June 7-8, 2018
St. John's
Newfoundland
Canada

2018 International Workshop on Environmental Genomics

Building a Path to Regulatory Acceptance of Evidence from Ecogenomics

Workshop Summary Report

Avery McCarthy & Nicole Fahner

CEGA & eDNAtec Inc.

9/14/2018

IWEG 2018 SPONSORS

GenomeAtlantic

Life Sciences. Life Solutions.

illumina®

EXECUTIVE SUMMARY

The third meeting of the International Workshop on Environmental Genomics established the conference as a regular event. Workshop participants from around the world congregated in St. John's, NL, Canada this spring to share the progress in ecogenomics from their different nations and economic sectors. Members from the oil and gas industry, environmental consulting agencies, academia, and regulatory agencies discussed advances in the field, recent technological innovations, opportunities for applied research, and acceleration of technological readiness. Continuing from previous meetings, the barriers to widespread implementation of this technology and potential solutions were at the core of many talks. Other focal points of the workshop were novel projects applying environmental genomics in the field, advances in bioinformatic analyses, and first steps towards protocol validation and standardization. Notably, the consortium created at the 2017 workshop, ICE-G, presented a research prioritization matrix from which a Joint Industry Program (JIP) will be initiated. This JIP will help formalize and guide collaborations between researchers and industry partners in environmental genomics moving forward.

The 4th International Workshop on Environmental Genomics will tentatively be held on June 12-13th, 2019, in St. John's, NL, Canada.

INTRODUCTION

The 2018 International Workshop on Environmental Genomics (IWEG), "*Building a Path to Regulatory Acceptance of Evidence from Ecogenomics"*, was held in St. John's, NL, Canada, on June 7th and 8th, 2018 and attended by approximately 40 participants. Most attendees were from North America but there were representatives from Asia, Australia, and Europe as well. IWEG is intended to facilitate communication and collaboration across economic sectors and there was roughly equal participation from industry, academia, regulatory agencies, and service companies (e.g. environmental consultants) at this year's meeting. The majority of these participants had experience with environmental genomics but roughly one in five attendees were new to the field. See the Appendix for detailed results from the surveys conducted throughout the workshop.

Presentations and discussions from the five workshop sessions revolved around the following central themes:

- *Environmental genomic research priorities are aligning across industry, academia, and regulatory agencies.* The ICE-G Industry sub-group is working to develop recommendations for priority research and development areas that will support the deployment of technology in the short-, medium-, and long-terms.

- *Environmental DNA analyses must be robust, reliable, and reproducible to garner confidence within regulatory agencies.* The IWEG group is focused on the development of SOPs, best practices, and reporting standards. In bioinformatics, software validation and standardization through best practices has begun and is essential to moving forward.

- *Research and development in environmental genomics is progressing rapidly*. This year's workshop highlighted an array of *in situ* eDNA projects including biodiversity assessment of fish species, environmental baseline surveys, and ongoing biomonitoring of oil and gas developments in marine and wetland ecosystems. Presentations on technological advances featured real-time qPCR detection in the field, DNA sequencing in the field, and the NovaSeq high-throughput sequencing platform.

- *Ease and affordability of environmental DNA sampling opens countless new opportunities for monitoring programs.* Sampling for environmental genomics is non-invasive, logistically easier, and less expensive than traditional sampling methods but provides demonstrably comparable results. Increased efficiency in the field means existing programs could expand the scale of their monitoring efforts, but more fundamental to its uptake, the technology has found applications where conventional monitoring techniques are challenging or prohibitively expensive such as sampling in remote or protected environments and detection of rare species.

- *Reference sequence libraries are the backbone of environmental genomics workflow.* Everyone benefits from expanding these databases. Researchers are addressing the challenges associated with poor curation of reference libraries to improve consistency and accuracy of results.

SESSION SUMMARIES

Session 1: Key Challenges Affecting eDNA Acceptance by Regulators

The first session provided an overview of regulations relevant to eDNA-based environmental surveys and the barriers to technology adoption by regulators. Representatives of three regulatory agencies gave talks on the subject: Willie Duncan (Scottish Environment Protection Agency), Wendy Monk (Environment and Climate Change Canada), and Dave Cote (Fisheries and Oceans Canada).

The need for standard operating procedures (SOPs) was one of the key challenges presented in this session that recurred in talks throughout the conference. For environmental monitoring applications, regulators require data that can be interpreted with a high degree of certainty and repeatability. If regulatory decisions get challenged there is loss of confidence, so consistency both within a single lab and across different labs is paramount. Best practices and SOPs are needed at all stages to standardize experimental design, sampling techniques, lab analyses, bioinformatics, and reporting. The intercalibration study, presented by Eric Stein (SCCWRP), showed that best practices are required for bioinformatic analyses to improve consistency of

results between labs. It is important to wait until SOPs are well-developed and validated before introducing these tools to regulatory frameworks, because regulators can lose trust if a new technology is not standardized and reproducible.

