MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 2039–2044

NEWS AND VIEWS

OPINION

Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing

DONALD J. BAIRD* and MEHRDAD HAJIBABAEI†

**Department of Biology, Environment Canada @ Canadian Rivers Institute, University of New Brunswick, Fredericton, New Brunswick, Canada E3B 6E1, †Department of Integrative Biology, Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada N1G 2W1*

Abstract

Biological monitoring has failed to develop from simple binary assessment outcomes of the impacted/unimpacted type, towards more diagnostic frameworks, despite significant scientific effort over the past fifty years. It is our assertion that this is largely because of the limited information content of biological samples processed by traditional morphology-based taxonomy, which is a slow, imprecise process, focused on restricted groups of organisms. We envision a new paradigm in ecosystem assessment, which we refer to as 'Biomonitoring 2.0'. This new schema employs DNA-based identification of taxa, coupled with high-throughput DNA sequencing on next-generation sequencing platforms. We discuss the transformational nature of DNA-based approaches in biodiversity discovery and ecosystem assessment and outline a path forward for their future widespread application.

Keywords: biodiversity discovery, biomonitoring, DNA barcoding, high-throughput gene sequencing, multiple stressors

Received 16 October 2011; revision received 16 December 2011; accepted 20 December 2011

Biomonitoring 1.0 and the development of diagnostic indicators

As the human impacts on our planet continue apace (Global Footprint Network, 2010), the need to increase the scale and frequency of observation of the biodiversity of natural ecosystems to support their wise use has never been more pressing. Trade globalization, coupled with the emergence of new technologies, and a changing climate are combining to increase the rate of movement of people and goods and to facilitate resource exploitation on a scale unprecedented in human history, causing widespread, irreversible environmental degradation. The daunting challenge that represents for ecosystem monitoring and assessment will require a revolution in monitoring technologies, driven by the need for new tools which will support more rapid, accurate and timely observations of ecosystem structure and function.

Current methods for ecosystem biomonitoring follow a traditional approach of limited, local-scale site sampling, followed by a lengthy period of processing and enumeration of sample taxonomic units, which can take months to years, and often generates data of low, often unverifiable taxonomic precision. Moreover, such biomonitoring is necessarily limited to observations on highly restricted sets of organisms (e.g. aquatic macroinvertebrates (Bonada *et al.* 2006; Magurran *et al.* 2010), algae (Reavie *et al.* 2010), microbes (Dequiedt *et al.* 2011), with little consistency in observation methods across ecosystem types. As a result, biomonitoring programs are generally restricted to simple binary outcomes (e.g. impacted/not impacted), and while these may be subdivided to report levels of impact (e.g. not impacted/moderately impacted/highly impacted) or gradients of effect, insight into the causes of impact is generally limited to conjectural expert judgment. This 'binary outcome' approach we henceforth designate as 'Biomonitoring 1.0', as it is the prevailing paradigm in biological monitoring programs today. A schematic illustration of this approach is given in Fig. 1. More recently, interest has focused on the development of diagnostic approaches for environmental monitoring, which aim to yield insight into

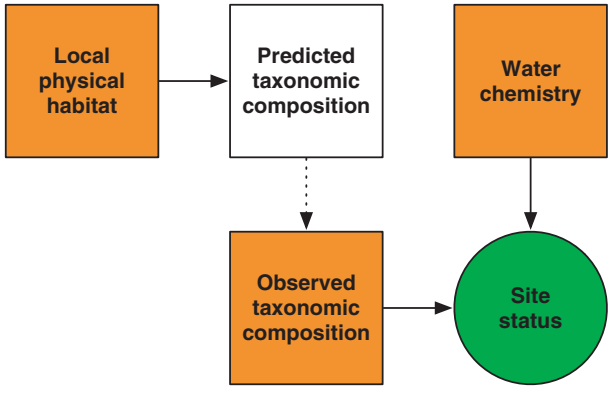

Fig. 1 Schematic representation of 'Biomonitoring 1.0' – the current norm for river biomonitoring. Observational data (orange boxes) are used to determine site status based on binary outcomes and draw inferences regarding putative causes of impact.

