DNA Metabarcoding

Mehrdad Hajibabaei

October 14, 2018
German satellites sense Earth's lumps and bumps

By Jonathan Amos
BBC Science Correspondent

10 October 2018

The map traces the variations in height across the Earth's land surfaces
DNA Metabarcoding workflow

Specimen: mixed environmental sample

Extract e-DNA and amplify/capture marker genes

HTS-NGS

DNA sequence libraries

Bioinformatics

Species identities and sequence counts

Sequence data for selected markers/DNA barcodes
Recovery

Resolution

Annotation & Analysis
Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis

Teresita M. Porter1,2 | Mehrdad Hajibabaei3

1Centre for Biodiversity Genomics, Biodiversity Institute of Ontario and Department of Integrative Biology, University of Guelph, Guelph, ON, Canada
2Natural Resources Canada, Great Lakes Forestry Centre, Sault Ste. Marie, ON, Canada
3Department of Biological Sciences, York University, Toronto, ON, Canada

Correspondence
Mehrdad Hajibabaei, Centre for Biodiversity Genomics & Department of Integrative Biology, University of Guelph, Guelph, ON, Canada.
Email: mihajib@uoguelph.ca

Abstract
The purpose of this review is to present the most common and emerging DNA-based methods used to generate data for biodiversity and biomonitoring studies. As environmental assessment and monitoring programmes may require biodiversity information at multiple levels, we pay particular attention to the DNA metabarcoding method and discuss a number of bioinformatic tools and considerations for producing DNA-based indicators using operational taxonomic units (OTUs), taxa at a variety of ranks and community composition. By developing the capacity to harness the advantages provided by the newest technologies, investigators can “scale up” by increasing the number of samples and replicates processed, the frequency of sampling over time and space, and even the depth of sampling such as by sequencing more reads per sample or more markers per sample. The ability to scale up is made possible by the reduced hands-on time and cost per sample provided by the newest kits, platforms and software tools. Results gleaned from broad-scale monitoring will provide opportunities to address key scientific questions linked to biodiversity and its dynamics across time and space as well as being more relevant for policymakers, enabling science-based decision-making, and provide a greater socio-economic impact. As genomic approaches are continually evolving, we provide this guide to methods used in biodiversity genomics.

KEYWORDS
biodiversity, bioinformatics, biomonitoring, DNA barcode, environment, high-throughput sequencing, metabarcoding, metagenomics

1 INTRODUCTION
Biodiversity encompasses the diversity of organisms, their relationships and their functions within ecosystems. Biodiversity assessment using traditional methods mainly involves identifying morphological characters whose states can be compared with taxonomic keys for species identification. It is not uncommon that characters needed for taxonomic assignment are not present or are difficult to discern even for highly experienced taxonomists. The process of identifying smaller taxa such as insects or microscopic organisms from environmental samples often continues well beyond the field collection season, and the results are taxonomic assignments with varying degrees of resolution that is dependent on the availability of taxonomic keys, expertise of the taxonomist and condition of the sample. The issues vary by organism: for insects, damaged or juvenile specimens may not contain the characters needed for identification (Sweeney, Bathc, Jackson, & Dawley, 2011) for fungi, bacteria and other microscopic organisms, it may be the difficulty in isolating and culturing individuals or the collection of samples in life stages that lack the characters needed for identification (Bride & Sparrow, 2003). These are just a few of the impediments to the morphology-based identification process that also affect downstream users of taxonomic data for

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. Molecular Ecology Published by John Wiley & Sons Ltd


wileyonlinelibrary.com/journal/mec | 333
Target Libraries: Bioindicators & Endangered Species

Porter and Hajibabaei, 2018
Automated High-throughput Metabarcode Classification
The reference condition approach
1987 to 1988

Canadian Aquatic Biosurveillance Network

RCBA
Changing how we approach biomonitoring

**Biomonitoring 1.0**
- Restricted information content
- Family- and genus-level ID
- Only part of sample is processed
- Limited to fall sampling (i.e. 3 months)

**Biomonitoring 2.0**
- Rich information content
- Genus- and species-level ID
- Whole sample is processed
- Year-round, consistent inter-annual observation
Biomonitoring 2.0

Alberta Oil Sands

Peace Athabasca Delta

Bennett Dam
Marl ponds, Wood Buffalo National Park, Canada. Photo: DJ Baird
Multiple habitats and sample types

