# Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild *Prochlorococcus*

Nadav Kashtan *et al.*

*Science* **344**, 416 (2014);
DOI: 10.1126/science.1248575

# Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild *Prochlorococcus*

Nadav Kashtan,[1]* Sara E. Roggensack,[1] Sébastien Rodrigue,[1,2] Jessie W. Thompson,[1] Steven J. Biller,[1] Allison Coe,[1] Huiming Ding,[1,3] Pekka Marttinen,[4] Rex R. Malmstrom,[5] Roman Stocker,[1] Michael J. Follows,[6] Ramunas Stepanauskas,[7] Sallie W. Chisholm[1,3]*

Extensive genomic diversity within coexisting members of a microbial species has been revealed through selected cultured isolates and metagenomic assemblies. Yet, the cell-by-cell genomic composition of wild uncultured populations of co-occurring cells is largely unknown. In this work, we applied large-scale single-cell genomics to study populations of the globally abundant marine cyanobacterium *Prochlorococcus*. We show that they are composed of hundreds of subpopulations with distinct "genomic backbones," each backbone consisting of a different set of core gene alleles linked to a small distinctive set of flexible genes. These subpopulations are estimated to have diverged at least a few million years ago, suggesting ancient, stable niche partitioning. Such a large set of coexisting subpopulations may be a general feature of free-living bacterial species with huge populations in highly mixed habitats.

The cyanobacterium *Prochlorococcus* is the smallest and most abundant photosynthetic cell in the oligotrophic oceans, contributing substantially to global photosynthesis (*1*). A single species by traditional measures, *Prochlorococcus* can be divided into several major clades, or ecotypes, defined by the intergenic transcribed spacer (ITS) region of their ribosomal RNA (rRNA) genes. These ecotypes are physiologically distinct (*2*–*4*); display distinctive seasonal, depth, and geographic patterns (*3*); and, like other microorganisms (*5*–*10*), embody tremendous genotypic and phenotypic diversity (*4*). To begin to understand the scope and limits of ecologically meaningful diversity within the canonical *Prochlorococcus* ecotypes, we examined cell-by-cell genomic diversity within a small sample of seawater and explored how it shifts in a dynamic environment.

We applied single-cell genome sequencing (*11*–*14*) to wild *Prochlorococcus* cells from samples collected at the Bermuda-Atlantic Time-series Study (BATS) site at three separate times of year (November 2008, February 2009, and April 2009) (Fig. 1A) (*15*). Because light, temperature, nutrients, and co-occurring communities change with winter deep mixing (*15*, *16*) (Fig. 1A), cells experience substantial environmental changes over tens of generations, enough to cause shifts in abundance of ITS-defined ecotypes (*2*, *15*, *17*).

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT), 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [2]Département de Biologie, Université de Sherbrooke, 2500 Boulevard de l'Université, Sherbrooke, Québec J1K 2R1, Canada. [3]Department of Biology, MIT, 77 Massachusetts Ave, Cambridge, MA 02139, USA. [4]Helsinki Institute for Information Technology, Department of Information and Computer Science, Aalto University, Post Office Box 15400, FI-00076 Aalto, Finland. [5]Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. [6]Department of Earth, Atmospheric and Planetary Sciences, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [7]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME 04544, USA.

*Corresponding author. E-mail: chisholm@mit.edu (S.W.C.); nadav.kashtan@gmail.com (N.K.)