Correspondence: Donald J. Baird, Fax: +1 506 452 6271;
E-mail: djbaird@unb.ca

the importance of specific causal agents within complex stressor scenarios. While it is not the purpose of this short perspective to explore these approaches in any detail, it is important to consider the nature of the problem that they seek to address. In doing this, we draw mostly from the area of river monitoring with macroinvertebrates, as this is the most widely practiced biomonitoring technique (Bonada *et al.* 2006), although our arguments broadly apply to other ecosystem types and monitoring situations.

Complex stressor scenarios in ecosystems can be illustrated by the problem of attempting to determine the impact of pesticide entering a river flowing through an agricultural region (Fig. 2). While biomonitoring samples may indicate a change in the community relative to adjacent reference areas where agriculture is limited or absent, it is unwise to attribute cause to pesticides alone (if at all), as many other factors covary with agricultural land-use intensity and hence also with pesticide emissions (e.g. nutrient emissions, sediment from soil tillage, increased temperature from riparian deforestation are some examples). Teasing apart these multiple causes requires some knowledge of the relative sensitivity to the various stressor agents of each of the organisms, which make up the community. Some have attempted to do this using the response of sensitive organisms based on a priori knowledge of sensitivity to pesticides (e.g. Schriever & Liess 2007). While this approach is promising, it has yet to convincingly demonstrate that single agents can be isolated from the 'causal thicket' (*sensu* Harris & Heathwaite 2011; after Wimsatt 2007) of largely collinear stressor variables. This is because of the fact that while sampled species each exhibit unique features, they are often phylogenetically related, which places limitations on phenotypic expression, particularly within the highly restricted subsets of species that are enumerated in biomonitoring samples. For example, in Canadian rivers, such samples tend to be dominated by the larval stages of insects. As a result, when examining the responses of an assemblage of species to a stressor regime, there are rarely enough species within the local community pool to permit the differential responses of individual species to individual stressors to be resolved. In other words, the samples themselves do not possess sufficient 'resolvable information content', to simultaneously identify and isolate specific stressor responses from a background of natural variability and co-acting stressors. These approaches are further constrained by the fact that in most cases, the quality of data that can be obtained from biomonitoring samples is necessarily constrained by cost, generally resulting in samples being identified only to a very coarse taxonomic level (e.g. family-level identification) despite clear evidence that, even for binary outcome assessment, genus- or species-level identification will produce more robust assessment outcomes (Lenat & Resh 2001). It is our assertion that if it was possible to achieve a step-increase in data information content by observing and analysing responses of the entire biota, this could increase our ability to separate individual stressor responses from complex stressor scenarios. This is supported by two observations: (i) the ability to simultaneously study responses of a broader range of species from the Tree of Life will permit a broader range of biological receptor responses to be included in any analysis and (ii) the increase in information content will permit more computationally intensive methods to be applied, which have previously been

