

The Environmental Niche Atlas: Global Mapping of Microbial Functions (#3300)

Katherine S. Pollard (lead) and Jonathan A. Eisen (co-lead)

PROGRESS REPORT – JANUARY 15, 2013

Research Progress

Aim 1. Build automated tools to characterize protein families from metagenomic sequencing data.

SFams Protein Family Database

- We designed, implemented, and released **SFams**, a protein family relational database tailored to metagenomic sequence analysis (see publication for details). This resource is being used by several prominent environmental microbiology labs.
- Designed automation scripts to efficiently update SFams twice a year.
- Established a protocol for automatic identification of phylogenetic marker families (based on universality and evenness). Identified 40 universal marker families for the domains of Archaea and Bacteria, and 74 Bacteria-specific markers.

Metagenomic Read Classifier (MRC)

- Developed and implemented **MRC**, a metagenomic read classifier that quantifies the protein functions (e.g., SFams) present in a community.
- Conducted simulation study to assess the performance of MRC with different classification algorithms and experimental parameters. Found that short reads can be accurately classified (Figure 1). The BLAST and LAST algorithms are more sensitive and precise than profile-HMM alignment, in contrast to published results for full-length gene sequences.

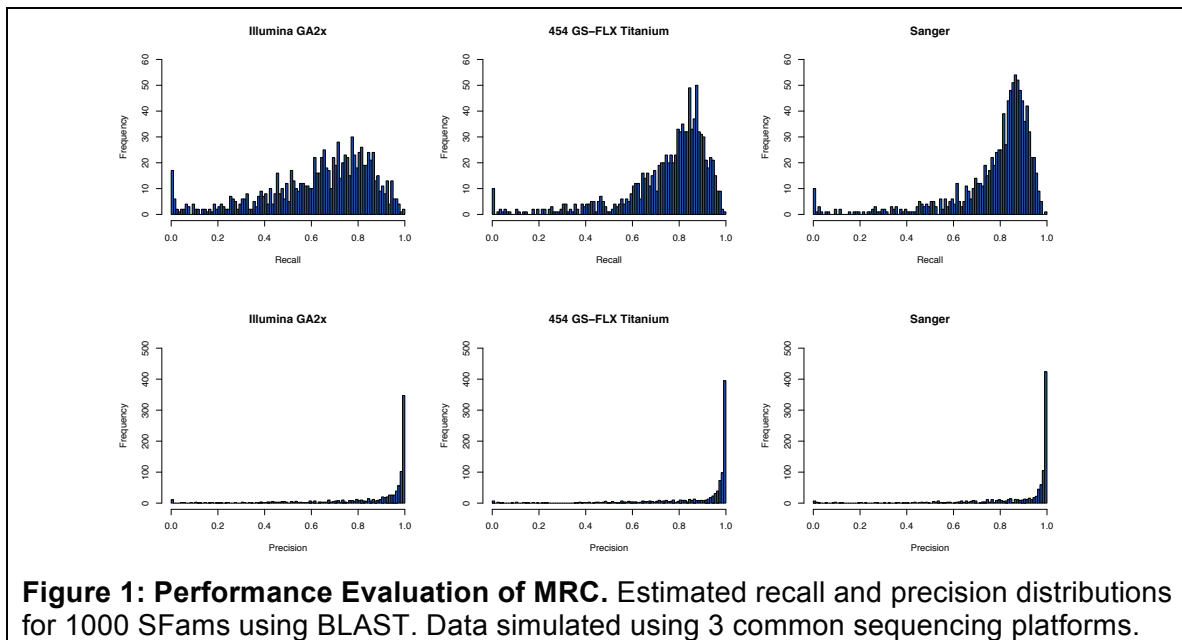


Figure 1: Performance Evaluation of MRC. Estimated recall and precision distributions for 1000 SFams using BLAST. Data simulated using 3 common sequencing platforms.

Aim 2. Predict niches of protein families from environmental data.

