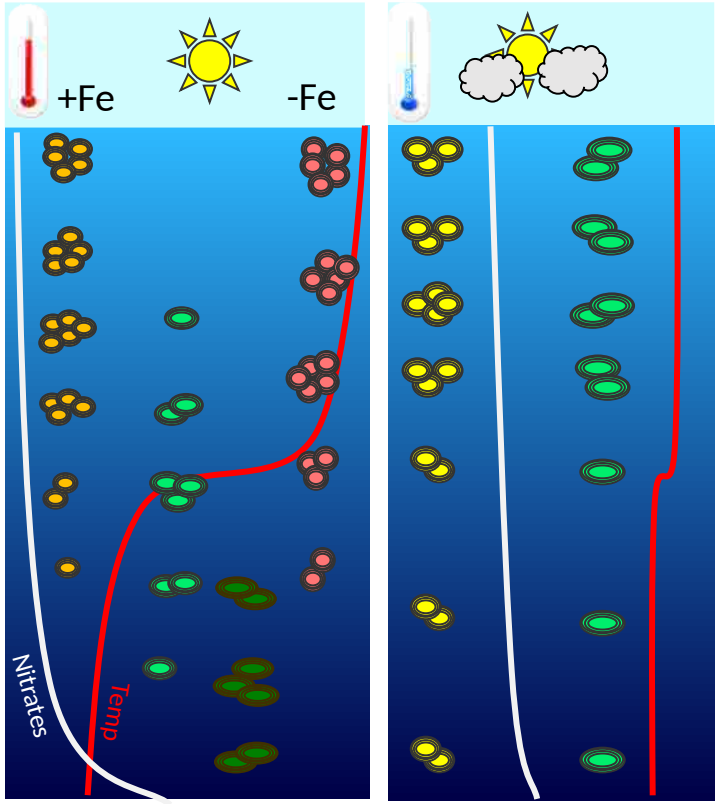
Jeremy W. Chandler,[1] Yajuan Lin,[2] P. Jackson Gainer,[1] Anton F. Post,[3] Zackary I. Johnson[2] and Erik R. Zinser[1]*

log-linear relationship between the HLII/HLI ratio and temp. : neither ecotype is outcompeted to (local) extinction.



Low Latitude
Oligotrophic, stratified waters

High Latitude
Nutrient-rich, mixed waters



Johnson et al., 2006; Malmstrom et al. 2010, Rusch et al. 2010

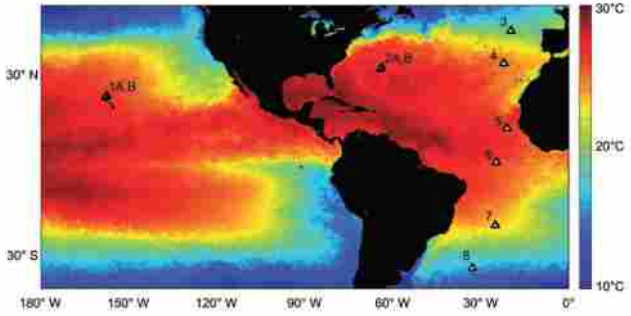
Modified from Partensky & Garczarek, ARMS (2010)

Genetic diversity within *Prochlorococcus*

Environmental Microbiology (2009) 11(4), 823–832

Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*

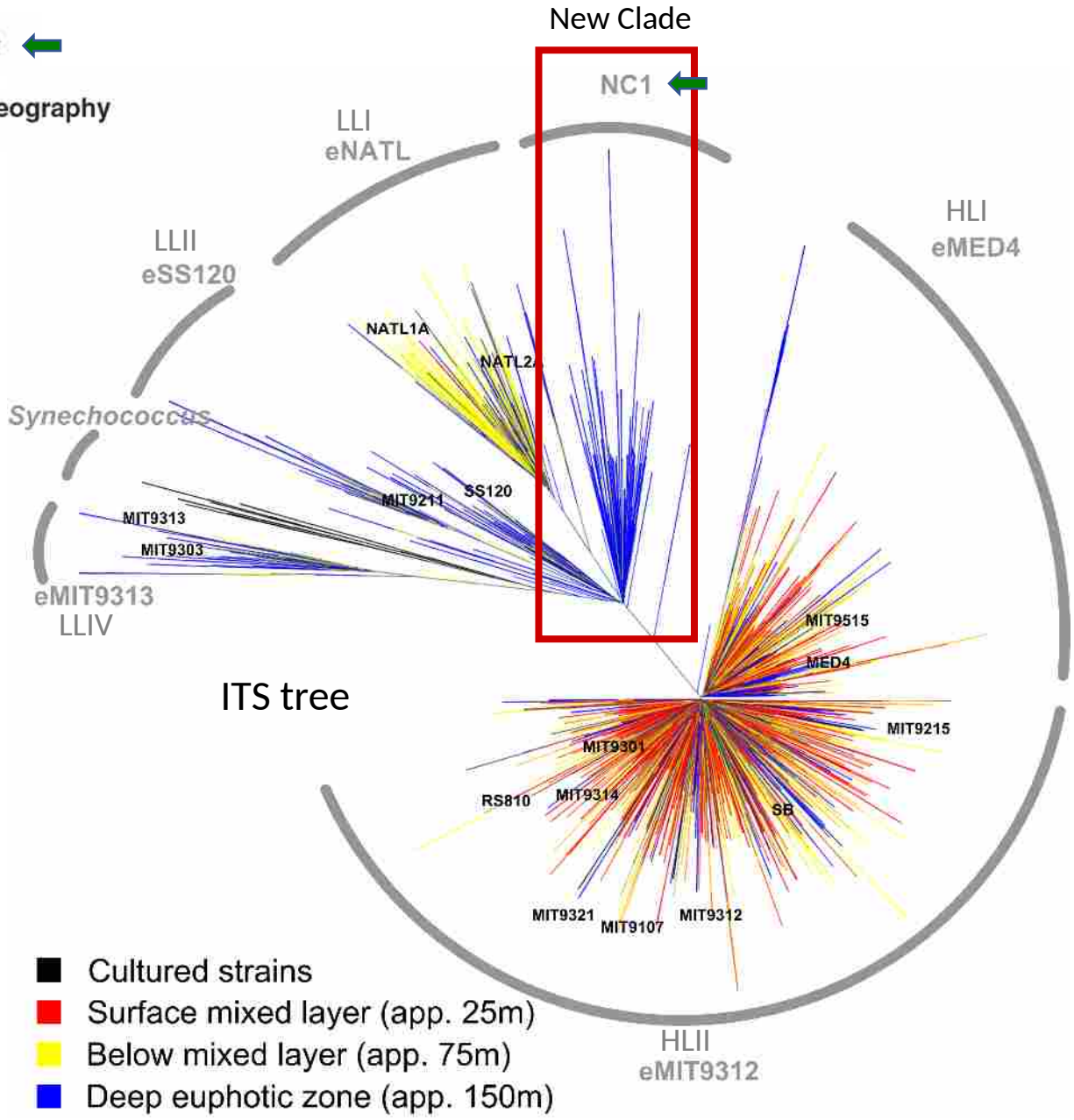
Adam C. Martiny,[1,2,3]* Amos P. K. Tai,[1†] Daniele Veneziano,[1] François Primeau[2] and Sallie W. Chisholm[1**]



The ITS region from *Proc*. populations from 10 sites in the Atlantic and Pacific Ocean was PCR-amplified, cloned and sequenced.

=> Detection in the deep euphotic zone of a new deeply branching LL group termed NC1 (for « Non-Cultured 1 ») and later renamed LLVII.

It comprised as much as 61.7% (mean= 23.1%) of the *Proc*. sequences detected in the lower euphotic zone.



Genetic diversity within *Prochlorococcus*

Environmental Microbiology Reports (2010) 2(6), 728–738

Novel lineages of *Prochlorococcus* thrive within the oxygen minimum zone of the eastern tropical South Pacific

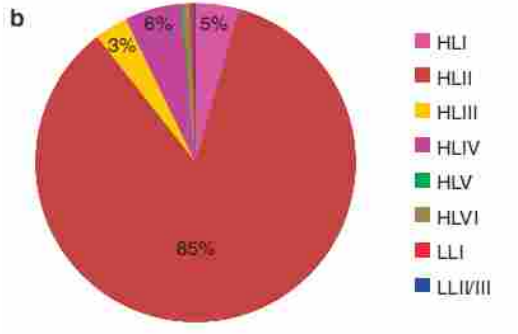
Paris Lavin,[1,2] Bernardo González,[3,4] J. Francisco Santibáñez,[2] David J. Scanlan[5] and Osvaldo Ulloa[2*]

The ISME Journal (2012) 6, 285–297

Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans

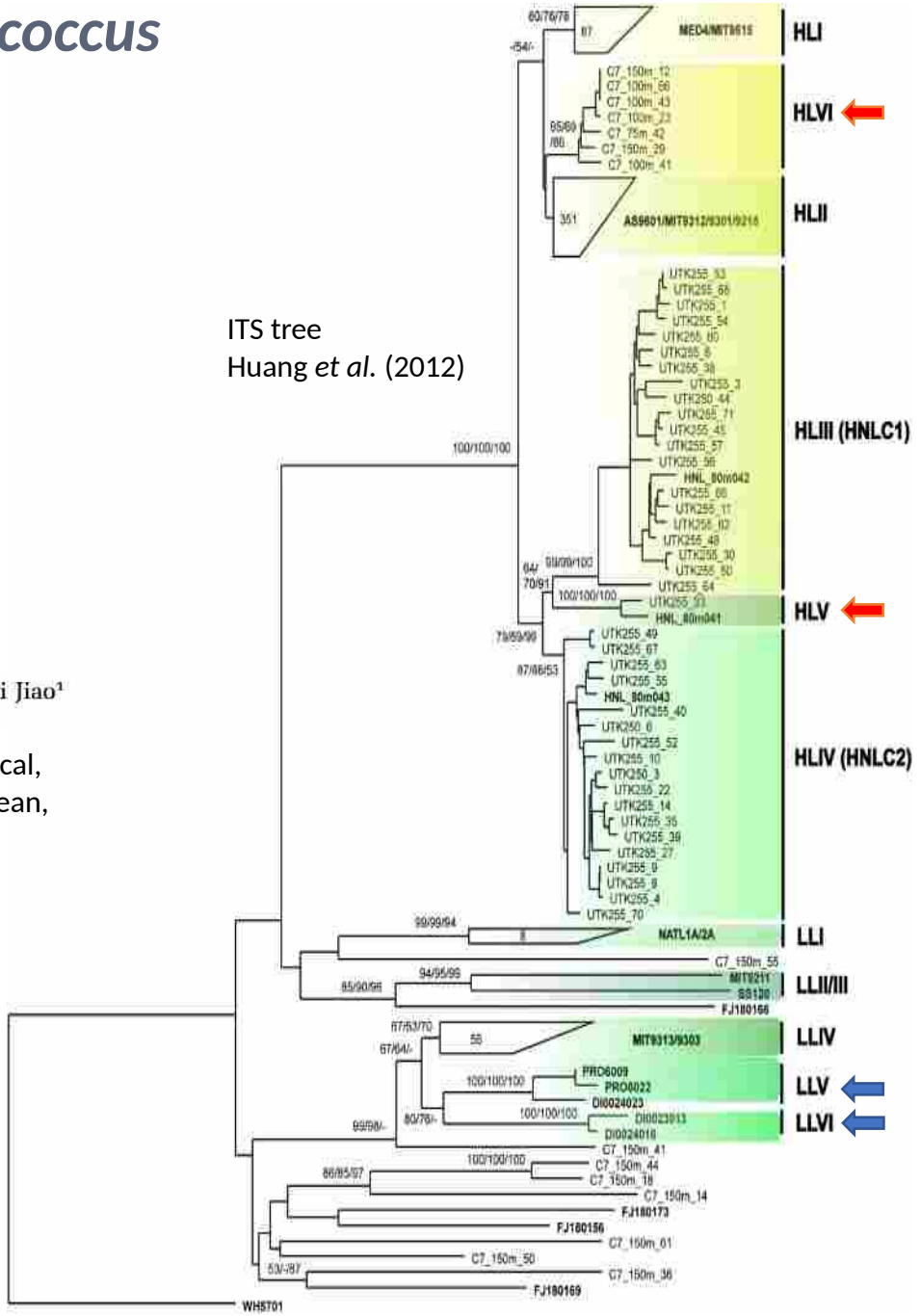
Sijun Huang[1,2], Steven W Wilhelm[3], H Rodger Harvey[4], Karen Taylor[4], Nianzhi Jiao[1] and Feng Chen[2]

DNA sequences from diverse marine ecosystems, including the tropical, subtropical and high-latitude N Atlantic Ocean, equatorial Pacific Ocean, subtropical S Pacific Ocean, S China Sea, Bering Sea and Chukchi and combined with previous sequences.



Classification of *Proc*. ITS sequences retrieved from the GOS database (only the surface was sampled)

ITS tree
Huang *et al*. (2012)



Some *Prochlorococcus* from Anoxic Marine Zones* possess complete phycobilisomes

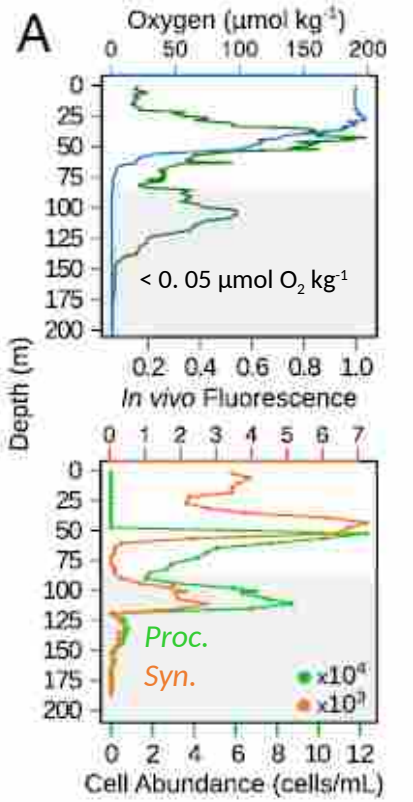
PNAS March 11, 2021 | 118 (11) e2025638118

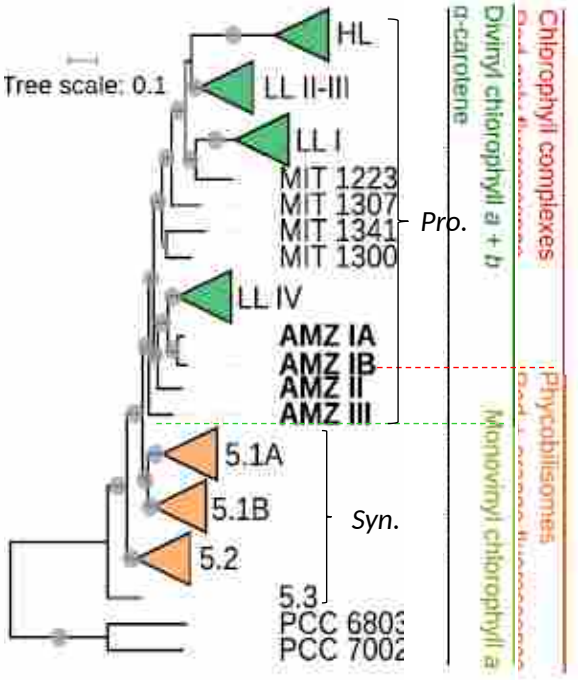
The cyanobacterium *Prochlorococcus* has divergent light-harvesting antennae and may have evolved in a low-oxygen ocean

Osvaldo Ulloa[a,b,1], Carlos Henríquez-Castillo[a,b,2], Salvador Ramírez-Flandes[a,b], Alvaro M. Plominsky[a,b,c,3], Alejandro A. Murillo[a,b,4], Connor Morgan-Lang[d], Steven J. Hallam[c,d,e,f,g], and Ramunas Stepanauskas[h]

*Use of new high sensitivity analytical technologies for measuring O_2 concentrations in OMZs indicated that OMZs are essentially anoxic marine zones (AMZs).

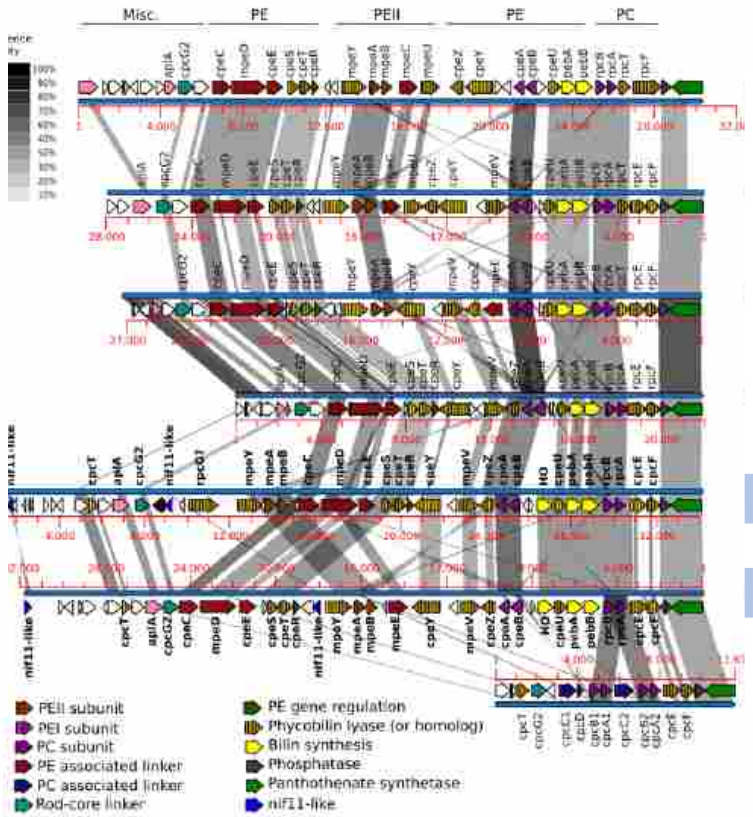
Ulloa *et al.* PNAS 2012 109:15596-160003





ML tree based on 49 concatenated ribosomal proteins

AMZ-I = LLV
AMZ-II = LLVI

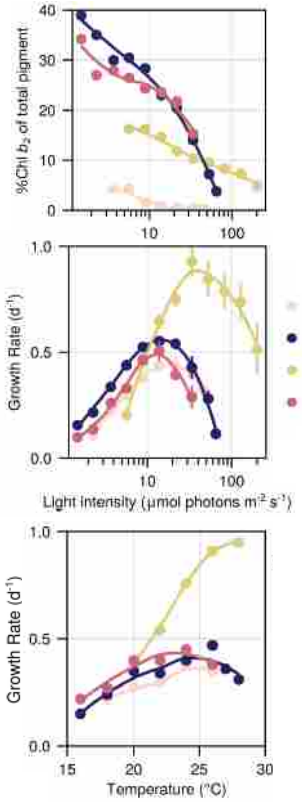


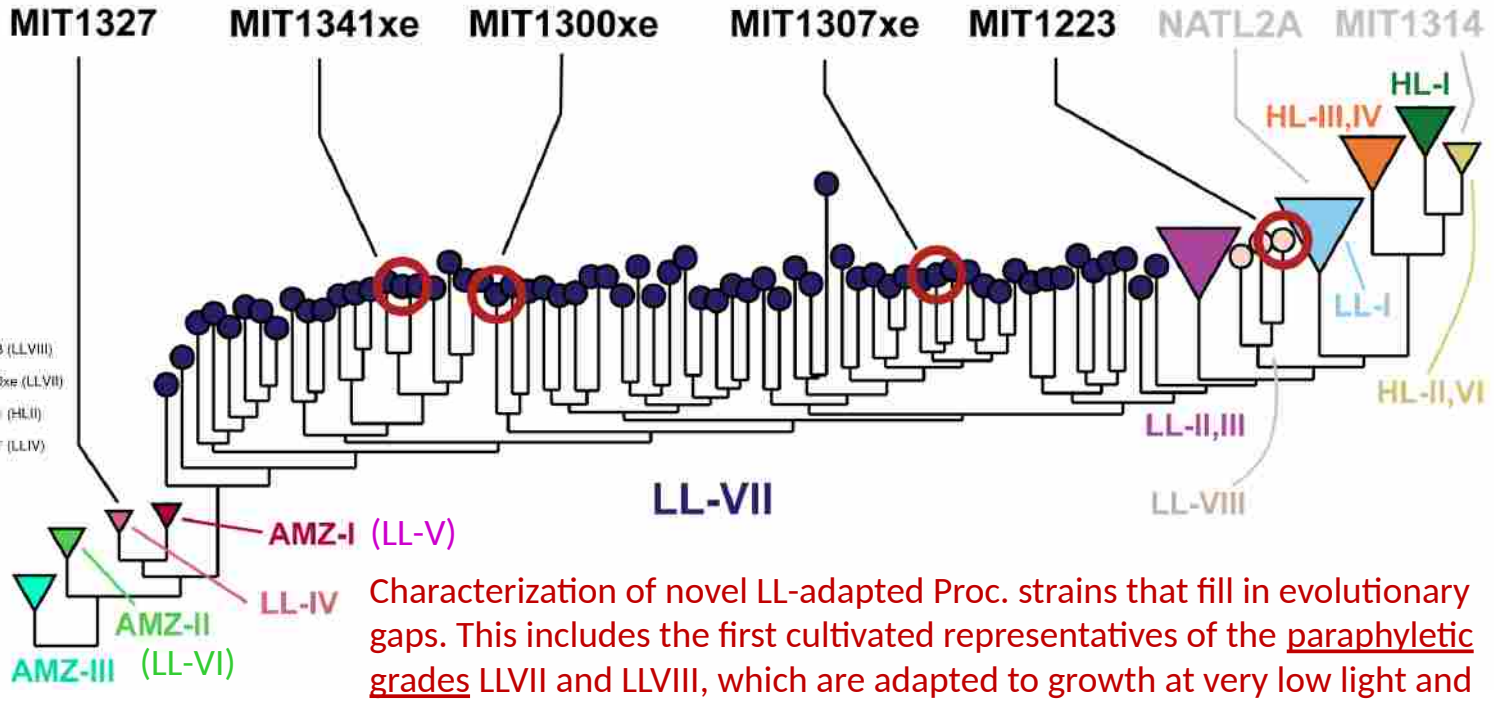
Genetic diversity within *Prochlorococcus*

mBio 2024 Volume 15 Issue 11

Novel isolates expand the physiological diversity of *Prochlorococcus* and illuminate its macroevolution

Jamie W. Becker,[1] Shaul Pollak,[1] Jessie W. Berta-Thompson,[1] Kevin W. Becker,[2] Rogier Braakman,[3] Keven D. Dooley,[1] Thomas Hackl,[1] Allison Coe,[1] Aldo Arellano,[1] Kristen N. LeGault.[1] Paul M. Berube.[1] Steven J. Biller.[1] Andrés Cubillos-Ruiz.[1] Beniamin A. S. Van Moov.[2] Sallie W. Chisholm[1,4]





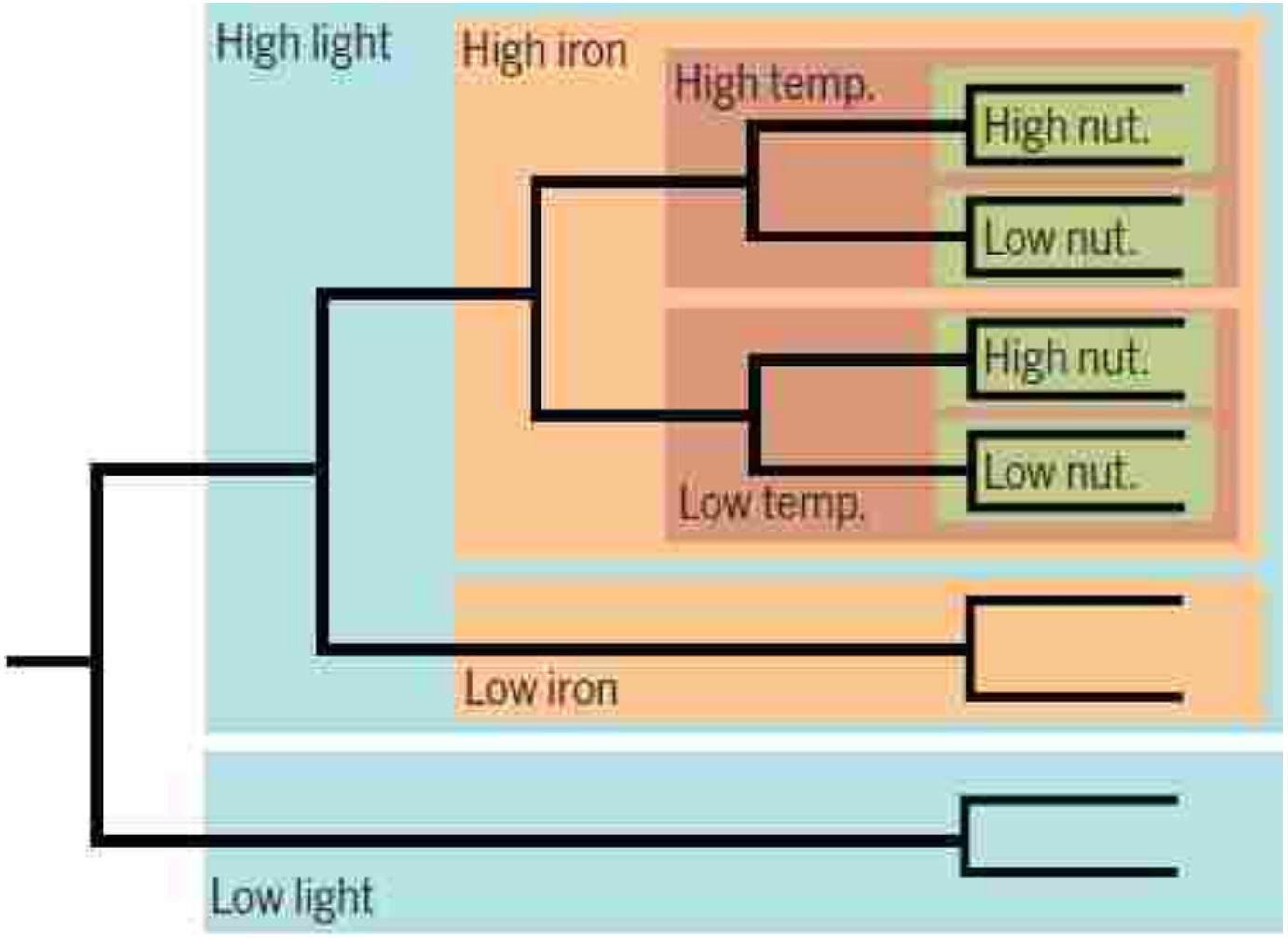
Phylogenomic tree using 44 marker genes

Characterization of novel LL-adapted Proc. strains that fill in evolutionary gaps. This includes the first cultivated representatives of the paraphyletic grades LLVII and LLVIII, which are adapted to growth at very low light and possess unique light-harvesting gene signatures and pigmentation. Their ITS proved particularly hard to align…

Despite its LL-adapted physiological characteristics, the LLVIII strain MIT1223 has a low chl b_2 content like HL-adapted strains, challenging the paradigm that all LL-adapted *Proc*. exhibit high chl b/a_2 ratios.

To date, still no culture representatives from HLIII through VI clades as well as LLV-VI and AMZ-III clades.

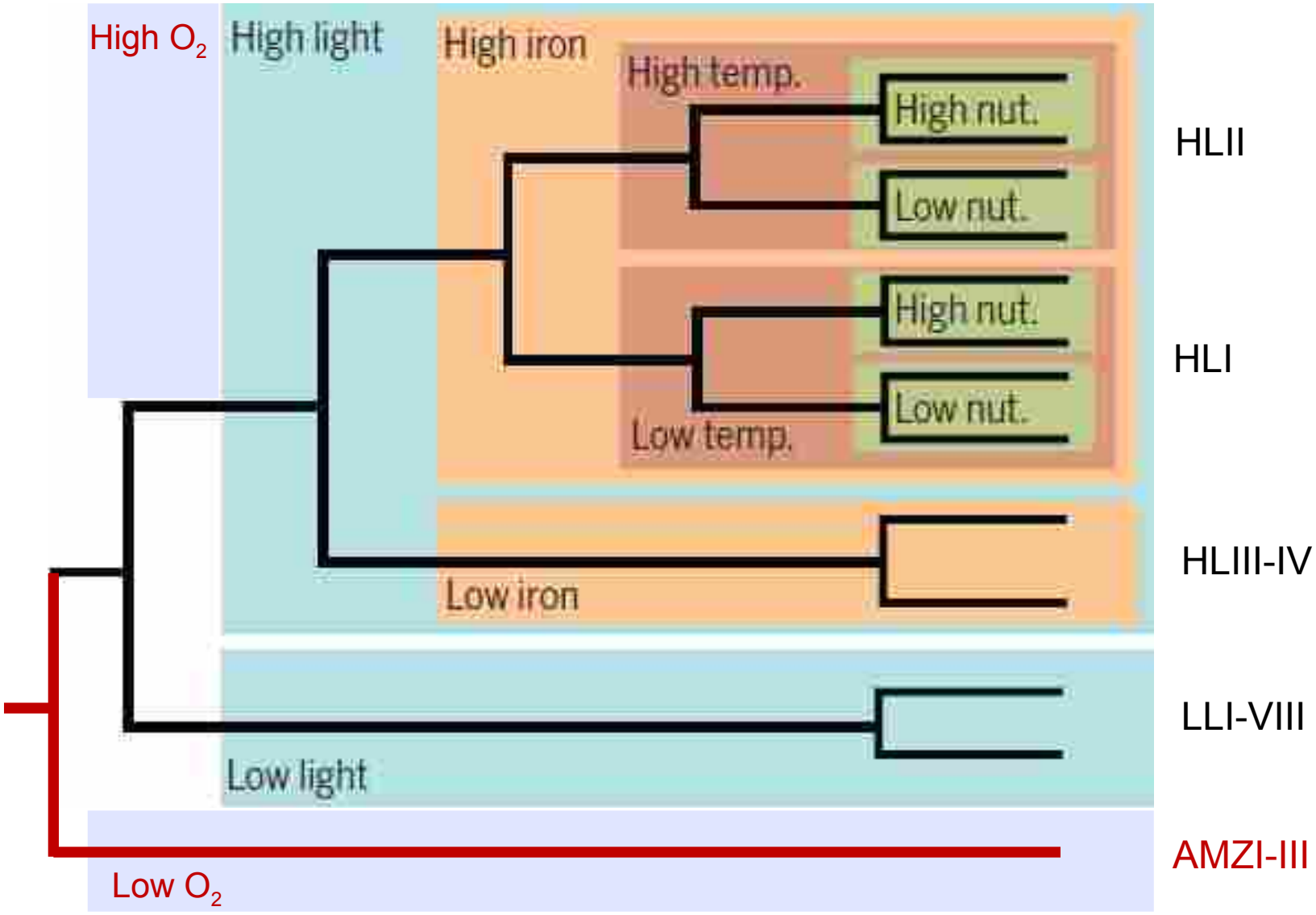
Prochlorococcus Evolution has been conditioned by different environmental factors



Martiny *et al.* 2015 *Science*

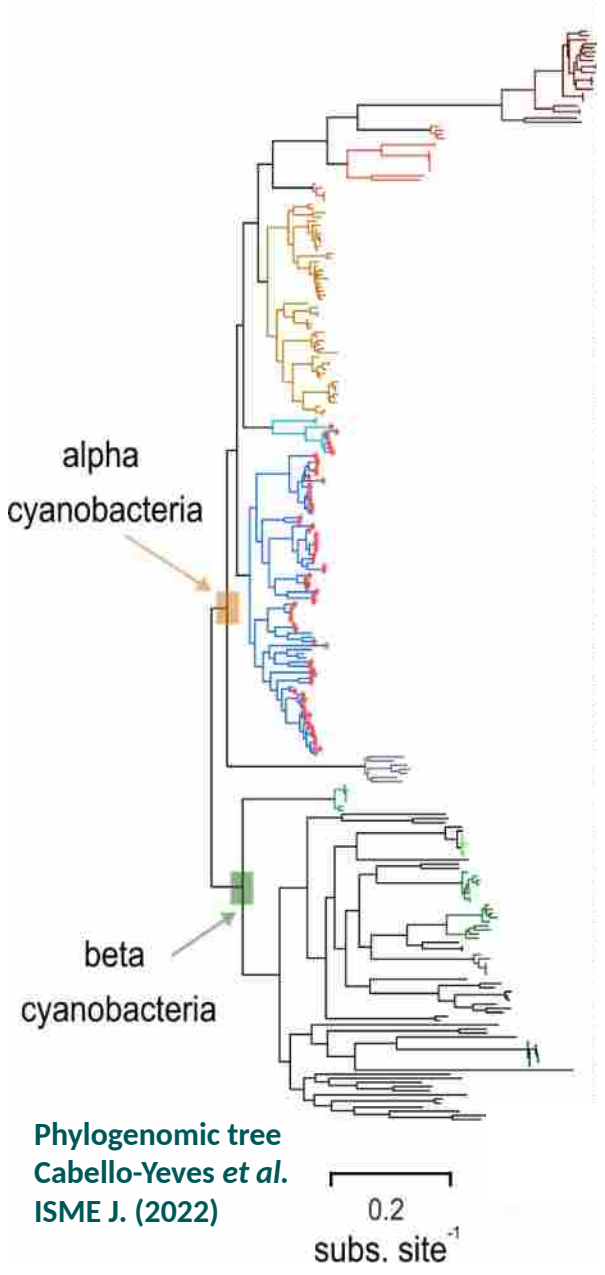
6 NOVEMBER 2015 • VOL 350 ISSUE 6261

Prochlorococcus Evolution has been conditioned by different environmental factors



Martiny *et al*. 2015 *Science* (modified)

Diversity within α-Cyanobacteria



Phylogenomic tree
Cabello-Yeves *et al.*
ISME J. (2022)

High light *Prochlorococcus*

Low light *Prochlorococcus*

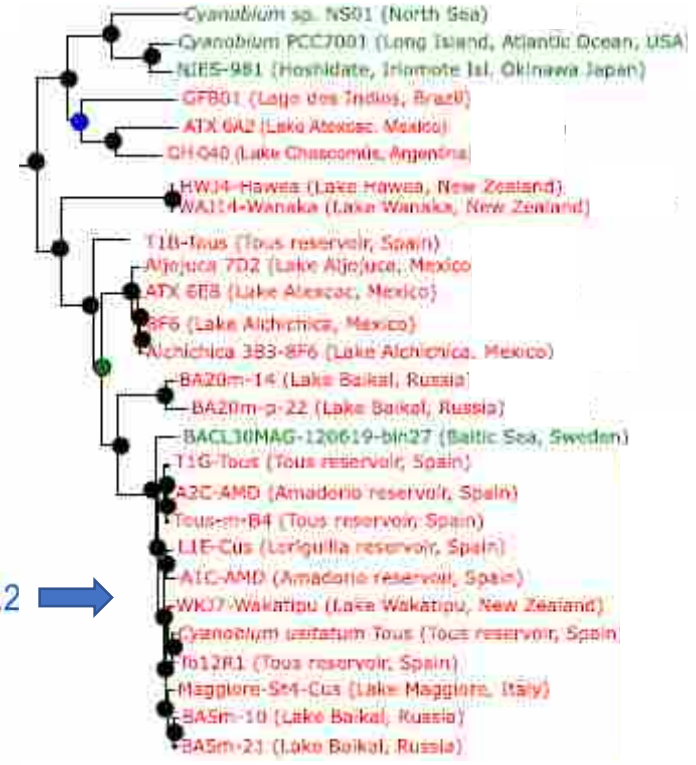
Synechococcus subcluster 5.1

Synechococcus subcluster 5.3

Synechococcus / Cyanobium subcluster 5.2

Ca.Synechococcus spongiarum

Synechococcus elongatus

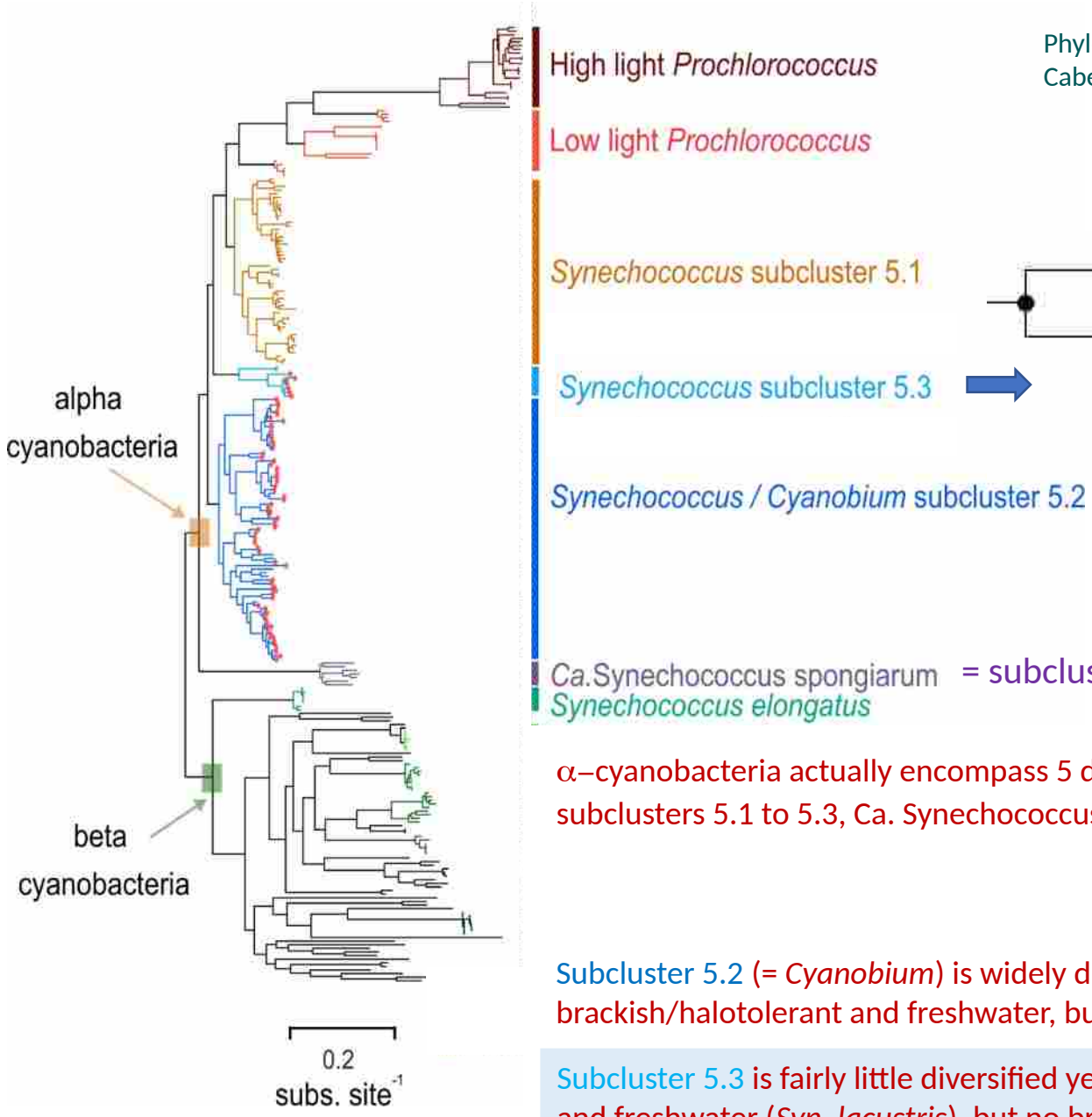


Phylogenomic tree (extract)
Cabello-Yeves *et al.* BMC Biol. (2022)

α–cyanobacteria actually encompass 5 deep monophyletic lineages called subclusters 5.1 to 5.3, Ca. Synechococcus spongarium (=SC 5.4?), and *Proc.*

Subcluster 5.2 (= *Cyanobium*) is widely diversified and encompasses brackish/halotolerant and freshwater, but no strictly marine representatives.

