University of Groningen

Novel isolates expand the physiological diversity of<i>Prochlorococcus</i>and illuminate its macroevolution

Becker, Jamie; Pollak, Shaul; Berta-Thompson, Jessie W.; Becker, Kevin W.; Braakman, Rogier; Dooley, Keven D.; Hackl, Thomas; Coe, Allison; Arellano, Aldo; LeGault, Kristen N.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Early version, also known as pre-print

Publication date:
2023

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Becker, J., Pollak, S., Berta-Thompson, J. W., Becker, K. W., Braakman, R., Dooley, K. D., Hackl, T., Coe, A., Arellano, A., LeGault, K. N., Berube, P. M., Biller, S. J., Cubillos-Ruiz, A., Mooy, B. A. S. V., & Chisholm, S. W. (2023). *Novel isolates expand the physiological diversity of<i>Prochlorococcus</i>and illuminate its macroevolution*. BioRxiv. https://doi.org/10.1101/2023.12.03.569780

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], Benjamin A. S. Van Mooy[2], and Sallie W. Chisholm[1,4].

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

[2]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

[3]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

[4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

Correspondence: JW Becker* or SW Chisholm, Departments of Civil and Environmental Engineering & Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Bldg 48-419, Cambridge, MA 02139, USA
E-mail: jamie.becker@alvernia.edu or chisholm@mit.edu

*Current address: Science Department, Alvernia University, Reading, PA, USA
#Current address: Division of Microbial Ecology, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria
+Current address: GEOMAR – Helmholtz Centre for Ocean Research Kiel, Kiel, Germany
†Current address: Department of Biological Sciences, Wellesley College, Wellesley, MA, USA

Running title: *Prochlorococcus* diversity and evolution

Originality-Significance Statement

The marine cyanobacterium, *Prochlorococcus*, is among the Earth's most abundant organisms, and much of its genetic and physiological diversity remains uncharacterized. While field studies help reveal the scope of diversity, cultured isolates allow us to link genomic potential to physiological processes, illuminate eco-evolutionary feedbacks, and test theories arising from comparative genomics of wild cells. Here, we report the isolation and characterization of novel low-light (LL) adapted *Prochlorococcus* strains that fill in multiple evolutionary gaps. These new strains are the first cultivated representatives of the LLVII and LLVIII paraphyletic grades of *Prochlorococcus*, which are broadly distributed in the lower regions of the ocean euphotic zone. Each of these grades is a unique, highly diverse section of the *Prochlorococcus* tree that separates distinct ecological groups: the LLVII grade branches between monophyletic clades that have facultatively particle-associated and constitutively planktonic lifestyles, while the LLVIII grade lies along the branch that leads to all high-light (HL) adapted clades. Characterizing strains and genomes from these grades yields insights into the large-scale evolution of *Prochlorococcus*.

The new LLVII and LLVIII strains are adapted to growth at very low irradiance levels and possess unique light-harvesting gene signatures and pigmentation. The LLVII strains represent the most basal *Prochlorococcus* group with a major expansion in photosynthetic antenna genes. Further, a strain from the LLVIII grade challenges the paradigm that all LL-adapted *Prochlorococcus* exhibit high ratios of chl b:a_2. These findings provide insights into major transitions in *Prochlorococcus* evolution, from the benthos to a fully planktonic lifestyle and from growth at low irradiances to the rise of the HL-adapted clades that dominate the modern ocean.

Abstract

Prochlorococcus is a diverse picocyanobacterial genus and the most abundant phototroph on Earth. Its photosynthetic diversity divides it into high- or low-light adapted groups representing broad phylogenetic grades - each composed of several monophyletic clades. Here we physiologically characterize four new *Prochlorococcus* strains isolated from below the deep chlorophyll maximum in the North Pacific Ocean and combine this information with genomic and evolutionary analyses. The isolates belong to deeply-branching low-light adapted clades that have no other cultivated representatives and display some unusual characteristics. For example, despite its otherwise low-light adapted physiological characteristics, strain MIT1223 has low chl b_2 content similar to high-light adapted strains. Isolate genomes revealed that each strain contains a unique arsenal of pigment biosynthesis and binding alleles that have been horizontally acquired, contributing to the observed physiological diversity. Comparative genomic analysis of all picocyanobacteria reveals that Pcb, the major pigment carrying protein in *Prochlorococcus*, greatly increased in copy number and diversity per genome along a branch that coincides with the loss of facultative particle attachment. Collectively, these observations add support to the current macroevolutionary model of picocyanobacteria, where niche constructing radiations allowed ancestral lineages to transition from a particle-attached to planktonic lifestyle and broadly colonize the water column, followed by adaptive radiations near the surface that pushed ancestral lineages deeper in the euphotic zone resulting in modern depth-abundance profiles.

Introduction

The picocyanobacterium *Prochlorococcus* is numerically the most abundant photoautotrophic organism in the global ocean (Flombaum *et al.*, 2013). It represents an enormous and genetically diverse collective of cells (Kashtan *et al.*, 2014) tuned to thrive under particular environmental conditions (Kettler *et al.*, 2007; Biller *et al.*, 2014). *Prochlorococcus* cells have traditionally been divided into high-light (HL) and low-light (LL)-adapted groups based on their optimal light intensity for growth, which generally corresponds to the depths at which they display their maximal abundance (Moore *et al.*, 1995, 1998; West and Scanlan, 1999). Finer scale niche partitioning among lineages has been linked to other abiotic factors, including temperature, trace metal and inorganic nitrogen acquisition, and vertical mixing (Bouman *et al.*, 2006; Johnson *et al.*, 2006; Zinser *et al.*, 2007; Thompson *et al.*, 2021). Over the years since its discovery (Chisholm *et al.*, 1988), environmental sequencing efforts throughout the global ocean have continued to uncover novel genetic diversity within the *Prochlorococcus* collective (Coleman and Chisholm, 2007; Berube *et al.*, 2018; Biller, Berube, *et al.*, 2018).

Within the broadly defined HL and LL-adapted groups of *Prochlorococcus*, monophyletic clades have been designated based on sequence similarity of the 16S/23S rRNA intergenic spacer sequence (ITS) (Rocap *et al.*, 2002; Biller *et al.*, 2015). Laboratory isolates of *Prochlorococcus* have existed since shortly after its discovery (Chisholm *et al.*, 1992; Partensky *et al.*, 1993), and while more than 100 isolates are now available, environmental sequencing continues to reveal groups within the *Prochlorococcus* collective for which no representative isolates exist. These include the HLIII, HLIV, and HLV lineages typically found in high-nutrient, low-chlorophyll environments (Rusch *et al.*, 2010; West *et al.*, 2010; Huang *et al.*, 2011; Malmstrom *et al.*, 2012), the HLVI lineage (Huang *et al.*, 2011), and the basal LLV/AMZ1, LLVI/AMZII, and AMZIII lineages associated with oxygen minimum zones (Lavin *et al.*, 2010; Ulloa *et al.*, 2021). Furthermore, quantitative PCR based abundance measurements typically underestimate the total number of *Prochlorococcus* cells *in-situ* when compared to flow cytometry counts, particularly for samples taken at or near the base of the euphotic zone. Because PCR primers are highly specific and are based on the known diversity of *Prochlorococcus*, this suggests a significant fraction of LL-adapted lineages are not represented in culture collections (Ahlgren *et al.*, 2006; Zinser *et al.*, 2006; Martiny *et al.*, 2009). For some of these uncultivated groups, little is known beyond the fact that they exist, precluding a deeper understanding of the evolutionary history of *Prochlorococcus*.

One particularly abundant, deeply branching uncultivated group originally called NC1 (Martiny *et al.*, 2009) and later renamed to LLVII (Biller *et al.*, 2015) is situated phylogenetically between two broad groups - LLIV and LLII/III - associated with different lifestyles (Figure 1). The transition between these two groups had major implications for the evolution and ecology of ancient *Prochlorococcus*. That is, the more basal LLIV lineage is capable of particle attachment, while the LLII/III clade marks the earliest diverging lineage of fully planktonic cells that characterize the rest of the *Prochlorococcus* tree (Capovilla *et al.*, 2023). The LLVII group was

4

first reported in a 2009 clone library study where they comprised as much as 61.7% (average 23.1%) of the *Prochlorococcus* sequences detected in the lower euphotic zone (140 - 160 m) in the North Atlantic and North Pacific Oceans (Martiny *et al.*, 2009). This uncultivated group has also been detected at and below 100 m in the Red Sea, where they were the majority (58%) of LL-adapted sequences recovered (Shibl *et al.*, 2014) and in the lower euphotic zone of the western Pacific Ocean and South China Sea (Huang *et al.*, 2011; Jiao *et al.*, 2013). Therefore, the physiology and genomic content of the LLVII lineage are particularly intriguing as they could shed light on niche differentiation among LL-adapted groups and macroevolution of the *Prochlorococcus* collective.

The current view of this evolutionary history is that *Prochlorococcus* diverged from its sister lineage, *Synechococcus*, ca. 420-300 million years ago (Fournier *et al.*, 2021; Capovilla *et al.*, 2023). Combining genomic, physiological, metabolic, elemental, and environmental data suggests a macroevolutionary model regarding the diversification of major *Prochlorococcus* lineages (Braakman *et al.*, 2017; Berube *et al.*, 2019; Capovilla *et al.*, 2023) in which the picocyanobacterial ancestor had a mostly particle-attached lifestyle that was lost in the ancestor of all LLVII. We speculate that this LLVII ancestor was committed to a purely planktonic lifestyle, broadly colonizing the water column. Over time adaptive radiations that allowed cells to devote more photosynthetic electron flux to nutrient uptake, effectively drawing surface nutrient levels lower and lower, made surface waters uninhabitable to ancestral lineages (Braakman *et al.*, 2017). This resulted in modern ecotype depth-abundance distributions (Johnson *et al.*, 2006; Zinser *et al.*, 2006; Malmstrom *et al.*, 2010; Thompson *et al.*, 2021), where more deeply diverging ecotypes are differentially more abundant at greater depths, where photosynthetic electron flux is lower and does not allow full drawdown of nutrient levels (Braakman *et al.*, 2017). While understanding deep-time evolution is inherently inferential, studying modern deep-branching lineages has the potential to inform us about the genomic content and physiology of ancient common ancestors, and further constrain possible scenarios.

Here, we study the physiology and genomics of deep-branching *Prochlorococcus* lineages and use their properties to refine our synoptic macroevolutionary perspective. We describe a culturing pipeline explicitly targeting the acquisition of deep-branching low-light adapted lineages, resulting in the isolation of four novel strains belonging to basal lineages. The isolates were physiologically characterized in terms of broad niche-determining traits such as light and temperature optima for growth, photosynthetic capacity, and pigment content. We sequenced their genomes and examined genetic diversity related to pigment biosynthesis and photosynthetic antenna genes of the *pcb* family, which are the primary chlorophyll-binding proteins in *Prochlorococcus*. We conclude by merging the physiological properties and genomic information of these strains with the current macroevolutionary model of forces that drove *Prochlorococcus* diversification in ancient oceans.

Results and Discussion

Phylogenetic placement of isolates

Four *Prochlorococcus* strains were isolated from below the deep chlorophyll maximum and 1% light level (Table 1) using cultivation methods designed to enrich LL-adapted cells. The four strains proved difficult to identify using standard sequence alignment of their ITS regions; they had < 70% nucleotide identity to strains from established clades and each other (Figure S1). These poor alignments precluded the generation of ITS-based phylogenetic trees with reliable bootstrap values, thus we moved to sequence alignments generated using 44 marker genes with conserved synteny and compared them with 1200 non-identical *Prochlorococcus* isolate genomes, SAGs, and MAGs. The four isolates belong to two distinct paraphyletic regions lacking any cultivated representatives (Figure 1). One of these regions, encompassing strains MIT1300xe, MIT1307xe and MIT1341xe (where an xe suffix designates xenic cultures, and all other cultures are axenic), falls between the LLII/III and LLIV clades and includes genomes with ITS sequences that resemble those of the NC1 group (later renamed LLVII) discovered in environmental clone libraries (Martiny *et al.*, 2009; Biller *et al.*, 2015). Our isolates, along with SAGs from this group (Berube *et al.*, 2018), confirm hypotheses that these lineages are not monophyletic (Martiny *et al.*, 2009; Huang *et al.*, 2011). The fourth isolate (MIT1223) resides in a paraphyletic region located between the LLI and LLII/III clades.

The fact that all four of these isolates were obtained from the base of the euphotic zone and maintained under low irradiances in the laboratory, along with their phylogenetic placement with previously identified LL-adapted clades, supports hypotheses that these lineages are LL-adapted (Martiny *et al.*, 2009; Shibl *et al.*, 2014). Here, we refer to the paraphyletic region between the LLII/III and LLIV clades (encompassing strains MIT1300xe, MIT1307xe and MIT1341xe) as the LLVII grade and the paraphyletic region between the LLI and LLII/III clades (encompassing strain MIT1223) as the LLVIII grade (Figure 1).

To further understand these novel groups, we analyzed pigment content and growth rate effects of light and temperature on one isolate from each grade alongside representative HL and LL-adapted strains from previously described monophyletic clades. Strains MIT1314 (HLII) and MIT1327 (LLIV) were chosen to represent two extremes of the *Prochlorococcus* collective life-history spectrum. MIT1327 (LLIV) represents a deep-branching low-light adapted strategy that can be particle-associated and thrives in the nutrient rich conditions of the lower water-column (Capovilla *et al.*, 2023), while MIT1314 (HLII) represents a recently diverging high-light adapted strategy with a streamlined genome that is common in the oligotrophic open ocean (Moore and Chisholm, 1999). To understand how our novel isolates, which are phylogenetically placed between these two representative strains, lie between the two life-history extremes, we focused our experiments on MIT1223 from the LLVIII grade and MIT1300xe as a representative of the LLVII grade.

Photophysiology

Light-dependent growth rates

Growth rates as a function of light intensity were similar for MIT1223, MIT1300xe, and MIT1327 (LLIV clade), but distinct from MIT1314 (HLII clade) - supporting the notion that the paraphyletic LLVII and LLVIII grades comprise LL-adapted lineages (Figure 2A). MIT1223, MIT1300xe, and MIT1327 grew at nearly maximal rates at irradiance levels that were too low to support the growth of MIT1314. Conversely, MIT1314 grew well at the highest irradiance (200 μmol photons m^{-2} s^{-1}), while the other strains did not grow reliably at intensities ≥ 73 μmol photons m^{-2} s^{-1} (Table 2A).

MIT1300xe (LLVII grade) sustained growth at extremely low light intensities, with a growth rate 1.6 times faster than MIT1327 (LLIV clade) - the only other strain capable of growth at 1.5 μmol photons m^{-2} s^{-1} - reflecting MIT1300xe's ability to thrive at the base of the euphotic zone where light limits primary production. MIT1300xe also grew well over a broad range of light intensities (1.5 to 65 μmol photons m^{-2} s^{-1}) consistently faster than the other LL-adapted strains (Figure 2A). We note, however, that MIT1300xe was the only xenic strain tested and we cannot rule out heterotrophic cells influencing these results due to *Prochlorococcus* mixotrophy (Wu *et al.*, 2022) and reduction of oxidative stress (Morris *et al.*, 2008; Sher *et al.*, 2011; Coe *et al.*, 2016; Ma *et al.*, 2017). Both MIT1300xe and MIT1223 grew better at higher light intensities than MIT1327, suggesting that members of the LLVII and LLVIII grades may tolerate higher irradiances than those of the LLIV clade. Interestingly, despite originating from depths that typically experience < 5 μmol photons m^{-2} s^{-1}, all four strains achieved their maximum growth rates at similar light intensities (13.9 - 33.8 μmol photons m^{-2} s^{-1}). However, the maximal growth rate of MIT1314 (HLII clade) was 1.7 times faster than the maximal growth rate of any LL-adapted strain (Table 2A). These results reinforce the well-established notion that adaptation to high-light conditions probably occurred only once and relatively late during *Prochlorococcus* evolution (Kettler *et al.*, 2007).

Photosynthetic properties

Fast repetition rate fluorometry was used to compare photosynthetic efficiencies and the functional size of light-harvesting antennae among strains when grown at various irradiances. Photosynthetic quantum efficiency (F_V/F_M) generally decreased with increasing light intensity in all four strains, with MIT1327 (LLIV clade) exhibiting the steepest decline, especially at light intensities > 8.9 μmol photons m^{-2} s^{-1} (Figure 3A). MIT1300xe (LLVII grade) had the same photosynthetic quantum efficiency when grown at 1.5 to 21.6 μmol photons m^{-2} s^{-1}, but declined at higher intensities, which is in agreement with its consistently high growth rate over this irradiance range when compared to other LL-adapted strains. Interestingly, MIT1300xe is the only strain tested that displayed maximum photosynthetic efficiency when growing at its maximum rate, likely due to the presence of heterotrophs reducing oxidative stress. MIT1327 (LLIV clade) was the most photosynthetic efficient strain at low light intensities, while

7

MIT1300xe (LLVII grade) was the most photosynthetic efficient strain at the intermediate light intensities. Photosystem II (PSII) functional absorption cross-sections (σPSII) were largest at intermediate light intensities, with a general decline in σPSII values for all strains at light intensities greater than those at which the maximum growth rate was observed, potentially indicating light-induced damage to PSII at higher irradiances (Figure 3B). MIT1300xe had larger σPSII values than the other LL-adapted strains at all light intensities, an advantage that is consistent with its faster growth rates.

