From molecules to monitoring: integrating genetic tools into freshwater quality assessments

Beentjes, K.K.

Citation

Beentjes, K. K. (2021, April 8). *From molecules to monitoring: integrating genetic tools into freshwater quality assessments*. Retrieved from https://hdl.handle.net/1887/3158798

Version: Publisher's Version

License: Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden

Downloaded from: https://hdl.handle.net/1887/3158798

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle https://hdl.handle.net/1887/3158798 holds various files of this Leiden University dissertation.

Author: Beentjes, K.K.
Title: From molecules to monitoring: integrating genetic tools into freshwater quality assessments
Issue Date: 2021-04-08

CHAPTER 6

General discussion and synthesis

Kevin K. Beentjes

KEY FINDINGS

1. For relatively simple water quality assessments such as the ecological quality ratio scoring, abundances in taxon data are of limited influence.
2. Taxonomic sorting prior to DNA analysis reduces the impact of preferential amplification, as data from complex sample mixtures with uneven biomass distributions between various taxa are often dominated by reads belonging to a single taxon. This shows that, even with universal primers, the effects of primer bias are still significant.
3. Differences in community composition caused by small-scale temporal turnover are equal to or larger than those caused by heterogeneity. Sampling replicates over time are more important for insight into the total diversity than spatial replicate sampling.
4. Multi-marker eDNA impact assessments across trophic levels prove to be a more comprehensive indicator of impacts on the food web and provide more information on a higher taxonomic resolution, whilst uncovering similar impact patterns as more cumbersome morphological surveys.

While there is an ever-increasing number of publications on the possibilities and limits of environmental DNA and DNA metabarcoding in biodiversity monitoring, it has become clear that these new techniques will most likely never truly conform to the needs of the traditional monitoring schemes. However, recent insights and developments have shown that there is merit in molecular biomonitoring. Developing a better understanding of the ecology of eDNA, as well as getting a grip on the effects of different choices in the field, lab, and analysis is paramount to making molecular tools successful. Implementation of DNA-based techniques such as eDNA sampling and metabarcoding can never work when the mechanisms behind the techniques are not understood properly, and there are still knowledge gaps in both methods and applications of eDNA (Garlapati et al. 2019). Closing these gaps is important, especially when dealing with regulations such as the EU-WFD. While there will always be some differences between studies that are introduced by sampling design, laboratory protocols, and analysis pipelines, understanding the causes of these differences will lead to better documentation of protocols and help in the intercalibration of studies.

6.1 SAMPLING STRATEGIES

The ecology of eDNA as discussed in Chapter 1 has its implication on practice as well, mainly on the sampling strategy deployed in the field, but also on downstream processes. There are numerous papers that describe other critical considerations during the entire pipeline from field sampling and sample processing to analysis and reporting of results (Goldberg et al. 2016, Dickie et al. 2018, Harper et al. 2019a, Zinger et al. 2019, Nicholson et al. 2020). While eDNA sampling seems relatively straightforward, especially compared to kicknet sampling, the reality is slightly more complicated. Environmental DNA concentrations are usually low, especially where it concerns rare species or alien species in early stages of invasion (Tréguier et al. 2014). One of the general conclusions that can be drawn from the eDNA literature is that larger volumes of water are preferable, although increased volumes do not always lead to better detection rates (Mächler et al. 2016). The need for spatial replicates is evident, but the replicate sampling strategy needs to be adapted to the specific questions and target organisms, much like traditional monitoring efforts. This not only concerns the number or replicates, but also the distance between them and the structure of sampled habitats (Lugg et al. 2018, Grey et al. 2018). Sampling sediment can be an alternative to water sampling, since a large proportion of the eDNA will settle to the sediment. Concentrations of eDNA in sediment can be higher than concentrations in the water column (Turner et al. 2015), but can represent a more historical signal.