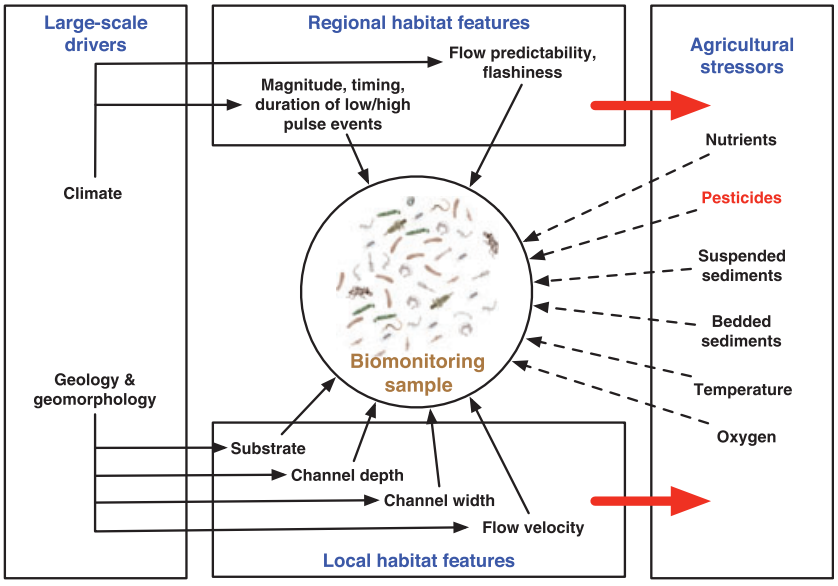

Fig. 2 Schematic representation of the influence of a single stressor (pesticide emissions) on the structure of a river macroinvertebrate community, placed in the context of other co-acting large-scale, regional-scale and local-scale influences, together with other examples of co-acting agricultural stressors. The large red arrows also emphasize the additional influence of natural habitat factors on the stressor regime, in addition to their direct influences on community composition, emphasizing the need to consider indirect and direct effects when examining causality in multiple stressor scenarios.

unworkable because of overfitting issues (e.g. Tetko *et al.* 1995).

The above arguments demonstrate an urgent need to develop biomonitoring sample processing methods, which can extract significantly more information, to begin the development of robust diagnostic assessment tools. Biomonitoring 1.0 has struggled to deliver such methods, and the approaches being developed using traditional data are, in our view, unlikely to yield significant new breakthroughs in terms of diagnostic assessment in the short-to-medium term. This is supported by a recent review of the biomonitoring science by Jones *et al.* (2010), who note that our knowledge of the responses of most biota to anthropogenic stressors is 'still lacking' and that '...other approaches...may have to be used to achieve the required goals *[of the European Water Framework Directive]* within the available time frame'. In this context, if we rule out the option of using morphological taxonomy to generate species-level information for selected taxa groups (e.g. macroinvertebrates) as impractical, and the possibility of combining multiple groups analysed in this way as doubly so, are there other potential approaches which might be explored?

Environmental barcoding through next-generation sequencing (NGS)

The advent of next-generation sequencing technologies (for a review, see Shokralla *et al.* 2012 in this issue), coupled with the rapid advance of DNA- and RNA-based techniques for taxonomic identification offers a possible solution to the limitations of Biomonitoring 1.0. Microbial ecologists pioneered the use of NGS-based metagenomics studies as previously used cloning-based metagenomics techniques had limitations in throughout cost and biases associated with them. These studies span from applications in exploring and discovering microbial groups in various environments (e.g. Sogin *et al.* 2006) to the analysis of human microbiome in various clinical settings (e.g. Ley *et al.* 2006). Many studies, subsequently, have used NGS approaches in the analysis of biodiversity in various habitats and taxonomic groups from all domains of life (see several articles published in this special issue). With the advancement of methodologies in data generation and analysis, NGS tools can become more feasible in clinical or environmental studies. Recently, we demonstrated the potential to use these technologies to extract species-level information on key bioindicator insect species from standard river biomonitoring samples, using a combination of cytochrome c oxidase (COI) DNA barcodes linked to a locally generated barcode reference library (Hajibabaei *et al.* 2011). While our proof of concept study was limited to aquatic insects, it clearly has wider application: NGS platforms offer a huge potential increase in the potential information that can be generated. In the case of the Roche 454 system, with approximately 1 M sequencing reads per run, it is now possible to consider analysis of environmental samples in terms of the simultaneous analysis of different taxonomic groups. This alone offers the potential to transform biomonitoring from a binary assessment tool into a new set of tools that uses this highly enriched data source to move from binary assessment to diagnostic assessment and also to provide a rich source of new data for the purposes of biodiversity observation. As an example, a river biomonitoring macroinvertebrate sample is generally resolved in the 10^1–10^2 range for families. If extra effort were applied, this might increase to the low end of the 10^2–10^3 range, but would not be cost-effective or practical, as few laboratories would be capable of this level of taxonomic competency. Extracting nucleotide sequence information from the same sample could easily and consistently yield species-equivalent operational taxonomic units (OTUs) in the 10^3–10^4 range (encompassing all biota from microbes to metazoa) at a comparable cost. This represents an increase in a hundredfold to a thousandfold in sample information content, which is a sufficient basis to begin development of a new generation of biomonitoring tools. That this can also be cost-effective is supported by our own experience that the cost of sequencing a standard river biomonitoring sample using 454 pyrosequencing was ca. $1000 in 2008, but has reduced to ca. $500 in 2011, as a result of improvements in plate design and whole-sample tagging methods (unpublished data). New high-throughput sequencing platforms are likely to reduce this further, as costs per megabase continue to fall at an increasing rate (http://genome.gov/sequencingcosts – queried 14/12/2011).