The photosynthetic properties of MIT1300xe (LLVII) place it as an intermediate in the *Prochlorococcus* collective life-history spectrum, consistent with its phylogenetic placement between the LLIV clade and other ecotypes (Figure 1). MIT1300xe exhibited the greatest photosynthetic quantum efficiency at intermediate light levels (Figure 3A) and σPSII values between those of the LLIV and HLII strains at all light intensities (Figure 3B). This suggests that the photosynthetic apparatus of MIT1300xe may represent a transitional state between LLIV and later diverging lineages, or that it occupies a niche where these intermediate photosynthetic parameters are advantageous. MIT1223ax (LLVIII) however, exhibited low photosynthetic quantum efficiencies despite σPSII values similar to the LLIV strain, suggesting physiological tradeoffs that limit photosynthetic growth efficiency may be a requirement for *Prochlorococcus* cells with with a relatively small genome size and low GC content (Table 3; (Biller *et al.*, 2014)) to grow at low irradiances (Figure 2A).

Pigment-driven photoacclimation

Prochlorococcus cells tune their pigment content to irradiance (Partensky *et al.*, 1993; Moore *et al.*, 1995), often revealed as an anticorrelation between chlorophyll (red) fluorescence per cell and growth irradiance in flow cytometric signatures. HL and LL-adapted strains are known to display different slopes, where pigment fluorescence decreases more drastically for LL-adapted cells with increasing light intensity (Moore *et al.*, 1995; Moore and Chisholm, 1999). We found this slope to be intermediate for MIT1300xe (LLVII grade), which is similar to previous observations of cells from the LLII/III clade (Figure 4A, see also Moore and Chisholm, 1999). Surprisingly, despite being adapted to grow at low light intensities, MIT1223 (LLVIII grade) appeared similar to MIT1314 (HLII clade) in this regard (Figure 4A). In terms of yellow fluorescence, which is mainly mediated by phycoerythrin in *Prochlorococcus*, the HLII strain (MIT1314), which lacks phycoerythrin biosynthetic potential (Hess *et al.*, 1999), represents the lowest extreme, while the LLIV strain (MIT1327) represents the highest. The novel LLVII and LLVIII isolates are intermediate in terms of slope and total yellow fluorescence/cell levels, recapitulating their evolutionary location between the basal LLIV and the more recently emerged HL-adapted clades (Figure 4B).

All four strains tended to increase in size with increasing irradiance (Figure 4C). As expected, based on their clade and genome size, MIT1327 (LLIV clade) cells were the largest cells at all light intensities (Cermak *et al.*, 2016). The forward angle scatter per cell of MIT1223

(LLVIII grade) was similar to that of the HLII strain at all irradiance levels (Figure 4C), providing further evidence that MIT1223 does not follow all canonical descriptions of a LL-adapted *Prochlorococcus* cell. These results provide potential clues about the period of evolutionary tinkering that occurred during the divergence of LLIV and other *Prochlorococcus* - suggesting that changes to the physiology and pigments that occurred as cells adapted to a planktonic lifestyle may have occurred in distinct stages.

To explore whether or not the unexpected properties of MIT1223's photophysiology - i.e., chlorophyll (red) fluorescence and forward angle scatter per cell - were unique to the LLVIII grade, we examined *Prochlorococcus* strain NATL2A from the LLI clade (Moore *et al.*, 1998; Berube *et al.*, 2014; Biller *et al.*, 2023) acclimated to 4 irradiance levels that fall within the middle of the range tested for the other strains. NATL2A grew faster than the other LL-adapted strains (MIT1223 and MIT1327) at higher light intensities (≥ 26.3) but slower than the HL-adapted strain (MIT1314) (Figure S2A). This finding supports the notion that LLI clade *Prochlorococcus* can tolerate higher irradiance levels than other LL-adapted strains - perhaps due in part to the presence of a photolyase gene and an abundance of high light-inducible genes (Coleman and Chisholm, 2007; Malmstrom *et al.*, 2010). Despite its unique light-dependent growth rates (Figure S2A), the *in-vivo* absorption spectra, median chlorophyll (red) fluorescence per cell, and median forward angle scatter per cell of NATL2A (LLI clade) were quite similar to those of MIT1223 (LLVIII grade) and MIT1314 (HLII clade), and distinct from MIT1327 (LLIV clade) (Figure S2 B,C,E). The median yellow (primarily phycoerythrin) fluorescence per cell signature of NATL2A was lower than MIT1223 and similar to that of MIT1314 (Figure S2D). While a detailed pigment analysis of *Prochlorococcus* strains from the LLI clade is needed to confirm these absorption and flow cytometry results, our experiments indicate that despite its phylogenetic placement among LL-adapted clades, NATL2A, like MIT1223, possesses many physiological characteristics that are traditionally associated with HL-adapted strains (e.g. lower red and yellow fluorescence per cell, reduced forward angle light scatter, and likely lower chl $b:a_2$ ratios as indicated by *in-vivo* absorption spectra).

In-vivo absorption spectra provide further evidence that the pigmentation of MIT1223 is unique among the LL-adapted strains. *Prochlorococcus* cells have traditionally been divided into two major groups based on chlorophyll content with HL-adapted cells exhibiting low total chl *b* (chl b_1 + chl b_2) to chl a_2 ratios and LL-adapted cells exhibiting high chl $b:a_2$ ratios (Partensky *et al.*, 1993; Moore *et al.*, 1998; Moore and Chisholm, 1999). Comparing spectra for all strains grown at the same low light intensity the second peak at ca. 480 nm (chlorophyll b_2) is visible for MIT1327 (LLIV) and MIT1300xe (LLVII), with higher absorbance than the peak at ca. 450 nm (chlorophyll a_2) (Figure 5). In contrast, the chl b_2 peak for MIT1314 (HLII) is lower than its chl a_2 peak, matching the canonical definition of HL-adapted strains (Partensky *et al.*, 1993; Moore *et al.*, 1998). The absorption spectra of MIT1223 (LLVIII) is anomalous for a LL-adapted strain, with no discernible chl b_2 peak and a faint peak at ca. 470 nm - likely due to a combination of

zeaxanthin and α/β-carotene (Moore *et al.*, 1995) - providing further support for atypical pigmentation in the LL-adapted MIT1223 strain.

We next analyzed pigments using HPLC and MS, which confirmed the *in-vivo* absorption spectra and provided additional information (Figures S3, S4). Strikingly, the chl b_2 values for MIT1223 were the lowest of the four strains, including HL-adapted MIT1314 (Figure 6A). LL-adapted strains MIT1327 (LLIV) and MIT1300xe (LLVII) contained appreciable amounts of chl b_2 relative to the total measurable pigment content (4 - 39%), especially at lower light intensities, while the relative abundance of chl b_2 in MIT1314 (HLII) was lower at all light intensities where both strains could grow (5 - 16%; Figure 6A).

As expected, the ratio of chl $b:a_2$ decreased with increasing light intensity for all strains. MIT1300xe (LLVII) showed ratios > 1 when grown at low intensities (Figure 6B), similar to previous findings for strains from the LLII and LLIV clades (Partensky *et al.*, 1993; Moore and Chisholm, 1999; Partensky *et al.*, 1999). These high ratios likely facilitate growth at very low irradiances (Moore and Chisholm, 1999). However, the strains tested here still do not have ratios as high as those reported for wild *Prochlorococcus* cells (Goericke and Repeta, 1993), thus the influence of complex heterotrophic partnerships on this ratio may be at play here; the highest chl $b:a_2$ ratios *in-vitro* have been measured for xenic cultures (Moore *et al.*, 1995; Moore and Chisholm, 1999; this study). MIT1300xe also demonstrated a remarkable capacity for photoacclimation with a range of chl $b:a_2$ ratios spanning more than one order of magnitude (Figure 6B). MIT1327 (LLIV) had the next highest chl $b:a_2$ ratios; however, these ratios were < 1, even when growing under the lowest intensities (Figure 6B). As expected from previous studies, MIT1314 (HLII) had lower chl $b:a_2$ ratios than MIT1300xe and MIT1327 at all light levels with less photoacclimation over the range of light levels tested (Figure 6B). The chl $b:a_2$ ratios of MIT1223 (LLVIII) were the lowest of any strain tested - as much as 80 fold lower than MIT1300xe at the same light level. MIT1223 displayed little photoacclimation, and chl $b:a_2$ ratios for this strain dropped below 0.01 at the highest light intensities (Figure 6B). MIT1223 is therefore an exception to the long-standing paradigm that LL-adapted strains contain higher ratios of chl $b:a_2$ than HL-adapted strains.

Ratios of zeaxanthin to chl a_2 generally increased with increasing light intensity, supporting its photoprotective role (Partensky *et al.*, 1993; Moore *et al.*, 1995) (Figure 6C). MIT1327 had the highest total extractable pigment per cell (2.6 - 9.0 fg/cell) compared to the other three strains tested and MIT1223 had values at or below those for MIT1314 when grown at the same irradiance (Figure 6D). Other primary pigments detected in all four strains consisted of α/β-carotene (indistinguishable with our method), pheophytin a_2, Mg 3,8 divinyl pheoporphyrin (DVP) a_5, and an unknown carotenoid (Figures S3, S4, S5), all of which are consistent with previous reports analyzing the pigment content of *Prochlorococcus* cells (Goericke and Repeta, 1992; Partensky *et al.*, 1993). Based on its exact mass, the unknown carotenoid has a sum formula of $C_{40}H_{56}O$, which matches with both cryptoxanthin and β-carotene 5,6-epoxide, however, the MS^2 fragment data does not appear to match published spectra for either

compound (Figure S5). It has been previously suggested that *Prochlorococcus* may contain cryptoxanthin (Goericke and Repeta, 1992), but due to the MS^2 results we obtained, we do not feel comfortable making a definitive identification, and additional information such as nuclear magnetic resonance analysis is needed to elucidate the structure of this carotenoid. We also putatively identified the compound all-*trans*-retinal in all four strains due to an exact match (retention time and mass) to a standard (Figure S6). However, this identification remains ambiguous, given that the abundance was low, and therefore MS^2 spectra could not be obtained. We feel the putative identification of this compound is worth noting, however, as the potential ability of *Prochlorococcus* to produce all-*trans*-retinal could have important implications for syntrophic interactions in the marine environment, given the importance of this molecule for proteorhodopsin-based phototrophy (DeLong and Béjà, 2010; Gómez-Consarnau *et al.*, 2019).

Thermophysiology

Temperature-dependent growth rates also revealed similar trends for MIT1223, MIT1300xe, and MIT1327 - which were quite distinct from that of MIT1314 - suggesting the paraphyletic LLVII and LLVIII regions comprise lineages of *Prochlorococcus* that are adapted to colder conditions found deeper in the water column near the base of the euphotic zone - the waters from which they were isolated (Figure 2B). While all of the LL-adapted strains grew much slower than the HL-adapted strain at temperatures $\geq$ 24 ºC, MIT1300xe was still able to grow at the highest temperature at which the HL-adapted strain was able to grow, 2 ºC warmer than the upper limit of the other LL-adapted strains (Table 2B). Again, we cannot rule out the potential influence of heterotrophic bacterioplankton, but note that MIT1327 (LLIV clade) grew faster than MIT1300xe at temperatures $\leq$ 24 ºC, despite being axenic. Consistent with the light physiology experiments, MIT1223 generally grew slower than the other strains regardless of temperature, but it could tolerate colder conditions than the HL-adapted strain (Table 2B).

Photosynthetic antenna evolution

Genomes of the four isolates ranged from 1.8 - 2.0 Mb in size with 35.7 - 41.4 %GC, intermediate to HL-adapted strains and those from the LLIV clade (Biller *et al.*, 2014) (Table 3). We examined light-harvesting genes to place these new genomes in context and explore a genetic basis for their observed physiological properties. Unlike other cyanobacteria, most *Prochlorococcus* ecotypes do not produce phycobilisomes as light-harvesting antennae, and instead rely on pigment-binding *pcb* proteins to channel photons to their photosystem II core (Roche *et al.*, 1996). This property gives *Prochlorococcus* its distinct fluorescence and absorption profiles and is one of the hallmark adaptations that allowed this group to colonize the entire photic zone (Ting *et al.*, 2002). Because these antenna proteins are the major pigment-containing elements in their photosystems, a plausible hypothesis is that the photophysiological features of these novel LLVII/VIII isolates are due – at least in part – to

11

unique combinations of pigment and antenna. Thus, we focused our genomic comparisons on these features.

In agreement with previous reports (Ulloa *et al.*, 2021), our analysis shows that most picocyanobacteria encode the iron-inducible *isiA* gene, and *Synecococcus* and the basal *Prochlorococcus* groups (AMZII and AMZIII) encode the full suite of phycobilisome genes, but no *pcb* genes. *pcbA,* the first *pcb* gene to diverge from *isiA*, first appears in the ancestor of the LLIV/AMZ-I radiation (Figure 7C). A comparison of the genomes from 1,691 *Synechococcus* and *Prochlorococcus* picocyanobacteria (Methods) reveals a major expansion in *pcb* gene copy number and diversity for the LLVII grade that is conserved in later branching LL-adapted clades, but lost in the ancestor of all HL-adapted cells, which only contain *pcbA* (Figure 7). The different ecotypes of *Prochlorococcus* have distinct repertoires of antenna genes, suggesting the observed genetic changes occurred in the ancestor of the group, and that selection has been strong enough to retain these changes.

Examining the evolution of individual antenna genes reveals a complicated history of divergence and horizontal transfer of multiple genes (Figures S7, S8). In some cases (such as *pcbD*), MIT1223 (LLVIII) shares alleles with LLVII strains, while in others (e.g., *pcbB*), it is closer to LLI strains. Interestingly, MIT1223 never seems to share antenna related alleles with HL strains (Figures S7, S8), suggesting that its "HL-like" characteristics are due to other genetic determinants such as regulation of pigment biosynthesis.

These findings give further support for the important role photophysiology played in the evolution of *Prochlorococcus* (Biller *et al.*, 2015), in that major radiations involved significant changes to the configuration of the photosynthetic machinery – more specifically in the light-harvesting apparatus. This began with the complete loss of phycobilisomes in the ancestor of LLIV/AMZI, replaced by the more streamlined *pcb* apparatus (Figure 7, diagram and total number of proteins needed by each antenna type). This transition was capitalized upon by later-diverging lineages, showing remarkable variation in the diversity and copy-number of *pcb* and pigment biosynthesis genes.

Macroevolutionary implications

Our current understanding of the evolution of *Prochlorococcus* is built upon the notion of an ancestor that once thrived at all depths of the sunlit ocean (Braakman *et al.*, 2017). Evolutionary radiations followed, and different lineages emerged with greater relative fitness in shallower waters due to smaller cell size, more streamlined genomes, reconfigured central metabolism, and more optimal elemental compositions. This led to the drawing down of nutrient levels in shallow waters and a displacement of ancestral lineages to deeper waters (Braakman *et al.*, 2017). Supporting this view is the distribution of the nitrate reductase gene *narB* in picocyanobacteria, as *narB* is vertically inherited and present in marine *Synechococcus*, LLI, and HL ecotypes, but absent in basal *Prochlorococcus* lineages (Berube *et al.*, 2019). This is consistent with the idea that the *Prochlorococcus* ancestor had a broad distribution in the water column and encoded

12

narB, which was lost in basal lineages as they were displaced to deeper waters where it was no longer advantageous due to lower light levels and higher levels of other nitrogen sources. Recently, an extension of this model was proposed based on the surprising finding that the basal lineages encode chitinases and that LLIV isolates express functional chitinases and attach to, and use, chitin (Capovilla *et al.*, 2023). Under this extension, the *Prochlorococcus* ancestor was particle-associated, and this trait was retained in basal lineages because they still led a facultative particle-associated lifestyle at the base of the euphotic zone, as evident from their enzyme activity and ability to attach to particles and form aggregates. Later radiations, such as the LLVII radiation then fully committed to the planktonic lifestyle and their diversification followed the "top-down" model described above and argued in detail in (Braakman *et al.*, 2017).

Our analysis here provides further independent support for the extended model in Capovilla *et al.*, 2023. First, it establishes that LLVII isolates are indeed LL-adapted. Second, the massive expansion in *pcb* gene copy number and allelic diversity in the ancestor of the LLVII grade are consistent with the idea that the cells in this radiation switched to a fully planktonic lifestyle. A commitment to the planktonic state, where particle-derived carbon inputs are reduced due to loss of attachment capacities would entail a higher dependence on photosynthetically derived energy and carbon, exerting a strong selective pressure to optimize light-gathering (Capovilla *et al.*, 2023). This is supported by the following observations regarding the physiological effect of Pcb proteins in low-light conditions. Different Pcb proteins are known to associate with both photosystems and are the major carriers of chlorophyll in photosystem II, enhancing its light-gathering capacity considerably (Bibby *et al.*, 2003). In *Prochlorococcus* strains that carry multiple *pcb* copies, total Pcb levels are higher compared to strains with fewer copies (MacGregor-Chatwin *et al.*, 2019), and at least some copies are differentially regulated and are expressed in response to different nutrient limitation conditions. This differential regulation possibly provides specific properties such as variations in trace-element utilization that are beneficial under distinct nutrient limiting conditions (Bibby *et al.*, 2003). The distinct physiology and unexpected pigmentation of the LLVIII isolate provide an additional opportunity to investigate the evolutionary journey of *Prochlorococcus* from particle attachment to a free-living lifestyle with greater autotrophic dependence and eventual ability to colonize the upper euphotic zone.

Conclusions

The four isolates described here help resolve two diverse, largely unexplored regions within the LL-adapted ecotypes of the *Prochlorococcus* collective. The isolates belong to two distinct paraphyletic regions in an area of the *Prochlorococcus* phylogeny with long branch lengths (i.e. the paraphyletic LLVII and LLVIII regions, along with the LLII/III clade). The area is flanked by three monophyletic clades (LLI, AMZI, and LLIV) containing lineages with shorter branch lengths. The inclusion of single cell genomes and metagenomes from field samples provides culture-independent evidence supporting these placements.