Once eDNA water samples have been collected, they should be filtered in the field or preserved until they can be processed in the laboratory. Various methods to preserve water have been published, but the efficacy of each of those methods is often only poorly tested, and in many publications not compared to other standard methods. Directly filtering water in the field has become more standard (Pilliod et al. 2013, Turner et al. 2014b, Hinlo et al. 2017, Grey et al. 2018), especially since there have been various papers that describe methods that allow for better preservation of filters in buffer (Renshaw et al. 2015), and prevention of cross-contamination between samples due to filter handling in the field by using enclosed filters (Spens et al. 2017, Thomas et al. 2019). Filtration methods have been examined in great detail, and many different filter types, pore sizes, and filter processing techniques have been compared in a broad range of studies using both single-species detection (Eichmiller et al. 2016) and metabarcoding for a variety of target organisms (Djurhuus et al. 2017, Li et al. 2018a, Majaneva et al. 2018, Jeunen et al. 2019). Generally, larger pore sizes seem to be optimal regarding the balance between volume of water filtered and DNA yielded from the filters.

6.2 REGARDING REPLICATES

As shown in Chapter 4, replicate sampling strategies are important to capture the full diversity of organisms in an aquatic environment. Especially in lentic systems, where dispersal of eDNA appears limited, spatial replicates are necessary to pick up these oft local signals (Thomsen et al. 2012b, Evans et al. 2017b, Grey et al. 2018, Lawson Handley et al. 2019). The data from Chapter 4, where eDNA samples were collected weekly during 20 consecutive weeks, in a total of six locations in two dune lakes, shows the heterogeneous distribution of environmental DNA signals in space and time. Looking at the three spatial replicate samples taken in each lake, the majority of observations (57.8%) were limited to MOTUs only found in single replicate sample.

One of the main reasons for the research presented in Chapter 4 was to delve into the temporal patterns of environmental DNA. While there is often specific focus on the inclusion of spatial replicates during sampling in the field, studies looking at seasonal variation are usually limited to larger temporal scales (e.g winter versus summer) (Chain et al. 2016, Guardiola et al. 2016). Various papers examining single species or a select number of taxa have already shown that detection rates for these taxa can vary throughout the year (Stoeckle et al. 2017, Buxton et al. 2017). Few studies, however, have been performed on small-scale temporal differences in water bodies, looking at intervals of months (Bista et al. 2017, Rees et al. 2017) or even weeks (Sigsgaard et al. 2017). Our data shows that the effects of turnover on a fine scale (weekly sampling) contributes as much to the overall observed diversity as community heterogeneity (as observed with spatial replicate sampling). Dissimilarity is increased over larger intervals, where it contributes more to the total observed diversity than spatial replicates for intervals of more than two weeks. The dissimilarity between taxonomic composition of the two lakes studied in this study showed a linear increase with increased intervals (Figure 4.4). This suggests that any study comparing diversity across different sites is susceptible to inflated dissimilarities when study sites are sampled mere weeks apart. Moreover, weekly samples during the 20-week period on average only contained 20.7% of the total observed MOTU diversity (Figure 6.1A), with an average 6.5% of those observations limited to single time point (Figure 6.1B). High seasonal diversity has since also been described for metabarcoding of bulk-collected samples in freshwater streams (Zizka et al. 2020), furthermore stressing the importance of sampling moment in comparative studies. However, that study also found that for the relatively simple water quality assessment scores, seasonal variation, much like abundance data, is of limited influence.