Biomonitoring 2.0: consistently observed biodiversity data on an epic scale?

Internationally, efforts to monitor biodiversity have struggled to deliver the scale of observation necessary to make confident statements on the state of global ecosystems based on the full breadth of their biological diversity. Monitoring remains highly skewed towards population-focused assessments of charismatic megafauna, as these groups are readily observed and quantified. However, our knowledge of the distribution and abundance of species in the lower strata of ecological food webs remains woefully inadequate to monitor status and trends, except on very small spatial scales. For example, in a recent authoritative review by Butchart *et al.* (2010) on global biodiversity status and trends, the only available indicator by which to assess global biodiversity status of freshwater ecosystems was water chemistry. The increasing prevalence of 'global biodiversity status and trends' articles in major science journals, websites and international reports is belied by some inconvenient scientific realities: trends are often based on aggregated study data collected using inconsistent methods, across a range of differing time periods, with little or no formal quality control. In short, we are making policy decisions based on highly limited data, which may significantly constrain or otherwise bias policy development. For example, we still make conservation decisions that are ultimately species-focused, but which imply nonexistent knowledge of ecosystem goods and services, in terms of the roles and functions of species who provide them. There

is an urgent need to redress the balance of ecosystem observation towards these less charismatic species, while sustaining efforts to preserve species that are well-known and cherished by society – as Bowen (1999) states: 'Perpetuating species without ecosystems makes as much sense as preserving ecosystems without species'.

The advent of NGS tools to consistently and cost-effectively extract information from environmental samples offers the beginning of a solution to this problem. Until recently, this technology had largely been applied to the issue of exploring microbial diversity (e.g. Sogin *et al.* 2006; Zinger *et al.* 2011), where traditional taxonomic approaches have limited application. Other categories of applications involve exploring ancient biota using short diagnostic DNA sequences generated by NGS tools from ice cores or permafrost (e.g. Willerslev *et al.* 2007) and diet analysis for various groups of organisms from faeces or stomach contents (e.g. Valentini *et al.* 2009). Additionally, a number of recently published studies have illustrated the application of this technology for bulk specimens of metazoan origin with linkage to environmental assessment: Creer *et al.* (2010) explored the use of NGS approaches to estimate biodiversity in marine meiofaunal communities, and Chariton *et al.* (2010) have explored a similar approach in marine sediments, where they explored the use of NGS data to study the impact of contaminants on the biota of Australian coastal areas. Hajibabaei *et al.* (2011) have also explored this approach in the context of river biomonitoring, where NGS was used to extract standard COI DNA barcode sequences to obtain species-level information from standard biomonitoring samples to contrast patterns of taxon occurrence in urbanized and conservation habitats. What these three studies illustrate is, it is possible to extract rich biodiversity data from a standard biomonitoring sample using next-generation sequencing. Moreover, they emphasize that the data generated in such studies have new properties: they contain a mix of 'named' OTUs (i.e. those DNA sequences that can be linked to a Linnean taxonomic name from an relevant database such as GenBank or the Barcode of Life Database) and 'unnamed' OTUs (i.e. DNA sequences that can be placed in a phylogenetic context, but that have not previously been deposited in a database). For certain, well-studied taxonomic groups such as biomonitoring indicator species, these online barcode databases are being rapidly populated for less-studied groups, and generally for lower metazoa, protists and microbes, DNA barcode/Linnean taxonomic name linkages are not likely to be available in the near future. The use of such data for bioassessment purposes constitutes a challenge in terms of statistical analysis and interpretation and illustrates the transformative nature of DNA-based biomonitoring techniques in relation to morphological taxonomy-based approaches. While it may be possible to align DNA-based observations with traditional taxonomic observations, the potential of this new approach should not be limited by attempting to fit it into the Biomonitoring 1.0 schema. DNA-based biodiversity information extracted from biomonitoring samples offers a step change in the immediacy, accuracy and quantity of observable information, which can be obtained without sacrificing current biomonitoring infrastructure investment. Where changes in sample collection protocols may be necessary to avoid