A representative from the LLVII grade displayed photophysiology properties and genomic characteristics that were generally intermediate to representatives from the LLIV and HLII clades - two extremes of the *Prochlorococcus* collective life-history spectrum. A representative from the LLVIII grade displayed many photophysiology properties indicative of a LL-adapted strain, however its pigmentation and some genomic characteristics resemble that of a HL-adapted strain, challenging several longstanding paradigms regarding light-adaptation among the *Prochlorococcus* collective. Preliminary work with strain NATL2A (LLI) indicates these anomalies may also extend to members of the LLI clade. Our findings blur the distinction between these historically defined ecotypes and suggest pigmentation and pigment driven photoacclimation are likely unreliable indicators of *Prochlorococcus'* optimal irradiance levels for growth and ecotype designation.

  Prochlorococcus lineages from regions of the phylogenetic tree with longer branch lengths - and therefore higher genetic divergence from nearby lineages - may reflect divergent evolution among LL-adapted *Prochlorococcus* in which a confluence of environmental variables, including light, has driven substantial diversification from ancestral strains. This is in contrast to the more rapid adaptive radiation events observed within the monophyletic LLI and LLIV clades, which may be a product of finer-scale niche-partitioning among lineages. As the LLVII and LLVIII grades are further examined in the context of evolutionary transitions and the rise of *Prochlorococcus*, the high diversity and long branch lengths contained within these groups may reflect prolonged periods of evolutionary tinkering that occurred during those transitions. Future investigations into the unique photophysiological features we have identified in individual strains, and the genomic and genetic diversity within the grades as a whole, could thus provide important new clues as to how *Prochlorococcus* evolved from its chitin particle-attached ancestors to the planktonic cells that dominate the extant oceans (Braakman *et al.*, 2017; Capovilla *et al.*, 2023).

Experimental Procedures

Isolation and identification of novel *Prochlorococcus* strains
Prochlorococcus strains mentioned in this study (Table 1) were isolated from the lower euphotic zone at Station ALOHA in the subtropical North Pacific Ocean (22.75ºN 158ºW). Strain MIT1223 was isolated from 175 m on September 12, 2012, during the HOE-DYLAN cruise, while all other strains were isolated from 150 m on June 2, 2013, during the HOE-PhoR I cruise. The presence and complexity of *Prochlorococcus* populations were monitored over time by flow cytometry (BD/Cytopeia Influx). The seawater base of all media was sterilized by 0.2 µm filtration followed by autoclaving and prepared in acid-washed autoclaved borosilicate glass or polycarbonate tubes.

  MIT1223 was derived from an initial enrichment of unfiltered raw seawater amended with 16 µM $NaNO_3$, 1 µM $NaH_2PO_4 \cdot H_2O$, and the trace metal mix used in Pro99 medium (Moore

14

et al., 2007) reduced 10-fold. The enrichment was initially maintained for ca. 1 year by serial passage at 22 - 24 ºC and ≤ 8 µmol photons m^{-2} s^{-1} on a 14:10 h light:dark cycle, while inorganic nutrient conditions were increased to 80 µM NaNO$_3$ and 5 µM NaH$_2$PO$_4 \cdot$H$_2$O before acclimating to a modified Pro99 medium (800 µM NH$_4$Cl replaced by 800 µM NaNO$_3$) and transitioning to higher irradiance - 19 ± 2 µmol photons m^{-2} s^{-1} on a 14:10 h light:dark cycle. Several months later, flow cytometry monitoring revealed the presence of larger picoeukaryote-like cell populations in addition to several *Prochlorococcus* populations. The enrichment was subsequently passed through a 0.8 µm polycarbonate filter (Nucleopore, Whatman/GE) by gentle gravity filtration to remove the larger cells. This enrichment was then maintained by serial passage for another 15 mo before flow cytometry monitoring revealed an apparent unialgal *Prochlorococcus* population, and sequencing revealed the ITS sequence of what we now refer to as strain MIT1223, a member of the paraphyletic LLVIII grade.

Enrichments for *Prochlorococcus* conducted on the HOE-PhoR I cruise in 2013 were prepared as previously described (Cubillos-Ruiz *et al.*, 2017; Becker *et al.*, 2019). MIT1314 and MIT1327 are derived from an enrichment amended at sea with Pro2 medium nutrients (Moore *et al.*, 2007) plus 1 µM thiosulfate, while MIT1300xe, MIT1307xe, and MIT1341xe are derived from an enrichment from the same original 1.0 µm filtered water sample amended with nitrite as the sole exogenous nitrogen source (Table 1). These enrichments and their derivatives from serial passage were maintained at 22 - 25 ºC on a 14:10 h light:dark cycle for 3 mo before transitioning to continuous light. Enrichments were maintained at very low irradiances to select for LL-adapted cells, with maximum light exposure of ca. 1 µmol photons m^{-2} s^{-1}.

After an initial incubation period of 1 mo, the enrichment that ultimately yielded MIT1314 and MIT1327 was transferred into Pro2 medium plus 1 µM thiosulfate to match the original amendment based on Hawaii surface seawater collected during HOE PhoR I. Two weeks later, a subculture was transitioned to 0.2µm filtered and autoclaved Sargasso seawater amended with Pro99 nutrients (Moore *et al.*, 2007). Enrichments were maintained by serial passage in their respective media - MIT1314 derives from the Pro99 line, while MIT1327 derives from the Pro2 line. After 4 mo of serial passage, dilution to extinction in ProMM medium was applied to the Pro2 enrichment, yielding the LLIV clade strain MIT1327 (Cubillos-Ruiz *et al.*, 2017). After 5 mo in the laboratory, a subculture of the Pro99 enrichment was moved to higher light (10 ± 2 µmol photons m^{-2} s^{-1}) to select for a subset of the *Prochlorococcus* diversity observed by flow cytometry. After six months at this irradiance, ITS screening revealed the HLII clade strain MIT1314.

After an initial incubation period of 7 weeks, the enrichment yielding MIT1300xe, MIT1307xe, and MIT1341xe was maintained by serial passage in Pro99 medium (Moore *et al.*, 2007) with a Sargasso seawater base. After 4 mo, replicate subcultures were established and maintained as separate transfer series, with one ultimately resulting in MIT1307xe and the other resulting in MIT1300xe and MIT1341xe. After 1 yr in the laboratory, these enrichments still contained diverse *Prochlorococcus* populations based on flow cytometry and ITS PCR

sequencing. Once again, a subculture was transitioned to higher light (10 ± 2 µmol photons m^{-2} s^{-1}), and after 1 yr, ITS screening revealed the sequence of what we now refer to as strain MIT1300xe. Flow cytometry monitoring of the two enrichments maintained below 1 µmol photons m^{-2} s^{-1} revealed the presence of LLIV-like *Prochlorococcus* populations and additional *Prochlorococcus* populations with distinct fluorescence signatures. These enrichments were then passed through a 0.8 µm polycarbonate filter (Nucleopore, Whatman/GE) by gentle gravity filtration in an attempt to remove the larger LLIV-like cells. After several months, flow cytometry monitoring revealed an apparent unialgal *Prochlorococcus* population in each, and ITS screening revealed the sequence of what we now refer to as strains MIT1307xe and MIT1341xe.

MIT1314, MIT223, and MIT1327 were purified using a high-throughput dilution to extinction method, and the purity of these isolates was confirmed by flow cytometry and a suite of purity test broths (ProAC, ProMM, and MPTB) as previously described (Berube *et al.*, 2014; Cubillos-Ruiz *et al.*, 2017; Becker *et al.*, 2019). Dilution to extinction purification was repeatedly attempted on the other strains without success.

Initial strain identification was performed by PCR screening of the ITS gene. In brief, 1 ml of culture (ca. 10^6 to 10^8 cells) was centrifuged for 15 to 30 min at 16,000 x g to form a pellet. The majority of supernatant was removed before centrifugation again at 16,000 x g and removal of all residual seawater before resuspending the pellet in 25 to 100 µl Tris-HCl (pH 8.0). Cells were then lysed by boiling at 95 °C for 10 min before centrifugation for 5 min at 16,000 x g at 4 °C to pellet cell debris. The supernatant was removed and used directly in a PCR screen with primers (ITS-F: 5'-CCGAAGTCGTTACTYYAACCC-3' and ITS-R: 5'-TCATCGCCTCTGTGTGCC-3'), which target the internal transcribed spacer (ITS) region of *Prochlorococcus* as previously described (Rodrigue *et al.*, 2009; Berube *et al.*, 2015).

Growth rate experiments

The physiological response of four *Prochlorococcus* isolates (MIT1314, MIT1223, MIT1300xe, and MIT1327) to light availability and temperature was measured in a series of growth experiments. All cultures were maintained in Pro99 medium with a Sargasso surface seawater base in acid-washed autoclaved borosilicate glass tubes.

For light-dependent growth experiments (except NATL2Aax), duplicate tubes of each strain were acclimated to target light levels under continuous light and monitored weekly using a light meter (LI-COR LI-250A) connected to a spherical micro quantum sensor (Walz US-SQS/L). 95% of irradiance measurements were within 10% of the target light level for strains MIT1223, MIT1300xe, and MIT1327 and within 16% of the target light level for MIT1314 (Table S1;Figure S9). Temperature was monitored daily and held stable at 24 ± 1 °C. Balanced growth was confirmed by examining growth rates over time at each condition. Attempts were made to acclimate each strain to increasingly higher and lower light intensities until reproducible growth could not be achieved. Growth was monitored daily by flow cytometry (see below) and

16

bulk chlorophyll fluorescence (10AU model, Turner Designs) for 4 - 6 days. Bulk chlorophyll fluorescence readings were blank subtracted using the mean of triplicate medium blanks, and any readings corresponding to lag phase or early onset of stationary phase were identified by eye and removed prior to calculating mean growth rates (Table 2A). For light-dependent growth experiments with NATL2Aax, triplicate tubes were acclimated to 4 target light levels (10, 20, 45, and 96 ± 0.5 µmol photons m^{-2} s^{-1}) on a 14:10 h light:dark cycle. Growth was monitored daily by flow cytometry (see below) and bulk chlorophyll fluorescence (10AU model, Turner Designs) for 8 - 16 days, and mean growth rates were calculated from blank subtracted fluorescence readings taken in exponential phase.

For temperature-dependent growth experiments, duplicate tubes of each strain were acclimated to specific temperatures ($\pm 0.5°C$) and grown at (76 ± 1 µmol photons m^{-2} s^{-1} for MIT1314; 20 ± 1 µmol photons m^{-2} s^{-1} for MIT1223, MIT1300xe, and MIT1327) on a 14:10 h light:dark cycle. Acclimation was defined as described above, and attempts were made to acclimate each strain to increasingly higher and lower temperatures until reproducible growth could not be achieved. Growth was monitored daily, and mean growth rates were calculated as described above (Table 2B).

Flow cytometry and fast repetition rate fluorometry

Cell concentrations were determined using a Guava Technologies easyCyte 12HT flow cytometer (EMD Millipore). Samples were diluted in sterilized Sargasso seawater to ensure <500 cells $µl^{-1}$ to avoid coincidence counting and were run with only the blue (488 nm) excitation laser enabled for maximum power. Technical duplicate *Prochlorococcus* populations were resolved for each biological sample based on their red (695/50 nm) emission parameters for enumeration. A fluorescent bead reagent (easyCheck Beads, EMD Millipore) was run daily as a reference, and to verify instrument performance. Median forward angle scatter, chlorophyll (695/50 nm), and yellow (583/26 nm) fluorescence per cell were determined for each population.

The maximum quantum efficiency of photochemistry in PSII (Fv/Fm) and the functional absorption cross-section of PSII (σPSII) were measured using fast repetition rate fluorometry (FRRF) on a FIRe fluorometer instrument (Satlantic) as previously described (Biller, Coe, *et al.*, 2018). Samples were stored in the dark for 15 min prior to loading into the instrument and Pro99 medium was used for blank measurements.

Absorption spectra and pigment measurements

In-vivo absorption spectra were obtained using a Beckman DU 800 spectrophotometer (Beckman Coulter Inc.) in absorbance scan mode from 400 to 700 nm at 1.0 nm intervals and a scan speed of 1200 nm/min and Pro99 medium was used for blank measurements.

For pigment sampling, duplicate *Prochlorococcus* cultures, as well as technical duplicates from each biological replicate, were filtered using vacuum filtration (ca. -200 mm Hg) onto 47 mm diameter 0.2 µm hydrophilic Durapore filters (Millipore). Samples were immediately

flash-frozen and stored in liquid nitrogen (-196 °C) until processing. Pigments were extracted using a modified Bligh and Dyer protocol (Popendorf *et al.*, 2013) with DNP-PE-$C_{16:0}$/$C_{16:0}$-DAG (2,4-dinitrophenyl phosphatidylethanolamine diacylglycerol; Avanti Polar Lipids, Inc., Alabaster, AL) used as an internal standard. Filter blanks and *Prochlorococcus* growth media blanks were extracted and analyzed alongside samples. The total lipid extract was analyzed by reverse-phase high-performance liquid chromatography (HPLC) mass spectrometry (MS) on an Agilent 1200 HPLC coupled to a Thermo Fisher Exactive Plus Orbitrap high-resolution mass spectrometer (ThermoFisher Scientific, Waltham, MA, USA). HPLC and MS conditions were as previously described (Hummel *et al.*, 2011; Collins *et al.*, 2016). In brief: 20 μl were injected onto a C8 Xbridge HPLC column (particle size 5 μm, length 150 mm, width 2.1 mm; Waters Corp., Milford, MA, USA). Lipids were eluted at a flow rate of 0.4 ml min^{-1} using the following gradient with eluent A (water with 1% 1M ammonium acetate and 0.1% acetic acid) and eluent B (70% acetonitrile, 30% isopropanol with 1% 1M ammonium acetate and 0.1% acetic acid): 45% A was held for 1 min, from 45% A to 35% A in 4 min, from 25% A to 11% A in 8 min, from 11% A to 1% A in 3 min with an isocratic hold until 30 min. Finally, the column was equilibrated with 45% A for 10 min. ESI source settings were: Spray voltage, 4.5 kV (+), 3.0 kV (-); capillary temperature, 150 ℃; sheath gas and auxiliary gas, both 21 (arbitrary units); heated ESI probe temperature, 350 °C. Mass data were collected in full scan while alternating between positive and negative ion modes. Pigments were identified using retention time, accurate molecular mass, and isotope pattern matching of proposed sum formulas in full-scan mode and tandem MS (MS^2) fragment spectra of representative compounds. For each MS full scan, up to three MS^2 experiments targeted the most abundant ions with N2 as collision gas. The scan range for all modes was 100-1500 *m/z*. The mass spectrometer was set to a resolving power of 140,000 (FWHM at *m/z* 200), leading to an observed resolution of 75,100 at *m/z* 875.5505 of our internal standard, DNP-PE. Exact mass calibration was performed by weekly infusing a tune mixture. Additionally, every spectrum was corrected using a lock mass, providing real-time calibrations. To validate the accuracy and reliability quantification, quality control samples of known composition spiked with lipid standards were interspersed with the samples as described previously (Collins *et al.*, 2016). Pigment abundances were corrected for the relative response of commercially available standards. The abundances of chlorophylls and their associated compounds were corrected for the response of a chlorophyll *a* standard and carotenoid pigments using a *β*-carotene standard. All standards were purchased from Sigma Aldrich (St. Louis, MO, USA). Individual response factors were obtained from external standard curves by triplicate injection of a series of standard mixtures ranging from 0.15 to 40 pmol on column per standard. Our method's use of external standards was validated in a study that compared lipid quantitation against internal, isotope-labeled standards (Becker *et al.*, 2018). Data were corrected for differences in extraction efficiency using the recovery of the DNP-PE internal standard.

Phylogenetic analysis

18

Maximum likelihood trees were reconstructed using FastTree (Price *et al.*, 2010) from 44 concatenated single-copy core proteins that have conserved local synteny (see below) in 1691 non-redundant Picocyanobacterial genomes (1201 *Prochlorococcus* and 490 *Synechococcus*). 4512 genomes matching either *Prochlorococcus* or *Synechococcus* in their description were downloaded from NCBI and IMG on December 7 2022. To filter out very closely related sequences, we used FastANI (Jain *et al.*, 2018) to calculate all pairwise whole-genome average nucleotide identity (ANI) values. From these results a network was created where nodes represent genomes and links represent >99.99% identity over 95% of the longer genome in the pair. This network had 1961 connected components, and the longest genome from each component was picked as the representative, resulting in the final genome set used in downstream analysis. Gene calling was performed using Pyrodigal v2.0.2 (Larralde, 2022) and protein sequences were clustered using MMseqs2 v14-7e284 (Steinegger and Söding, 2017) based on 50% identity and 80% coverage. Initial protein homology clusters were then split based on their genomic surroundings into groups that share at least 3 of the 10 protein families in their immediate genomic vicinity. The final protein clusters therefore share both homology and local synteny. We found 44 final clusters that are present in >50% of analyzed genomes, and calculated the protein alignment of each protein cluster using mafft v7.310 (Katoh and Standley, 2013). We then transformed the protein alignments to nucleotide alignments and concatenated all individual alignments to one long alignment that was used for phylogenetic inference with FastTree v2.1.10 (Price *et al.*, 2010).

Genome sequencing and analysis

Closed genome sequences of MIT1223, MIT1300xe, MIT1307xe, and MIT1341xe were obtained as previously described for MIT1314 (Becker *et al.*, 2019). In brief, genomic DNA was extracted from concentrated cells using the MasterPure complete DNA and RNA purification kit (Epicentre) and sequenced using P6 chemistry on a RS II instrument (Pacific Biosystems). Library construction and sequencing were performed at the University of Massachusetts Medical School Deep Sequencing and Molecular Biology Core Laboratory. The average insert range for libraries constructed using a standard SMRTbell Template Prep Kit 1.0 and Sequencing Primer v3 (Pacific Biosciences) was 21 − 28 kb. Reads were assembled *de novo* into a single contig for each strain using the RS_HGAP_Assembly.2 protocol within the SMRT Analysis software (v2.0; Pacific Biosciences) before circularization using the Geneious sequence analysis package (V7.1, Biomatters) and polishing using Quiver and the RS_Resequencing.1 protocol within the SMRT Portal software. The average coverage for each assembly ranged from 73X to 335X (Table 3). The draft genome of MIT1327 was obtained as previously described (Cubillos-Ruiz *et al.*, 2017) and consists of 29 scaffolds with a N50 of 328.4 kbp and an average coverage of 204X (Table 3). All genomes have been deposited in GenBank at the National Center for Biotechnology Information (NCBI) and the Joint Genome Institute's Integrated Microbial Genomes (IMG) system (see Table 3 for IDs) and annotated using the IMG Annotation Pipeline version 4

(Markowitz *et al.*, 2013; Huntemann *et al.*, 2016). Sequenced isolates were included in the genome collection described above.

