

mARS Vision Paper

building the Microbial Antarctic Resource System



December 2013



mARS

creating the Microbial Antarctic Resource System

Vision Paper

Alison Murray¹, Anton Van de Putte², Nabil Youdjou³, Bruno Danis⁴

1: Division of Earth and Ecosystem Sciences, Desert Research Institute, United States of America

2: biodiversity.aq, Royal Belgian Institute of Natural Sciences, Belgium

3: biodiversity.aq, Université Libre de Bruxelles, Belgium

4: Marine Biology Lab, Université Libre de Bruxelles, Belgium

About

As a outcome of the mARS Workshop hosted at the Belgian Science Policy Office (BELSPO, Brussels) in May 2012, the mARS Workshop held during the SCAR Open Science Conference (Portland, OR) in July 2012 and the mARS Workshop hosted at the Université Libre de Bruxelles in December 2013, this Vision Paper is an initial attempt to scope out the Microbial Antarctic Resource System, an information system dedicated to serving the community of Antarctic microbial scientists. It is intended to prepare the grounds for the deployment of this initiative.

For citation of this work, please see the Licensing section at the bottom of the document.

Scope

The Microbial Antarctic Resources System (mARS) is envisioned as an information system dedicated to facilitate the discovery, access and analysis of geo-referenced, molecular microbial diversity (meta)data generated by Antarctic researchers, in an Open fashion. The scope of diversity will encompass all free-living and host-associated virus, Bacteria, Archaea, and singled-celled Eukarya.

mARS focuses on past, present and future works. It offers a community-driven platform for scientists to publish, document, analyse and share their (meta)data with the broad community for science, conservation and management purposes, in the spirit of the Antarctic Treaty.

mARS is composed of interoperable modules, iteratively building the microbial component of the biodiversity.aq infrastructure.

To the best extent possible, the wishes of the community regarding mARS functionalities will be reflected in the flexibility and scalability of the system. Feedback is expected from the users community in order to align their needs and the functionalities of mARS.

The development of mARS arises as a legacy of the International Polar Year (IPY).

Steps

At this point of the design of mARS, 4 incremental steps are envisioned to reach the objectives of the system.

Step 1: Data description and discovery

This step will capture information about molecular microbial diversity research efforts that are being or have been conducted by the Antarctic research

community. The results of step 1 will facilitate communication and collaboration, augment comparative biodiversity studies, and provide a legacy-discoverable resource to advance science, conservation awareness and management. The scope of the information that can be entered in the IPT encompasses present, past, or future studies involving marker gene surveys (e.g. 16S or 18S rRNA, functional genes), or meta “omic” projects from natural samples in Antarctic habitats, enrichment or pure culture efforts.

Using the biodiversity.aq Integrated Publishing Toolkit (IPT) - a user-friendly project metadata entry portal - ensures optimal interoperability with other information systems (e.g. GBIF, OBIS, GenBank) through adoption of data standards such as MiMARKS, and enhanced discoverability by search engines. The IPT allow allows the user to control the visibility of his/her (meta)data.

Step 2: Habitat and Microbial Sequence Metadata Entry *(MiMARKS and mARS Sequence Set Template)*

Secondly, users will be invited to upload habitat and molecular methods-specific (meta)data pertaining to the samples and the related sequencing data (including accession numbers) using standardized templates (respectively MiMARKS Googlesheet and mARS Sequence Set Spreadsheet) accessible on the mARS website. These templates can readily be shared with your collaborators and it works with GenBank submission tools (Sequin and WebIN).

Used together, and uploaded with the corresponding IPT metadata entry (as described in Step 1), these templates will describe geo-referenced physiochemical information that relates to Antarctic microbial diversity studies as well as the matching sequencing information.

Thus, these integrated (meta)data recording efforts can have multiple outcomes and will serve the objective of reporting environmental data to biodiversity.aq. By harboring this information directly at biodiversity.aq, Antarctic scientists, and the global scientific community will have the information archived and accessible through common language queries using the biodiversity.aq data portal. mARS data submissions can also serve the requirements of National Antarctic Programs for data reporting.

