

JULY 2016

Options for a New Integrated Natural Resource Monitoring Framework for Wales

Project Document

Briefing note: The Potential for Molecular Genetic Identification of Biodiversity across the Welsh Biosphere

Welsh Government Contract No. C147/2010/2011

Agreed Additional Work Requirement Dated 8th March 2016

NERC CEH Project: NEC05945

How to cite this document: Creer, S., Griffiths, R., Hatton-Ellis, T.W., Jones, D.L. (2016) **Options for a New Integrated Natural Resource Monitoring Framework for Wales; Phase 1, Project Document - Briefing note: The Potential for Molecular Genetic Identification of Biodiversity across the Welsh Biosphere;** Report to Welsh Government (Contract reference: C147/2010/11; Agreed Additional Work Requirement Dated 8th March 2016). NERC/Centre for Ecology & Hydrology (NERC CEH Project: NEC05945)

Options for a New Integrated Natural Resources Monitoring Framework for Wales

Project Document - Briefing note:

The Potential for Molecular Genetic Identification of Biodiversity across the Welsh Biosphere

S. Creer (College of Natural Sciences, Bangor University)
R. Griffiths (CEH)
T.W. Hatton-Ellis (NRW)
D.L. Jones (College of Natural Sciences, Bangor University)

July 2016

Intentionally blank

Briefing note:

The Potential for Molecular Genetic Identification of Biodiversity across the Welsh Biosphere

Introduction

In order to monitor and evaluate the biological condition of our nation's natural resources and determine how they are affected by environmental and management change, there is a pressing need to assess the composition and diversity of organisms across the breadth of life in both space and time (e.g. bacteria, fungi, invertebrates, fish etc). Traditionally, this national-scale monitoring has been operationally limited by the difficulties in identifying and counting different taxa, both of which incur significant resource constraints (i.e. manpower, cost). For many taxonomic groups, the skills base to effectively and consistently monitor a diverse range of organisms may be inadequate or even completely lacking. Advances in molecular biology now provide alternative new approaches that can revolutionise how biodiversity is monitored in a comprehensive way across the whole of the Welsh landscape.

The molecular genetic toolbox

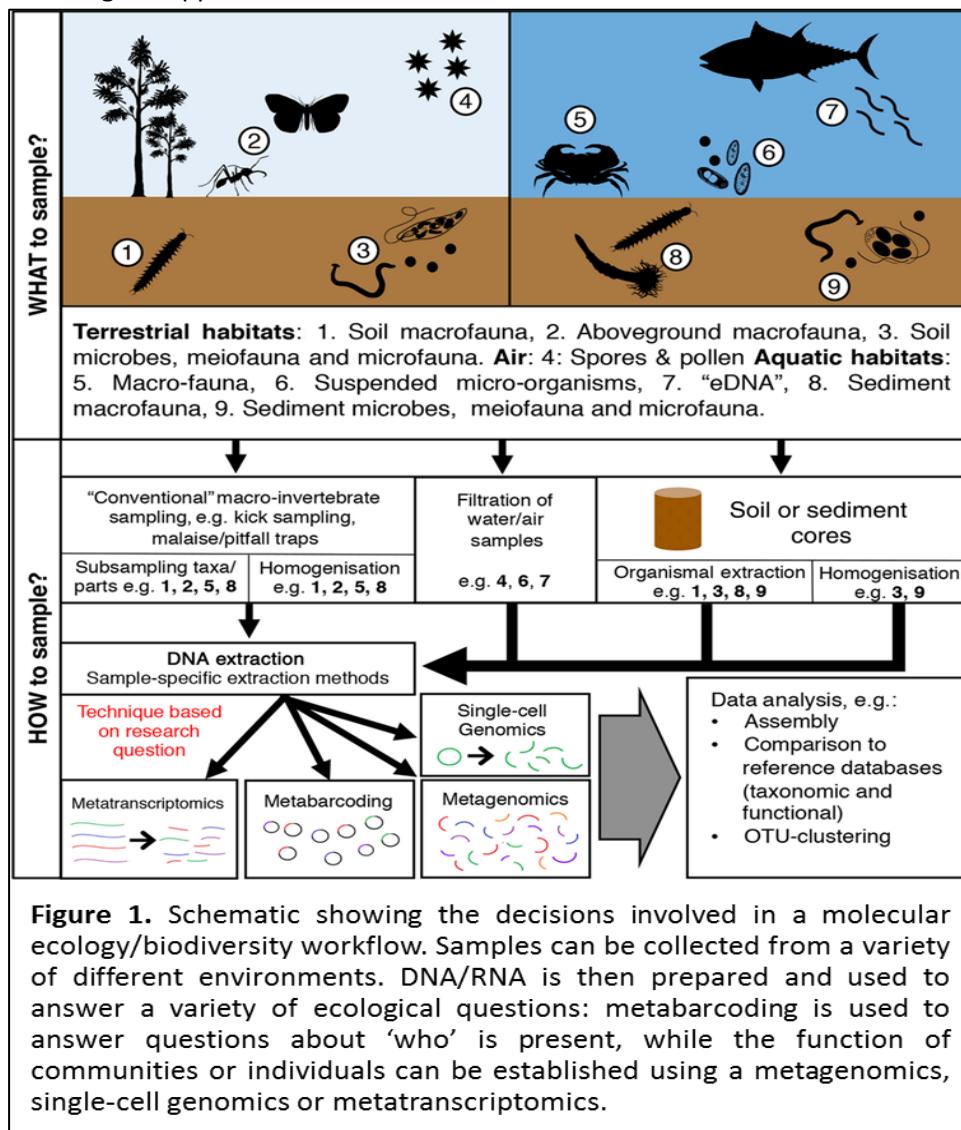
For many years, out of necessity, researchers in the field of microbiology have been using molecular approaches to assess the biodiversity of communities using genetic approaches. However, the relatively high cost of such work has tended to restrict its use to the research community or to more specialist applications. Recent developments in sequencing technologies have greatly increased the accessibility and hence attractiveness of this technology, including its use in assessing the biodiversity of larger taxa.