Despite these difficulties in implementation, eDNA is well situated to overcome ecological sampling challenges such as rare species detections, species misidentifications, sampling in harsh environments, and non-invasive sampling. Environmental DNA methods are applicable to a broader set of environments where traditional sampling is not possible or prohibitively expensive. The technology shines in applications where no other methods are feasible.

Several studies presented at IWEG directly compared eDNA methods to traditional methods which is a common approach used to validate new technologies. Results in these and similar studies repeatedly show both eDNA metabarcoding and morphological analyses have incomplete taxa detection, for different reasons, but researchers are now starting to question if these direct comparisons are meaningful or if they just compare apples to oranges. Instead, users need to evaluate which method is best for each biological question. Comparison to morphological data, however, remains important in terms of uptake into established regulatory regimes and the continuation of long-term projects.

Preservation of time series data was also a concern. Comparing data over long periods of time – typical of biomonitoring applications – may be challenging since the technology is rapidly evolving and tools and methods are continuously updating. Biobanking environmental samples or DNA for re-analysis in the future may solve this issue but there are logistical and cost concerns that need to be addressed.

Session 2: 'Early Adoption' Case Studies

As environmental DNA monitoring technologies continue to develop rapidly, some projects are testing these technologies in the field: Willie Duncan (Scottish Environment Protection Agency) applied eDNA methods to detect fish species in Lake Windermere, where long term datasets from traditional gill-netting methods are available; Wendy Monk (Environment and Climate Change Canada) applied the technology to explore spatial and temporal patterns in the Peace-Athabasca

Delta; Michael Bunce (Curtin University) measured conservation baselines on Cocos Island and pre-operational baselines for oil and gas sites in Kazakhstan, Bulgaria, and Finland; and Tristan Cordier (University of Geneva) conducted environmental DNA metabarcoding in the vicinity of three oil platforms in the North Adriatic Sea.

A number of exciting technological advancements were also highlighted: Marc Skinner (Stantec) applied qPCR in the field to detect of target species; Dave Cote (Fisheries and Oceans Canada) introduced plans to test a field qPCR assay for salmon; John Love (University of Exeter) performed DNA extraction and sequencing in the field; and Mehrdad Hajibabaei (University of Guelph; Centre for Environmental Genomics Applications) used the Illumina NovaSeq platform to increase sequencing depth by an order of magnitude, greatly increasing the number of species identified.

Session 3: Bioinformatics and Data Analysis

IWEG 2018 participants delved into the technical details behind environmental DNA sequence data analysis. Eric Stein (SCCWRP) presented results from an intercalibration study comparing bioinformatics pipelines between labs using a single data set with one DNA marker. Over the course of the study, best practices were identified by iteratively refining steps contributing to disagreement between the pipelines including: assignment of sequences to samples, rules used to generate OTUs, and methodology for taxonomic assignment. Future work needs to expand the comparison to other gene markers and sample types, and to include new bioinformatics methods not based on OTU generation. Intercalibration studies like this help align methods by establishing best practices that generate more consistent results between labs.

Paul Greenfield (CSIRO) detailed the challenges in amplicon metagenomics analysis. Selection of marker genes, primers, clustering methods, taxonomic classifiers were all important and several best practices were recommended: (1) use multiple marker genes and multiple primer sets; (2) test primer universality and specificity; (3) fine-tune trimming and pairing protocols to recover high quality sequences; (4) follow appropriate bioinformatics methodologies for broad surveys versus targeted searches for specific organisms; and (5) evaluate reference database quality and

limitations – or even compare results using multiple reference databases. Outside these best practices, distinguishing rare organisms from artefacts remains a challenge.

There were exciting bioinformatics advances leading to novel methods: Mehrdad Hajibabaei (University of Guelph; CEGA) advanced DNA metabarcode classification with automated high-throughput methods; Tristan Cordier (University of Geneva) used machine learning to classify eDNA; Wendy Monk (Environment and Climate Change Canada) introduced applications of network-based monitoring to look at changes in the structures of food webs; and Anthony Chariton (Macquarie University) used ecological networks as indicators for environmental conditions.

Session 4: Developing a Collaborative Research Network to Address Sectoral Needs

The ICE-G Industry Sub-group is working to align the interests of the various stakeholders in the Consortium. The Sub-group will commission a whitepaper with participation from academia, government, and industry, to propose priority research and development areas to accelerate the deployment of eDNA-based technologies in the short-, medium-, and long-terms. The timeline for the development of this whitepaper is approximately six months.