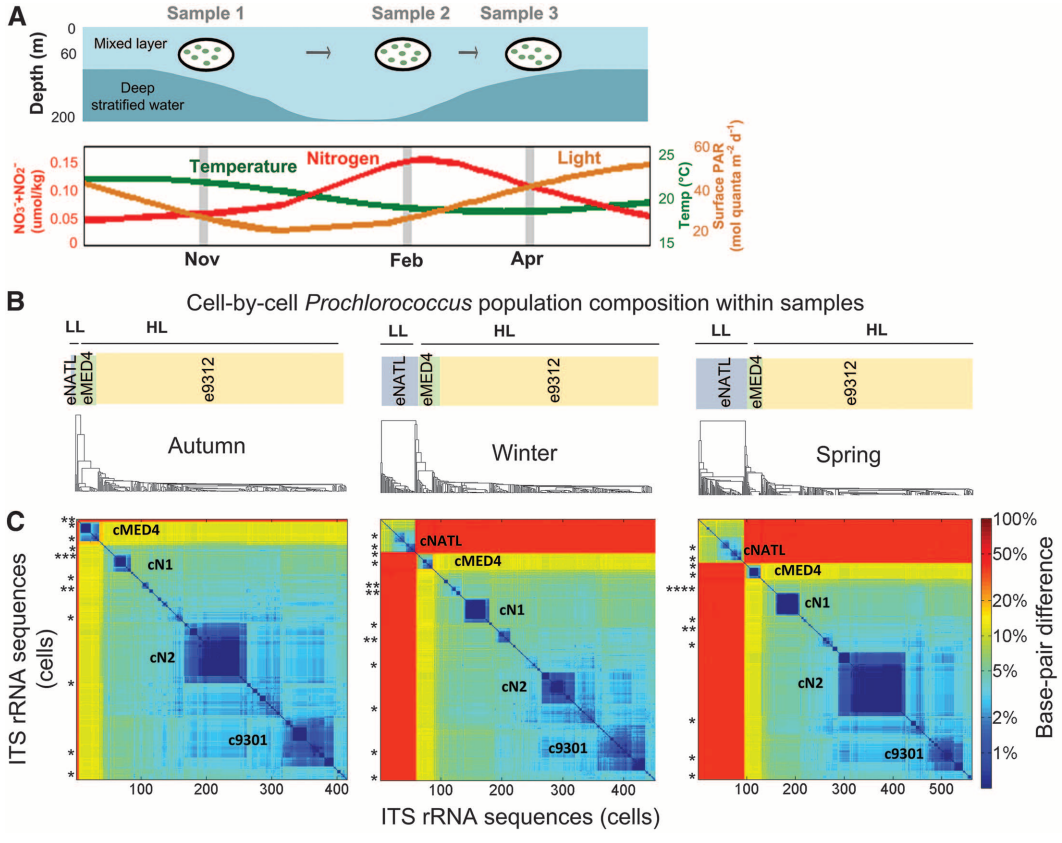


**Fig. 1. Cell-by-cell *Prochlorococcus* population composition in samples from three separate times of year at the BATS site.** Cells were collected within the mixed layer at 60 m depth in November 2008, February 2009, and April 2009 [see (*15*)]. (**A**) Schematic of seasonal dynamics at BATS and sampling design. (Top) A typical mixed-layer depth profile and context of our three samples. (Bottom) Typical average dynamics of light [smoothed mean surface photosynthetically active radiation (PAR) from 2004 to 2009], temperature, and nitrogen (within the mixed layer, averaged from 1999 to 2009) experienced by cells (*15*). Winter deep mixing brings cold nutrient-rich water to the surface. (**B**) Phylogenetic trees from pairwise genetic distances of ITS-rRNA sequences of individual cells from each sample [based on multiple alignment (*15*)]. The relevant sub-tree range of the known ecotypes (*2*) is marked above each tree if cells belonging to that ecotype were found, as is the division into low-light–adapted (LL) and high-light–adapted (HL) groups (*2*). (**C**) Heat maps describing the pairwise distance matrix between ITS-rRNA sequences of individual cells from each sample. Rows and columns are arranged according to the order of leaves of the trees shown in (B). The color map represents genetic distances as a percentage of base substitutions per site (log scale), such that the blue blocks identify very closely related ITS ribotypes. ITS sequences from cultured isolates with completely sequenced genomes are denoted by asterisks centered on the relevant line. Names of the largest clusters are marked in bold (e.g., cN2).

Flow sorting and DNA amplification (*11*–*14*) of more than 1000 co-occurring *Prochlorococcus* cells allowed us to explore the cell-by-cell genomic composition of these wild populations. We were able to identify coherent subpopulations at the whole-genome level and their relationship to those defined by the ITS region, explore finely resolved diversity patterns within and between subpopulations, and examine shifting abundances with seasonal changes in the habitat.

We first examined the population composition by sequencing the ITS regions of hundreds of *Prochlorococcus* cells in each sample, revealing the presence of finely resolved clusters within the broadly defined ecotypes (Fig. 1B). The populations were composed of tens to hundreds of nearly identical ITS clusters (>98% similar) within the coarse-grained ecotypes (Fig. 1, B and C). The relative abundance of cells belonging to the different clusters changed with season (Fig. 1, A to C) (*15*), suggesting shifts in their relative fitness in response to environmental changes.

To study the fine-scale genomic variation and compare it with the ITS-defined clusters, we sequenced the partial genomes (representing, on average, 70% of the total genome) of 90 individual cells (30 per sample) from the largest nearly identical ITS cluster, cN2 (Figs. 1C and 2), as well as 6 cells from two other clusters, cN1 and c9301. For each time of year, cells were randomly selected for genome sequencing from within the major ITS ribotypes (>99% similar) within cluster cN2 (C1 to C5) (30 cells), as well as from c9301-C8 and cN1-C9 (one cell each), as detailed in (*15*). We used a modified mediator genome reference assembly approach (*15*, *18*) to analyze between-cell variation in the partial genomes recovered. The topologies of the ITS and genomic trees were highly congruent (Fig. 2), indicating that ITS sequences can serve as a proxy for genome sequences in *Prochlorococcus* at a much finer level of resolution than previously demonstrated (*4*, *19*). The genomic data further revealed that the largest cluster cN2 is divided into five major clades [C1 to C5 (Fig. 2)] and a few additional minor clades represented by only one cell each. The delineation of clades C1 to C5 was highly robust and also observed in trees constructed from genomic position subsets (figs. S1 and S2).

To explore the evolutionary forces that shaped the cN2 C1 to C5 clades, we examined differences in nucleotide sequences within and between clades. For example, the C1 and C3 subpopulations (Fig. 2B) differ in 52,885 dimorphic single-nucleotide polymorphisms (SNPs), which represent 3.2% of their genomes (Fig. 3A, blue). The dimorphic SNPs between C1 and C3 are scattered across the genomes, occurring in 1519 out of 1974 genes (most of them core genes); 8% of these SNPs are found in intergenic regions (9% of the genome is noncoding). Of the intragenic SNPs, 37% are nonsynonymous, thus affecting the amino acid sequences of the proteins they encode. In contrast to the scattered nature of the sequence variation between the C1 and C3 clades, the polymorphism within them is confined to a few regions of the genome (Fig. 3A, black), indicating that most regions along the genome are conserved within clades and are different between them (*15*), which is true for all pairwise comparisons within C1 to C5 (figs. S3 and S4).

