Freshwater biomonitoring in the Information age
Francois Keck, Valentin Vasselon, Kalman Tapolczai, Frédéric Rimet, Agnes Bouchez

To cite this version:
Francois Keck, Valentin Vasselon, Kalman Tapolczai, Frédéric Rimet, Agnes Bouchez. Freshwater biomonitoring in the Information age. Frontiers in Ecology and the Environment, Ecological Society of America, 2017, 15 (5), pp.266-274. 10.1002/fee.1490. hal-01604509

HAL Id: hal-01604509
https://hal.archives-ouvertes.fr/hal-01604509
Submitted on 26 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License
Freshwater biomonitoring in the Information Age

François Keck1,2*, Valentin Vasselon1, Kálmán Tapolczai1, Frédéric Rimet1, and Agnès Bouchez1

Freshwaters worldwide face serious threats, making their protection increasingly important. Freshwater monitoring has historically produced valuable data and continues to develop. Rapid improvements to biomolecular techniques are revolutionizing the way scientists describe biological communities and are bringing about major changes in biomonitoring. Combined with high-throughput sequencing, DNA metabarcoding is fast and cost-effective, generating massive amounts of data. In a world with numerous ecological threats, “big data” constitute a tremendous opportunity to improve the efficiency of biological monitoring. These fundamental changes in biomonitoring will require freshwater ecologists and environmental managers to reconsider how they handle large amounts of data.

Front Ecol Environ 2017; 15(5): 266–274, doi:10.1002/fee.1490

Human activities have broadly affected freshwater ecosystems, especially since the Industrial Revolution. Over the past 50 years, however, policy makers and citizens have become more attuned to environmental issues. This has led to the development of important governmental programs to assess and limit ecological impacts of human activities (Figure 1). In this context, one objective of environmental managers is to evaluate how water quality changes over time. Bioindicator organisms are commonly used for this purpose, based on the premise that the presence or absence of certain biological communities at a given site reflects its environmental quality.

Freshwater biomonitoring has a long tradition in the field of ecology. A century of research has led to substantial improvements in understanding how human disturbances can shape biological communities. Based on this knowledge, many approaches have been developed to estimate environmental quality from the richness, diversity, structure, and functioning of these communities (Jørgensen et al. 2010). These widely used methods are based on solid theoretical grounds and are known to perform quite well. Most of them commonly require a taxonomical description of the community. Hence, freshwater biomonitoring essentially consists of collecting individual organisms, performing taxonomic identification, and using inventories to estimate the environmental condition of a given site. However, traditional biomonitoring also faces recurrent criticisms, mainly related to taxonomic identification relying on morphological criteria, a process that is time-consuming, complex, and technically demanding (Mandelik et al. 2010). These limits inevitably restrict the number of sites that can be monitored and the frequency of controls.

During the past decade, the idea arose that DNA analyses could advantageously replace morphological methods to identify species (Hebert et al. 2003). Metabarcoding was developed as a set of techniques to identify multiple taxa simultaneously from an environmental sample with standard genetic markers (Taberlet et al. 2012; Panel 1 and Figure 2). This has led to the idea of “Biomonitoring 2.0”, which offers novel perspectives for monitoring environmental communities (Baird and Hajibabaei 2012). In this paper, we explain why and how metabarcoding will profoundly change the nature of data produced by biomonitoring. We examine these changes in the general context of massive data production – so-called “big data”, a topic that is the subject of increasing interest in biology (Marx 2013). We show why this big data revolution holds promise for ecological assessment purposes. Finally, we highlight three challenges posed by big data for metabarcoding and propose a framework that takes them into account. We illustrate our point with examples taken from freshwater monitoring, where metabarcoding is developing rapidly (Hajibabaei et al. 2011; Kermarrec et al. 2014). Nevertheless, the ideas discussed could be extended and applied to a broader context.