By focusing on a range of genetic source material (e.g. community-level or environmental DNA [eDNA]), habitats, and spatial scales, we can now characterise entire communities more easily and cheaply across a wide range of taxonomic groups. The purpose of this paper therefore is to provide a succinct summary of the different molecular approaches suitable for the assessment of biodiversity and showcase the ecological research opportunities afforded by contemporary DNA sequencing. The text is derived primarily from Creer et al. 2016. An ecologist's guide to sequence based identification of biodiversity available Online Open from <http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12574/abstract> and augmented with relevant case studies throughout.

Genomic, community, or environmental DNA?

For the field ecologist, we can define many forms of DNA. Genomic DNA is extracted from a single individual (or from a collection of individuals belonging to the same species). Community DNA consists of genomic fragments from many individuals representing a mix of different species. Community DNA is isolated from organisms in bulk samples, but separated from their habitat (e.g. soil, sediment, river benthos). Community DNA extracts have important potential in ecological studies, especially for biomonitoring purposes, since the focus is on the extant community. Environmental DNA (eDNA) (Figure 1) is isolated directly from an

environmental sample without first isolating any type of organism (e.g. soil, sediment, faeces, water, air, etc.). One of the most powerful aspects of eDNA analysis is the ability to sample biodiversity that is not easily sampled by other means or requires complicated procedures to extract organisms of interest (e.g Tullgren funnel extraction of soil fauna, or filtering organisms from aqueous material). The combination of genomic, community and environmental DNA therefore provide a variety of sources of biodiversity information that can be analysed using the approaches here on.



Current and potential applications

Researchers have used eDNA methods for fundamental research into the diversity of life and its function in a variety of habitats as well as to answer ecological questions relating to environmental or management change. More recently, the methodologies have been used in larger scale survey and monitoring to establish broader drivers of microbial diversity (e.g GB Countryside Survey 2007, see ukso.org; Glastir Monitoring and Evaluation Programme - GMEP). For larger organisms, contemporary eDNA analyses have already been extensively implemented for detecting invasive species in aquatic environments using species-specific markers and more recently for reliable detection of fish and/or amphibian communities. In rivers, eDNA can even represent information that is integrated over large spatial areas due to the transport of DNA downstream and is an area (in addition to marine ecosystems) currently benefiting from investment from

NERC Highlight Topic Funding (<http://mefgl.bangor.ac.uk/news/can-we-use-edna-as-an-environmental-magnifying-glass-24870>). Marine sediments have provided eDNA and community DNA for analysing the pollution impact on biodiversity. It is also possible to collect plant eDNA from the air, from faeces, or from pollinators (e.g. honey bees). Ancient DNA from locations such as lake beds or permafrost offers a window into past communities.

One important advantage of eDNA approaches, is that DNA can be stored in small volumes and archived for future use. For instance, at CEH a DNA archive is available for over 1000 soil samples collected across Britain in 2007 and a further 750 samples collected across Wales in 2012-2016. Whilst this was initially used for a microbial survey, the development of new markers means that the samples can now be probed for a variety of other taxa. Coupled with long term and large scale monitoring, these technologies potentially allow for investigations into the spread of invasive or pathogenic taxa over time (e.g. insect vectors of disease; livestock pathogens; microbial human and plant pathogens; non-native plants etc., see Case Study 1).