This emerging pattern was further supported by a standard measure of genetic differentiation between populations, the fixation index ($F_{ST}$) (*20*), applied at gene-by-gene resolution to the five cN2 clades, C1 to C5 (Fig. 3, B and C). Seventy-five percent of the core genes had high $F_{ST}$ values (>0.8), (Fig. 3, B and C) (*15*), meaning different clades contained significantly different alleles. Some of the differentiated core genes have functions involved in the interaction between the cell and environmental stimuli [e.g., transporters, genes that affect oxidative stress responses, and cell surface biosynthesis and modification (Data S1)]; that is, they are not all simply "housekeeping genes" that control central metabolism. For example, alleles of phosphoglucosamine mutase, which is involved in the biosynthesis of outer membrane lipopolysaccharides (*21*), differ by an average of 10% of their amino acid sequences (Fig. 3C), with substitutions in the hydrophilic center of the enzyme (*21*), possibly affecting its specificity and kinetics.

We next asked whether different clade subpopulations carry distinct sets of flexible genes. Using de novo assemblies to capture regions unmapped by the reference assemblies (*15*), we found that each subpopulation carries a small set of distinct genes, typically in the form of cassettes within genomic islands (Table 1). Cassettes containing genes in the glycosyltransferase family account for much of the gene content variation between these clade subpopulations (Table 1 and table S1). The gene content in these cassettes suggests involvement in outer membrane modifications, possibly affecting phage attachment (*22*), recognition by grazers (*23*), cell-to-cell communication, or interactions with other bacteria (*24*).

We conclude that these clade subpopulations have distinct "genomic backbones" (and are

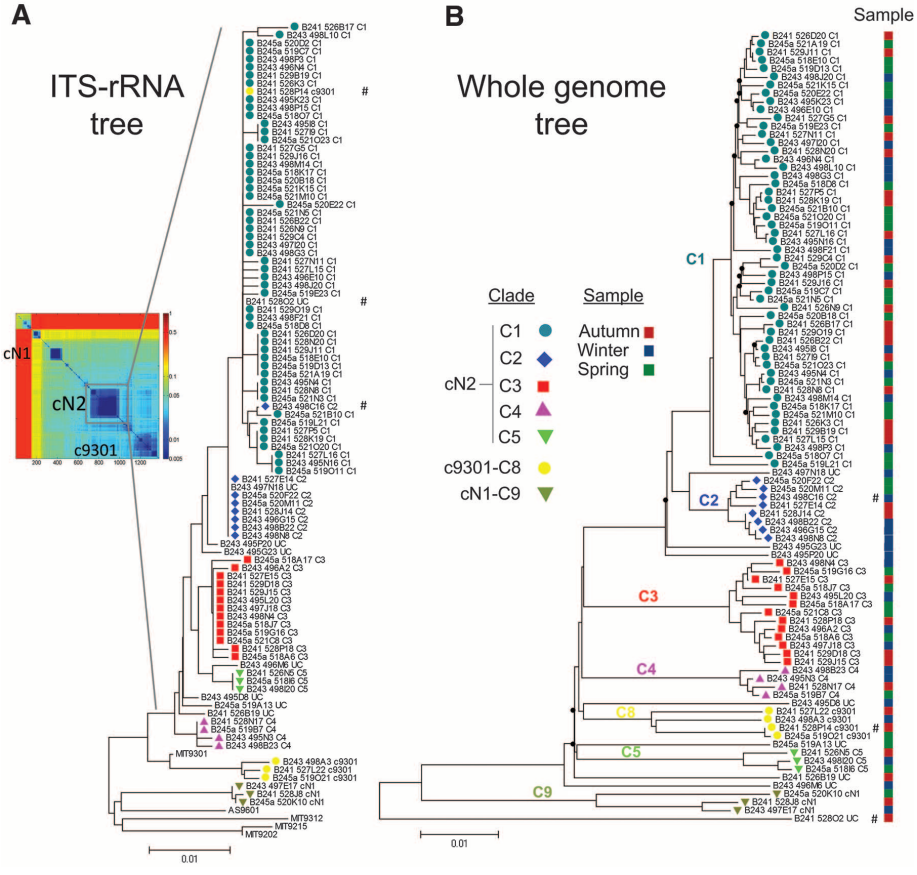


**Fig. 2. ITS-rRNA sequence and whole-genome neighbor-joining phylogenetic trees at a fine resolution of diversity.** (**A**) Phylogenetic tree based on ITS-rRNA sequences of 96 single cells (90 cN2 ribotypes, three cN1 ribotypes, and three c9301 ribotypes), as well as additional five high-light–adapted cultured strains. (**B**) Phylogenetic tree of the 96 single cells based on whole-genome sequences. The colored symbols to the left of the leaf labels in (A) and (B) represent the different clades depicted from the deep branches observed in the whole-genome tree. The sample origin of each cell is marked with red, blue, and green squares (representing autumn, winter, and spring, respectively) on the right. Distance units are base substitutions per site (see scale bar) (*15*). Bootstrap values <80 are marked as black dots on the internal nodes in (B) (fig. S1). Cells marked with # fall into an ITS clade that differs from the genome-defined clade. Neighbor-joining trees in (A) and (B) were constructed using p-distance.