In addition to biological replicates in the field, technical replicates during

extraction (Lanzén et al. 2017) and especially PCR are often highlighted in literature focusing on sampling methodology. PCR replicates are often common practice in qPCR detection studies looking at single species (e.g. Biggs et al. 2015, Agersnap et al. 2017, Harper et al. 2018), but are starting to be used in metabarcoding studies as well (Civade et al. 2016, Alberdi et al. 2018). Especially when working with low quantities of eDNA, the heterogeneity of the DNA extract can cause issues. While the dissimilarities between PCR replicates were not as large as the dissimilarities between spatial and temporal replicates (Figure 4.3), including these replicates does increase

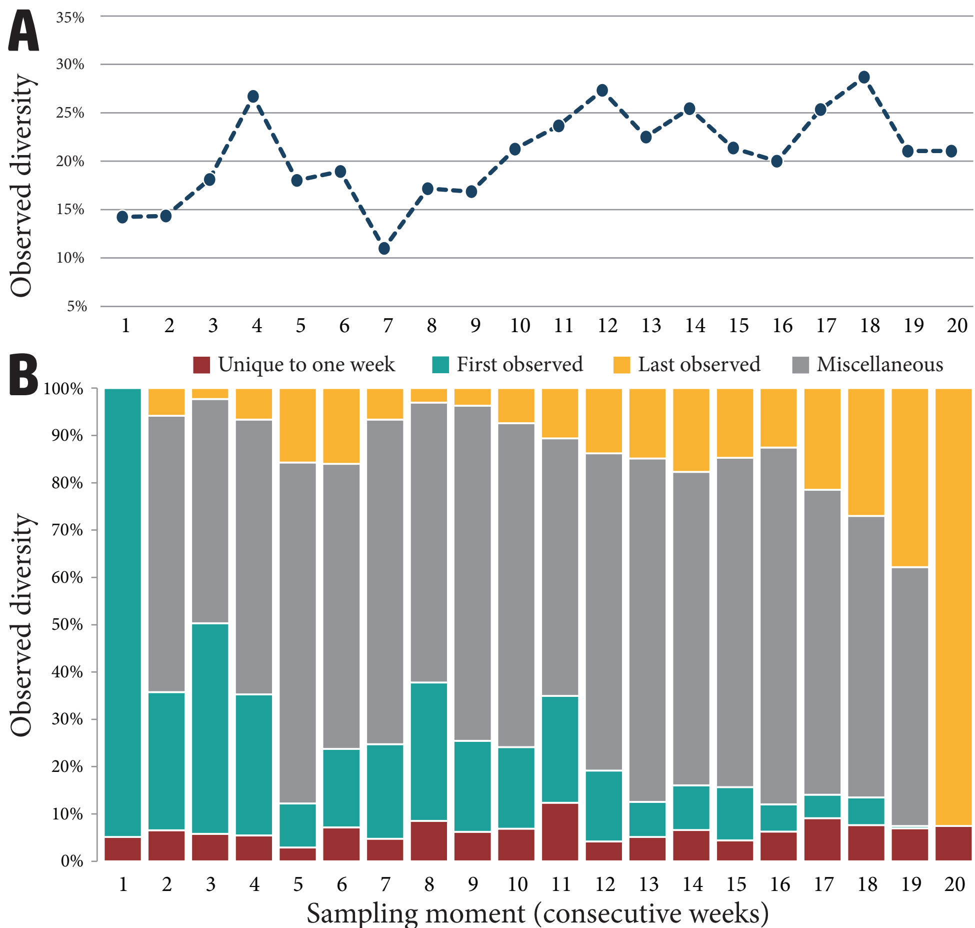

FIGURE 6.1. The distribution of MOTUs in data from Chapter 4, with (A) the number of observed MOTUs expressed as a fraction of the number of MOTUs found in total across the 20-week sampling period, and (B) the distribution of MOTUs observed each week based on when each MOTU was observed across the 20-week sampling period. These are either unique observations for a single time point, and MOTUs that were either not observed before or after each time point.

the observed diversity, as many rare species are often found only in a single PCR replicate. Data from Chapter 4 shows that 40.2% of all observed MOTUs have only been found in a single PCR replicate, whereas only 27.4% of MOTUs were observed in all three PCR replicates. Including PCR replicates also decreases the uncertainty in detection rates (Mächler et al. 2016). To circumvent the need for qPCR replicates in order to get accurate concentration measures in single species detection studies, droplet digital PCR (ddPCR) has been employed with promising results (Doi et al. 2015, Uthicke et al. 2018).