Case study 1: Soil biomonitoring

Soils are one of the most biodiverse habitats, and traditional methods for reliable sampling and taxonomic characterisation are under-representative. Most studies to date have focussed on microbial communities since they represent the bulk of the soil diversity and biomass, as well as playing key roles in important processes such as carbon storage, nutrient cycling and regulating greenhouse gas emissions. Much of this diversity cannot be assessed using traditional culturing and so prior to the implementation of molecular methods our knowledge of the true extent of soil diversity was limited; and our understanding of biodiversity distribution and ecological drivers of spatial patterns was almost non-existent.

The application of molecular approaches to large-scale soil surveys, has revealed much new information on the broad drivers of bacterial biodiversity. For instance, as part of the GB scale “Countryside Survey” CEH provided a molecular assessment of the bacterial communities across England, Wales and Scotland and revealed strong relationships with the same geological and climatic features that determine the distributions of plant communities. Importantly, this revealed that at the broad level, we can make certain predictions as to the type of bacterial communities found in different climatic and geological settings; and also infer likely effects of land management based on direct effects on soil edaphic conditions. Subsequent research further confirmed this by producing detailed predictive maps of bacterial distributions (see the UK soils portal: ukso.org), utilising the modelled relationships between bacterial biodiversity and habitat type obtained from the remote sensed UK Land Cover Map and existing geological maps.

A key challenge is how to implement soil eDNA approaches for a wider variety of taxa and to use the information to inform on ecosystem services. For instance, using the same DNA resources from the Countryside Survey researchers have used a specific qPCR assay to report on the distribution of *Mycobacterium avium* ssp. *Paratuberculosis*, a soil borne animal pathogen. Such quantitative assays could be developed for other taxa such as other human or plant pathogens, and then applied to DNA resources from large-scale surveys. Equally, the use of high throughput sequencing assessing the diversity of broad specificity marker genes (as now implemented e.g in GMEP) may also provide relative abundances on specific taxa of interest. There is currently much research on the use HTS approaches to quantify the diversity of soil mesofauna, particular members of which are considered soil “ecosystem engineers”; and wider taxa also play a large role in soil decomposition processes. Traditional methods for the enumeration of soil mesofauna involve complex and biased specimen extraction, as well as labour intensive taxonomic characterisation and so there is considerable hope that eDNA methods may overcome these issues. Potential barriers to implementation include: choice of marker gene to provide reliable taxonomic ID for a wide range of mesofauna; poor molecular records of known species in nucleotide databases; and sampling issues with respect to adequate representative coverage (DNA is often extracted from <0.5g of soil in large scale surveys, meaning potential “catch” may be limited). Nevertheless, these issues are likely to be overcome in the near future, given the considerable speed of progress and research effort in this area.

What are the advantages and disadvantages of molecular genetic approaches for national scale monitoring in Wales?

Most monitoring essentially boils down to five general questions:

1. What species is / are present?
2. Where are they?
3. When were they recorded?
4. How much / many of them were there?
5. What does this tell us about environmental quality?

The first three questions are more or less essential for monitoring to have any real usefulness. The fourth question is useful in most circumstances though it can be challenging to collect in many circumstances, resulting only in presence / absence data. The last question is the most important of all as it connects the data to environmental management and policy questions. In many cases specific monitoring tools exist that integrate all five questions (see Case Study 3). In this section we compare molecular methods in general terms with current methods against the criteria above.

What species are present?

Although most biological recording is carried out at the species level, a significant amount of recording also takes place at higher taxonomic levels such as the genus or even family. As discussed, molecular methods are generally predicted to be more effective at species detection than conventional methods. They are capable of correctly detecting species at lower abundances than is normally possible; detecting a wider range of taxa than conventional methods from a single sample, and have the potential to identify taxa that cannot be identified at all using existing methods (e.g. different life history stages and difficult to identify species). There are also limitations in the genetic databases used to identify environmentally retrieved sequences, since the majority of global species have yet to be sequenced. As a result, not all molecular sequences can as yet be assigned a taxon name, but will instead be assigned a taxonomy according to the most closely related taxon in the reference database. Assigning identities to sequences derived from community/ eDNA is implicit, and therefore, a unified stance on building specific DNA reference data bases is of utmost importance. Augmenting the existing databases with the necessary records can be achieved at low cost per species (e.g. £10-£15 per species). A significant advance in Wales has been the collation of plant barcodes for the majority of Welsh and UK flowering plants, covering 1,479 UK native flowering plant species <http://www.gardenofwales.org.uk/science/barcode-wales/> – an invaluable resource for the future of botanical, pollinator and allergenic health research in Wales, that is already drawing in substantial RCUK funding (<http://mefgl.bangor.ac.uk/news/new-1-2m-nerc-grant-aims-to-revolutionise-pollen-forecasting-24704>).

