2016 International Workshop On Environmental Genomics
Understanding The State of the Science and Research Opportunities

Workshop Summary Report

St. John’s, Newfoundland, Canada,
June 2nd and 3rd, 2016

Summary Report prepared by:
Ian King, eDNAtec

SPONSORS

petroleum
RESEARCH
newfoundland & labrador

GenomeAtlantic
Life Sciences. Life Solutions.

RDC
RESEARCH & DEVELOPMENT CORPORATION
NEWFOUNDLAND AND LABRADOR

ExxonMobil

Canada

Executive Summary

The application of environmental genomics methods (e.g. environmental DNA, DNA metabarcoding) to biodiversity assessment and biomonitoring is in the exponential phase of development, due in part to both the increasing capacity and decreasing costs of DNA sequencing. In some studies, these methods have been shown to have the potential to save both time and money, and provide greater sensitivity compared to traditional methods of morphological identification and counting of individuals. There are still a number of theoretical and technical issues that need to be addressed, which will then facilitate the implementation of these methods for broader industrial and regulatory use.

Participants at the workshop proposed a number of "next steps" to help foster greater collaboration, which will be crucial to facilitate the mainstream application of environmental genomics.

- Creation of a consortium/working group
- Creation of Bioinformatics Data Quality Assessment Guidelines
- Annual Environmental Genomics Workshop
- Dedicated resource to collate information on protocols and standards at all stages
- Prioritization of top 5-10 burning questions to be distributed leading to the writing of a collective white paper
- Inter-laboratory comparison, with co-analysis of some samples, and paper demonstrating applications
- More comparison/calibration studies and a data repository
- Broadening the scope of studies to a more systematic approach – translate impacts to a wider audience

Based on the success of this first workshop, a 2nd International Workshop On Environmental Genomics is planned for June, 2017, to be held again in St. John's.

Introduction

The 2016 “International Workshop On Environmental Genomics: Understanding The State of the Science and Research Opportunities” was held in St. John’s, Newfoundland, Canada, on June 2nd and 3rd, 2016.

The workshop was attended by approximately 45 participants from around the world. There were 37 participants from North America (US and Canada), 3 from Europe (France and Switzerland), 2 from South America (Brazil), 2 from Asia (China and Japan), and 1 from Australia (see Figure 1).

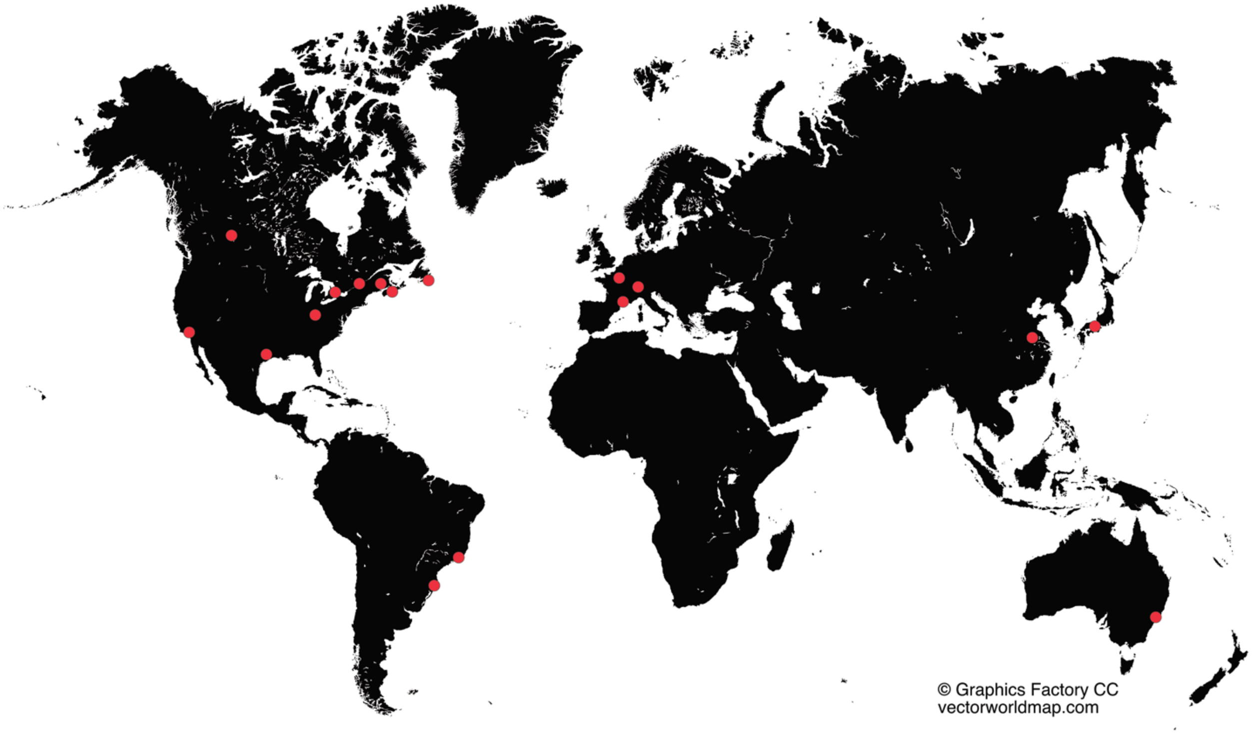


Figure 1. Map showing regional diversity of participants at the 2016 International Workshop on Environmental Genomics.

As well as representing various geographic locations, participants also represented various sectors. As part of a real-time introductory survey of 38 respondents, 11 gave their primary affiliation as being in academia, 10 from the oil and gas industry, 10 from governmental institutions, 2 from non-governmental organizations (NGOs), and 5 identified as “Other”, which represented individuals from service companies and funding agencies (see Appendix for these and other results of surveys conducted throughout the workshop).

The workshop brought together a diversity of scientists and engineers with various levels of expertise in the field of “environmental genomics” - when asked how familiar they were with the term and field of “environmental genomics”, 61% of respondents indicated that they were either “Expert in the field” or that they “Can make informed decisions”, while 39% indicated that they had “Recently heard/learned about it”.

The workshop was organized in generic themes with a set of presentations from various experts and facilitated discussions dispersed throughout the two-day effort. A summary of key discussion points and of each session is provided in this report.

Session Summaries

Session 1: Environmental Genomics Overview

This session provided an overview of the potential—and pitfalls—of the application of environmental genomics. The presentations included considerations of the theoretical underpinnings of the science of biodiversity assessment through to the practical applications of this research to a number of diverse geographical locations and environmental settings and issues. These included monitoring the ecological impact of salmon aquaculture in Europe (Jan Pawlowski – University of Geneva, Switzerland), characterizing the ecology around oil sands mining in Canada (Mehrdad Hajibabaei – University of Guelph, Canada), and assessing the ecological effect of pollutants in urban rivers in China (Xiaowei Zhang – Nanjing University, China), all of which showed that monitoring environmental change is a global issue that can be tackled by environmental genomics.

Common to the variety of examples of the use of environmental genomics was the recognition that these methods provide the possibility of a more cost-effective, quicker, and reproducible way of surveying biodiversity and assessing environmental impact. It was also clear that improvements in database robustness and standardization are crucial to the implementation of these methods.

Session 2: Environmental Genomics Research

The first session was formulated to give a high-level overview of the topics that were to be discussed during the workshop. This set the stage for Session 2, which included more in-depth discussions of research projects involving environmental genomics. Something that was clear from the first session was emphasized in the presentations in the second – these applications of environmental genomics and environmental DNA are potentially relevant to a number of fields and stakeholder groups. Examples of industrial applications were provided by Charles Greer (National Research Council of Canada) presenting on microbial diversity associated with oil and gas condensate in the marine environment, Brandon Booker (University of Alberta, Canada) speaking about detecting Boreal Toad in reclamation sites in Alberta, and Afonso Bainy (Federal University of Santa Catarina, Brazil), whose research is on studying the effects of produced water on scallops through transcriptomics. There were also many examples of the use of these pioneering approaches to the monitoring and protection of rare, threatened, and endangered species, such as the Japanese giant salamander (Satoshi Yamamoto – Kobe University, Japan), for which environmental DNA information was used in ecological niche modeling to better assess the range of this rare species.

