MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 2045–2050

NEWS AND VIEWS

OPINION

Towards next-generation biodiversity assessment using DNA metabarcoding

PIERRE TABERLET,* ERIC COISSAC,* FRANÇOIS POMPANON,* CHRISTIAN BROCHMANN† and ESKE WILLERSLEV‡

**Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, F-38041 Grenoble Cedex 9, France, †National Centre for Biosystematics, Natural History Museum, University of Oslo, PO Box 1172 Blindern, NO-0318 Oslo, Norway, ‡Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, DK-1359 Copenhagen K, Denmark*

Virtually all empirical ecological studies require species identification during data collection. DNA metabarcoding refers to the automated identification of multiple species from a single bulk sample containing entire organisms or from a single environmental sample containing degraded DNA (soil, water, faeces, etc.). It can be implemented for both modern and ancient environmental samples. The availability of next-generation sequencing platforms and the ecologists' need for high-throughput taxon identification have facilitated the emergence of DNA metabarcoding. The potential power of DNA metabarcoding as it is implemented today is limited mainly by its dependency on PCR and by the considerable investment needed to build comprehensive taxonomic reference libraries. Further developments associated with the impressive progress in DNA sequencing will eliminate the currently required DNA amplification step, and comprehensive taxonomic reference libraries composed of whole organellar genomes and repetitive ribosomal nuclear DNA can be built based on the well-curated DNA extract collections maintained by standardized barcoding initiatives. The near-term future of DNA metabarcoding has an enormous potential to boost data acquisition in biodiversity research.

Keywords: capture probes, DNA-based species identification, next-generation DNA sequencing

Received 1 October 2011; revision received 14 December 2011; accepted 16 December 2011

Correspondence: Pierre Taberlet, Fax: +33(0)4 76 51 42 79; E-mail: pierre.taberlet@ujf-grenoble.fr

Introduction

DNA sequencing technologies have undergone impressive improvements with the recent emergence of next-generation sequencing (NGS) platforms (Shendure & Ji 2008; Glenn 2011). These new platforms can provide billions of sequence reads in a single experiment, which corresponds to an improvement of at least five orders of magnitude when compared to traditional Sanger sequencing using capillary electrophoresis. Such a dramatic leap in sequencing capacity has the potential to revolutionize many areas of scientific inquiry. However, while the immediate impact on genomics and transcriptomics is obvious, the impact of the new technologies on biodiversity research is more difficult to assess. In this opinion paper, we examine the current impact of NGS for biodiversity research and point to future trends in high-throughput species identification using degraded DNA from multiple species in environmental samples.

DNA-based species identification

Virtually all empirical ecological studies require species identification during data collection. Identification typically relies on easily observable morphological characteristics, but various DNA-based strategies have been developed for those in cases where morphology-based identification proved problematic (Fig. 1). The first methods were proposed at the end of the 1980s and were based on DNA hybridization (Southern blots) either by using specific probes (e.g. Gale & Crampton 1987; Gibson *et al.* 1988) or by prior restriction enzyme digestion and electrophoresis (restriction fragment length polymorphism, RFLP; e.g. Curran & Webster 1987). With the invention of PCR-based amplification of DNA (Mullis & Faloona 1987; Saiki *et al.* 1988) and the design of universal primers (e.g. Kocher *et al.* 1989; Taberlet *et al.* 1991), species identification moved towards being based on direct sequencing (e.g. Cronin *et al.* 1991) or hybridization (e.g. Teletchea *et al.* 2008) of PCR products. Initially, many different nuclear and organellar DNA regions were targeted for DNA amplification and sequencing, but the approach has now been standardized and designated 'DNA barcoding' (Hebert *et al.* 2003). This initiative is supported by the Consortium for the Barcode of Life (CBoL; http://www.barcodeoflife.org/). The standardized DNA barcodes are a 658-bp region of the mitochondrial cytochrome c oxidase I gene (COI) for animals (Hebert *et al.* 2003), and two 500- to 800-bp plastid fragments of the large subunit of ribulose 1,5-bisphosphate carboxylase gene (*rbc*L) and the maturase K gene (*mat*K) for plants (Hollingsworth *et al.* 2009). Whereas only a few articles dealing specifically with DNA-based species identification were available