6.3 ADDRESSING ABUNDANCE

Interpretation of molecular data requires some insight into the behavior of DNA, especially where it concerns environmental DNA. For example, caution has to be taken when interpreting DNA-based data in terms of specimen abundances. Several papers that use single-species detection show good correlations between biomass of target organisms and DNA concentrations from standardized eDNA samples (Pilliod et al. 2013, Klymus et al. 2015, Uthicke et al. 2018, Spear et al. 2020). To translate such concentration measures back to actual present biomass or specimen counts remains difficult, and would require extensive calibration efforts, or extensive sampling regimes (e.g. Levi et al. 2019). DNA concentrations do, however, allow for the inference of differences in specimen abundance between different locations, providing useful information for water management (Lacoursière-Roussel et al. 2015).

For metabarcoding, however, it is more complicated. The preferential amplification of certain taxa over others (PCR bias), can cause difficulties in the interpretation of molecular data and the comparison to morphological assays. A study by Elbrecht and Leese (2015) showed that sequence abundance varied by up to four orders of magnitude between species sequenced from bulk samples, even though the input biomass was comparable for all species. Preferential amplification lies at the basis of most difficulties surrounding the inference of abundances from molecular data, but additionally also impacts the detection of rarer species in complex mixtures with uneven biomass distributions. This disbalance in biomass causes that large proportions of the DNA extract come from only few large specimens, and it often happens that rare species are lost in the background, especially when sequencing depth is not sufficient. In Chapter 3 we looked into the effects of taxonomic sorting on the detection of taxa. Samples from WFD monitoring sites were homogenized using a blender, but with a prior sorting into six taxonomic groups as provided by the monitoring agency. Pooling of DNA extracts prior to PCR and sequencing

was performed to simulate a situation where all organic material was homogenized without any sorting. The idea behind the analysis of sorted samples stemmed from previous results obtained during pilot studies, where samples were often dominated by a single species or taxon. The sorting resulted in much higher estimated diversities, both in terms of MOTUs and formal taxa (Figure 3.1), with 46.5% increase in taxon recovery. As expected, the pooled samples were often dominated by certain taxa, be it one of the six groups (15 out of 24 samples), or even a single taxon (eleven samples) (Supplementary Figure 3.3). Sorting bulk samples into size fractions prior to homogenization, to prevent large specimens from dominating the DNA data, provided similar results to our findings, with significant increases in taxon recovery (Elbrecht et al. 2017b, 2020). Such methods may be more feasible in practice.

The use of eDNA metabarcoding is even more prone to complications, because there often is no original biomass to compare read data with, and research has shown that eDNA often has a very local and heterogeneous distribution in lentic waters. Some studies with relatively few taxa in aquarium setups have shown modest relations between biomass and read abundances (Evans et al. 2016), and read abundances can still be informative in comparative studies, such as those that evaluate fish community assemblages along a river (Pont et al. 2018). Preferential amplification certainly affects eDNA studies too, in some cases even preferentially amplifying taxa that are not the intended target, as witnessed by the many non-macrofaunal taxa amplified using the primers in Chapter 4, which were developed as universal macrofauna primers. Environmental DNA studies seem hampered more than helped by such universal primers, as they tend to pick up a lot of unintended "bycatch".