Session 5: Key Research Needs

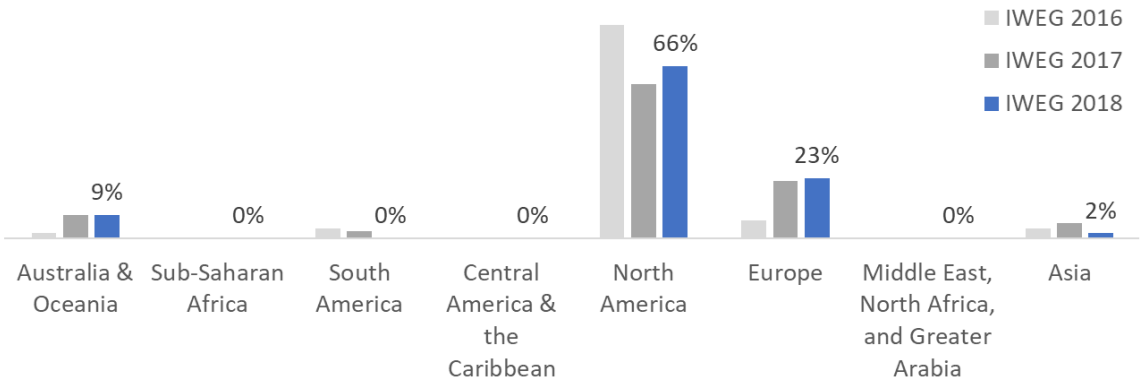
Throughout the meeting several technical challenges were identified: DNA persistence and transport in the environment is poorly understood; quantification of DNA remains intractable (although it was noted that many important biological questions do not require quantitation); choice of markers and primers has a large impact on the results obtained—particularly with respect to the specificity, resolution, detection sensitivity – but is not standardized across studies, and reference database coverage is lacking. The importance of high-quality reference databases recurred in several talks, especially with respect to rare and endangered species.

Discussions touched on the importance of taxonomists in moving eDNA technology forward. If unknown OTUs are flagged for morphological and taxonomical studies, any newly described sequences or species add value to future eDNA studies and would mutually benefit everyone.

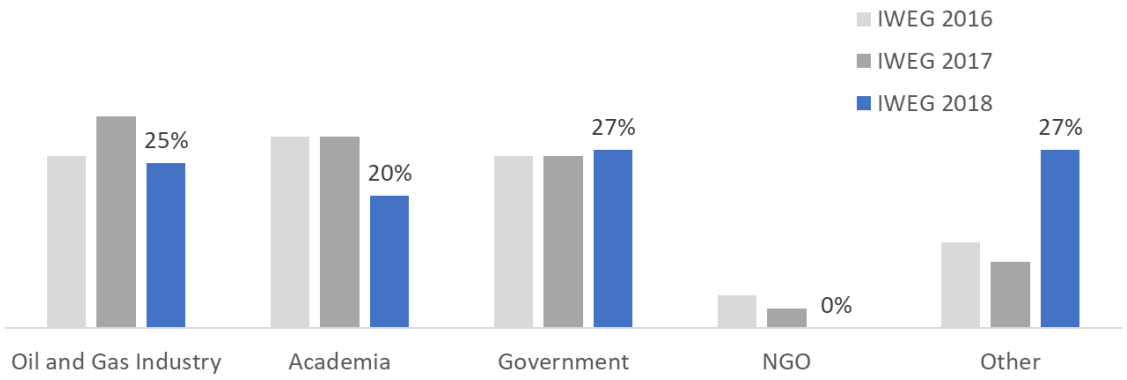
Finally, workshop participants discussed the importance of including postdocs and graduate students in future conferences. Postdocs and graduate students help advance the field and play an essential role in transferring skills and expertise between labs.

APPENDIX – RESULTS OF REAL-TIME WORKSHOP SURVEY QUESTIONS

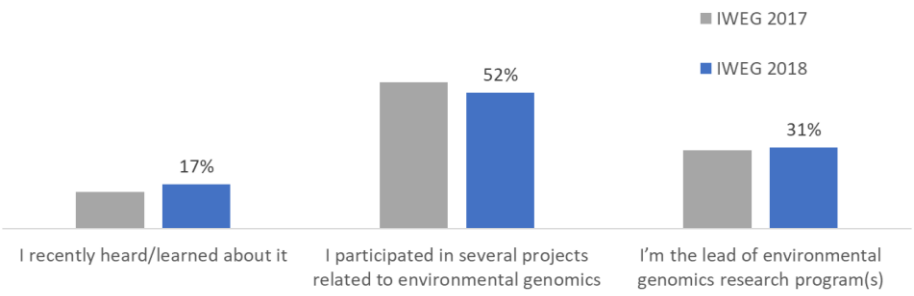
1. Which region of the world are you from?