All these examples demonstrated the versatility of environmental genomics approaches when employed in settings where traditional methods would be difficult (logistically, time-wise, money-wise), if not impossible. This session also highlighted the important consideration of the detectability of the organism(s) being studied. Species that are rare and difficult to detect by traditional methods lend themselves to these new genomics tools, especially the analysis of environmental DNA.

Session 3: Data Analysis & Interpretation and Session 4: Biodiversity Surveys and Genomics

Sessions 3 and 4 gave more specific examples of the comparison of biodiversity information from conventional *vs.* genomics-based methods and the need to further develop methodologies. In general, most of the studies presented concluded that genomics data were comparable to conventional biodiversity data. These studies also identified challenges in data analysis, and addressed questions such as the benefits of the use of operational taxonomic units *vs.* species names for data comparison.

In a presentation by Lucie N'Guessan (ExxonMobil Upstream Research Company, USA), it was shown that the use of multiple markers can provide more information on a wider taxonomic breadth to achieve the greatest diversity coverage. Jan Pawlowski (University of Geneva, Switzerland) gave an example of a setting—the abyssal deep-sea—where little is known about the biological diversity, and which would be impossible to study using classical means. In this case, environmental genomic methods are the only way to get at a biodiversity assessment of such challenging locations. Anne Basseres (Total S.A., France) and Sophie Arnaud-Haond (IFREMER, France) also showed complementarity of multiple barcodes in their research, and presented that a preliminary internal study has shown the high potential of a metabarcoding approach and helped identified gaps in associated analytical protocols. Mauro Rebelo (Universidade Federal do Rio de Janeiro, Brazil) presented the results of a comparative study of marine benthos from the Campos basin. Although only one marker was used in this study, a good coverage of organisms that had been identified morphologically was achieved. In order to address gaps in species detection due to the limitations of PCR, Mauro Rebelo suggested a PCR-free approach for environmental genomics. Another issue that was raised in the presentation by Delina Lyon (Shell Oil Company, USA) was the importance of collecting as much relevant metadata as possible along with samples to help with tracking and to help explain unexpected results when they arise. As well, she drew attention to a limited understanding of the longevity of environmental DNA in sediment (and other sample types), and the limitations in the various databases used for matching DNA sequences obtained from samples.

As was made clear from the presentations in these two sessions, issues related to data analysis are a bottleneck for the implementation of environmental genomics methods on a more mainstream scale. These issues include, for example, the determination of appropriate statistical analysis methods for genomics data, comparability of genomic data to the traditional data types that have already been collected in baseline studies, and standardization of bioinformatics parameters.

Session 5: From R&D to Applications

This session examined the crucial step of moving from the development and improvement of environmental genomics and environmental DNA methods to their adoption by government and regulatory agencies, with speakers from a number of these agencies from around the world (Donald Baird – Environment and Climate Change Canada; Erik Pilgrim – US Environmental Protection Agency; Anthony Chariton – CSIRO Oceans and Atmosphere, Australia; and Eric Stein – Southern California Coastal Water Research Project, USA).

Common to all these presentations were views of the challenges to the implementation of metabarcoding as a routine observation method: DNA metabarcoding lacking a consensus approach—in both field and laboratory protocols, pseudoabsences in metabarcoding data, DNA metabarcoding not supporting quantification and not having been tested for reproducibility, and lack of co-ordination among studies.

Governments, industry, and other agencies are amenable to the inclusion of these new methods in their protocols, policies, and regulations as they have the potential to save time and money. It is important then that the methodologies are sufficiently robust to meet the level of scrutiny to which they will be subjected before approval.

This session also highlighted the differences in the regulatory uptake of environmental genomics around the world. For example, Australia has been more pro-active by having environmental genomics approaches incorporated in various guideline documents and Canada follows closely behind with environmental genomics having been incorporated into the Integrated Monitoring Plan for the Oil Sands. The US has a more precautionary approach to integrating environmental genomics in various projects.

Interactive Discussions Audience survey and Table discussions

Throughout the workshop there were real-time survey sessions that included questions designed to stimulate thinking about some of the issues that need to be addressed in both the research stage and the application stage of environmental genomics. Interactive discussion sessions also included an opportunity for participants to discuss some of the survey questions amongst the people at each table before contributing their consensus opinions to the general discussion, thereby helping facilitate input on questions from all participants. These included questions, for example, about the differences in environmental DNA degradation in various samples (soil, sediment, freshwater, seawater), the choice of genetic marker to be used in “well-studied” *vs*. “frontier” regions, and opinions on when/if environmental genomics would become the primary approach to characterize and monitor biodiversity in marine/aquatic *vs.* terrestrial environments (see Appendix). There were a number of survey questions that dealt with definitions of terms commonly encountered in the general field of “environmental genomics”, including “genomics”, “metagenomics”, “barcoding”, “metabarcoding”, “bioinformatics”, and “environmental DNA”. Respondents did well choosing a common answer from the survey options provided (see Appendix), but it became clear from the subsequent discussions that there are degrees of nuance, subtlety—and even confusion and disagreement—in the definitions that need to be considered and clarified as the field progresses.

A number of important points came out of the survey and discussions. These included the observation that in a biomonitoring program or project, the main costs and logistic challenges come from the sampling, transport, etc., not the sequencing. As well, multiple genetic markers provide more information, but there is the challenge in optimizing the information signal-to-noise ratio, and there is the necessity to harmonize the choice of marker depending on the target. As well, the phrasing of the research question being asked is important (eg. biomonitoring *vs.* biodiversity assessment/characterization), as consistency is most important for biomonitoring so as not to include false positives, etc. Finally, there is a need to put the new methods in the context of existing methods. It was pointed out that there has to be an acknowledgement that current/conventional methods are far from perfect and subject to uncertainty, despite the opinion that current methods are more precise than they are. When applying new methods, such as environmental genomics, attempts must be made to set up a “standardized” method that is not as subjective as the traditional method, which is dependent on the observer – this is where the proper use of controls and improved QA/QC is crucial.

This led to a very fruitful discussion amongst all participants about the feasibility of standardization, with some participants expressing the opinion that it is not possible to have one standard for everything, as each case/setting may need different markers, extraction methods, etc. Others felt that we can start by developing a “good practices” approach—however compromised—to get things started, which can be refined as required by each case. It was also suggested that another immediate thing we can do is to standardize metadata collection by, for example, employing a MIxS (“Minimum Information about any X Sequence”) standard.

Main Themes

One of the common themes emerging from the workshop was that there is tremendous potential of the application of environmental genomics and environmental DNA methods to biodiversity assessment/characterization and biomonitoring—representing savings of time and money, improved identification, automation—but as each presentation demonstrated, there are still a number of theoretical and technical issues that need to be addressed to help facilitate the implementation of these methods for greater industrial and regulatory use, whether they are for detecting individual species of interest or inventorying communities.

The theoretical issues include the need for clear definitions of some of the terminology that is used—often interchangeably—in the general field of “environmental genomics”. This would allow for greater clarity in research and improved comparability. As well, there needs to be consideration of how classical ecological and environmental theory and methods such as taxonomic indices are modified and adapted to include some of the unique characteristics of genomic-derived data, such as the use of the concept of “Operational Taxonomic Units” (OTUs).

On the technical side, there needs to be investment in expanding the sequence databases that are used for taxonomic identification. Participants also recognized that bioinformatics analyses are often the bottleneck for research advancements, so there is a need for improvements in standardization, reproducibility, and bioinformatics training (see “Next Steps” below).

Next Steps

To close out the workshop, each table of participants was asked to list their choices for the most important “next steps” to be taken for the continued development, improvement, and implementation of environmental genomics and environmental DNA methods. All suggestions were combined, duplicates were removed, and the remaining ones were voted on by all participants who were asked to rank their top 3 choices (Figure 2). Votes were weighted (3 for first choice, 2 for second, and 1 for third choice), and the results of the vote were as follows (weighted votes)

1) Creation of a consortium/working group (59)
2) Creation of Bioinformatics Data Quality Assessment Guidelines (29)
3) Annual Environmental Genomics Workshop (27)
4) Dedicated resource to collate information on protocols and standards at all stages (16)
5) Prioritization of top 5-10 burning questions to be distributed leading to the writing of a collective white paper (16)
6) Inter-laboratory comparison, with co-analysis of some samples, and paper demonstrating applications (13)

7) More comparison/calibration studies and a data repository (10)

8) Broadening the scope of studies to a more systematic approach – translate impacts to a wider audience (1)

The workshop concluded with agreements by participants to coordinate further planning over the next year culminating in a second workshop in June, 2017.