The proteomes of all genomes were annotated using the eggnog v6 database (Hernández-Plaza *et al.*, 2022, 6). Input proteins were competitively aligned to the database using MMseqs2 with a minimum identity of 30% and 50% coverage. The annotations of the best annotated hit of each input protein was then transferred to the input protein. For the analysis in figure 7, antenna proteins in the "Photosynthesis Proteins" kegg pathway (ko00194) were extracted based on their eggnog-derived KO numbers, and their presence/absence in genomes was tabulated and plotted on the phylogenetic tree. For each KO the phylogenetic tree was built by aligning KO members using mafft and building a tree using FastTree as above. Gene trees were rooted using the minimal ancestor deviation criterion (Tria *et al.*, 2017).

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Ahlgren, N.A., Rocap, G., and Chisholm, S.W. (2006) Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environ Microbiol* **8**: 441–454.

Becker, J.W., Hogle, S.L., Rosendo, K., and Chisholm, S.W. (2019) Co-culture and biogeography of *Prochlorococcus* and SAR11. *ISME J* 1–14.

Becker, K.W., Collins, J.R., Durham, B.P., Groussman, R.D., White, A.E., Fredricks, H.F., et al. (2018) Daily changes in phytoplankton lipidomes reveal mechanisms of energy storage in the open ocean. *Nat Commun* **9**: 5179.

Berube, P.M., Biller, S.J., Hackl, T., Hogle, S.L., Satinsky, B.M., Becker, J.W., et al. (2018) Single cell genomes of *Prochlorococcus*, *Synechococcus*, and sympatric microbes from diverse marine environments. *Sci Data* **5**: 180154.

Berube, P.M., Biller, S.J., Kent, A.G., Berta-Thompson, J.W., Roggensack, S.E., Roache-Johnson, K.H., et al. (2014) Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *ISME J* **9**: 1195–1207.

Berube, P.M., Coe, A., Roggensack, S.E., and Chisholm, S.W. (2015) Temporal dynamics of *Prochlorococcus* cells with the potential for nitrate assimilation in the subtropical Atlantic and Pacific oceans. *Limnol Oceanogr* **61**: 482–495.

Berube, P.M., Rasmussen, A., Braakman, R., Stepanauskas, R., and Chisholm, S.W. (2019) Emergence of trait variability through the lens of nitrogen assimilation in *Prochlorococcus*. *Elife* **8**:.

Bibby, T.S., Mary, I., Nield, J., Partensky, F., and Barber, J. (2003) Low-light-adapted *Prochlorococcus* species possess specific antennae for each photosystem. *Nature* **424**: 1051–1054.

Biller, S.J., Berube, P.M., Berta-Thompson, J.W., Kelly, L., Roggensack, S.E., Awad, L., et al. (2014) Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*. *Sci Data* **1**: 140034.

Biller, S.J., Berube, P.M., Dooley, K., Williams, M., Satinsky, B., Hackl, T., et al. (2018) Marine microbial metagenomes sampled across space and time. *Sci Data* **5**: 180176.

Biller, S.J., Berube, P.M., Lindell, D., and Chisholm, S.W. (2015) *Prochlorococcus*: the structure and function of collective diversity. *Nat Rev Micro* 1–15.

Biller, S.J., Coe, A., Arellano, A.A., Dooley, K., Silvestri, S.M., Gong, J.S., et al. (2023) Environmental and Taxonomic Drivers of Bacterial Extracellular Vesicle Production in Marine Ecosystems. *Appl Environ Microbiol* **89**: e00594-23.

Biller, S.J., Coe, A., Roggensack, S.E., and Chisholm, S.W. (2018) Heterotroph interactions alter *Prochlorococcus* transcriptome dynamics during extended periods of darkness. *mSystems* **3**: 782–18.

Bouman, H.A., Ulloa, O., Scanlan, D.J., Zwirglmaier, K., Li, W.K.W., Platt, T., et al. (2006) Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* **312**: 918–921.

Braakman, R., Follows, M.J., and Chisholm, S.W. (2017) Metabolic evolution and the self-organization of ecosystems. *Proceedings of the National Academy of Sciences* **114**: E3091–E3100.

Capovilla, G., Braakman, R., Fournier, G.P., Hackl, T., Schwartzman, J., Lu, X., et al. (2023) Chitin utilization by marine picocyanobacteria and the evolution of a planktonic lifestyle. *Proc Natl Acad Sci USA* **120**: e2213271120.

Cermak, N., Becker, J.W., Knudsen, S.M., Chisholm, S.W., Manalis, S.R., and Polz, M.F. (2016) Direct single-cell biomass estimates for marine bacteria via Archimedes' principle. *ISME J* **11**: 825–828.

Chisholm, S.W., Frankel, S.L., Goericke, R., Olson, R.J., Palenik, B., Waterbury, J.B., et al. (1992) *Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and *b*. *Arch Microbiol* **157**: 297–300.

Chisholm, S.W., Olson, R.J., Zettler, E.R., Goericke, R., Waterbury, J.B., and Welschmeyer, N.A. (1988) A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**: 340–343.

Coe, A., Ghizzoni, J., LeGault, K., Biller, S., Roggensack, S.E., and Chisholm, S.W. (2016) Survival of *Prochlorococcus* in extended darkness. *Limnol Oceanogr* **61**: 1375–1388.

Coleman, M.L. and Chisholm, S.W. (2007) Code and context: *Prochlorococcus* as a model for cross-scale biology. *Trends in Microbiology* **15**: 398–407.

Collins, J.R., Edwards, B.R., Fredricks, H.F., and Van Mooy, B.A.S. (2016) LOBSTAHS: An Adduct-Based Lipidomics Strategy for Discovery and Identification of Oxidative Stress Biomarkers. *Anal Chem* **88**: 7154–7162.

Cubillos-Ruiz, A., Berta-Thompson, J.W., Becker, J.W., van der Donk, W.A., and Chisholm, S.W. (2017) Evolutionary radiation of lanthipeptides in marine cyanobacteria. *Proc Natl Acad Sci USA* **114**: E5424–E5433.

DeLong, E.F. and Béjà, O. (2010) The Light-Driven Proton Pump Proteorhodopsin Enhances Bacterial Survival during Tough Times. *PLoS Biol* **8**: e1000359.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., et al. (2013) Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci USA* **110**: 9824–9829.

Fournier, G.P., Moore, K.R., Rangel, L.T., Payette, J.G., Momper, L., and Bosak, T. (2021) The Archean origin of oxygenic photosynthesis and extant cyanobacterial lineages. *Proceedings of the Royal Society B: Biological Sciences* **288**: 20210675.

Gómez-Consarnau, L., Raven, J.A., Levine, N.M., Cutter, L.S., Wang, D., Seegers, B., et al.

(2019) Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Sci Adv* **5**: eaaw8855.

Hernández-Plaza, A., Szklarczyk, D., Botas, J., Cantalapiedra, C.P., Giner-Lamia, J., Mende, D.R., et al. (2022) eggNOG 6.0: enabling comparative genomics across 12 535 organisms. *Nucleic Acids Research* gkac1022.

Hess, W.R., Steglich, C., Lichtle, C., and Partensky, F. (1999) Phycoerythrins of the oxyphotobacterium *Prochlorococcus marinus* are associated to the thylakoid membrane and are encoded by a single large gene cluster. *Plant molecular biology* **40**: 507–521.

Huang, S., Wilhelm, S.W., Harvey, H.R., Taylor, K., Jiao, N., and Chen, F. (2011) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* **6**: 285–297.

Hummel, J., Segu, S., Li, Y., Irgang, S., Jueppner, J., and Giavalisco, P. (2011) Ultra Performance Liquid Chromatography and High Resolution Mass Spectrometry for the Analysis of Plant Lipids. *Front Plant Sci* **2**:.

Huntemann, M., Ivanova, N.N., Mavromatis, K., Tripp, H.J., Paez-Espino, D., Palaniappan, K., et al. (2016) The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Standards in Genomic Sciences* **10**: 86.

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**: 1–8.

Jiao, N., Luo, T., Zhang, R., Yan, W., Lin, Y., Johnson, Z.I., et al. (2013) Presence of *Prochlorococcus* in the aphotic waters of the western Pacific Ocean. *Biogeosciences Discuss* **10**: 9345–9371.

Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Woodward, E.M.S., and Chisholm, S.W. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**: 1737–1740.

Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., et al. (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344**: 416–420.

Katoh, K. and Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**: 772–780.

Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., et al. (2007) Patterns and Implications of Gene Gain and Loss in the Evolution of *Prochlorococcus*. *PLoS Genetics* **3**: 2515–2528.

Larralde, M. (2022) Pyrodigal: Python bindings and interface to Prodigal,an efficient method for gene prediction in prokaryotes. *JOSS* **7**: 4296.

Lavin, P., González, B., Santibáñez, J.F., Scanlan, D.J., and Ulloa, O. (2010) Novel lineages of *Prochlorococcus* thrive within the oxygen minimum zone of the eastern tropical South Pacific. *Environmental Microbiology Reports* **2**: 728–738.

Ma, L., Calfee, B.C., Morris, J.J., Johnson, Z.I., and Zinser, E.R. (2017) Degradation of hydrogen peroxide at the ocean's surface: the influence of the microbial community on the realized thermal niche of *Prochlorococcus*. *ISME J* 1–12.

MacGregor-Chatwin, C., Jackson, P.J., Sener, M., Chidgey, J.W., Hitchcock, A., Qian, P., et al. (2019) Membrane organization of photosystem I complexes in the most abundant phototroph on Earth. *Nat Plants* **5**: 879–889.

23

Malmstrom, R.R., Coe, A., Kettler, G.C., Martiny, A.C., Frias-Lopez, J., Zinser, E.R., and Chisholm, S.W. (2010) Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *ISME J* **4**: 1252–1264.

Malmstrom, R.R., Rodrigue, S., Huang, K.H., Kelly, L., Kern, S.E., Thompson, A., et al. (2012) Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J* **7**: 184–198.

Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., et al. (2013) IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res* **42**: D560–D567.

Martiny, A.C., Tai, A.P.K., Veneziano, D., Primeau, F., and Chisholm, S.W. (2009) Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ Microbiol* **11**: 823–832.

Moore, L.R. and Chisholm, S.W. (1999) Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol Oceanogr* **44**: 628–638.

Moore, L.R., Coe, A., Zinser, E.R., Saito, M.A., Sullivan, M.B., Lindell, D., et al. (2007) Culturing the marine cyanobacterium *Prochlorococcus*. *Limnol Oceanogr Methods* **5**: 353–362.

Moore, L.R., Moore, L.R., Goericke, R., Goericke, R., and Chisholm, S.W. (1995) Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Mar Ecol Prog Ser* **116**: 259–275.

Moore, L.R., Rocap, G., and Chisholm, S.W. (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**: 464–467.

Morris, J.J., Kirkegaard, R., Szul, M.J., Johnson, Z.I., and Zinser, E.R. (2008) Facilitation of robust growth of *Prochlorococcus* colonies and dilute liquid cultures by "helper" heterotrophic bacteria. *Applied and Environmental Microbiology* **74**: 4530–4534.

Partensky, F., Hoepffner, N., Li, W.K.W., Ulloa, O., and Vaulot, D. (1993) Photoacclimation of *Prochlorococcus* sp. (Prochlorophyta) strains isolated from the North Atlantic and the Mediterranean Sea. *Plant Physiol* **101**: 285–296.

Popendorf, K.J., Fredricks, H.F., and Van Mooy, B.A.S. (2013) Molecular Ion-Independent Quantification of Polar Glycerolipid Classes in Marine Plankton Using Triple Quadrupole MS. *Lipids* **48**: 185–195.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**: e9490.

Rocap, G., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Applied and Environmental Microbiology* **68**: 1180–1191.

Roche, J.L., van der Staay, G.W.M., Partensky, F., Ducret, A., Aebersold, R., Li, R., et al. (1996) Independent evolution of the prochlorophyte and green plant chlorophyll a/b light-harvesting proteins. *Proceedings of the National Academy of Sciences* **93**: 15244–15248.

Rodrigue, S., Malmstrom, R.R., Berlin, A.M., Birren, B.W., Henn, M.R., and Chisholm, S.W. (2009) Whole Genome Amplification and De novo Assembly of Single Bacterial Cells. *PLoS ONE* **4**: e6864.

Rusch, D.B., Martiny, A.C., Dupont, C.L., Halpern, A.L., and Venter, J.C. (2010) Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proc Natl Acad Sci USA* **107**: 16184–16189.

Sher, D., Thompson, J.W., Kashtan, N., Croal, L., and Chisholm, S.W. (2011) Response of *Prochlorococcus* ecotypes to co-culture with diverse marine bacteria. *ISME J* **5**: 1125–1132.

Shibl, A.A., Thompson, L.R., Ngugi, D.K., and Stingl, U. (2014) Distribution and diversity of *Prochlorococcus* ecotypes in the Red Sea. *FEMS Microbiol Lett* **356**: 118–126.

Steinegger, M. and Söding, J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**: 1026–1028.

Thompson, A.W., Kouba, K., and Ahlgren, N.A. (2021) Niche partitioning of low‑light adapted *Prochlorococcus* subecotypes across oceanographic gradients of the North Pacific Subtropical Front. *Limnol Oceanogr* **66**: 1548–1562.

Ting, C.S., Rocap, G., King, J., and Chisholm, S.W. (2002) Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends in Microbiology* **10**: 134–142.

Tria, F.D.K., Landan, G., and Dagan, T. (2017) Phylogenetic rooting using minimal ancestor deviation. *Nat Ecol Evol* **1**: 1–7.

Ulloa, O., Henríquez-Castillo, C., Ramírez-Flandes, S., Plominsky, A.M., Murillo, A.A., Morgan-Lang, C., et al. (2021) The cyanobacterium *Prochlorococcus* has divergent light-harvesting antennae and may have evolved in a low-oxygen ocean. *Proc Natl Acad Sci USA* **118**: e2025638118.

West, N.J., Lebaron, P., Strutton, P.G., and Suzuki, M.T. (2010) A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME J* **5**: 933–944.

West, N.J. and Scanlan, D.J. (1999) Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Applied and Environmental Microbiology* **65**: 2585–2591.

Wu, Z., Aharonovich, D., Roth-Rosenberg, D., Weissberg, O., Luzzatto-Knaan, T., Vogts, A., et al. (2022) Single-cell measurements and modelling reveal substantial organic carbon acquisition by *Prochlorococcus*. *Nat Microbiol* **7**: 2068–2077.

Zinser, E.R., Coe, A., Johnson, Z.I., Martiny, A.C., Fuller, N.J., Scanlan, D.J., and Chisholm, S.W. (2006) *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Applied and Environmental Microbiology* **72**: 723–732.