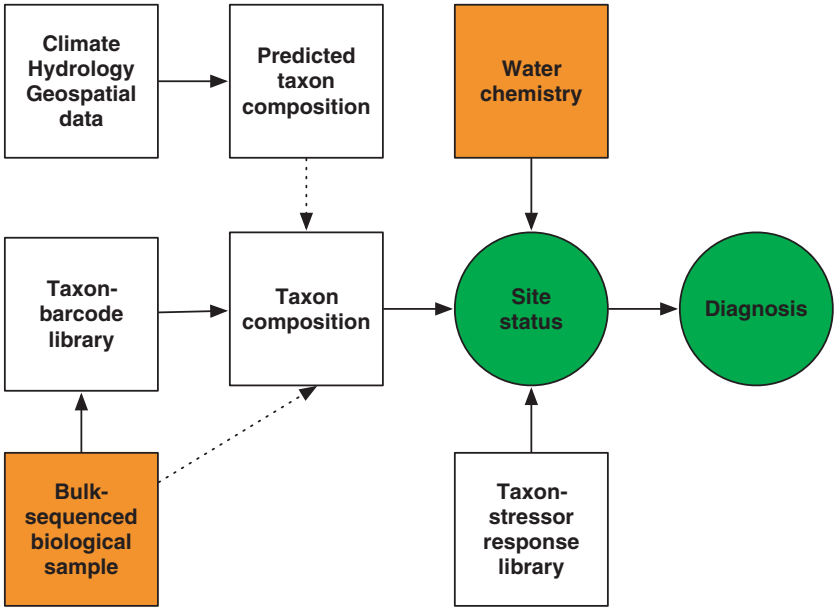

Fig. 3 Schematic representation of 'Biomonitoring 2.0' – a new proposed schema for ecosystem biomonitoring. Biological samples are subject to high-throughput gene sequencing, linked to DNA/RNA barcode libraries, allowing a more complete profile of biodiversity content, generating a rich data source for effects interpretation and causal analysis. The dashed line between the bulk sample and taxon analysis boxes indicates the potential for OTU assignment without reference to barcode libraries (e.g. Fitch & Margoliash 1967).

cross-contamination between sites, these more rigorous requirements actually go with the grain of current thinking in biosecure sampling protocols (e.g. Bothwell & Spaulding 2008).