Diversity within α-Cyanobacteria



Phylogenomic tree
Cabello-Yeves *et al.* ISME J. (2022)



Phylogenomic tree (extract)
Cabello-Yeves *et al.* BMC Biol. (2022)

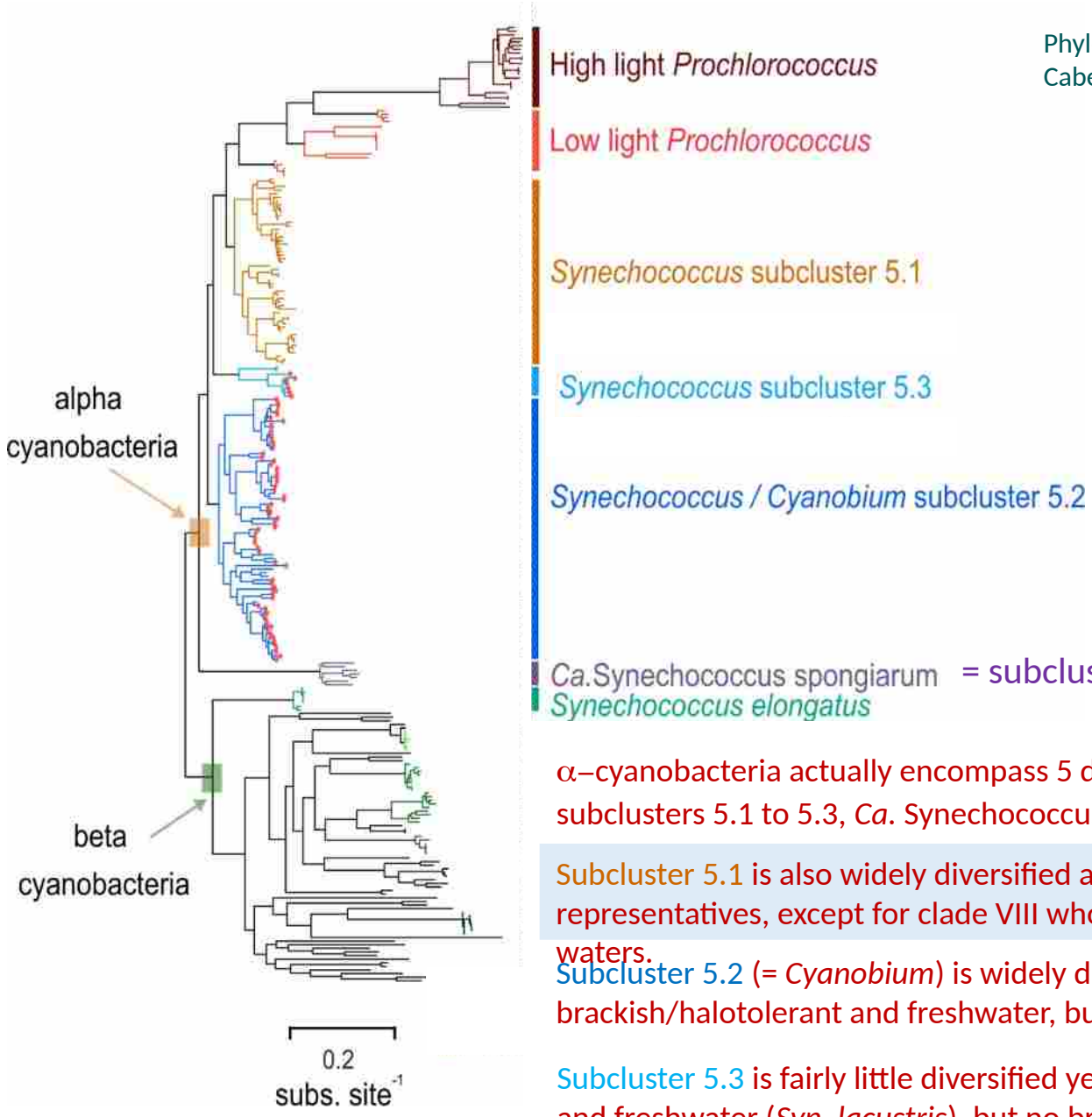
= subcluster 5.4 ?

α–cyanobacteria actually encompass 5 deep monophyletic lineages called subclusters 5.1 to 5.3, Ca. Synechococcus spongarium (=SC 5.4?), and *Proc.*

Subcluster 5.2 (= *Cyanobium*) is widely diversified and encompasses brackish/halotolerant and freshwater, but no strictly marine representatives.

Subcluster 5.3 is fairly little diversified yet encompasses both strictly marine and freshwater (*Syn. lacustris*), but no brackish representatives.

Diversity within α-Cyanobacteria



Phylogenomic tree
Cabello-Yeves *et al.* ISME J. (2022)

= subcluster 5.4 ?

α–cyanobacteria actually encompass 5 deep monophyletic lineages called subclusters 5.1 to 5.3, *Ca.* Synechococcus spongarium (=SC 5.4?), and *Proc.*

Subcluster 5.1 is also widely diversified and encompasses mainly strictly marine representatives, except for clade VIII whose members are found in brackish waters.

Subcluster 5.2 (= *Cyanobium*) is widely diversified and encompasses brackish/halotolerant and freshwater, but no strictly marine representatives.

Subcluster 5.3 is fairly little diversified yet encompasses both strictly marine and freshwater (*Syn. lacustris*), but no brackish representatives.

Genetic diversity within *Synechococcus* SC 5.1

The ISME Journal (2012) 6, 285–297

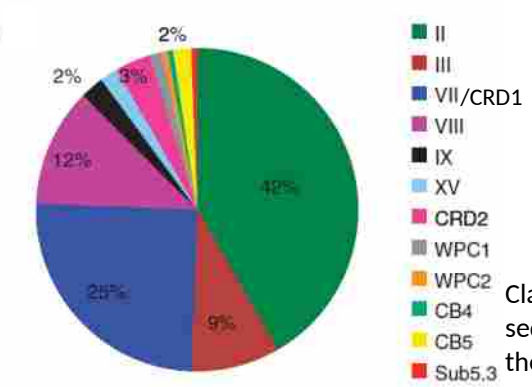
Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans

Sijun Huang[1,2], Steven W Wilhelm[3], H Rodger Harvey[4], Karen Taylor[4], Nianzhi Jiao[1] and Feng Chen[2]

The number of clades within SC 5.1 varies between 16-20, depending on the marker. In this tree, Huang et al. found 18 clades but clades VII and CRD1 are merged, whereas they correspond to different clades.

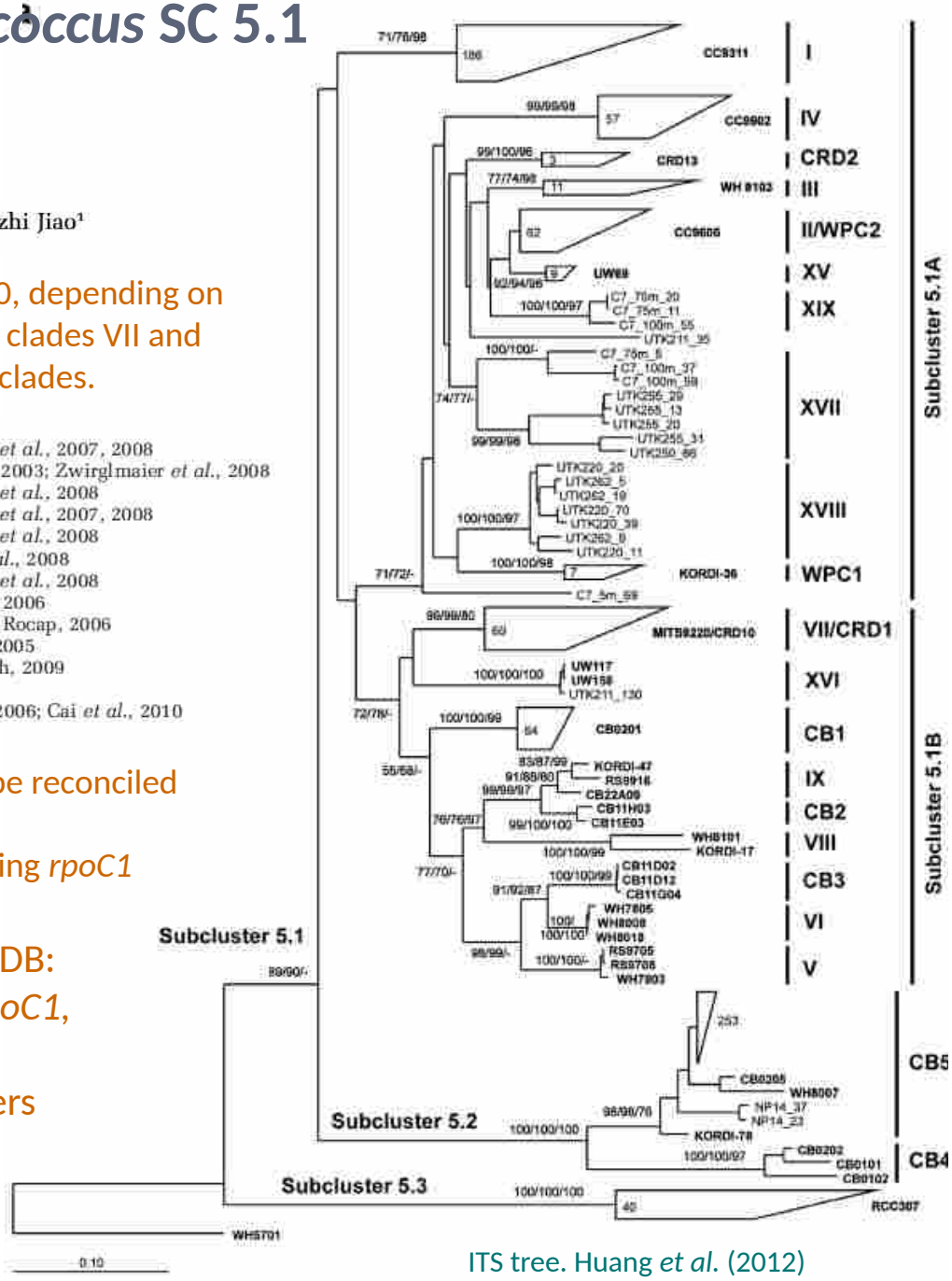
Subcluster 5.1		
I	Temperate to polar waters, co-occurrence with clade IV	Zwirglmaier *et al.*, 2007, 2008
II	Tropical/subtropical waters	Fuller *et al.*, 2003; Zwirglmaier *et al.*, 2008
III	Oligotrophic waters	Zwirglmaier *et al.*, 2008
IV	Temperate to polar waters, co-occurrence with clade I	Zwirglmaier *et al.*, 2007, 2008
V–VII	Oceanic waters	Zwirglmaier *et al.*, 2008
VIII	Hypersaline waters	Dufresne *et al.*, 2008
IX	Rarely detected	Zwirglmaier *et al.*, 2008
XI–XIV	Gulf of Aqaba subsurface water (predominant)	Penno *et al.*, 2006
XV/XVI	Sargasso Sea (isolates and environmental sequences)	Ahlgren and Rocap, 2006
CRD1/CRD2	Costa Rica upwelling dome (predominant)	Saito *et al.*, 2005
WPC1/WPC2	East China Sea and East Sea (isolates and environmental sequences)	Choi and Noh, 2009
CB1–CB3	Chesapeake Bay, summer	Chen *et al.*, 2006; Cai *et al.*, 2010

Clades that have different names between markers can be reconciled using genomes, since they contain all markers:
e.g., CRD2 defined using ITS = EnvB using *petB* = FLU1 using *rpoC1*



Classification of *Syn.* ITS sequences retrieved from the GOS database

⇒ CyanoMarks DB:
16S rRNA, ITS, *rpoC1*, *petB*, *psbO*, pigment markers



ITS tree. Huang *et al.* (2012)

Genetic diversity within *Synechococcus* SC 5.1

PNAS | August 6, 2019 | vol. 116 | no. 32 | 15973–15978

Single-cell genomics unveiled a cryptic cyanobacterial lineage with a worldwide distribution hidden by a dinoflagellate host

Takuro Nakayama[a,1], Mami Nomura[b,2], Yoshihito Takano[c], Goro Tanifuji[d], Kogiku Shiba[b], Kazuo Inaba[b], Yuji Inagaki[e], and Masakado Kawata[a]

Pelagic heterotrophic dinoflagellates of the genus *Ornithocercus* possess *Syn.*-like symbionts in a crown-shaped extension of the thecal plate which forms an extracellular chamber.

Phylogenetically these symbionts fall at the basis of SC 5.1.

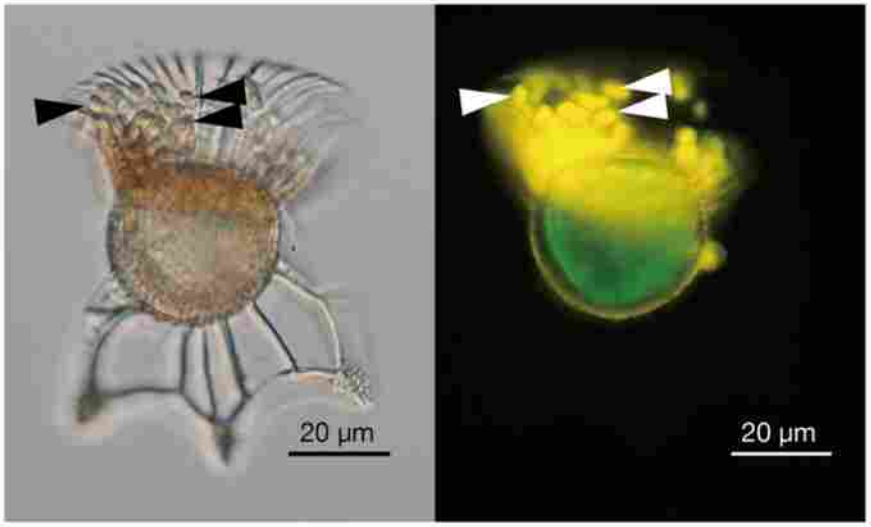
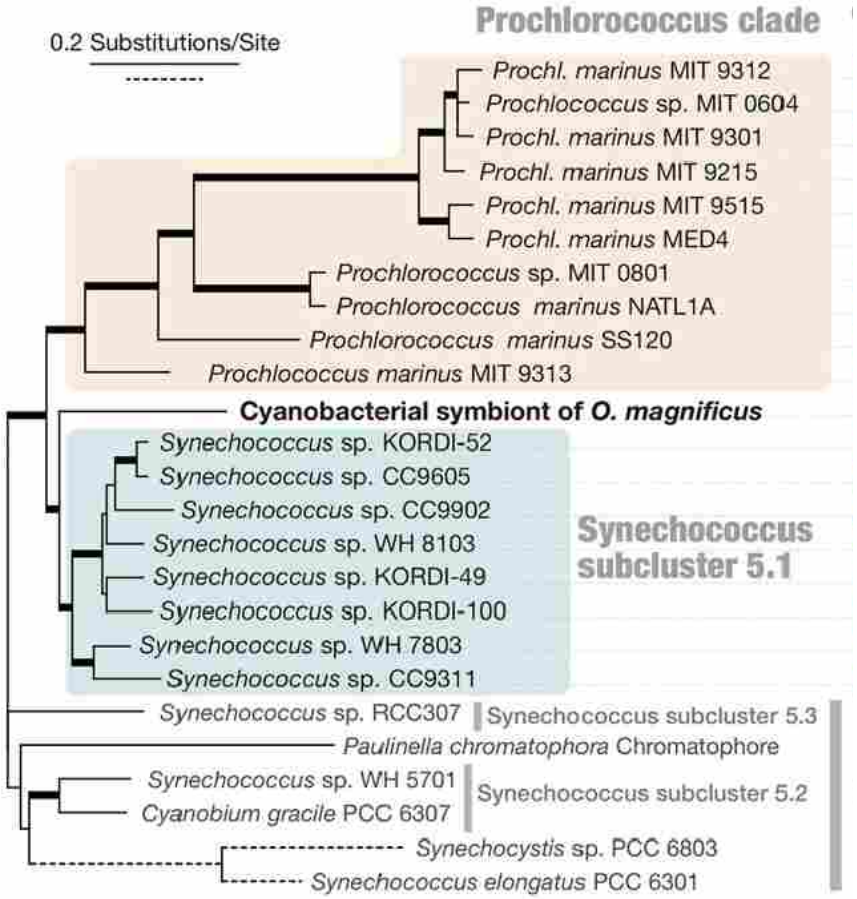


Fig. 1. Microscopy images of an *Ornithocercus magnificus* dinoflagellate with its cyanobacterial symbiont. (*Left*) Bright-field image of *O. magnificus* (*Right*) Blue-light epifluorescence



Phylogenomic tree
Nakayama *et al.* (2019)

Croucher course 2 outline

I. Genetic diversity and evolution of picocyanobacteria

II. Ecological distribution of picocyanobacteria: are clades equivalent to ecotypes ?

III. Outcomes from comparative genomics

IV. Outcomes from metagenomics

Ecological distribution of *Synechococcus* SC 5.1 clades

Environmental Microbiology (2008) 10(1), 147–161

Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes

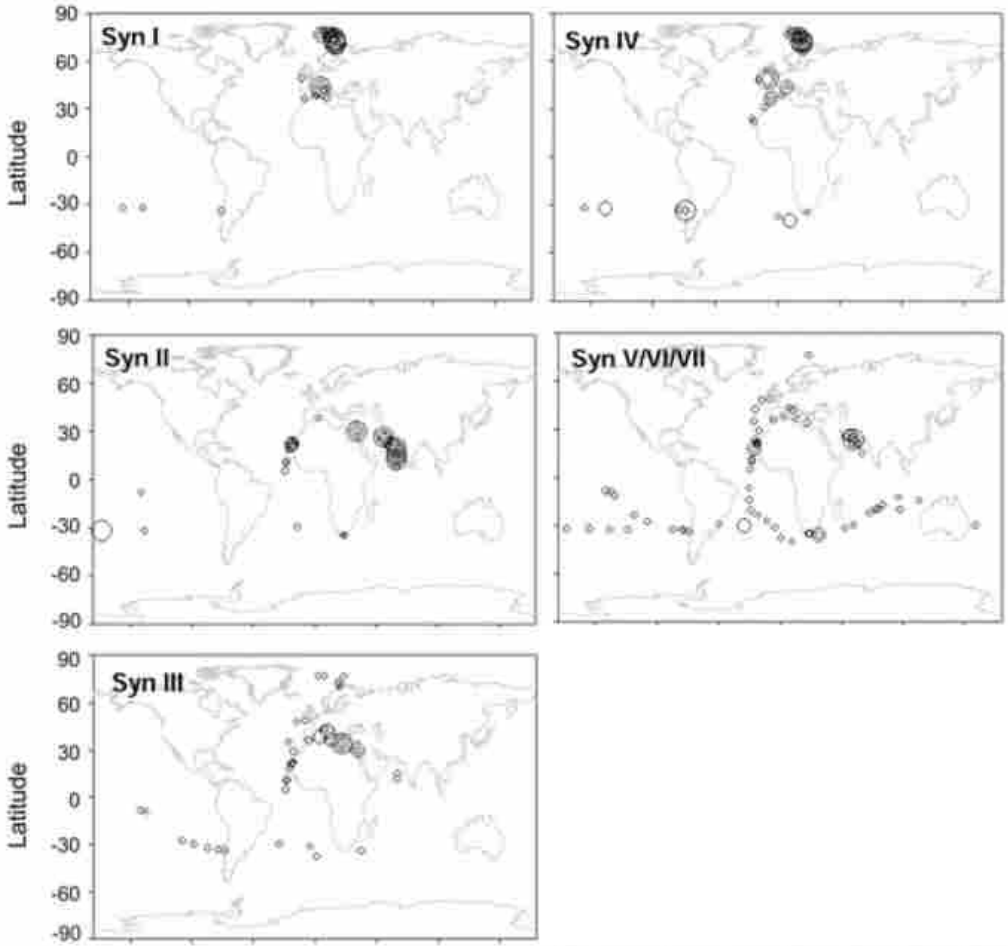
Katrin Zwirglmaier,[1] Ludwig Jardillier,[1]
Martin Ostrowski,[1] Sophie Mazard,[1]
Laurence Garczarek,[2] Daniel Vaulot,[2] Fabrice Not,
Ramon Massana,[3] Osvaldo Ulloa[4] and
Dave J. Scanlan[1]*

Table 1. Details of cruises; for cruise tracks see Fig. 1.

Cruise	Date	Geographic region	Latitude/longitude	Sampling	Reference
1 – AMBITION	September 2001	Arabian Sea	1°S–26°N, 56–67°E	11 stations, 6 depths	Fuller *et al.* (2006)
2 – AMT15	September–October 2004	Atlantic Meridional Transect	49°N–40°S, 25°W–10°E	21 stations, 6 depths	Zwirglmaier *et al.* (2007)
3 – ARCTIC	August 2002	Arctic Ocean	70–79°N, 3–23°E	11 stations, 5 depths	This study
4 – BEAGLE	August 2003–January 2004	Circumnavigation of the southern hemisphere	20–32°S	15 stations, surface only	Bouman *et al.* (2006) and this study
5 – BIOSOPE	October–December 2004	Eastern South Pacific Ocean	8–34°S, 141–72°W	11 stations, 6 depths	This study
6 – VANC10MV	May–June 2003	Indian Ocean	35–12°S, 24–113°E	15 stations, 6 depths	This study
7 – PROSOPE	September–October 1999	Mediterranean Sea	31–42°N, 10°W–22°E	9 stations, 5 depths	Garczarek *et al.* (2007)
8 – RED SEA	February 1999–January 2000	Red Sea	29°N, 34°E	One station, 6 depths over an annual cycle	Fuller *et al.* (2005)

Method: Relative clade abundance determined using dot blot hybridization or 16S rRNA probes for 8 cruises.

- Clades I and IV co-occur at high latitudes (>30°N/S) *predominantly in coastal areas*.
- Clade II is their (sub)tropical counterpart occurring between 30°S-30°N, *usually in coastal/continental shelf zones*.
- Clades III and V/VI/VII occur in moderate abundance.
- Clades VIII, IX and X occur in low abundance.



Ecological distribution of *Synechococcus* SC 5.1 clades

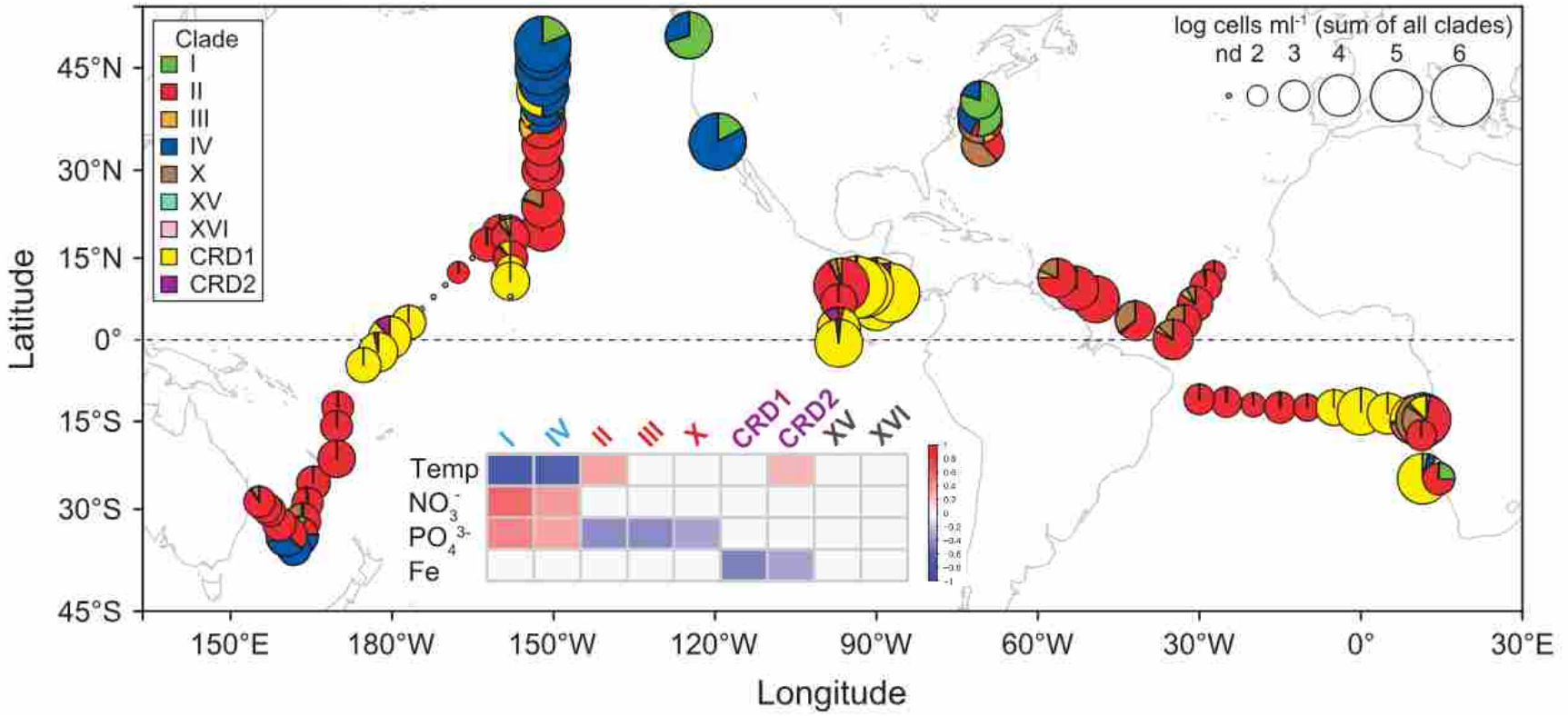
The ISME Journal (2016) 10, 333–345

Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron

Jill A Sohm[1,4], Nathan A Ahlgren[1,2,4], Zachary J Thomson[2], Cheryl Williams[2], James W Moffett[1], Mak A Saito[3], Eric A Webb[1] and Gabrielle Rocap[2]

Method: Abundance of 10 *Syn*. Clades measured by quantitative qPCR of ITS amplicons in surface waters of the Atlantic and Pacific Oceans.

- Clades I and IV prevalent in colder, mesotrophic waters.
- Clades II, III and SC 5.3 (= clade X) dominated in warm, oligotrophic waters.
- Clades CRD1 and CRD2 restricted to low-Fe areas.



Ecological distribution of *Synechococcus* SC 5.1 clades

PNAS | Published online June 2, 2016 | E3365–E3374

Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria

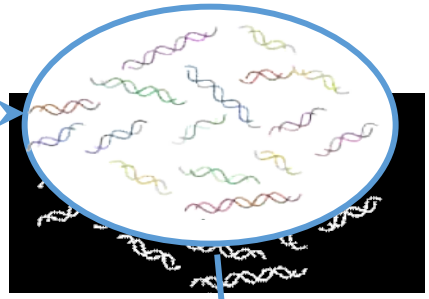
Gregory K. Farrant[a,1,2], Hugo Doré[a,1], Francisco M. Cornejo-Castillo[b], Frédéric Partensky[a], Morgane Ratin[a], Martin Ostrowski[c], Frances D. Pitt[d], Patrick Wincker[e], David J. Scanlan[d], Daniele Iudicone[f], Silvia G. Acinas[b], and Laurence Garczarek[a,3]



DNA/RNA extraction

Complete sequencing

METAGENOME



in silico extraction



miTag

petB (cytochrome b_6)

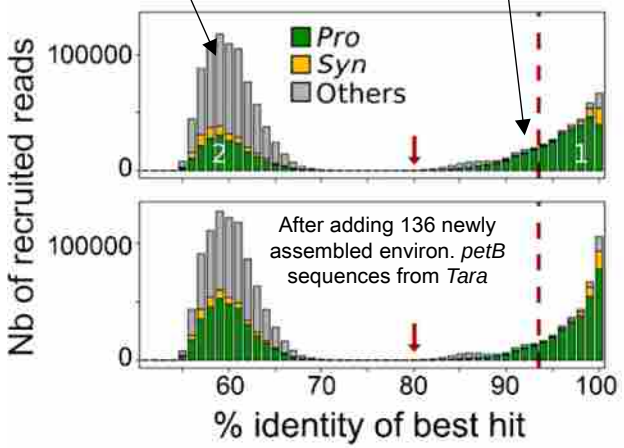
petB reference database
(585 non-redundant sequences incl. outgroups)

Retrieve only tags assigned to *Proc.* or *Syn.*

Assign reads to *Proc.* or *Syn.* clades

Non-specific

Field *Proc.* or *Syn petB*

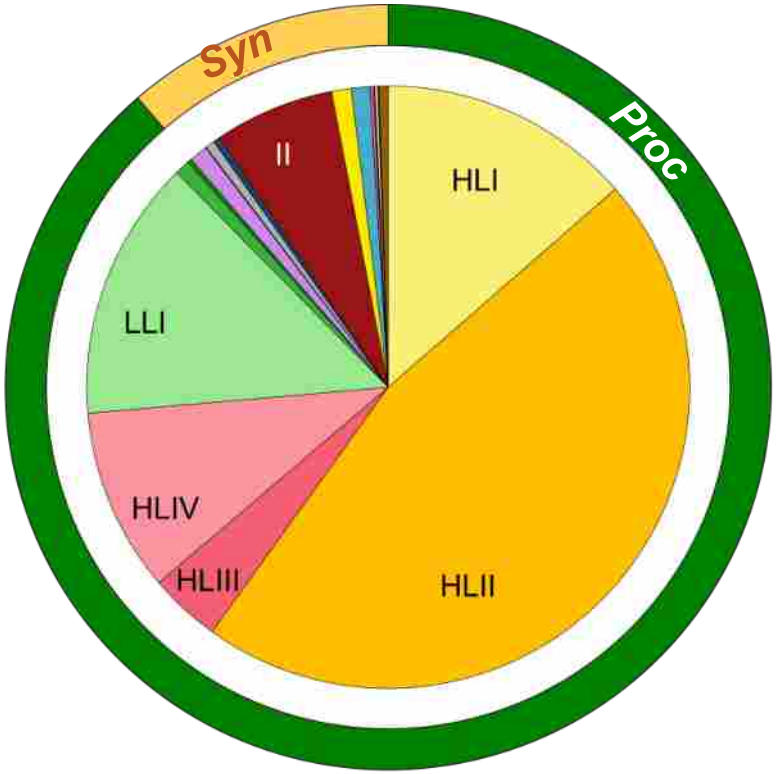


against the *petB* reference database

Farrant, Doré *et al.* (2016)

Relative clade abundance of *Prochlorococcus* and *Synechococcus*

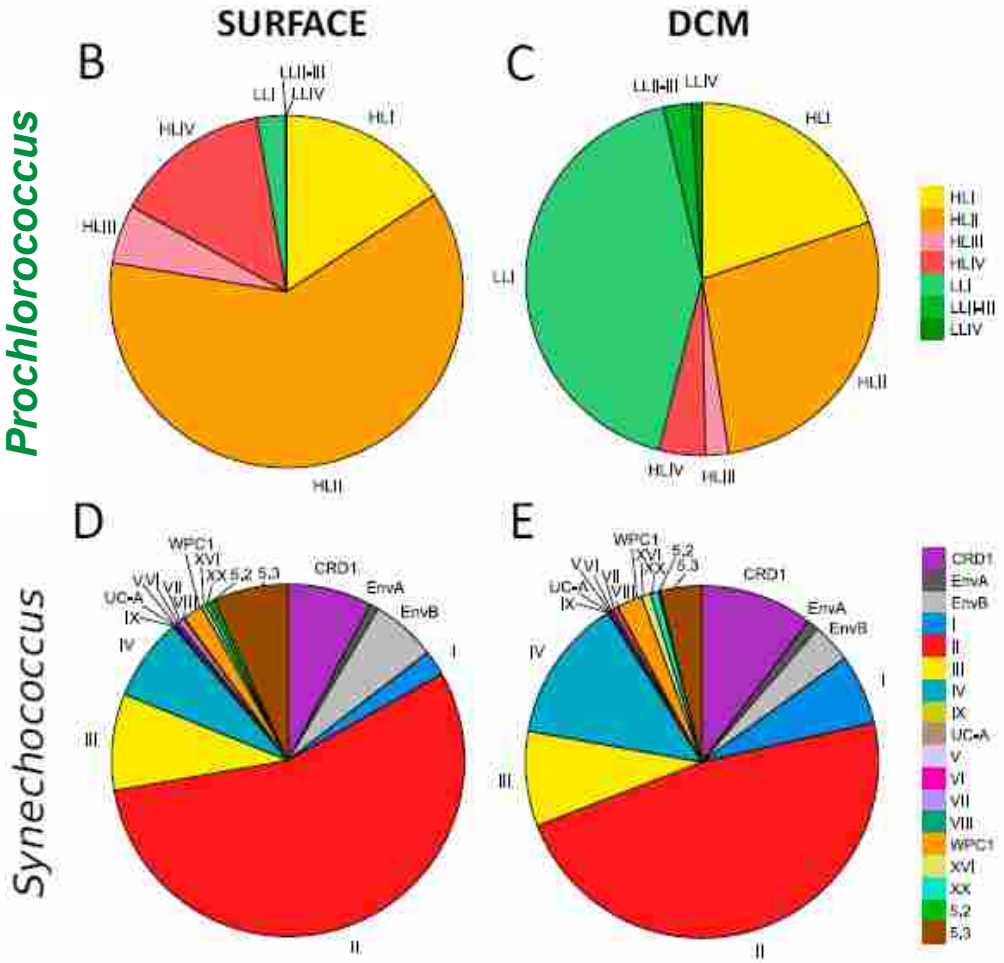
SURFACE + DCM



Large dominance (90%) of *Proc.* over *Syn.* because *Tara* Oceans essentially sampled open ocean waters.

Essentially HLI-IV ecotypes in surface waters.

LLI ecotype dominates at the DCM (few LLI-IV) located deeper in the euphotic zone

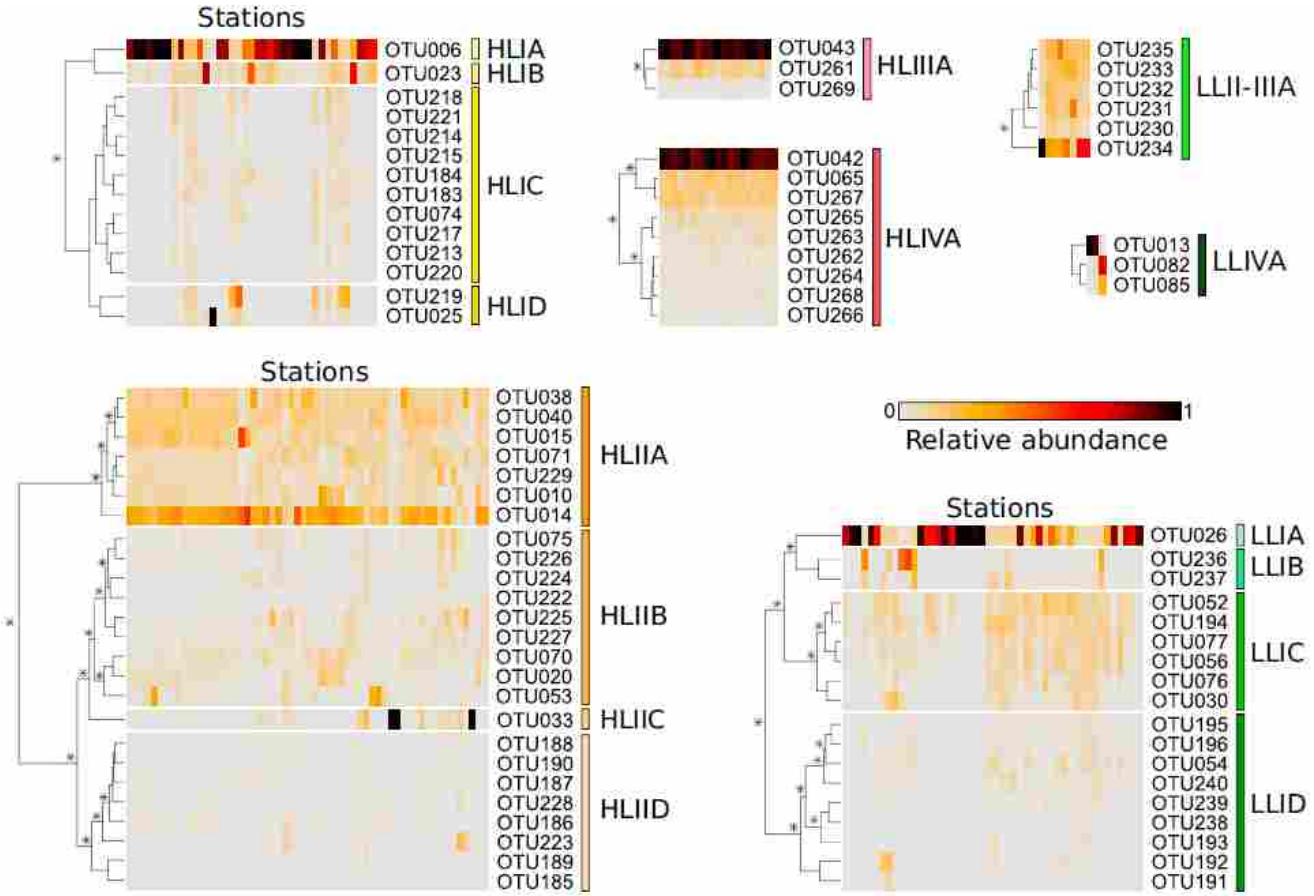


Farrant, Doré *et al.* 2016 *PNAS*

Defining « Ecologically Significant Taxonomical Units » within Clades

Method: Determination of ESTUs based on the distribution patterns of within-clade 94% OTUs.

At each station, the number of reads assigned to a given OTU was normalized by the total number of reads assigned to the clade at this station. OTUs were then hierarchically clustered (Bray–Curtis distance) according to their distribution pattern and those sharing a significantly similar pattern were grouped into ESTUs.

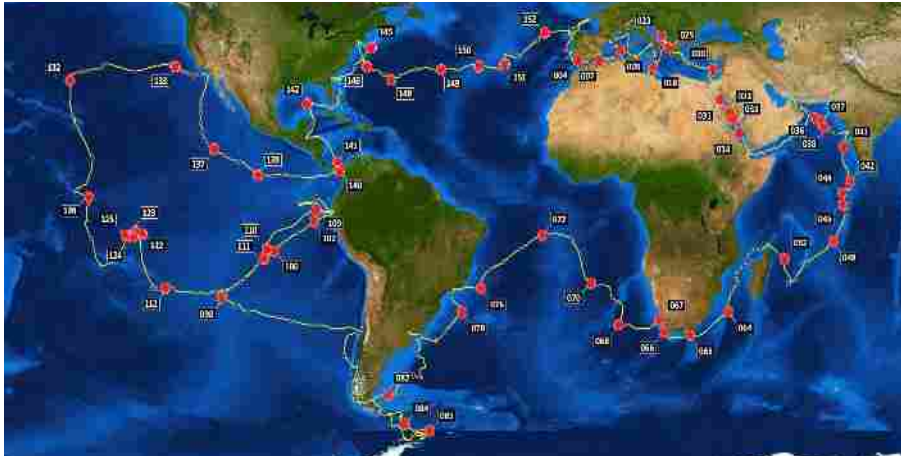


Distribution pattern of *petB* OTUs (94% ID)

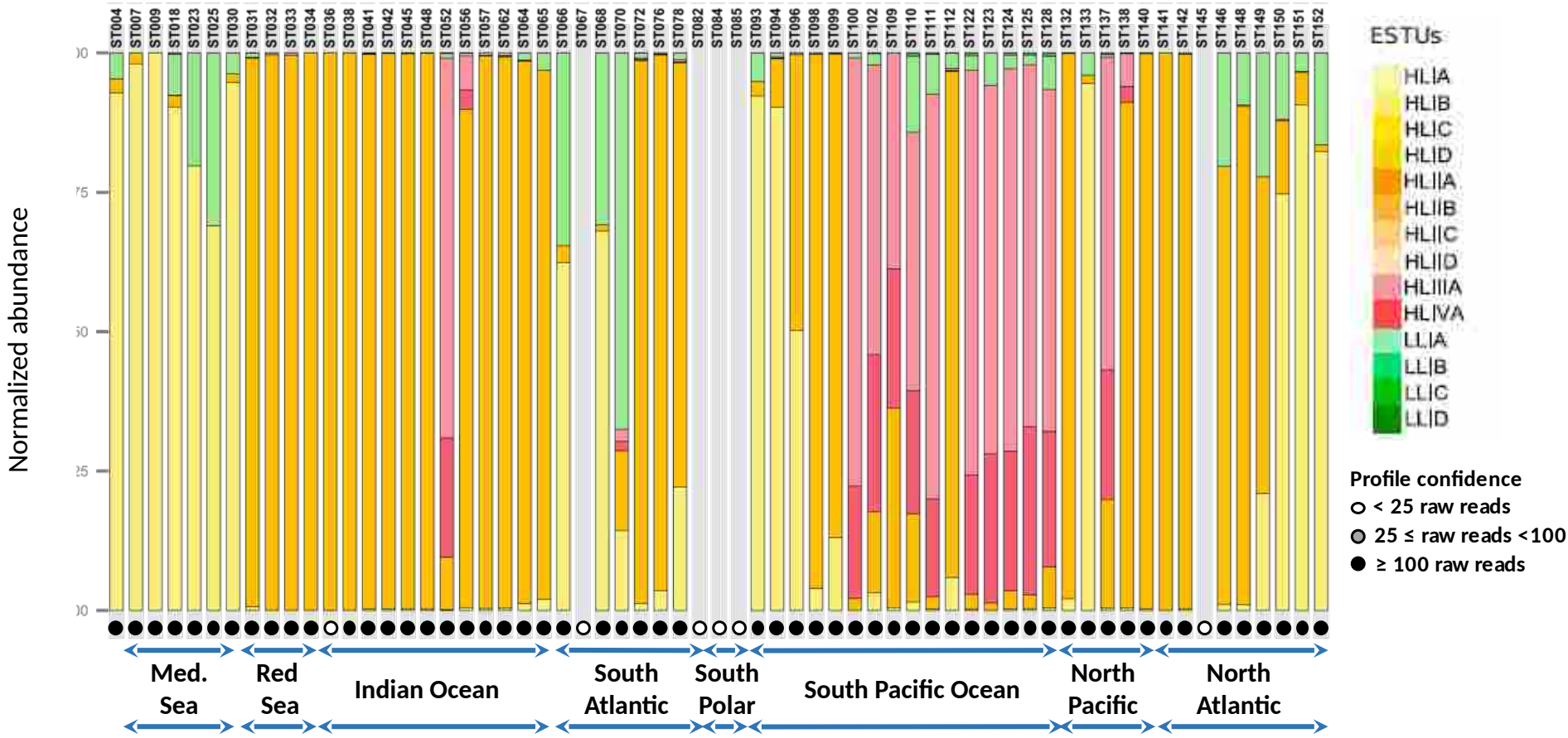
Prochlorococcus clades encompass between 1 and 4 ESTUs.