Zinser, E.R., Johnson, Z.I., Coe, A., Karaca, E., Veneziano, D., and Chisholm, S.W. (2007) Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr* **52**: 2205–2220.
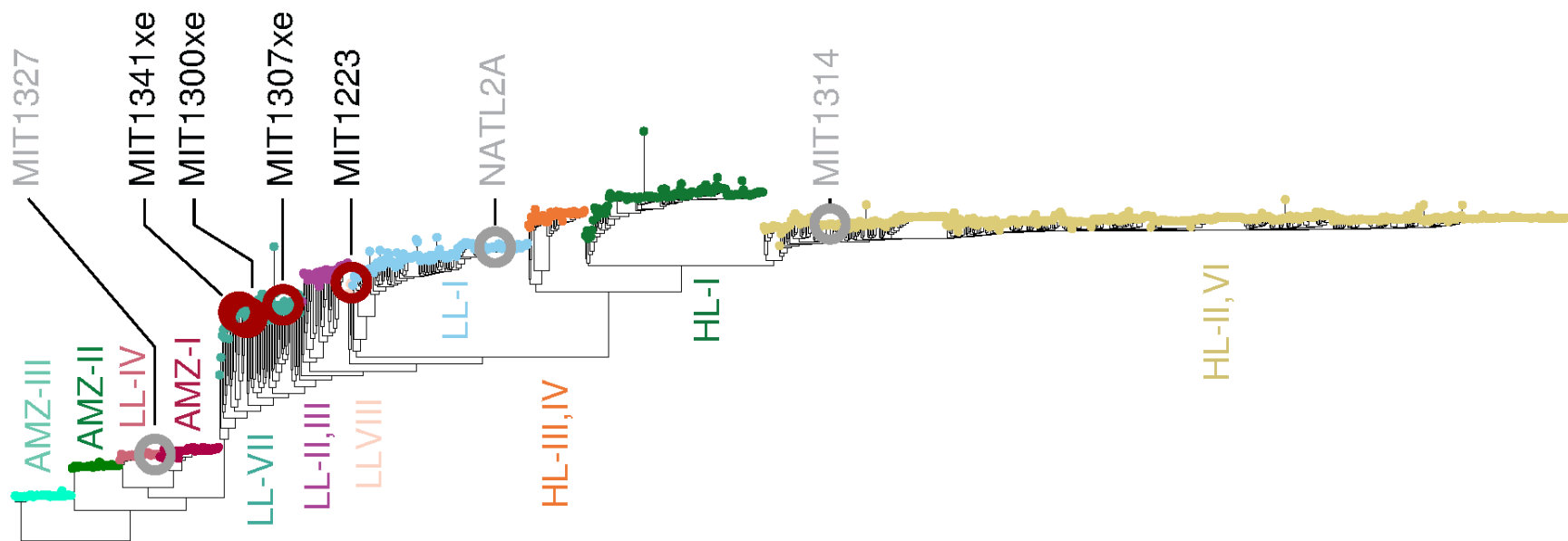
Figure 1. Placement of the isolates used in this study in a phylogeny of 1200 non-identical Prochlorococcus genomes. The genome database contains isolate genomes, MAGs, and SAGs with a wide range of genome completeness, isolation sources, and sequencing methods (see methods). The tree was built using FastTree2 (Price *et al.*, 2010) on the concatenated alignment of 44 synteny-conserved orthologous genes (methods). Major clades are represented by tip color. Novel isolates (black text, red circles) and representative strains (gray text and circles) are indicated on the tree.
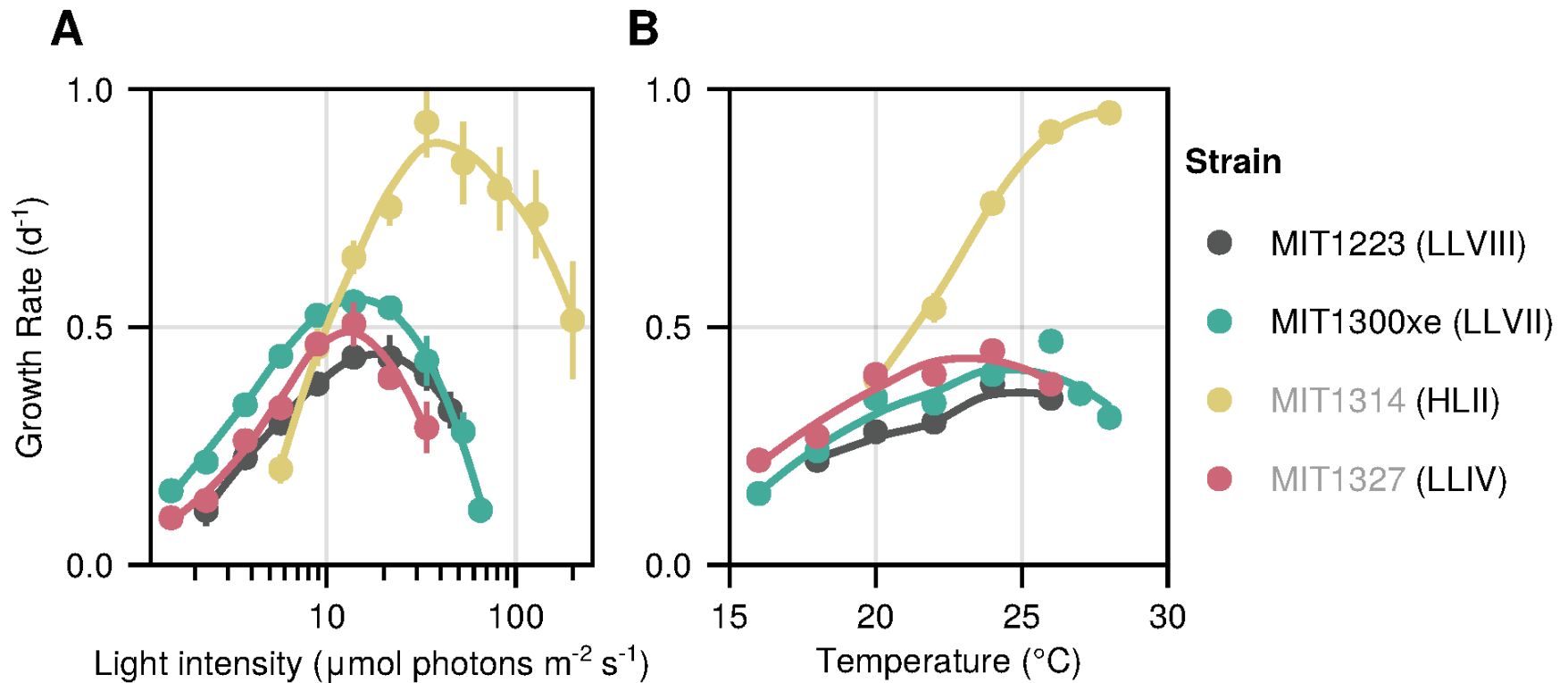
Figure 2. The light and temperature dependent growth rate of novel *Prochlorococcus* isolates. Growth rate as a function of light intensity (A) and temperature (B). Regressions were produced by locally weighted smoothing (LOESS). Circles represent the mean (± s.d.) of biological replicate cultures acclimated to each condition (see Table 2). Experiments in panel A were conducted at 24 ± 1 °C and experiments in panel B were done at 76 ± 1 μmol photons m⁻² s⁻¹ for MIT1314; 20 ± 1 μmol photons m⁻² s⁻¹ for MIT1223, MIT1300xe and MIT1327 on a 14:10 h light:dark cycle. Error bars are smaller than the size of the symbols where not visible. Novel isolate names are marked with black text, while previously described strains are listed in gray. The clade/grade designation of each strain is indicated in parenthesis.
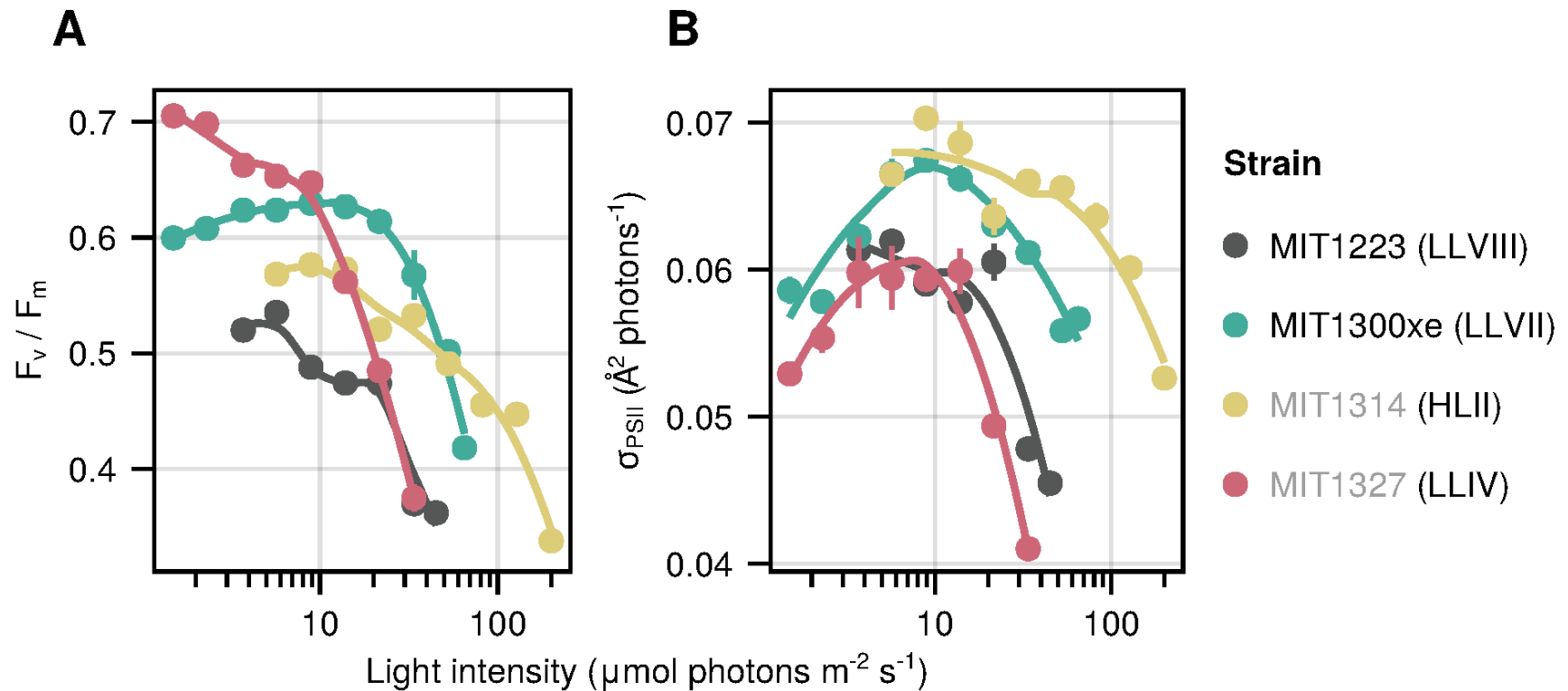
Figure 3. Photosynthetic parameters of novel isolates as a function of growth light intensity. Photosynthetic quantum efficiency (A) and functional absorption cross-section (σPSII) (B). Regressions produced by locally weighted smoothing (LOESS). Circles represent the mean (± s.d.) of biological replicate cultures acclimated to each condition sampled over time (n = 6 to 10). Error bars are smaller than the size of the symbols where not visible. Novel isolate names are marked with black text, while previously described strains are listed in gray. The clade/grade designation of each strain is indicated in parenthesis.
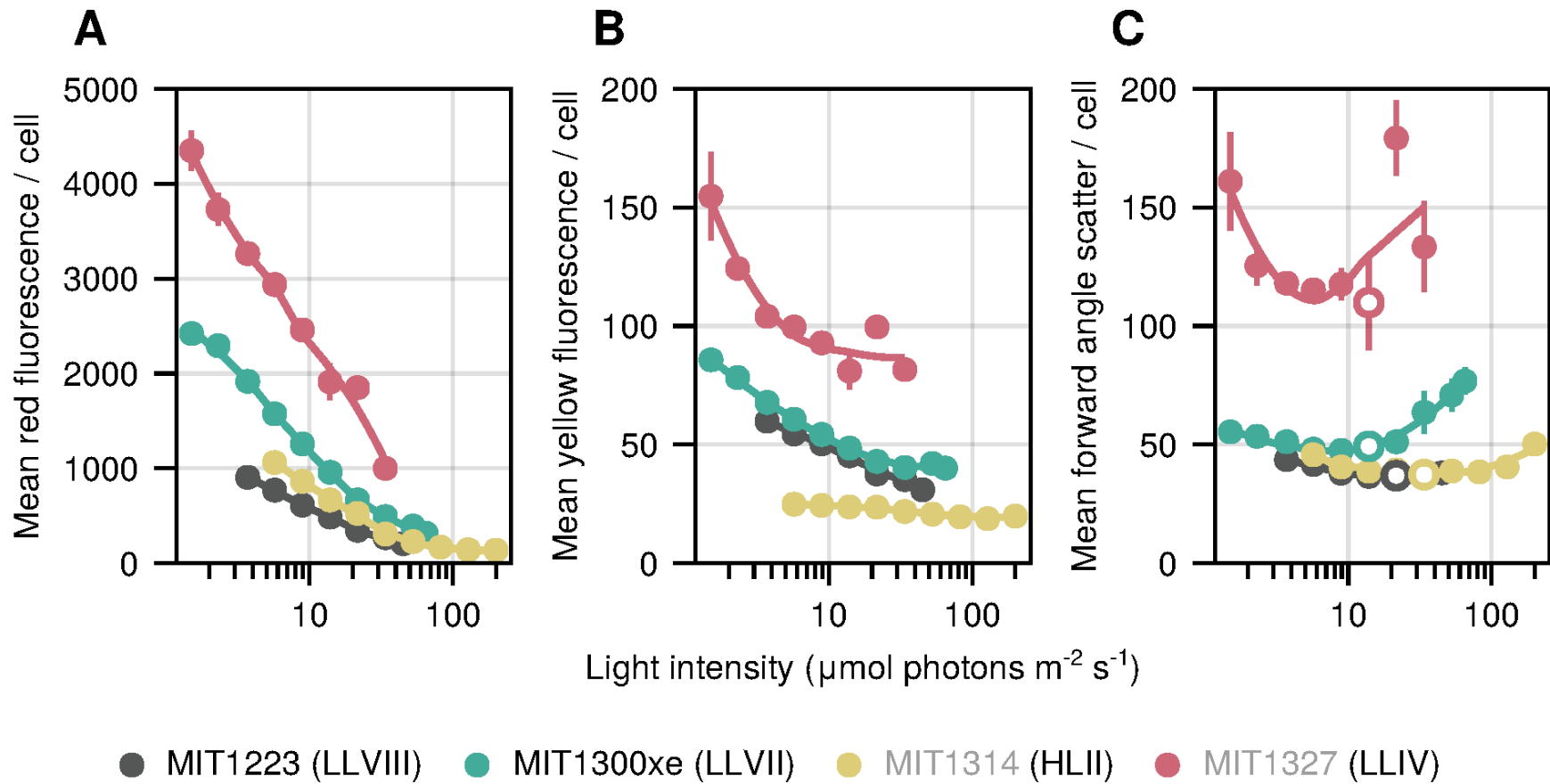
Figure 4. Fluorescence and light scattering properties of novel isolates as a function of growth light intensity. Flow cytometric analyses (A) red (chlorophyll) and (B) yellow (phycoerythrin) fluorescence per cell, and (C) forward light scatter (proxy for cell size). Symbols represent the mean (± s.d.) of biological replicate cultures acclimated to each condition. Error bars are smaller than the size of the symbols where not visible. Open circles in (C) denote the light level at which maximum growth rate was achieved for each strain. Novel isolate names are marked with black text, while representative strains are listed in gray. The lineage of each strain is indicated in parenthesis
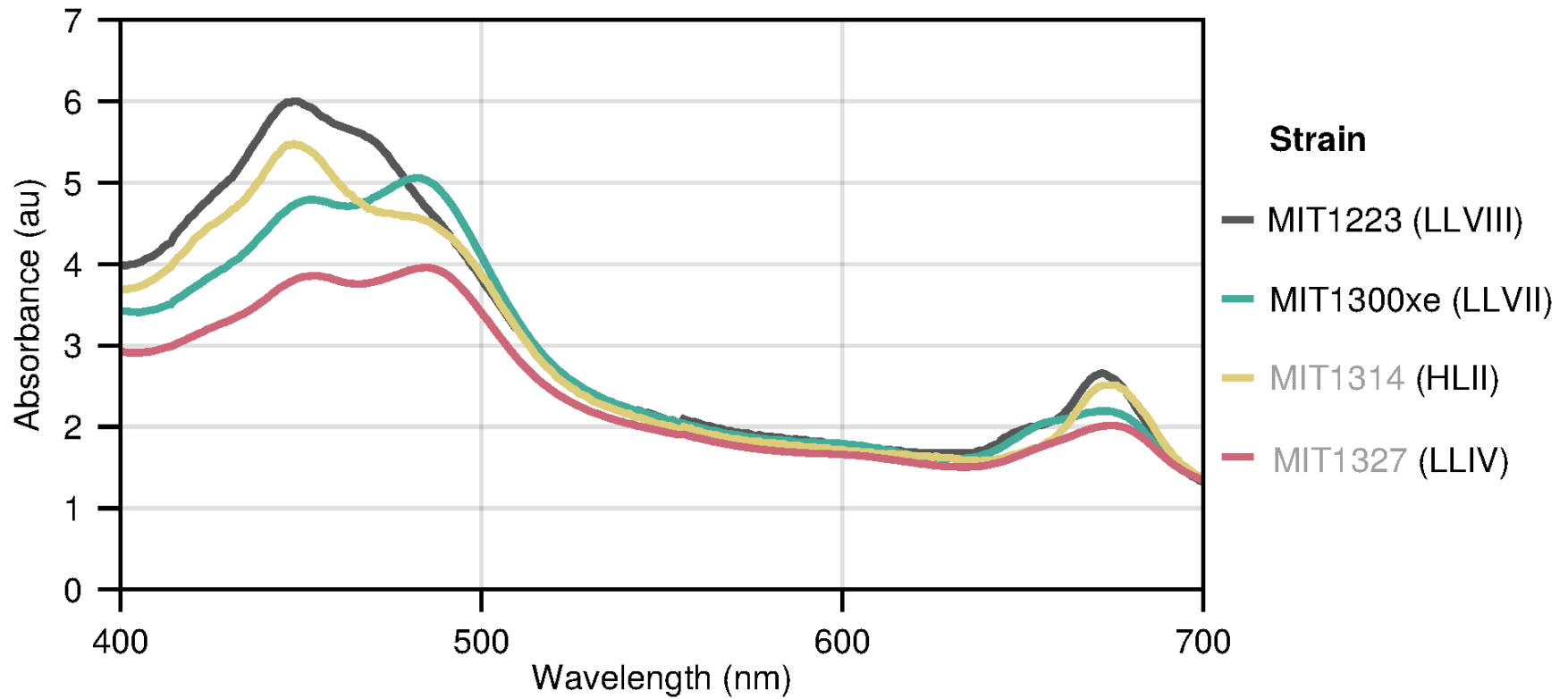
Figure 5. *In vivo* absorption spectra of novel isolates compared to representatives from established LL and HL-adapted clades. Spectra measured for log phase cultures growing at 5.7 μmol photons m^{-2} s^{-1}. Novel isolate names are marked with black text, while representative strains are listed in gray. The lineage of each strain is indicated in parenthesis.
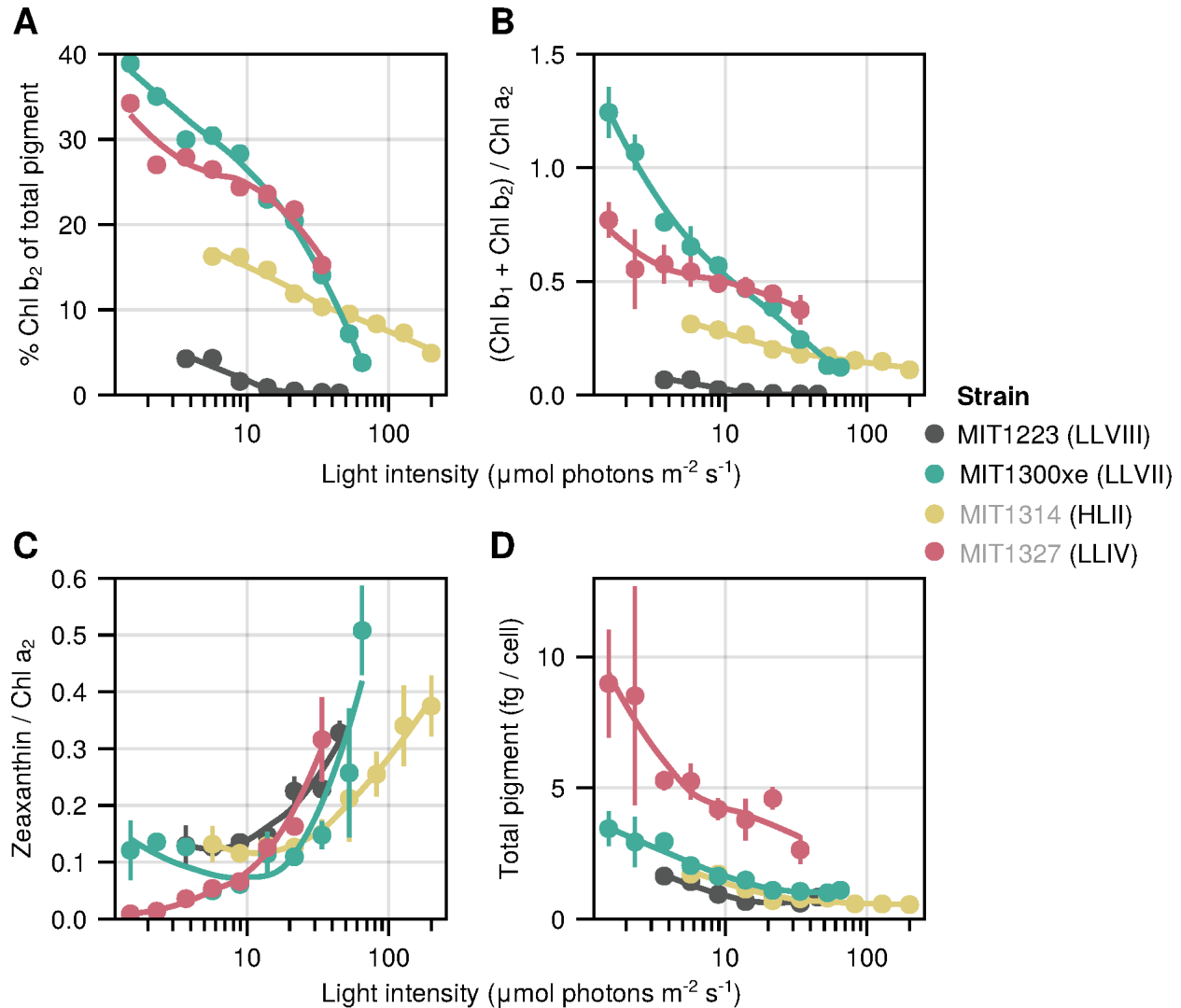
Figure 6. Chlorophyll content of the novel isolates as a function of growth light intensity. Percent of chlorophyll b2 from the total extractable pigment (A), chlorophyll b/a2 ratios (B), zeaxanthin/a2 ratios (C), and total extractable pigment per cell (D). Values are means (± s.d.) of biological replicate cultures acclimated to each condition sampled over time (n = 2 to 7). Lines represent local Loess fits. Novel isolate names are marked with black text, while representative strains are listed in gray. The lineage of each strain is indicated in parenthesis