The greater detection power of molecular methods (especially eDNA) has significant potential for species monitoring, especially at low abundances or in environments that are difficult to observe / sample cost-effectively using other methods. Species detection records generated through the analysis of eDNA can then be used to target other forms of survey and management actions (see Case Study 2). Examples of relevant policy applications include:

- Detection of rare and priority biodiversity (e.g. Section 42 species) in order to focus management action, planning decisions or further survey;
- Detection of invasive species in order to facilitate eradication at an early stage, before the species becomes established (of interest to a range of stakeholders and Dŵr Cymru, Welsh Water);
- Broad scale monitoring of biodiversity patterns in poorly sampled environments such as soils and marine ecosystems
- Understanding the relationship between environmental stressors and biodiversity indicator species

Case Study 2: eDNA as a tool for detecting Great Crested Newt

Great Crested Newt is a globally threatened species that is strictly protected by UK and European Law, but is locally quite common in parts of England and Wales. Adult newts enter the water in spring to breed and remain until early summer when they return to land. The larvae may be present in the pond at any time of year but are difficult to detect using conventional surveys. Traditional surveys use a combination of trapping and searching by torchlight when the newts are active, but this is a relatively labour-intensive process and can only be carried out at certain times of year. In addition, a relatively high rate of false negatives means that several surveys are required before newts can be declared to be absent.

These constraints are a problem for developers in areas where Great Crested Newts are present, because they can cause substantial delays and additional costs to projects. By collecting water samples and testing them for great crested newt eDNA, an approach developed by the Freshwater Habitats Trust can now be used to correctly identify ponds where newts are present or absent with a much higher success rate than previously. This provides decision makers with the information they need much more quickly, thus reducing costs to developers and facilitating conservation of this threatened amphibian. Natural England and Defra have now adopted this eDNA test as part of the formal process for consenting developments where Great Crested Newts are likely to be present.

Where are they?

Since molecular methods frequently sample remains or traces of organisms, there is an additional complication in linking a molecular record to an actual occurrence of a living organism. In more stable, static environments (e.g. soils, ponds and lakes) this is unlikely to be an issue, but in more mobile environments such as rivers or the sea the potential for eDNA transport is more significant. Current research programmes are studying the transport of eDNA in rivers and the marine environment to better understand the effects of this.

Most current monitoring records the presence of living or dead organisms to at least a 6 figure grid reference (i.e. 10m²), though data may be analysed at lower resolutions such as 10km² for simplicity. From a regulatory perspective, it is not usual to analyse biological data at a coarser grain than this and so ongoing research will illustrate the different scales over which eDNA/conventional analyses integrate biodiversity information in relation to existing approaches. However, potential monitoring issues related to the spatial scale over which eDNA analyses may reflect broader biodiversity could be overcome by adjusting the sampling technique (for example collecting community DNA instead of free eDNA). Conversely, the scale at which eDNA analyses may reflect biodiversity could offer additional

insights in relation to our understanding of broader, catchment scale level biodiversity in relation to environmental pressures/land-use. Nevertheless, such insights currently fall outside the remit of standard monitoring approaches.

When were they recorded?

Existing methods generally record sightings on a daily basis, though this has become lost by some datasets (e.g. Atlas data). As with spatial resolution, since molecular methods may be sampling traces of organisms rather than the organisms themselves, there can be additional uncertainty.

DNA can persist in the environment for some time depending upon the habitat and conditions, and therefore it is possible that organisms recorded in a sample were actually present weeks, months or even (in some cases, such as ancient sediments) decades, or centuries ago. As with transport, this is an area that requires further study and will be highly substrate dependent. It is well documented that small and fragmented DNA can become bound to sediments and persist for substantial periods of time. Nevertheless, initial work suggests that DNA samples correspond reasonably well with seasonal variation as measured by conventional methods. In general the detection of older DNA requires more specialised techniques and therefore its effect on regular sampling is likely to be small. Consequently, it is likely that DNA records are likely to reflect timescales that are ecologically relevant for the vast majority of applications. However, where very fine resolution is required (i.e. less than two weeks), molecular techniques are likely to lose resolution.

How much / many?

Although not essential for all applications, estimations of abundance greatly increase the value of most biological data. Molecular data is not truly quantitative and will likely never give exact estimates of abundance in terms of biomass of any given species. It is also important to acknowledge the confounding issue of the occurrence of different life history stages, eggs, larvae and adult phases, contributing to the molecular genetic signal. Given the importance of this issue, especially in relation to biomonitoring (e.g. EU Water Framework Directive), estimates of habitat quality and understanding ecological interactions, there are very few studies that provide adequate data contrasting molecular data with abundance estimated using conventional methods. However, in most cases conventional methods of estimating abundance are also relatively imprecise, hence the frequent use of broad abundance classes rather than absolute numbers. For different reasons, abundance estimates of different species may also be biased using either conventional or molecular methods.

