June 8-9, 2017
St. John's,
Newfoundland,
Canada

2017 International Workshop on Environmental Genomics

Transitioning to Real World Applications for Environmental Characterization and Monitoring

Workshop Summary Report

Centre for Environmental Genomics Applications

Avery McCarthy & Nicole Fahner
CEGA & eDNAtec Inc.
8/1/2017

SPONSORS

petroleum
RESEARCH
newfoundland & labrador

GenomeAtlantic

RDC
RESEARCH & DEVELOPMENT CORPORATION
NEWFOUNDLAND AND LABRADOR

Canada

ExxonMobil

EXECUTIVE SUMMARY

For the second year in a row, the International Workshop for Environmental Genomics brought together researchers from the oil and gas industry, academia, and regulatory agencies from around the world to discuss the current status of this emerging technology and identify opportunities for innovative applications. The workshop, which was held in St John's, NL, Canada, facilitated an inclusive discussion on the pace of deployment of environmental genomic tools and the possible barriers and solutions to bring this technology to the mainstream. Discussions touched on several topics including the numerous regional and national-level projects applying environmental genomics in the field, the continuously evolving bioinformatics pipelines, and the need to validate and standardize protocols. At the end of this 2017 event, workshop participants supported the establishment of an International Consortium for Environmental Genomics (ICE-G) to tackle the challenges in generating real-world genomics-based assessment programs and effectively communicating complex genomic data to decision makers. The consortium will address these technical and logistical challenges and develop solutions to facilitate the implementation of environmental genomics in broad industrial and regulatory use.

The 3rd International Workshop on Environmental Genomics will be held on June 7-8th, 2018, in St. John's, NL, Canada.

INTRODUCTION

The 2017 International Workshop on Environmental Genomics: "Transitioning to real-world applications for environmental characterization and monitoring" was held in St. John's, NL, Canada, on June 8th and 9th 2017.

The workshop was attended by approximately 40 participants from around the world, with representatives from the oil and gas industry (32%), academia (29%), and government agencies (26%). There were also representatives from non-governmental agencies, service companies, and funding agencies (13%). Of all the workshop participants, 52% had participated in projects related to environmental genomics, 31% were the lead of an environmental genomics research program, and 17% had recently learned about environmental genomics (see Appendix for these and other results of surveys conducted throughout the workshop).

The workshop was organized in five sessions with presentations from experts and round table discussions among participants. Through this dialogue, several recurring themes emerged:

Main themes

- *Biomonitoring practices are slow to change*: Time-series datasets using conventional morphological methods span back many decades. While expensive and time consuming, these methods are deep-seated in current regulatory frameworks, making it challenging to create some space for novel approaches such as environmental genomics. Industry relies on these conventional, established methods to comply with regulations which vary around the world. The introduction of new data types may not always be welcome depending on the level of familiarity with environmental genomics or the progressive culture in different regulatory agencies.

- *Results from environmental genomics pilot projects are advancing the field:* Environmental genomics methods are now being tested or applied at the regional and national level. Pilot projects undertaken in the academic, industry, and regulatory sectors show environmental DNA (eDNA) techniques to be more sensitive, faster, and in many cases less expensive than conventional biomonitoring. This cost-effectiveness can be further optimized if sampling strategies are tailored to the needs of eDNA technologies rather than conventional methods.

- *Biodiversity and ecosystem indicators and indices inform mitigation strategies:* Reduction of multidimensional data with indicators and indices needs to strike a balance between providing a lot of information and being simple enough to understand easily. With the advent of environmental genomics and the DNA sequence data type associated with it, we need to develop new ways to simplify data for easy comparison and interpretation that are compatible with the requirements of decision makers.

- *Aiming for the low-hanging fruit can facilitate uptake:* There has been some success introducing environmental genomics to solve novel or challenging problems that could not be addressed with conventional methods. Environmental genomics is well suited to applications in remote locations and situations where rapid biological information is required. The technology can also be introduced alongside traditional methods to add value.

- *Bioinformatics software validation and standardization remains a challenge*: A multitude of bioinformatics software options and pipelines are already available to process genomics data. Many of these were developed for specific medical or microbiology applications so they may not be suitable for environmental genomics without adaptations. This area of research has been identified as a priority and a joint effort is in place to facilitate an iterative calibration and validation process that can begin to address some pressing data processing questions.

- *Reference sequence libraries are essential to environmental genomics technology*: There is a need to expand reference libraries and to preserve existing databases and related resources.