Step 3: Georeferenced-molecular sequence database integration

In this step, sequence data files produced by different technologies (e.g. Sanger sequencing, 454, Illumina, Ion Torrent) will be linked back to the relevant entries as described in steps 1 and 2. mARS will provide indexed searching capabilities and geo-server links to DNA sequence data from Antarctic studies that have been deposited in public repositories, providing rapid access to this information through the biodiversity.aq data portal.

There is currently no exhaustive resource that provides this level of information from a geo-referenced perspective. The Antarctic scientific community is actively engaged in molecular microbial diversity and genomic surveys in both terrestrial and marine realms. mARS provides a unique resource to harness this information.

Step 4: Processing batch sequence data

Circum-Antarctic microbial diversity

As the primary mandate of biodiversity.aq is to provide the scientific community access to Antarctic diversity information, biodiversity.aq staff will process the microbial diversity information referenced in mARS for selected, highly used regions of marker genes (for each domain of life) generated through both Sanger sequencing studies and NGS efforts in order to provide the users with a window into the microbial diversity present in Antarctica.

The procedures for data processing will be endorsed by and shared with the scientific community. The results will be accessible using the biodiversity.aq search engine via BLAST, GEOBLAST or taxon search, making sequence-based Antarctic microbial diversity information discoverable through the biodiversity.aq data portal. Currently, the suite of sequence analysis tools enveloped in Mothur are envisioned as the main data processing pipeline for this step and step 4. A full pipeline has been developed following the standard operating procedures available for Mothur and have been tested on different data sets.

The pipeline and output files including alignments, OTUs and non-redundant sequence data sets, OTU occurrence files, taxonomic assignments and diversity statistics etc. will be made available through the biodiversity.aq website for representative data sets (e.g. all Bacteria 454 and Illumina-tag 16S rRNA gene surveys; all eukaryal 18S rRNA 454 and Illumina-tag surveys).

Note that execution of this step is contingent up on resources to support data analyst activities.

It is also envisioned that the [mARS](#) user community who publish new microbial diversity surveys using meta(data) are available in [mARS](#), can make the outputs of such efforts available through the web portal.

Perspectives

The [mARS](#) initiative brings innovative perspectives to Antarctic microbial biodiversity research and its applications. Once [mARS](#) reaches full operability it is envisioned that new research areas in both basic and applied areas will be significantly enabled. For example, biogeography, bioprospecting, environmental impact, species introductions, and climatechange-related studies will be made possible using a data-driven approach accessible through [mARS](#).

Also, [mARS](#) will allow the consolidation of a new community within [SCAR](#) and new perspectives for collaboration within and beyond [SCAR](#). There is also significant potential for expanding the model for genetic work carried out on all organisms, allowing integrated studies on Antarctic biodiversity.

Relevant links

mARS website: <http://mars.biodiversity.aq/>

mARS sequence set template: <https://docs.google.com/spreadsheet/ccc?key=0AsB0418meERPdHpYaGZaMENmQXpVN0xXeVpPeFdIOWc&usp=sharing#gid=0>

mARS howtos: mars.biodiversity.aq/Howto.html

biodiversity.aq (@biodiversityaq):

General website: <http://www.biodiversity.aq>

Integrated Publishing Toolkit: <http://ipt.biodiversity.aq>

Data Portal: <http://data.biodiversity.aq>

RDP Metadata (MiMARKS): <http://rdp.cme.msu.edu/misc/googleSheetsHelp.jsp>

GBIF Genomic biodiversity data: <http://community.gbif.org/pg/groups/22216/genomic-biodiversity-data/>

Initial examples of Metadatasets:

Sanger Sequencing: http://ipt.biodiversity.aq/resource.do?r=andee3_archaeobacterial_communities

Next Generation Sequencing:

[http://ipt.biodiversity.aq/resource.do?](http://ipt.biodiversity.aq/resource.do?r=antarctic_peninsula_bacterioplankton_diversity)

[r=antarctic_peninsula_bacterioplankton_diversity](http://ipt.biodiversity.aq/resource.do?r=antarctic_peninsula_bacterioplankton_diversity)

Culture Collection:

<http://ipt.biodiversity.aq/manage/resource.do?r=culturetest>

Licensing of this work



the Microbial Antarctic Resource System by Murray A, Van de Putte A, Youdjou N, Danis B is licensed under a Creative Commons Attribution 4.0 International License.

Please read and follow the [Polar Information Commons Ethics and Norms of Behavior](#) when contributing or using mARS data.