How to sequence a glacier: a computational biologist's rough guide

Sulbha Choudhari, Ruchi Lohia, Andrey Grigoriev

Rutgers University, 315 Penn St, Camden NJ 08055, USA, agrigoriev@camden.rutgers.edu

This talk will describe the motivation behind sequencing a glacial environmental sample, challenges in collecting and sequencing it as well as the insights from the analysis of the results and comparisons with other metagenomes.

Glaciers are the beacons of the global climate change and the simplicity of glacial life makes studying its metagenome attractive. Glacial ecosystems comprise relatively few microbial inhabitants (e.g., algae, bacteria, fungi, archaea); some temperate glaciers along the Pacific Coast of North America support glacier ice worms (Annelida: Enchytraeidae), snow Collembola (Arthropoda) and possibly representatives from a handful of other animal phyla. Culturing ~1 ml of surface ice collected from Byron Glacier, AK identified a mixture of ~10 single-celled microbes (algae, bacteria, fungi), most of which were undescribed species (1). One can thus expected to collect with a metagenomics approach a nearly complete set of species in a full minimal ecosystem with representatives of all domains of life. Other estimates also suggested relative simplicity of glacial life. An early report (2) listed 354 algal and cyanobacterial species, 77 fungal species, and 35 bacterial species that occur in snow. A direct metagenomic analysis of a glacier in German Alps (3) identified 72 bacterial OTUs (operational taxonomic units) at 97% identity cut-off. Comparing these numbers to many metagenomics projects currently underway, which suffer from high complexity of the studied environments, it seems feasible to characterize and model the ecosystem in its entirety (in contrast to, say, much richer soil samples).

Harsh environment represents a challenge for human collection of samples but there is no shortage of material: glacier ice water is not a limiting factor and glacier ice is a suitable medium for large scale DNA/RNA extraction and metagenomics following its concentration by passage over a suitable filter. We utilized this approach for our initial characterization of a prokaryotic subset of the ecosystem (4). Snow and ice was collected from an avalanche cone

at elevation 154 m above the sea level (snow depth 2 m) and filtered with a 0.22 μm filter to remove larger particles. Following the amplification of the 16S rRNA V4 region (5) and sequencing on an Illumina HiSeq platform in the Earth Microbiome Project (6), we obtained a total of ~140,000 reads of 151 bp in length (~25,000 reads with Phred score ≥20).

Strikingly, our analysis identified a much richer and more diverse prokaryotic community than previously estimated. At the species level (97% identity) we have detected almost 2,500 OTUs, yet we have not reached species saturation. Similar to many other metagenomic studies, most of the diversity is contributed by uncultured species and the largest phylum is Proteobacteria (40%). On the other end of the spectrum, among the rare representatives we also identified archaeal 16S sequences, which so far had escaped detection in glaciers of the Nothern Hemisphere. We performed extensive comparisons to other related metagenomes using diversity ratios, word frequencies and principal component analysis and detected intriguing patterns suggesting metagenome-specific biases.

We will further discuss how the design of metagenomic experiments may influence their results and potential biases in the data. We will also describe the follow-up work on addressing the energy metabolism in cold environments by means of the genome and pathway analysis.

A.G. acknowledges support of the NSF grant DBI-1126052.

1. M.J. Napolitano, D.H. Shain**.** (2004) Four kingdoms on glacier ice: convergent energetic processes boost energy levels as temperatures fall, *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271:**S273-S276.
2. E. Kol**.** (1968) *Kryobiologie: Biologie und Limnologie des Schnees und Eises. 1: Kryovegetation*, vol. 24. (Schweizerbart'sche (Nagele und Obermiller), Stuttgart).
3. C. Simon, et al. (2009) Phylogenetic Diversity and Metabolic Potential Revealed in a Glacier Ice Metagenome, *Applied and Environmental Microbiology* **75:**7519-7526.
4. S. Choudhari, et al. (2013) Metagenome Sequencing of Prokaryotic Microbiota Collected from Byron Glacier, Alaska, *Genome announcements (in press)*.
5. J.G. Caporaso, et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms, *ISME J* **6:**1621-1624.
6. J.A. Gilbert, et al. (2010) Meeting report: the terabase metagenomics workshop and the vision of an Earth microbiome project, *Standards in genomic sciences* **3:**243.