Farrant, Doré *et al.* (2016)

Distribution of *Synechococcus* ESTUs along the *Tara* Oceans transect

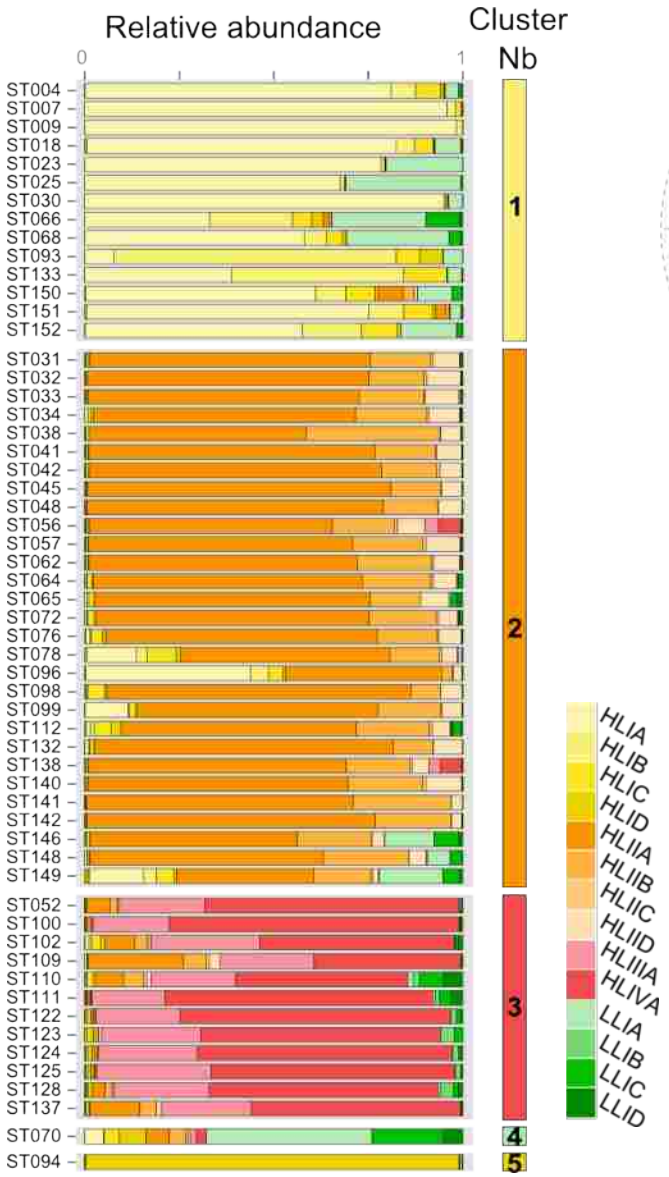


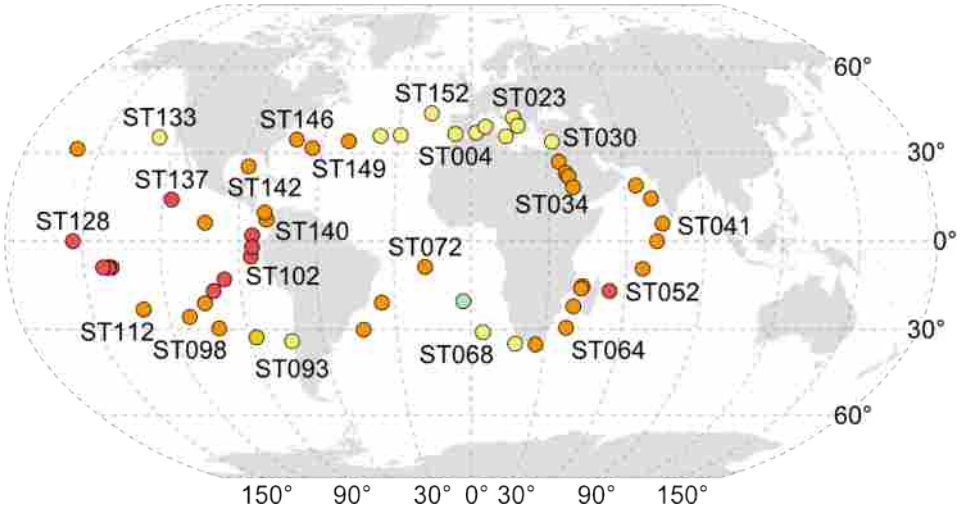
Prochlorococcus

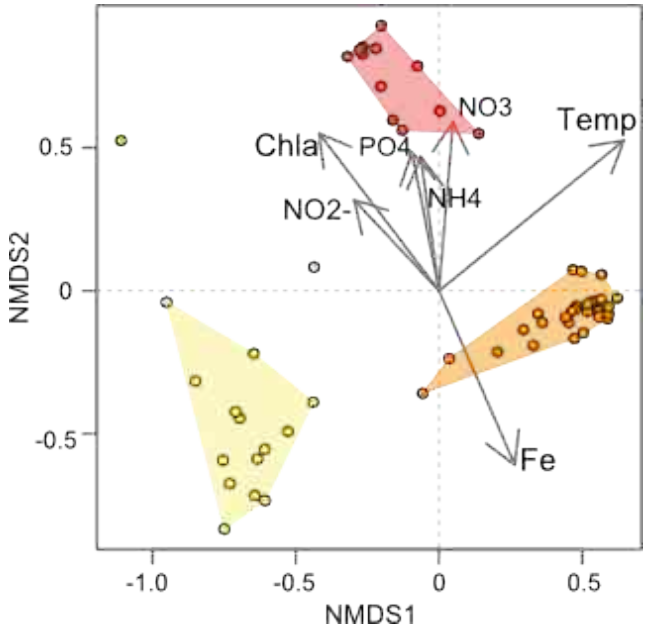


Farrant, Doré *et al.* (2016)

Global niche distribution of *Prochlorococcus* ESTUs

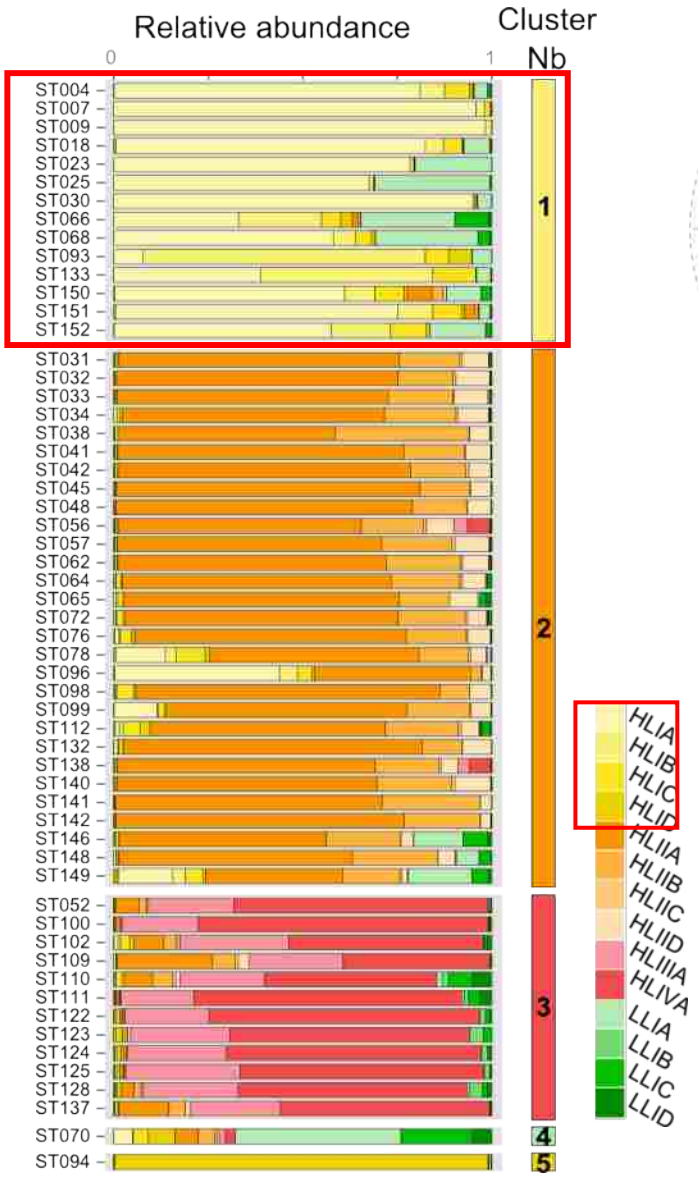


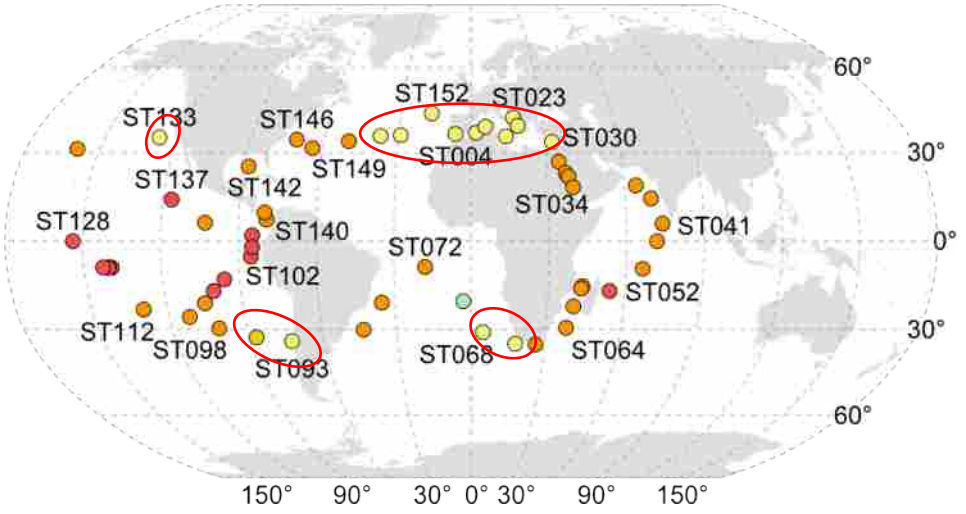


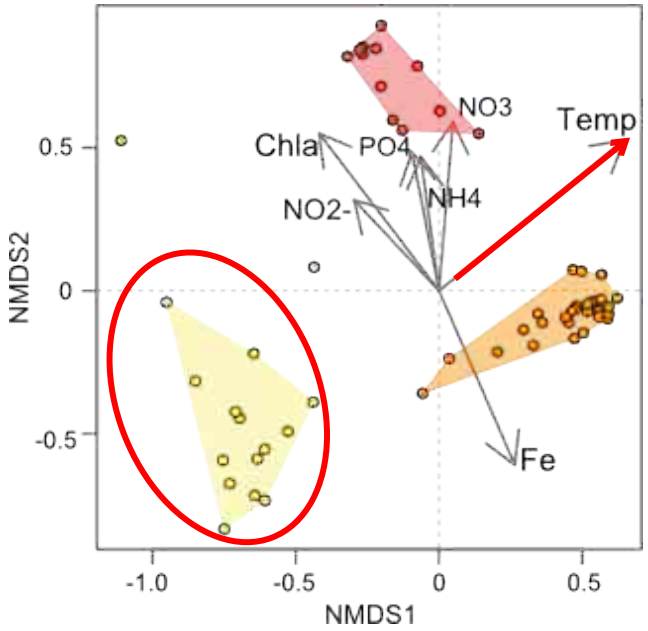


Farrant, Doré *et al.* (2016)

Global niche distribution of *Prochlorococcus* ESTUs



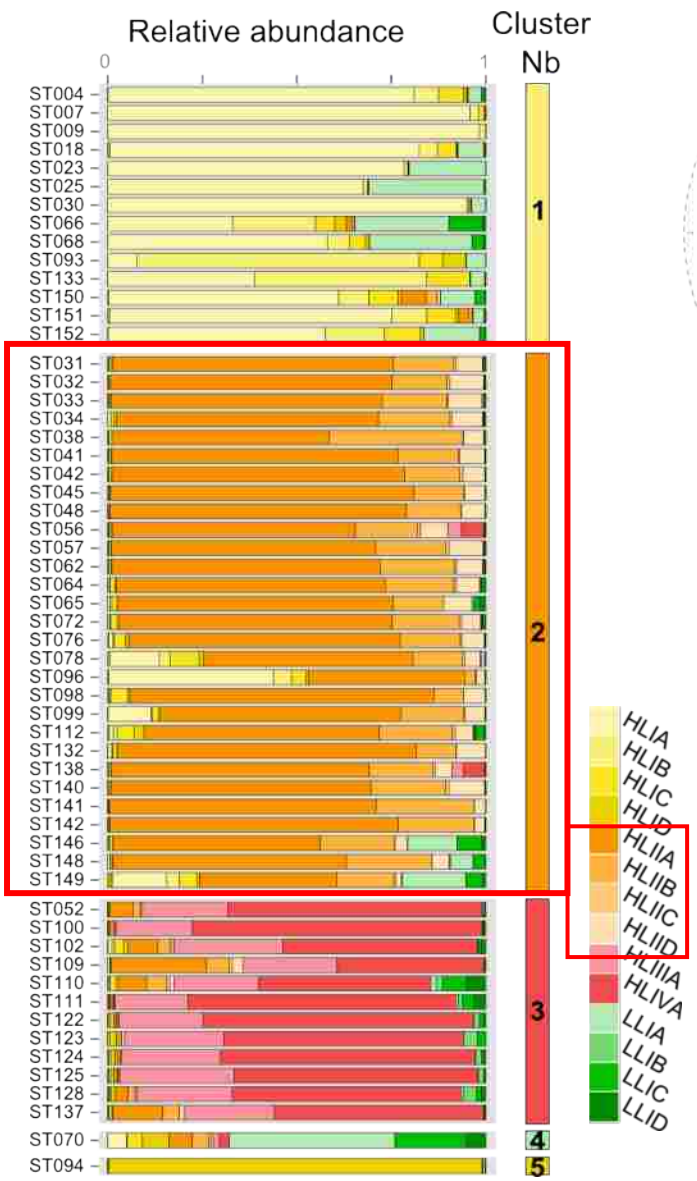


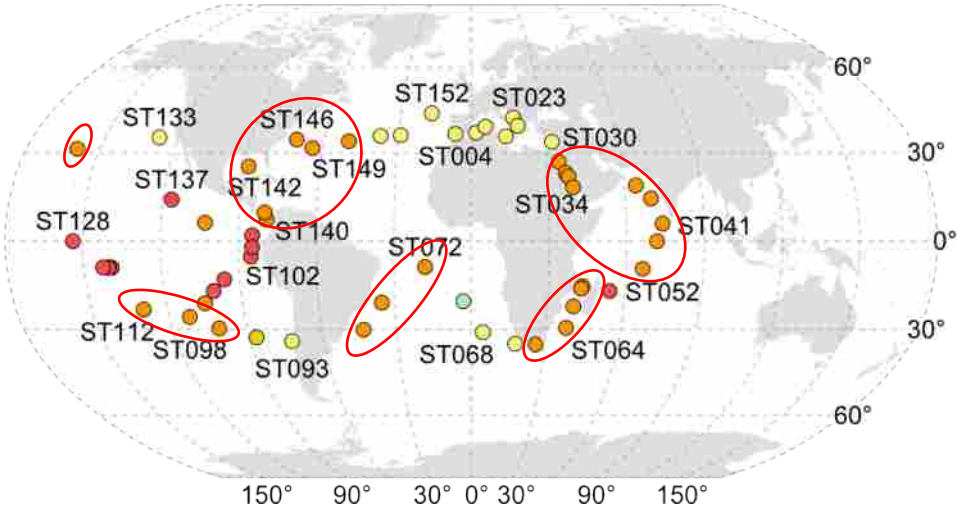


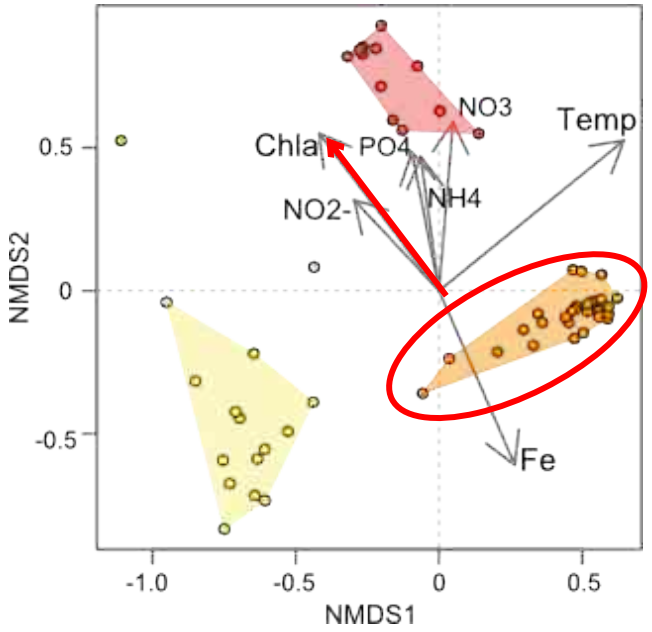
HLIA: **Cold** and temperate waters

Farrant, Doré *et al.* (2016)

Global niche distribution of *Prochlorococcus* ESTUs





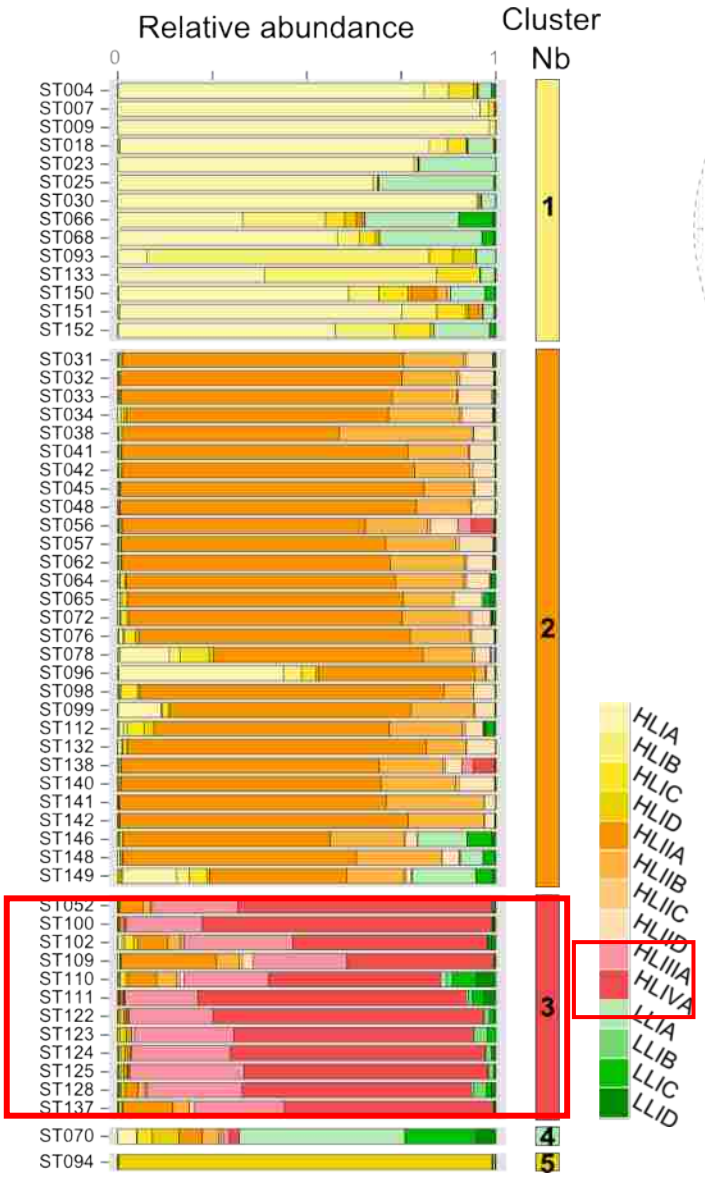


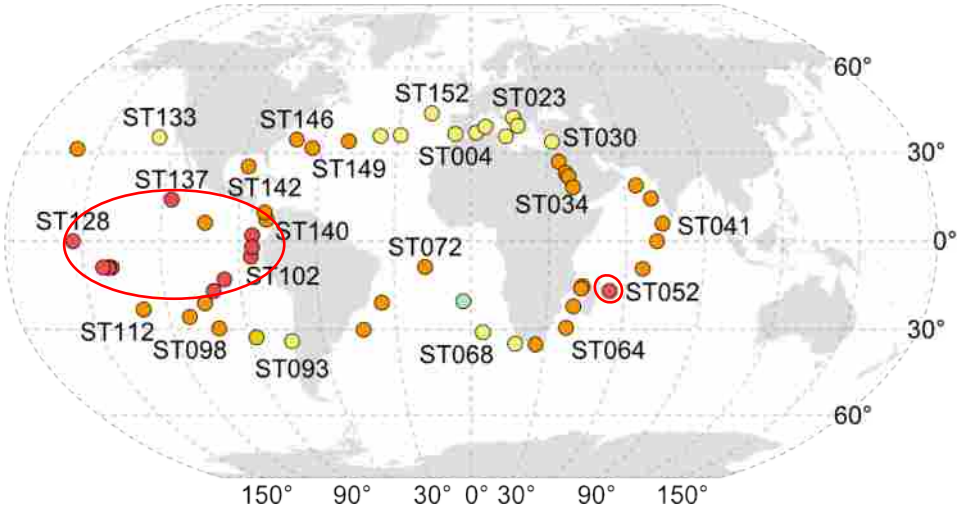
HLIA: Cold and temperate waters

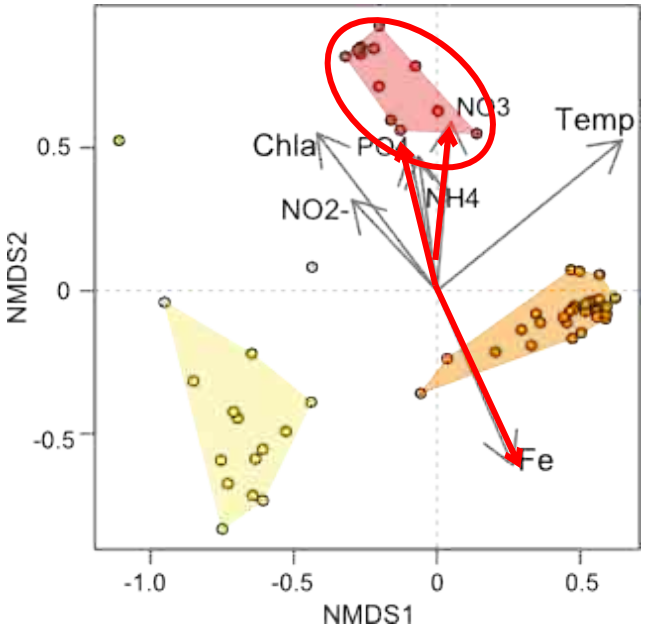
HLIIA: Warm oligotrophic Fe-replete waters

Farrant, Doré *et al.* (2016)

Global niche distribution of *Prochlorococcus* ESTUs







HLIA: Cold and temperate waters

HLIIA: Warm oligotrophic Fe-replete waters

HLIIIA/HLIVA: Warm N- and P-rich Fe-poor waters

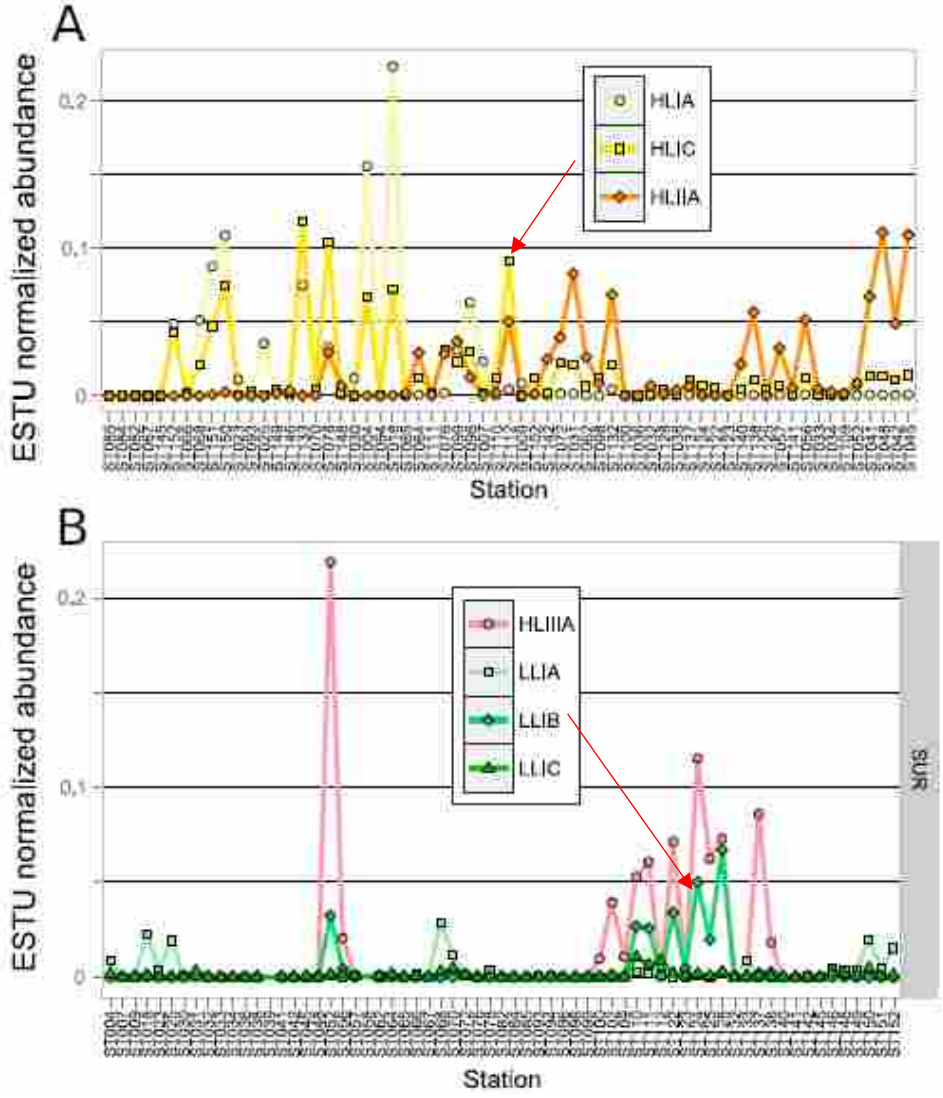
Farrant, Doré *et al.* (2016)

Atypical distribution of some minor *Prochlorococcus* ESTUs

Even though all clades comprised one major ESTU (generally called « A », e.g. HLIA, HLIIA, …) that behaves like previously known for the whole clade, the study revealed minor ESTUs with a strikingly different distribution, such as:

- HLIC has a wide niche/temperature range: it co-occurs with HLIA at high latitude and is the major HLI population in warm oligotrophic waters where HLIIA dominates the *Proc.* community.

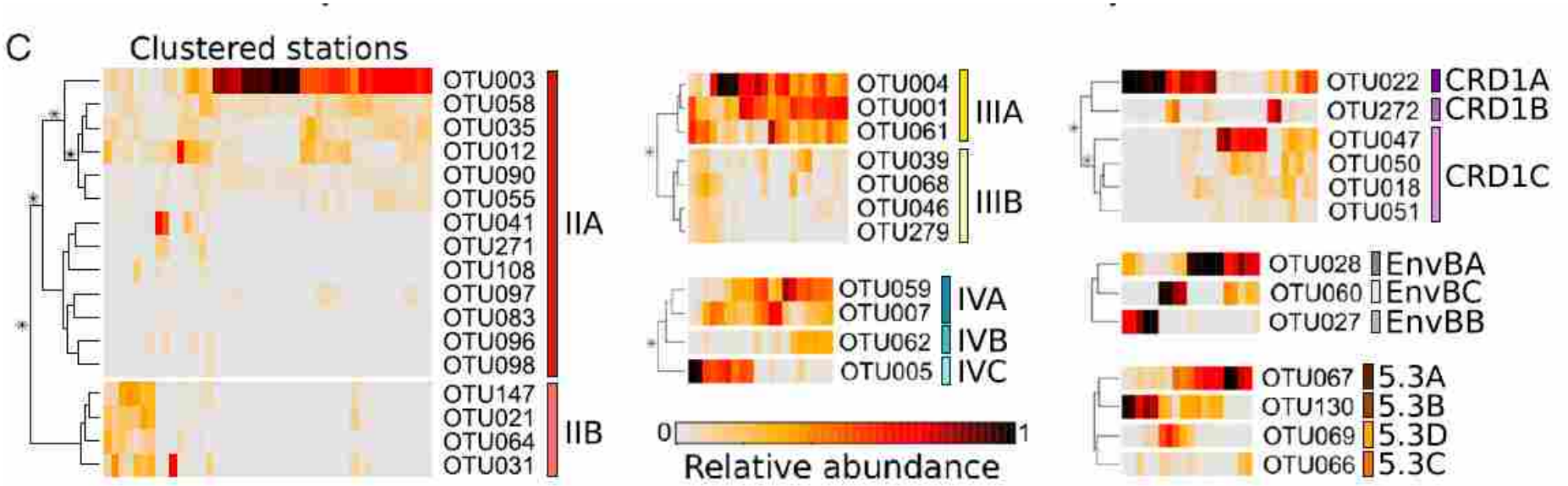
- LLIB, entirely comprised of newly assembled *petB* sequences, dominates the LLI population in low-iron (HNLC) areas where HLIIIA and HLIVA co-dominate in surface.



Defining « Ecologically Significant Taxonomical Units » within Clades

Method: Determination of ESTUs based on the distribution patterns of within-clade 94% OTUs.

At each station, the number of reads assigned to a given OTU was normalized by the total number of reads assigned to the clade at this station. OTUs were then hierarchically clustered (Bray–Curtis distance) according to their distribution pattern and those sharing a significantly similar pattern were grouped into ESTUs.

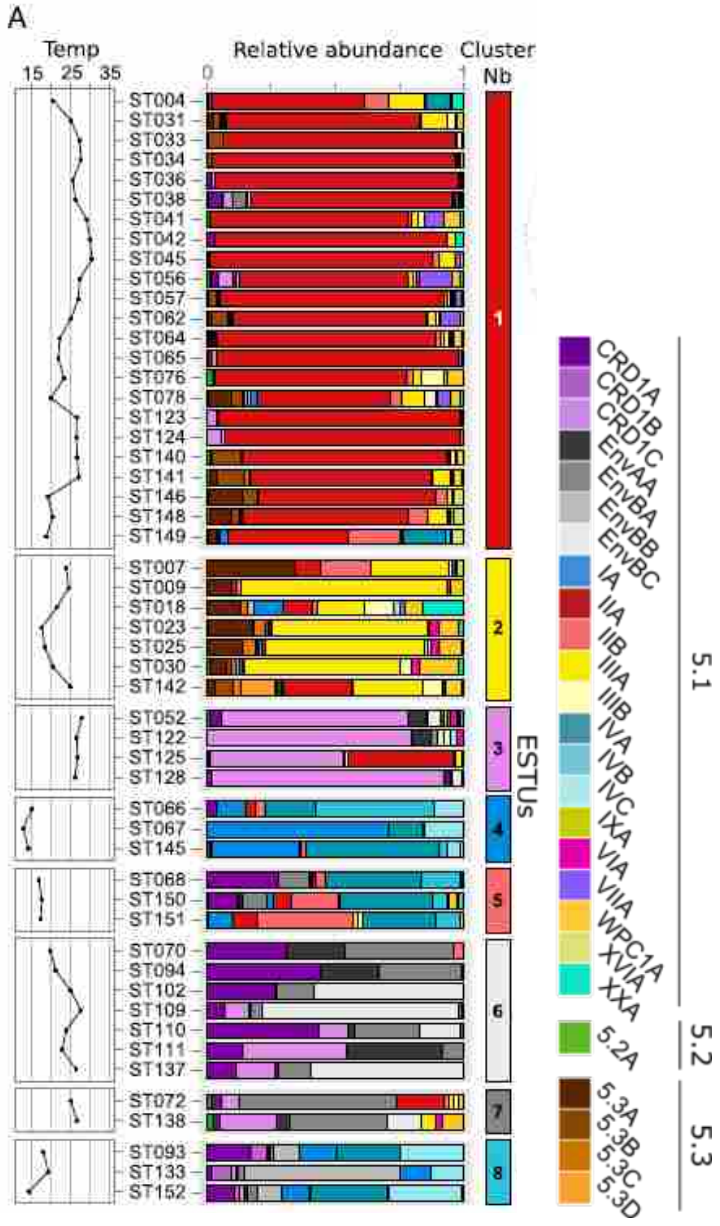


Distribution pattern of *petB* OTUs (94% ID)

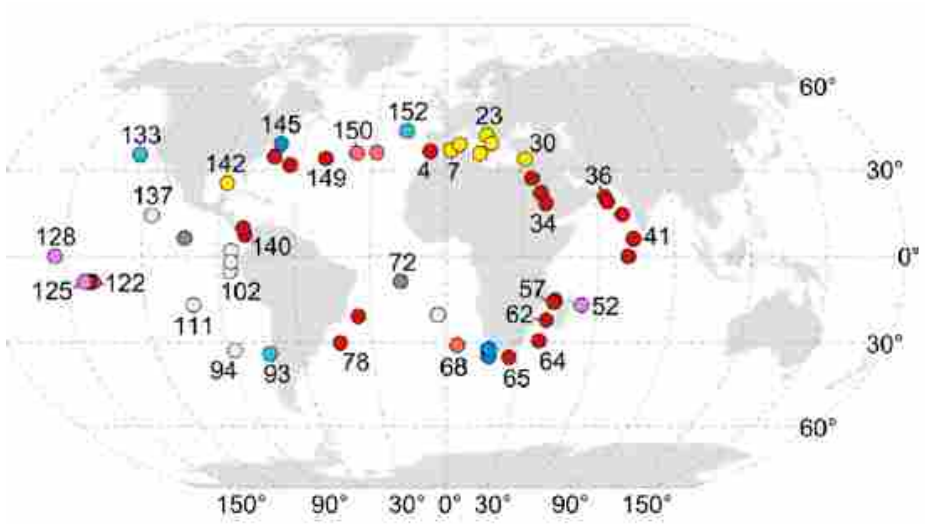
Most major *Syn*. clades encompassed two or more ESTUs.

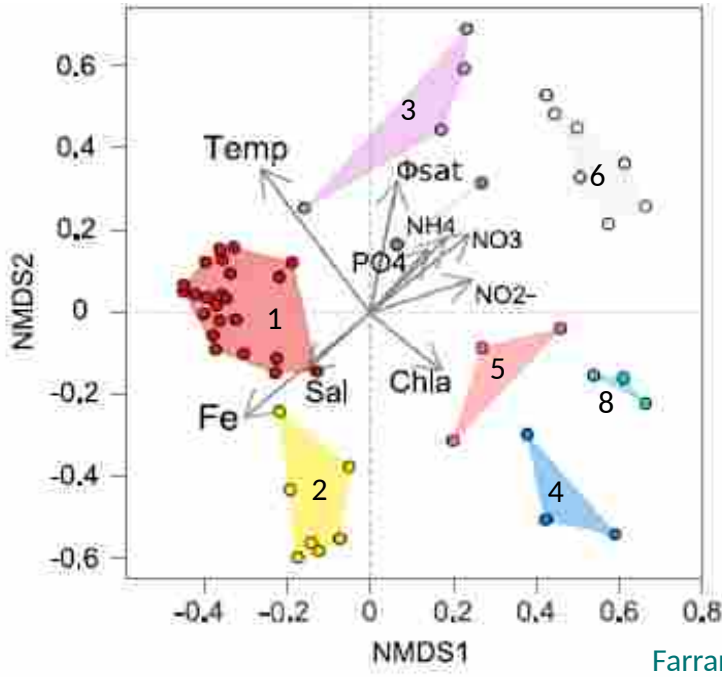
Farrant, Doré *et al.* (2016)

Global niche distribution of *Synechococcus* ESTUs



Clustering (Bray-Curtis distance) of stations based on ESTU Assemblage

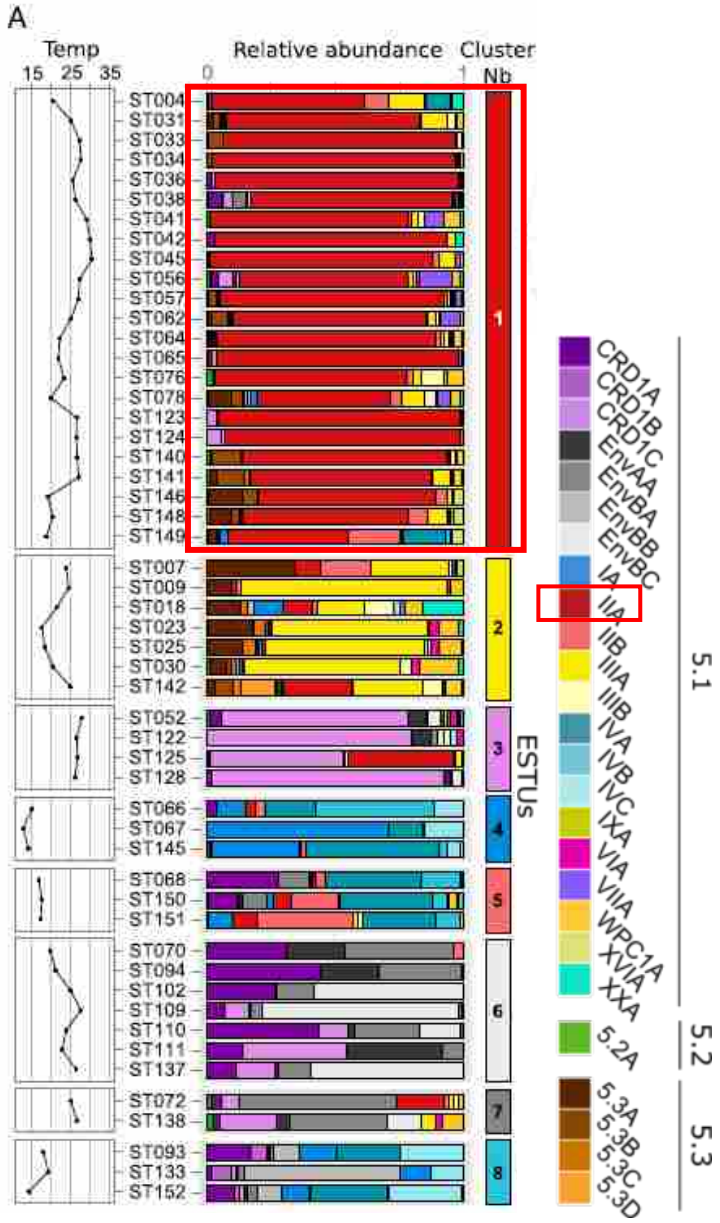




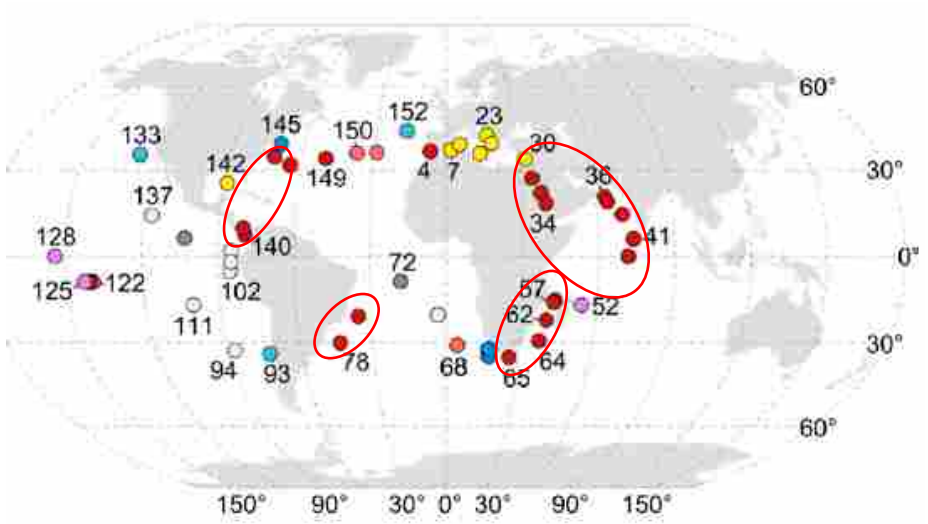
NMDS analysis of stations with physico-chemical parameters

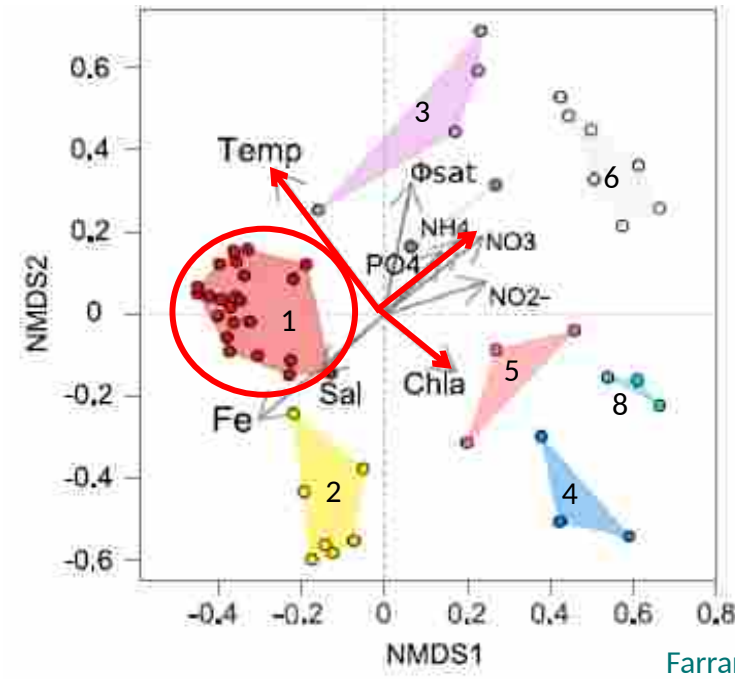
Farrant, Doré *et al.* (2016)