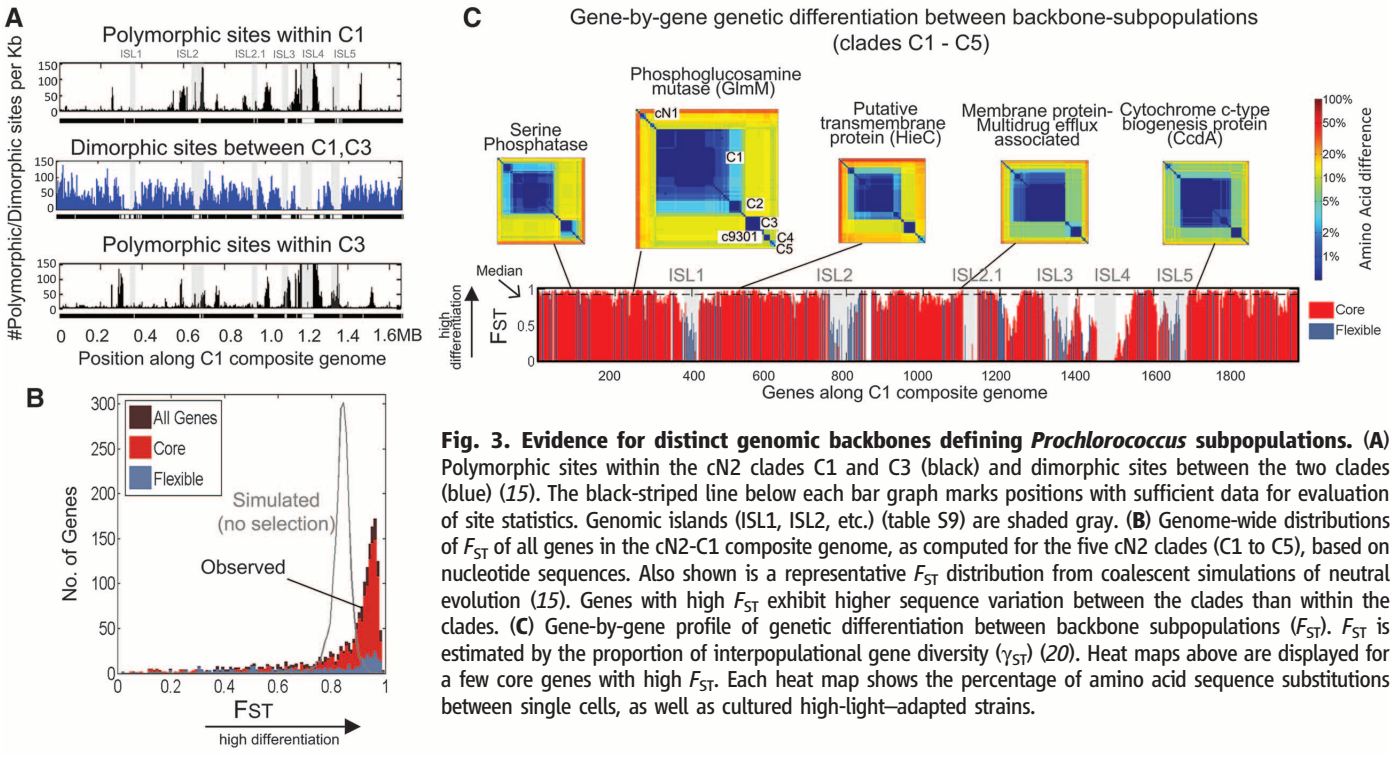


**Fig. 3. Evidence for distinct genomic backbones defining *Prochlorococcus* subpopulations.** (**A**) Polymorphic sites within the cN2 clades C1 and C3 (black) and dimorphic sites between the two clades (blue) (*15*). The black-striped line below each bar graph marks positions with sufficient data for evaluation of site statistics. Genomic islands (ISL1, ISL2, etc.) (table S9) are shaded gray. (**B**) Genome-wide distributions of $F_{ST}$ of all genes in the cN2-C1 composite genome, as computed for the five cN2 clades (C1 to C5), based on nucleotide sequences. Also shown is a representative $F_{ST}$ distribution from coalescent simulations of neutral evolution (*15*). Genes with high $F_{ST}$ exhibit higher sequence variation between the clades than within the clades. (**C**) Gene-by-gene profile of genetic differentiation between backbone subpopulations ($F_{ST}$). $F_{ST}$ is estimated by the proportion of interpopulational gene diversity ($\gamma_{ST}$) (*20*). Heat maps above are displayed for a few core genes with high $F_{ST}$. Each heat map shows the percentage of amino acid sequence substitutions between single cells, as well as cultured high-light–adapted strains.