In a nutshell:
- DNA metabarcoding and high-throughput sequencing methods produce massive quantities of data and will markedly change freshwater biomonitoring
- Molecular methods propel biomonitoring into the Information Age and bring exciting new opportunities to make ecological monitoring more effective and relevant
- Genetic “big data” challenge scientists to think differently about the way that biological monitoring information is analyzed; we propose and discuss alternatives to the classical taxonomic affiliation approach to process bioassessment metabarcoding data

1UMR CARRTEL, Institut National de la Recherche Agronomique, Université Savoie Mont Blanc, Thonon-Les-Bains, France; 2Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden *(francois.keck@gmail.com)
Biomonitoring as a source of massive data

Characterizing ecological quality from biological entities has produced important sources of data since the first attempts to do so at the beginning of the 20th century. This is because biomonitoring largely consists of sampling, identifying, enumerating, and reporting biological organisms. The saprobic system for organic pollution assessment developed by Kolkwitz and Marsson (1908, 1909) is often cited as the first bioassessment tool in freshwaters and uses 298 plant species and 527 animal species as indicator organisms. Methods soon diversified thereafter, and specific biological groups (fishes, macroinvertebrates, algae) have been employed. Increasing stringency in precision requirements has led to more powerful and sophisticated tools, based on hundreds of families and thousands of species.

The amount of data produced has increased rapidly because biomonitoring is rarely done in isolation, but instead is replicated across space (through a network of sites; eg along a river, within a watershed) and over time (long-term monitoring). Since the 1970s, general awareness of ecological issues has grown, and biomonitoring has been increasingly implemented and incorporated into legal frameworks for fresh waters, such as the Clean Water Act (CWA, 1972) in the US and the Water Framework Directive (WFD, 2000) in Europe. This guarantees the abundant production of data with respect to recognized standards.

However, biomonitoring methods are expected to change considerably in coming years. After a century of classifying taxa based on morphological criteria, species can now be identified through the use of DNA barcodes (Hebert et al. 2003); for definitions of selected specialist terms used throughout, see Panel 1. The introduction of high-throughput sequencing (HTS; Shokralla et al. 2012) coupled with the development of extended reference databases (Ratnasingham and Hebert 2007; Benson et al. 2008) and efficient bioinformatics tools (eg Schloss et al. 2009) have enabled the production of reliable and cost-effective community inventories from environmental DNA (Chariton et al. 2015; Gibson et al. 2015; Pawlowski et al. 2016). While numerous issues and technical limitations remain (DNA spatial transfer and persistence over time, polymerase chain reaction [PCR] amplification biases, sequencing errors, chimeras, quantification; see also Coissac et al. 2012 and Shokralla et al. 2012), methods are improving quickly and metabarcoding is expected to be an increasingly important component of biomonitoring in the future.

The progressive adoption of metabarcoding for taxonomical identification will substantially increase the volume of data produced by biomonitoring activities and modify the characteristics of these data (Dafforn et al. 2016). It is often stated that characteristics of big data fulfill five “Vs”: volume, velocity, variety, variability, and value (Fan and Bifet 2013). Biomonitoring data will likely meet these five criteria in unprecedented ways in the coming years.

Volume

The amount of data acquired from biomonitoring is expected to increase very quickly. HTS techniques are
developing rapidly and have extremely high-throughput (Figure 3d). With the development of standardized protocols, the processing rate will also probably increase considerably and allow more sites to be surveyed and with greater frequency. Finally, assessments that rely on morphological criteria alone tend to underestimate species diversity, whereas the level of diversity detected by genetic methods tends to be much higher, especially for microbial communities (Caron et al. 2009), leading to larger inventory tables.

**Velocity**

Traditional monitoring requires experts to undertake a long and laborious process of taxonomically identifying collected biota. Consequently, one site is typically monitored seasonally or yearly. With metabarcoding and HTS techniques, however, the identification process is automated and faster. This will allow sites to be monitored at a finer time scale and to approach real-time monitoring.

---

**Panel 1. Biomonitoring and metabarcoding**

The biological monitoring of freshwater systems is traditionally based on the morphological identification of indicator species, which provides information on the ecological status of their environment. Instead of relying on morphological features (e.g., size, shape) to perform species identification, which requires specialized knowledge of taxonomic groups, small DNA fragments—about 300 base pairs in length, known as DNA barcodes—can be used (Hebert et al. 2003). This identification approach is termed DNA barcoding. Existing DNA barcode reference databases are based on different genes (including COI, 18S, and rbcL) and link species taxonomy to DNA barcodes. While DNA barcoding is useful for identifying individual specimens, its application to community-level samples (i.e., multiple species) was difficult because it required sorted samples or even isolating and cultivating individuals. This challenge was overcome through a metagenomic method called metabarcoding, which allows for the detection of all species found in one sample directly from their DNA barcode sequences using a single workflow. The DNA is extracted directly from the sample, followed by the amplification and sequencing of the targeted DNA barcode (Figure 2). Using bioinformatics tools, DNA barcodes are compared to those contained in a reference database to identify the species composition within the sample.