Figure 3 illustrates an alternative schema for ecosystem monitoring (using river monitoring as an example), which we designate as 'Biomonitoring 2.0'. The key advance offered by this new approach is the harnessing of the information-rich biodiversity data stream, which can be delivered now by existing next-generation sequencing platforms towards the investigation of cause in complex environmental stressor scenarios. We believe that this approach is set to revolutionize not only biodiversity observation, but also poised to transform the management of human impacts on the biosphere, through the provision of much more nuanced and robust advice regarding prioritization of amelioration and remediation efforts. However, there are major challenges ahead: much work remains to be done regarding the reliability and repeatability of DNA-based taxonomic assignment. Moreover, a shift from sampling individual organisms towards sampling their DNA offers new opportunities for the future such as the development of real-time automated monitoring systems, while at the same time requires careful consideration of data interpretation. DNA can persist beyond the lifespan of an individual (Dejean *et al.* 2011) and dealing with such 'zombie DNA' currently poses a challenge in data interpretation. In addition to taxonomic identification, there is a general desire to generate information on the relative abundance of organisms in biomonitoring samples, and while there is some evidence that this might be possible (Hajibabaei *et al.* 2011), it remains challenging, particularly given the risks of bias from differential PCR amplification (Polz & Cavanaugh 1998). The advent of new sequencing protocols and platforms (e.g. Clarke *et al.* 2009) may eliminate this problem altogether, however. In addition, once the taxonomic/phylogenetic profiles of communities are characterized accurately, sequencing at the transcriptome level offers the possibility to tease apart functional diversity among different communities (Bailly *et al.* 2007; Urich *et al.* 2008). These efforts can link taxon-based biomonitoring to trait-based biomonitoring (Menezes *et al.*, 2011) where in this case, traits can be defined as transcriptome profiles or specific transcript changes across communities or environments.

A path forward for Biomonitoring 2.0?

Demonstrating the value of an entirely novel approach requires proof of concept. While this process has begun, and Hajibabaei *et al.* (2011) have demonstrated what is possible, there is still a long road ahead. Discussions are continuing regarding harmonization of approaches, including selection of candidate genes to provide a comprehensive coverage of biota and the need to link ongoing ecosystem-focused projects currently underway in Canada, Australia, UK and the USA to focus on a common goal. Also of importance is the development of the bioinformatics and ecoinformatics tools to permit seamless integration and intepretation of the 'big data' generated by gene-based observation. Some ideas on how to bring together work in these areas was recently highlighted by Baird *et al.* (2011), but much work remains to be done, particularly in terms of how to deal with mixed data sources comprising named taxa and OTUs in the development of diagnostic indices, and also on the practical aspects of how to collect, preserve and subsequently analyse field samples in a manner, which is compatible with current and future DNA analysis methods. Despite these challenges, the fields of gene-based biodiversity discovery and biomonitoring diagnostics development are entering an exciting and rapidly accelerating phase, which must move forward in a spirit of international collaboration if it is to deliver on its promising beginnings.

Acknowledgements

This work was supported by the Government of Canada through funds from Environment Canada, Genome Canada and the Ontario Genomics Institute (OGI-050) and by NSERC Canada.

References

Bailly J, Fraissinet-Tachet L, Verner MC (2007) Soil eukaryotic functional diversity: a metatranscriptomic approach. *ISME Journal*, **1**, 632–642.

Baird DJ, Baker CJO, Brua RB *et al.* (2011) Towards a knowledge infrastructure for traits-based ecological risk assessment. *Integrated Environmental Assessment and Management*, **7**, 209–215.

Bonada N, Prat N, Resh VH, Statzner B (2006) Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. *Annual Review of Entomology*, **51**, 495–523.

Bothwell ML, Spaulding SA (2008) Proceedings of the 2007 International Workshop on *Didymosphaenia geminata*. *Canadian Technical Report of Fisheries and Aquatic Sciences*, **2795**, xxxv + 58p.

Bowen BW (1999) Preserving genes, species or ecosystems? Healing the fractured foundations of conservation policy. *Molecular Ecology*, **8**, S5–S10.

Butchart SM, Walpole M, Collen B *et al.* (2010) Global biodiversity: indicators of recent declines. *Science*, **328**, 1164–1168.

Chariton AA, Court LN, Hartley DM *et al.* (2010) Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Frontiers in Ecology and the Environment*, **8**, 233–238.

Clarke J, Wu H-C, Jayasinghe L *et al.* (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology*, **4**, 265–270.