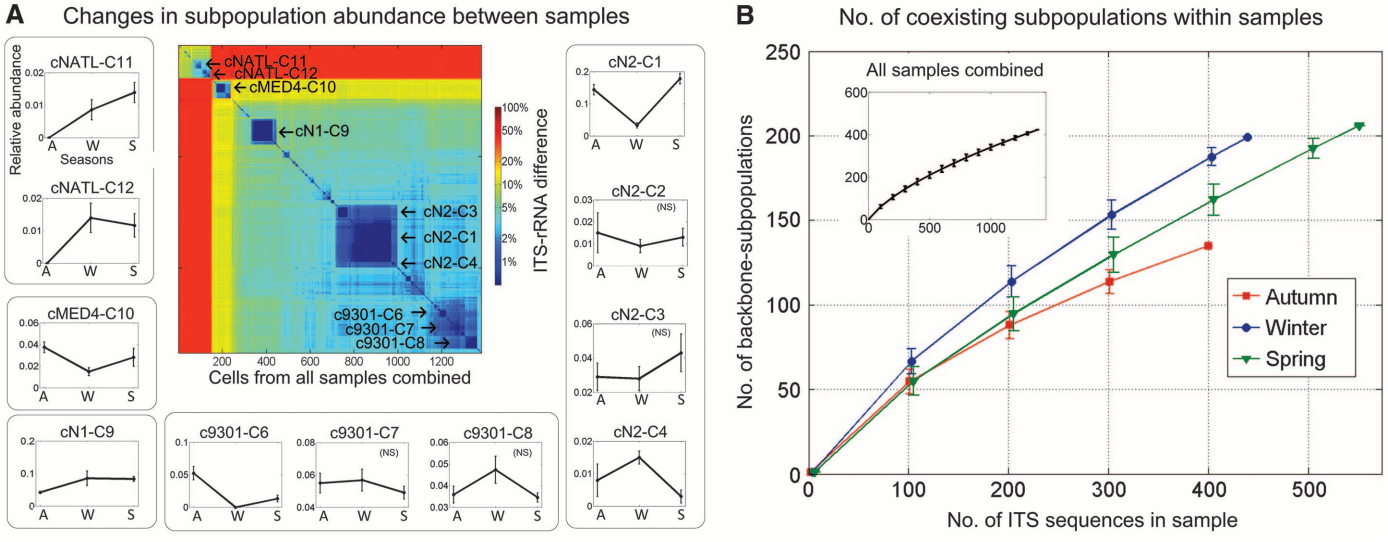


**Fig. 4. Abundance profiles of backbone subpopulations and the estimated number of coexisting subpopulations within samples.** (**A**) Relative abundance profiles of the 11 largest backbone subpopulations in our samples within the ITS clusters cNATL, cMED4, cN1, cN2, and c9301. A, autumn; W, winter; S, spring. Backbone names are marked near the relevant cluster on the ITS heat map. Backbone subpopulations were predicted by 99% ITS similarity for the full set of 1381 ITS sequences. Error bars represent SEM. NS, no significant changes between seasons (false discovery rate, $\alpha = 0.05$) (*15*). (**B**) Rarefaction curves estimating the number of coexisting backbone subpopulations within samples (*15*). Backbones predicted as in (A). Error bars represent 95% confidence intervals. (Inset) Rarefaction curve of all samples combined.

henceforth referred to as "backbone subpopulations") consisting of highly conserved (within subpopulation) alleles of the majority of core genes and a small distinct set of flexible genes that is linked with a particular backbone. This covariation between the core alleles and flexible gene content, and its fine scale resolution, represents a new dimension of microdiversity within wild *Prochlorococcus* populations. It is noteworthy that similar patterns have been identified in cultured isolates and metagenomic assemblies within coexisting members of a few other microbial species with very different ecologies (*5*–*10*, *25*), suggesting that differentiated genomic backbones may be a feature of diverse types of microbial populations.

At a finer resolution of diversity, we observed that cells within the five cN2 backbone subpopulations differ by 19,000 nucleotide positions on average, in comparison to 77,000 positions between backbone subpopulations (equivalent to 1.2 and 4.7% of the genome, respectively) (Fig. 2B). The most similar pairs of individual cell genomes in our samples differ in a few hundred base pairs [close to the detection limit when one considers single-cell processing and sequencing error (*15*)]; some of these pairs likely have identical

**Table 1. Flexible gene cassettes associated with different cN2 backbone subpopulations highlighting gene content that may contribute to ecological differentiation.** GT, glycosyltransferase; ABC-T, adenosine triphosphate–binding cassette (ABC) transporter; HLIP, high-light–inducible protein; CO, Cytochrome oxidase c subunit VIb; HlpA, histone-like protein; CpsL, polysaccharide biosynthesis protein. In the "Selected gene annotations" column, numbers before gene annotation refer to number of that type of gene. A complete list of the genes in each cassette is described in table S1 (*15*).

| Clade | Cassette ID | Position | No. of genes in cassette | Selected gene annotations | Cassette function |
|---|---|---|---|---|---|
| **cN2-C1** | CST_I | Island 2.1 | 4 | HLIP, CO | Redox stress response |
| | CST_II | Island 4 | 7 | 3GT, ABC-T | Outer membrane modification |
| **cN2-C2** | CST_II | Island 4 | 7 | 3GT, ABC-T | Outer membrane modification |
| **cN2-C3** | CST_III | Island 1 | 2 | 2GT | Outer membrane modification |
| **cN2-C4** | CST_I | Island 2.1 | 4 | HLIP, CO | Redox stress response |
| | CST_IV | Island 4 | 14 | 3GT, HlpA, CpsL | Outer membrane modification |
| **cN2-C5** | CST_V | Island 4 | 5 | 2GT | Outer membrane modification |

gene content (*15*). Except for these few pairs, each cell carries at least one gene cassette not found in any other. In some cases, a few closely related cells (a subclade) within backbones share a distinct gene cassette. Among these genes are, again, glycosyltransferase genes, as well as transporters and genes involved in nucleotide binding and processing. In a few cases, cells from different backbone subpopulations carry similar flexible gene cassettes [e.g., high-light–related genes (Table 1) and phosphonate related genes], demonstrating the combinatorial nature of backbones and flexible genes.

