Ad-hoc GEOTRACES Data Management Meeting: Biological Parameters

WHOI, Woods Hole, MA, USA November 22-23, 2015

PARTICIPANTS

in person: Sara Baldwin Collins (MIT, US), Paul M. Berube (MIT, US), Abby Bull (BODC, UK), Danie Kinkade WHOI EarthCube), Julie LaRoche (U. Halifax, Canada), Maite Maldonado (UBC, Canada) Matt McIlvin (WHOI), Mak Saito (WHOI), Alessandro Tagliabue (U. Liverpool), Benjamin Twining (Bigelow Laboratory, US)
via Skype: Andy Allen (UCSD, US) Philip Boyd (U. Tasmania, Australia), Andrew Bowie (U. Tasmania, Australia)

MINUTES FROM MEETING

Introduction

BioGeotraces was initiated as an effort within Geotraces to complement the "trace elements and their isotopes" data with biological data sets. The biological data sets were carefully selected, aiming for those that help explain trends of distribution and speciation of trace elements in the sea. This effort was initially led by Phil Boyd.

BioGeotraces was also initiated as a program of opportunity: chemists on board of Geotraces cruises volunteer to collect samples for biological parameters. In order to acknowledge the efforts of these volunteers, a database was set up to record a) the kind of biological samples collected, b) the volunteers involved in the collection, and c) the laboratory responsible for the analyses.

In addition, many Geotraces Process Studies focused on specific biological parameters that enhance our understanding of trace metal biogeochemical cycles. Access to these biological parameters data sets, from both the Geotraces cruise sections and the Process Studies, is thought to be useful to interpret TEI results. Thus, Geotraces International Data Assembly Center (GDAC) is happy to accept a limited data set of biological parameters. The parameters approved by Geotraces SSC, as of July 2015, are: a) picoplankton flow cytometry counts, quantitation of of specific genes [eg. nif gene by **q**uantitative **P**olymerase **C**hain **R**eaction (qPCR)], and HPLC pigments. Despite this effort, no biological data sets have been submitted to GDAC, nor did the 2014 Intermediate Data Product (IDP) include any biological data. The next intermediate data product will be released in August 2017. This is a unique opportunity for the BioGeotraces community to make an effort to submit data. To make this happens, a meeting of PIs, who are likely to submit biological data, was called in WHOI at the end of November of 2015.

Major goals for this meeting:

1. Identify biological data sets to be submitted for the IDP2017

2. Prioritize these data sets, in case that GDAC can not handle all the data sets for the IDP2017
3. Identify additional biological parameters for submission to GDAC, past the IDP2017 deadline
4. Come up with intercalibration protocols among laboratories: including collection protocols, and standarized sample processing protocols
5. Come up with possible strategies for a future stand alone BioGeotraces Program
6. Update methodologies on Geotraces website: http://www.geotraces.org/science/biological-parameters/225-appendix-detailed-measurements

Below, we highlight the achievements for each goal, as well as some discussion on challenges and future vision.

1. Using a table (see end of document), we

    a) identified ***biological parameter data sets*** to be included in the IDP2017
    b) identified ***PI leaders*** for intercalibration efforts and the gathering of data
    c) included ***comments*** for many data sets, regarding samples to focus on, units, etc.
    d) identified ***cruises*** to focus on, prioritizing cruises where many biological parameters are available
    e) briefly mentioned intercalibration efforts

2. ***Biological parameters were prioritized*** using the following criteria: data availability, coverage and novelty, as well as direct link to the Geotraces TEIs. From high to low priority , the ranking was:
    1) *Prochlorococcus* ecotype distribution (extensive data, linkages to TEIs distributions) and nif gene (extensive data, clear link to TEIs, existing science questions) ;
    2) Targeted metaproteomics of biogeochemically relevant proteins (iron, nitrogen stress biomarkers; metalloenzyme concentrations)
    3) HPLC pigments (biomass, wide coverage) and targeted protein abundance (exciting, close connection to TEIs, potential to derive rates);
    4) 18s eukaryotic biodiversity / community structure (exciting, wide Arctic coverage, demonstration of molecular intercalibration);
    5) 16s prokaryotic biodiversity / community structure (exciting, wide Arctic coverage, demonstration of molecular intercalibration), and single cell metal quotas (close connection to TEIs and proteomics);
    6) *Prochloroccocus*, *Synechoccocus,* and heterotrophic bacteria counts (limited coverage, data not on hand); and
    7) Active fluorescence (some issues with time of day, dark adapted cells, etc.).