- *Effective communication underlies acceptance of environmental genomics*: Decision makers, regulators, and other stakeholders need to be able to understand environmental genomics data to effectively use the information it provides.

SESSION SUMMARIES

Session 1: Current monitoring practices for industry and regulators

The first session provided context for where environmental genomics technology would be most useful given current monitoring practices employed by industry and regulators.

Industry representatives spoke on Biodiversity and Ecosystem Services (BES) management frameworks used in the oil and gas industry (Paola Pedroni – Eni, Italy) and ongoing offshore environmental effects monitoring practices (Mark Skinner – Stantec, Canada). These talks emphasized the reduction of complex BES and other environmental data into simple actionable categories that can be understood by decision makers. At their core, BES indicators are derived from biodiversity data trends, and the choice of specific parameters in environmental effects monitoring are driven in part by regulatory bodies and the lag time between data collection and environmental impact management. Some regulators have shown interest in genomic methods which offer biodiversity data in shorter turn-around times than conventional methods.

Representatives of three regulatory agencies spoke about genomic biomonitoring techniques within their agencies: Alex Bush (Environment and Climate Change Canada), Erik Pilgrim (US Environmental Protection Agency), and Sunniva Aagaard (Norwegian Environment Agency). Even though environmental genomics approaches like DNA metabarcoding offer improved discrimination between ecological communities through cost efficient increases in taxonomic breadth and taxonomic resolution compared to conventional surveys, agencies are reluctant to change established monitoring practices without rigorous and time-consuming direct comparisons of methods. Instead, adoption of environmental genomics has been more successful where it has first been introduced to solve novel biomonitoring problems. For example, it was used to collect data on organisms that are difficult to identify using morphological characters (e.g. larval fish) and to monitor locations that were cost-prohibitive with conventional sampling approaches. There are still technical challenges in collecting and interpreting eDNA data, so validation of this new technology in controlled studies and real monitoring scenarios continues to be a priority.

Session 2: Pilot Projects

Environmental DNA monitoring technologies are in a period of rapid development, and many exciting pilot projects were highlighted at this years' workshop. Alex Bush (Environment and Climate Change Canada) used environmental DNA metabarcoding in the assessment of rivers, and advanced this technology with the use of occupancy models. Thierry Baussant (International Research Institute of Stavanger, Norway) showcased the Monterey Bay Aquarium Research Institute's (MBARI) Environmental Sample Processor (ESP), a DNA-based platform for offshore environmental monitoring. With near real-time targeted species detection capability, the ESP could facilitate increased monitoring of remote and harsh marine environments. Michael Bunce (Curtin University, Australia) reviewed his eDNA research program exploring the sensitivity and robustness of eDNA experimental design factors such as choice of substrate, use of positive controls, and selection of molecular targets. His lab is also testing applications such as temporal analysis of biodiversity from ice cores, biomonitoring using gut contents, endangered and invasive species detections, and ballast water treatment monitoring. Chris Prosser (ExxonMobil) presented the results of his study looking at the migration of eDNA in soil in response to bioturbidation and Toshifumi Minamoto (Kobe University, Japan) demonstrated the power of environmental DNA methods to predict aquatic species habitat including new breeding grounds of critically endangered species.

Significant research resources are devoted to these pilot projects, but without effective communication, decision makers cannot recognize the progress in environmental genomics methods and their widespread utility. It is also important, however, to be transparent with decision makers and accurately represent the current limitations of the technology. Some workshop participants suggested using eDNA in conjunction with current methods either as a triage approach or as an add-on while it is being developed. Workshop participants agreed that future pilot projects would be most valuable if they focus on questions or applications that have immediate utility and would be more easily accepted in monitoring programs (i.e. they fill a gap or provide information not previously available). Consultation between researchers, industry members, and regulators will help identify the various technology uptake opportunities.

Session 3: Technical and bioinformatics advances

Session three delved into the technical details behind environmental DNA sequence data analysis. In this session Greg Singer (eDNAtec, Canada), Robert Beiko (Dalhousie University, Canada), and Josh Steele (Southern California Coastal Water Research, USA), provided an overview of metagenomic bioinformatic pipelines and advances in DNA sequence data analysis. There is an abundance of decision points in bioinformatic pipelines where users must make their own judgement calls. Users select reference sequence libraries, assign cutoff values for filters, and make numerous decisions in the overall design of an analysis pipeline. In addition, innovative statistical models and algorithms for sequence analysis are being introduced from other fields. Jan Pawlowski (University of Geneva, Switzerland) described how his group has been applying supervised machine learning to predict ecological status from eDNA data.