With appropriate molecular genetic sampling design and / or lab analysis it is often possible to gain insights into the relative abundance of communities that correlates well with measurements using conventional methods, suggesting that molecular methods are capable of estimating abundance at least in general terms. Further work is needed in this area, but initial results suggest that in many cases, molecular methods have the potential to estimate abundance to a comparable level of accuracy and precision to conventional methods. However, where estimates of habitat structure (e.g. zonation, cover / extent, mapping) are required, molecular methods will in general be unsuitable or more expensive than conventional methods.

Environmental Quality

Initially, molecular methods are being used to transpose existing biomonitoring tools in order to provide more cost-effective ways to measure pressures (Case Study 3). However, it is already apparent

that molecular methods are detecting a much wider range of taxa than was previously possible. Once suitable research datasets across a range of different habitats and pressure gradients are available, it should be possible to construct more reliable pressure indices and also to objectively measure pressures and trends that were not previously quantifiable. This more applied use of molecular methods is of particular importance, and should be a key research goal.

Despite the massive potential of eDNA approaches for rapid and cost-effective widespread monitoring of biodiversity; there are limitations to the current approaches. A number of these are outlined in Figure 2. First and foremost, this is a recent but rapidly developing technology and particularly for larger organisms there is currently no consensus on which marker gene is most appropriate to reliably discriminate between recognised taxa. Whilst this is a highly active current area of research, it is likely that there may never be a single marker gene which gives reliable species level discrimination across the variety of life. The extent to which this is a problem depends upon the purpose of data collection but may to some extent restrict the opportunity for integrating different molecular datasets if a standardised marker system has not been established. Similar problems already exist for some conventional datasets. Further advancements in sequencing technologies may overcome these limitations to a certain extent, when it is possible to sequence larger proportions of the genomes present in eDNA.

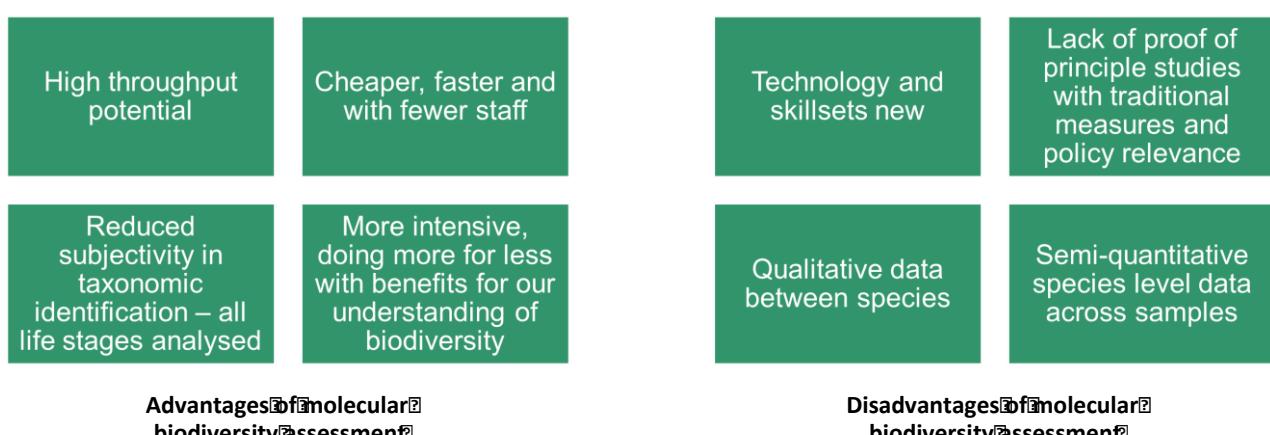


Figure 2. The advantages of molecular biodiversity assessment are primarily related to resource efficiency in implementation and a reduced level of uncertainty. Such power leverages the ability to analyse more biodiversity from more samples, with associated benefits for assessing the relationship between biodiversity and ecosystem health, and/or doing the same job with less resource. Conversely, critics often focus on the quantitative nature of the data, potential biases and technical artefacts and how relevant the different approaches are to traditional approaches associated with policy relevance.

Molecular methods deal with minute quantities of DNA and the risk of contamination is therefore a significant one. In many cases it is not easy to identify the source of any contamination. Scrupulous quality control mechanisms would be needed to ensure that correct conclusions are reached, especially if the outcome of the test affects livelihoods.

Case Study 3: Molecular Approaches for Water Framework Directive Monitoring

The Water Framework Directive (WFD) requires member states to develop and use monitoring tools to assess the ecological quality of their freshwater and inshore marine waters. These tools cover a range of taxon groups and should respond to environmental pressures such as nutrients. The Directive includes quite a detailed framework on the nature and structure of these tools including the taxon groups to be monitored, general parameters to be assessed (e.g. diversity and abundance) and procedures for ensuring comparability among member states. Tools exist for a wide range of taxon groups including macroinvertebrates, aquatic plants, diatoms, phytoplankton and fish. WFD monitoring occurs in a network of thousands of sample points throughout the UK in and therefore is very costly.