Global niche distribution of *Synechococcus* ESTUs



Clustering (Bray-Curtis distance) of stations based on ESTU Assemblage



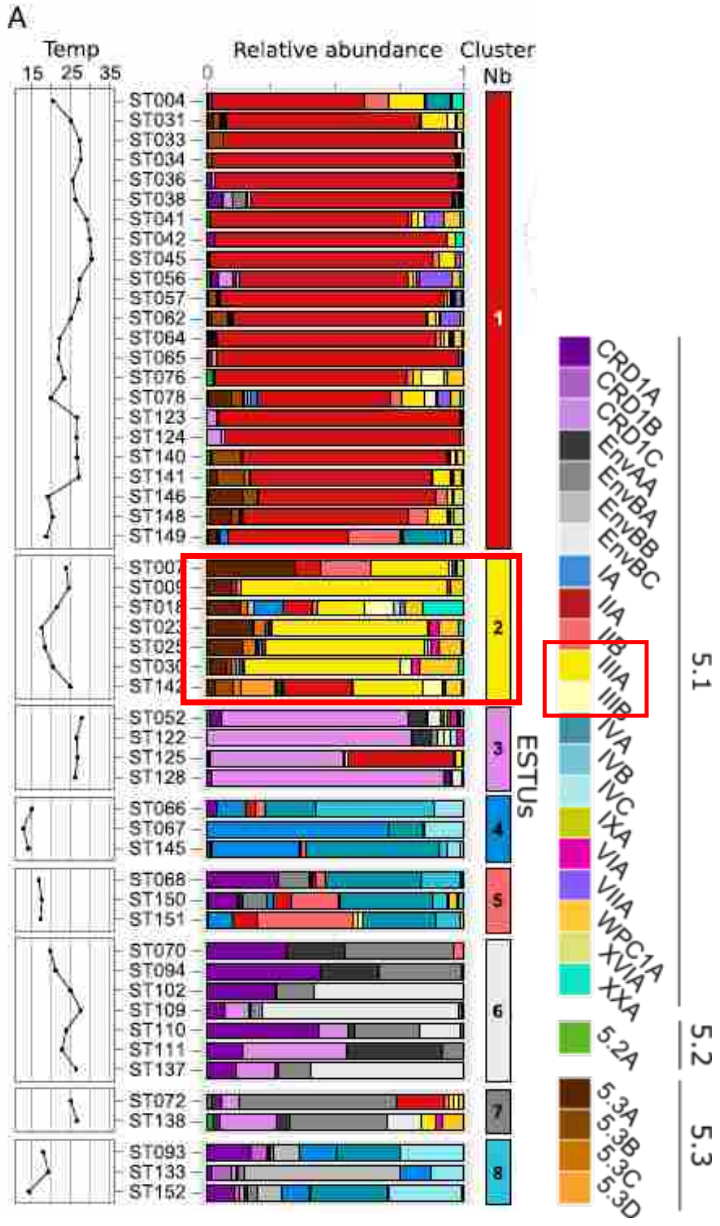


NMDS analysis of stations with physico-chemical parameters

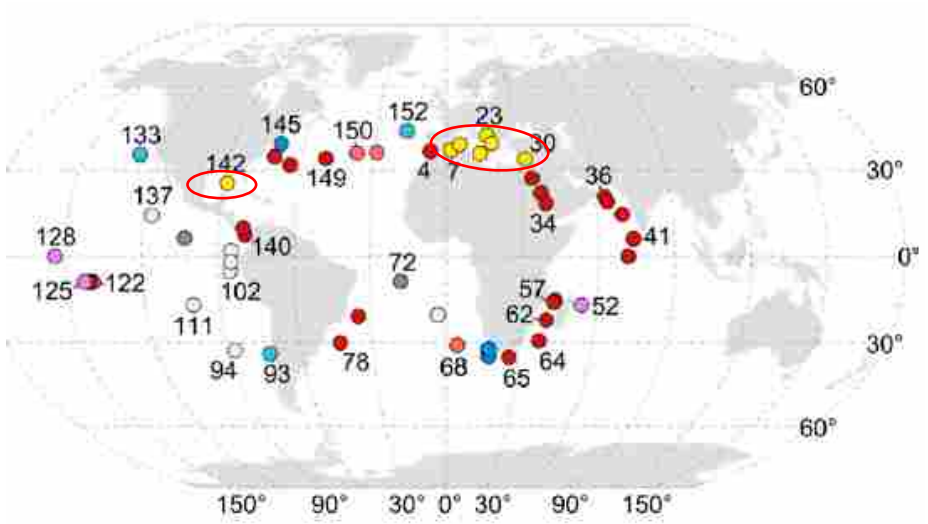
✓ **ESTU IIA-dominated assemblage** :
Warm
Chl- and N-poor
Fe-replete waters

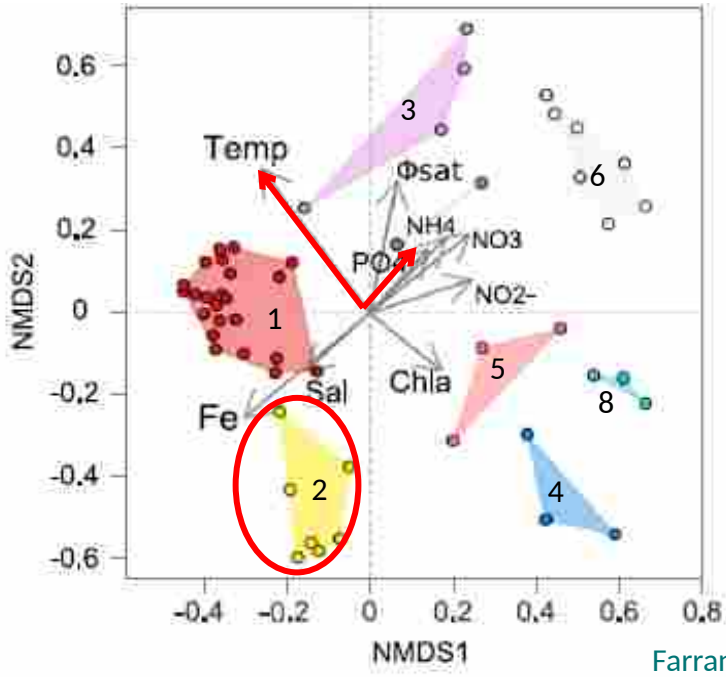
Farrant, Doré *et al.* (2016)

Global niche distribution of *Synechococcus* ESTUs



Clustering (Bray-Curtis distance) of stations based on ESTU Assemblage



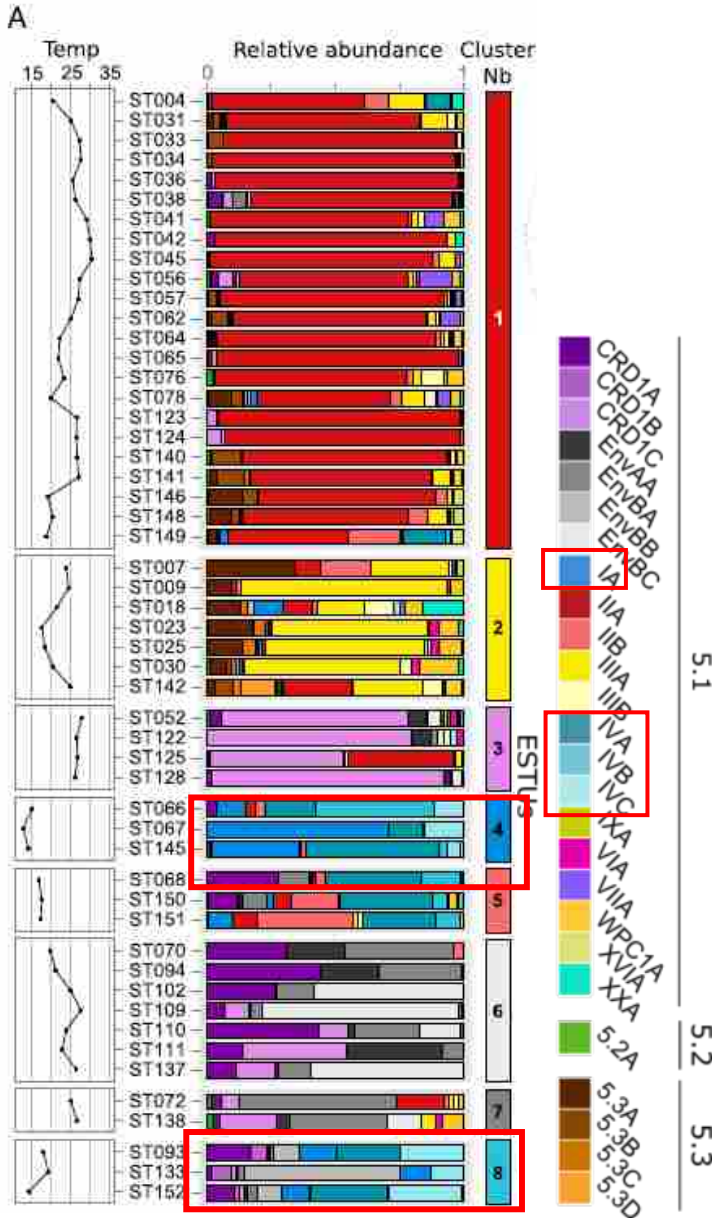


NMDS analysis of stations with physico-chemical parameters

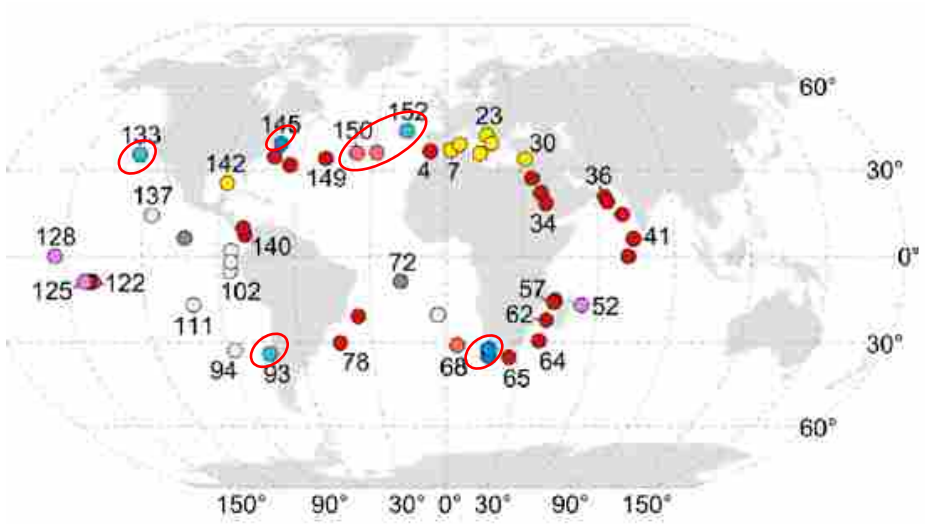
✓ **ESTU IIIA-dominated assemblage** :
Temperate
P-poor
Fe-replete waters

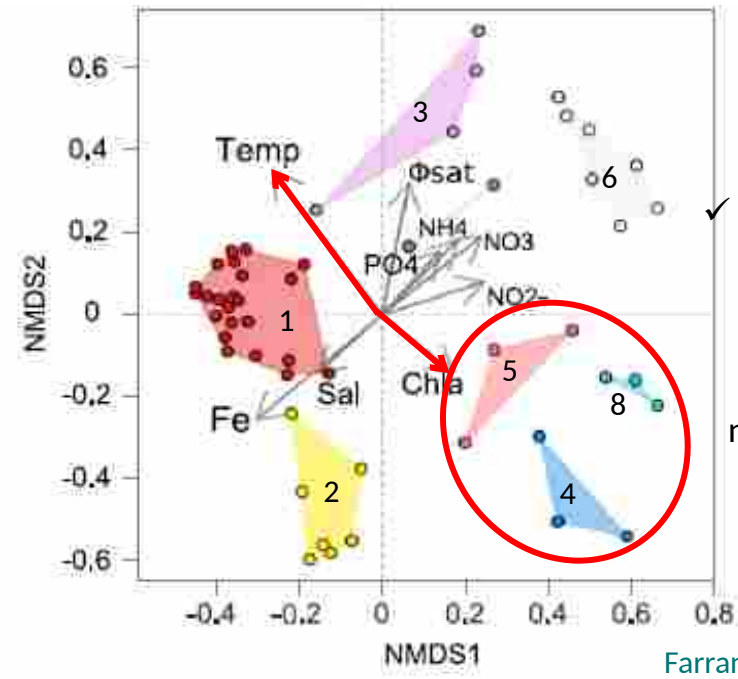
Farrant, Doré *et al.* (2016)

Global niche distribution of *Synechococcus* ESTUs



Clustering (Bray-Curtis distance) of stations based on ESTU Assemblage



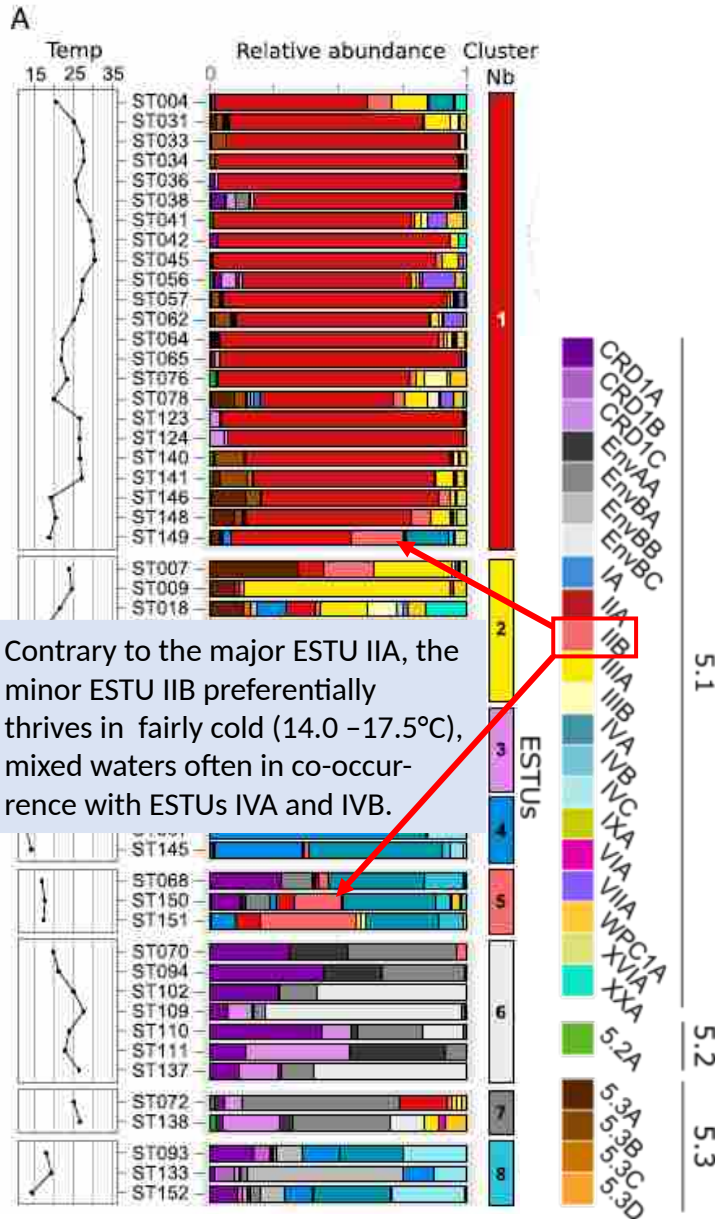


NMDS analysis of stations with physico-chemical parameters

✓ **ESTU IA and/or IVA-dominated assemblages :**
Cold
Chl-rich
nutrient-rich waters

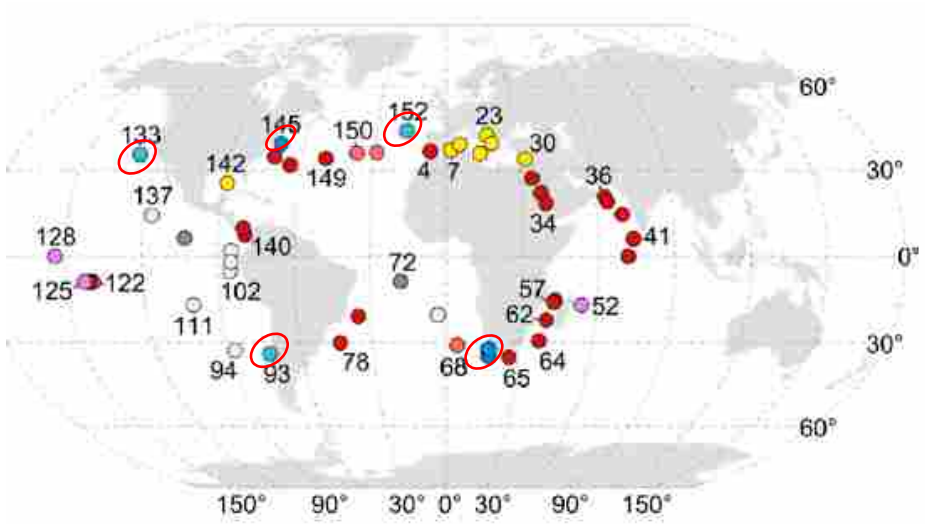
Farrant, Doré *et al.* (2016)

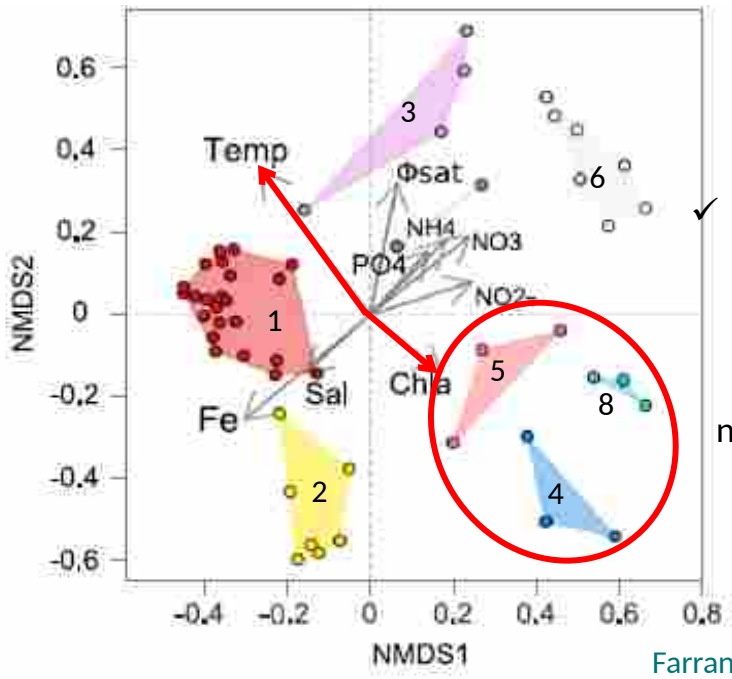
Global niche distribution of *Synechococcus* ESTUs



Clustering (Bray-Curtis distance) of stations based on ESTU Assemblage

Contrary to the major ESTU IIA, the minor ESTU IIB preferentially thrives in fairly cold (14.0 –17.5°C), mixed waters often in co-occur-rence with ESTUs IVA and IVB.



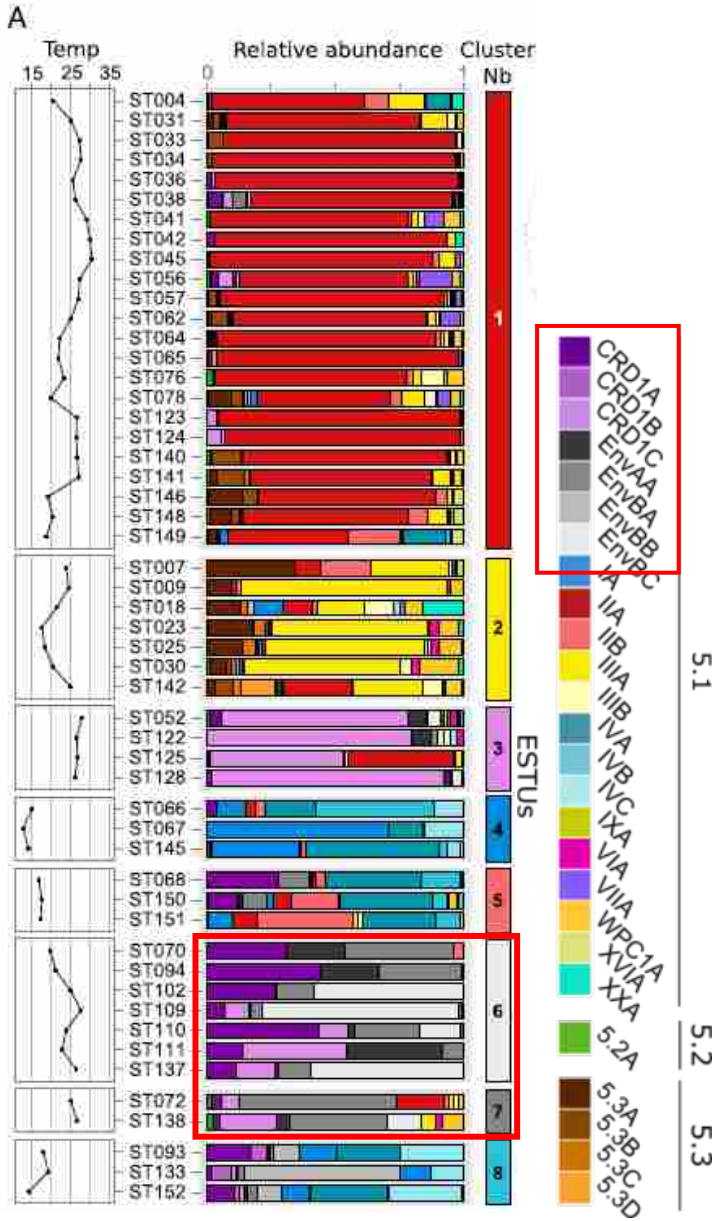


NMDS analysis of stations with physico-chemical parameters

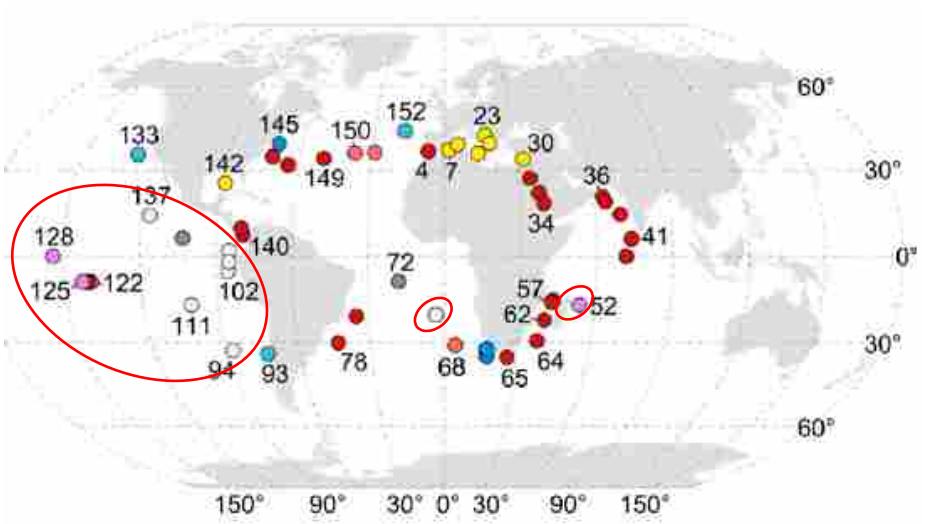
✓ **ESTU IA and/or IVA-dominated assemblages** :
Cold
Chl-rich
nutrient-rich waters

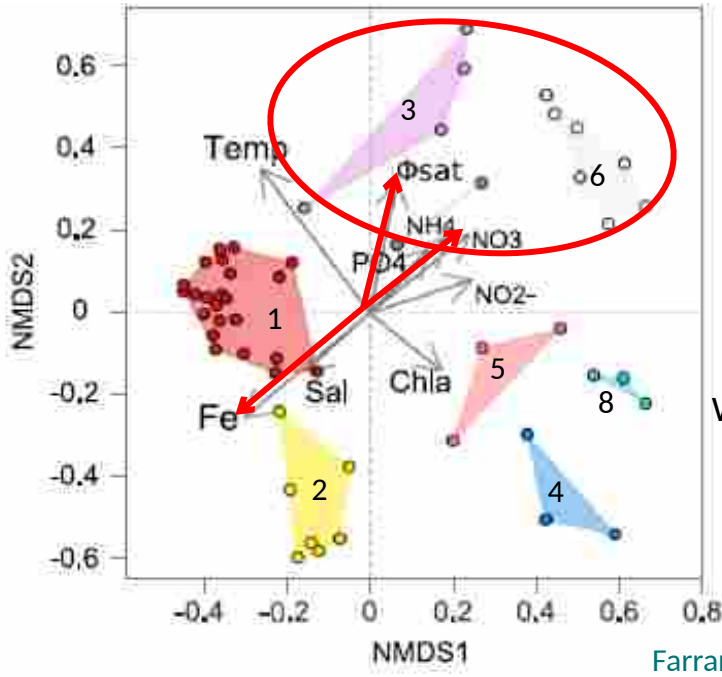
Farrant, Doré *et al.* (2016)

Global niche distribution of *Synechococcus* ESTUs



Clustering (Bray-Curtis distance) of stations based on ESTU Assemblage



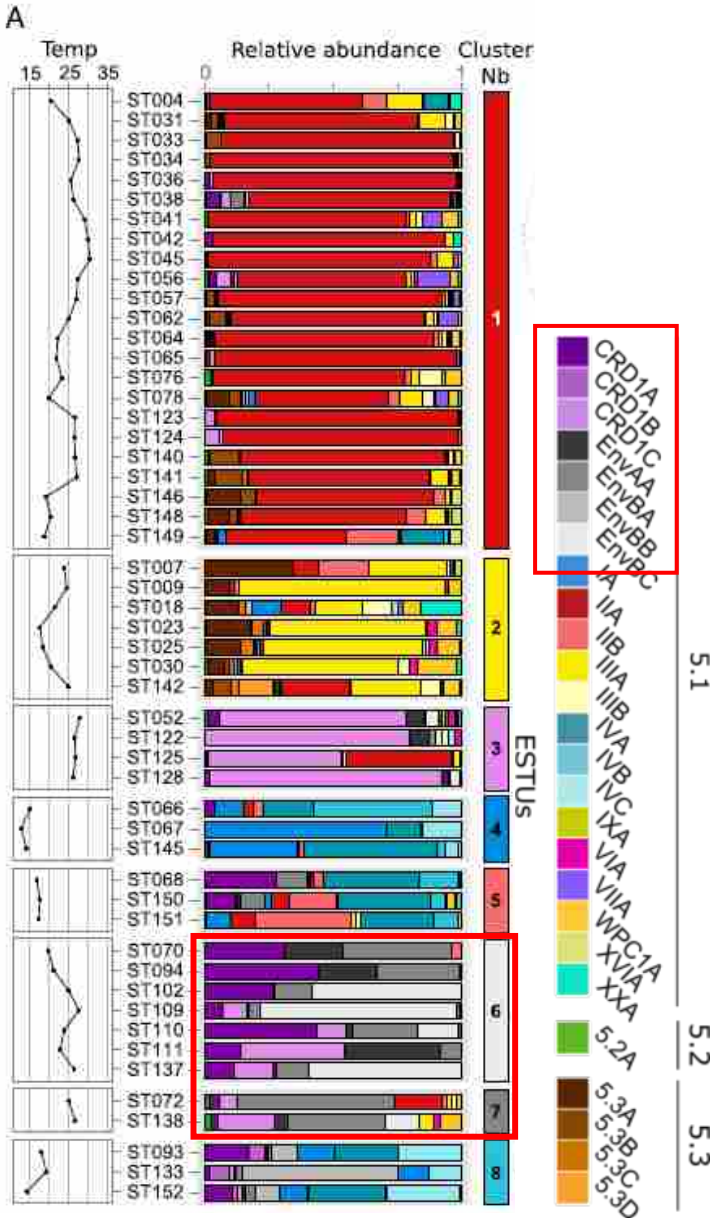


NMDS analysis of stations with physico-chemical parameters

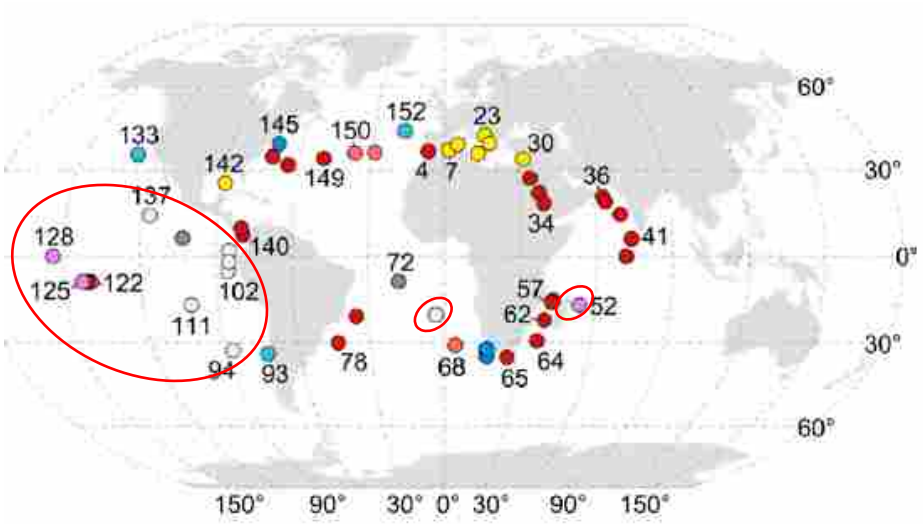
✓ **CRD1A and/or EnvA/EnvB-dominated assemblages** : Warm or temperate N and P-rich Fe-poor waters

Farrant, Doré *et al.* (2016)

Global niche distribution of *Synechococcus* ESTUs



Clustering (Bray-Curtis distance) of stations based on ESTU Assemblage



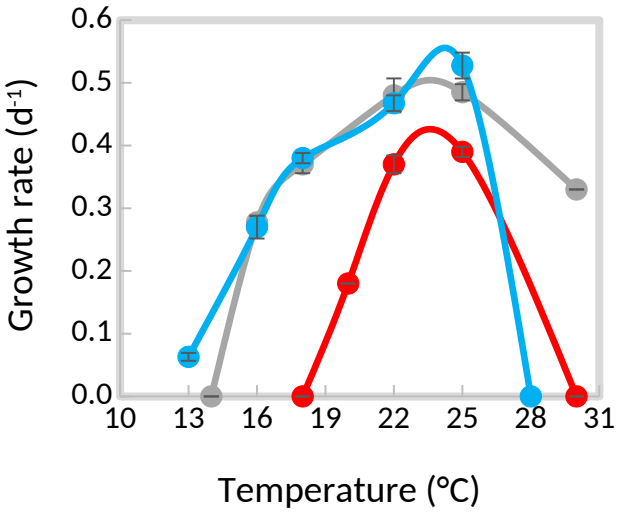
❑**3 distinct CRD1 ESTUs:**

CRD1C/EnvBC: warm HNLC regions

CRD1B/EnvBB: cold upwelling/coastal waters

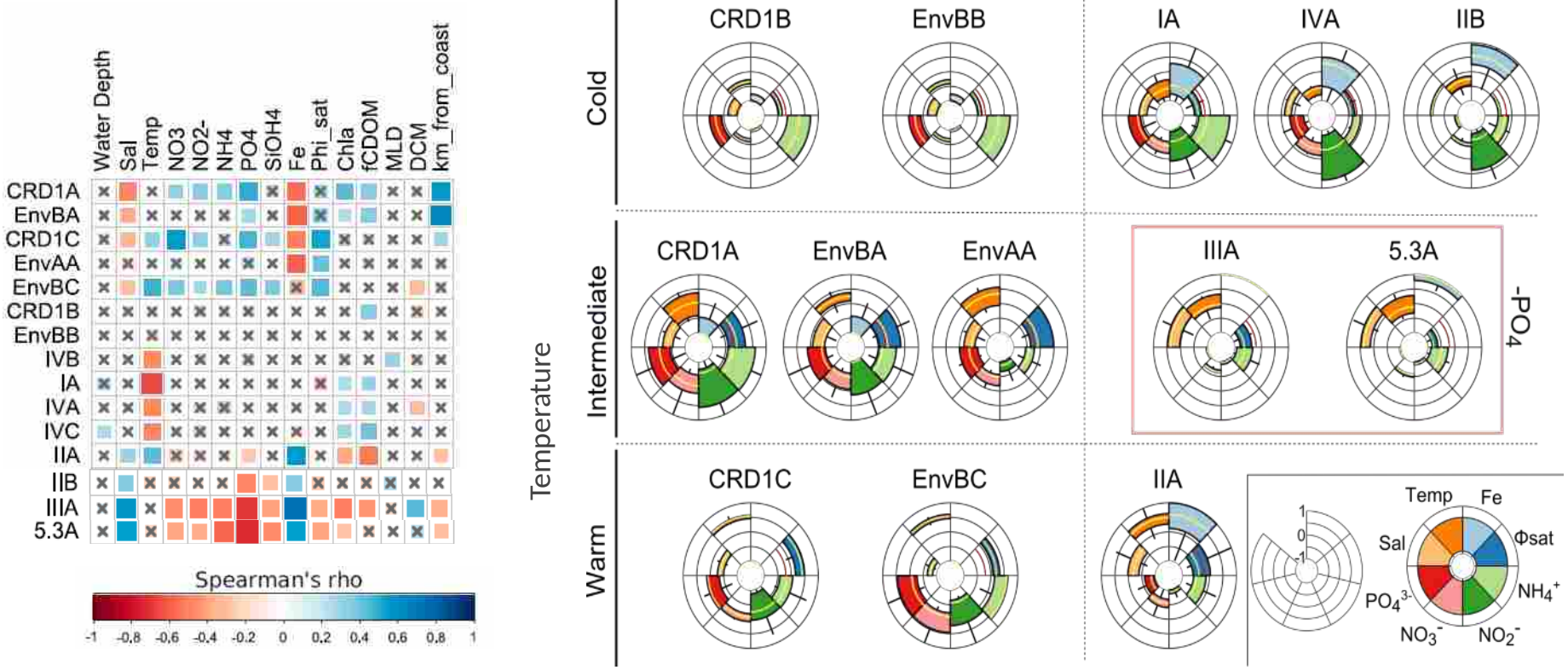
CRD1A/EnvBA: both areas

Farrant, Doré *et al.* (2016)



Ferrieux *et al. Front. Microbiol.* (2022)

Realized environmental niches of *Synechococcus* ESTUs



Although most previous field diversity studies on picocyanobacteria focused on clades, generally considered as distinct “ecotypes”, this study showed that ESTUs appear to provide a better appraisal of “ecotypes” than do clades.

ESTU distribution patterns shows the tight control of *Proc.* and *Syn.* communities by abiotic parameters (mainly, temperature, Fe- and P- availability).

Croucher course 2 outline

I. Genetic diversity and evolution of picocyanobacteria

II. Ecological distribution of picocyanobacteria: are clades equivalent to ecotypes ?

III. Outcomes from comparative genomics

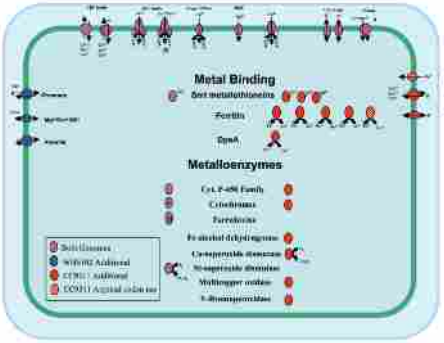
IV. Outcomes from metagenomics

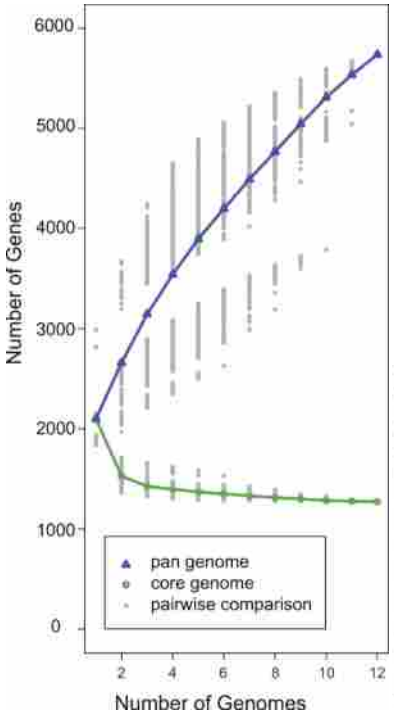
Prochlorococcus and *Synechococcus* : more and more genomes !

PNAS | September 5, 2006 | vol. 103 | no. 36 | 13555–13559

Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment

Brian Palenik*, Qinghu Ren†, Chris L. Dupont*, Garry S. Myers†, John F. Heidelberg†, Jonathan H. Badger†, Ramana Madupu†, William C. Nelson†, Lauren M. Brinkac†, Robert J. Dodson†, A. Scott Durkin†, Sean C. Daugherty†, Stephen A. Sullivan†, Hoda Khouri†, Yasmin Mohamoud†, Rebecca Halpin†, and Ian T. Paulsen†‡





PLOS Genetics

December 2007 | Volume 3 | Issue 12 | e231

Patterns and Implications of Gene Gain and Loss in the Evolution of *Prochlorococcus*

Gregory C Kettler, Adam C Martiny, Katherine Huang, Jeremy Zucker, Maureen L Coleman, Sebastien Rodrigue, Feng Chen, Alla Lapidus, Steven Ferriera, Justin Johnson, Claudia Steglich, George M Church, Paul Richardson, Sallie W Chisholm

Table 1. General Characteristics of the *Prochlorococcus* and *Synechococcus* Isolates Used in This Study

Cyanobacterium	Isolate	Light Adaptation	Length (bp)	GC %	Number of Genes*	Isolation Depth	Region	Date	Accession Number	Reference
Prochlorococcus	MED4	HL(I)	1,657,990	30.8	1,929	5m	Med. Sea	Jan. 1989	BX548174	[7,38]
	MIT9515[b]	HL(I)	1,704,176	30.8	1,908	15m	Eq. Pacific	Jun. 1995	CP000552	[18]
	MIT9301[b]	HL(II)	1,642,773	31.4	1,907	90m	Sargasso Sea	Jul. 1993	CP000576	[18]
	AS9601[b]	HL(II)	1,669,886	31.3	1,926	50m	Arabian Sea	Nov. 1995	CP000551	[21]
	MIT9215[b]	HL(II)	1,738,790	31.1	1,989	5m	Eq. Pacific	Oct. 1992	CP000825	[19]
	MIT9312	HL(II)	1,709,204	31.2	1,962	135m	Gulf Stream	Jul. 1993	CP000111	[4,60]
	NATL1A[b]	LL(I)	1,864,731	35.1	2,201	30m	N. Atlantic	Apr. 1990	CP000553	[20]
	NATL2A[b]	LL(I)	1,842,899	35	2,158	10m	N. Atlantic	Apr. 1990	CP000095	[22]
	SS120	LL(II)	1,751,080	36.4	1,925	120m	Sargasso Sea	May 1988	AE017126	[8,26]
	MIT9211[b]	LL(III)	1,688,963	38	1,855	8.1m	Eq. Pacific	Apr. 1992	CP000878	[19]
	MIT9303[b]	LL(IV)	2,682,807	50.1	3,022	100m	Sargasso Sea	Jul. 1992	CP000554	[4]
	MIT9313	LL(IV)	2,410,873	50.7	2,843	135m	Gulf Stream	Jul. 1992	BX548175	[4,27]
Synechococcus	CC9311	Syn.	2,606,748	52.5	3017	95m	Calif. Current	1993	CP000435	[9]
	CC9902	Syn.	2,234,828	54.2	2504	5m	Calif. Current	1999	CP000097	Palenik, unpublished data
	WH8102	Syn.	2,434,428	59.4	2782		Sargasso Sea	Mar. 1981	BX548020	[2,36]
	CC9605	Syn.	2,510,659	59.2	2991	51m	Calif. Current	1996	CP000110	Palenik, unpublished data

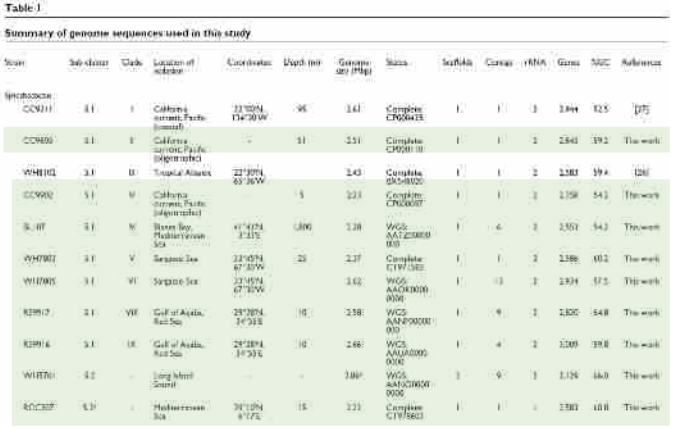
12 *Proc.* & 11 *Syn.* genomes in 2008

As of today >2,000 more *Proc.* & *Syn.* genomes, MAGs and SAGs available in public databases!