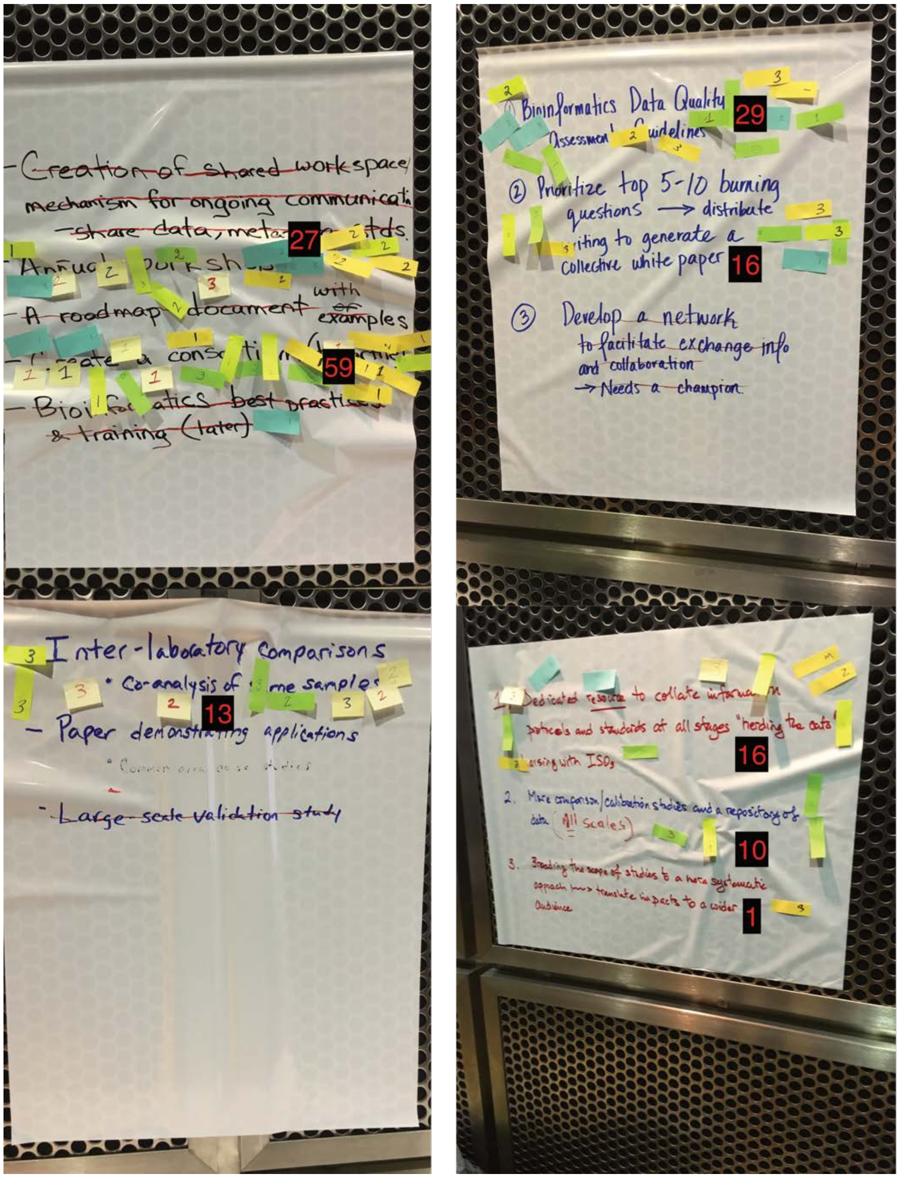
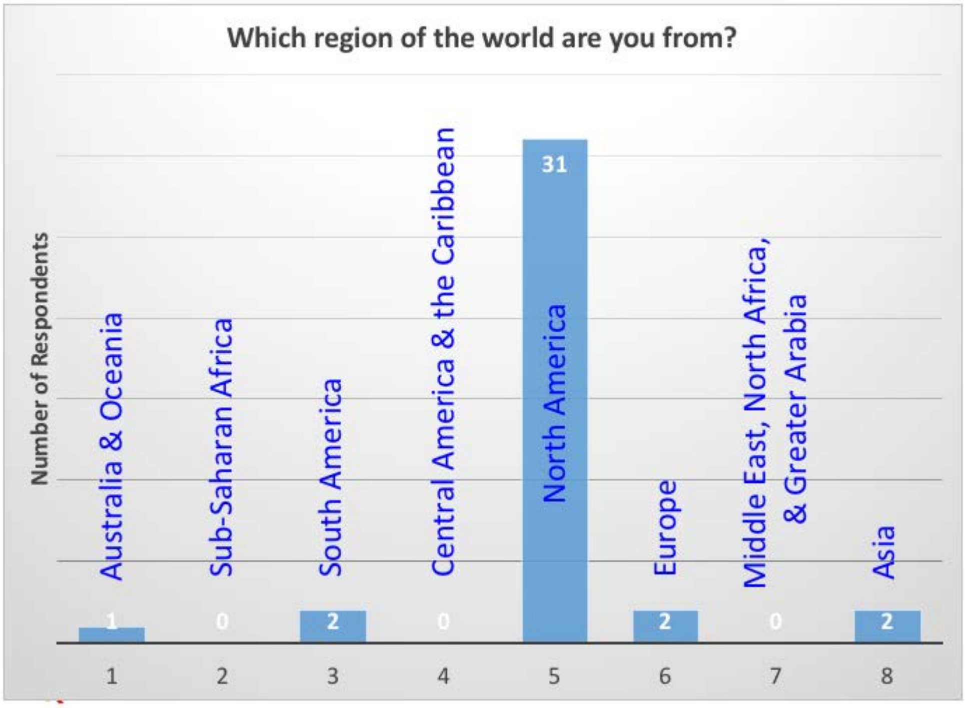
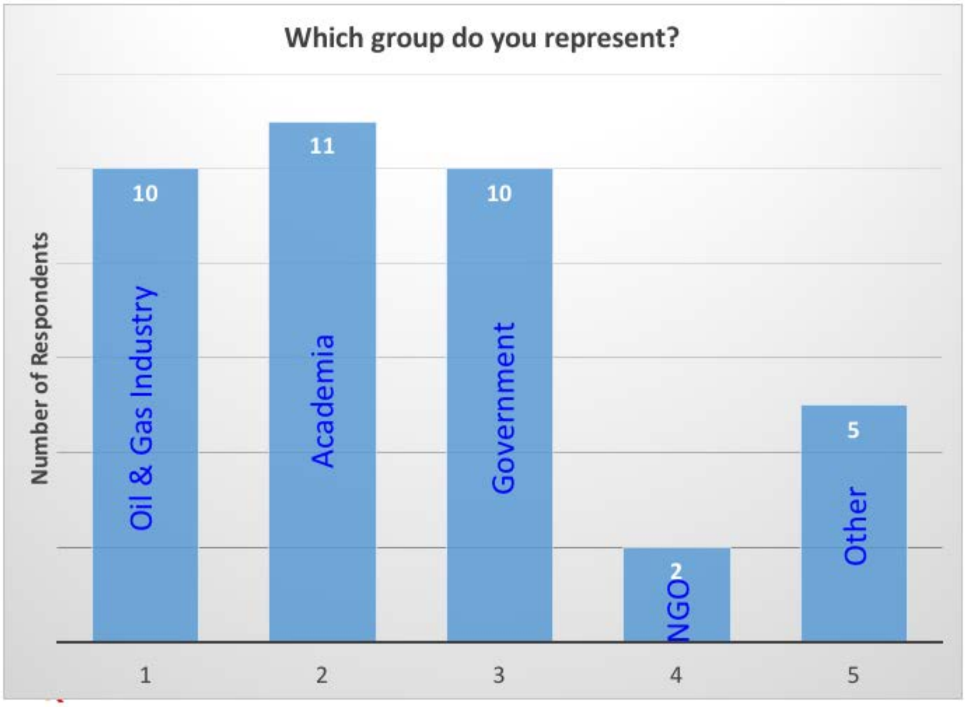
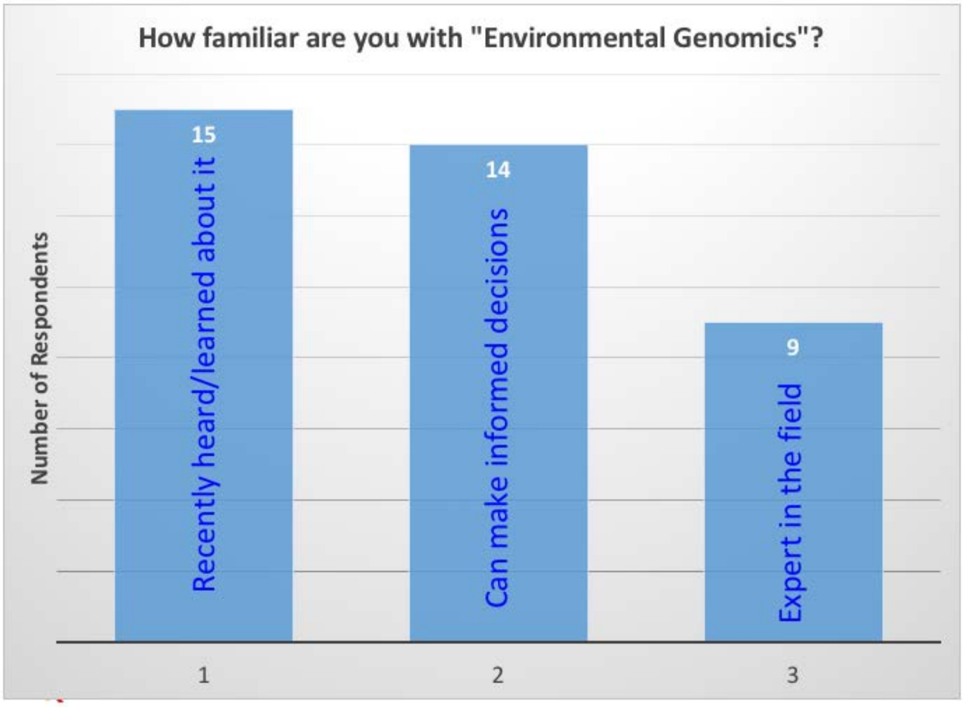