Environmental DNA was defined by Taberlet et al. (2012) as the “DNA that can be extracted from environmental samples (such as soil, water, or air), without first isolating any target organisms.” This includes DNA from microorganisms and free DNA. The free part of environmental DNA may be used to detect the presence of invasive species (Ficetola et al. 2008) or to monitor rare and indicator species (Mächler et al. 2014). Microorganisms present in environmental samples (e.g., bacteria, fungi, and diatoms) enable the use of longer DNA barcodes (Taberlet et al. 2012) and facilitate access to uncultured taxa. For example, diatom molecular inventories can be used to calculate a quality index that indicates the ecological status of the sampled river (Kermarrec et al. 2014; Viscò et al. 2015). Precision and reliability of the species list obtained from DNA metabarcoding depend on the completeness and reliability of the reference database.

The development of high-throughput sequencing (HTS) enables the rapid and inexpensive sequencing of hundreds of environmental samples at a time, making the incorporation of the DNA metabarcoding into biomonitoring programs possible.

---

**Figure 2.** Several steps are required to perform DNA metabarcoding: (i) the sampling of environmental samples (e.g., sediment, biofilm, water) or the creation of bulk samples (mix of individual specimens); (ii) the extraction of the DNA; (iii) the amplification of a DNA barcode specific to the targeted community using polymerase chain reaction (PCR) techniques; (iv) the sequencing of the amplicons (amplified DNA barcodes); and (v) the taxonomic assignment of the DNA reads (amplicon sequences) using bioinformatics and a reference database (database connecting DNA barcode sequences to their taxonomic identity). Total environmental DNA comprises “endogenous” DNA from living organisms and “exogenous” free DNA.
Variety

Biomonitoring elicits multiple types of data. Community inventories generally come in the form of presence-absence or count data tables. Environmental managers often prefer to rely on multiple biological indicators (e.g., fishes and macroinvertebrates) to monitor multiple sources of impairment. Moreover, assessment methods commonly integrate physical and chemical data, which may also constitute big data, especially when recorded with remote sensors and with high frequency. Metabarcoding will also make it possible to work with genetic data and phylogenies (Hajibabaei et al. 2007).

Variability

Biomonitoring data are valuable when there is variability in community structures between reference and impacted sites (Jorgensen et al. 2010). With the use of DNA, finer-scale taxonomic characterization of communities can be achieved. Thus, with appropriate analyses, it will be possible to differentiate communities in a subtler way (Stein et al. 2014a) and to gain capacity in distinguishing the effects of various pressures.

Value

Data produced by biomonitoring are used to assess environmental quality. Many applications could be enhanced with big data, including monitoring over space and time; examining multi-trophic food web structure; and assessing the effects of pollution, environmental restoration, and invasive species. Moreover, biomonitoring data are often exploited by ecologists for purposes other than environmental assessment, such as studying biodiversity patterns or validating theoretical models (Lovett et al. 2007; Lindenmayer and Likens 2010).

Modern techniques and big data

Increasing the number of indicators

The modern concept of biomonitoring – as implemented in the WFD and CWA – is to use biological indicators accompanied by hydromorphological and physicochemical measurements (Ibáñez et al. 2010). For example, the WFD’s bioindicators (or biological quality elements [BQEs]) are fishes, macrophytes, macroinvertebrates, benthic diatoms, and phytoplankton. Each of these indicators presents advantages (e.g., diversity, ubiquity, ecological importance) and disadvantages (sampling difficulties, lack of metrics) (Resh 2008). Each BQE can indicate different pressures and provide complementary information (Passy et al. 2004; Figure 4). Thus, the overall quality assessment of an aquatic ecosystem is based on the results of all BQEs. In the WFD, the “one out all out” (OOAO) rule states that the worst status of the BQEs used in the assessment determines the final status of the ecosystem. However, in practice, using all BQEs for a sampled site is seldom or only partly achieved because of both financial and logistical constraints (Birk et al. 2012).