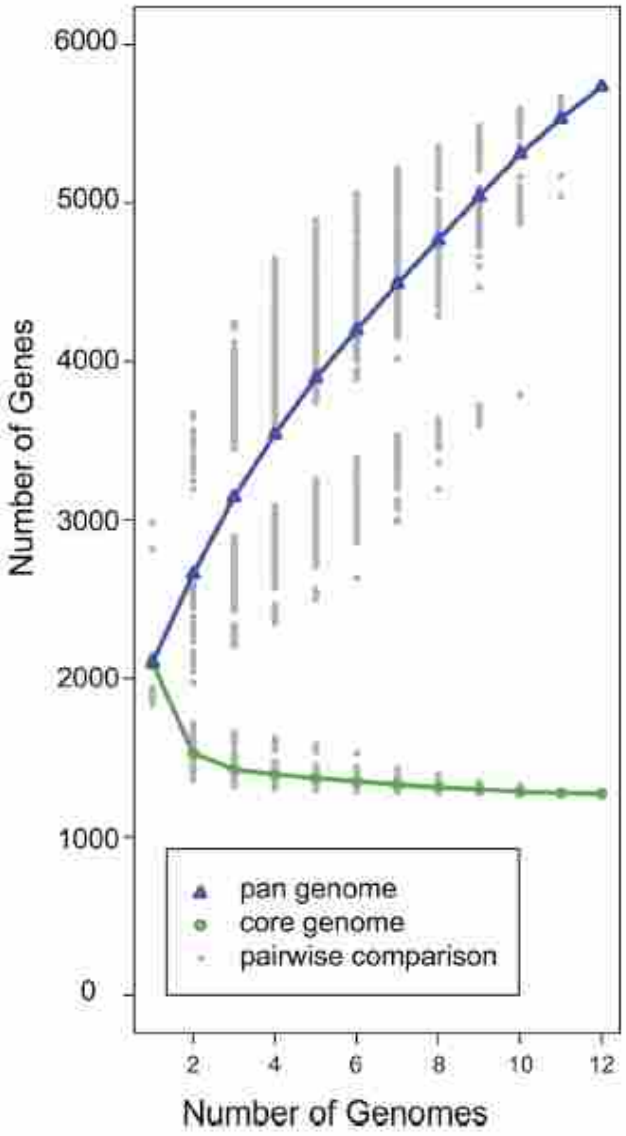
Genome Biology 2008, **9**:R90

Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria

Alexis Dufresne*†, Martin Ostrowski*‡, David J Scanlan§, Laurence Garczarek¨, Sophie Mazard£, Brian P Palenik§, Ian T Paulsen¶, Nicole Tandeau de Marsac¥, Patrick Wincker¤, Carole Dossat¤, Steve Ferriera**, Justin Johnson**, Anton F Post††, Wolfgang R Hess‡‡ and Frédéric Partensky¨



Core, accessory and pan genomes



Core genome: The pool of genes shared by all *Proc.* :
1,273 genes for 12 genomes
Predicted size of the core genome of ca. 1,250 genes

Accessory (or flexible) genome: Number of genes that are present in some but not all strains of *Proc.*:
4,463 genes for 12 genomes
=> ca. 160 new genes for each new genome (5-8%)

Pan-genome: The global gene repertoire of *Proc.*: core genome + accessory genome:
5,736 genes for 12 genomes
Gene accumulation curve far from saturated:
Proc. pan-genome estimate: 57,792 genes

Kettler et al., 2007, *PloS Genetics*

Re-evaluation with 41 *Proc.* genomes:

- Core genome: about 1,000 genes
- Pan-genome: 84,872 genes.

Biller et al. 2015 *Nature Microbiol. Rev.*

Core, accessory and pan genomes

frontiers in Microbiology 2020 | Volume 11 | Article 567431

Evolutionary Mechanisms of Long-Term Genome Diversification Associated With Niche Partitioning in Marine Picocyanobacteria

*Hugo Doré[1], Gregory K. Farrant[1], Ulysse Guyet[1], Julie Haguait[2], Florian Humily[1], Morgane Ratin[1], Frances D. Pitt[3], Martin Ostrowski[3†], Christophe Six[1], Loraine Brillet-Guéguen[4,5], Mark Hoebeke[4], Antoine Bisch[4], Gildas Le Corguillé[4], Erwan Corre[4], Karine Labadie[6], Jean-Marc Aury[6], Patrick Wincker[7], Dong Han Choi[8,9], Jae Hoon Noh[8,10], Damien Eveillard[2,11], David J. Scanlan[3], Frédéric Partensky[1] and Laurence Garczarek[1,11]**

The proportion of accessory genes (including unique genes) is highly variable from genome to genome.

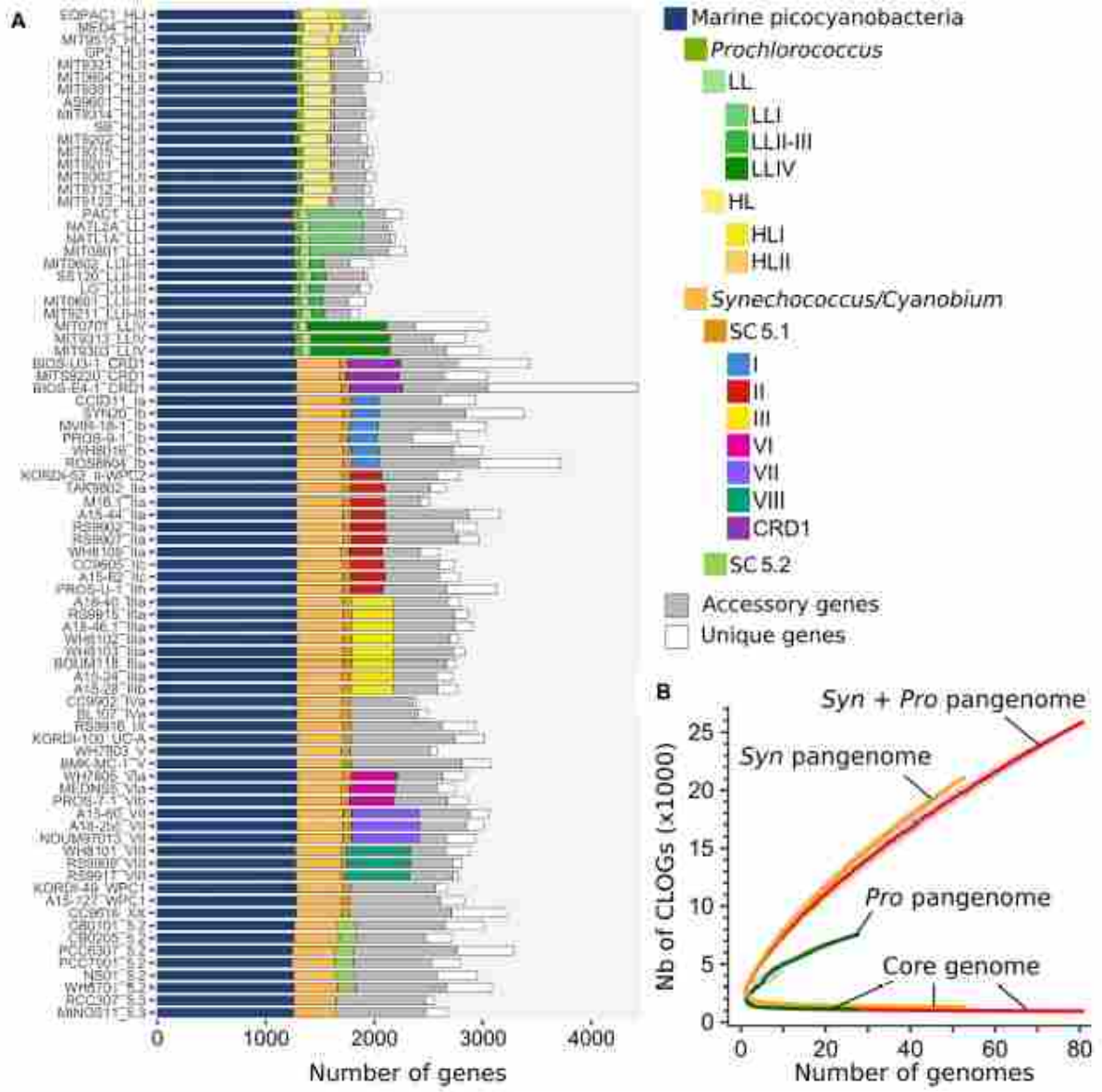
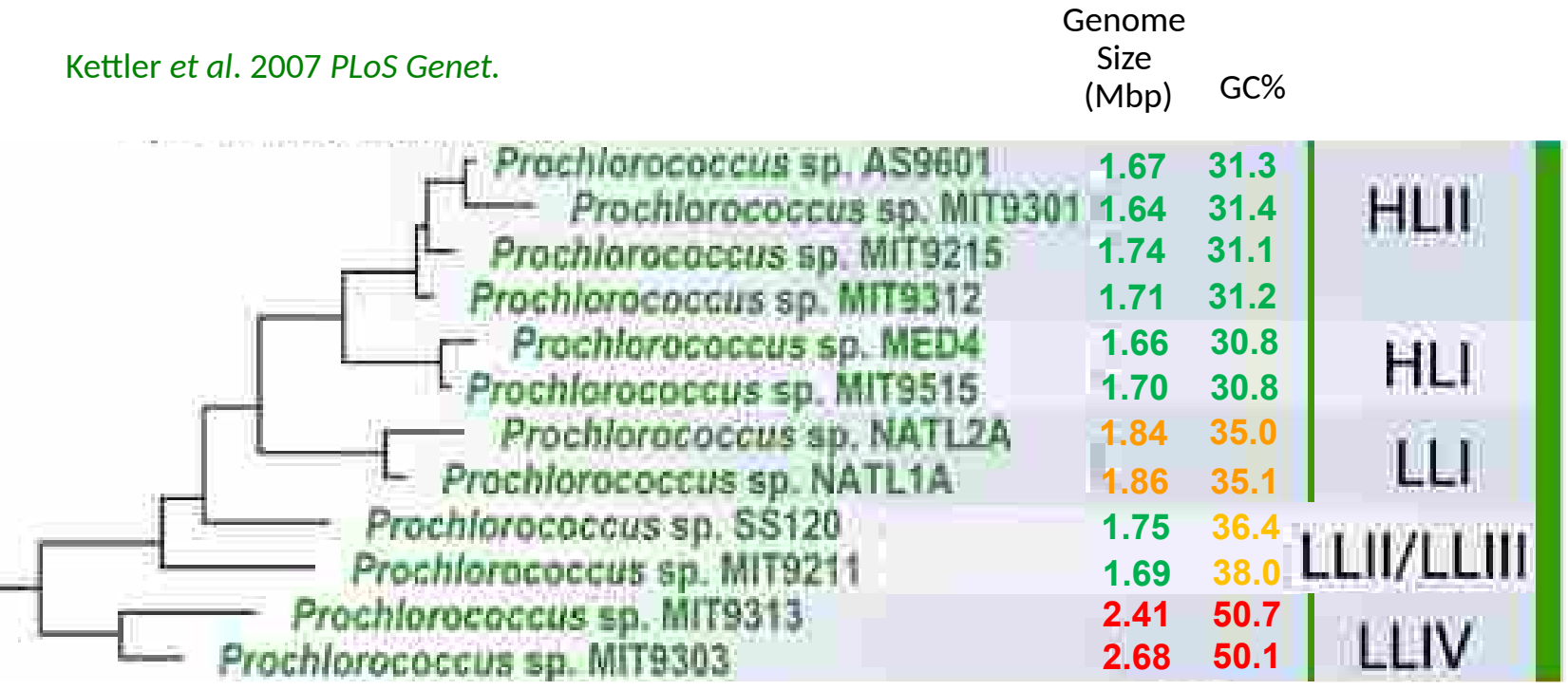


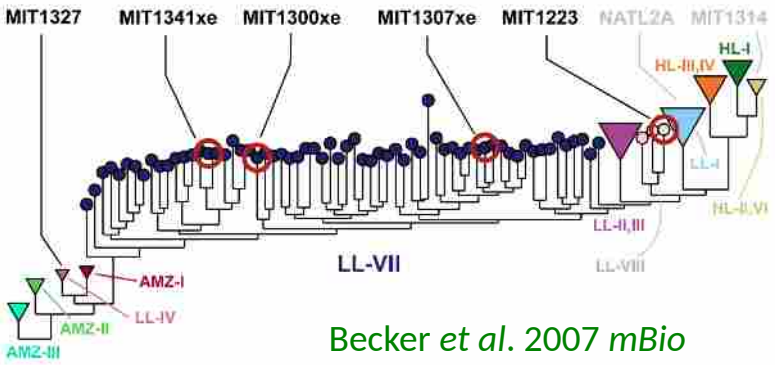
FIGURE 4 | Core, accessory and pan genomes of marine picocyanobacteria

Variability of *Prochlorococcus* Genome Characteristics: Evidence for Genome Streamlining and GC% Drift

Kettler *et al*. 2007 *PLoS Genet.*



Strain	Clade or grade	Assembly size (bp)	% GC
MIT1314	HLII	1,704,447	31.2
MIT1223[a]	LLVIII	1,795,922	35.7
MIT1300xe[a]	LLVII	1,855,146	41.4
MIT1307xe[a]	LLVII	2,032,419	39.9
MIT1341xe[a]	LLVII	1,937,096	40.1
MIT1327	LLIV	2,587,389	50.3



Becker *et al*. 2007 *mBio*

Consequences of Genome Streamlining

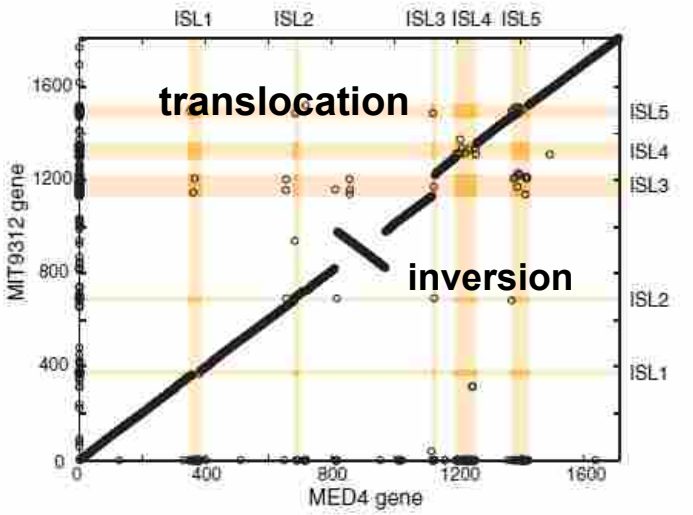
Based on an estimated genome size of 2.5 Mbp for the ancestral *Prochlorococcus* genome and a gene density of 1 gene every 874 bp, one may assess the total gene loss for HL and LLII/III strains as ~930 genes (or ~740 for NATL strains).

However, only ~101 genes are present both in all *Synechococcus* and *Prochlorococcus* LLIV strains but absent in other *Prochlorococcus* (Dufresne et al. 2008 Genome Biol.)

=> Different ecotypes have lost (or retained) different sets of genes…

Role of Genomic Islands in Niche Adaptation

Genomic differences between *Proc.* MED4 (HLI) and MIT9312 (HLII)



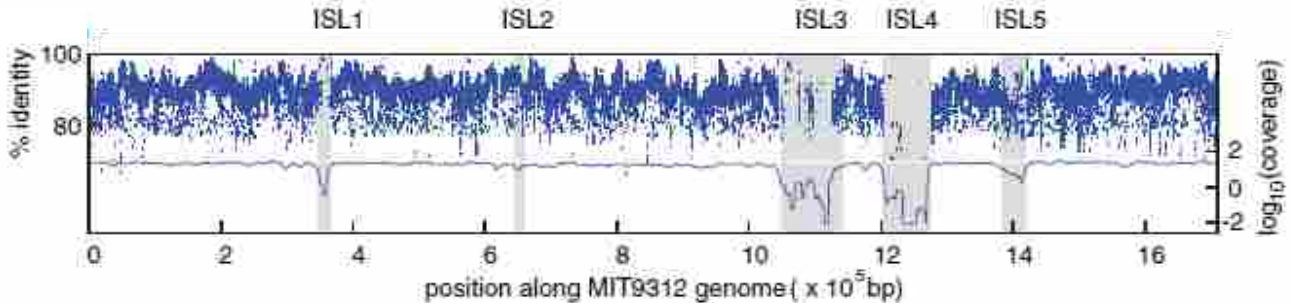
Whole-genome alignment

SCIENCE 24 MARCH 2006 VOL 311

Genomic Islands and the Ecology and Evolution of *Prochlorococcus*

Maureen L. Coleman,[1] Matthew B. Sullivan,[1] Adam C. Martiny,[1] Claudia Steglich,[1*] Kerrie Barry,[2] Edward F. DeLong,[1] Sallie W. Chisholm[1†]

Recruitment of Global Ocean Survey (GOS) reads against the MIT9312 genome



- Presence of five genomic islands containing accessory and unique genes :

  => Privileged areas for laterally transferred genes.

- Genomic island genes comprise about 25% of all genes of a genome.
- Genomic islands harbor > 2/3rd of all accessory genes of the *Prochlorococcus* pangenome.

Hackl et al., 2023, *Cell*

Comparative genomics of picocyanobacteria

frontiers in Microbiology September 2020 | Volume 11 | Article 567431

Evolutionary Mechanisms of Long-Term Genome Diversification Associated With Niche Partitioning in Marine Picocyanobacteria

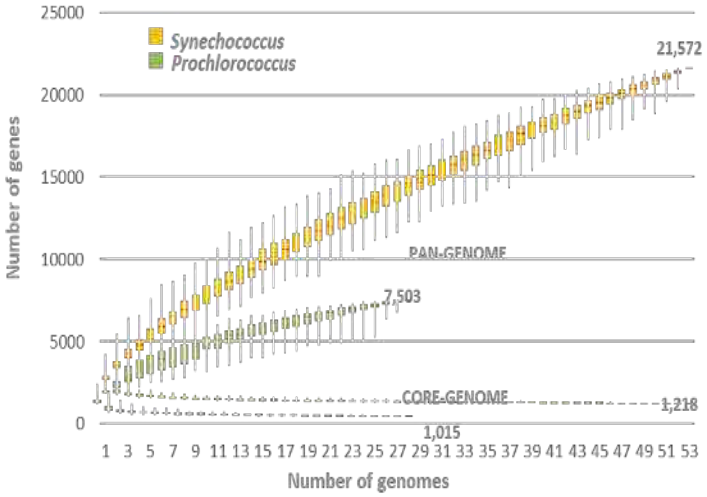
*Hugo Doré[1], Gregory K. Farrant[1], Ulysse Guyet[1], Julie Haguait[2], Florian Humily[1], Morgane Ratin[1], Frances D. Pitt[3], Martin Ostrowski[3†], Christophe Six[1], Loraine Brillet-Guéguen[4,5], Mark Hoebeke[4], Antoine Bisch[4], Gildas Le Corguillé[4], Erwan Corre[4], Karine Labadie[6], Jean-Marc Aury[6], Patrick Wincker[7], Dong Han Choi[8,9], Jae Hoon Noh[8,10], Damien Eveillard[2,11], David J. Scanlan[3], Frédéric Partensky[1] and Laurence Garczarek[1,11]**

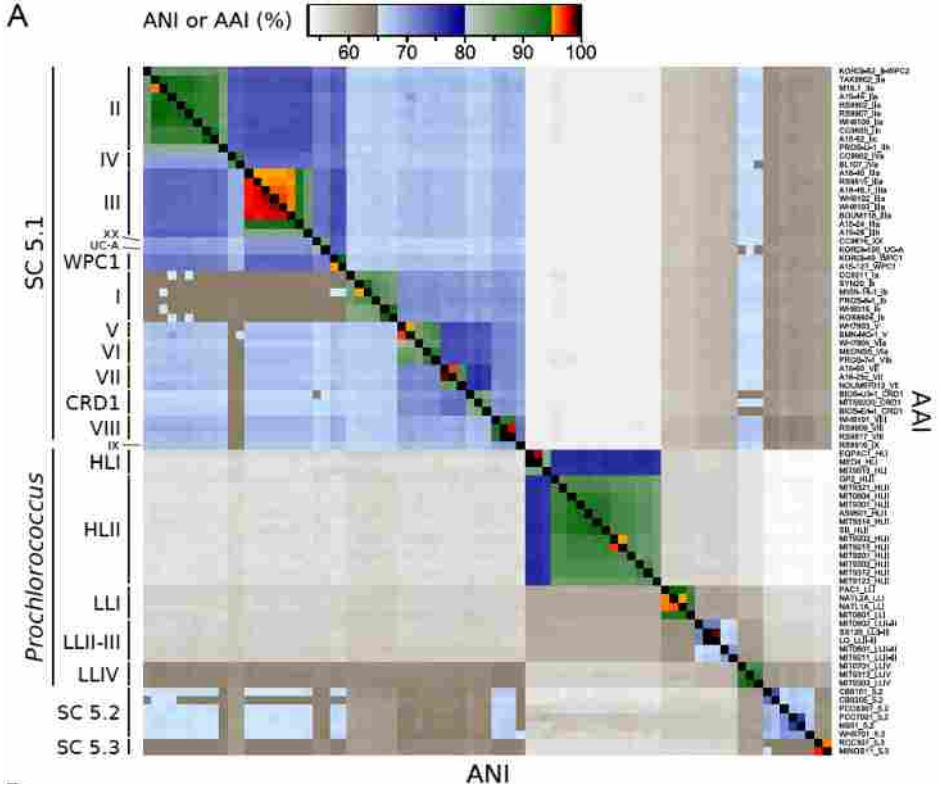




H. Doré G. Farrant U. Guyet

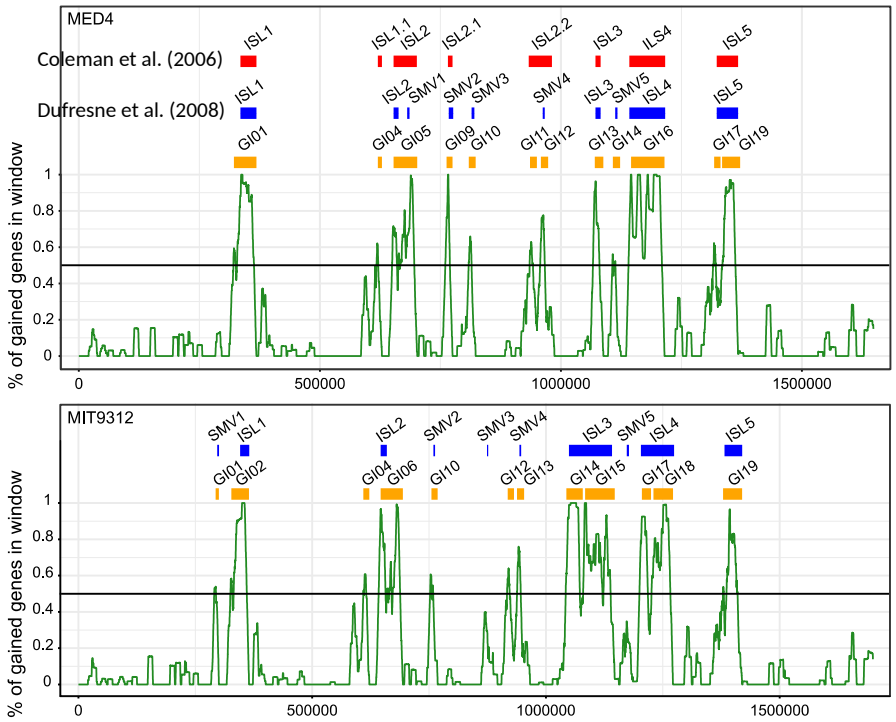
- 81 high-quality, non redundant genomes
  - 28 *Prochlorococcus*
  - 53 *Synechococcus/Cyanobium*





➡ High genomic diversity in terms of sequence identity e.g. Clade II (10 strains): most of them share less than 94 % ID

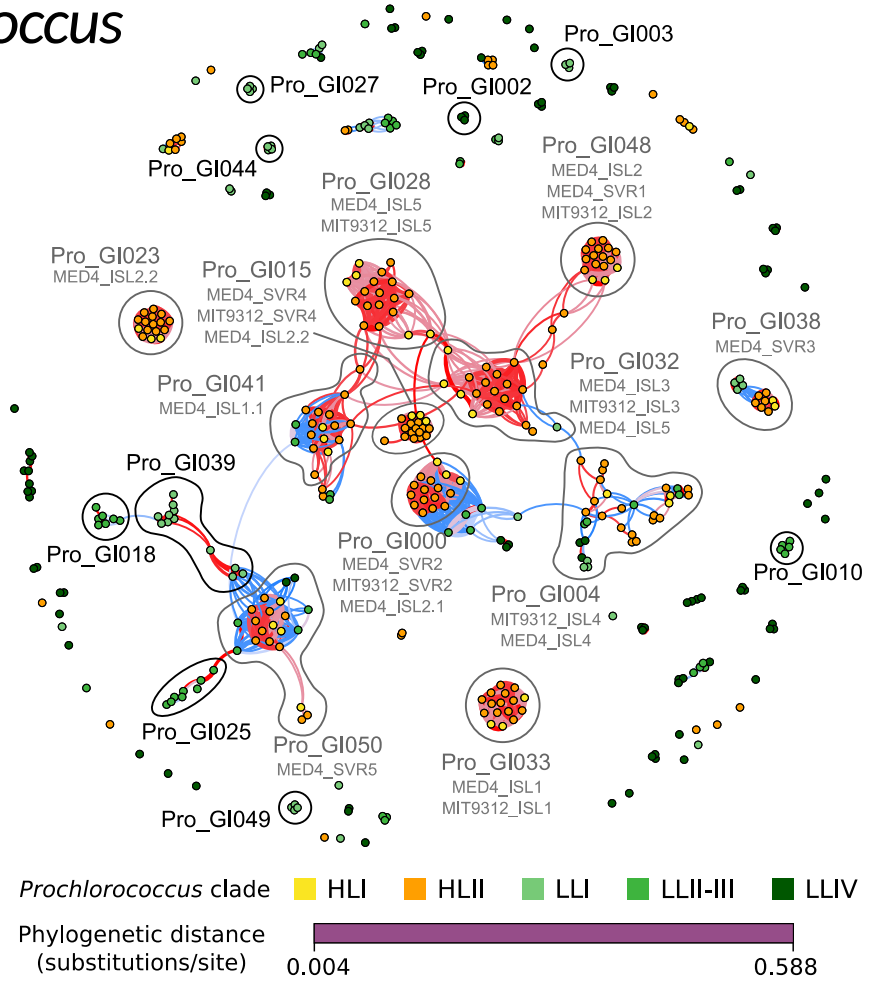
Genomic islands network in *Prochlorococcus*



Green line: % of gained genes in 10 kb windows w 100 bp step

Black line: 50% cut-off applied to delineate genomic islands

Genomic islands can be determined automatically using sliding windows (size 10 kbp, interval 100 bp) as regions where the ratio of nt from gained CDS to total coding nt > 50%.



Network based on the partial similarity of gene contents between islands shared by pairs of genomes (edge when 2 islands shared at least 10% of their gene content

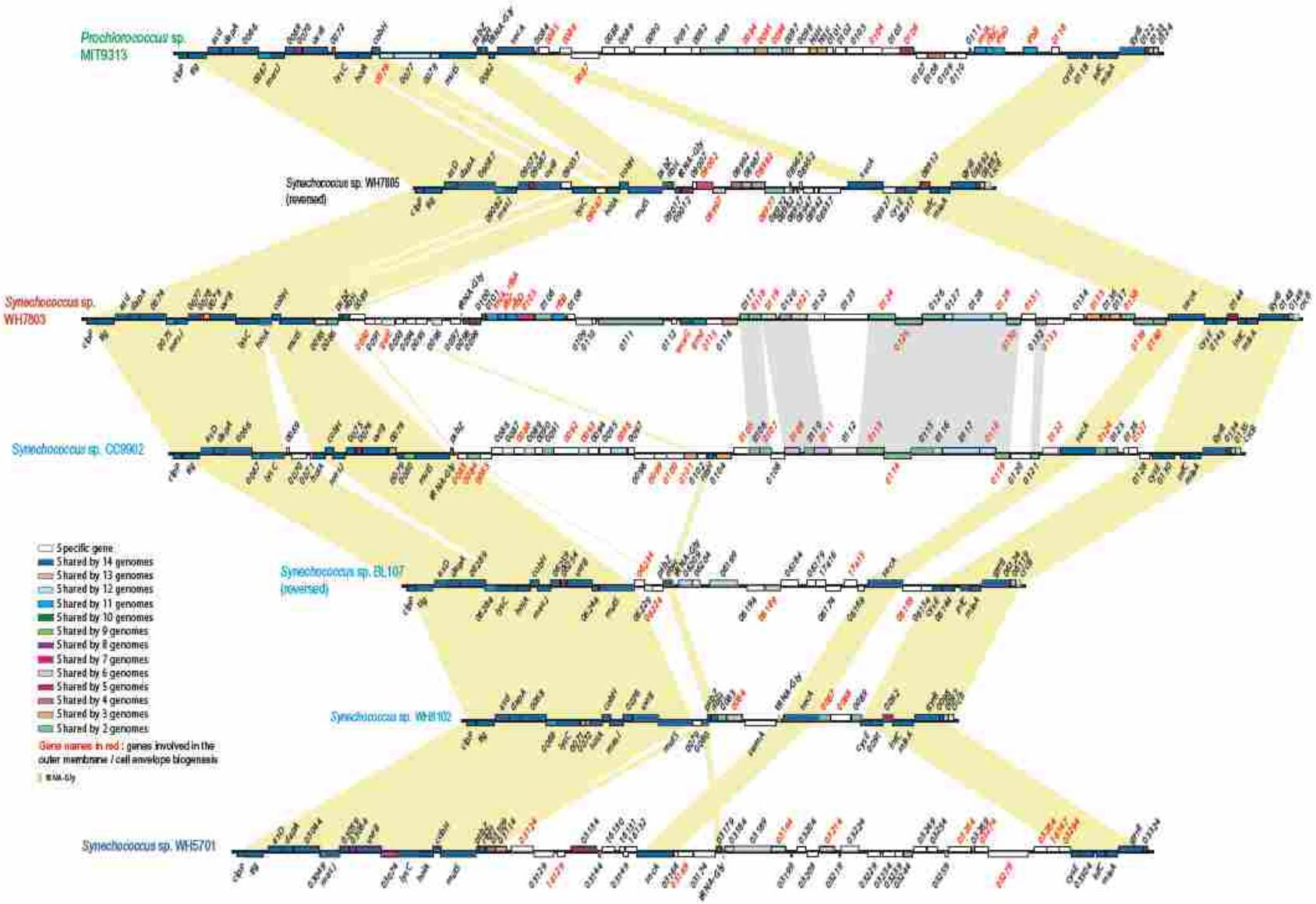
Node = one genomic island in a given strain

Red edges = short phylogenetic distance

Blue edges = long phylogenetic distance (≠ clades)

Doré et al. (2020) Front Microbiol

‘Shared’ gene islands between two genomes share only a subset of genes



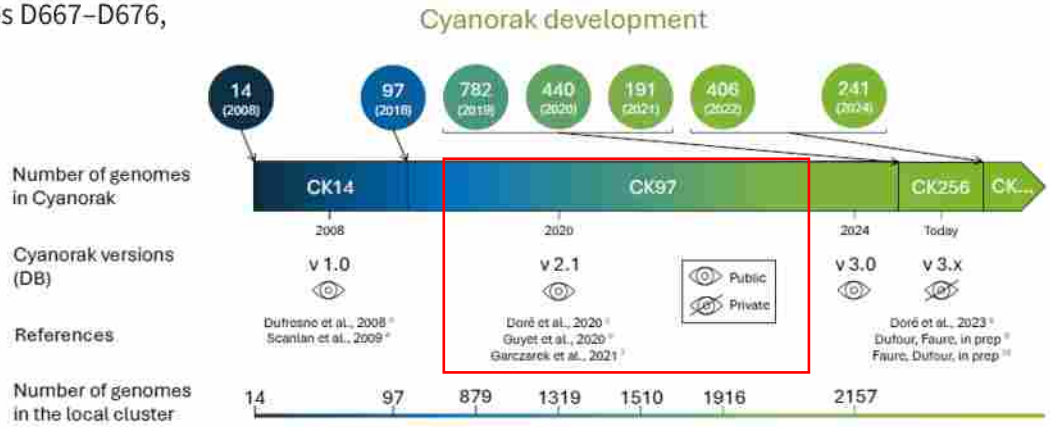
Dufresne et al., *Genome Biology*, 2008

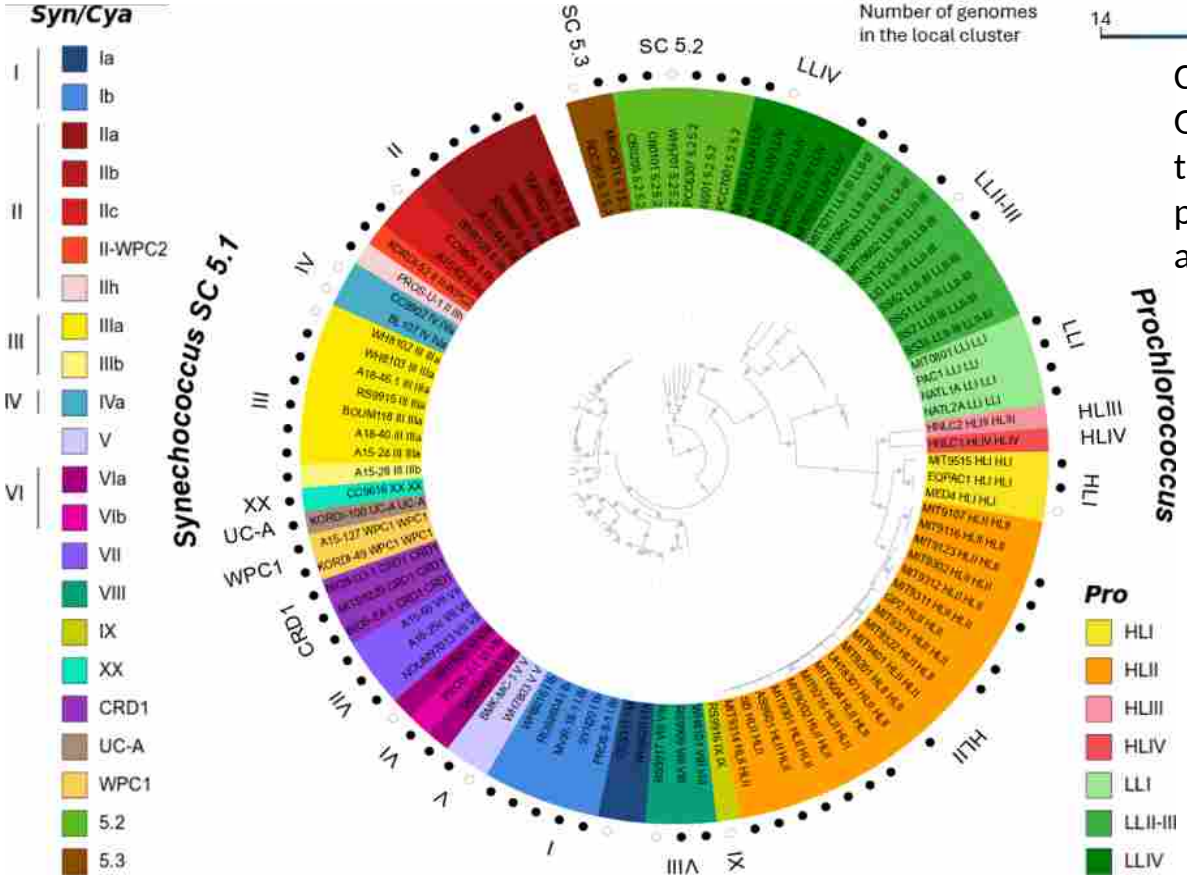
A dedicated database for comparative genomics of picocyanobacteria

Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021, Pages D667–D676,

Cyanorak v2.1: a scalable information system dedicated to the visualization and expert curation of marine and brackish picocyanobacteria genomes

Laurence Garczarek ✉ , Ulysse Guyet , Hugo Doré , Gregory K Farrant , Mark Hoebeke , Loraine Brillet-Guéguen , Antoine Bisch , Mathilde Ferrieux , Jukka Siltanen , Erwan Corre , Gildas Le Corguillé , Morgane Ratin , Frances D Pitt , Martin Ostrowski , Maël Conan , Anne Siegel , Karine Labadie , Jean-Marc Aury , Patrick Wincker , David J Scanlan , Frédéric Partensky





Cyanorak is a database of Cluster of Likely Orthologous Genes (CLOGS) developed by the ECOMAP team and local bioinformatics platform ABIMS since 2008. CLOG numbers are maintained over the time.

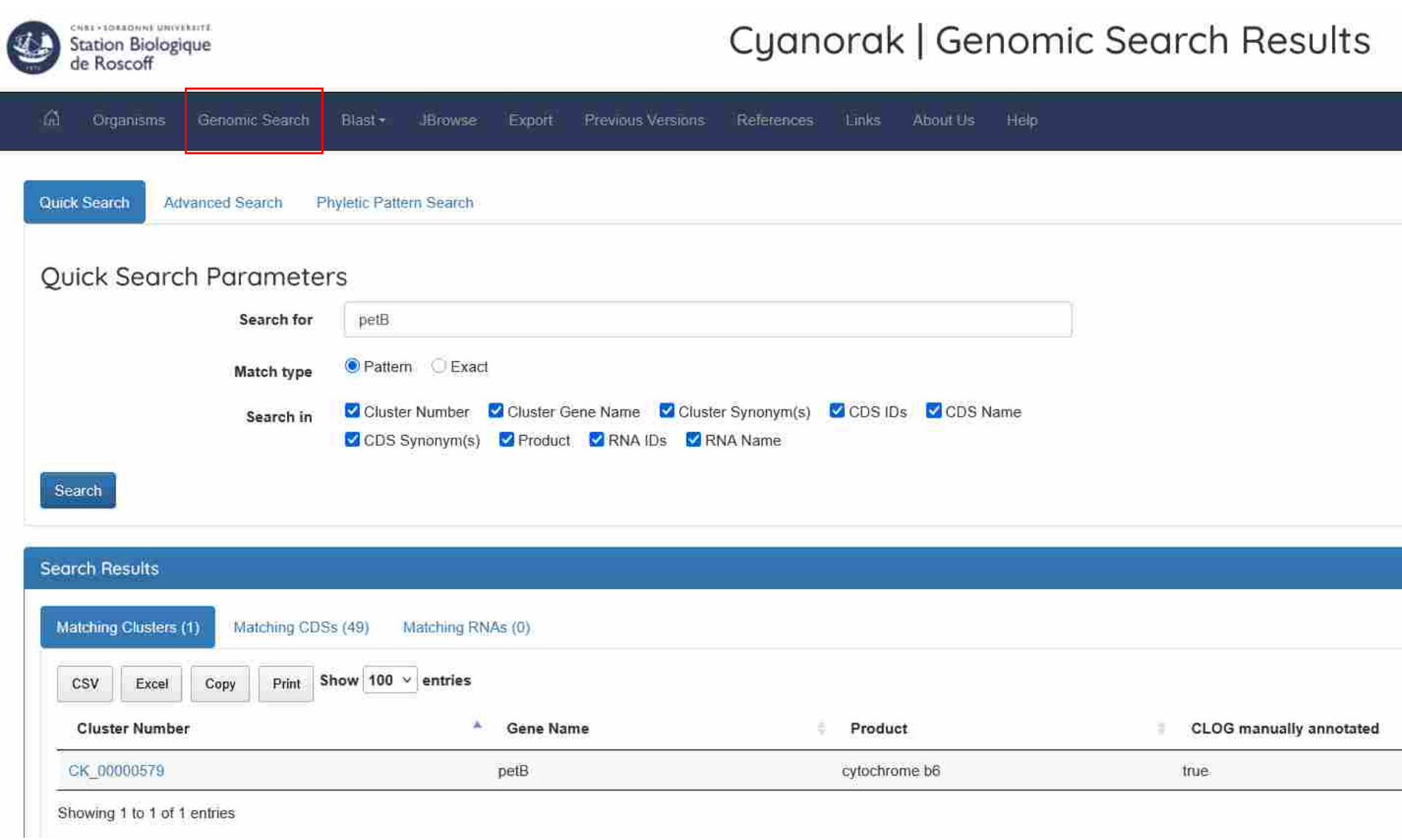


Version 3.0 (April 2024) is accessible here:

https://cyanorak.sb-roscoff.fr/ck_97/

Cyanorak is used for comparative genomics and as a reference database for annotating both new genomes and metagenomes (detailed demonstration upon request!)