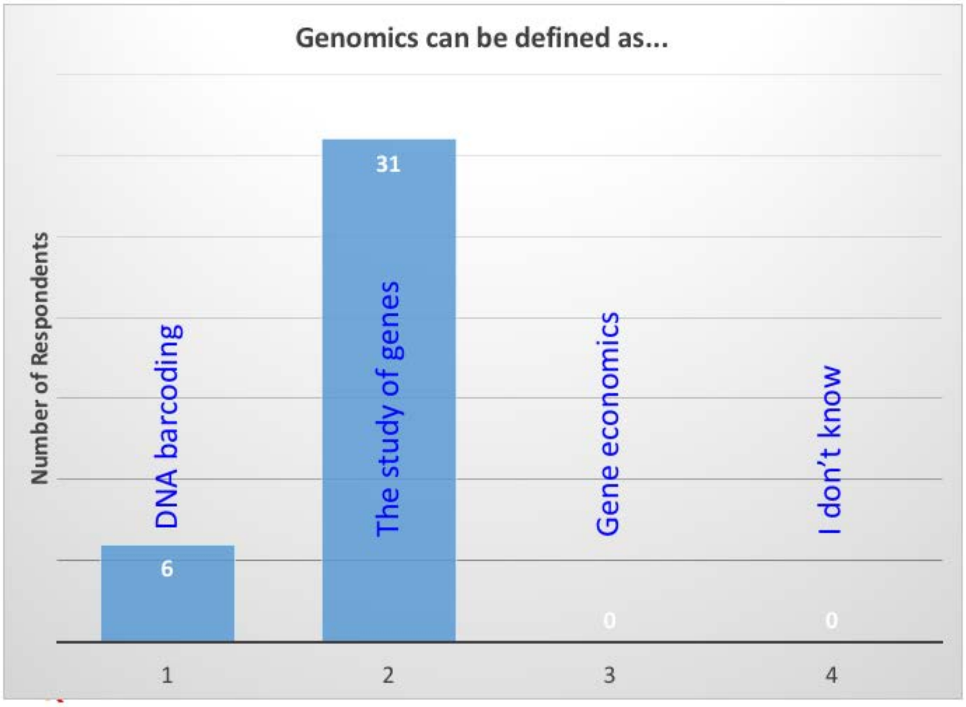
Figure 2. Results of the voting on “Next Steps” in collaborations on environmental genomics.

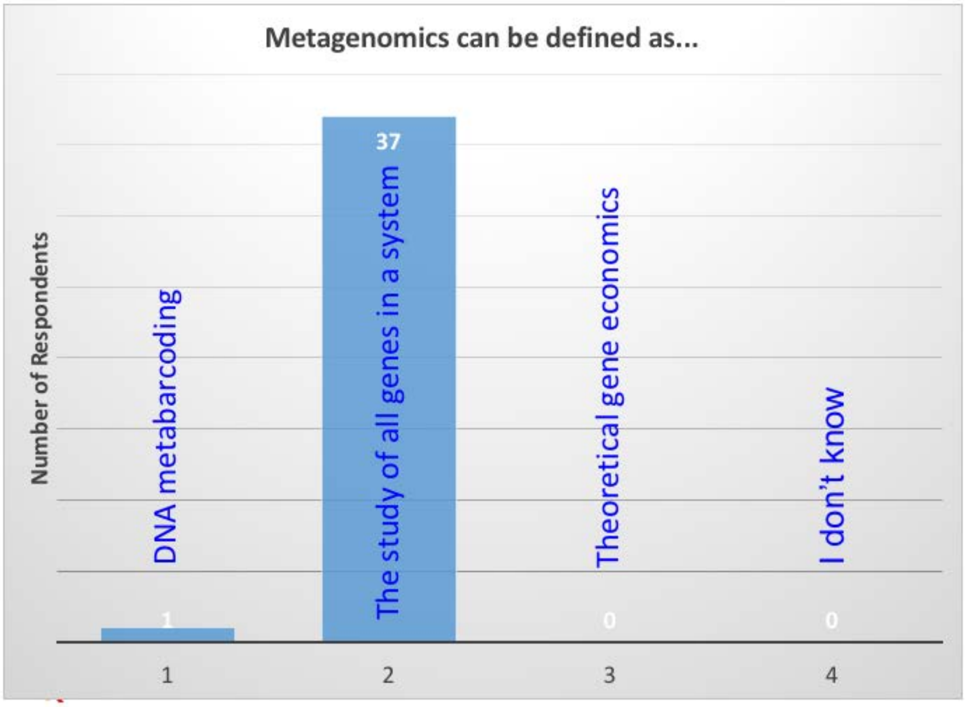
Appendix – Results of real-time workshop survey questions