Fortunately for many water quality indices, abundance has proven to be of limited importance. As shown in Chapter 2, there is a strong correlation between ecological quality scores calculated with and without specimen abundances in the Dutch WFD system. Similar results have been obtained for quality scoring for freshwater systems in New Zealand (Wright-Stow & Winterbourn 2003) and Germany (Buchner et al. 2019), as well as for the AZTI's Marine Biotic Index, a marine scoring system used by many European countries (Aylagas et al. 2014). For other assessments of biodiversity, however, abundances are often an important parameter, and using presence/absence data can overestimate the importance of rare species (Deagle et al. 2019). Alternatives like shotgun sequencing seem to provide decent correlations between biomass and read abundance (Bista et al. 2018). They do introduce other difficulties though, since usually only the mitochondrial DNA is usable for matching to reference sequences, which only amounts to at most 0.5-1.0% of the read data. Calibration studies evaluating PCR bias in NGS via qPCR (Pawluczyk et al. 2015) are cumbersome for

studies with a potentially large number of taxa (most macroinvertebrate assessments) and require *a priori* knowledge of the species composition of a sample. Using internal standards with known concentrations in metabarcoding might at least allow for the different samples to be more comparable to each other regardless of differences in sequencing depth (Ushio et al. 2018).

6.4 MOTUs VERSUS TAXA

The other main challenge of DNA-based data, besides the difficulties in interpreting read abundances, lies with the inferences of taxonomic composition of samples. Sequence data is usually clustered before analysis and comparison to references, but there are several different clustering approaches that all produce their own MOTUs (molecular taxonomic operational unit) (Clare et al. 2016). The most common methods use clustering based on dissimilarity percentage thresholds between reads, others take abundances of exact sequence variants into account (Edgar 2016, Rognes et al. 2016). MOTU clustering can introduce either overestimation of diversity by creating separate clusters for taxa with high intraspecific variation, or underestimation due to lumping taxa with low interspecific variation, and finding a balance between these can be difficult (Alberdi et al. 2018).

Chapter 3 illustrates this potential overestimation of richness, where the number of MOTUs found in each sample was approximately five-fold the number of taxa (211.4 versus 40.8 on average), and the correlation between the numbers of MOTUs and the numbers of morphological taxa was weak ($r = 0.365$) and above all not significant ($p = 0.072$) (Figure 3.1). The difference, however, was exacerbated by the fact that many specimens in the morphological analysis had not been identified up to species level, and higher-level taxonomic observations are likely to represent clusters of multiple taxa. We also found various MOTUs that had the same taxonomic assignments, suggesting cryptic species or haplotype diversity is present in many taxa, which remain undetected during morphological analysis (Gibson et al. 2015, Elbrecht et al. 2018a). When specifically looking at the number of taxa observed with morphology and molecular analysis, the correlations were stronger ($r = 0.662$), although in this case there were still issues with differences in depth of taxonomic assignment for various groups. For example, various dipterans were not identified to species level with morphology, but were with DNA metabarcoding, whereas the COI fragment was unable to differentiate between morphologically different species of leeches.

Even for those taxonomic groups that are well-covered in the (public) DNA reference libraries, reference specimens usually do not cover the entire geographic

range in which they occur. These references may thus not represent the full genetic diversity of species, especially over larger geographical ranges (Bergsten et al. 2012, Baselga et al. 2013). The limited dispersal of many aquatic taxa has led to highly structured populations in insular freshwater systems (Strayer 2006), where even species that are assumed to be common and widespread show high levels of cryptic diversity (Alp et al. 2012, Sworobowicz et al. 2015). Metabarcoding studies can expose this intraspecific genetic variation (Elbrecht et al. 2018a, Sigsgaard et al. 2020, Chapter 3), but may also lead to overestimations of diversity when using MOTUs that cannot be linked to reference databases (Brown et al. 2015). Attempts to identify such MOTUs at a higher taxonomic level may introduce a lot of noise into a dataset (Berney et al. 2004). In an ideal situation all the possible species are represented in the DNA reference libraries, and all sequencing reads can be matched directly to one of the potential species. This might work to a certain extent for metabarcoding of relatively simple bulk samples, but complex samples or environmental DNA samples will always contain sequences that cannot be linked directly to a reference, especially when using universal primers or when analyzing microorganisms (e.g Chapters 4 and 5). Even for commonly studied taxon groups such as the freshwater macroinvertebrates, or marine macrobenthos, DNA reference libraries are far from complete (Wangensteen et al. 2018, Weigand et al. 2019) (see also Figure 1.3). Especially uncommon species, which are also often missed by morphological surveys (Jackson et al. 2014), might be absent from the reference libraries.