The UK group tasked with managing the technical development of the freshwater ecological tools is currently investigating and assessing the options for using eDNA in WFD monitoring in partnership with the scientific community. The ecological tools vary substantially in their suitability for eDNA conversion and cross-calibration is required to ensure that comparable results are obtained to the existing method. One, the rivers diatom tool, is likely to be operationally ready within the next year; others such as the lakes diatom tool and one of the lakes invertebrate tools show promise. eDNA is also allowing the development of a cost-effective and non-damaging lake fish tool; existing methods require the use of gill nets which kill large numbers of fish and are very ineffective at detecting many species.

Initial estimates suggest that, where suitable, eDNA methods are around 30-40% cheaper than existing WFD methods, though this is highly scale-dependant. In future it may be possible to identify sensitive taxa from groups that cannot be identified using traditional methods and therefore improve the power of our biomonitoring tools.

Although the operating cost of molecular methods is low, the setup cost is high. Purpose built laboratories with highly trained personnel are required. This initial investment can be recouped by economies of scale, but due to the large size of the investment, the adoption of eDNA techniques is likely to be most effective if undertaken collaboratively at a UK level.

Finally, there is a general concern that eDNA could discourage appreciation of nature and ecosystems by taking a very technocratic but less aesthetic view of nature that minimises time spent in the field, over time produces a more deskilled workforce and produces a disconnect between ecologists and the environments they study. Certainly molecular methods place a much greater emphasis on laboratory work than many other methods, although conventional approaches such as sorting invertebrates or counting diatoms using a microscope is also heavily laboratory based. However, the existing taxonomic skills base for many groups is already inadequate or non-existent, resulting in very poor national coverage. In essence, this means that conservation of them is weak or non-existent and national experts become overburdened with identifying individual specimens rather than studying their taxonomy or ecology. Used correctly, molecular methods have the potential to break these logjams and facilitate new insights into ecosystems and the biodiversity they support. In turn, this creates the potential to create new and better tools for measuring the status of our environment.

What could the technology deliver within a five-year timescale?

Via the GMEP program, we have already implemented proof of principle studies to assess terrestrial microbial (in particular, bacteria and fungi) biodiversity across the Welsh landscape and this could be enhanced to include quantitative measures where appropriate. For example, the DNA archive could be used at any time to assess the distribution of specific organisms (e.g. the causative agent for bovine TB, *Mycobacterium bovis* in soil). Dŵr Cymru, Welsh Water also have requirements to develop pathogen (e.g. *E.coli* and *Cryptosporidium*) and odour imparting organism (e.g. *Geosmin* and *Methylisoborneol*) detection assays.

Molecular methods have also been used extensively in freshwaters, where they are beginning to contribute directly to statutory processes. As well as their use for detecting great crested newts, work in France demonstrates that they can be used to monitor whole amphibian and fish communities in ponds and rivers. Since most amphibian species receive some form of protection in the UK, such a test is highly relevant. Moreover, in the UK, some WFD methods are also being transposed to eDNA and are likely to become operational within the foreseeable future. Were funds to be made available, single species or whole community tests for rare species or even entire communities would be a reasonable aspiration within a relatively short timescale.

The potential for molecular approaches for assessment of biomonitoring of larger taxa, with concomitant benefits for efficiency and resource use is particularly high. The current Welsh Government/DEFRA annual spend on evidence is in the region of £200 million, with approximately 35% of this attributed to statutory reporting. With the appropriate level of resource and collaboration with existing stakeholder bodies and/or research organisations, we would be in a position to identify how molecular genetic approaches for biodiversity assessment could enhance existing approaches employed in the freshwater (lentic and lotic), marine and terrestrial biomes. Moreover, we would also be able to employ a cost benefit analysis (including socio economic considerations), what level of information can be obtained from the different approaches and how molecular approaches could be incorporated into statutory reporting. Many of these goals are being replicated across Europe by distributed networks of researchers seeking to enhance the way that we assess the relationship between biodiversity and ecosystem health (<http://www.cost.eu/COST Actions/ca/CA15219?management>).

Costs

As per traditional analyses, samples still need to be collected from the field and so if downstream analyses are necessary, the costs for the field component are roughly similar whichever approach is used. Typically, the costs of extracting DNA from a sample is £6, while commercial suppliers offer custom biodiversity sequencing at £40-80 per sample assay. The establishment of a bespoke facility would require careful costing but it is highly likely that the cost-benefit incurred for an original start-up would be favourable for larger scale operations, with diminishing returns for smaller-scale operations/assessments. For the latter, using existing facilities or commercial providers could be economically preferential alternatives.