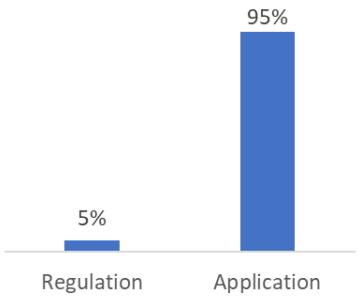
2. What group do you represent?



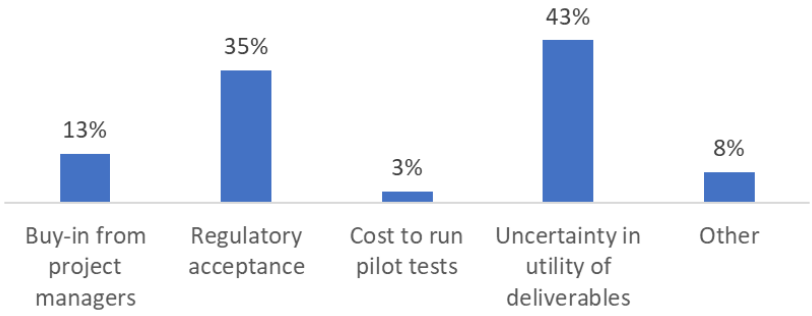
3. How familiar are you with "Environmental Genomics"?



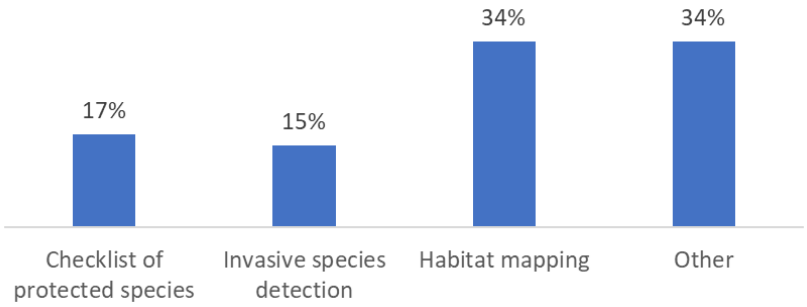
4. Which comes first? Regulation or application?



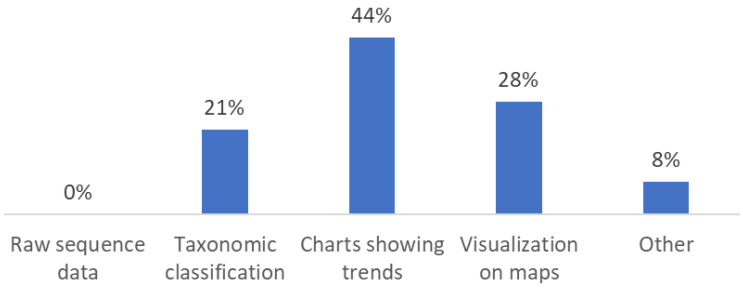
5. What are the biggest barriers to introducing environmental genomics in new and existing environmental assessment and monitoring efforts?



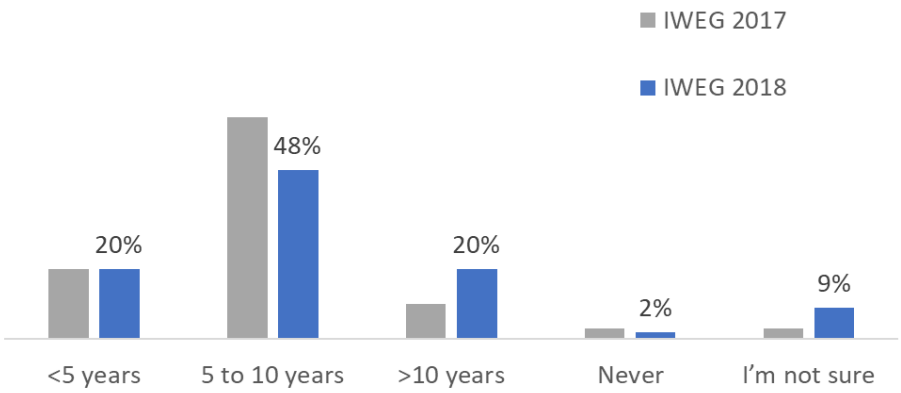
6. From a regulatory perspective, what is the highest priority?



7. What type of reporting do you find most useful?



8. When do you see environmental genomics becoming the primary approach to characterize and monitor biodiversity in the marine/aquatic environment?



9. When do you see environmental genomics becoming the primary approach to characterize and monitor biodiversity in the terrestrial environment?