There is a trade-off between the ease of sampling and the ease of identifying organisms with respect to the average size of different BQEs (Figure 4). Groups of organisms with larger individual body size (typically fishes) are more difficult to sample representatively and collect, whereas smaller or microscopic organisms such as macroinvertebrates or benthic diatoms are relatively easy to collect by sampling the substrate directly. On the other hand, larger organisms are easier to manipulate and identify. For fishes and macrophytes, identification is performed in situ, whereas macroinvertebrates, benthic diatoms, and phytoplankton require arduous laboratory-based work (chemical treatment, microscopy). Modern molecular techniques appear to offer a promising solution to the trade-off between the ease of sampling and identifying organisms.
Covering a larger diversity

In traditional biomonitoring, taxonomical identification is rarely performed at the most precise levels of specificity because doing so is cost-prohibitive. DNA metabarcoding, however, could reveal diversity at the finest level for a fraction of that cost. With appropriate libraries, DNA barcodes can be linked to a Linnaean taxonomic name. The precision of taxonomic affiliation depends on the selected barcode and the availability of data in the reference libraries. By using correctly populated libraries, it is possible to reach the species level (e.g., Hajibabaei et al. 2011; Kermarrec et al. 2014) with less ambiguity and discrepancy than with classical microscopy, where species-level identification is often extremely laborious and even impossible at some development stages. However, data derived from DNA carry much more information than taxonomic names alone. Baird and Hajibabaei (2012) emphasized that genetic techniques have far more potential for identifying taxa than the traditional approach of relying on morphological characteristics. DNA-based techniques should facilitate working at the infra-species level and ultimately at the nucleotide level. It will therefore be possible to disentangle cryptic species complexes and to perform population-level analyses. Having the capacity to monitor diversity at so many levels should also promote the development of very sensitive tools to monitor the effects of specific types of pollution on various biota.
the DNA sequence has appeared as a promising alternative unit. Scientists have tried to integrate genetic sequences in the classical taxonomy, with varying degrees of success (Padial et al. 2010). However, in the context of biomonitoring, the question remains, whether the traditional Linnaean binomial species name affiliation still makes sense within a full molecular approach.

Typically, DNA reads provided by HTS are clustered into molecular operational taxonomic units (MOTUs), which are in turn converted to species units through the use of a bioinformatic workflow and a DNA reference database. The conversion from DNA reads to species units is not without drawbacks: for instance, selected barcodes may be associated with incorrect taxonomic affiliations, genetic information may be lost (unaffiliated reads are discarded), and rare species are often insufficiently studied. This approach is suitable if the reference database is sufficiently comprehensive, but this is rarely the case because of the high species diversity and the time and effort required to sequence organisms’ barcodes. Previously undescribed species are also frequently detected from genetic data, while formal taxonomic description can be a very long process (Goldstein and DeSalle 2011). Moving to full molecular biomonitoring will allow for much more data to be used, beyond that limited strictly to taxonomic assignments. The greatest challenge is to develop new, high-quality indices based on DNA reads and environmental information. Three alternative but complementary approaches are described below and are represented in Figure 5.

Developing MOTU-based indices

Biomonitoring assumes that the presence or absence of particular taxa at a site of interest is indicative of distinct environmental conditions at that site. Thus, in traditional biological assessments, an ecological profile associated with each taxon is required. Pawlowski et al. (2016) suggested calibrating MOTU-based indices with traditional indices computed from simultaneously conducted morphology-based identifications. However, the traditional indices could be easily adapted to the new molecular approach by computing the indices directly from the reads clustered in MOTUs (Steele et al. 2011). This approach would require databases associating reads, MOTUs, and their responses to environmental stressors (Figure 5). Thus, the MOTU-based indices approach is expected to be fully functional when ecological profiles for clusters of reads are estimated directly from previous molecular inventories; this will require substantial work in addition to data compilation and sharing. As a first step, known ecological profiles for taxa can be transferred to MOTUs.

Using phylogeny to include rare species

DNA metabarcoding can reveal a wealth of diversity, but the lack of taxon–stressor response libraries is problematic. Given that ecological profiles are usually estimated from in situ observations of general disturbances or from laboratory bioassays for specific substances, such libraries are restricted to common species and to a few types of disturbances. Rare species are often ignored (Guénard et al. 2011), and the effects of specific compounds remain poorly understood (Schwarzenbach et al. 2006).