If backbone subpopulations have differential fitness, we would expect their relative abundance to change with changing environmental conditions (Fig. 1). Accordingly, the majority of the largest subpopulations exhibited significant seasonal abundance variation (Fig. 4A), higher than expected by chance (*15*), consistent with the hypothesis that this reflects selection, but more data are needed to draw that conclusion. Backbone subpopulations maintain their genomic composition between seasons (tested for C1) (*15*), which we would expect, as the establishment of new mutations and the acquisition and loss of genes are not likely to be in play on these time scales (*15*).

The congruency of genomic and ITS phylogenies in *Prochlorococcus* at both coarse (*4*, *19*) and fine resolution (Fig. 2) suggests that ITS-ribotype clusters coincide, in most cases, with distinct genomic backbones (*15*). This allowed us to estimate the number of coexisting backbone subpopulations in our samples through rarefaction analysis, revealing at least hundreds of coexisting subpopulations with distinct backbones (Fig. 4B) in each sample. These backbone subpopulations are estimated to have diverged at least a few million years ago (*15*), suggesting ancient, stable niche partitioning. That they have different alleles of genes associated with environmental interactions, carry a distinct set of flexible genes, and differ in relative abundance profiles as the environment changes suggests strongly that they are ecologically distinct.

Enormous population sizes and immense physical mixing probably played a role in the evolution of diverse genomic backbones in *Prochlorococcus*. A simple fluid mechanics model bridging the micrometer and kilometer scales for a typical ocean suggests that just-divided cells will be centimeters apart within minutes, tens of meters apart within an hour, and a few kilometers apart within a week (*15*). Thus, *Prochlorococcus* populations are expected to be well mixed over large water parcels (~10 $km^2$ area by 3 m depth) on ecologically relevant time scales (~1 week) (*15*). This mixing and a stable collective *Prochlorococcus* population density of $10^7$ to $10^8$ cells $liter^{-1}$ (*17*) make the size of each backbone subpopulation in such parcels enormous (>$10^{13}$ cells) (*15*). The effective population size is arguably close to this census population size (*15*), implying that *Prochlorococcus* evolution is governed by selection, not genetic drift [based on population genetics theory (*26*)]. Consistent with this argument, the difference in the observed $F_{ST}$ distribution from that estimated for no selection (Fig. 3B) provides further evidence that the differentiation of genomic backbones in *Prochlorococcus* is a product of selection (*15*).

The correlation between phylogeny and flexible gene content (Table 1, tables S1 and S13, and fig. S5) leads us to propose that the emergence of a genomic backbone is initiated by the acquisition of a beneficial flexible gene cassette, followed by slow fine-adjustment of the core gene alleles to the new niche dimension afforded by the acquired cassette. Given the huge effective population size, even extremely weak fitness differentials among alleles (*27*) can facilitate fine-adjustment of core genes (*15*) over the millions of years of evolution after divergence.

The diverse set of hundreds of subpopulations with distinct genomic backbones probably plays an important role in the dynamic stability of the *Prochlorococcus* "collective" in the global oceans (fig. S6). Small fitness differentials, niche differentiation, and selective phage and grazer predation, in the context of temporal and spatial environmental variation, help to explain their coexistence (*28*, *29*). On seasonal time scales, the *Prochlorococcus* collective maintains a relatively stable population size through temporal and local adjustments in the relative abundance of backbone subpopulations (Figs. 1C and 4A and fig. S6D). On longer time scales (decades to millions of years), the collective may respond to shifting selective pressures through the exchange of gene cassettes between and within backbone subpopulations, and through the evolution of the backbones themselves. The coherence of the collective population holds as long as subpopulations do not diverge to the point where they are no longer able to exchange flexible genes and backbone extinction and emergence rates are relatively balanced. If *Prochlorococcus* backbone subpopulations were designated as distinct species (*30*), it would imply that the global collective is an assortment of thousands of species. It is likely that such a large set of coexisting subpopulations with distinct genomic backbones is a characteristic feature of free-living bacterial species with very large population sizes living in highly mixed habitats.