Rocher wetland, Peace-Athabasca Delta
Biomonitoring 2.0
MORPHOTAXONOMY: lowest taxonomic level

DNA METABARCODING: genus-level taxa
MORPHOTAXONOMY: lowest taxonomic level

SPACE: Site

DNA METABARCODING: genus-level taxa

DNA METABARCODING: family-level traits

DNA METABARCODING: genus-level taxa
21.9% of DNA metabarcoded taxa found only in Peace Delta sites

23.0% of DNA metabarcoded taxa found only in Athabasca Delta sites

11.3% of morpho taxa found only in Athabasca Delta sites

14.8% of morpho taxa found only in Peace Delta sites
DNA METABARCODING:
genus-level taxa

DNA METABARCODING:
family-level traits

MORPHOTAXONOMY:
lowest taxonomic level

SPACE:
Site

TIME:
Annual
Morphotaxonomy

2012: 71.2%
2013: 63.2%
2014: 66.0%
2015: 64.8%

DNA metabarcoding

2012: 31.1%
2013: 19.4%
2014: 20.2%
2015: 31.9%
What did we learn from the case study?

- Added taxonomic resolution from DNA metabarcoding provides increased resolution of alpha, beta, and gamma biodiversity

- Increased spatial discrimination over more traditional methods but less inter-annual variation as compared to morphotaxonomy
Benthic DNA metabarcoding for forest ecological integrity

- Hearst Forest, Ontario. Land base of 1,231,707 million hectares
- Over 10,000 km² of productive forest
- Collaboration with Canadian Forest Services of Natural Resources Canada
- Environmental gradients & different management regimes
- 23 watersheds, Illumina MiSeq 100K reads/per sample
Benthic DNA metabarcoding and environmental variables

Emilson et al. 2017
Ethanol metabarcoding vs morphotaxonomy

\[ r = 0.81 \]

\[ r = 0.82 \]
Volunteers shimmy and shuffle in the river for science

WWF

SEPTEMBER 26, 2017

Four volunteers became citizen scientists on the picturesque shore of the Ottawa River by donning hip-waders and gloves before kicking up the surface of the water to collect benthic invertebrates (bugs!). Led by Living Lakes Canada experts, the volunteers joined World Wildlife Fund Canada’s David Miller in practicing a kick-test shimmy on shore before venturing into Remic Rapids to do their part in assessing the health of the Ottawa River watershed to help reverse the decline of wildlife in the ecosystem.
Water monitoring in the 21st Century: DNA and Citizen Science
National Freshwater DNA metabarcoding, launched October 01, 2018
illumina NovaSeq vs. MiSeq

25 million reads (MiSeq v3 flow cell)

\[ \times 52 \]

1.3 billion reads

*Or 1 NovaSeq S1 flow cell*

Per NovaSeq flow cell:
- S1: 1.3-1.6 billion reads
- S2: 3.3-4.1 billion reads
- S4: 8-10 billion reads
NovaSeq – a single flow cell
Acknowledgments

Michael Wright (Lab Manager-UGuelph)
Terri Porter (Research Scientist, UGuelph & NRCan)
Katie McGee (Postdoc-UGuelph)
Siavash Nazari (Graduate student - UGuelph)
Josip Rudar (Graduate student – Uguelph)
Joel Gibson (Former Postdoc – UGuelph)
Shadi Shokralla (Former Postdoc – UGuelph)
Ian King (Former Postdoc – UGuelph)
Donald Baird (Co-PI – UNB & Environment & Climate Change Canada)
Brian Golding (Co-PI – UMcMaster)
Daniel Peters (Scientist - Environment & Climate Change Canada)
Kristie Heard (Taxonomist - Environment & Climate Change Canada)
Adam Martens (Lab Tech – Environment & Climate Change Canada)
Daryl Halliwell (Lab Tech - Environment & Climate Change Canada)
Greg Singer (CEGA)
Nicole Fahner (CEGA)
Adam Bliss (Lab Tech - UNB)
Ronnie Campbell (Field Crew -Parks Canada)
Jeff Shatford (Field Crew – Parks Canada)
Colin Curry (PDF – UNB)
Steff Connor (Graduate Student – UNB)

Funding

- Environment & Climate Change Canada
- Genome Canada
- Ontario Genomics Institute
- Parks Canada
- MRI (Ontario)
- NSERC
- Canada Foundation for Innovation
- Ontario Ministry of Environment & Climate Change
- WWF Canada