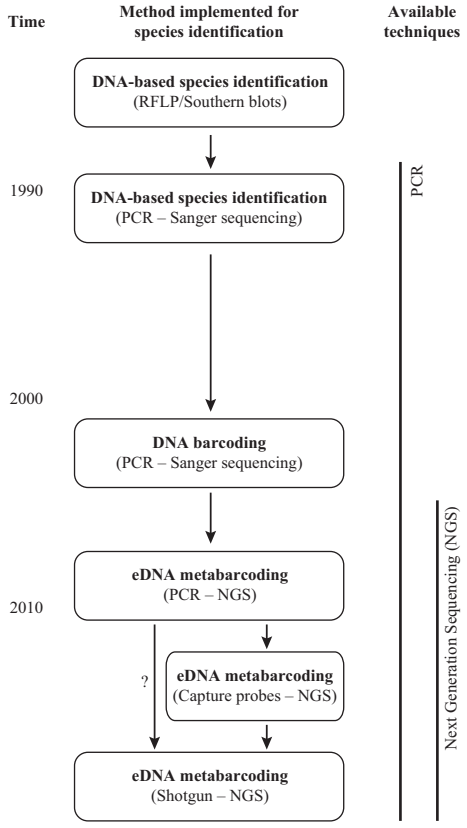

Fig. 1 DNA-based species identification. Past and current approaches, and possible future trends.

before 2003, hundreds of articles have been published since the emergence of the DNA barcoding concept. Clearly, standardization was an important step in the development of DNA-based species identification, and it has encouraged extensive international efforts to build taxonomic reference libraries of the standardized regions. However, the barcoding standards were designed to identify species from more or less intact DNA isolated from single specimens using Sanger sequencing, and focus more on the variability of the amplified region than on the nonvariability of the primer sites and the length of the targeted DNA region.

The emergence of DNA metabarcoding in relation to next-generation sequencing and to the needs of the scientific community

Here, we introduce the term 'DNA metabarcoding' to designate high-throughput multispecies (or higher-level taxon) identification using the total and typically degraded DNA extracted from an environmental sample (i.e. soil, water, faeces, etc.). Species identification from bulk samples of entire organisms (e.g. Chariton *et al.* 2010; Creer *et al.* 2010; Porazinska *et al.* 2010; Hajibabaei *et al.* 2011), where the organisms are isolated prior to analysis, can also be considered as DNA metabarcoding. Below, we will restrict our considerations to the analysis of environmental DNA (eDNA), because analysis of bulk samples has very relaxed technical constraints compared to that of environmental samples. Bulk samples are usually composed of a restricted taxonomic group and provide high-quality DNA allowing the use of a longer barcode, even the standardized ones. We will also emphasize that because the goal of DNA metabarcoding is to identify taxa, it should be clearly differentiated from metagenomics that 'describes the functional and sequence-based analysis of the collective microbial genomes contained in an environmental sample' (Riesenfeld *et al.* 2004).

The emergence of DNA metabarcoding was because of technology catching up to a scientific need. Standardized ('traditional') DNA barcoding does not fulfil all the needs of ecologists. As it is designed to identify single specimens with DNA that is more or less intact, it typically requires the isolation of a suitable specimen to be analysed, which is time-consuming and, for some taxonomic groups, difficult or virtually impossible. Consequently, standardized barcoding is limited in the number of specimens that it can identify. We must therefore accept that standardized DNA barcoding is not ideal for high-throughput species identification for use in ecological studies, although it has an obvious added value in many situations where classical species identification is difficult and in facilitating the discovery of new species. These limitations have been surmounted by the increasing availability of NGS machines that permit high-throughput techniques such as DNA metabarcoding. At the moment, sequencing platforms can produce up to 6 billions of sequence reads of 100 bp per run, with the possibility to implement paired-end experiments (Glenn 2011). Thus, it is not any more a problem to obtain several thousands of sequence reads per amplicon, and the length of the sequence reads is already fully compatible with the short fragment lengths required for eDNA metabarcoding. There is no doubt that the technology will improve still further. As a consequence, NGS has the potential to provide an enormous amount of information per experiment from in-depth sequencing of uniquely tagged amplicons (Binladen *et al.* 2007; Valentini *et al.* 2009). So, why not use eDNA to simultaneously identify many species in a single experiment? After some initial experiments based on PCR/cloning/sequencing (Willerslev *et al.* 2003, 2007), the approach using NGS has already demonstrated its potential, for analysing plant communities using soil samples (Yoccoz *et al.* 2012), for reconstructing past plant or animal communities using permafrost or ice samples (Haile *et al.* 2009; Sønstebø *et al.* 2010; Boessenkool *et al.* 2012; Jørgensen *et al.* 2012a,b; Epp *et al.* submitted), for tracking earthworms using soil samples (Bienert *et al.* 2012), for monitoring vertebrate biodiversity (Andersen *et al.* 2012), or for diet analysis using faeces or stomach content as a source of DNA (see review in Pompanon *et al.* 2012).