In addition to undescribed genetic diversity, many organisms carry pseudogenes that potentially also introduce overestimations of species richness in metabarcoding studies (Song et al. 2008, Buhay 2009). These nuclear-mitochondrial pseudogenes (NUMTs), are co-amplified with the target region. This effect becomes more profound for those species that are more abundant, as over-amplification tends to bring this signal, which often resides in the background, to light. For example, in the dataset for Chapter 4 we were able to identify a total of 21 putative pseudogene sequences in the muntjac control samples. After filtering the data, these samples contained 22 MOTUs, which were all identified by the LCA as either *Muntiacus* or a member of the subfamily Cervinae. One of these MOTUs represented 96.0% of all read data from the control samples and resulted in a 100% match with Sanger sequenced reference sequences of the *Muntiacus reevesi* sample. The second largest MOTU, which represented less than 0.5% of the control sample read data, only showed a 93% match with *M. reevesi*. As no other species were expected to be present in these control samples (DNA was extracted from *M. reevesi* blood samples), we postulate that the 21 additional MOTUs found in the control samples are pseudogenes. Read

errors seem unlikely, as most of the 21 MOTUs are present in all twelve independent PCR replicates, suggesting the signal was intrinsic to the control DNA sample. Four out of 21 MOTUs had indels causing frame shifts, twelve more had stop codons in their reading frame, meaning that there were five suspected pseudogenes that were not identifiable as pseudogenes based on their sequence alone. It is also difficult to define putative pseudogenes based on the likeness to highly-abundant actual biological signals, since they often differ substantially from their original sequence (93.2-82.6% pair-wise identity in case of the 21 muntjac sample MOTUs). This means that many DNA metabarcoding datasets are likely to have unrecognized pseudogenes present, an effect that is furthermore amplified in datasets with larger sequence data outputs. This is also supported by findings in Chapter 3, where 14 MOTUs assigned to *Asellus aquaticus* had significantly more reads that the 109 MOTUs assigned only to the genus *Asellus* via LCA (75,128 versus 1,768 reads on average), suggesting that at least some of the latter may have been pseudogenes. Large-scale DNA barcoding programs such as the Global Malaise Trap Program (Geiger et al. 2016) or BIOSCAN (Hobern 2020) could resolve some of these issues by creating large datasets with better geographical coverage directly linked to actual specimens, as it will help us understand which part of the observed diversity in metabarcoding studies translates to actual biological diversity.

On the other hand, when analyzing patterns in richness or diversity, the effects of these "fake" MOTUs are likely limited, similar to rare species. As seen in the impact assessment in Chapter 5, the ecotoxicological effects on MOTU richness were not any different from the effects on morphological richness estimates in the exact same study site. One could also argue that overestimations of richness are comparable within one study (i.e. the effect is the same for all samples). Response patterns based on this potentially overestimated diversity were also similar to those observed in morphological assessments (Figure 5.4), even though the molecular assessment used fewer replicates. In any case, it is difficult to correct for overestimations, as MOTUs can often not be distinguished as artefacts. However, with all the undescribed genetic diversity in many (common) taxa, "lumping" all these MOTUs leads to a loss in potentially interesting information. Much like specimen-based assessments take advantage of DNA-based identifications and may even lead to new species-specific insights on ecology (Jackson et al. 2014), metabarcoding studies will uncover genetic diversity that goes beyond currently recognized species. Studies have already shown that different clades, haplotypes or cryptic species can have different and even contrasting responses to environmental stressors, and many unidentified MOTUs can still provide informative response patterns to stressors (Macher et al. 2016,