Summary

In conclusion, molecular approaches provide new ways to assess biodiversity at a variety of spatial scales. In some cases, they are unlikely to replace existing and well-established survey approaches (e.g. for birds and butterflies). However, in many cases they can generate results that are comparable to, or better than existing methods at a lower cost, or over a longer survey season. For many taxon groups and environments (e.g. microbes, soils, lake fish), they provide fresh and transformative insights into Welsh biodiversity, replacing existing methods that were ineffective and costly.

The cost of analysis and manpower involved in these molecular approaches is often reduced in comparison to existing approaches. In addition, DNA samples of less than 1ml can be preserved indefinitely in an archive allowing targeting of specific questions as and when the policy need arises (e.g. to evaluate the presence of a pathogenic bacteria or insect disease vector). As an exemplar, GMEP has pioneered the use of molecular biodiversity assessment to assess the impact of land management on soil organisms in Wales and similar opportunities exist in the marine and freshwater biomes spanning the full spectrum of organismal diversity. Carefully designed strategic molecular sampling networks that take account of the minimum spatial, temporal and taxonomic requirements in the terrestrial, freshwater and marine environments could be used to provide Wales level biodiversity datasets for a range of operational, management and policy purposes, augmenting and in some cases replacing other data collection approaches. In addition, more specific eDNA tools and tests could be developed for specific policy drivers, as is already taking place for the Water Framework Directive.

The above notwithstanding, it is important to stress that molecular approaches remain one tool among many. In our view, they are a highly efficient, powerful and effective tool for many biodiversity related applications and we expect that they will become cheaper, more widely used, accessible and accurate with time. However, they are not a panacea: there will remain applications, environments and species where other approaches are more informative and/or cost-effective.

Further Reading

What key methods feature in using molecular approaches for biodiversity discovery?

Quantitative (qPCR) or real time PCR (rtPCR)

It is widely acknowledged that real-time or quantitative PCR (qPCR) represents the gold standard in both the qualitative and quantitative assessment of cells/biomass. Enhanced by recent recommendations for minimum quality, qPCR is widely employed at the diagnostic level and has been used extensively in the development of single species approaches to detect rare and endangered species via the analysis of aqueous eDNA. Of particular note is that eDNA evidence is now accepted at the statutory level to assess the presence of the endangered great crested newt for DEFRA. Nevertheless, qPCR is only useful for targeting specific taxa (either a “species” or broader taxonomic group), reducing its efficacy and raising costs when assessing the composition of diverse communities.

Marker Gene Assessment - Metabarcoding

Marker gene studies have become the most prevalent approach, typically relying on broad coverage PCR primers to amplify marker genes from environmental samples. Whilst not as directly quantitative as qPCR approaches, the main advantage is the rapid assessment of the change in relative abundances of a broader range of taxa. Currently implemented marker genes include the ribosomal rRNA marker for bacteria and some eukaryotes (though the validity as a species specific taxonomic marker is acknowledged as weak for the latter); the ribosomal RNA Internal Transcribed Spacer region (ITS) mainly for fungi but also wider eukaryotes; and the Cytochrome oxidase subunit 1 (COI) gene which is being touted as a universally informative marker for larger eukaryotes (see <http://boldsystems.org/>). Marker gene assessments are more generally known as ‘amplicon’, ‘metagenetic’, ‘metasystematic’ and metabarcoding sequencing among many others. The recent advancement which has facilitated the rise of these approaches is the development of high throughput sequencing technologies. These approaches allow the simultaneous analyses of several hundred PCR amplified DNA samples in a single assay; utilising complex but now well established bioinformatic approaches to essentially generate quality filtered tables of taxon abundances across many samples.

Metagenomics – environmental shotgun sequencing

Prokaryotic Communities

True ‘metagenomic’ approaches utilize random sequencing of genomic fragments isolated from environmental samples to elucidate both the taxonomic and functional genomic capability of a community. Shotgun sequencing can provide a complementary, independent method for assessing community diversity, additionally allowing for the capture of information from groups that are otherwise difficult to survey. Metagenomic data are typically used in two ways. The taxonomic component of shotgun sequencing can be used to identify organisms present in a sample, followed by ecologically informative analyses. Metagenomes can also be used to characterize the functional potential of microbial communities through investigation of their full genomic repertoire.

Microscopic and macroscopic eukaryotic communities

Environmental shotgun sequencing could resolve some of the biases prevalent in metabarcoding studies, particularly if it is used in conjunction with targeted genome sequencing. Accordingly, the sequencing of DNA from organelles is developing as an alternative: mitochondrial genomes for animals and chloroplast genomes for plants. Clearly, sequencing the genomes of mixed communities, compared to specific genetic loci, requires a huge increase in sequencing power and consequently a reduction in sample throughput. An alternative relies on using DNA capture array technology to target specific organelles. Here, arrays are designed from existing genomic organelle information which are used to hybridize and extract specific regions from genomic DNA, thereby reducing the size of the genomic target and increasing throughput. It is likely that different studies will utilize different approaches depending on budget, sample number, community composition and questions.