3. The ***additional parameters for submission past the IDP2017*** are : global proteome data, prokaryote and eukaryote metagenomes, rates of Fe uptake and N fixation (challenging to calibrate, but useful data).

4. ***Calibration efforts for the biological parameters*** are to be clasified into 2 categories : parameters with published, well established protocols and those parameters for which the

protocols are not well standardised because methodologies vary among labs, and/or are evolving quickly. During this meeting we identified PIs to led the intercalibration effort for each parameter. As a start, these PIs, in consultation with others, agreed to write the "state of the art" methodology for each parameter, as well as calibration protocols. This document will include discussion on accuracy and precision of the method, recommendations for the number of instrumental replicates, as well as true sample replicates.

The calibration effort for various parameters will be led by researchers in parenthesis : HPLC pigments (Maldonado), *Prochlorococcus* and *Synechoccocus* flowcytometry cell abundance (Chisholm & LaRoche), heterotrophic bacteria counts (LaRoche & Boyd), nif gene and 16S biodiversity (LaRoche), 18s biodiversity (Allen), targeted protein abundance (Saito), single cell metal quota (Twining), *Prochlorococcus* ecotype distributions (Chisholm), active fluorescence (Boyd).

Issues highlighted during the intercalibration dicussions :

- The amplification DNA region targeted for 18s eukaryotic biodiversity was the same in the LaRoche and Allen labs. Thus, their data will be combined and reprocessed to common **O**perational **T**axonomic **U**nit (OTU) groups. This will allow direct comparisons among the data sets. Different regions were used for the 16s, so the researchers will not be able to combine and reprocess the data. In the future, DNA regions to be amplified will be agreed upon.

- For intercalibration efforts among labs, splitting whole filter samples was recommended. These large volume samples might be easier to have, once the WHOI autonomous vehicle (Clio) for large volume particles collection is operating. Using split samples, labs could compare extraction protocols, and come up with standard pre-extraction protocols (eg. synthetic transcript additions), as well as agree on cDNA reagents, specific primers and standards (qPCR). Ideally, in the case of 16s and 18s, same variable region should be identified and targeted for amplification.

5. Come up with possible strategies to ***create a stand alone BioGeotraces Program***. We created a Google Drive to share relevant documents for BioGeotraces participants. An overview of some of the BioGeotraces projects can be found in the presentations by participating groups. Note that each participating group was assigned a task for the presentation (see schedule), so not all presentations were necessarily a summary of their activities.

   The TARA program focused on describing marine microbial biodiversity and identifing new microbes. BioGeotraces could focus on microbial community functioning, as well as ecosystem functioning. We discussed focusing on temporal changes to better understand biological variability, and targeting oceanic regions that show gradients in TEI and biological parameters. Compliance cruises that measure a series of bioactive TEIs and macronutrients were recommended. Having good core sampling gear, as the WHOI Clio Biogeochemical Vertical AUV sampling vehicle under development, is essential for the success of the program. Creating a SCOR BioGeotraces working group was suggested as a possible initial strategy. Some of the deliverables could include: a) rolling out technology to developing countries, b) develop a BioGeotraces data product, c) establish

a visiting scholar program to train other BioGeotraces researchers, d) strengethen bioinformatic and observational capabilities to allow mining of big data sets, e) strengthen links with modelers. M. Maldonado will approach Ed Urban with the idea of this working group.