A

PE
PC
AP

Thylakoid membrane

Photosystem II proteins

B

Photosystem II proteins

Pcb proteins

Chloropyll a$_2$, b

C

Chlorophyll a/b-binding, light-harvesting

Pcb proteins

pcbH
pcbF
pcbE
pcbD
pcbB
pcbA
isiA

Phycobilisome Billin

cpeZ
cpeY
cpeU
cpeT

Phycobilisome Phycoerythrin (PE)

cpeR
cpeD
cpeC
cpeB
cpeA

Phycobilisome Phycocyanin (PC)

cpcG
cpcF
cpcE
cpcD
cpcC
cpcB
cpcA

Phycobilisome Allophycocyanin (AP)

apcF
apcE
apcD
apcC
apcB
apcA

LLVII

AMZ III
AMZ II
LL IV
AMZ I
LL II/III
LL I
HL III,IV
HL I
HL II/VI

Synechococcus

Prochlorococcus

Figure 7. Genomic analysis of photosynthetic pigment biosynthesis and antenna genes across picocyanobacteria. Simplified structure of the phycobillisome (A) and prochlorosome (B) with the major pigments indicated by their abbreviations. (C) Distribution of light-harvesting antennae genes across the phylogeny of marine picocyanobacteria. The tree at the bottom represents the picocyanobacterial phylogeny as in Figure 1, with *Synechococcus* lineages added (see methods). Major lineages are depicted by colored text and corresponding tip point colors, except for *Synechococcus* which has no colored tips, and LLVII which is highlighted with a colored background in the tree and in the matrix. Each row in the matrix represents a single KEGG orthologue (and *isiA*), and colored bars signify the presence of a gene (row) in a genome (column). Vertical lines separate Syn, AMZ-III/II, LLIV/AMZ-I, LLVII, LLI-III, and HL groups. The variation in fine-scale presence/absence patterns observed within clades are most likely due to the inclusion of partial genomes in our database.

Strain	Clade or Grade	Sampling Date	Sampling Depth (m)	Isolation Medium	Pre-filtration	Purification Method	Light (µmol photons m^{-2} s^{-1})	Notes
MIT1314	HLII	June 2013	150	Pro2 + 1 µM sodium thiosulfate	Gravity (1.0 µm)	dilution to extinction	0.3 - 40	
MIT1223	LLVIII*	Sept. 2012	175	16 µM NO_3^- 1 µM PO_4^{3-} 1/10th Pro99 metals	None	dilution to extinction	3 - 19	Additional filtration (0.8 µm)
MIT1300xe	LLVII*	June 2013	150	15 µM NO_2^- 1 µM PO_4^{3-} Pro2/Pro99 trace metals	Gravity (1.0 µm)	non axenic	0.3 - 12	
MIT1307xe	LLVII*	June 2013	150	15 µM NO_2^- 1 µM PO_4^{3-} Pro2/Pro99 trace metals	Gravity (1.0 µm)	non axenic	0.3 - 12	Additional filtration (0.8 µm)
MIT1341xe	LLVII*	June 2013	150	15 µM NO_2^- 1 µM PO_4^{3-} Pro2/Pro99 trace metals	Gravity (1.0 µm)	non axenic	0.3 - 12	Additional filtration (0.8 µm)
MIT1327	LLIV	June 2013	150	Pro2 + 1 µM sodium thiosulfate	Gravity (1.0 µm)	dilution to extinction	3 - 12	

Table 1. Origin of the *Prochlorococcus* strains described in this work. *Denotes strains representing lineages with no prior isolates. The isolation location for all strains was Station ALOHA in the subtropical North Pacific Ocean (22.75 ºN 158 ºW). The light column indicates the range of irradiance values experienced by the cells during the journey from enrichment to unialgal isolate and the medium column the nutrient concentrations if they differ from Pro2 and Pro99 media recipes as described in (Moore *et al.*, 2007).

A

Strain	Light Intensity (μmol photons m^{-2} s^{-1})															
	1.5	**2.3**	**3.7**	**5.7**	**8.9**	**13.9**	**21.6**	**33.8**	**40**	**45**	**52.7**	**65**	**73**	**82.2**	**128**	**200**
MIT1314	ND	ND	ND	0.20 ± 0.03 (12)	0.46 ± 0.04 (20)	0.65 ± 0.04 (20)	0.75 ± 0.04 (18)	0.93 ± 0.07 (17)	-	-	0.84 ± 0.09 (20)	-	-	0.79 ± 0.09 (15)	0.74 ± 0.09 (15)	0.51 ± 0.12 (32)
MIT1223	ND	0.11 ± 0.03 (7)	0.23 ± 0.01 (20)	0.30 ± 0.02 (28)	0.38 ± 0.02 (23)	0.44 ± 0.02 (22)	0.44 ± 0.05 (32)	0.40 ± 0.03 (22)	-	0.32 ± 0.04 (9)	ND	ND	-	ND	-	-
MIT1300xe	0.16 ± 0.00 (10)	0.22 ± 0.00 (10)	0.34 ± 0.01 (12)	0.44 ± 0.01 (14)	0.52 ± 0.01 (16)	0.55 ± 0.02 (18)	0.54 ± 0.02 (16)	0.43 ± 0.05 (18)	-	-	0.28 ± 0.04 (14)	0.12 ± 0.03 (9)	ND	ND	-	-
MIT1327	0.10 ± 0.01 (4)	0.14 ± 0.01 (30)	0.26 ± 0.01 (16)	0.33 ± 0.01 (14)	0.46 ± 0.01 (18)	0.51 ± 0.05 (38)	0.39 ± 0.02 (12)	0.29 ± 0.05 (26)	ND	ND	ND	-	-	-	-	-

B

Strain	Temperature (°C)									
	15	**16**	**18**	**20**	**22**	**24**	**26**	**27**	**28**	**29**
MIT1314	-	-	ND	0.39 ± 0.00 (2)	0.54 ± 0.03 (3)	0.76 ± 0.00 (2)	0.91 ± 0.01 (2)	-	0.95 ± 0.00 (2)	ND
MIT1223	-	ND	0.22 ± 0.01 (2)	0.28 ± 0.01 (2)	0.30 ± 0.00 (3)	0.38 ± 0.01 (3)	0.35 ± 0.01 (2)	ND	-	-
MIT1300xe	ND	0.15 ± 0.00 (2)	0.24 ± 0.01 (2)	0.35 ± 0.01 (2)	0.34 ± 0.00 (3)	0.40 ± 0.00 (2)	0.47 ± 0.01 (2)	0.36 ± 0.00 (2)	0.31 ± 0.00 (2)	ND
MIT1327	ND	0.22 ± 0.00 (2)	0.27 ± 0.00 (2)	0.40 ± 0.00 (2)	0.40 ± 0.00 (3)	0.45 ± 0.01 (2)	0.38 ± 0.00 (2)	ND	-	-

Table 2. Mean growth rates (d^{-1}) ± s.d. for *Prochlorococcus* strains acclimated to different light intensities (A) and temperatures (B). ND (not determined) indicates that the strain was tested at this condition, however a consistent reproducible growth rate could not be achieved. A dash indicates that growth at this condition was not tested. Parentheses indicate the number of biological replicates included in calculations.

Strain	Clade or Grade	Assembly Size (bp)	% GC	# Scaffolds	Average Coverage	# Coding Sequences	JGI Genome ID	NCBI BioSample #
MIT1314	HLII	1,704,447	31.2	1	73	1,926	2681813573	SAMN38315734
MIT1223	LLVIII	1,795,922	35.7	1	236	1,939	2681813568	SAMN38315731
MIT1300xe	LLVII	1,855,146	41.4	1	186	1,965	2681813570	SAMN38315732
MIT1307xe	LLVII	2,032,419	39.9	1	317	2,140	2681813572	SAMN38315733
MIT1341xe	LLVII	1,937,096	40.1	1	335	2,037	2681813574	SAMN38315735
MIT1327	LLIV	2,587,389	50.3	29*	204	2,642	2681812949	SAMN04490364

Table 3. Genome characteristics and assembly statistics for *Prochlorococcus* strains in this study. *N50 for this assembly was 328.4 kbp.

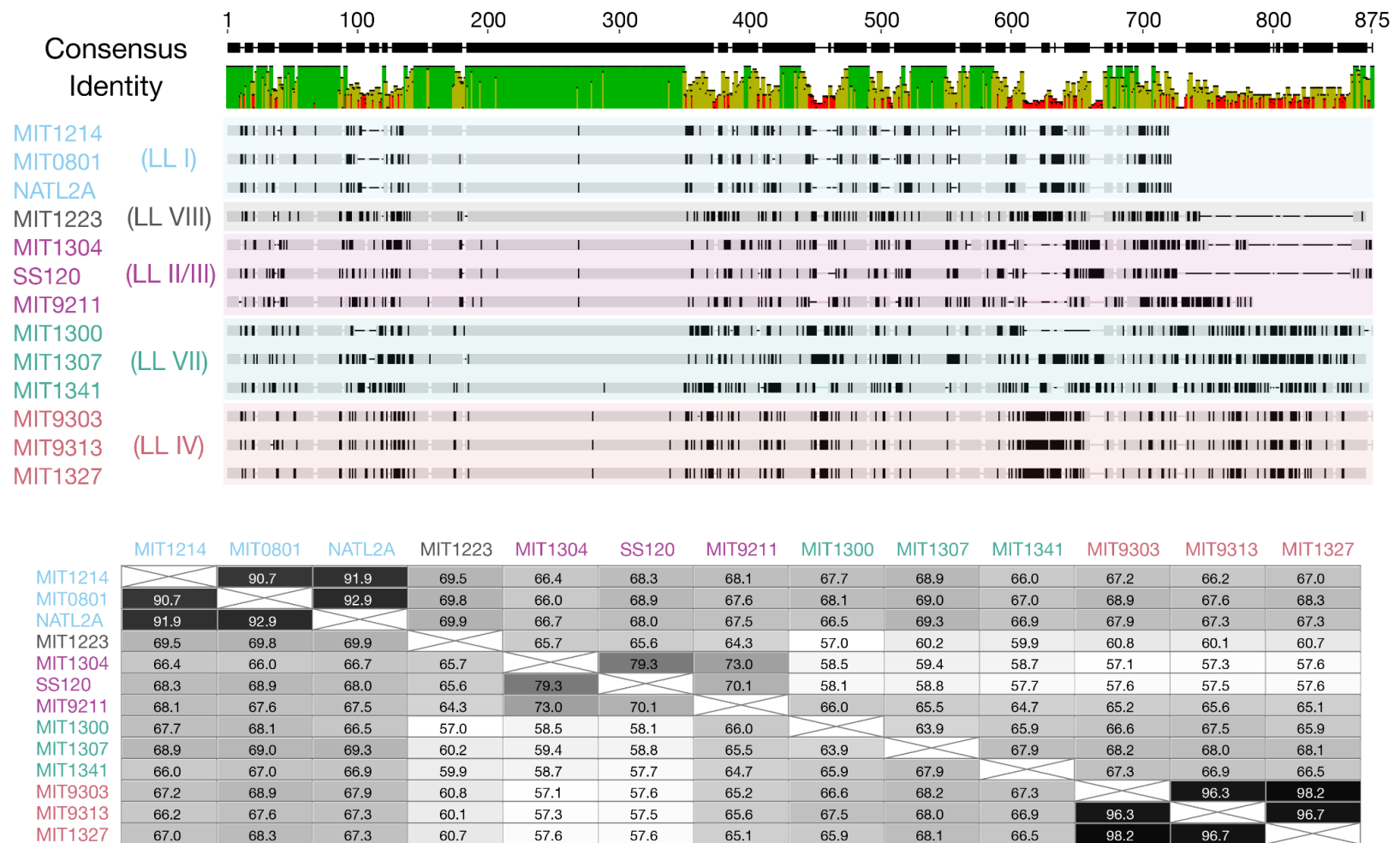|	MIT1214	MIT0801	NATL2A	MIT1223	MIT1304	SS120	MIT9211	MIT1300	MIT1307	MIT1341	MIT9303	MIT9313	MIT1327
MIT1214		90.7	91.9	69.5	66.4	68.3	68.1	67.7	68.9	66.0	67.2	66.2	67.0
MIT0801	90.7		92.9	69.8	66.0	68.9	67.6	68.1	69.0	67.0	68.9	67.6	68.3
NATL2A	91.9	92.9		69.9	66.7	68.0	67.5	66.5	69.3	66.9	67.9	67.3	67.3
MIT1223	69.5	69.8	69.9		65.7	65.6	64.3	57.0	60.2	59.9	60.8	60.1	60.7
MIT1304	66.4	66.0	66.7	65.7		79.3	73.0	58.5	59.4	58.7	57.1	57.3	57.6
SS120	68.3	68.9	68.0	65.6	79.3		70.1	58.1	58.8	57.7	57.6	57.5	57.6
MIT9211	68.1	67.6	67.5	64.3	73.0	70.1		66.0	65.5	64.7	65.2	65.6	65.1
MIT1300	67.7	68.1	66.5	57.0	58.5	58.1	66.0		63.9	65.9	66.6	67.5	65.9
MIT1307	68.9	69.0	69.3	60.2	59.4	58.8	65.5	63.9		67.9	68.2	68.0	68.1
MIT1341	66.0	67.0	66.9	59.9	58.7	57.7	64.7	65.9	67.9		67.3	66.9	66.5
MIT9303	67.2	68.9	67.9	60.8	57.1	57.6	65.2	66.6	68.2	67.3		96.3	98.2
MIT9313	66.2	67.6	67.3	60.1	57.3	57.5	65.6	67.5	68.0	66.9	96.3		96.7
MIT1327	67.0	68.3	67.3	60.7	57.6	57.6	65.1	65.9	68.1	66.5	98.2	96.7	