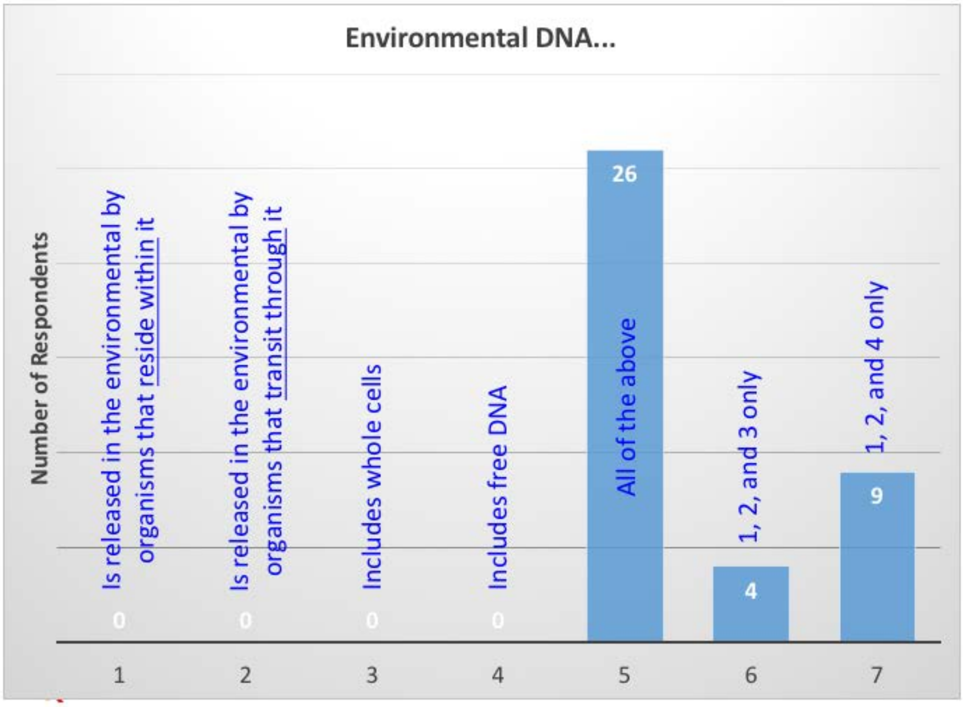


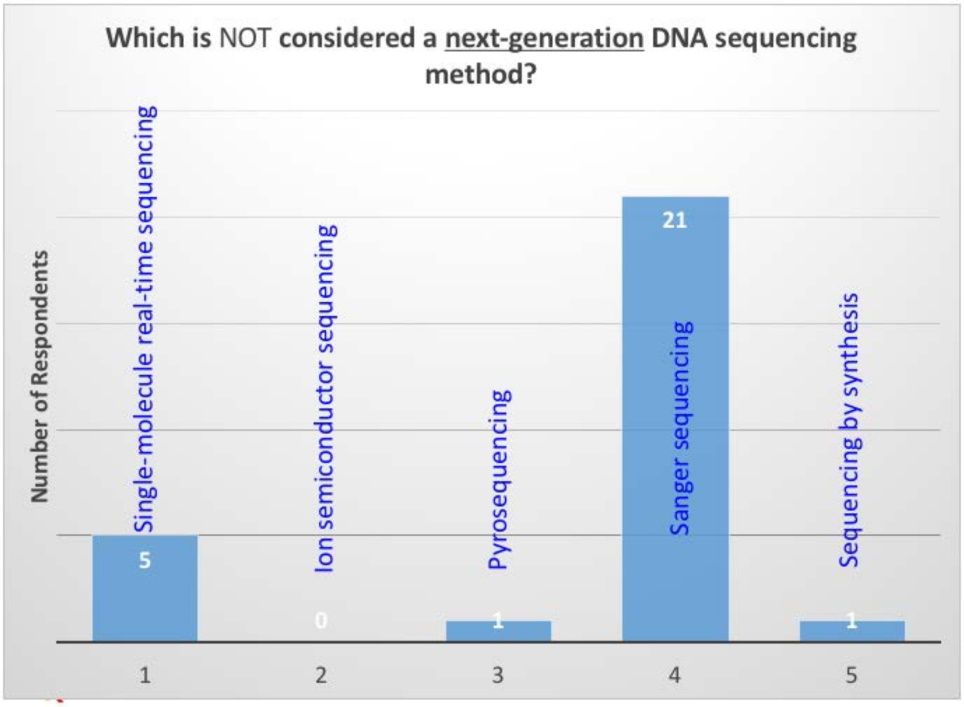


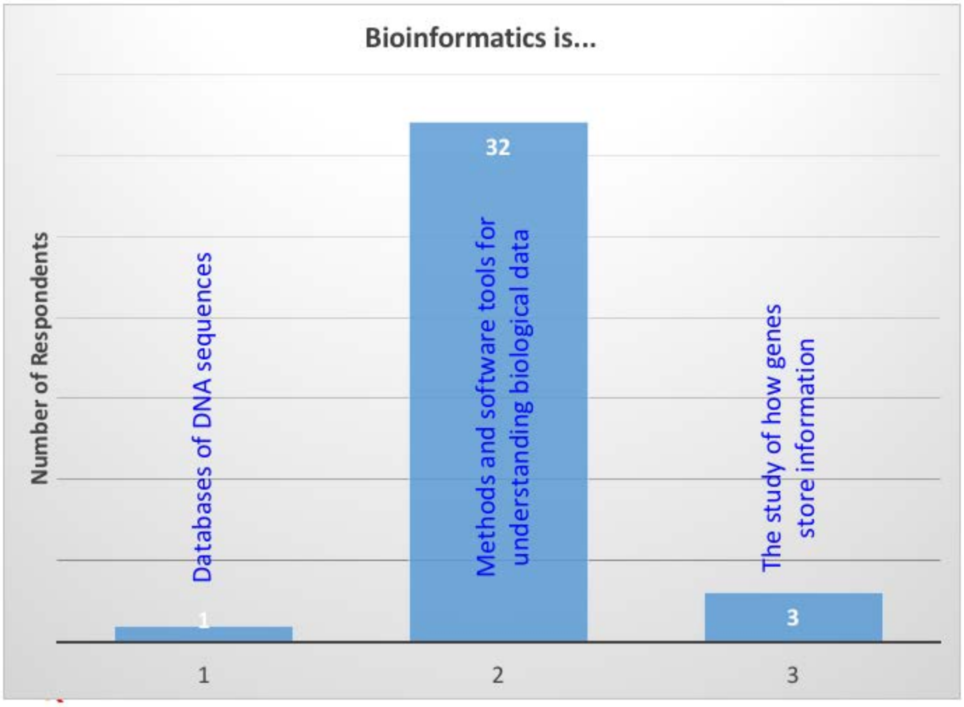

How familiar are you with "Environmental Genomics"?
Number of Respondents
15
Recently heard/learned about it
14
Can make informed decisions
9
Expert in the field
1
2
3

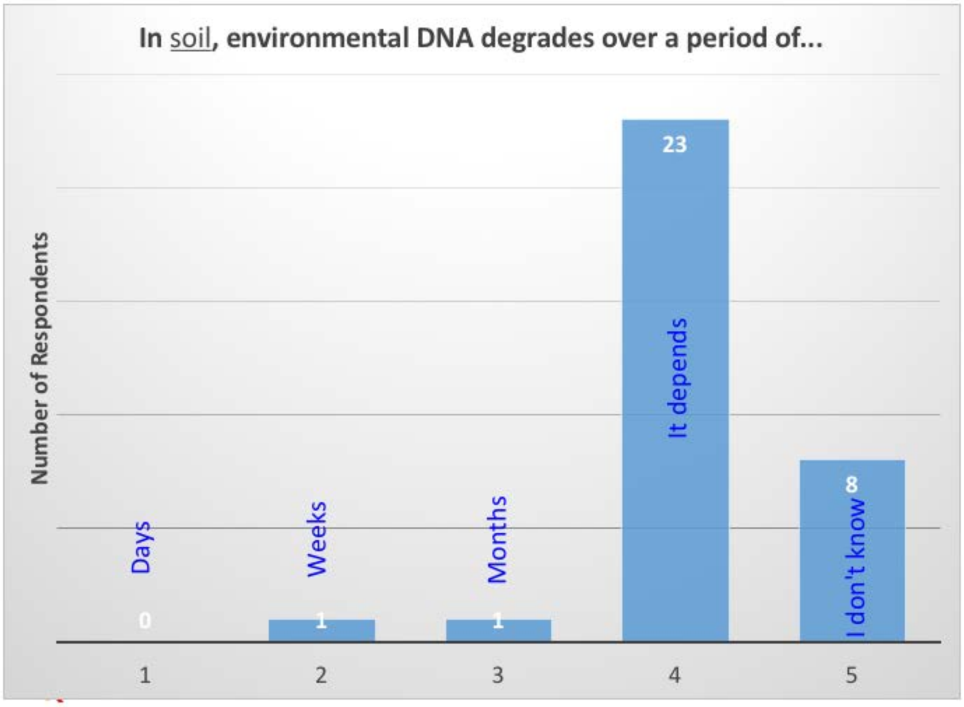

Genomics can be defined as...
Number of Respondents
DNA barcoding
6
31
The study of genes
Gene economics
0
I don't know
0
1
2
3
4