One elegant way to solve these problems could involve phylogenetic methods harnessing the principle that species’ tolerances are the legacy of evolution (Keck et al. 2016). The increasing availability of DNA sequences and computational power (Figure 3) should allow for the establishment of large and robust phylogenies. Then, if adequately long and informative (thereby excluding short
fragments and degraded DNA), reads can be inserted in the reference phylogeny using a posteriori replacement algorithms (Matsen et al. 2010; Berger et al. 2011). Finally, recent approaches to predict species’ tolerances based on information available from other species and their respective phylogenetic positions (Guénard et al. 2013) could be used to estimate an ecological profile for a given read (Figure 5). Routine inclusion of such phylogenetic-based methods in biomonitoring would help to account for the immense diversity uncovered by DNA barcoding and the thousands of toxicants in the environment.

**Machine learning techniques for ecological assessment**

Analyzing and extracting valuable information from massive datasets can be extremely challenging. This has encouraged the development of machine learning methods, which use a set of statistical algorithms designed to recognize complex patterns in vast quantities of data. These methods include modern algorithms for classification, such as random forest, gradient boosting, support vector machines, and neural networks (Hastie et al. 2009). Machine learning approaches are fully data-driven and do not rely on any theoretical models (Breiman 2001). This system fits particularly well with the goals of biomonitoring, where the first aim is not necessarily to understand and explain the ecological processes leading to a given observation. In an applied context, correlation approaches are interesting because the final aim is to assess the state of the environment. This does not imply that machine learning should be used indiscriminately, but that these techniques are fully compatible with the ecological monitoring philosophy.

Machine learning methods have a broad range of applications. In biomonitoring, they may be used with different kinds of inputs for site classification, analyses of spatial networks of sites, and time-series forecasting. However, the most anticipated application of machine learning for biomonitoring is the processing of genetic data. The ultimate aim is for algorithms to classify a new site directly from the bulk of DNA reads just by identifying genetic patterns learned from previous experience.

The same data can be interpreted in various ways if analyzed by different algorithms programmed with different training for different purposes (e.g., detection of eutrophication, effects of toxicants, or changes in flow regime). A set of sophisticated algorithms should enable scientists to monitor the effects of complex combinations of stressors on the environment. Such approaches are needed in view of multiple global threats (Vörösmarty et al. 2010). Furthermore, these methods should be implemented for massive datasets and communicate with holistic and integrative algorithms for automated and autonomous monitoring systems. In contrast to other more established fields in biology (Marx 2013), bioassessment is just beginning to face the problems associated with massive datasets. Scientists will need to begin collaborating more closely with experts in computer science and applied mathematics to benefit from big data, and to develop new ways to communicate results to managers (Panel 2).

### Conclusions

With the development of DNA metabarcoding, traditional environmental monitoring is experiencing a period of transformation, one outcome of which will be the need to deal with unprecedented amounts of data. Ascertaining the technical requirements to obtain and analyze data is just a part of the challenge. In contrast to scientists from other disciplines, ecologists have a relatively poor culture of data sharing, despite opportunities for making big data more accessible (Reichman et al. 2011; Hampton et al. 2013). However, there are signs that this is starting to change. Making biomonitoring big data freely available will potentially allow a range of new applications such as meta-analyses and large-scale analyses of biodiversity. Metabarcoding data are particularly relevant in this case because genetic data are highly comparable. Scientists and resource...