Beermann et al. 2018, 2020, Chapter 5). These studies indicate that even with an incomplete reference library or unresolved cryptic species complexes, DNA barcodes provide higher-resolution taxonomic information that can be used for assessments. Furthermore, the "taxonomy-free" studies that have been performed with diatoms already show that MOTU-based data sets can provide ecological status assessments similar to traditional surveys. Chapter 5 illustrates that eDNA metabarcoding data showed the same impact patterns in response to agricultural stressors, on different trophic levels. Such multi-marker impact assessments also provide more information on a higher taxonomic resolution, even if not all of the MOTUs can be assigned to taxa at this point in time. This corroborates the findings in marine aquaculture impact studies, which showed that different markers could accurately predict stressor impact and even outperform the more cumbersome traditional methods (Cordier et al. 2019).

6.5 FUTURE PERSPECTIVES

Despite the many ongoing discussions around the technical considerations when dealing with metabarcoding data, there seems to be some consensus on its usefulness to provide information on species occurrences and changes thereof. The number of new papers on the topic is rapidly growing (Figure 1.2), and many researchers have shifted the focus of their work towards the implementation of molecular tools in actual monitoring. There is need for a solid foundation of scientific research that directly compares traditional monitoring with new methodologies, as this is the work where similarities and differences between the "old" and the "new" come to light. Some traditional monitoring systems are more easily supplanted by DNA-based techniques, such as impact assessments or the detection of invasive species. Others are currently too heavily adapted to traditional monitoring and information that is hard or impossible to obtain from molecular data, such as the age and/or size distributions of fish, but many molecular techniques will at least provide complementary data to obtain a more complete insight into the ecosystem.

The main issue with traditional monitoring, especially the monitoring as prescribed by the WFD (and its national interpretations), is that it is set up to deal with the shortcomings of the traditional techniques. This is why WFD monitoring mainly focusses on groups that are identifiable by light microscopy and relatively easy to collect. We must, however, not try to make new techniques compatible with imperfect existing systems. The potential of molecular techniques has been proven in many scientific papers, and leaves ample room to develop new monitoring schemes

that fully harness the power of these DNA-based tools. Better insights into the genetic diversity of many species is direly needed, as this will allow for the improvement of diversity measures based on molecular data. It will allow us to find a balance between the potential overestimation caused by artefactual sequence data and the underestimation caused by lumping potentially informative cryptic taxa into a single entity. This is especially relevant since the trend in DNA metabarcoding studies moves towards more and more stringent clustering methods.

That is not where the work stops, however. With more insight into the genetic diversity of taxa, we will also be able to do a more in-depth exploration of the ecological meaning of such genetic variants. As shown in a few studies, these cryptic species or (sub)populations related to genetic variations can have very different responses to environmental stressors, which has a considerable impact on stressors assessments. This would be a lot of work, and we seem to have barely scratched the surface. The genetic diversity, and its ecological diversity, also showcase the continued importance of taxonomists, which are sorely needed not only to fill and quality-check the ever-growing DNA references libraries, but also to find what this genetic diversity uncovered by next-generation sequencing means in the field. The technological advances in environmental DNA and metabarcoding studies are meaningless without taxonomic and ecological knowledge to translate sequences into an understanding of the ecosystem. Multi-trophic analyses of communities show there are cascading effects in food webs, and they not only provide information on the composition of an ecosystem, but also its interconnectedness and, more importantly, its functioning.

Future developments in ecological assessments will have to focus more on ecosystems truly as systems, rather than just a collection of taxa. Additionally, one of the most important challenges for scientists in the coming years is to also translate the findings from these new DNA-based monitoring methods into useful information for monitoring agencies and policy makers. They are the ones that need to be convinced of the merits of molecular monitoring at this point, as it seems that the scientific community has all but embraced the techniques.