Figure S1. MAFFT (V7.54; (Katoh and Standley, 2013) alignment (top) of internal transcribed spacer (ITS) regions derived from *Prochlorococcus* isolates representing established clades (LLI, LLII/III, LLIV) and novel grades (LLVII, LLVIII) created using default settings in the Geneious software package (V10.2.6; (Kearse *et al.*, 2012). A percent identity matrix for the alignment is also included (bottom).
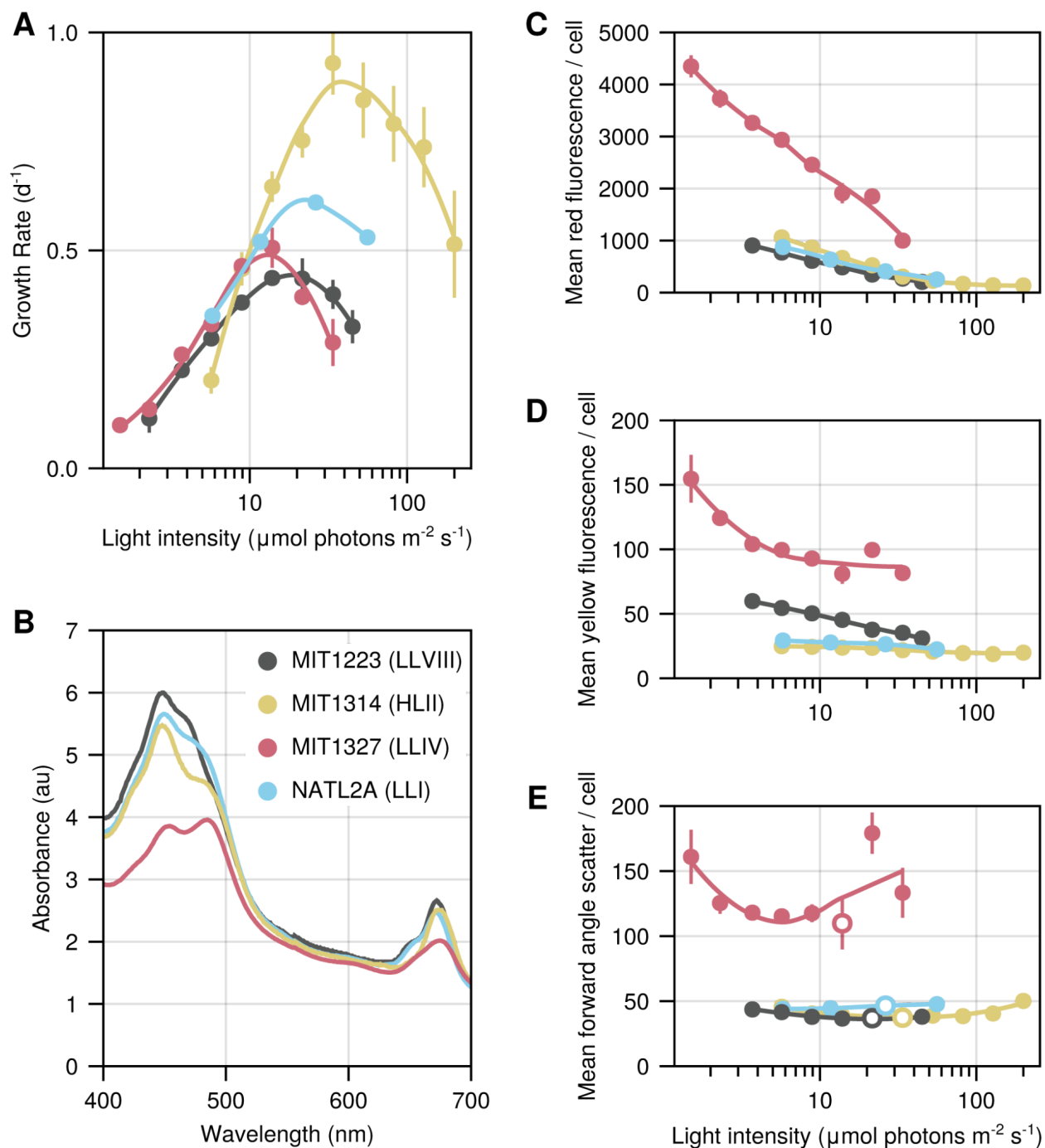
Figure S2. Photophysiology comparison with NATL2A (LLI clade). Light-dependent growth rates (A), *in vivo* absorption spectra for log phase cultures of different strains growing at 5.7 - 5.8 μmol photons m^{-2} s^{-1} (B), and flow cytometry parameters red fluorescence (C), yellow fluorescence (D), forward scatter (E) for NATL2A - a well studied LL-adapted strain of *Prochlorococcus* (Scanlan *et al.*, 1996; Biller *et al.*, 2015; Berube *et al.*, 2019) - for comparison with the novel strains reported in this study. Open circles denote the light level at which maximum growth rate was achieved for each strain (E) Data for MIT1314, MIT1223, and MIT1327 are reproduced from figures 2, 4, and 5.
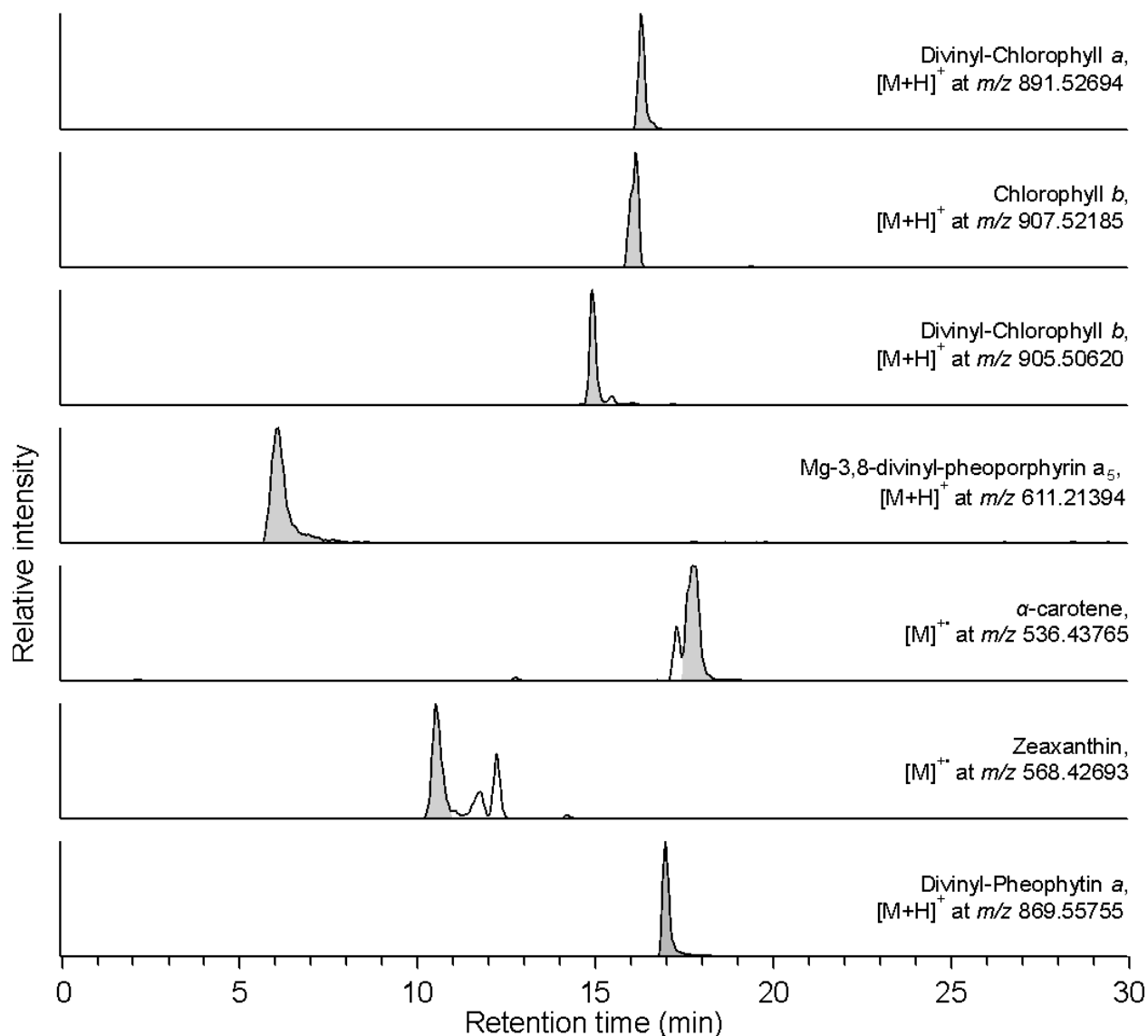
Figure S3. HPLC-Orbitrap-MS chromatograms shown as extracted ion chromatograms (EICs) of identified pigments in *Prochlorococcus* MIT1314 (HLII clade) acclimated to an irradiance of 21.6 μmol photons m^{-2} s^{-1}.

Divinyl-chlorophyll *a*
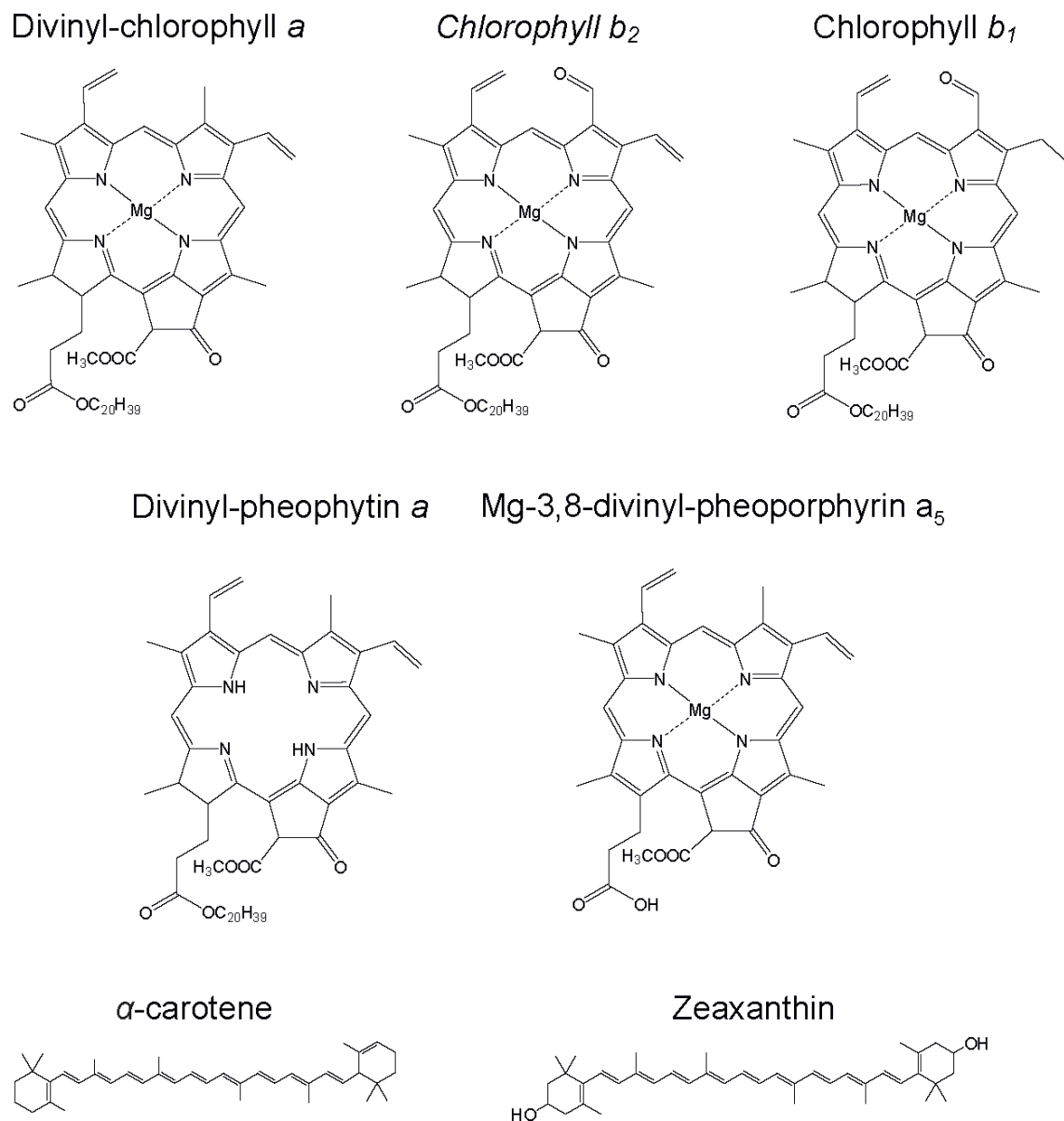
Chlorophyll b₂

Chlorophyll *b₁*

Divinyl-pheophytin *a*

Mg-3,8-divinyl-pheoporphyrin a₅

α-carotene

Zeaxanthin

Figure S4. Molecular structures of pigments identified in *Prochlorococcus* strains MIT1327, MIT1300xe, MIT1223, and MIT1314.

a

Unknown carotenoid,
EIC, $[M+H]^+$ at *m/z* 553.44039,
proposed sum formula: $C_{40}H_{56}O$

b

MS^2 spectrum of unknown carotenoid
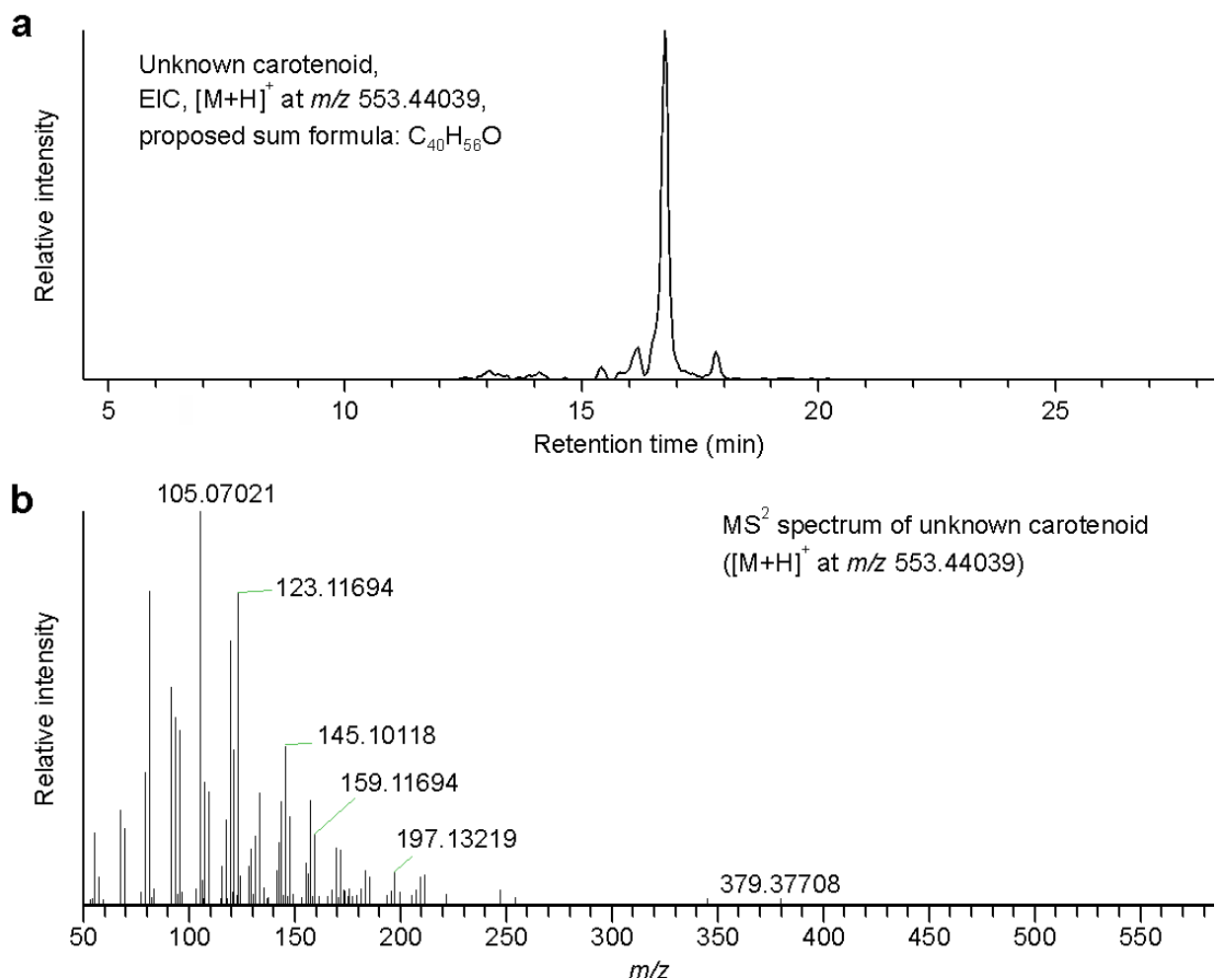($[M+H]^+$ at *m/z* 553.44039)