---

**Panel 2. Communication with managers**

Molecular methods constitute a new paradigm in freshwater ecosystem assessment. Environmental managers who are accustomed to traditional biological assessments and who are not familiar with genetics and molecular methods may be initially reluctant to adopt these approaches or may need training in order to do so. The widespread use of metabarcoding in biomonitoring depends on how these new tools will be implemented in future environmental assessment programs. Thus, new ways to communicate with resource managers must be developed. Communication should emphasize the benefits of metabarcoding, as well as explain the basics of genetics and the vocabulary of metabarcoding and HTS to managers in order to empower them to understand, interpret, communicate, and benefit from the results of metabarcoding. However, we must also acknowledge difficulties, such as the challenges associated with machine learning. Although it is important that biomonitoring tools are derived from sound theoretical concepts in ecology, because machine learning often operates as a black box (i.e., the user does not understand how the algorithm works), it might be hard to relate results to environmental health and key stressors. The implementation of such new environmental assessment frameworks will therefore take time and require a close collaboration between scientists and managers. Knowledge and experience gained over many years must not be lost and traditional approaches should continue to be used, at least for the purposes of comparison and discussion.
managers must work together to create effective networks and to develop dedicated sharing platforms. Indeed, the technical solutions discussed in this paper require substantial quantities of data and supporting infrastructures. Sharing platforms should be accessible to citizens and ecologists and would provide both raw and processed data as well as metadata. Raw data can be re-used with new bioinformatic workflows and statistical methods, while processed data are important for non-specialists and to help inform citizens (Soranno et al. 2015). If we can make public – and make sense of – the terabytes of data that ecological assessments will produce in the foreseeable future, the entry of biomonitoring into the Information Age will be a genuine success.

Acknowledgements

We thank A Franc for constructive comments and I Domaizon for insightful discussion on metabarcoding terminology.

References

Baird DJ and Hajibabaei M. 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Mol Ecol 21: 2039–44.

Benson DA, Karsch-Mizrachi I, Lipman DJ, et al. 2008. GenBank. Nucleic Acids Res 36: D25–30.

Berger SA, Krompass D, and Stamatakis A. 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Systems Biol 60: 291–302.

Birk S, Bonne W, Borja A, et al. 2012. Three hundred ways to assess Europe’s surface waters: an almost complete overview of biological methods to implement the Water Framework Directive. Ecol Indic 18: 31–41.

Breiman L. 2001. Statistical modeling: the two cultures. Stat Sci 16: 199–231.

Caron DA, Countway PD, Savai P, et al. 2009. Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. Appl Environ Microb 75: 5797–808.

Chapman T. 2003. Lab automation and robotics: automation on the move. Nature 421: 661–66.

Chariton AA, Stephenson S, Morgan MJ, et al. 2015. Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. Environ Pollut 203: 165–74.

Cosinac E, Riaz T, and Puillandre N. 2012. Bioinformatic challenges for DNA metabarcoding of plants and animals. Mol Ecol 21: 1834–47.

Dafforn KA, Johnston EL, Ferguson A, et al. 2016. Big data opportunities and challenges for assessing multiple stressors across scales in aquatic ecosystems. Mar Freshwater Res 67: 393–413.

De Queiroz K. 2007. Species concepts and species delimitation. Syst Biol 56: 879–86.

Fan W and Bifet A. 2013. Mining big data: current status, and forecast to the future. SIGKDD Explorations 14: 1–5.

Ficetola GF, Miuud C, Pompanon F, and Taberlet P. 2008. Species detection using environmental DNA from water samples. Biol Lett 4: 423–25.

Gibson JF, Shokralla S, Curry C, et al. 2015. Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. PLoS ONE 10: e0138432.

Goldstein FZ and DeSalle R. 2011. Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. Bioessays 33: 135–47.

Guénard G, Legendre P, and Peres-Neto P. 2013. Phylogenetic eigenvector maps: a framework to model and predict species traits. Methods Ecol Evol 4: 1120–31.

Guénard G, von der Ohe PC, de Zwart D, et al. 2011. Using phylogenetic information to predict species tolerances to toxic chemicals. Ecol Appl 21: 3178–90.

Hajibabaei M, Shokralla S, Zhou X, et al. 2011. Environmental metabarcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6: e17497.

Hajibabaei M, Singer GAC, Hebert PDN, and Hickey DA. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet 23: 167–72.

Hampton SE, Strasser CA, Tewksbury JJ, et al. 2013. Big data and the future of ecology. Front Ecol Environ 11: 156–62.

Hastie T, Tibshirani R, and Friedman J. 2009. The elements of statistical learning: data mining, inference, and prediction. 2nd edn. New York, NY: Springer.

Hebert PDN, Cywinska A, Ball SL, and deWaal JR. 2003. Biological identifications through DNA barcodes. P Roy Soc Lond B Bio 270: 313–21.

Ibáñez C, Caiola N, Sharpe P, and Trobajo R. 2010. Ecological indicators to assess the health of river ecosystems. In: Jørgensen SE, Xu F-L, and Costanza R (Eds). Handbook of ecological indicators for assessment of ecosystem health. Boca Raton, FL: CRC Press.