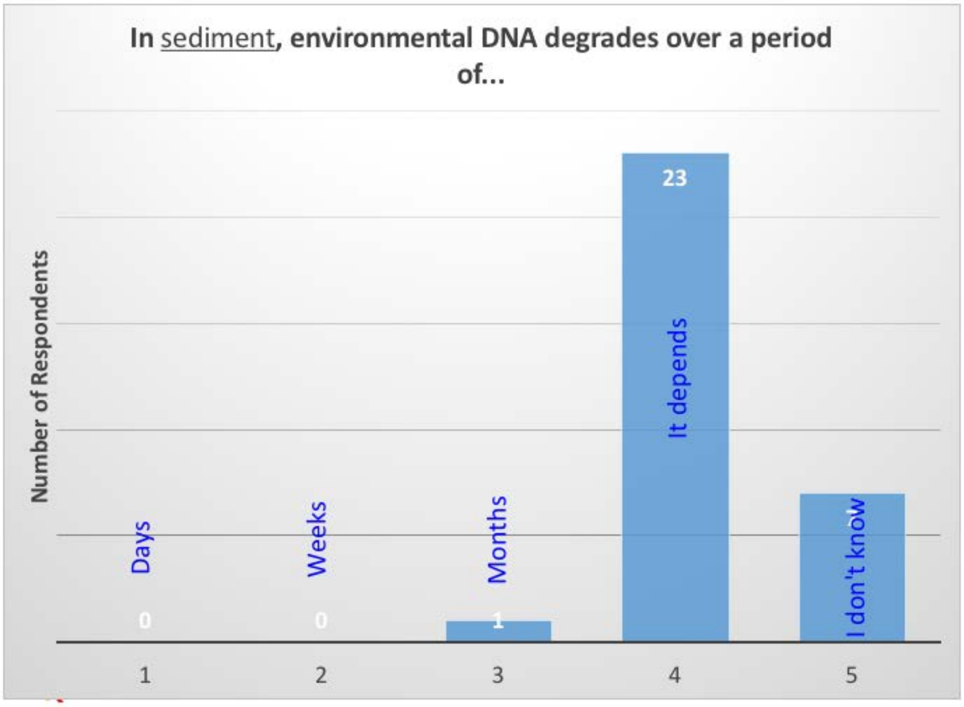

Metagenomics can be defined as...
Number of Respondents
DNA metabarcoding
The study of all genes in a system
Theoretical gene economics
I don't know
1
37
0
0
1
2
3
4

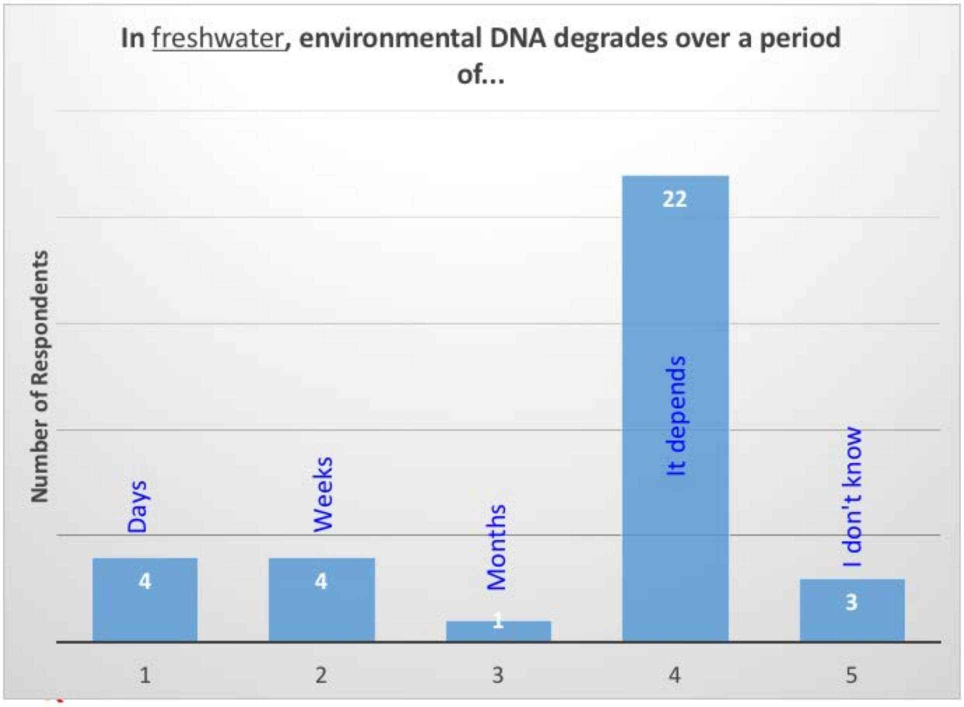

Environmental DNA...
Number of Respondents
Is released in the environmental by organisms that reside within it
Is released in the environmental by organisms that transit through it
Includes whole cells
Includes free DNA
All of the above
1, 2, and 3 only
1, 2, and 4 only
0
0
0
0
26
4
9
1
2
3
4
5
6
7

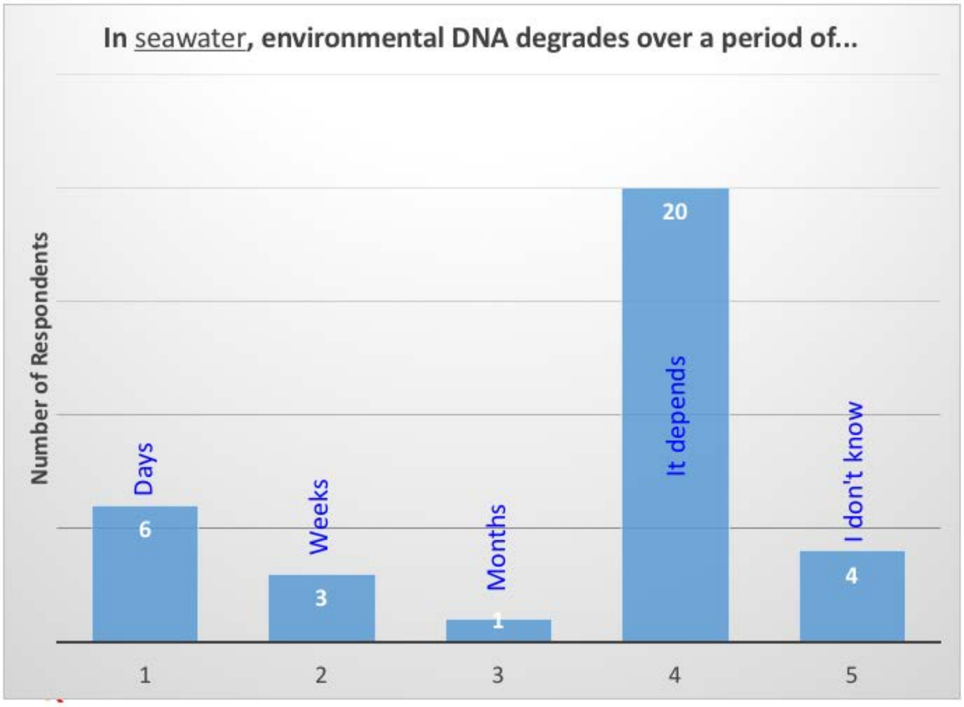

Which is NOT considered a next-generation DNA sequencing method?
Number of Respondents
Single-molecule real-time sequencing
5
Ion semiconductor sequencing
0
Pyrosequencing
1
Sanger sequencing
21
Sequencing by synthesis
1
1
2
3
4
5