Figure S5. HPLC-Orbitrap-MS chromatogram shown as extracted ion chromatogram (EIC) of an unknown pigment ($[M+H]^+$ at *m/z* 553.44039) in *Prochlorococcus* MIT1314 acclimated to an irradiance of 21.6 µmol photons m^{-2} s^{-1} (A). The proposed sum formula ($C_{40}H_{56}O$) based on the accurate mass in full scan is also shown. MS^2 spectrum of the unknown carotenoid ($[M+H]^+$ at *m/z* 553.44039) (B). We observed a dominant fragment ion at *m/z* 123.117 in the MS^2 spectrum, which is typically associated with α-carotene (Rivera *et al.*, 2014) as the fragment represents the ε ring. The second dominant fragment at *m/z* 105 is usually associated with epoxy xanthophylls (Rivera *et al.*, 2014).
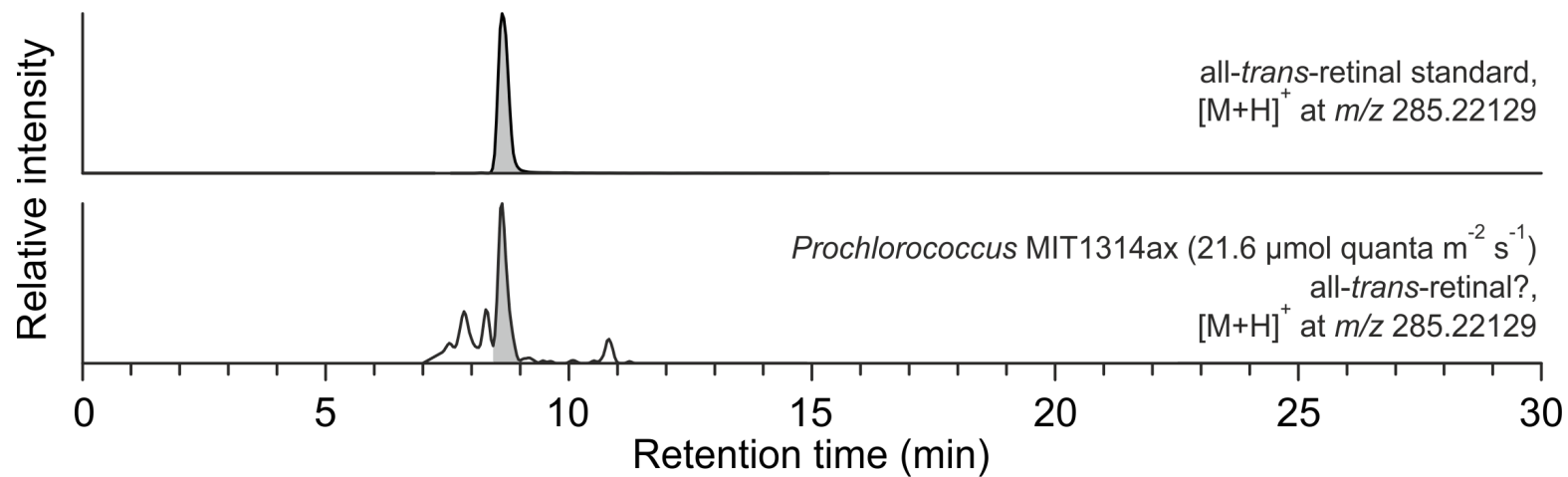
Figure S6. HPLC-Orbitrap-MS chromatogram shown as extracted ion chromatogram (EIC) of an all-*trans*-retinal standard ([M+H]$^+$ at *m/z* 285.22129) and an EIC of the same mass in *Prochlorococcus* MIT1314 acclimated to an irradiance of 21.6 μmol photons m^{-2} s^{-1}. The match of the accurate mass together with the retention time indicates the presence of all-*trans*-retinal in *Prochlorococcus*, but further identification is needed to confirm this identification. Due to its low abundance, we were not able to obtain MS^2 spectra of the putative retinal
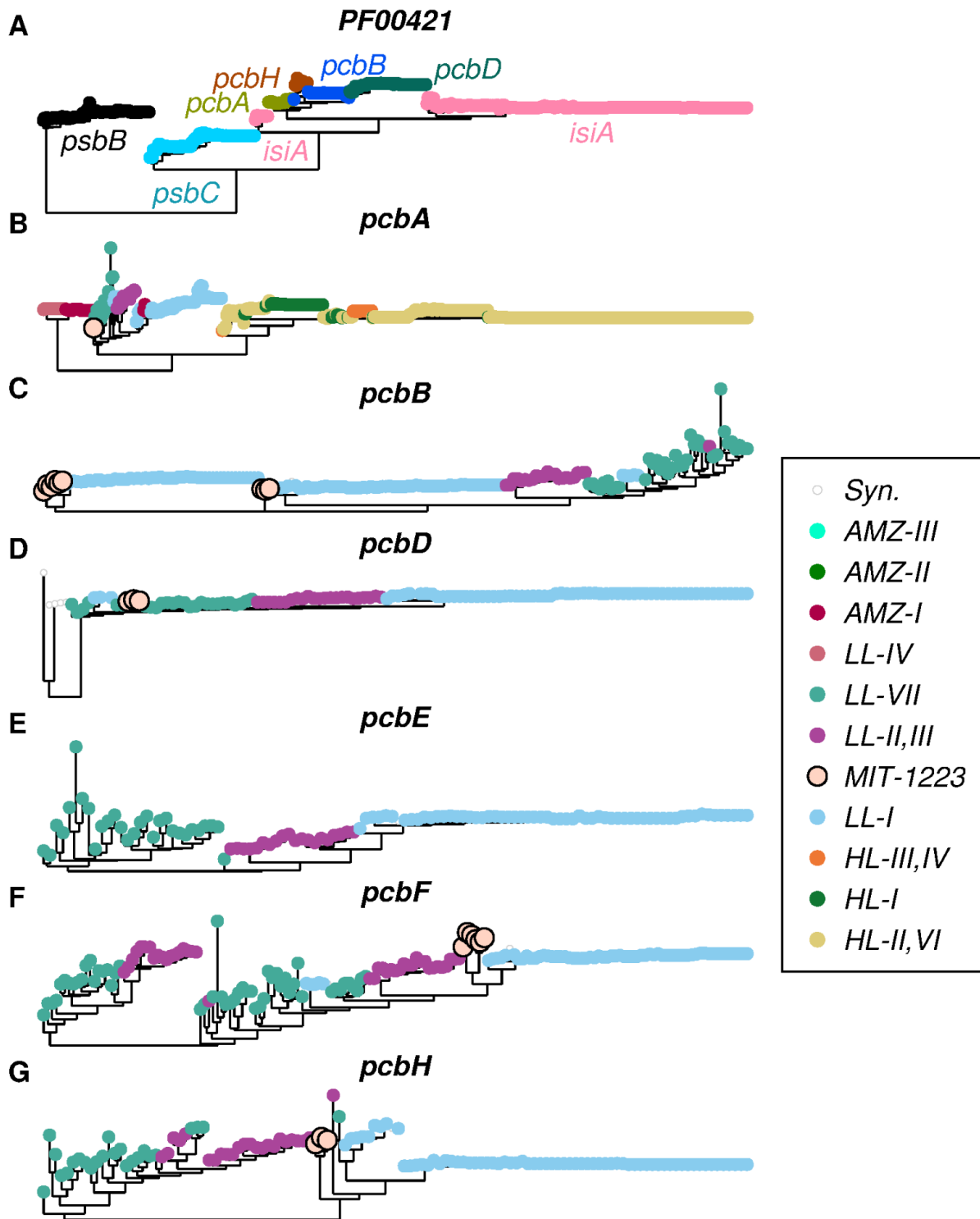
Supplementary Figure 7: Trees of picocyanobacterial genes that contain the pfam PF00421 (Photosystem II protein) domain. (A) Sequences were extracted based on the presence of the PF00421 domain. Colors represent assigned annotation, as used in Figure 8. pcbE/F proteins only contain weak hits to PF00421, and were excluded from the tree. For psbB/C, 100 sequences

were chosen at random, as they are all monophyletic. (B-G) trees of individual genes with tip colors depicting ecotype as shown in the legend. (B) pcbA mostly fits the species phylogeny and shows very different alleles for Basal (LL-IV, and AMZ-I), LL, and HL clades. (C) pcbB is found in LL only. MIT1223 together with LL-I. LL-I also had a duplication, which MIT1223 lacks. (D) MIT1223 embedded in LL-VII. (E) agrees with the species tree. (F) MIT1223 closest to LL-I. (G) MIT1223 has a divergent allele, dissimilar to both LL-VII and LL-I.
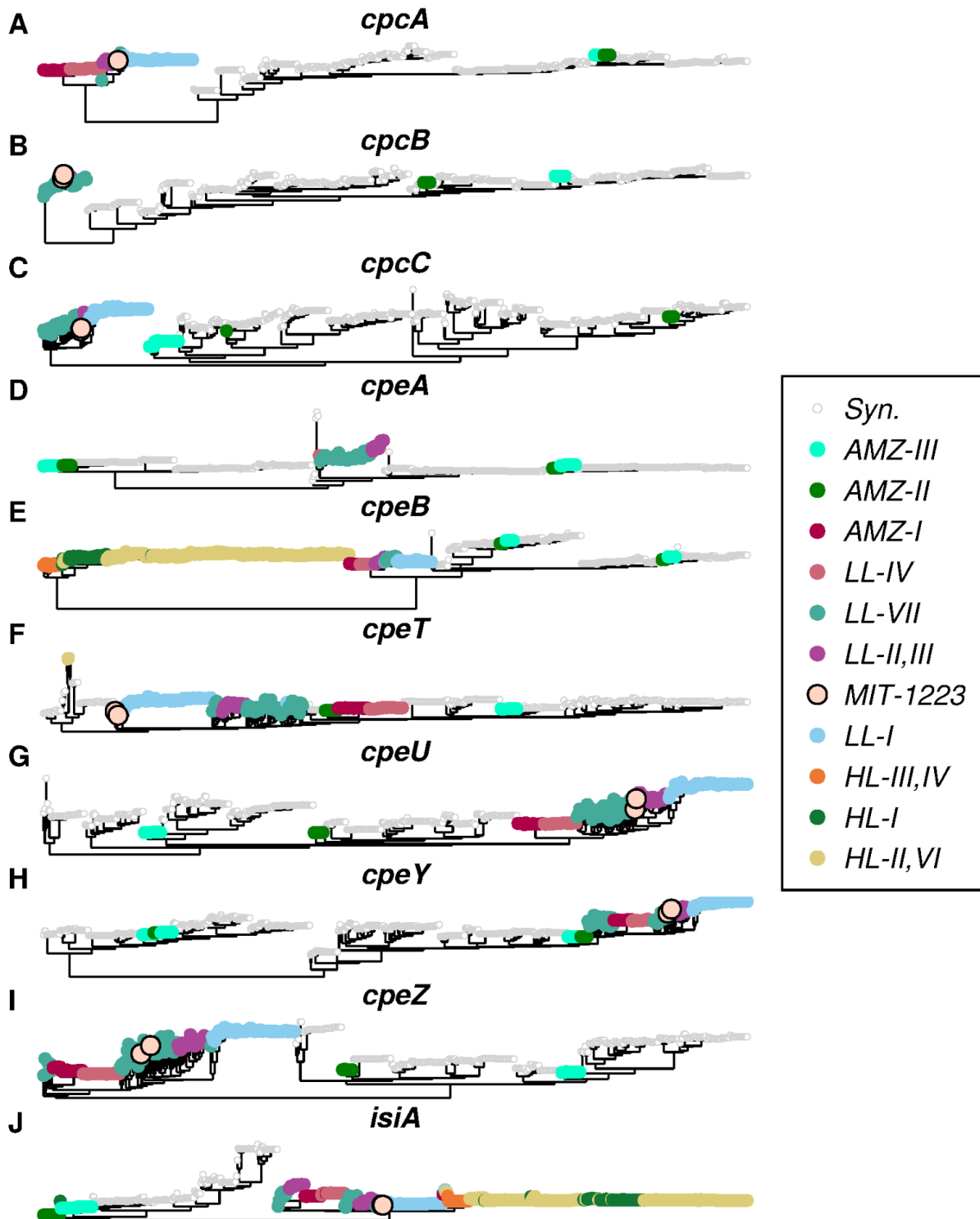
Supplementary Figure 8: Trees of phycobilisome associated genes. Tip colors represent ecotype as denoted in the legend. (A) *Prochlorococcus* allele is completely different from *Synechococcus*, except for AMZ-III, and II, which share the *Synechococcus* allele. (B) MIT1223

embedded in LL-VII. *Prochlorococcus* allele is completely different from *Synechococcus*, except for AMZ-III, and II, which share the *Synechococcus* allele. (C) MIT-1223 has a divergent allele, dissimilar to both LL-VII and LL-I. (D) Duplication in *Synechococcus*. AMZ-II and III acquired both copies. LL-VII and LL-II,III have a different allele that diverged from the ancestor of *Synechococcus* genes before their duplication. (E) HL allele is completely different from rest. Basal *Prochlorococcus* (AMZ-I, LL-IV, LL-II.III, LL-I, and LL-VII) have a different allele than *Synechococcus*, which has a duplication that is shared with AMZ-II and III. (F) MIT1223 has a divergent allele, dissimilar to both LL-VII and LL-I. Basal *Prochlorococcus* (AMZ-I to AMZ-III, and LL-IV) are embedded in *Synechococcus* diversity. (G) *Prochlorococcus* portion agrees with the species tree. AMZ-II and AMZ-III each have a different *Synechococcus* allele. (H) Duplication in *Synechococcus*. AMZ-II, and II have both *Synechococcus* alleles. *Prochlorococcus* diversity agrees with the species tree, and branches off from one of the *Synechococcus* alleles. (I) Three clades in *Synechococcus*. AMZ-II, and AMZ-III each have a different *Synechococcus* allele. MIT1223 embedded in LL-VII diversity. (J) *Prochlorococcus* completely separated from *Synechococcus*. Complex history in *Pro*. MIT1223 and some LL-VII share the LL-I allele.
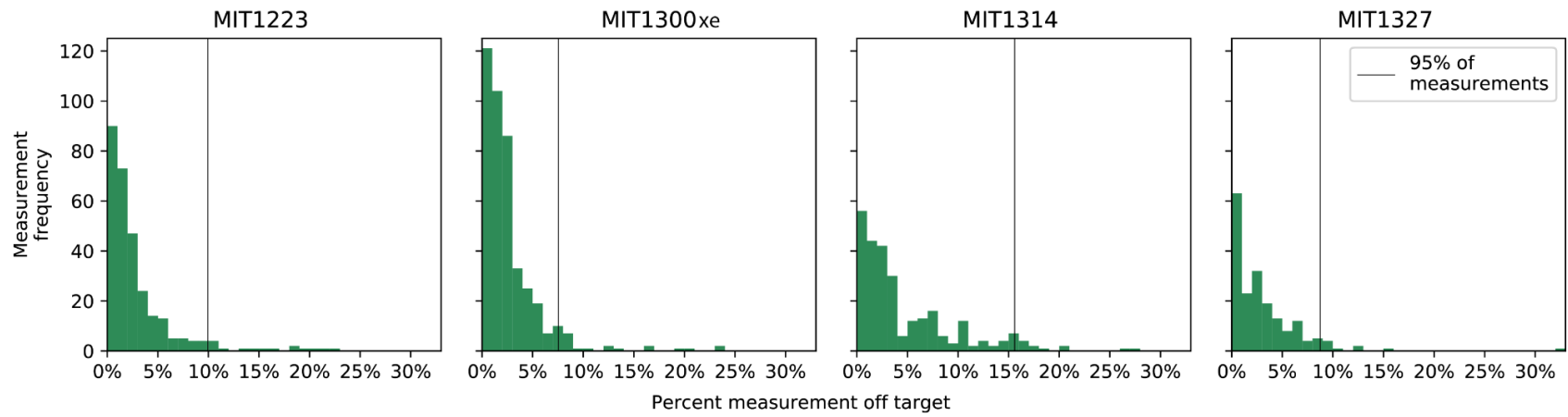
Figure S9. Histograms showing percent deviation from target light level for irradiance measurements taken over the course of light-dependent growth experiments.

Target light level	Median light level	Mean light level	Standard deviation	Minimum value	Maximum value	CV	Mean % of target	2.5th percentile	97.5th percentile
1.5	1.5	1.5	0.1	1.0	1.6	5.5	98.0	1.3	1.6
2.3	2.3	2.3	0.1	1.8	2.8	5.3	98.8	1.9	2.5
3.7	3.7	3.7	0.2	2.9	4.5	5.7	99.6	3.2	4.0
5.7	5.7	5.7	0.3	4.8	6.5	4.7	100.0	5.2	6.3
8.9	8.8	8.7	0.4	7.3	9.8	4.2	98.2	7.7	9.3
13.9	13.8	13.7	0.6	11.6	16.2	4.3	98.8	12.3	14.7
21.6	21.4	21.4	1.1	16.6	26.0	5.3	99.0	18.5	23.2
33.8	33.7	33.5	1.6	27.0	39.1	4.7	99.2	29.4	36.4
40.0	39.6	39.2	0.9	38.0	40.1	2.2	97.9	38.0	40.1
45.0	44.5	44.2	2.0	38.6	48.0	4.5	98.3	39.2	47.1
52.7	52.0	51.9	2.6	44.0	66.5	4.9	98.4	46.9	55.7
65.0	63.2	63.6	3.0	57.0	76.0	4.8	97.8	57.6	69.8
73.0	75.3	75.1	3.3	70.0	81.0	4.4	102.8	70.2	80.8
82.2	82.0	81.5	5.7	65.7	104.4	7.0	99.2	69.2	90.3
128.0	129.5	128.0	8.7	108.0	151.2	6.8	100.0	109.5	149.2
200.0	192.0	191.5	10.2	166.0	212.0	5.3	95.7	170.2	210.9

Table S1. Summary statistics for light measurements (µmol photons m^{-2} s^{-1}) taken over the course of the light-dependent growth rate experiments. Measurements were taken weekly and minor adjustments were made to maintain target levels. Two of the target light levels listed above (40.0 and 73.0 µmol photons m^{-2} s^{-1}) were used for acclimation purposes – no samples were taken at these levels.

References

Berube, P.M., Rasmussen, A., Braakman, R., Stepanauskas, R., and Chisholm, S.W. (2019) Emergence of trait variability through the lens of nitrogen assimilation in *Prochlorococcus*. *eLife* **8**: 10679.

Biller, S.J., Berube, P.M., Lindell, D., and Chisholm, S.W. (2015) *Prochlorococcus*: the structure and function of collective diversity. *Nat Rev Microbiol* 1–15.

Katoh, K. and Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**: 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Moore, L.R., Coe, A., Zinser, E.R., Saito, M.A., Sullivan, M.B., Lindell, D., et al. (2007) Culturing the marine cyanobacterium *Prochlorococcus*. *Limnol Oceanogr Methods* **5**: 353–362.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**: e9490.

Rivera, S.M., Christou, P., and Canela-Garayoa, R. (2014) Identification of carotenoids using mass spectrometry. *Mass Spectrom Rev* **33**: 353–372.

Scanlan, D.J., Hess, W.R., Partensky, F., Newman, J., and Vaulot, D. (1996) High degree of genetic variation in *Prochlorococcus* (Prochlorophyta) revealed by RFLP analysis. *Eur J Phycol* **31**: 1–9.