Keck F, Rimet F, A Franc, and Bouchez A. 2016. Phylogenetic signal in diatom ecology: perspectives for aquatic ecosystems monitoring. Ecol Appl 26: 861–72.

Kermarrec L, A Franc, and Marques JC. 2010. Application of indicators for the assessment of ecosystem health. In: Jørgensen SE, Xu F-L, and Costanza R (Eds). Handbook of ecological indicators for assessment of ecosystem health. Boca Raton, FL: CRC Press.

Kolkwitz R and Marsson M. 1909. Ökologie der tierischen Saprobien. Beiträge zur Lehre von der biologischen Saprobien. Ber Deut Bot Ges 26: 861–72.

Lindenmayer DB and Likens GE. 2010. The science and application of ecological monitoring. Biol Conserv 143: 1317–28.

Lovett GM, Burns DA, Driscoll CT, et al. 2007. Who needs environmental monitoring? Front Ecol Environ 5: 253–60.

Mächler E, Deiner K, Steinmann P, and Altermatt F. 2014. Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. Freshwater Sci 33: 349–63.

Mandelik Y, Roll U, and Fleischer A. 2010. Cost-efficiency of biodiversity indicators for Mediterranean ecosystems and the effects of socio-economic factors. J Appl Ecol 47: 1179–88.

Marx V. 2013. Biology: the big challenges of big data. Nature 498: 255–60.

Matsen F, Kodner R, and Armbrust EV. 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11: 538.

Padial JM, Miralles A, la Riva ID, and Vences M. 2010. The integrative future of taxonomy. Front Zool 7: 1–14.

Passy SL, Bode RW, Carlson DM, and Novak MA. 2004. Comparative environmental assessment in the studies of benthic diatom, macroinvertebrate, and fish communities. Int Rev Hydrob 89: 121–38.

Pawlowski J, Lejezerowicz F, Apotheloz-Perret-Gentil L, et al. 2016. Pro taxonomic practice: determination, discovery, and description. Bioessays 33: 135–47.
Ratnasingham S and Hebert PDN. 2007. BOLD: the Barcode of Life Data system (www.barcodinglife.org). Mol Ecol Notes 7: 355–64.
Reichman OJ, Jones MB, and Schildhauer MP. 2011. Challenges and opportunities of open data in ecology. Science 331: 703–05.
Resh VH. 2008. Which group is best? Attributes of different biological assemblages used in freshwater biomonitoring programs. Environ Monit Assess 138: 131–38.
Schloss PD, Westcott SL, Ryabin T, et al. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75: 7537–41.
Schwarzenbach RP, Escher BI, Fenner K, et al. 2006. The challenge of micropollutants in aquatic systems. Science 313: 1072–77.
Shokralla S, Spall JL, Gibson JF, and Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. Mol Ecol 21: 1794–805.
Steele JA, Countway PD, Xia L, et al. 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J 5: 1414–25.
Stein ED, White BP, Mazor RD, et al. 2014a. Does DNA barcoding improve performance of traditional stream bioassessment metrics? Freshwater Sci 33: 302–11.
Stein ED, Martinez MC, Stiles S, et al. 2014b. Is DNA barcoding actually cheaper and faster than traditional morphological methods? Results from a survey of freshwater bioassessment efforts in the United States. PLoS ONE 9: e95525.
Soranno PA, Cheruvillil KS, Elliott KC, and Montgomery GM. 2015. It’s good to share: why environmental scientists’ ethics are out of date. BioScience 65: 69–73.
Taberlet P, Coissac E, Hajibabaei M, and Rieseberg LH. 2012. Environmental DNA, Mol Ecol 21: 1789–93.
van Dijk EL, Auger H, Jaszczyszyn Y, and Thermes C. 2014. Ten years of next-generation sequencing technology. Trends Genet 30: 418–26.
Visco JA, Apothéloz-Perret-Gentil L, Cordonier A, et al. 2015. Environmental monitoring: inferring the diatom index from next-generation sequencing data. Environ Sci Technol 49: 7597–605.
Vörösmarty CJ, McIntyre PB, Gesnner MO, et al. 2010. Global threats to human water security and river biodiversity. Nature 467: 555–61.