Multigene environmental DNA data analysis for New Zealand genomic observatory

Alexei Drummond [1], Dong Xie [2], Department of Computer Science, Andrew Dopheide [3] School of Biological Sciences, NZGO team [4] Genomicobservatory, University of Auckland

New Zealand genomic observatory

The project aims to produce a comprehensive phylogenetic and environmental characterization of the terrestrial species in a well-defined New Zealand model ecosystem using modern sequencing, informatics, niche modelling and field ecology approaches. The project is a collaboration between several universities and research centres in New Zealand and Australia, the Department of Conservation, Auckland Council, and iwi, and internationally as part of the Network of Genomic Observatories. It also aims to provide a long-term research program structure for collaborative, interdisciplinary research projects at the intersection of ecology, evolutionary biology and genomics.
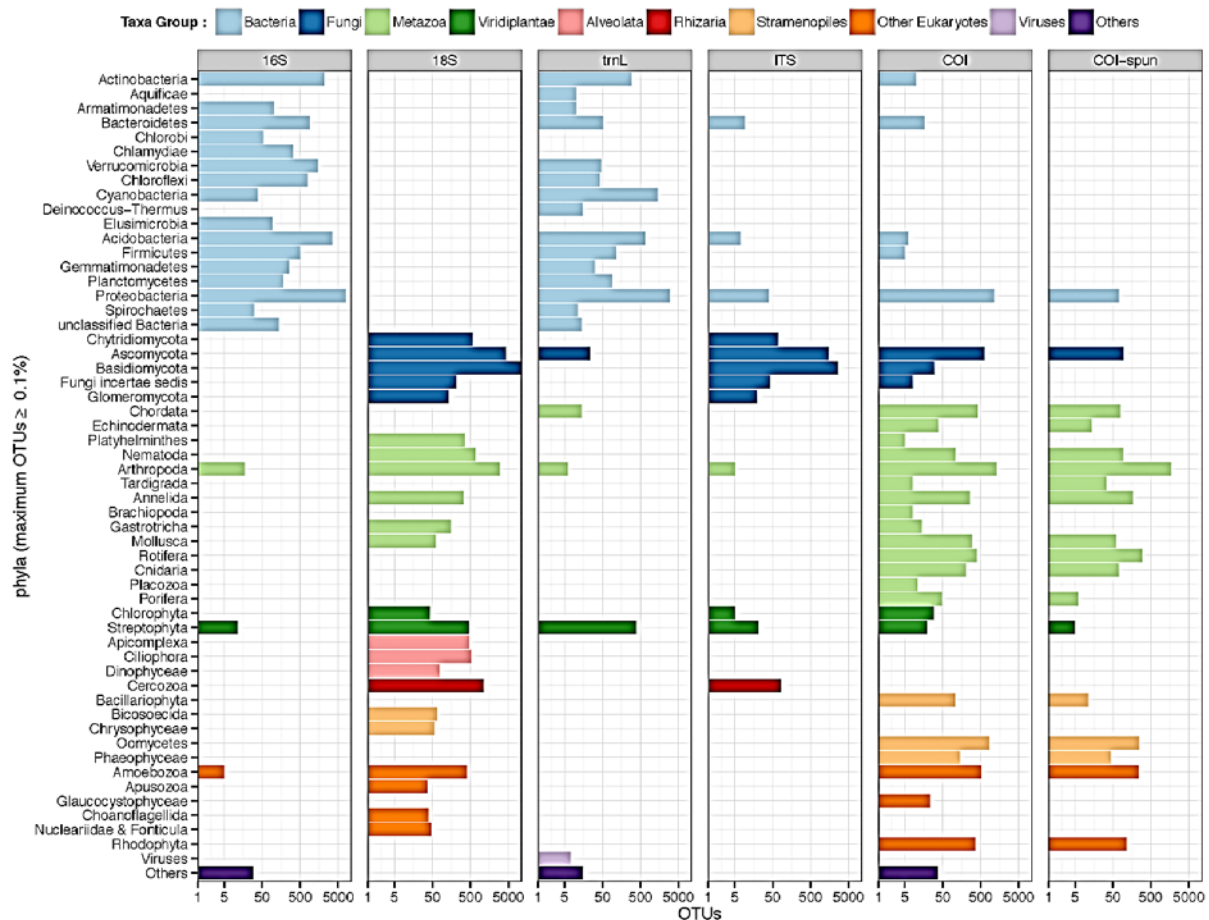
The percentage of OTUs at the 97% clustering threshold assigned to phyla. Unclassified OTUs, OTUs containing low-complexity sequences, and OTUs from phyla that are represented by less than 0.1% of the OTUs are grouped into the "Others" category.

Centre for eResearch
University of Auckland
Private Bag 92019
Auckland 1142
New Zealand

Physical address:
Room G021, Building 409
24 Symonds Street
Auckland 1010, New Zealand
www.eresearch.auckland.ac.nz

Contact:
Professor Mark Gahegan,
Director
m.gahegan@auckland.ac.nz
DDI +64 9 9231891

Marcus Gustafsson, Manager
m.gustafsson@auckland.ac.nz
DDI +64 9 923 2231

In the project, we are able to measure broad diversity of eukaryotes from soil using an environmental DNA approach. Environmental DNA (eDNA) approaches typically focus on microbial communities within the soil and tend to use single gene marker regions. Here we evaluate a suite of DNA markers coupled with Next Generation Sequencing (NGS) that span across the tree of life. Sequences analysis, such as Operational taxonomic units (OTUs) identification by molecular markers, taxonomic assignment, and biodiversities estimation, is a main part of this evaluation.

The raw reads file in a FASTQ format was then passed into a UPARSE pipeline (Edgar, 2013) to identify OTUs, which includes quality filtering, length truncation (300 bp), dereplication, abundance sorting, OTU clustering, chimera filtering and mapping OTUs. The outputs of the pipeline were a FASTA file containing OTU sequences and a mapping file between OTUs and reads for each given OTU clustering threshold. The community matrix was created from the mapping file by retrieving the site information added in the sequence label previously, and the matrix described species abundance (OTU counts) according to sampling sites.

Jost's biodiversities (Jost 2006) are respectively calculated regarding community matrices of six eDNA methods using R package *vegetarian* (Charney and Record 2012). Rarefraction curves for diversities are further estimated using a 97% threshold for OTU identification by subsampling the minimum number of OTUs of sampling sites (subplots) using R ecology package *vegan* (Oksanen et al 2013). BLAST+ was used to classify the taxonomy of OTUs, and the classification result was interpreted to taxonomic assignment by phyla.

Project webpage is http://www.genomicobservatory.cs.auckland.ac.nz. and database is https://data.genomicobservatory.cs.auckland.ac.nz.

Contact

Alexei Drummond , Department of Computer Science, University of Auckland. Email: alexei@cs.auckland.ac.nz