However, there are significant constraints when designing an eDNA metabarcoding study. First, eDNA is often highly degraded, and long fragments of several hundreds of base pairs cannot be reliably amplified (Willerslev *et al.* 2004; Hansen *et al.* 2006). Second, because many species have to be amplified in the same PCR experiment, it is extremely

important that the primers used for the amplification be highly versatile, that is, that they amplify many different target molecules with the same efficiency, without missing species containing target sequences that do not match well with the primers. For these two reasons, the standardized DNA barcodes cannot reliably be used for analysing eDNA, and new DNA metabarcodes must be designed according to the aim of each study. Several bioinformatic tools have recently been designed for finding the most optimal metabarcodes and primers (ecoPrimers program; Riaz *et al.* 2011) and for testing their characteristics (ecoPCR program; Bellemain *et al.* 2010; Ficetola *et al.* 2010).

Current limitations of eDNA metabarcoding

Despite its potential power for high-throughput species identification, eDNA metabarcoding as implemented today has several limitations. One such limitation is its current dependency on PCR. The requirement for a DNA amplification step in the eDNA metabarcoding approach has many drawbacks.

First, PCR can introduce errors during the amplification (Cline *et al.* 1996), both substitutions and insertions/deletions. In metabarcoding studies, errors observed in the output sequences have three origins: degradation of template DNA (comparable to that observed in ancient DNA experiments), errors during amplification and errors during sequencing, the latter varying among different sequencing technologies. Errors produced during PCR appear to be abundant relative to sequencing errors, which occur at a rate of about 0.25% per nucleotide for the Roche 454 and Illumina sequencing platforms, the Roche 454 being more prone to errors in homopolymers (Glenn 2011). The use of a proofreading polymerase might reduce substitution errors, but will not solve the main problem of polymerase slippage during the elongation (Hauge & Litt 1993; Litt *et al.* 1993; Murray *et al.* 1993). Such slippage artefacts produce length variation in homopolymers, resulting mainly in shorter artefactual copies.

The second drawback with using PCR on eDNA is its reliance on finding a suitable metabarcode, namely, one that possesses a short variable DNA region suitable to target a particular taxonomic group, flanked by two highly conserved regions of ~20 bp to anchor the primers. The length constraint to allow working with highly degraded DNA will also reduce the taxonomic resolution of metabarcodes compared with that of the 500–800 bp long standardized barcodes. Nevertheless, suitable metabarcoding markers have been relatively easy to find, for example, for vascular plants and bryophytes (Taberlet *et al.* 2007; Epp *et al.* submitted), vertebrates (Riaz *et al.* 2011; Shehzad *et al.* 2012), birds (Epp *et al.* submitted), earthworms (Bienert *et al.* 2012) and fungi (Epp *et al.* submitted). Nonetheless, some taxonomic groups are recalcitrant, and it is difficult to find suitable metabarcodes on mitochondrial DNA, even when using the ecoPrimers program (Riaz *et al.* 2011). This is the case, by way of example, for nematodes that have very divergent mitochondrial genomes, making the identification of a short variable region flanked by two conserved regions difficult.

The third problem with PCR is attributed to the need to analyse different groups of organisms (archaea, bacteria, fungi, plants, arthropods, vertebrates, etc.) separately. As a consequence, it is extremely difficult to assess the relative proportion of each group within a DNA extract.