There was a consensus among workshop participants that the multidimensional “big data” results from eDNA studies are difficult to communicate, but summary figures and tables need to be easy to understand by decision makers outside the field. A discussion is required between industry, regulators, and academics to determine how to simplify data reporting while still maintaining the richness of information afforded by genomics. It is important to balance familiar data output styles with new data presentation tools like interactive or multidimensional representations. To help with these issues, workshop participants had the following suggestions: (A) Present the familiar figures that are easy to interpret, but make the complete dataset available for situations when people want to see the details. (B) Use GIS platforms to help organize environmental genomics data together with other relevant information. (C) When reporting, it must be possible to translate the data into actionable information. Typically, it was articulated that decision makers are not interested in the details of the science; the scientists need to translate the raw data into something that decision makers can act on. Data report generators would therefore need to be very clear about their audience when selecting specific reporting formats.

There were discussions throughout the workshop debating whether complex biodiversity information should be reduced to a few indicators or summary diversity indices. Appropriate indicators must be selected for each distinct habitat and some felt it may be possible to move away from this reductionism in the future with the assistance of environmental genomics approaches. Conversely, indicators can allow for prioritization of information and direct assessments towards conservation priorities or where potential impacts of projects are most severe.

Session 4: Integrating environmental genomics into existing paradigms

The workshop highlighted many cases where environmental genomics technology has been applied to obtain comprehensive biological assessments and to monitor biodiversity. Environmental DNA is being tested or applied at the national level in the USA (Erik Pilgrim – USEPA) and Norway (Sunniva Aagaard - Norwegian Environment Agency). Michael Bunce (Curtin University - Australia) presented applications of environmental DNA monitoring biodiversity in Coral Bay and Baffin Island. Alex Bush (Environment and Climate Change Canada) used metabarcoding for the assessment of rivers in Eastern Canada. Juliette Leyris (Statoil) and Mehrdad Hajibabaei (University of Guelph, Canada) described how environmental DNA is also being applied in offshore monitoring scenarios. Traditional biomonitoring methods only monitor a limited set of species but rapid advances in sequencing technology mean DNA-based biodiversity assessments are suitable to global industries who operate in different jurisdictions that are interested in monitoring a variety of taxonomic groups across many sites.

Integrating environmental genomics into existing paradigms is an ongoing effort. DNA metabarcoding provides many advantages over traditional morphological identification: it is less expensive, it is quicker, and it offers better taxonomic resolution. Nevertheless, certain information gathered from conventional surveys, like abundance, is collected differently using environmental genomics. Perceived differences in the type or format of the data obtained slow uptake of new technology despite the advantages. To facilitate uptake, the benefits of DNA-based methods have been demonstrated in applications where conventional biodiversity assessments are most difficult. This approach has garnered interest among industry members and regulators but challenges to widespread integration of environmental genomics exist: sampling methods need optimization, reference databases must be expanded, detection sensitivities need to be established, and communication with stakeholders and regulators should be improved. While challenges remain, significant progress has been made since the last workshop, and the momentum of the environmental genomics technology is growing within academia, industry, and regulatory bodies.

Session 5: The environmental genomics consortium

The consortium gained an official name by popular vote: The International Consortium for Environmental Genomics (ICE-G). The information currently hosted on the PRNL webpage will be migrated to a dedicated ICE-G webpage once it is set up.

Workshop participants discussed the possibility of starting a new graduate training program where PhD students rotate through environmental genomics labs in different sectors. PhD students would gain specialized skills in this field while facilitating knowledge transfer between labs and this knowledge transfer could accelerate advances in research. Funding might be available through different research councils (e.g. NSERC in Canada) to support this type of training program, making it a cost-effective way to recruit talented minds. Participants from last year's workshop found that it might be easier for organizations or businesses to sponsor a PhD student or postdoctoral scientist to do research at their facility rather than have organizations fund research projects at external labs.

Finally, a list of resources available within different environmental genomics groups may be useful to the consortium. It was suggested that a record of each group's sampling or analysis capabilities, the capacity of their facilities, the types of habitats they specialize in, the types of systems they work with, etc. could be valuable for establishing collaborations.

The next workshop is scheduled for June 7-8th 2018 in St. John's, NL.

