

How many ecotypes make up the world's most abundant photosynthetic organism?

Maike Morrison, Mark Kirkpatrick | Department of Integrative Biology | maike.morrison@utexas.edu

Background

- The world's most abundant photosynthetic organism is a blue-green alga called *Prochlorococcus marinus* (Fig 2, 3) which has a global population size of 10^{27} and is responsible for about 25% of global primary production¹.
- As with other microorganisms, ecotypes (genetically distinct variants within a species) in *Prochlorococcus* are conventionally defined by arbitrary thresholds of genetic similarity among individuals.
- We explored the performance of the program Bayesian Phylogenetics and Phylogeography² (BPP) as an alternative ecotype delimitation method for *Prochlorococcus*.
- BPP implements Bayesian inference methods on genomic sequence data under the multispecies coalescent model and is traditionally used for eukaryotic species delimitation.

Research Questions

- How many ecotypes make up *Prochlorococcus*?
- Does BPP perform well on data from a non-recombining, superabundant organism? Are its results robust to perturbations of the initial parameters?

Methods

We used genetic data from 101 *Prochlorococcus marinus* cells, which were published by Kashtan et al³.

The BPP analysis proceeds as follows:

- We assign individuals to small, indivisible populations (each leaf in Fig 1).
- We provide a rooted “guide tree” (Fig 1) which gives an initial phylogeny for these populations.
- BPP uses Markov Chain Monte Carlo (MCMC) algorithms that consider jumps to different guide tree topologies and ecotype delimitations.
- BPP outputs posterior probability distributions for several results, such as the total number of ecotypes and individual ecotype delimitations.

We used BPP to separately analyze both the ITS rRNA sequences and the partial genome data published by Kashtan et al³. The results discussed here used Figure 1 as a guide tree and assigned populations by dividing each blue clade in half and assigning all other lineages to their own populations.

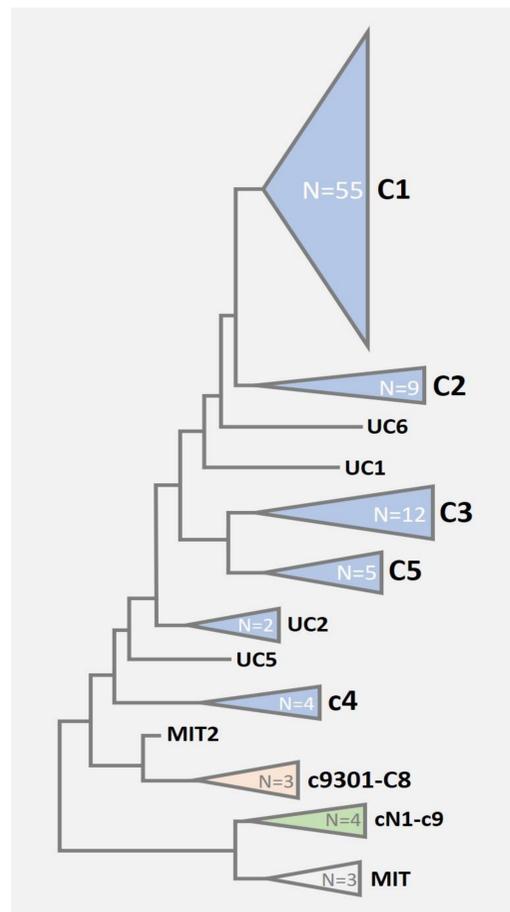


Figure 1 - Schematic representation of the phylogenetic relationships among *Prochlorococcus* clades. Triangles represent clades with 99% genetic similarity in the ITS sequence. The number of samples in each clade is written. Blue clades and “UC” cells are from the cN2 ecotype, orange is from the c9301 ecotype, and green is from the cN1 ecotype. Samples labeled “MIT” are cultured strains.

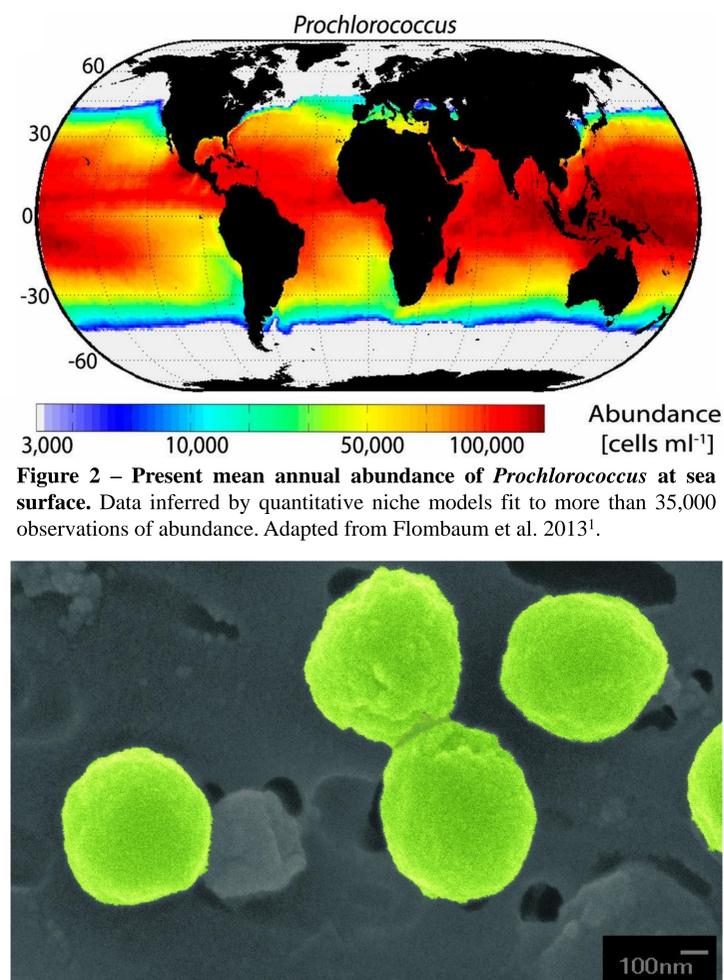


Figure 2 – Present mean annual abundance of *Prochlorococcus* at sea surface. Data inferred by quantitative niche models fit to more than 35,000 observations of abundance. Adapted from Flombaum et al. 2013¹.