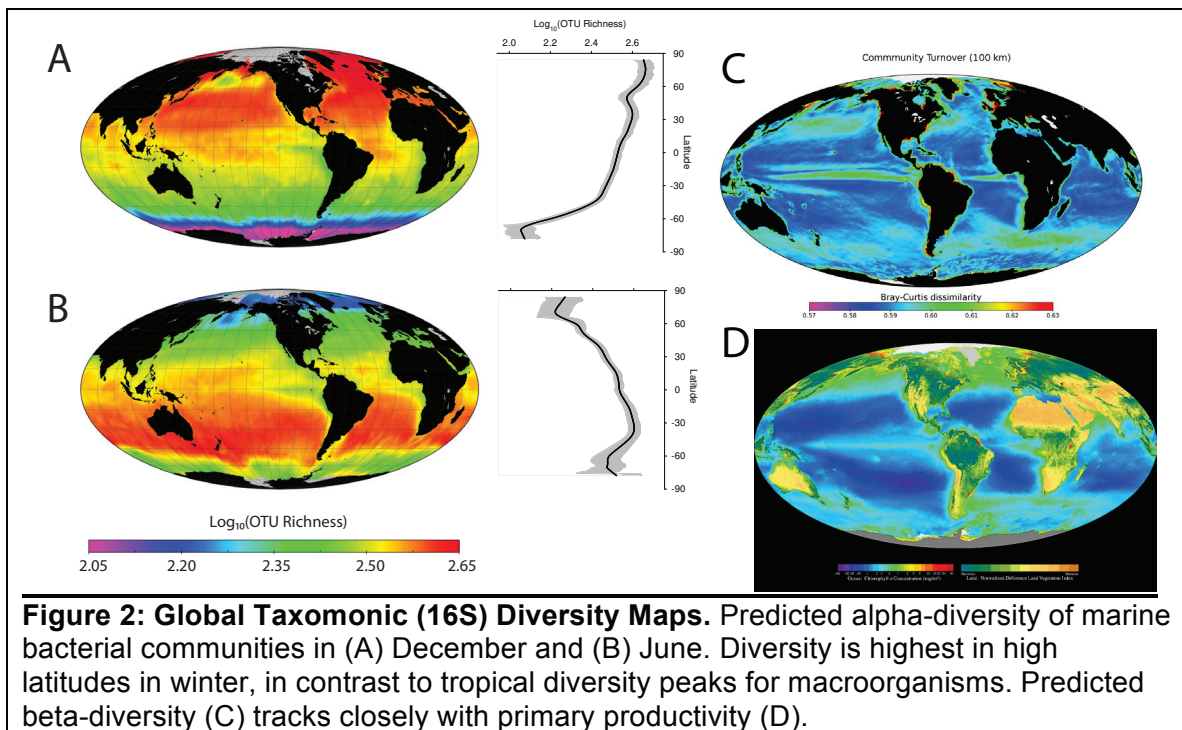
Niche Mapping

- Implemented software for taxonomic and functional alpha-diversity (richness) prediction using niche modeling.
- Extended this approach to quantify beta-diversity (turnover).

- Developed methods for drawing diversity heat maps.
- Designed a clustering approach to locate ecoregions (i.e., similar communities) based on alpha- and beta-diversity maps.

Aim 3. Analyze niche maps to characterize global functional diversity.

- Conducted a search for all publicly available marine metagenomic datasets. Downloaded these to the project server.
- Identified a paucity of global shotgun metagenomics data. Working with GBMF to identify cruises and investigators that are generating this data.
- Proceeded to apply our methods to available data:
 - **SFams/MRC**: Processed the Global Ocean Survey metagenomic libraries. Found that our procedure more thoroughly characterizes the functional profiles of these communities than do other commonly used resources and approaches.
 - **Niche Mapping**: Predicted global marine bacterial taxonomic diversity using the MICROBIS data set, collected as part of the ICoMM sampling initiative, which consists of bacterial 16S rDNA sequences from approximately 230 sampling locations around the global ocean. This data enables taxonomic, but not functional, diversity prediction (Figure 2).
- Investigated the possibility of conducting functional (protein family) diversity studies by extrapolating from 16S data using sequenced genomes. Morgan Langille (former iSEEM member) implemented PI-CRUSTS, software for this purpose. We may employ PI-CRUSTS if we do not get access to shotgun metagenomic data sets in the next year.



Publications

Directly supported by this grant.

- Sharpton TJ, Jospin G, Wu D, Langille MG, Pollard KS, Eisen JA (2012). Sifting through genomes with iterative-sequence clustering produces a large, phylogenetically diverse protein-family resource. **BMC Bioinformatics** 13: 264. PMID: 23061897. <http://www.biomedcentral.com/1471-2105/13/264/>
- Kembel SW, Wu M, Eisen JA, Green JL (2012). Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. **PLoS Computational Biology** 8(10): e1002743. PMID: 23133348. <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1002743>
- Eisen JA. Phylogenetic and phylogenomic approaches to analysis of microbial communities. In "The Social Biology of Microbial Communities – A Report from the National Academy of Sciences Forum on Microbial Threats." [PDF](#).

Supported in part by this grant but also supported by other funds.

- Jiang X, Langille MGI, Neches RY, Elliot MA, Levin SA, Eisen JA, Weitz JS, Dushoff J. 2012. Functional biogeography of ocean microbes: dimension reduction of metagenomic data identifies biological patterns across scales. **PLoS ONE** 7(9): e43866. doi:10.1371/journal.pone.0043866. PMID: 23049741.
- Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. **Proc Natl Acad Sci U S A. Ahead of print 12/31**. [PDF](#).

Talks

- *July 2012*: Katie Pollard – Invited talk at the Joint Statistical Meetings, San Diego, CA.
- *July 2012*: Jonathan Eisen. *Phylogenomic approaches to functional prediction*. At AFP "Automated functional prediction" submeeting during ISMB 2012. [Slideshow on Youtube](#). [Slideshare](#).
- *September 2012*: Katie Pollard – Seminar at the VanBUGS cross-institutional bioinformatics group, Vancouver, BC.
- *September 2012*: Jonathan Eisen. *Phylogenomic approaches to the study of microbial diversity*. Keynote Talk at Lake Arrowhead Microbial Genomes Meeting. [Slidehsare](#). [Slideshow on Youtube](#).
- *September 2012*: Jonathan Eisen. *Phylogenomic approaches to the study of microbial diversity*. At Bay Area Illumina Users Meeting. [Slideshare](#). [Slideshow on Youtube](#).
- *October 2012*: Tom Sharpton – Seminar at the Center for Genome Research and Biocomputing, Oregon State University.
- *October 2012*: Jonathan Eisen. *Sequencing and Microbes, the Now Generation*. At UC Davis Food for Health Symposium. [Slideshow on Youtube](#).
- *November 2012*: Katie Pollard – Invited talk at Systems Biology Symposium, Johns Hopkins Medical School, Baltimore, MD.