### References and Notes


1. F. Partensky, W. R. Hess, D. Vaulot, *Microbiol. Mol. Biol. Rev.* **63**, 106–127 (1999).
2. L. R. Moore, G. Rocap, S. W. Chisholm, *Nature* **393**, 464–467 (1998).
3. Z. I. Johnson *et al.*, *Science* **311**, 1737–1740 (2006).
4. G. C. Kettler *et al.*, *PLOS Genet.* **3**, e231 (2007).
5. J. Grote *et al.*, *MBio.* **3**, e00252-12 (2012).
6. D. E. Hunt *et al.*, *Science* **320**, 1081–1085 (2008).
7. S. L. Simmons *et al.*, *PLOS Biol.* **6**, e177 (2008).
8. H. Cadillo-Quiroz *et al.*, *PLOS Biol.* **10**, e1001265 (2012).
9. A. Gonzaga *et al.*, *Genome Biol. Evol.* **4**, 1360–1374 (2012).
10. R. T. Papke *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14092–14097 (2007).
11. S. Rodrigue *et al.*, *PLOS ONE* **4**, e6864 (2009).
12. T. Kalisky, P. Blainey, S. R. Quake, *Annu. Rev. Genet.* **45**, 431–445 (2011).
13. R. Stepanauskas, *Curr. Opin. Microbiol.* **15**, 613–620 (2012).
14. R. S. Lasken, *Nat. Rev. Microbiol.* **10**, 631–640 (2012).
15. Materials and methods are available as supplementary materials on *Science* Online.
16. A. F. Michaels *et al.*, *Deep Sea Res. Part I* **41**, 1013–1038 (1994).
17. R. R. Malmstrom *et al.*, *ISME J.* **4**, 1252–1264 (2010).
18. O. Wurtzel, M. Dori-Bachash, S. Pietrokovski, E. Jurkevitch, R. Sorek, *PLOS ONE* **5**, e15628 (2010).
19. M. Mühling, *Environ. Microbiol.* **14**, 567–579 (2012).
20. M. Nei, in *Human Genetics, Part A: The Unfolding Genome*, B. Bonné-Tamir, T. Cohen, R. M. Goodman, Eds. (Alan R. Liss, New York, 1982), p. 167.
21. R. Mehra-Chaudhary, J. Mick, L. J. Beamer, *J. Bacteriol.* **193**, 4081–4087 (2011).
22. S. Avrani, O. Wurtzel, I. Sharon, R. Sorek, D. Lindell, *Nature* **474**, 604–608 (2011).
23. J. Pernthaler, *Nat. Rev. Microbiol.* **3**, 537–546 (2005).
24. F. Malfatti, F. Azam, *Aquat. Microb. Ecol.* **58**, 1–14 (2009).
25. U. Dobrindt, B. Hochhut, U. Hentschel, J. Hacker, *Nat. Rev. Microbiol.* **2**, 414–424 (2004).
26. J. F. Crow, M. Kimura, *An Introduction to Population Genetics Theory* (Harper & Row, New York, 1970).

27. R. D. Barrett, D. Schluter, *Trends Ecol. Evol.* **23**, 38–44 (2008).
28. A. D. Barton, S. Dutkiewicz, G. Flierl, J. Bragg, M. J. Follows, *Science* **327**, 1509–1511 (2010).
29. F. Rodriguez-Valera *et al.*, *Nat. Rev. Microbiol.* **7**, 828–836 (2009).
30. C. C. Thompson *et al.*, *Microb. Ecol.* **66**, 752–762 (2013).

**Acknowledgments:** We thank S. Itzkovitz, P. H. R. Calil, D. Sher, R. Milo, P. M. Berube, A. P. Yelton, R. Braakman, and particularly M. F. Polz for comments on the manuscript. We thank the Bermuda Atlantic Time-series Study for sample collection, the Bigelow Laboratory Single Cell Genomics Center for single-cell sorting and whole-genome amplification, and the BioMicroCenter facility at MIT for their contributions to the generation of genomic data. N.K. acknowledges the Rothschild Foundation (Yad Hanadiv) and the National Oceanic and Atmospheric Administration "Climate and Global Change" Postdoctoral Research Fellowships. This work was supported by grants to S.W.C. from the NSF Evolutionary Biology Section and Biological Oceanography Section, the NSF Center for Microbial Oceanography Research and Education (C-MORE), the U.S. Department of Energy (DOE)–GTL, and the Gordon and Betty Moore Foundation Marine Microbiology Initiative; to R. Stepanauskas from the NSF Biological Oceanography Section; and to R.R.M from the DOE (contract number DE-AC02-05CH11231). Genomic data have been deposited in National Center for Biotechnology Information GenBank under accession numbers KJ477896 to KJ479276 and JFKN00000000 to JFOE00000000. Additional data files have been deposited to Dryad (doi:10.5061/dryad.9r0p6).

**Supplementary Materials**
www.sciencemag.org/content/344/6182/416/suppl/DC1
Materials and Methods
Figs. S1 to S21
Tables S1 to S13
References (*31*–*91*)
Data S1

18 November 2013; accepted 20 March 2014
10.1126/science.1248575

# Structure-Guided Transformation of Channelrhodopsin into a Light-Activated Chloride Channel

**Andre Berndt,[1]* Soo Yeun Lee,[1]* Charu Ramakrishnan,[1] Karl Deisseroth[1,2,3]†**

**Using light to silence electrical activity in targeted cells is a major goal of optogenetics. Available optogenetic proteins that directly move ions to achieve silencing are inefficient, pumping only a single ion per photon across the cell membrane rather than allowing many ions per photon to flow through a channel pore. Building on high-resolution crystal-structure analysis, pore vestibule modeling, and structure-guided protein engineering, we designed and characterized a class of channelrhodopsins (originally cation-conducting) converted into chloride-conducting anion channels. These tools enable fast optical inhibition of action potentials and can be engineered to display step-function kinetics for stable inhibition, outlasting light pulses and for orders-of-magnitude-greater light sensitivity of inhibited cells. The resulting family of proteins defines an approach to more physiological, efficient, and sensitive optogenetic inhibition.**