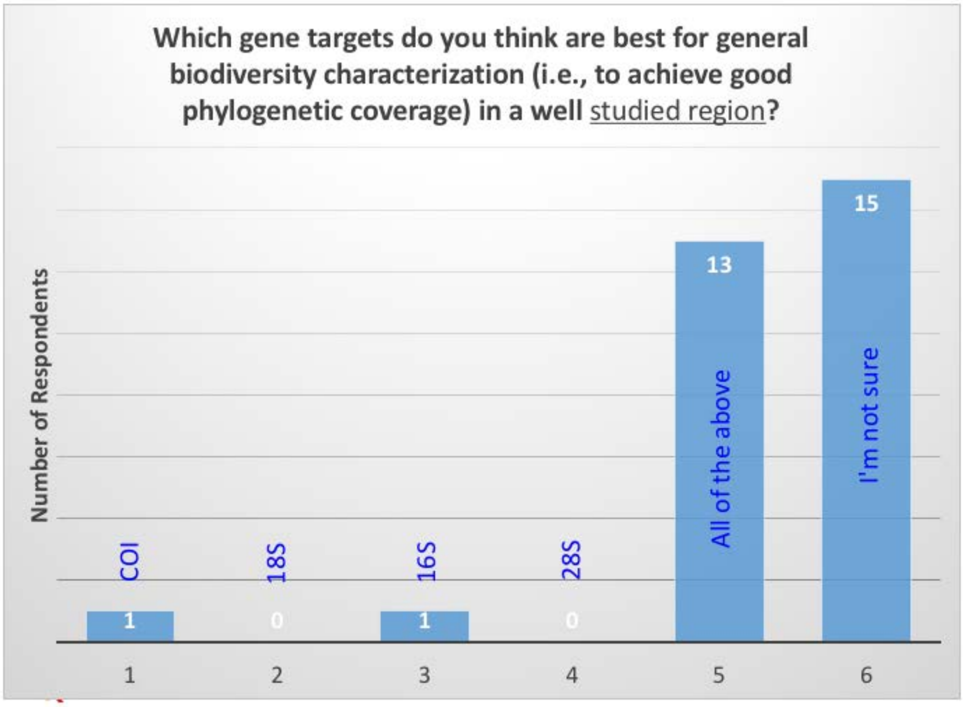

Bioinformatics is...
Number of Respondents
Databases of DNA sequences
1
Methods and software tools for understanding biological data
32
The study of how genes store information
3
1
2
3

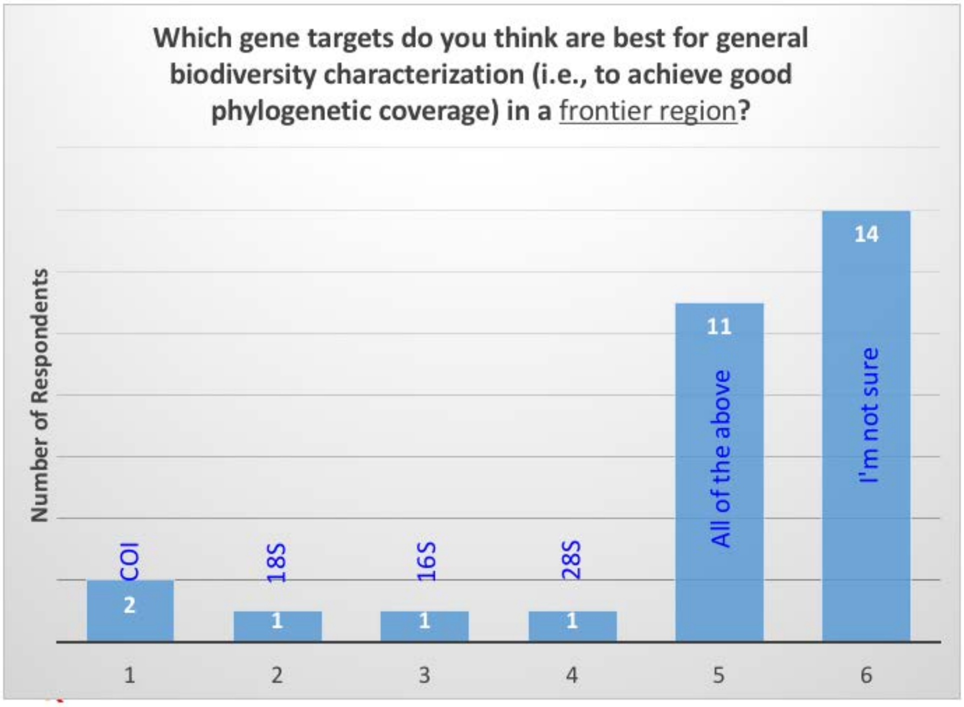

In soil, environmental DNA degrades over a period of...

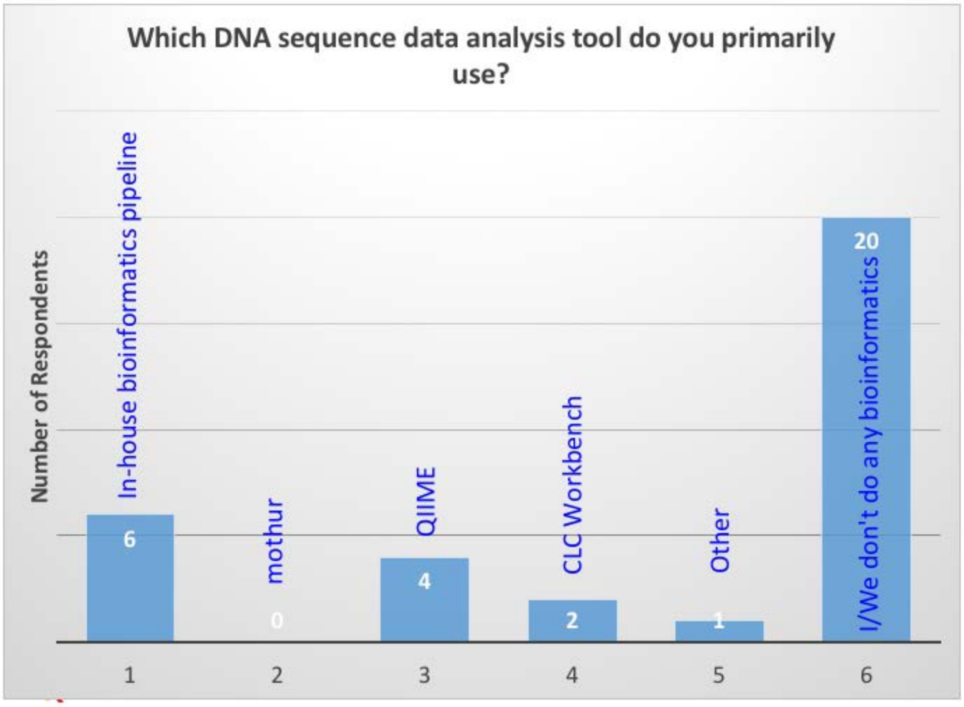

In sediment, environmental DNA degrades over a period of...

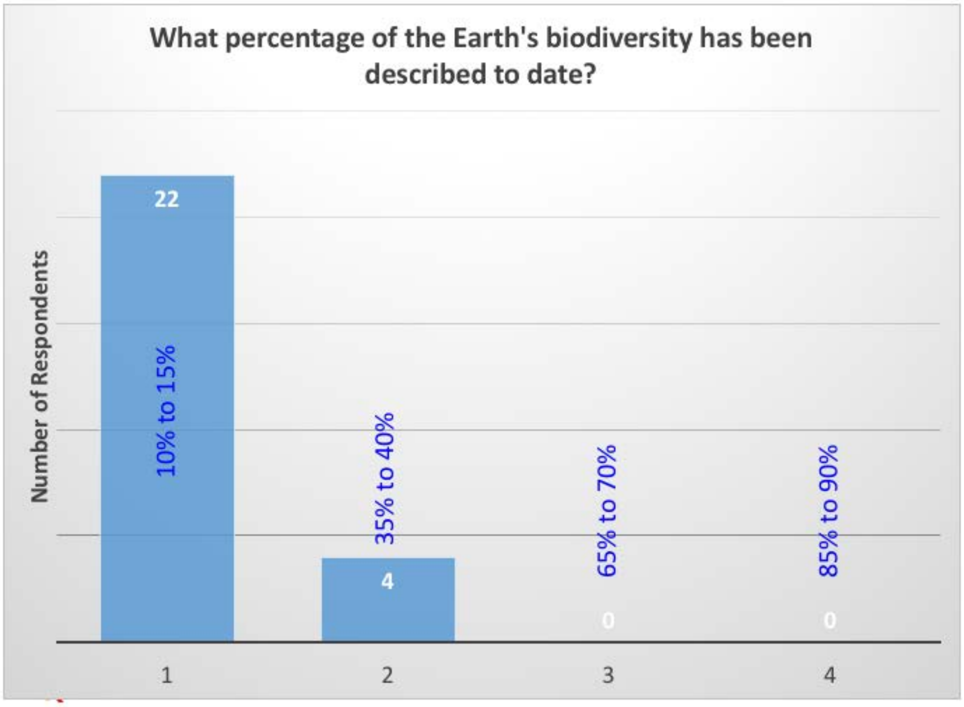

In freshwater, environmental DNA degrades over a period of...
Number of Respondents
Days
4
Weeks
4
Months
1
22
It depends
I don't know
3
1
2
3
4
5