APPENDIX – RESULTS OF THE SLI.DO SURVEY QUESTIONS

Name the IWEG Consortium

Votes	Name
6	International Consortium for Environmental genomics (ICE-G)
4	International Consortium for eDNA, Baselines, Environmental Regulation and Genomics (ICEBERG)
2	International Consortium for eDNA and Ecogenomics (ICEE)
1	IWEG consortium
1	GENvironmentOMICS
1	International Consortium of Environmental Genomics Applications (ICEGA)
1	EnvGenCon or EnviroGenCon
1	Consortium for Environmental Genomics Research
1	Tripartite Organization for Environmental Genomics Advancement

Question 1: Given the wide range of regulatory requirements worldwide what do you think is the most efficient path to bringing environmental genomics to the mainstream?

> A1. Development of an industry recommended practice (eg., via IOGP &/or IPIECA). Performance based regimes could adapt to this relatively quickly if they wished.
>
> A2. Regulators understanding the technology.

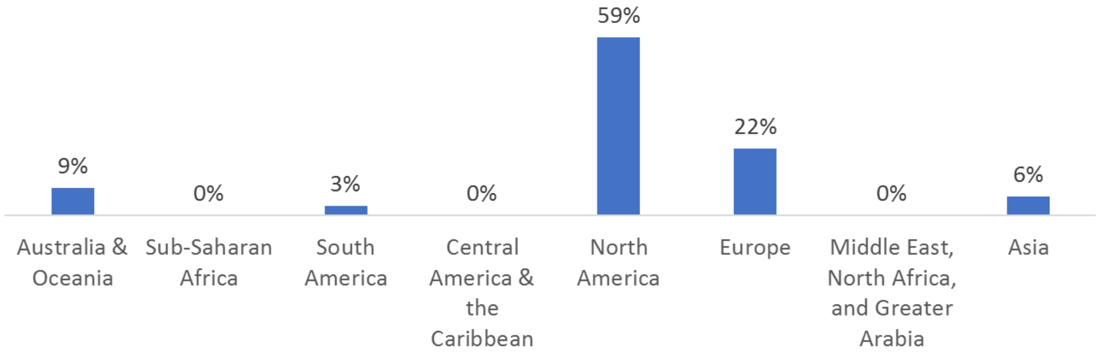
Question 2: What is the scope and objective of a pilot project? When does a pilot project transition to mainstream application? What sorts of pilot projects would help advance the technology to address the greatest gaps and the greatest needs?

> A1. Show that the technology is scalable. That you can analyze many samples. When results are replicable and when you can track back mistakes. The ones sequencing all they haystack and finding whatever needle you want.

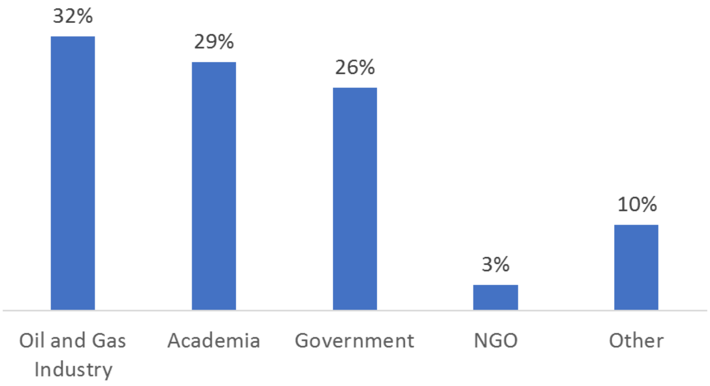
Note: Due to restricted WiFi access on the first day of the workshop, most participants chose to express their opinions on these topics during the discussion periods instead of online.

APPENDIX – RESULTS OF REAL-TIME WORKSHOP SURVEY QUESTIONS

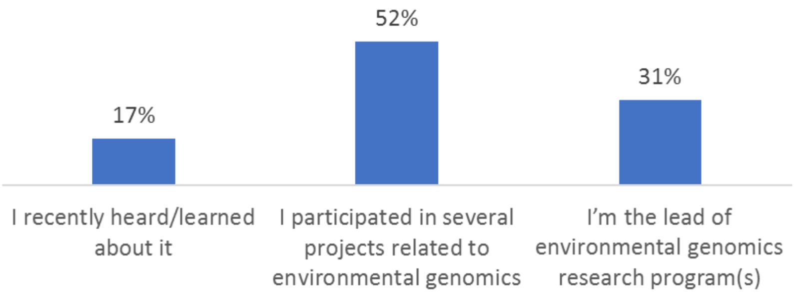
1. Which region of the world are you from?