Cyanorak Search Page



Cyanorak CLOG Page (top)

CLOG CK_00000579

Genome Context

Identifiers

Cyanorak V1 Cluster Number	579

Annotations

Gene Name	petB
Synonym(s)	
Product	cytochrome b6

Functional Categories

EggNOG	COG / NOG	COG1290 Cytochrome b subunit of the bc complex
	bactNOG	bactNOG05294 Component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome [...]
	cyaNOG	cyaNOG00104 Component of the cytochrome b6-f complex, which mediates electron transfer between [...]
		cyaNOG06334 Component of the cytochrome b6-f complex, which mediates electron transfer between [...]
	CyOG	CyOG00284
KEGG	EC Number	7.1.1.6 Plastoquinol--plastocyanin reductase
	K Number	K02635 petB; cytochrome b6
	TIGR Role	112 Energy metabolism / Electron transport
		164 Energy metabolism / Photosynthesis
	Cyanorak Role	J.3 Cytochrome b6/f complex
GO Terms	Biological Process	0009767 photosynthetic electron transport chain
		0022900 electron transport chain
	Cell. Component	0009512 cytochrome b6f complex
	Molecular Function	0045158 electron transporter, transferring electrons within cytochrome b6f complex of photosystem II activity

Protein Domains and Functional Sites

	TIGRFAM	
	PFAM	PF13631 Cytochrome b(N-terminal)/b6/petB
ProSite	Pattern	
	Profile	
	InterPro	IPR005797 Cytochrome b/b6, N-terminal

Cyanorak CLOG Page (bottom)

Phyletic pattern : Pattern of presence/absence of the gene in the different strains

Phyletic Pattern

Prochlorococcus																																
HL																																
HLI			HLII																		HLIII	HLIV	LLI				L			
EQPAC1	MED4	MIT9515	AS9601	GP2	MIT0604	MIT9107	MIT9116	MIT9123	MIT9201	MIT9202	MIT9215	MIT9301	MIT9302	MIT9311	MIT9312	MIT9314	MIT9321	MIT9322	MIT9401	SB	UH18301	HNLC2	HNLC1	MIT0801	NATL1A	NATL2A	PAC1	LG	MIT0602	MIT0603	SS120
Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Lba	Hba	Hba	Hba	Hba	Hba	Hba	Hba	Hba
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Synechococcus/Cyanobium																																									
Subcluster 5.1																																									
I							II										III								IV		V		VI			VII			VIII			IX	CRD1		
Ia		Ib					IIa					IIb	IIc		IIh	II-WPC2	IIIa							IIIb	IVa				VIa		VIb										
CC9311	WH8020	MVIR-18-1	PROS-9-1	ROS8604	SYN20	WH8016	A15-44	M16.1	RS9902	RS9907	TAK9802	WH8109	A15-62	CC9605	PROS-U-1	KORDI-52	A15-24	A18-40	A18-46.1	BOUM118	RS9915	WH8102	WH8103	A15-28	BL107	CC9902	BMK-MC-1	WH7803	MEDNS5	WH7805	PROS-7-1	A15-60	A18-25c	NOUM97013	RS9909	RS9917	WH8101	RS9916	BIOS-E4-1	BIOS-U3-1	MITS9220
3dA	3dA	3aA	3dA	3a	3a	3aA	2	3a	3c	3a	3a	3bB	3dB	3c	3dB	3bB	3c	3dB	3c	3c	3dB	3c	3dB	3c	3dA	3dA	2	3a	3c	2	2	3c	3c	3a	1	1	1	3dA	3cA	3dA	3dA
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

CDS (97)

CSV Excel Copy Print

Cyanorak ORF ID	Initial gene name	Initial product annotation
CK_Cya_CB0101_02683		cytochrome b6
CK_Cya_CB0205_04963		cytochrome b6
CK_Cya_NS01_02463	petB	cytochrome b6
CK_Cya_PCC6307_01601		cytochrome b subunit of the bc complex
CK_Cya_PCC7001_00188	qcrB	menaquinol-Cytochrome c reductase cytochrome b subunit
CK_Cya_WH5701_06120	WH5701_06120	cytochrome b6
CK_Pro_AS9601_03491	petB	cytochrome b6
CK_Pro_EQPAC1_01873		Cytochrome b6-f complex subunit

Export

Genes / Proteins

FASTA format : All genes (.fna) / Proteins (.faa)

Cyanorak Genomic Context

CLOG CK_00000579

Genome Context

Relative gene sizes : ON

Genomic context : 4 genes before and after the selected gene



Cyanorak : Organisms Page

Prochlorococcus | Synechococcus/Cyanobium

Show 100 entries

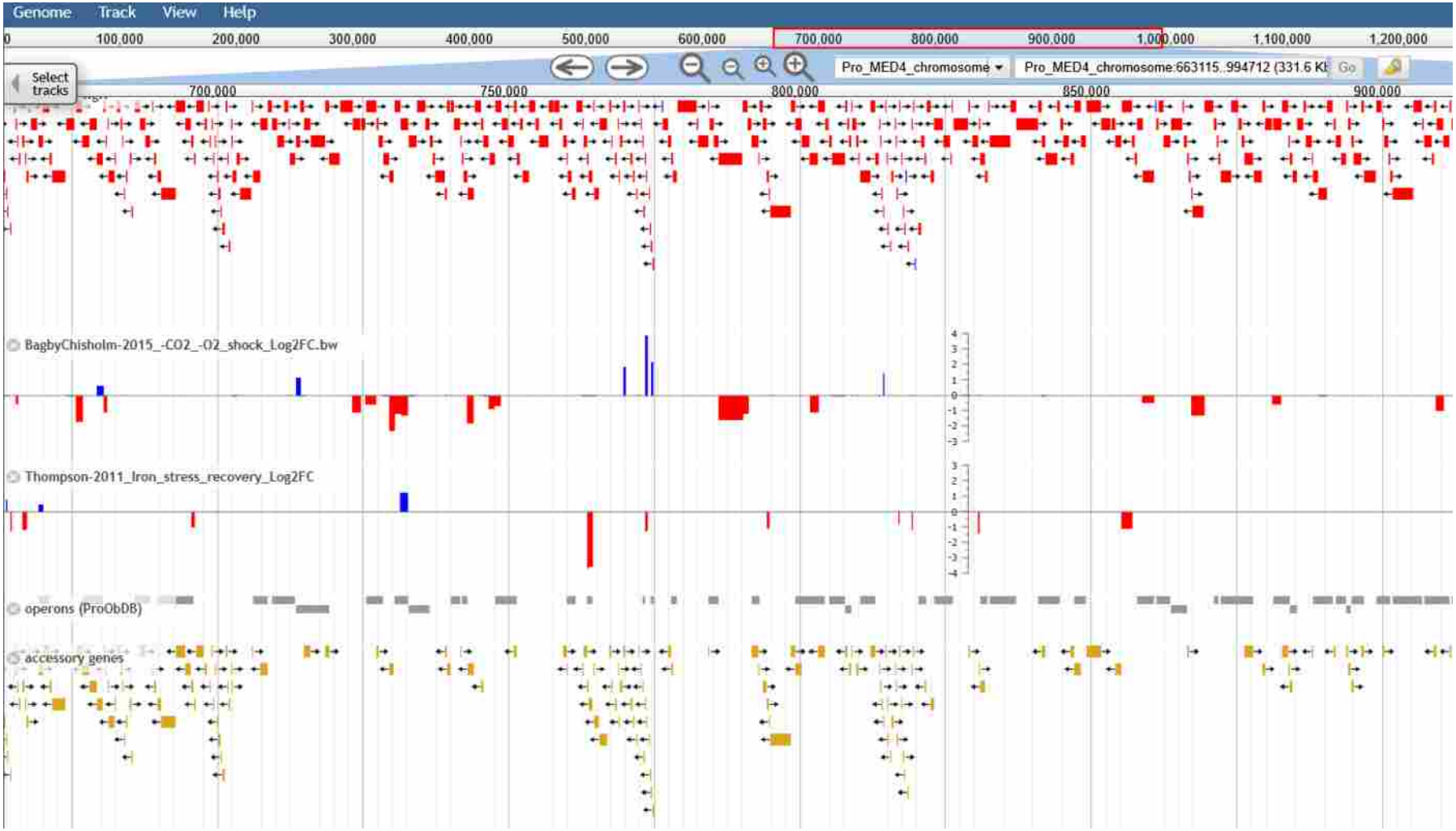
JBrowse	Name	Species	SubCluster/ Phototype	Clade	SubClade	Ecotype/ESTU	Pigment type	Sequencing center
J	Pro_AS9601	Prochlorococcus marinus	HL	HLII	HLII	HLIIA	Lba	JCVI
J	Pro_EQPAC1	Prochlorococcus marinus	HL	HLI	HLI	HLIA	Lba	MIT BioMicro Center
J	Pro_GP2	Prochlorococcus marinus	HL	HLII	HLII	HLIID	Lba	MIT BioMicro Center
J	Pro_HNLC1	Prochlorococcus marinus	HL	HLIV	HLIV	HLIVA	Lba	JCVI
J	Pro_HNLC2	Prochlorococcus marinus	HL	HLIII	HLIII	HLIIIA	Lba	JCVI
J	Pro_LG	Prochlorococcus marinus	LL	LLII	LLII	LLII-IIIA	Hba	MIT BioMicro Center
J	Pro_MED4	Prochlorococcus marinus	HL	HLI	HLI	HLIA	Lba	JGI
J	Pro_MIT0601	Prochlorococcus marinus	LL	LLIII	LLIII	LLII-IIIA	Hba	MIT BioMicro Center
J	Pro_MIT0602	Prochlorococcus marinus	LL	LLII	LLII	LLII-IIIA	Hba	MIT BioMicro Center



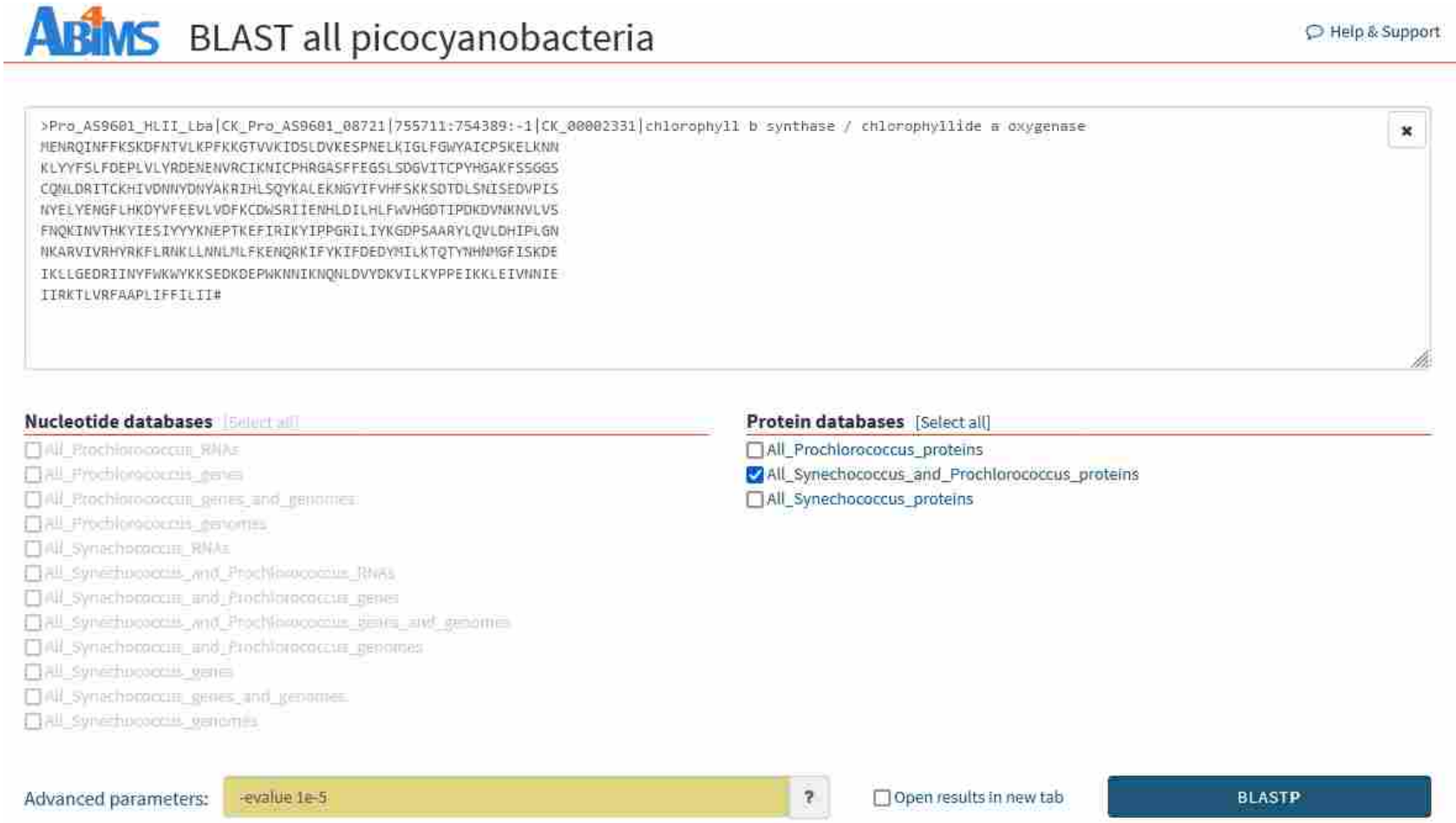
Click on genome name to access individual genome description

Click on J to access the Jbrowse viewer

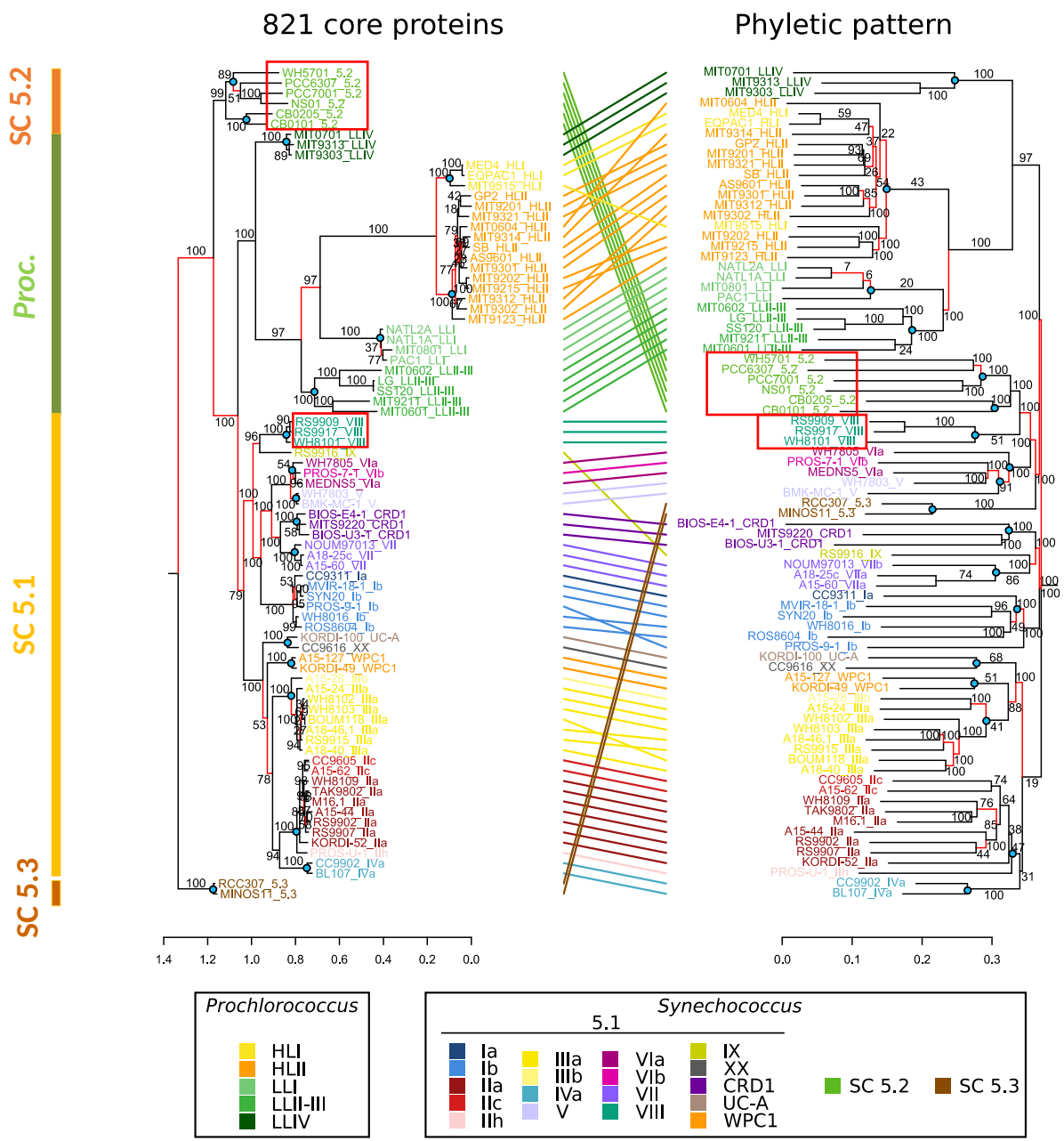
Cyanorak : Jbrowse Genome viewer



Cyanorak : Blast Tools (all and selection)



Comparing genetic diversity and gene content



- Similar topology
  - -> Gene content and protein sequences evolved concomitantly
  - -> Exception: brackish clade VIII strains phylogenetically distant from SC 5.2 but their gene content is closer than it is from other SC 5.1.
- Relative size of internal and terminal branches
  - -> More variability in sequence between clades
  - -> More variability in gene content between strains

Croucher course 2 outline

I. Genetic diversity and evolution of picocyanobacteria

II. Ecological distribution of picocyanobacteria: are clades equivalent to ecotypes ?

III. Comparative genomics

IV. Outcome from metagenomics

Genomic adaptation to nutrient depletion: Phosphorus

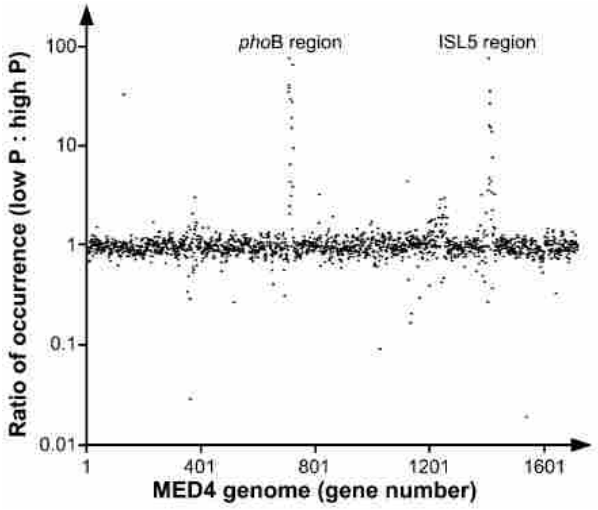
Environmental Microbiology (2009)

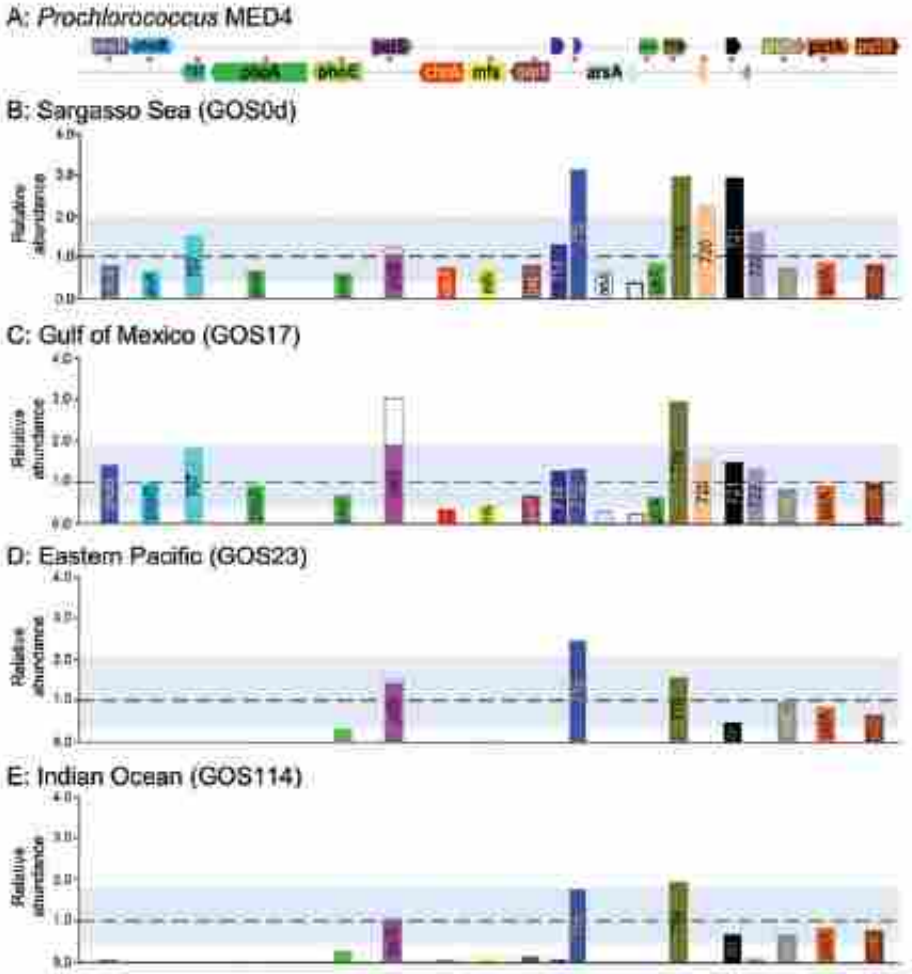
Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions

Adam C. Martiny,[1*] Ying Huang[2] and Weizhong Li[2]

Proc. populations thriving in P-poor areas (Sargasso Sea and Caribbean Sea) are enriched in genes allowing P transport and assimilation.

These genes are located in two genomic islands: the *phoB* region and MED4-ISL5, whose genes respond to P limitation.

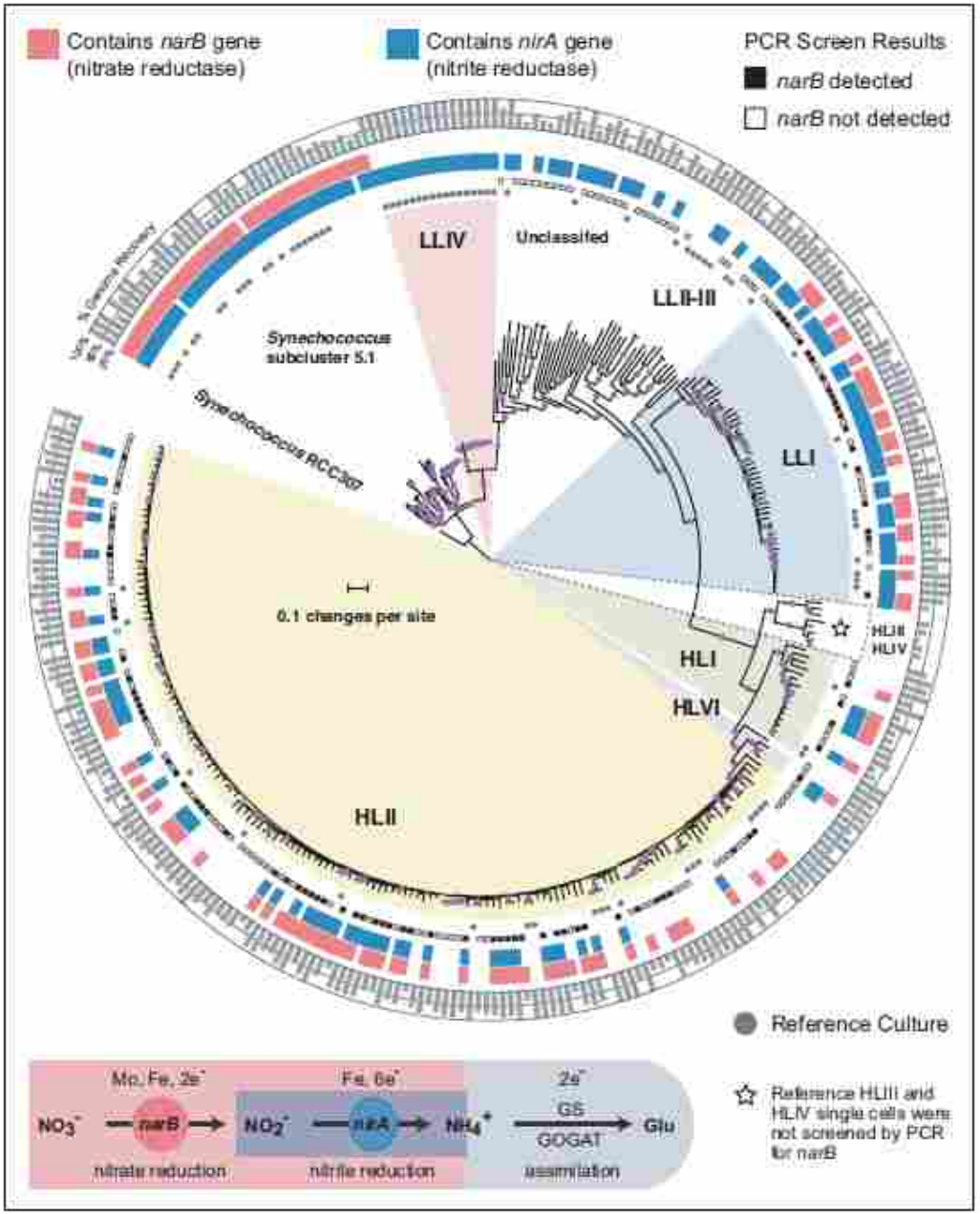




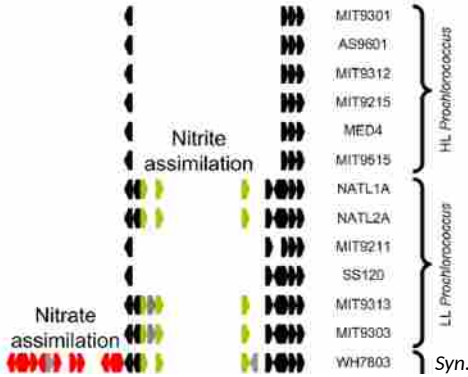


GOS = Global ocean survey

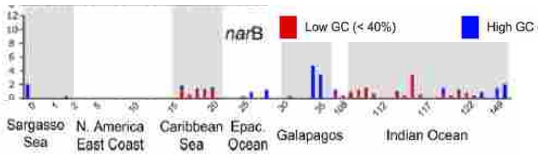
Genomic adaptation to nutrient depletion: Nitrogen



Proc. long thought to be unable to assimilate nitrates since none of the 12 first sequenced genomes had *narB*.



First *Proc. narB* sequences from the field retrieved from GOS metagenomes.

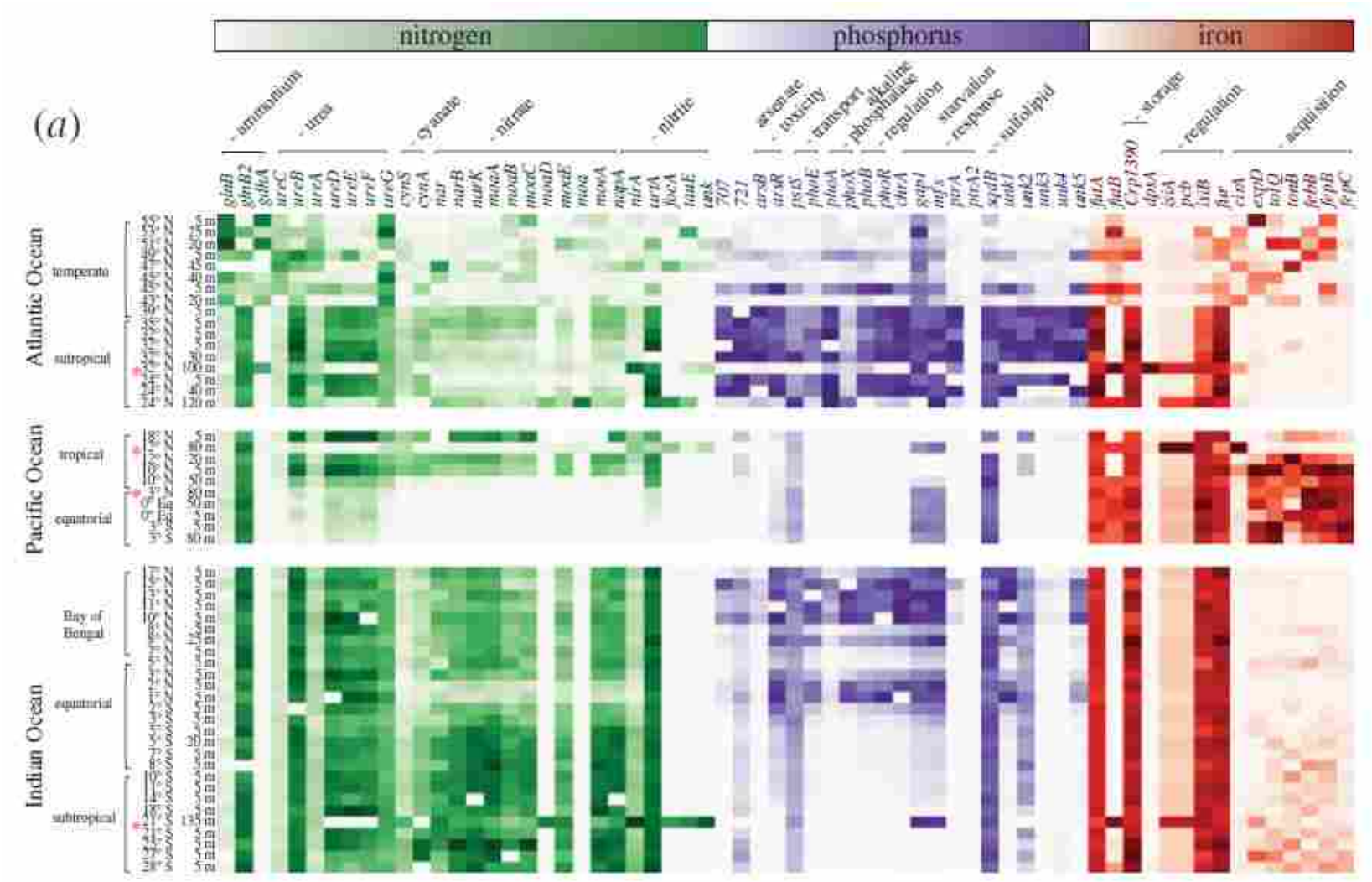


Martiny *et al.* 2009, *PNAS*

Analysis of hundred of *Proc.* genomes showed that while genes for nitrite assimilation are distributed across the tree, the full pathway of nitrate assimilation is restricted to particular clades or low-N habitats.

Berube et al. 2019, *eLife*

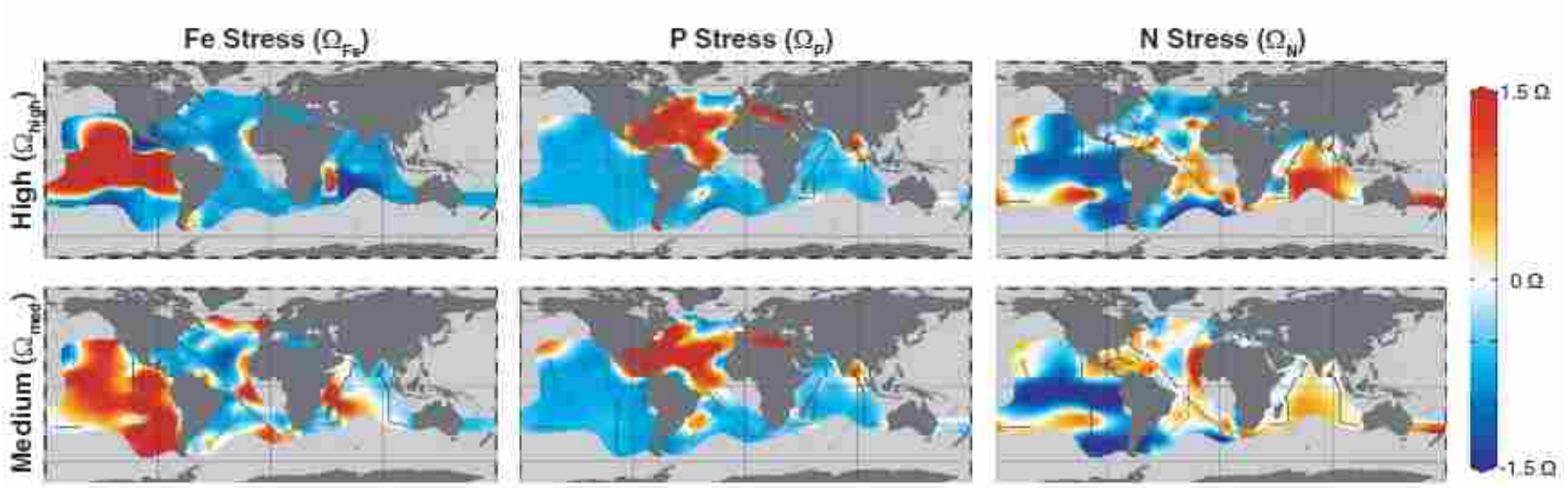
The distribution of *Prochlorococcus* accessory genes involved in nutrient metabolism can be used to determine which nutrient is limiting



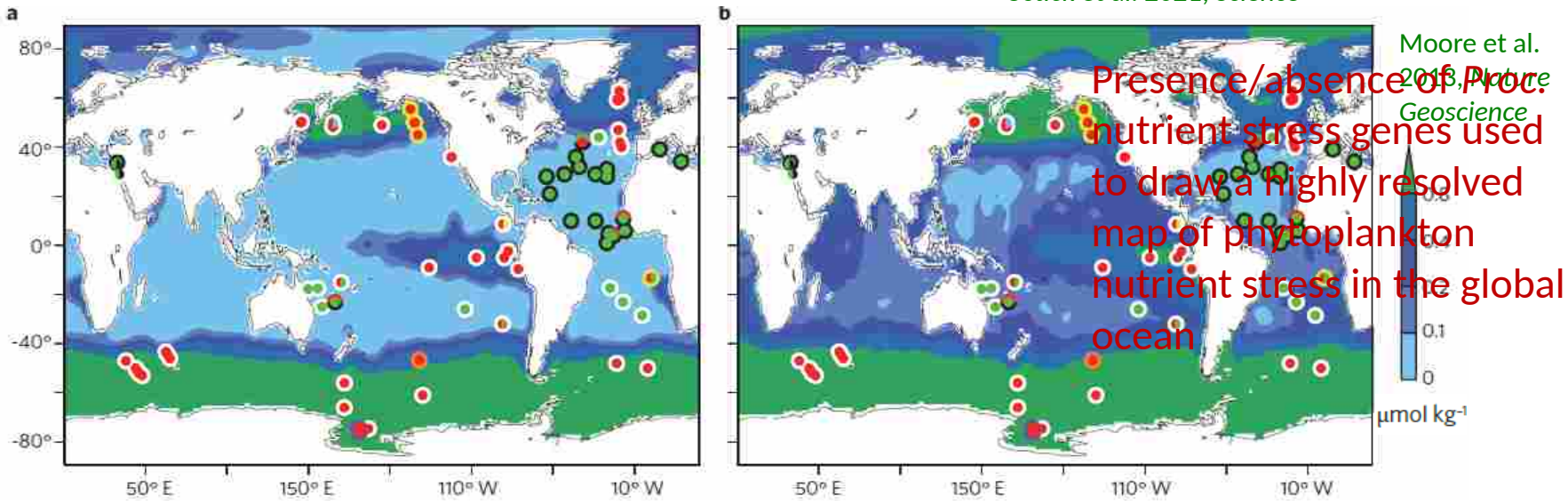
In many ecosystems, it is challenging to quantify microbial stress. A genomic approach can be used for understanding microbial responses to environmental changes.