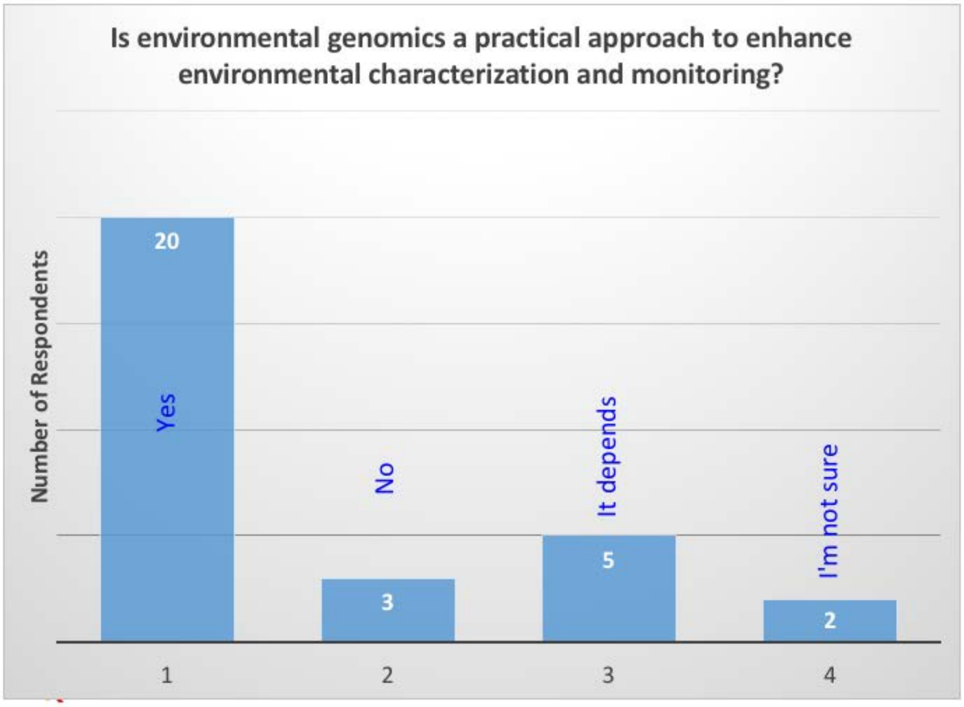

In seawater, environmental DNA degrades over a period of...
Number of Respondents
Days
6
Weeks
3
Months
1
20
It depends
I don't know
4
1
2
3
4
5

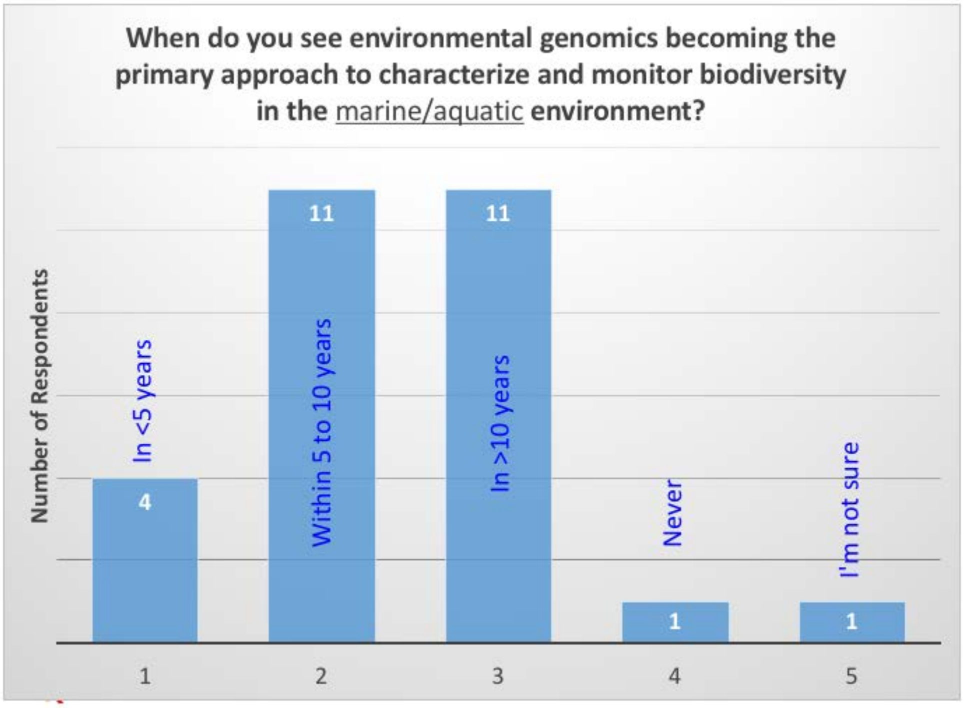

Which gene targets do you think are best for general biodiversity characterization (i.e., to achieve good phylogenetic coverage) in a well studied region?

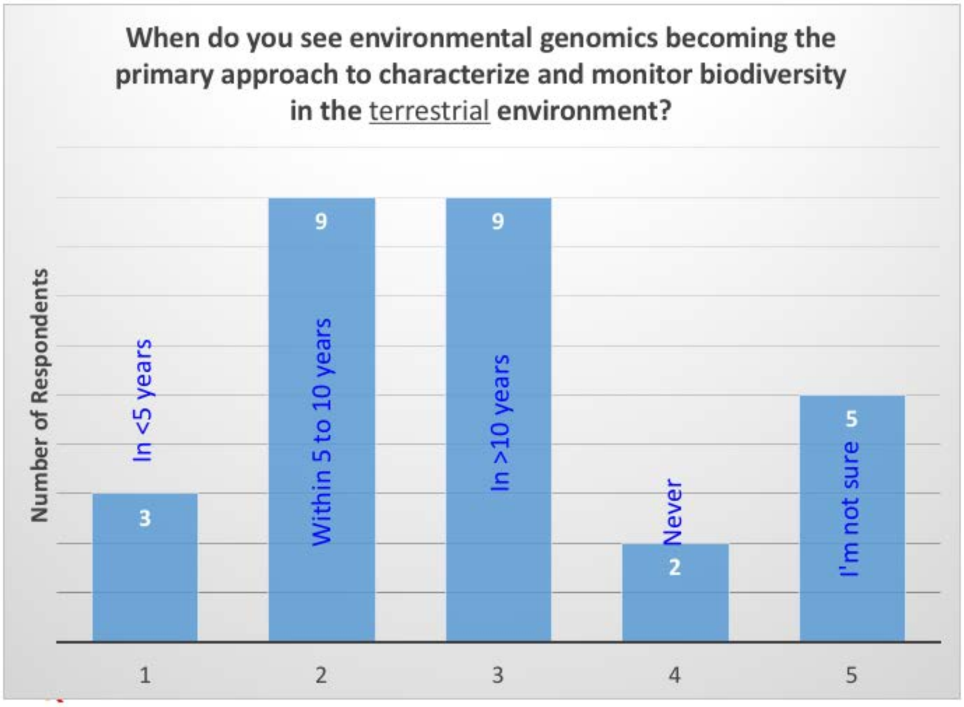

Which gene targets do you think are best for general biodiversity characterization (i.e., to achieve good phylogenetic coverage) in a frontier region?


Which DNA sequence data analysis tool do you primarily use?


What percentage of the Earth's biodiversity has been described to date?


Is environmental genomics a practical approach to enhance environmental characterization and monitoring?
Number of Respondents
20
Yes
3
No
5
It depends
2
I'm not sure
1
2
3
4


When do you see environmental genomics becoming the primary approach to characterize and monitor biodiversity in the marine/aquatic environment?
Number of Respondents
4
In <5 years
11
Within 5 to 10 years
11
In >10 years
1
Never
1
I'm not sure
1
2
3
4
5



When do you see environmental genomics becoming the primary approach to characterize and monitor biodiversity in the terrestrial environment?
Number of Respondents
In <5 years
3
Within 5 to 10 years
9
In >10 years
9
Never
2
I'm not sure
5
1
2
3
4
5