Analysis of large-scale patterns in phytoplankton diversity and community structure across the Pacific and Atlantic Oceans

Sophie Clayton (sclayton@uw.edu)

Project summary

Microscopic algae (called phytoplankton) form the base of the oceanic food chain, and are key players in the biogeochemical cycles of many climatically-active elements. Ecological theory predicts that diverse ecosystems are more stable, i.e. more resistant to stressors, than less diverse ecosystems. However data on the diversity of oceanic phytoplankton communities is very sparse as it typically depends on very labor-intensive methods (e.g. microscope identification, molecular sequencing). In order to understand how phytoplankton diversity may be affected by climate change, it is essential to have a baseline understanding of current patterns in diversity and how they relate to environmental conditions.

In this study, we will calculate indices of phytoplankton diversity using data collected using SeaFlow, a continuously sampling underway flow cytometer. This will produce diversity estimates at high resolution over large spatial scales, and across different seasons. We will adapt Li's (1997) cytometric diversity to better reflect the taxonomic diversity of phytoplankton observed with SeaFlow, and develop methods for integrating data from different instruments and cruises in such a way that they are comparable. Using data from the Pacific and Atlantic Oceans collected during 18 oceanographic cruises, we will conduct a meta-analysis of the patterns in cytometric diversity, and how these relate to other biotic and abiotic variables (e.g. temperature, salinity, density gradients, biomass).

Data

The data to be used for this study comes from 18 oceanographic cruises conducted over the last 4 years. It was collected underway using a continuous flow cytometer (SeaFlow) which measures the forward scatter and fluorescence of small phytoplankton cells in 3 minute increments. This consists of scatter, fluorescence and population assignment data for 1.7B particles measured over the course of the cruises. This data is contained in >50,000 binary .opp and .vct files. Concurrent measurements of the location and the ocean temperature and salinity from the ships' underway instruments are contained in .sds files. This gives a total of about 250 GB of data. All of the flow cytometry data from .opp and .vct files has already been uploaded to the UW's Myria database. The data from the .sds files is currently in the form of .csv files for each cruise, and will be uploaded to the Myria soon.

Key science questions

This project will explore the controls on phytoplankton diversity on an unprecedentedly large scale. Success will rely on developing a robust method for filtering data into subsets that are in some way comparable across cruises, or standardizing data from different cruises to make them comparable.

- Can we identify physical controls on patterns in diversity? More specifically, do regions of high diversity correspond to frontal zones (i.e. regions with strong horizontal density gradients)?
- What is the relationship between diversity and biotic factors (e.g. biomass)? Does it follow predictions from numerical models and ecological theory?

Publications

The main focus of the project will be a publication on the large-scale patterns in community structure and diversity across the different oceanic regimes reflected in the data set. We would also envision writing up a methods paper, which could be aimed at an oceanographic journal (e.g.

L&O Methods) or as a case study for a database/computer science workshop/journal. This would highlight the power of using database methods for analyzing large datasets in biological oceanography.

Key technical challenges

The technical challenges that we encounter in this project are of different kinds.

- Variable data quality. Data was collected on different cruises with three different SeaFlow instruments and different settings/sensitivity. Beads are used as an internal standard, and will be used to standardize data across cruises and filter the data down to a subset that can be used for the meta-analysis.
- Scalability of the method developed, which should be applicable to any other dataset of flow cytometry data to produce estimates of cytometric diversity.
- Multivariate analysis of the relationship between biotic and abiotic variables, as well as gradients in these properties. We expect relationships between variables, but they may not be generally applicable across all seasons or regions.

Timeline

- Fall 2014: Finish uploading all of the relevant data to the Myria database. Finalize quality control/standardization of files. Calculate cytometric diversity and conduct sensitivity analysis. Statistical analysis of the relationships between the diversity indices and biotic and abiotic variables.
- Winter 2015: Finalization of the data analysis and drafting of the main paper and methods paper.
- Spring 2015: Submit both finished papers.