6. The ***webpage http://www.geotraces.org/science/biological-parameters/225-appendix-detailed-measurements will be modified to include a simple list of BioGeotraces parameters*** that can be collected in routine cruises. This list will describe the expected data product, the amount of water required for sample collection, as well as the time required for these sample collections. Scientists interested in collecting samples for these parameters will contact Maite. Maite will then put the interested party in contact with the BioGeotraces PIs who make these measurements.

Challenges:

- A repository for the global proteome spectral raw data was recommended. A. Allen suggested Google Drive as a good option. Seeking support from the new NSF EarthCube program may also be useful.

- P. Chisholm lab will have their derived product for the metagenome ready by the end of 2016. The product will include most abundant proteins/enzymes with metal cofactors, as well as proteins involved in trace metal transport, efflux, and handling.

- Publising a *Nature Geosciences* article highlighting the following BioGeotraces findings: Ellwood (Fe & Zn stable isotopes in the Southern Ocean), Ben Twining (single cell quotas), and Chisholm (Prochlorococcus metagenome). Phil Boyd is taking the lead to contact the editor.

Action items:

Deadline January 22nd, 2016:

1. **All:** suggest to Abby "parameter naming" and "units" for BioGeotaces parameters, according to their own experience with parameter naming conventions. To help us with this, Abby has uploaded in the Google Drive "List of parameters in the GEOTRACES Intermediate Data Product 2014 v2" and the current Geotraces "Parameter Naming Conventions".
2. **All:** provide Abby with a typical excel data sheet for their parameter/s.

3. **Dani (BCO-DMO)**: to share taxonomic excel sheet

Deadline April 1th, 2016:

Data submission deadline to GDAC for guarantee publication in the IDP2017.

Please see Abby's presentation entitled: BioGEOTRACES_meeting_Nov_2015_final.pptx in the Google BioGeotraces Drive. Abby did a great job highlighting what we need to do, see slide "Demands on PIs" as well as "How can you help GDAC?" She has also loaded a template of for the metadata, see "MetadataSubmissionTemplate.xlsx"

Here is some info I extracted from her presentation:

When submitting data please consult the following webpage for information:
http://www.bodc.ac.uk/geotraces/data/submission/

Please see the 'general guide for data submission' page for further details on what metadata require submitting. There are metadata specific to CTD data submissions, and metadata specific to water sample data submissions.

Please fill out the dataset description metadata form with accurate and clear methodologies (or a reference to your methodology)/ analytical procedures/ equipment. This metadata dataset description form also states how the data should be reported.

The EVENT tables include information about when, where and how the sample was taken
The BOTTLE table holds metadata about the Niskin bottles

Deadline April 30th, 2016:

1. **All:** provide Maite with recommendations for methodologies and intercalibration protocols/efforts for each parameter (for further details see point 4 above). I would like to submit these documents to the Geotraces Standards and Intercalibration committee soon thereafter.
2. **All:** provide Maite with the necessary information to update the website :
   This could be three simple sentences describing a) the expected data product, b) the amount of water required for sample collection, and c) the time required for these sample collections.

Deadline July 1st, 2016:

Maite: submit to the S&I committee and the DMC committee a summary of the standards and intercalibration efforts for Biogeotraces parameters, and the BioGeotraces data submitted or to be submitted by Dec 2016. M. Saito or M. Maldonado will be suggested as possible new members of the S&I committee, for proper handling of the incoming BioGeotraces data sets.

Deadline Dec 1st, 2016:

Last data submission deadline to GDAC, but publication in the IDP2017 is no guaranteed.

Deadline Feb 27th, 2017 (estimated deadline, we need to consult with publisher):

Submisison deadline for ***special issue in Biogeosciences to highlight BioGeotraces findings***; to be published in the fall 2017. Contributions include: a) Atlantic data set for Prochlorococcus and Synechococcus (Chisholm), b) nif gene from Geovide and Arctic cruises (LaRoche), c) metalloproteomic data (Saito), d) single cell quotas (Twining) and proteomic data (Saito) modeled by Al Tagliabue.