Creer S, Fonseca VG, Pozarinska DL *et al.* (2010) Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology*, **19**(Suppl 1), 4–20.

Dejean T, Valentini A, Duparc A *et al.* (2011) Persistence of environmental DNA in freshwater ecosystems. *PLoS One*, **6**, e23398.

Dequiedt S, Saby NPA, Lelievre M *et al.* (2011) Biogeographical patterns of soil molecular microbial biomass as influenced by soil characteristics and management. *Global Ecology and Biogeography*, **20**, 641–652.

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. *Science*, **155**, 279–284.

Hajibabaei M, Shokralla S, Zhou X *et al.* (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One*, **6**, e17497. doi:10.1371/journal.pone.0017497.

Harris GP, Heathwaite L (2011) Why is achieving good ecological outcomes in rivers so difficult? *Freshwater Biology*, doi: 10.1111/j.1365-2427.2011.02640.x.

Jones JI, Davy-Bowker J, Murphy JF *et al.* (2010) Ecological monitoring and assessment in rivers. In: *Ecology of Industrial Pollution* (eds Batty L, Hallberg KB), pp. 126–146. Cambridge University Press, Cambridge, UK.

Lenat DR, Resh VH (2001) Taxonomy and stream ecology – the benefits of genus- and species-level identifications. *Journal of the North American Benthological Society*, **20**, 287–298.

Ley RE, Turnbaugh PJ, Klein S *et al.* (2006) Microbial ecology: human gut microbes associated with obesity. *Nature*, **444**, 1022–1023.

Magurran AE, Baillie SR, Buckland ST *et al.* (2010) Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends in Ecology and Evolution*, **25**, 574–582.

Menezes S, Baird DJ, Soares AMVM (2010) Beyond taxonomy: a review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring. *Journal of Applied Ecology*, **47**, 711–719.

Polz MF, Cavanaugh CM (1998) Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology*, **64**, 3724–3730.

Reavie ED, Jicha TM, Angradi TR *et al.* (2010) Algal assemblages for large river monitoring: comparison among biovolume, absolute and relative abundance metrics. *Ecological Indicators*, **10**, 167–177.

Schriever CA, Liess M (2007) Mapping ecological risk of agricultural pesticide runoff. *Science of the Total Environment*, **384**, 264–279.

Shokralla S, Spall JL, Gibson JF *et al.* (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, **21**, doi:10.1111/j.1365-294X.2012.05538.x.

Sogin ML, Morrison HG, Huber JA *et al.* (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proceedings of the National Academy of Sciences*, **103**, 12115–12120.

Tetko IV, Livingstone DJ, Luik AI (1995) Neural network studies. 1. Comparison of overfitting and overtraining. *Journal of Chemical Information and Computing Science*, **35**, 826–833.

Urich T, Lanzén A, Qi J *et al.* (2008) Simultaneous assessment of soil microbial community structure and function through analysis of the metatranscriptome. *PLoS One*, **3**, e2527.

Valentini A, Miquel C, Ali Nawaz M *et al.* (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Molecular Ecology Research*, **9**, 51–60.

Willerslev E, Cappellini E, Boomsma W *et al.* (2007) Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*, **317**, 111–114.

Wimsatt WC (2007) *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press, Harvard, Connecticut.

Zinger L, Amaral-Zettler LA, Fuhrman JA *et al.* (2011) Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS One*, **6**, e24570. doi:10.1371/journal.pone.0024570.

D.J.B. is an aquatic ecologist interested in the application of ecological knowledge in the assessment of human impacts on the freshwater environment; M.H. is a molecular evolutionary biologist interested in studying biodiversity in different ecological settings through comparative analysis of genomics information. He currently leads, the Biomonitoring 2.0 project (http://www.biomonitoring2.org), a large-scale effort that utilizes next-generation sequencing technologies for monitoring environmental change.

---

doi: 10.1111/j.1365-294X.2012.05519.x