Figure 3 – Colorized scanning electron microscope image of *Prochlorococcus* MIT9215. A sample from this strain is included in our analysis—specifically in the grey MIT clade depicted in Figure 1. Photo taken by Anne Thompson, Chisholm Lab, MIT.
<https://www.flickr.com/photos/prochlorococcus/3287700189/in/album-72157613990334681/>

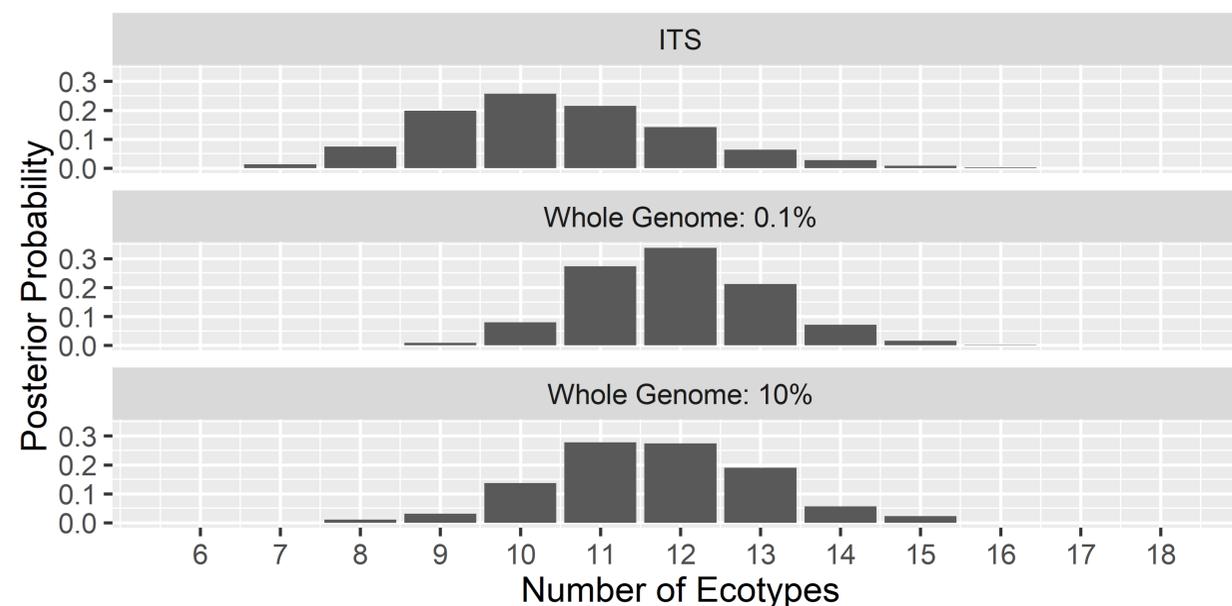


Figure 4 - Posterior probability distributions for the number of ecotypes delimited by BPP under our three primary analyses. Our BPP analysis using only the intergenic transcribed spacer (ITS) rRNA sequence data (549 bp) yielded a right-skewed distribution centered at about 10 ecotypes. However, when we used 0.1% (1,627 bp) and 10% (162,677 bp) of the whole genome data, the distributions are more symmetric and are both centered at about 12 ecotypes.

Results

- Ecotype delimitation results are largely robust to drastic perturbations of the BPP framework, such as shuffling the guide tree and changing the prior distribution for the number of ecotypes.
- When analyzing ITS data, BPP consistently merged all divided clades.
- When analyzing partial genome data, BPP was slightly less consistent in merging the divided clades (C1 and C2 were consistently reunified; C3 and C5 were consistently split).
- This discrepancy leads to a posterior distribution of ecotype counts that shifts to the right when the analysis uses partial genomic data instead of ITS data (Fig 4).

Conclusions

- There is evidence for many independently evolving lineages within *Prochlorococcus*—perhaps as many as 12 to 13.
- BPP is a viable tool for ecotype delimitation within a superabundant, non-recombining organism. Its results are largely robust to perturbations of initial parameters, such as the guide tree.

Acknowledgments:

Thank you to Dr. Mark Kirkpatrick and Michael Miyagi for their collaboration, advice, and encouragement.

References

1. Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincon, J., Zabala, L. L., Jiao, N., ... Martiny, A. C. (2013). Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences*, 110(24), 9824–9829. <https://doi.org/10.1073/pnas.1307701110>
2. Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, 61(5), 854–865. <https://doi.org/10.1093/czoolo/61.5.854>
3. Kashtan, N., Roggensack, S. E., Rodrigue, S., Thompson, J. W., Biller, S. J., Coe, A., ... Chisholm, S. W. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science*, 344(6182), 416–420. <https://doi.org/10.1126/science.1248575>