Garcia et al. 2020, *Phil Trans Roy Soc B*

The distribution of *Prochlorococcus* accessory genes involved in nutrient metabolism can be used to map nutrient stress type and severity



Ustick et al. 2021, *Science*



Moore et al. 2013, *Nature Geoscience*

Presence/absence of Proc. nutrient stress genes used to draw a highly resolved map of phytoplankton nutrient stress in the global ocean

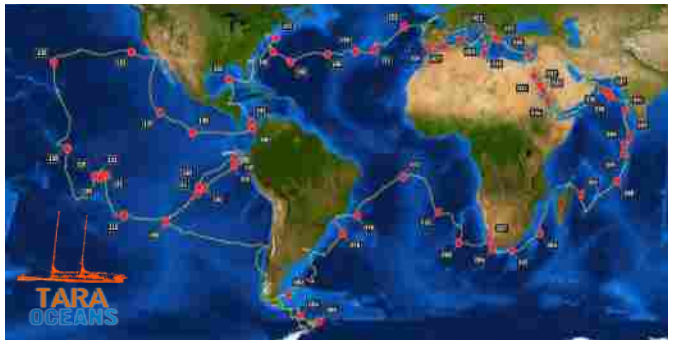
Background : annual mean surface concentrations of **nitrates (left)** and **phosphates (right)**
Primary (central circles) and secondary (outer circles) limiting nutrients: N (green), P (black), Fe (red)

Studying the functional diversity of picocyanobacteria using *Tara* Oceans metagenomes

The ISME Journal **17**, 720–732 (2023)

Differential global distribution of marine picocyanobacteria gene clusters reveals distinct niche-related adaptive strategies

Hugo Doré, Ulysse Guyet, Jade Leconte, Gregory K. Farrant, Benjamin Alric, Morgane Ratin, Martin Ostrowski, Mathilde Ferrieux, Loraine Brillet-Guéguen, Mark Hoebeke, Julka Siltanen, Gildas Le Corguillé, Erwan Corre, Patrick Wincker, David J. Scanlan, Damien Eveillard, Frédéric Partensky and Laurence Garczarek





Metagenome

↓

Recruitment against 256 reference genomes

↓

Taxonomic assignment to *Synechococcus* or *Prochlorococcus*

↓

Functional assignment (Cyanorak)

↓

Tables of genes relative abundance at each station



Cyanorak CLOGs with functional annotation



Prochlorococcus genes

Stations	CK_1	CK_2	CK_3	CK_n
ST1	2	10	0	2
ST2	0	5	20	2
ST3	5	5	15	5

Synechococcus genes

	CK_1	CK_2	CK_3	CK_n
ST1	1	5	2	7
ST2	5	2	10	3
ST3	1	3	12	3

Building ecologically representative clusters of adjacent genes (eCAGs)

Stations

	Gene 1	Gene 2	Gene 3	Gene n
ST1	2	10	0	2
ST2	0	5	20	2
ST3	5	5	15	5

Correlations between genes
based on shared distribution



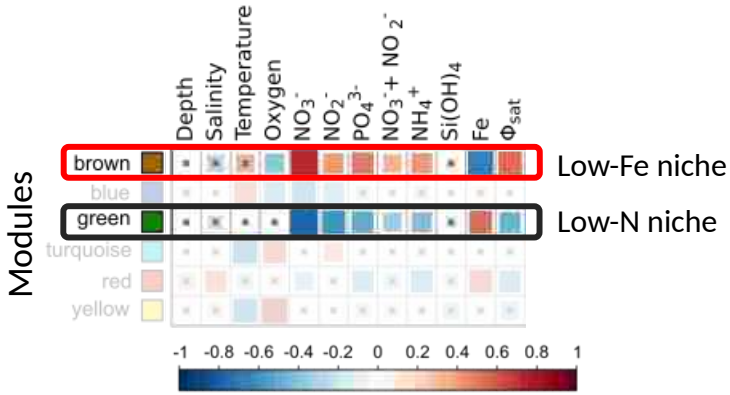
Correlation network

module detection

Groups of genes (or **modules**) displaying a similar distribution In the field



Niche-specific genes

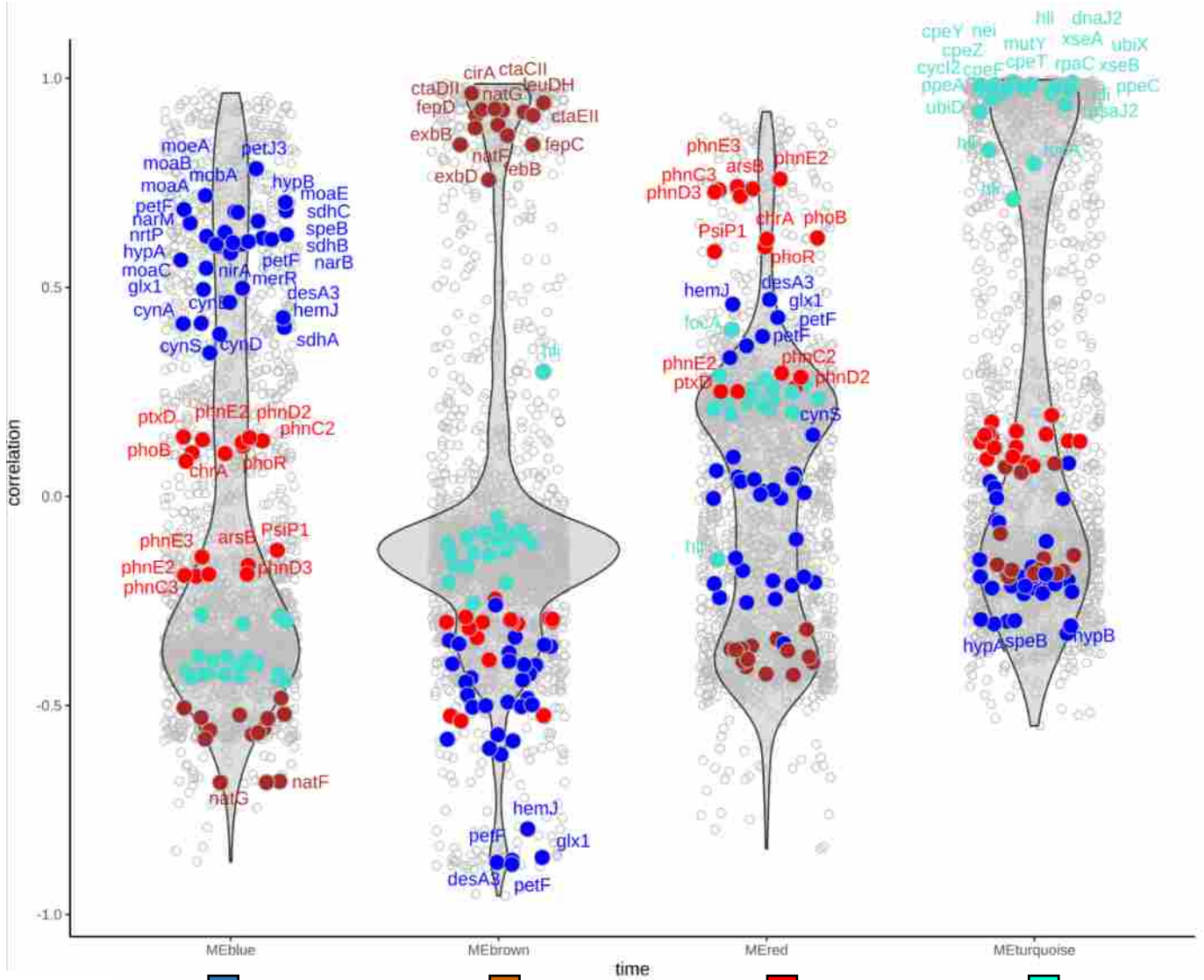


Low-Fe niche

Low-N niche

Doré et al. ISME J., 2023

Building ecologically representative clusters of adjacent genes (eCAGs)



Modules				
Env. Niches	-N, +Fe	-Fe	-P, cold, +Fe	cold, mixed waters
Major ESTUs:	HLIIA-D	HLIIIA/IVA; LLIB	HLIA	LLIA

Building ecologically representative clusters of adjacent genes (eCAGs)

Stations

	Gene 1	Gene 2	Gene 3	Gene n
ST1	2	10	0	2
ST2	0	5	20	2
ST3	5	5	15	5

Correlations between genes
based on shared distribution



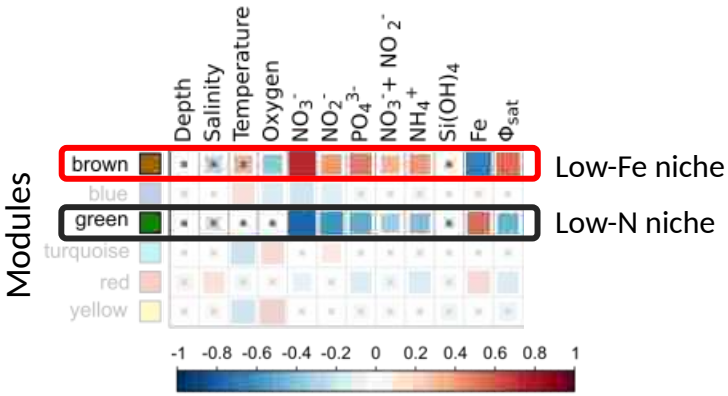
Correlation network

module detection

Groups of genes (or **modules**) displaying a similar distribution In the field



Niche-specific genes



Low-Fe niche

Low-N niche

Synteny in Cyanorak reference genomes

(gene adjacency matrix within each module)

eCAGS = clusters of genes that are both :

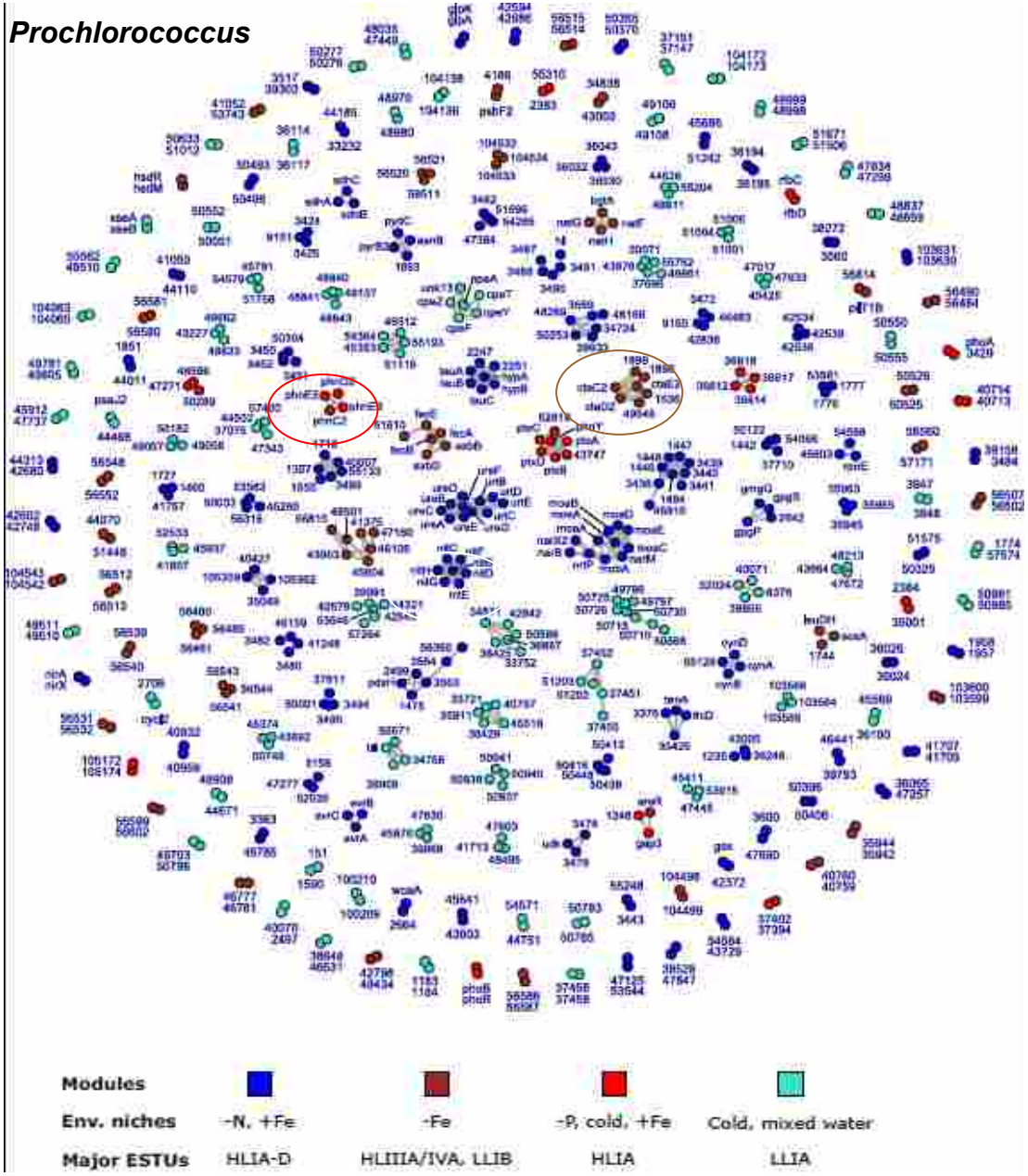
- adjacent in reference genomes,
- displaying a similar distribution *in situ*

→ potentially involved in the same metabolic pathway

Identification of not only single genes but entire gene clusters, potentially involved in the adaptation to specific ecological niches

Doré et al. ISME J., 2023

Genomic adaptation to ecological niches: a comprehensive study



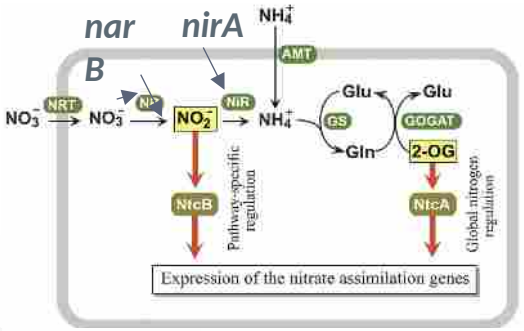
Network of accessory genes that are adjacent in reference genomes and found in specific niches in the field (**eCAGs**).

Some eCAGs involved in the metabolism of complex organic forms of nutrients => directly used by cells, for e.g. the biosynthesis of proteins or DNA or degraded into inorganic N and/or P forms.

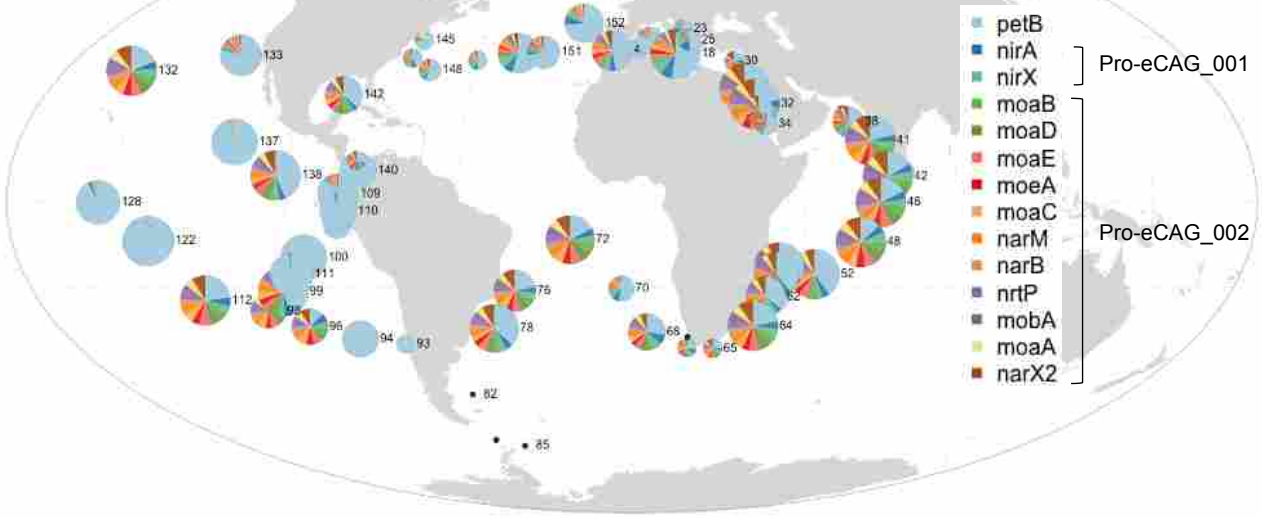
Doré, Guyet *et al*. ISME J. 2023

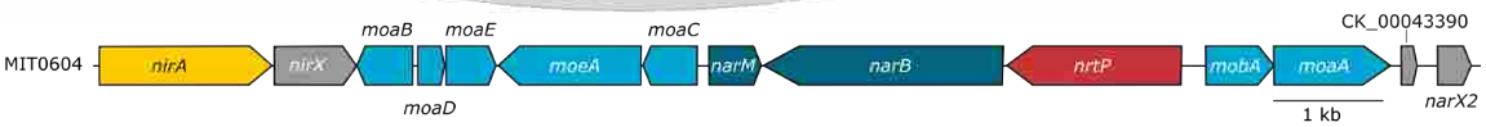
Genomic adaptation to N-limitation

Transport/assimilation of inorganic nitrate and nitrite



Prochlorococcus



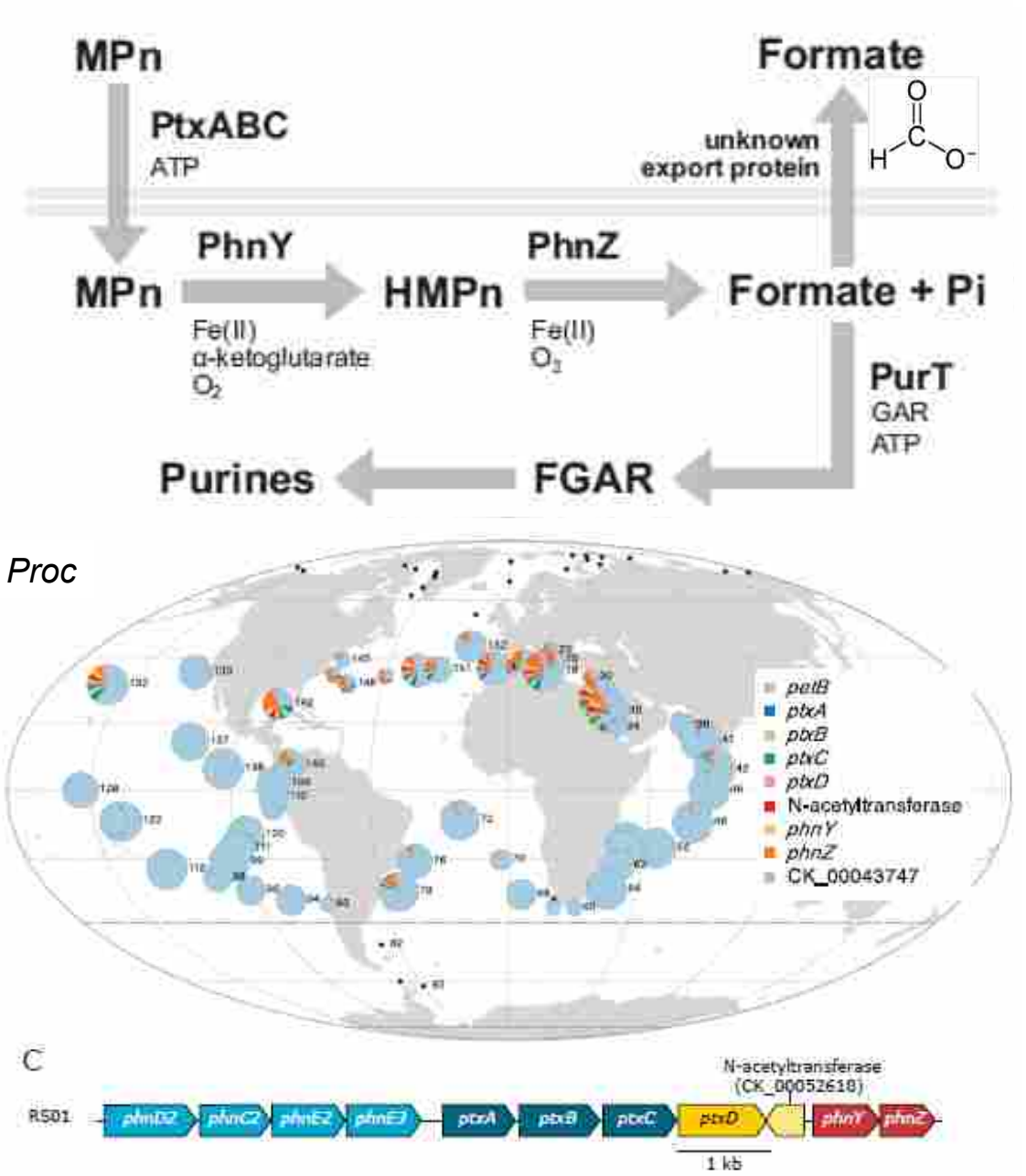


The *nirA* gene and *narB* operon are present in a large proportion of natural *Proc.* populations in N-poor areas, a much lower proportion in P-poor areas, and completely absent in low-Fe areas:

=> The only inorganic form of N for the latter is NH_4^+. But they also have access to organic N forms.

Doré, Guyet et al. ISME J. 2023

Genomic adaptation to P-limitation

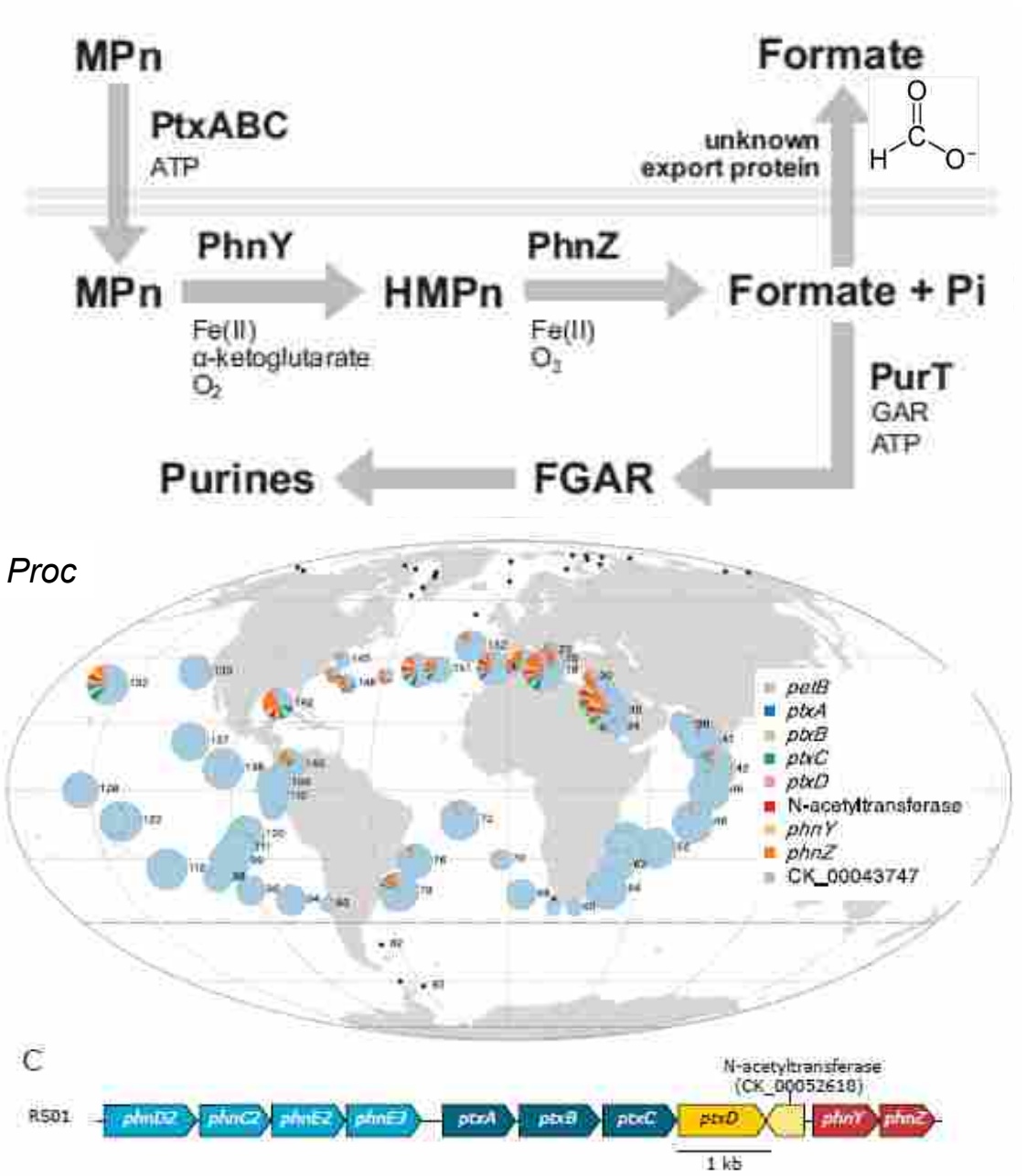


1) Most *Prochlorococcus* In P-limited areas possess an **accessory phosphonate ABC transporter (*phnC2-D2-E2-E3*),** present only in a few reference *Prochlorococcus* and no *Synechococcus* genomes. Phosphonates could be incorporated into cell surface phosphono-glycoproteins that may act to mitigate cell mortality by grazing and viral lysis.

2) Only a low proportion of the *Prochlorococcus* population in low-P areas possess the *phnYZ* operon, involved in C-P bond cleavage, the putative phosphite dehydrogenase PtxD as well as the phosphonate transporter PtxABC.

Doré, Guyet et al. ISME J. 2023

Genomic adaptation to P-limitation

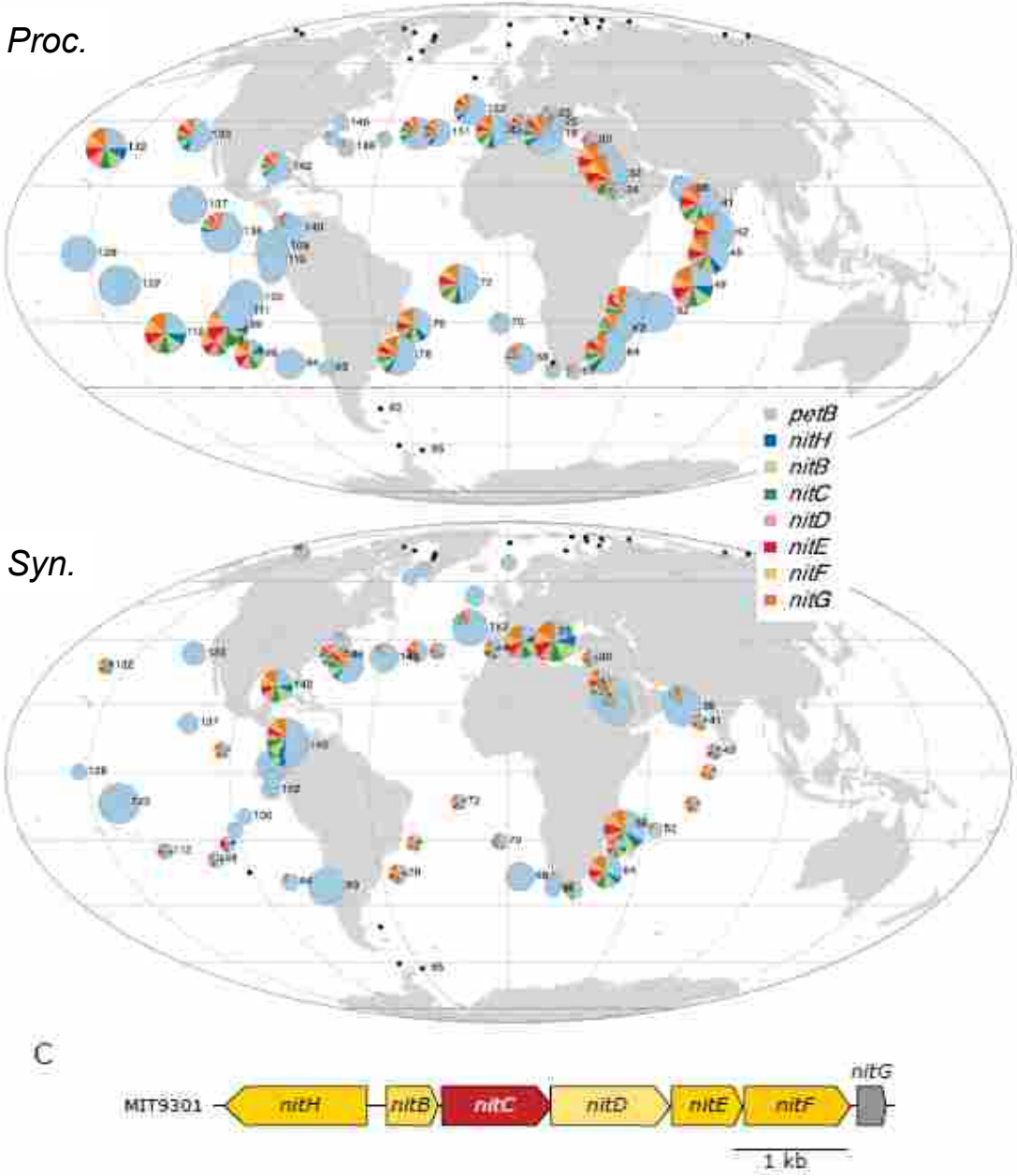


1) Most *Prochlorococcus* In P-limited areas possess an **accessory phosphonate ABC transporter (*phnC2-D2-E2-E3*),** present only in a few reference *Prochlorococcus* and no *Synechococcus* genomes. Phosphonates could be incorporated into cell surface phosphono-glycoproteins that may act to mitigate cell mortality by grazing and viral lysis.

2) Only a low proportion of the *Prochlorococcus* population in low-P areas possess the *phnYZ* operon, involved in C-P bond cleavage, the putative phosphite dehydrogenase PtxD as well as the phosphonate transporter PtxABC.

Doré, Guyet et al. ISME J. 2023

Genomic adaptation to N- and/or P-limitation



7 gene CAG encompassing the nitrilase gene (*nitC*)
=> most *Synechococcus* cells and a variable proportion of *Prochlorococcus* could use nitriles or cyanides in warm, Fe replete waters (e.g. low N areas of the Indian Ocean).

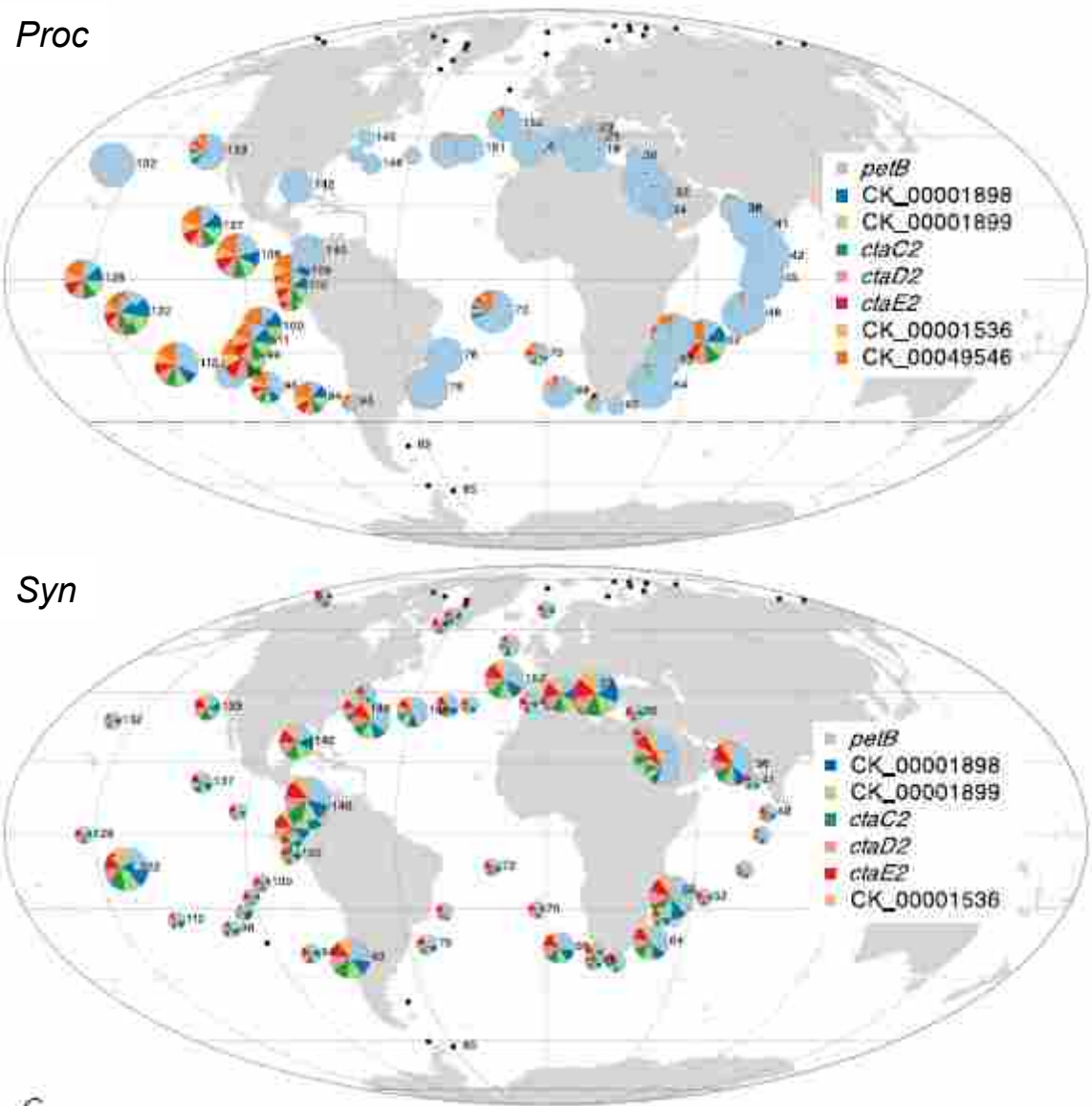
In *Pseudomonas fluorescens*, the whole operon (*nitHBCDEFG)* is upregulated in the presence of cyanide triggering an increase in the rate of NH_4^+ accumulation

=> cyanide could constitute an alternative nitrogen source in marine picocyanos as well though its source is yet unknown.

The operon might also be involved in cell detoxification.

Doré, Guyet et al. ISME J. 2023

Genomic adaptation to N- and/or P-limitation



The vast majority of *Proc* cells in low-Fe regions possess an eCAG encompassing the *ctaC2-D2-E2* operon, encoding the alternative respiratory terminal oxidase (ARTO), involved in the reduction of Fe(III) into Fe(II) prior to its transport through the plasma membrane.

⇒ ARTO is critical for *Proc.* adaptation to low-Fe areas while it is almost core in *Syn*.

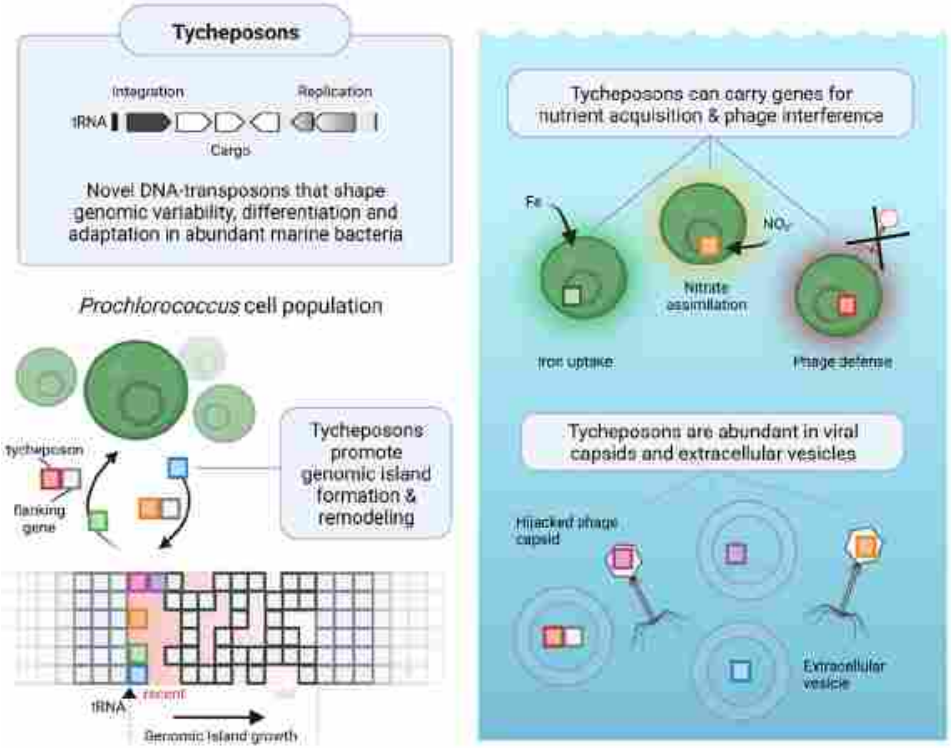
Doré, Guyet et al. ISME J. 2023

How is genome flexibility generated and how is it transferred between cells?

2023, Cell *186*, 47–62

Novel integrative elements and genomic plasticity in ocean ecosystems

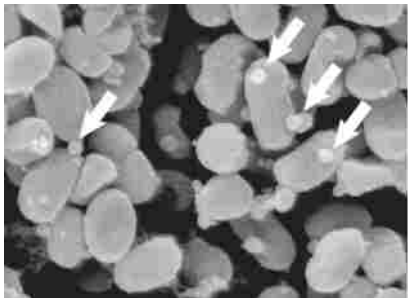
Thomas Hackl, Raphaël Laurenceau,
Markus J. Ankenbrand, ...,
Edward F. DeLong, Steven J. Biller,
Sallie W. Chisholm



SCIENCE VOL 343 10 JANUARY 2014

Bacterial Vesicles in Marine Ecosystems

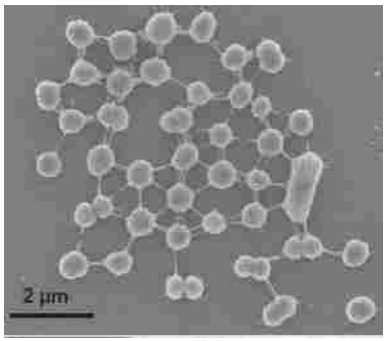
Steven J. Biller,[1*] Florence Schubotz,[2] Sara E. Roggensack,[1] Anne W. Thompson,[1] Roger E. Summons,[2] Sallie W. Chisholm[1,3*]



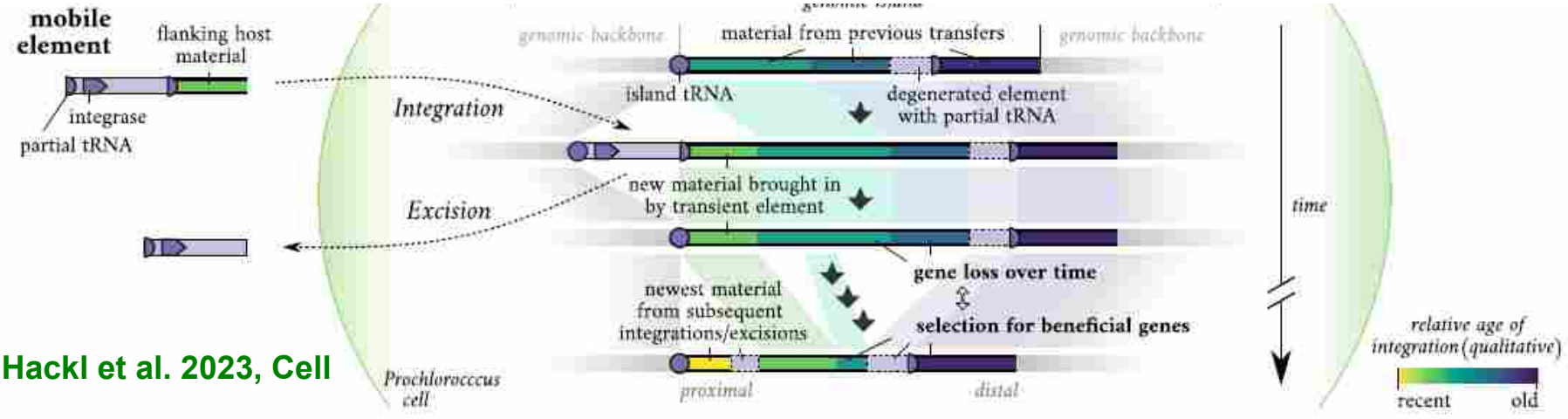
Sci. Adv. **10**, eadj1539 (2024)

Direct interaction between marine cyanobacteria mediated by nanotubes

Elisa Angulo-Cánovas[1], Ana Bartual[2], Rocío López-Igual[3]†, Ignacio Luque[3]†, Nikolai P. Radzinski[4], Irina Shilova[5], Maya Anjur-Dietrich[4], Gema García-Jurado[6], Bárbara Úbeda[2], José Antonio González-Reyes[7], Jesús Díez[1], Sallie W. Chisholm[4], José Manuel García-Fernández[1]*, María del Carmen Muñoz-Marín[1]*

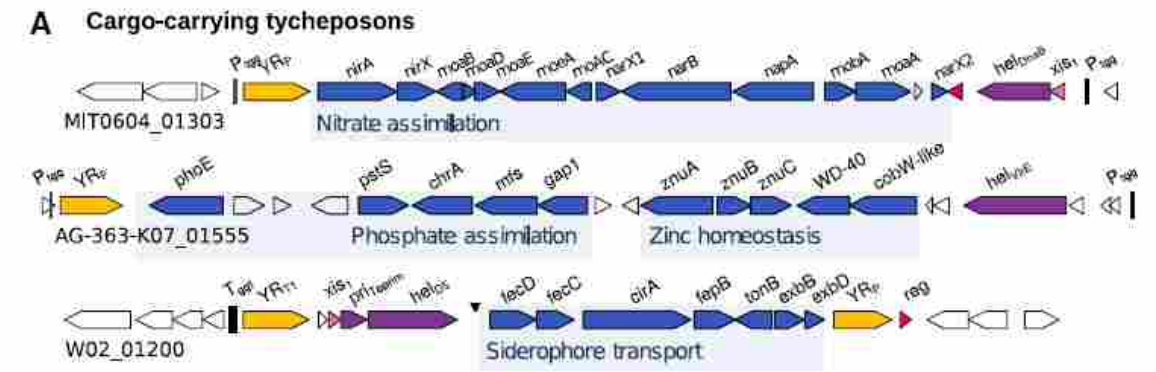


Generation of genomic islands : The role of “Tycheposons”



Hackl et al. 2023, Cell

- Tycheposons are novel, transient DNA transposons promoting genomic adaptation in *Proc.*
- About 4/5 of the tycheposons carry “cargo”, such as nutrient-acquisition gene cassettes



YR: tyrosine recombinase, a site-specific integrase

Target: one of 7 tRNAs

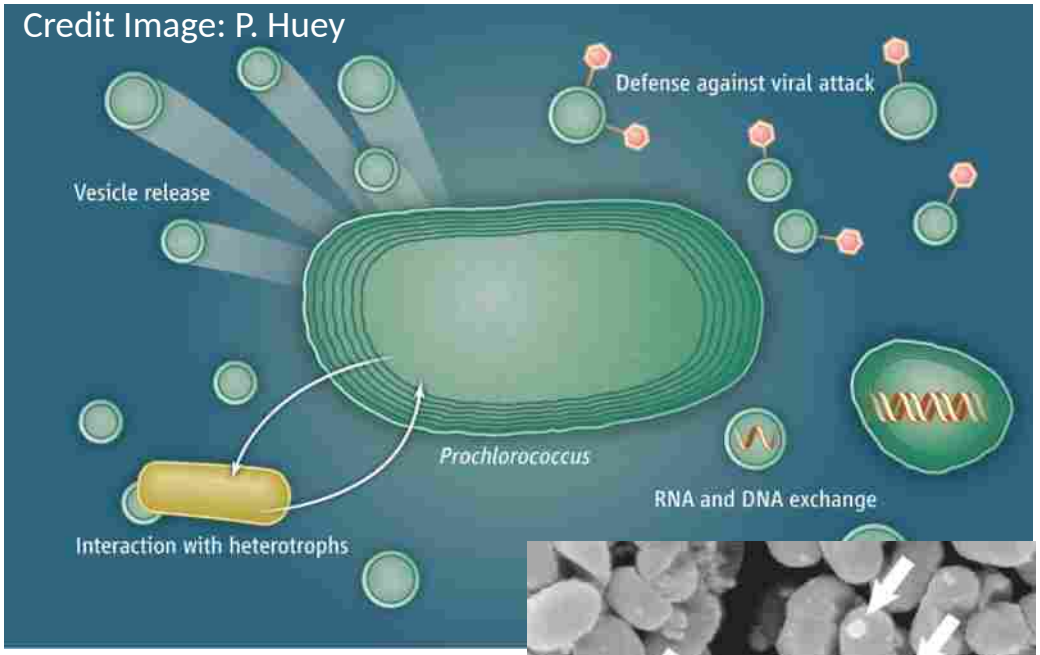
- Tycheposons accelerate genomic island formation and remodeling

Transfer of genetic material between *Prochlorococcus* cells ?

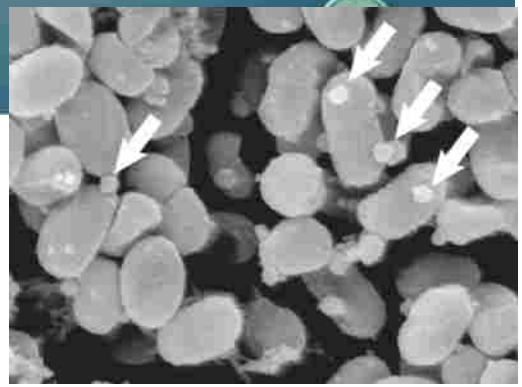
- *Prochlorococcus* have no plasmid nor pili (except LLIV) => no possible conjugation ?