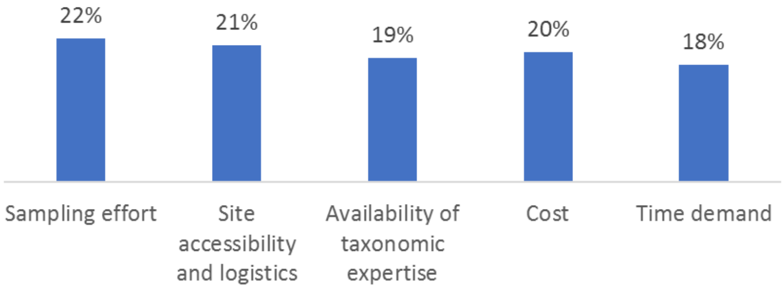
2. What group do you represent?



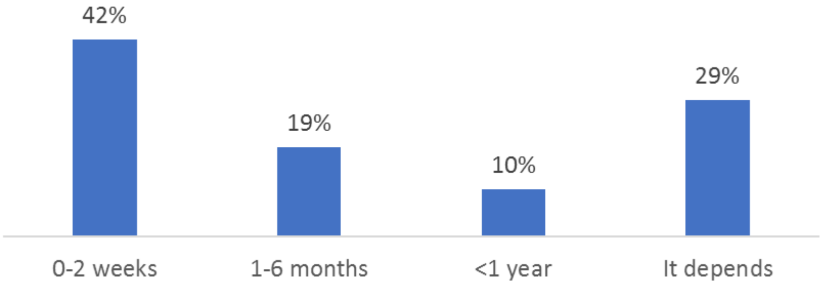
3. How familiar are you with "Environmental Genomics"?



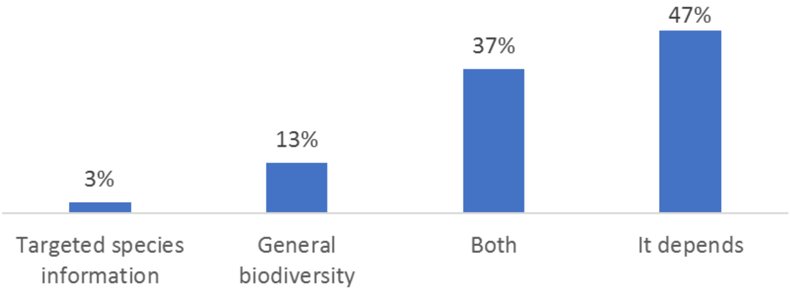
4. Rank the following components of environmental characterization and monitoring from most challenging (First) to least challenging (Last):



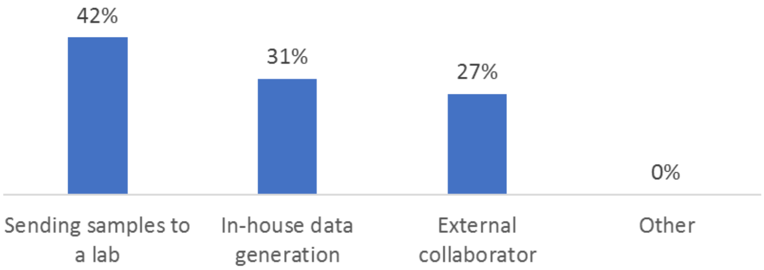
5. What is a desirable turn-around time for environmental monitoring data?



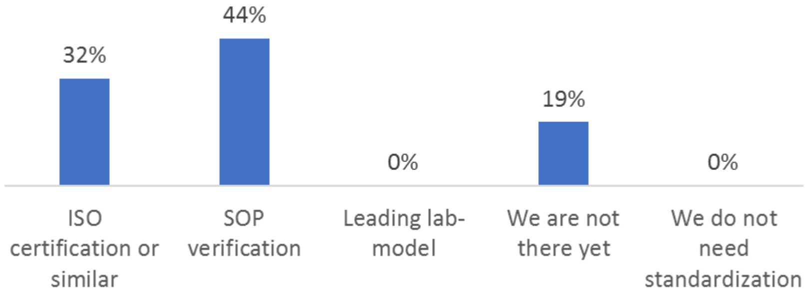
6. Which is more useful/valuable to you?



7. What is your preferred practice for data generation?



8. What level of standardization would be necessary to have established for the uptake of environmental genomics?



9. When do you see environmental genomics becoming the primary approach to characterize and monitor biodiversity in the (A) marine/aquatic environment? (B) terrestrial environment?