Future molecular approaches that are in development and may represent the future of the field include metatranscriptomics and targeted genome sequencing, but are not covered here, since they will less likely to become operational within the near to mid future.

Future potential for directly assessing functionality?

Given the vast amount of functions performed by biodiversity, and particularly microbial biodiversity, it is hard to directly infer that a change in abundance of a particular taxa will result in a change in functionality. Such questions are better addressed by directly addressing change in specific gene pathways, such as the genes responsible for the degradation of a carbon source, nitrification, or pathogenicity etc.. Molecular approaches based on sequencing the whole soil DNA pool (whole genome metagenomics) or total transcribed RNA (metatranscriptomics), and then counting reads annotated to functional categories offers a potentially more useful approach to directly addressing change in functionality. Despite advances in sequencing technology the costs required to conduct such analyses often restrict the analysis of 100s-1000s of samples, although it is likely that in the future sequencing costs will come down and such approaches will become more routine.

RNA or DNA?

It has long been acknowledged that DNA may be highly resistant to degradation and may persist in the environment for long periods. Therefore there have been numerous concerns that the detection of genes/organisms through DNA based approaches may not derive from functionally active organisms. For this reason, several studies have explored the sequencing of either the ribosomal RNA marker directly for taxonomic investigations; or transcribed RNA for functional studies (metatranscriptomics). Particularly for soil systems, there have been few studies which have reported major changes in the communities assessed by either the RNA and DNA approaches for taxonomic investigations. This is possibly because active organisms may also be numerically abundant; or alternatively because ribosomal RNA is also long lived in soil. Given this and also the greater degree of labour required for working with RNA due to the lack of high throughput approaches for extraction, it is unlikely in the near future that RNA methods will become routine for large scale monitoring. With respect to metagenomic approaches for functionality; DNA based methods are considered to give

information on functional potential, but again it is thought that targeting RNA directly (metatranscriptomics) will better reflect the genes which are functionally expressed at a given time. Unfortunately at present these methods are very much in their infancy; and so there are few studies which have directly compared the results from both methods. Undoubtedly direct sequencing of the total RNA pool should provide more information on both taxonomic identity (no use of specific rRNA primers) and functional genes; and so could be a good solution to addressing both taxonomy and function in the future once sequencing costs decrease.

Definitions at a glance

Amplicon sequencing. Targeted sequencing of an amplified marker gene.

Community DNA. Defined here as the DNA derived from many individuals representing several species.

Degenerate primers. A mixture of similar, but not identical oligonucleotide sequences used for amplicon sequencing where the targeted gene(s) is typically similar, but not identical.

Environmental DNA (eDNA). DNA isolated directly from an environmental sample (e.g. air, faeces, sediment, soil, water).

Genomic DNA. Defined here as the DNA derived from a single individual or from a collection of individuals of the same species.

Locus. The specific location of a gene or DNA sequence on a chromosome.

Marker gene. A gene or DNA sequence targeted in amplicon sequencing to screen for a specific organism group or functional gene.

Metabarcoding. Uses gene-specific PCR primers to amplify DNA from a collection of organisms or from environmental DNA. Another term for amplicon sequencing.

Metagenomics. The random sequencing of gene fragments isolated from environmental samples, allowing sequencing of uncultivable organisms.

Metatranscriptomics. Shotgun sequencing of total RNA from environmental samples. Techniques such as poly-A amplification or rRNA depletion are often used to target messenger (mRNA) transcripts to assess gene expression patterns in complex communities.

Polymerase chain reaction (PCR). Used to amplify a targeted piece of DNA, generating many copies of that particular DNA sequence.



BANGOR

Centre for Ecology & Hydrology
Environment Centre Wales
Deiniol Road
Bangor
Gwynedd
LL57 2UW
United Kingdom
T: +44 (0)1248 374500
F: +44 (0)1248 362133

EDINBURGH

Centre for Ecology & Hydrology
Bush Estate
Penicuik
Midlothian
EH26 0QB
United Kingdom
T: +44 (0)131 4454343
F: +44 (0)131 4453943

LANCASTER

Centre for Ecology & Hydrology
Lancaster Environment Centre
Library Avenue
Bailrigg
Lancaster
LA1 4AP
United Kingdom
T: +44 (0)1524 595800
F: +44 (0)1524 61536

WALLINGFORD - Headquarters

Centre for Ecology & Hydrology
Maclean Building
Benson Lane
Crowmarsh Gifford
Wallingford
Oxfordshire
OX10 8BB
United Kingdom
T: +44 (0)1491 838800
F: +44 (0)1491 692424