- Likely role of cyanophages to transfer genes between cells (auxiliary genes/tycheposons)

- Recent discovery of vesicle production and release by *Prochlorococcus* cells



Biller et al. (2014) *Science*

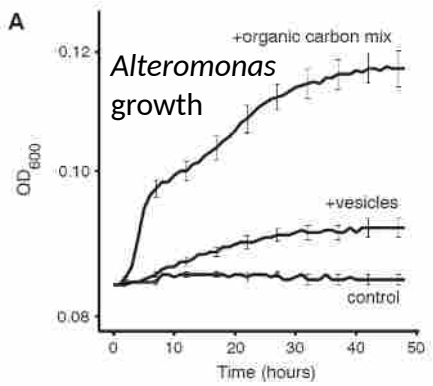


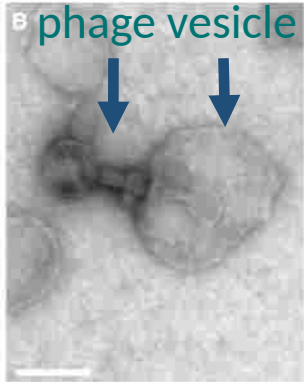
Vesicles contain :

- Periplasmic and outer membrane proteins
- Membranes
- DNA and RNA
- Tycheposon hallmark genes

Likely role:

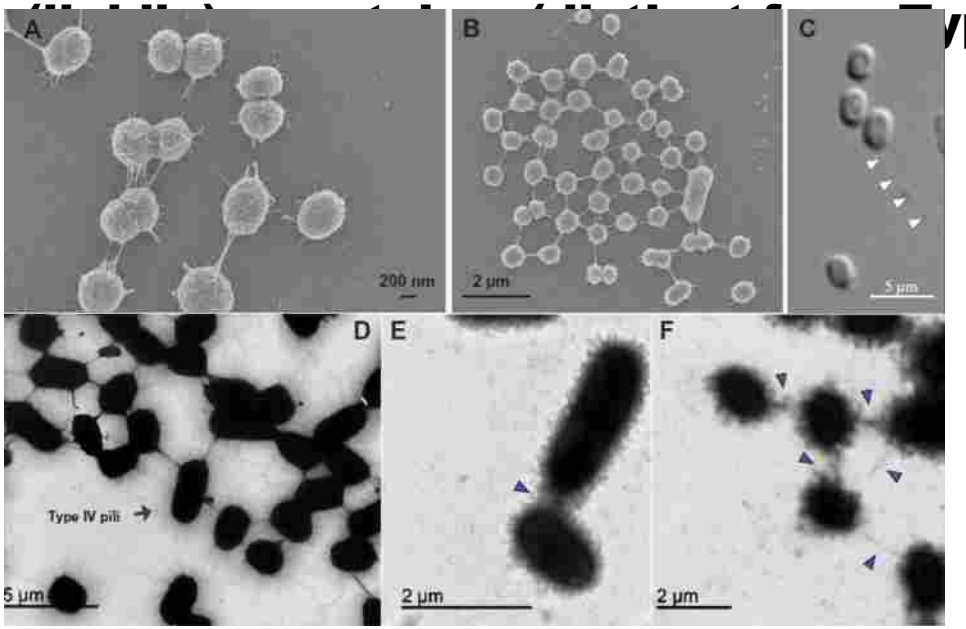
- Release of genetic material (lateral transfer)
- Release of nutrients for 'helper' bacteria (e.g. catalase-producing) ?
- Defense against viruses (lure)?





Transfer of genetic material between *Prochlorococcus* cells ?

Inter-and intra-genus exchange of cytoplasmic material can also occur directly between neighboring and distant cells of cyanobacteria mediated by inter-cellular membrane (tubular) nanotubes (distinct from Type

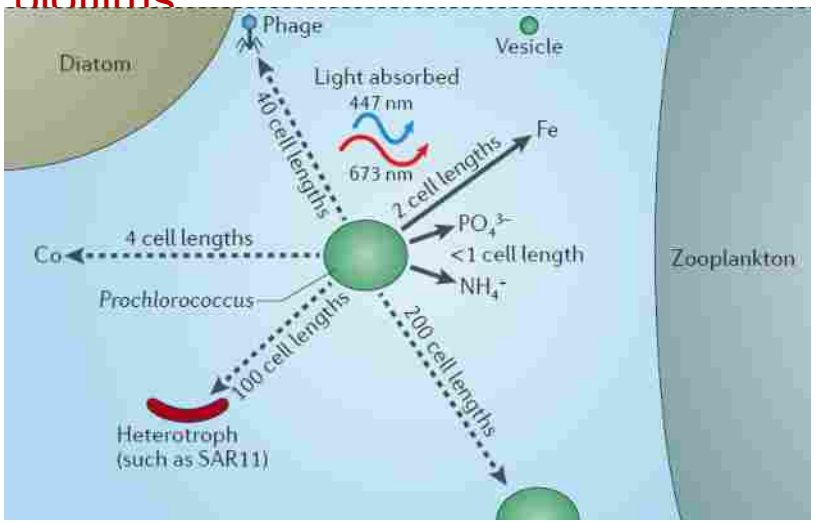


Nanotubes project from the cell surface at different positions, each mediating a connection with another cell, so that one cell can be connected to several others simultaneously.

Angulo-Cánovas et al. (2024) *Science Adv.*

These cell-cell connections were long overlooked since they are much less frequent in planktonic given how dilute they are.

So, in the extremely dilute oligotrophic oceans (ca. 200 cell length between two cells) nanotubes could be important only in some specific conditions, e.g. in marine snow, where the bacteria-bacteria separation on average is about 10 µm, or in marine biofilms.



Biller et al. (2014) *Nature Rev. Microbiol.*

Take home messages

- Picocyanobacteria are ideal models to understand the links between biology and ecology thanks to:
  - ✓ *in situ a*bundance and ubiquity, and so abundance in metagenomes
  - ✓ Presence in fairly 'easy-to-access' ecological niches: the lit layer of aquatic systems
  - ✓ Availability of numerous strains & reference genomes

- Picocyanobacteria display an extraordinary genetic and functional diversity to cope with the variety of niches they occupy in the Ocean, resulting from subtle combinations of temperature, light and nutrient gradients.

- This genomic diversity is reflected by the huge toolbox of accessory genes (the pan-genome), in which locally useful genes can be recruited by natural selection, and we are only starting to decipher what mechanisms picocyanobacteria actually use to do that.

- Many biologists are working all their carrer on single model organisms maintained for years in culture, such as *Synechocystis* PCC 6803. The lesson from studying picocyanobacteria is that cyanobacteria are a collective, not a single organism, and one cannot understand the collective by studying just one strain.

- Picocyanobacteria can be used in models of the current and future marine Ocean as biomarkers of the fate of phytoplankton communities in the context of global change, with *Proc*. being a marker of oligotrophic areas, and *Syn*. being a generalist and behaving more like larger phytoplankton species.

Acknowledgments

SBR

ECOMAP Team

Laurence Garczarek
Morgane Ratin

Alexis Dufresne
Gregory Farrant
Hugo Doré
Ulysse Guyet

Bioinformatics Core Facility

Mark Hoebeke
Jukka Siltanen
Erwan Corre
Loraine Guéguen

Other collaborators

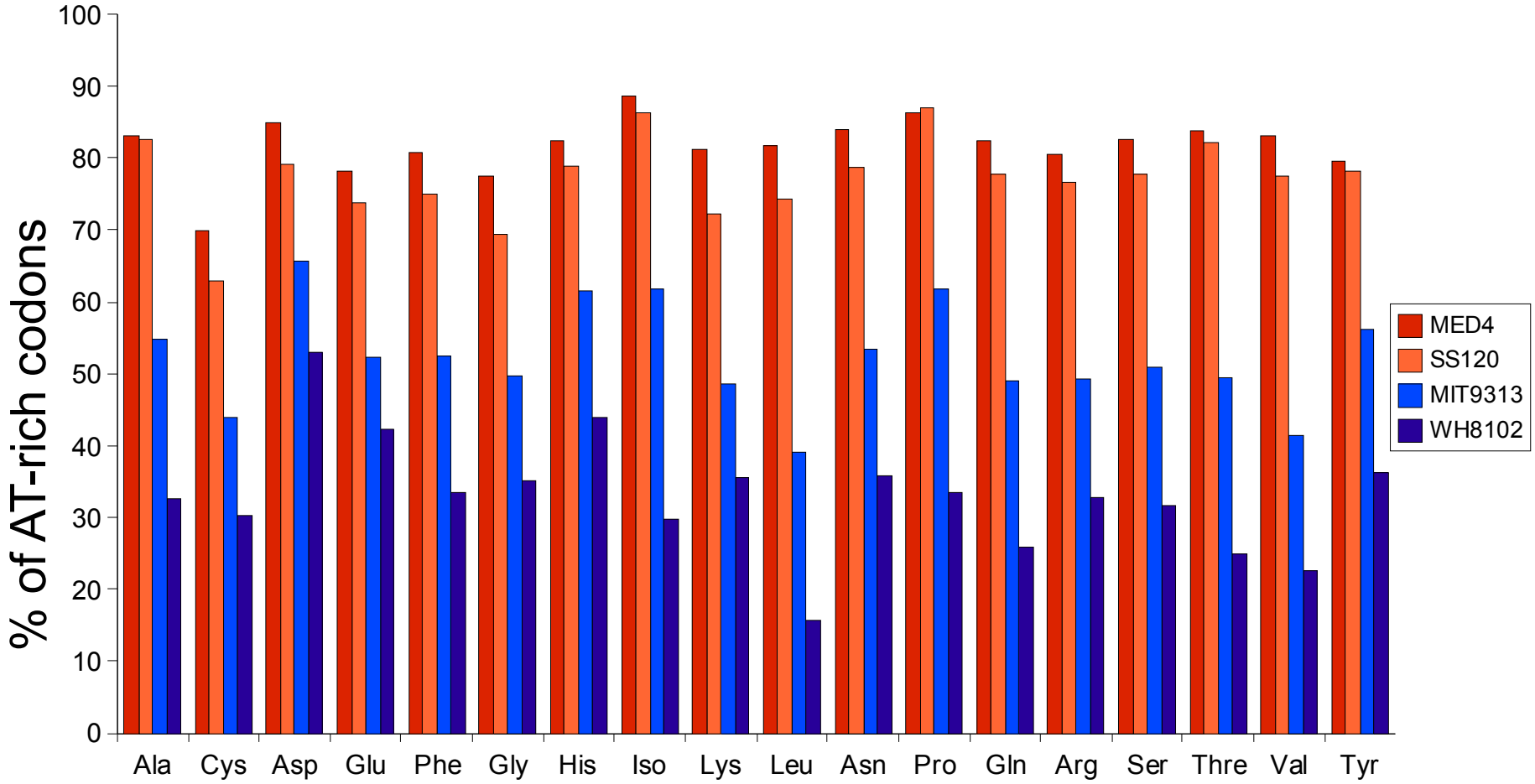
David J. Scanlan (Warwick Univ., UK)

Funding

Genoscope, France
Tara Oceans Foundation
ANR SAMOSA (2014-2018)
ANR CINNAMON (2018-22)
ANR TONGA (2019-22)
ANR EFFICACY (2020-2024)
ANR TaxCy (2024-2028)
Assemble Plus (2017-2022)

Thanks for your attention!

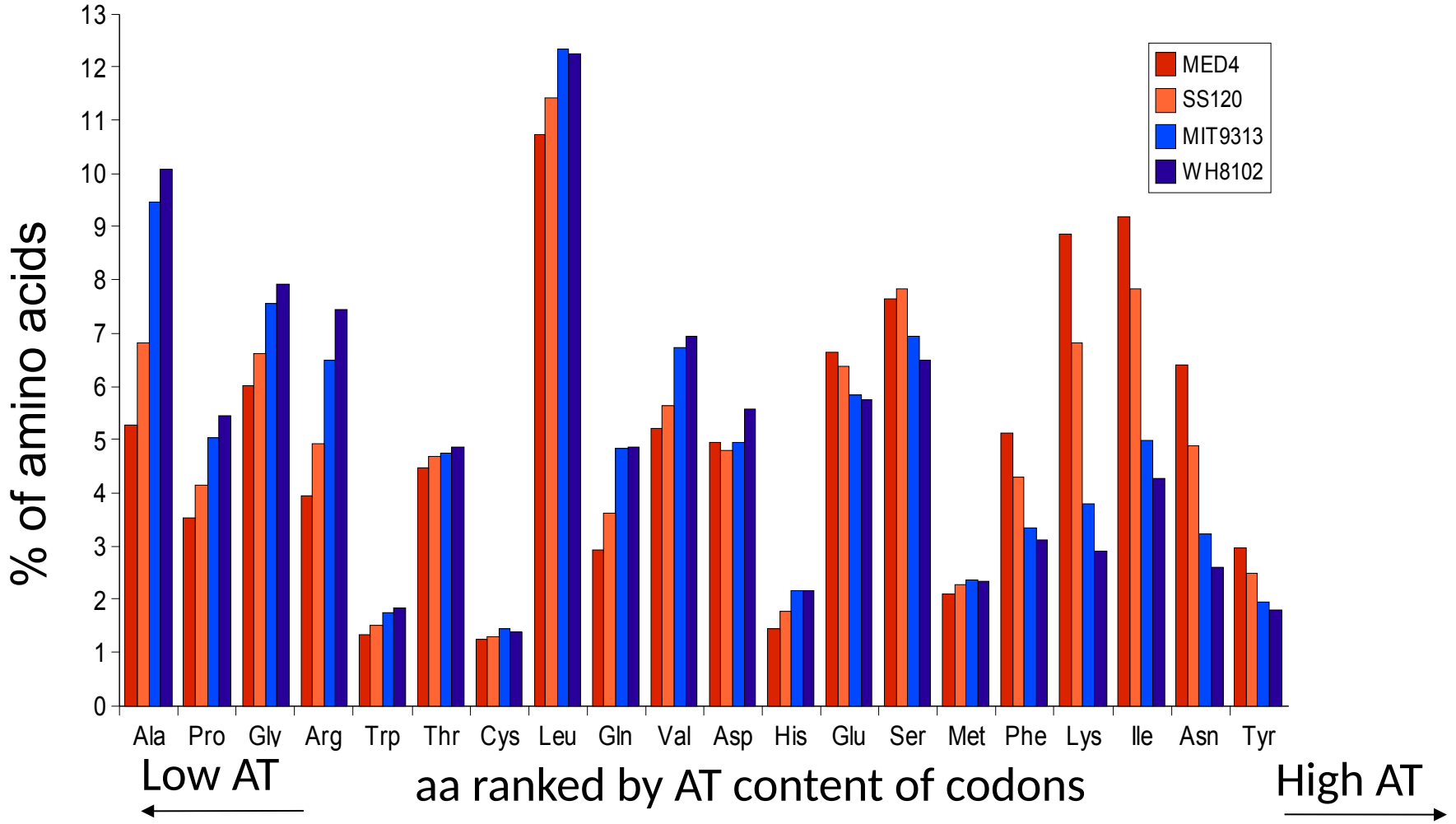
Codon Usage in Marine Picocyanobacteria



For any given amino acid, MED4 and SS120 always preferentially use AT-rich codons => the bias toward AT strongly influences codon usage

Dufresne et al., 2005, *Genome Biology*

Use of Amino acids in Marine Picocyanobacteria

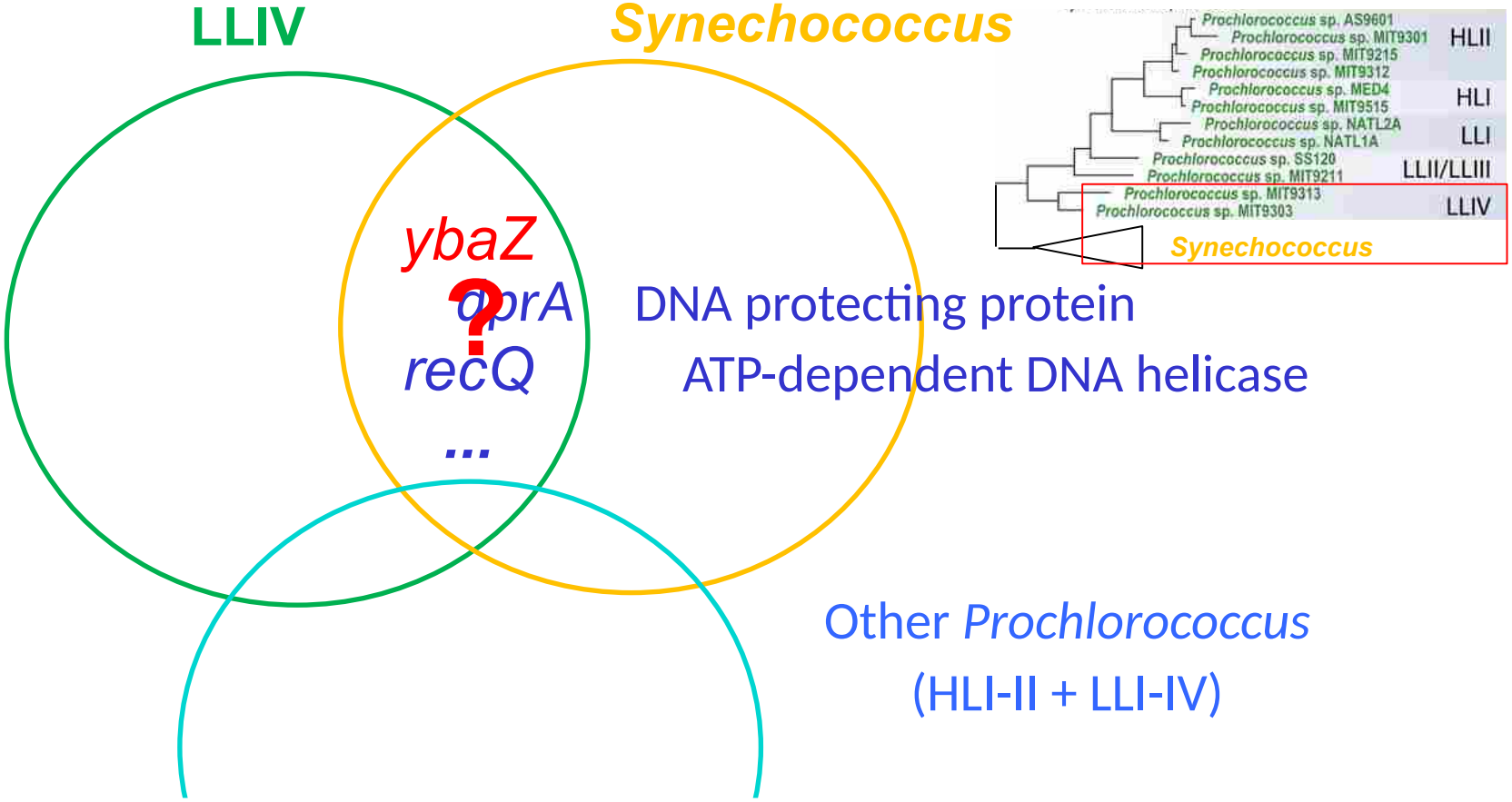


MED4 and SS120 use more AT-rich (and less GC-rich) amino acids than the two other strains

=> the bias toward AT strongly influences protein composition

Dufresne et al., 2005, *Genome Biology*

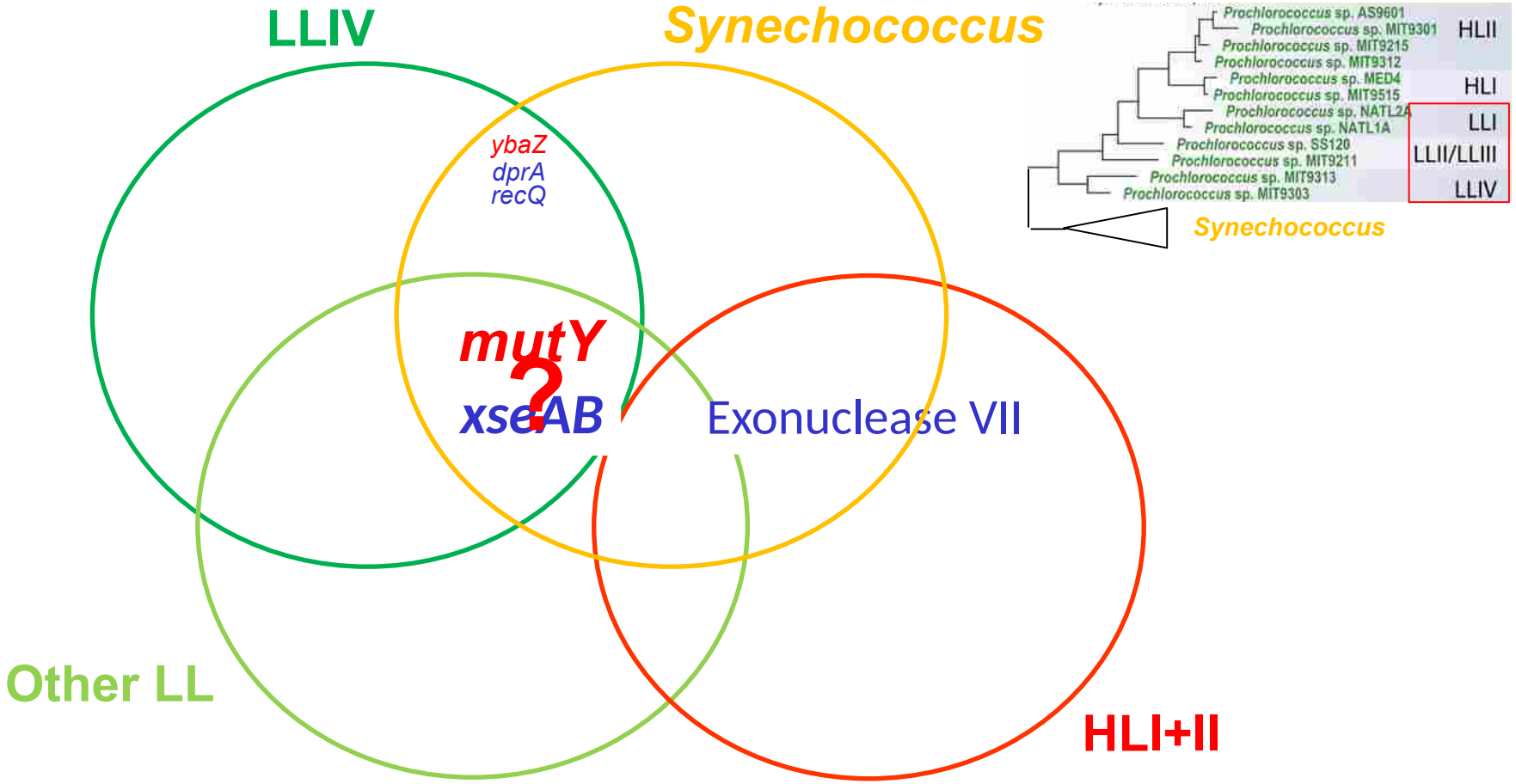
Differential Distribution of DNA Replication and DNA Repair Genes Among *Prochlorococcus* Ecotypes



ybaZ codes for an alkyl transferase involved in the repair of GC=>AT transversions, like ***ada*** and ***ogt***, which are mutator genes in *E. coli* (i.e., when inactivated, they provoke an increase of mutation rate).

Partensky & Garczarek (2010) *Ann. Rev. Mar. Sci.*

Differential Distribution of DNA Replication and DNA Repair Genes among *Prochlorococcus* Ecotypes

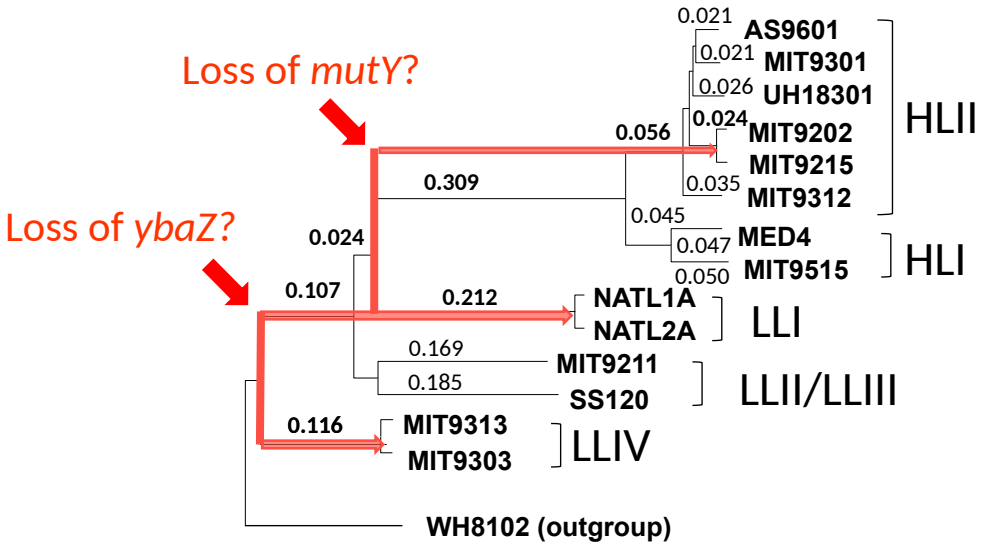


mutY codes for an A/G-specific adenine glycosylase, a mutator gene in *E. coli*
(repair of GC=>AT transversions)

Partensky & Garczarek (2010) *Ann. Rev. Mar. Sci.*

Gene Evolution Rate in *Prochlorococcus*

NJ tree using concatenated sequences of 1,014 core proteins from 15 genomes, including *Syn*. WH8102 as outgroup: branch lengths are proportional to aa substitution rate (i.e., the # of aa substitutions per site).



- LLI / LLIV ratio = 2.95
- HLII / LLI ratio = 1.84
- HLII / LLIV ratio = 5.51

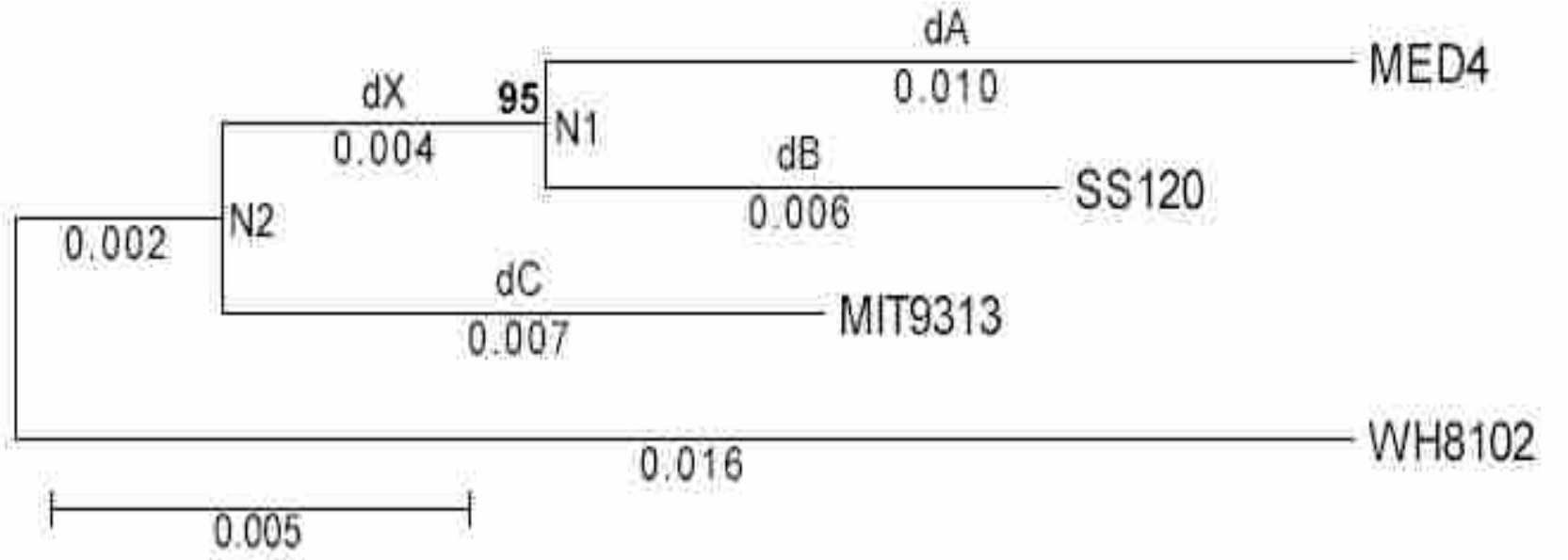
Data unpublished but derived from a similar analysis in Dufresne et al. (2005) using only 4 genomes.

Significant differences in the evolution rates of protein-coding genes between lineages:

=> A significant acceleration of mutation rates seemingly occurred twice during *Proc*. Evolution

In comparison, no such acceleration occurred within the *Syn*. lineage.

Pairwise Distance Computation using 16S rRNA gene

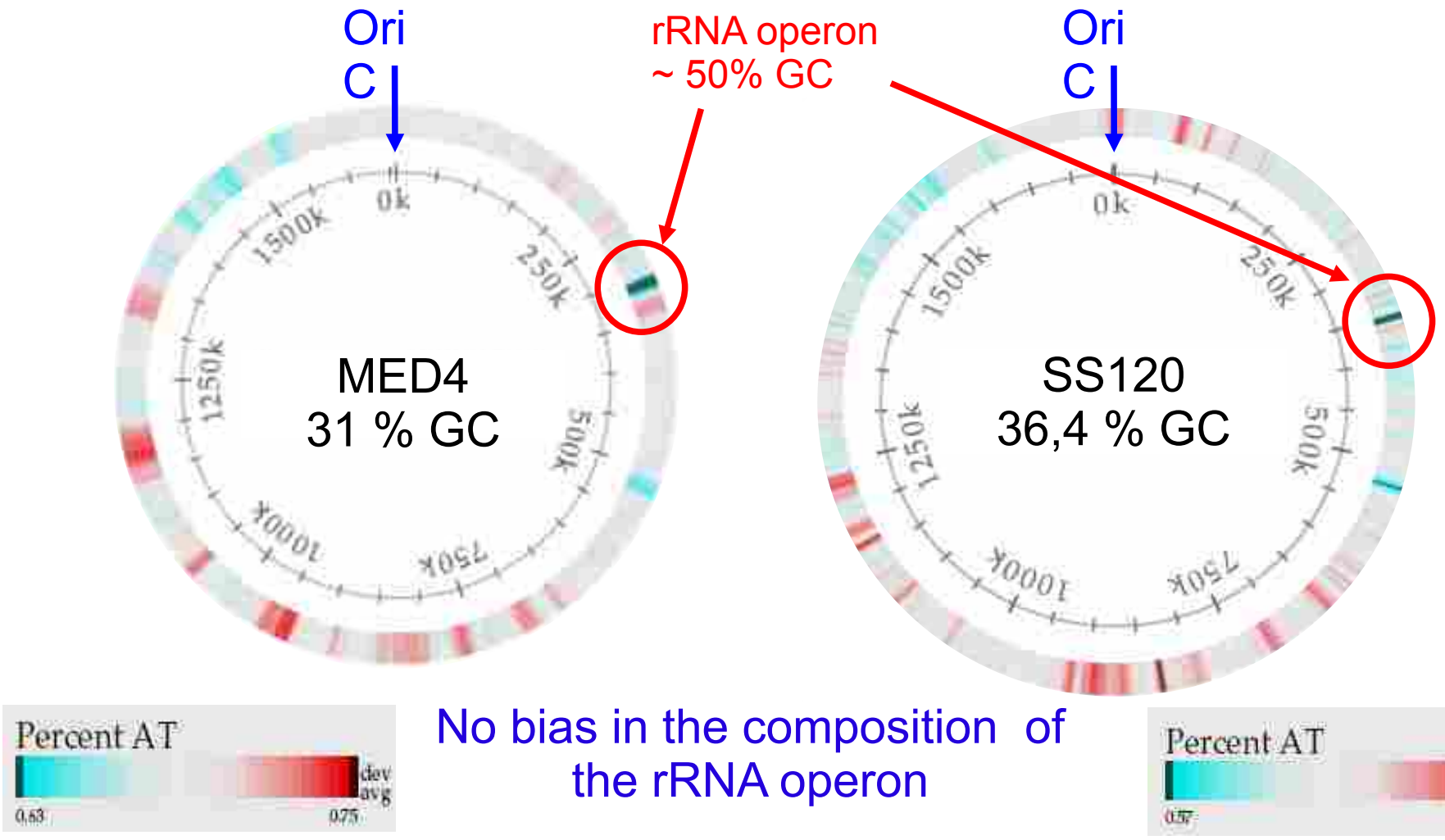


Evolution rates of the 16S rRNA gene in MED4, SS120 & MIT9313 are not statistically different (relative rate tests)

=> Hypothesis of a constant evolutionary rate valid for the 16S rRNA gene

Dufresne et al., 2005, *Genome Biology*

The rRNA operon, a GC-rich island in an AT-rich Sea



No bias in the composition of the rRNA operon

CBS Genome Atlas Database http://www.cbs.dtu.dk/services/GenomeAtlas/

Summary: Hypotheses for the occurrence of Genome Streamlining

The loss of DNA repair genes in some genotypes could have created 'mutator strains' (increased mutation rate) => selective advantage to conquer new ecological niches.

Several 'mutator' events could have occurred during the evolution of *Prochlorococcus:*

1) Differentiation of the ancestor of all streamlined genomes from the LLIV branch
=> Loss of *ybaZ*, potentially a 'strong' mutator gene

2) Differentiation of the HL branch
=> Loss of *mutY*, a 'mild' mutator gene

These events must have been followed each time by restoration of a normal mutation rate.

Dufresne et al. 2005, *Genome Biol.*

Partensky & Garczarek (2010) *Ann. Rev. Mar. Sci.*

'Genome streamlining was probably driven by strong, genome-wide selection for the removal of genes with only a small fitness benefit that was outweighed by the associated costs.'

Biller et al. (2015) Nature Reviews in Microbiology

Remarks about Genome Streamlining

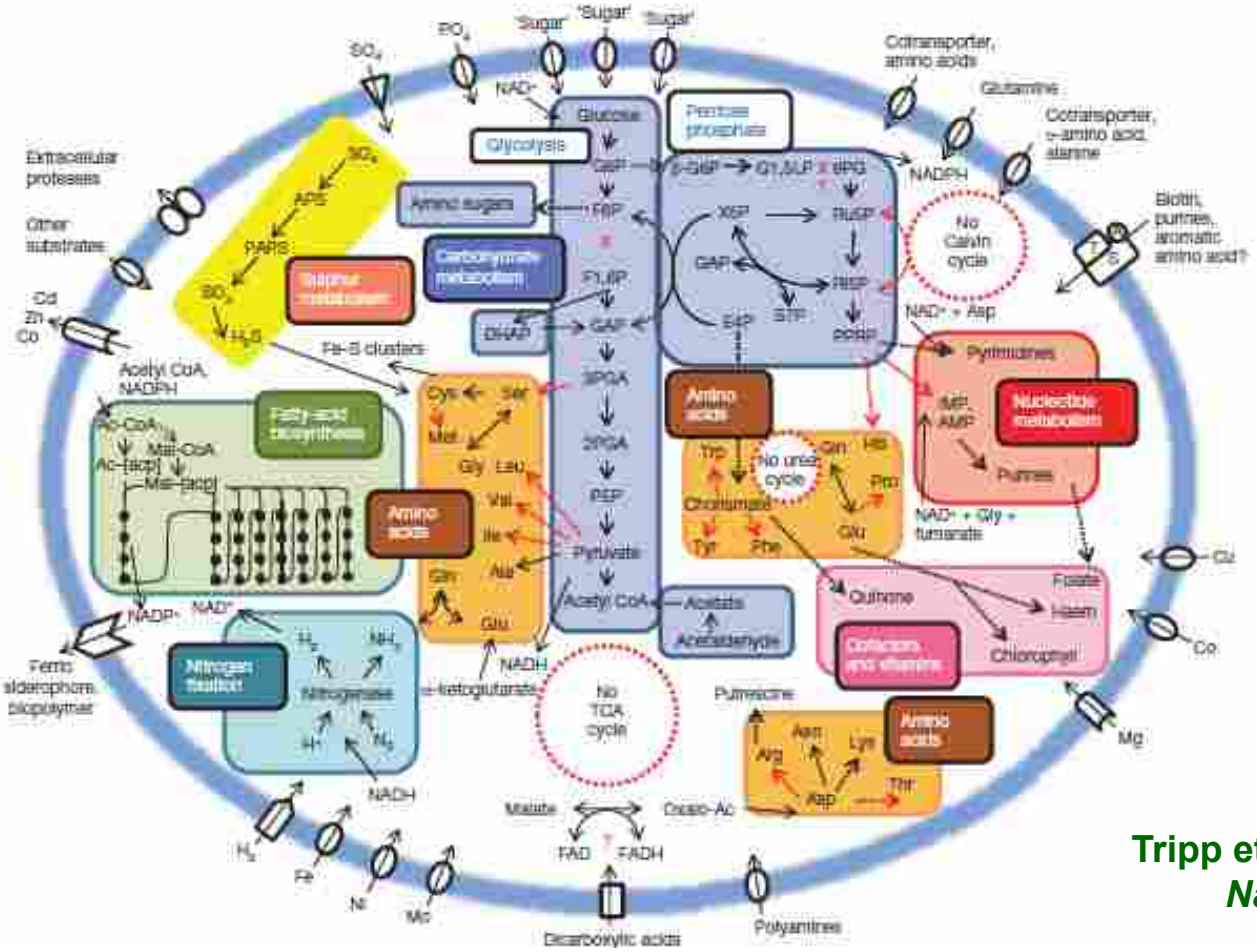
Prochlorococcus is not the only free-living prokaryote with reduced genome and GC%

=> SAR11, the most abundant bacteria in the ocean, particularly abundant in oligotrophic areas.

Similar genome evolution to that of host-dependent microorganisms (parasites, endosymbionts, etc.), but with different causes (rich vs. poor environment, small vs. large populations, strong vs. weak genetic drift, etc.)

=> ex: UCYN-A

An example of endosymbiont : UCYN-A, an anoxygenic cyanobacterium !



Tripp et al. (2010)
Nature

Reduced genome (1.44 Mb) compared to *Cyanothece* (> 5.4 Mb) but with :

- no PSII nor antenna genes => no O_2 produced!
- no genes for Calvin Cycle => no CO_2 fixation!
- no Krebs cycle genes => no ATP generated by oxidation of acetate).

⇒ UCYN-A: Photoheterotroph that uses PSI for energy generation ?

⇒ Symbiont of microalga sharing functions with its host (it fixes N_2 but uses organic C from its host).

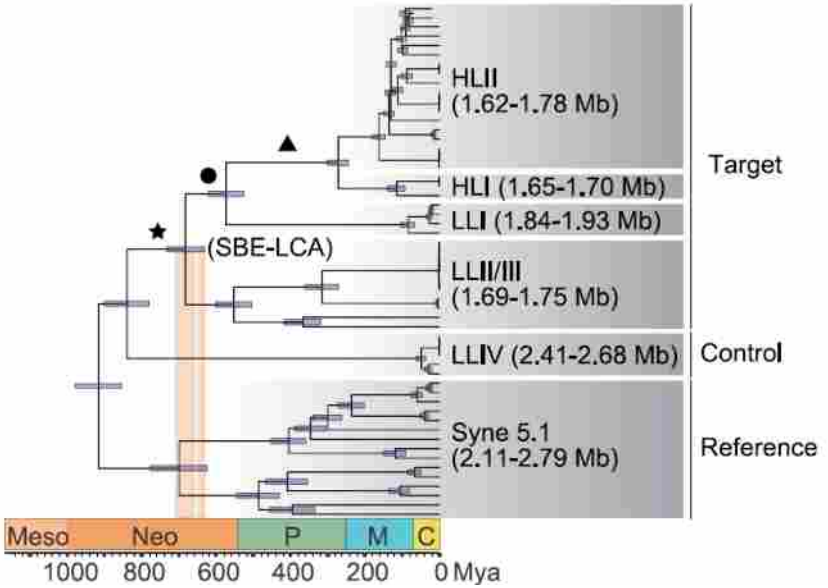
Recent alternative theory for *Prochlorococcus* genome streamlining

The ISME Journal, 2024, **18(1)**, 1–7

Genome reduction occurred in early *Prochlorococcus* with an unusually low effective population size

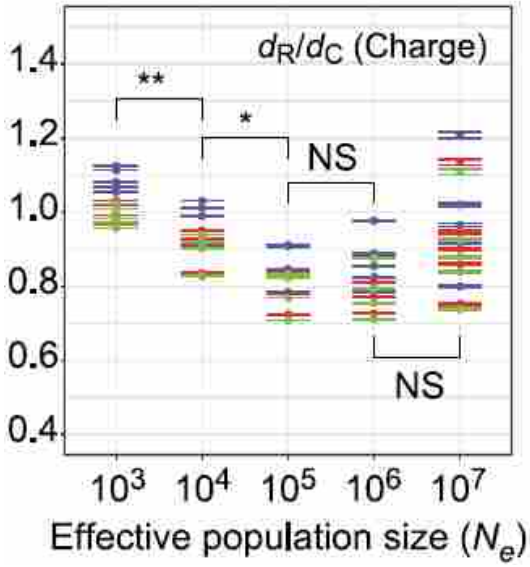
Hao Zhang[1,2,‡], Ferdi L. Hellweger[3,‡], Haiwei Luo[1,4,*]

School of Life Sciences, The Chinese University of Hong Kong



The "genome streamlining" theory posits that *Proc.* have extremely large effective population sizes (*Ne*) so that selection for metabolic efficiency acts to drive genome reduction. Because genome reduction largely took place anciently, this theory builds on the assumption that their ancestors' *Ne* was also large.

Method: Simulate changes in ratio of radical to conservative non-synonymous nt substitution rate (*dR/dC*) in a range of *Ne* in ancestral populations. This proxy showed decreases of *Ne* only when *Ne* falls to about 10^4–10^5 or lower, magnitudes characteristic of *Ne* of obligate endosymbiont species where drift drives genome reduction.



=> Simulations would support a scenario where the primary force of *Proc.* genome reduction was drift rather than selection.