The microbial opsins (*1*–*3*) used for optical control of genetically targeted cellular activity (*4*–*7*) include light-activated proton and $Cl^-$ pumps and the cation channels called channelrhodopsins (ChRs). ChRs are derived from algae (*3*, *8*–*10*) and, when expressed in neurons, can elicit precise action potential (AP) firing (*11*–*15*). ChRs conduct $K^+$, $Na^+$, protons, and $Ca^{2+}$ (*3*, *10*, *16*, *17*); because of this nonselective cation-conductance, ChRs display reversal potentials ($V_{rev}$) near 0 mV under physiological conditions and therefore depolarize neurons, leading to AP generation (*18*).

Direct light-triggered inhibition of neuronal activity is possible with inward-pumping $Cl^-$-transporting opsins and outward-pumping proton-transporting opsins (*10*); hyperpolarization to –150 mV or beyond can be achieved (*18*–*20*). However, pumps are inefficient in neural systems because only one ion is moved per photon and no input resistance decrease is elicited (failing to recruit the most potent mechanism of spiking inhibition). Moreover, because the pumps use energy to transport ions against electrochemical gradients, the creation of abnormal gradients is more likely (*18*). Last, pumps cannot take advantage of certain molecular engineering opportunities to achieve light sensitivity and long-term photocurrent stability enhanced by many orders of magnitude (but which depend on formation of a transmembrane pore) (*21*–*23*). Therefore, the creation of inhibitory channels has long been a central goal of optogenetics.

Given typical ion balance in neural systems, identification or creation of light-activated $K^+$ or $Cl^-$ channels could give rise to inhibitory optogenetic tools. ChRs can be engineered to alter kinetics, spectrum, and selectivity among cations (*10*, *24*, *25*). However, $V_{rev}$ has not been shifted sufficiently for nondepolarizing spike inhibition in neurons. We have designed a family of ChRs for $Cl^-$ permeability and capability to inhibit APs without depolarizing neurons to or beyond the AP-generation threshold.

Building on the high-resolution crystal structure of the ChR chimera C1C2 (*24*), we noted that the ion-selectivity pore of ChR is less ordered as compared with the well-defined symmetry of tetrameric $K^+$-selective channels such as KcsA and NaK2K (*26*–*31*). Therefore, we speculated that the specific cation selectivity of ChR is rather a result of negative electrostatic potential surrounding the pore and vestibule; for instance, the C1C2 structure shows seven glutamates framing the conduction pathway (*24*). We hypothesized that systematic replacement of such residues within or close to the pore according to structure-guided electrostatic modeling could reverse this polarity and create an inhibitory ChR, if it were possible to maintain proper protein folding, membrane expression, optical activation, and pore gating.

We initiated a broad structure-guided screen by introducing single site-directed mutations into C1C2 (Fig. 1A). We expressed all variants in cultured rat hippocampal neurons and tested photocurrents using whole-cell patch-clamp so as to ensure proper function in neurons (external/internal [$Cl^-$], 147 mM/4 mM). We quantified stationary photocurrent amplitudes across a range of holding potentials (Fig. 1B), with particular attention to $V_{rev}$, in order to identify permeability variants (Fig. 1C). C1C2 exhibits $V_{rev}$ of –7 mV under these conditions, which is typical for nonspecific cation channels (*16*, *26*, *32*, *33*). Certain mutations with powerful effects on $V_{rev}$ displayed concomitant adverse effects on photocurrent sizes (such as E136R and E140K) (Fig. 1B), and were not studied further (*34*). More promising mutations, such as N297Q and H173R, exhibited both potent currents and altered $V_{rev}$ (Fig. 1C) and were combined in a series of increasingly integrated mutations. The fivefold mutation T98S/E129S/E140S/E162S/T285N and fourfold mutation V156K/H173R/V281K/N297Q both displayed prominently-shifted $V_{rev}$ (in the range of –40 mV) while maintaining functionality (Fig. 1, D and E).

We next combined these constructs to generate a ninefold mutated variant with contiguous shifts in expected electrostatic potential distribution (Fig. 2A and fig. S1) (*24*). We expressed the ninefold variant in human embryonic kidney (HEK) 293 cells to test both $V_{rev}$ and permeability under controlled ion composition and optimized voltage clamp settings (Fig. 2B). We mapped photocurrents over a broad range of membrane potentials (Fig. 2C) (from –75 mV to +55 mV) (*35*). Under these conditions (external/internal [$Cl^-$], 147mM/4mM), the combined ninefold mutation exhibited $V_{rev}$ of –61 mV, which is far more negatively shifted than was the C1C2 backbone or either parental 4× or 5× construct (Fig. 2D). Despite this major change in functionality, both peak and stationary photocurrents remained fast and robust (predicting suitability for optogenetics, especially because this channel could also recruit a reduced-membrane resistance mechanism for spiking inhibition), and the original blue light–activation spectrum of C1C2

[1]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. [2]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94305, USA. [3]Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA.

*These authors contributed equally to this work.
†Corresponding author. E-mail: deissero@stanford.edu