Another important limitation of eDNA metabarcoding is the need for high-quality taxonomic reference databases, that is, libraries containing the targeted sequences of the relevant species obtained from sequencing taxonomically verified and curated specimens that are secured in long-term storage (such as the library provided for circumarctic vascular plants by Sønstebø *et al.* 2010). As high-quality reference libraries are time- and resource-consuming to build, the ideal situation would have been to design short metabarcodes within the standardized barcodes. Such an approach would benefit from the enormous effort made by the International Barcode of Life (http://ibol.org/) to build reference libraries containing several hundred thousand species. Unfortunately, designing suitable short metabarcodes within the standardized barcodes is a very difficult task. This is because the primer targets of the standardized barcodes are protein-coding. Even if it were possible to find conserved regions at the amino acid level, the corresponding DNA sequence (i.e. the target sequence of the primers) is typically variable, especially at every third nucleotide of each codon. Although primers targeting such regions can amplify DNA from single specimens relatively efficiently, they do not provide equally good matches to all target sequences derived from a bulk or environmental sample (i.e. a mixture of many species), introducing bias in DNA amplifications. They are therefore not suitable for DNA metabarcoding.

However, a considerable part of the investment needed to build reference libraries goes to specimen collection, identification by taxonomic experts, databasing, DNA extraction, and curation and storage of the reference specimens and DNA extracts. The enormous collection of DNA extracts built by standardized barcoding initiatives can therefore also serve as a resource for eDNA metabarcoding, as new target regions can be sequenced based on the same DNA extracts.

Finally, identifying a species from a single organellar marker is problematic, as discussed for standardized barcoding (see e.g. Will *et al.* 2005; Rubinoff *et al.* 2006). In this respect, eDNA metabarcoding suffers from the same drawback as standardized barcoding. Interspecific mitochondrial and plastid introgressions are common and well documented (Rieseberg & Soltis 1991; Petit & Excoffier 2009), and species that are young in evolutionary terms have often not yet achieved monophyly for these markers (i.e. incomplete lineage sorting), which can potentially lead to erroneous identifications.

The future of eDNA metabarcoding

Based on the above discussion, it becomes clear that the main limitations of eDNA metabarcoding are its depen-

dency on PCR and on the requirement for considerable investment to build comprehensive taxonomic reference libraries.

How to avoid PCR in eDNA metabarcoding experiments? One possibility would be to replace the PCR by a capture approach (Fig. 1; see e.g. Hodges *et al.* 2007; Briggs *et al.* 2009; Hodges *et al.* 2009; Avila-Arcos *et al.* 2011) using oligonucleotides (=capture probes) that target conserved regions. As single conserved regions are often scattered (i.e. not as pairs separated by short distances as required to design metabarcodes for PCR), the possibility of finding suitable capture probes is much higher than that of finding suitable PCR-based metabarcodes. Furthermore, as many capture probes can be used in the same experiment, this strategy creates the possibility to analyse hundreds of short DNA regions flanking the capture probes, all at once. This means that not only can many different taxonomic groups be analysed using the same eDNA extract and in the same capture experiment, but several capture probes can be used for the same taxonomic groups, targeting both organellar and nuclear DNA regions. Thus, a standardized set of capture probes has the potential to decipher, in a single experiment, most of the species contained in an eDNA extract, including those of archaea, bacteria, protists, fungi, plants and animals. We suggest focusing on repetitive DNA, that is, plastid, mitochondrial and nuclear ribosomal DNA to increase the proportion of informative fragments in the eDNA extract.

A simpler possibility to avoid PCR would be to directly sequence the eDNA extract (Fig. 1) with NGS platforms, which can produce several billions sequence reads per experiment (e.g. using the Illumina HiSeq 2000 platform). Such experiments would be similar to the current metagenomic approach (e.g. Riesenfeld *et al.* 2004; Tringe *et al.* 2005), except that the goal would not be to assemble whole microorganism genomes or to find the most common functional genes, but to use repetitive DNA to identify the different species. However, at the moment, we do not know the proportion of potentially informative sequence reads (i.e. the proportion of mitochondrial, chloroplast and nuclear ribosomal DNA) that is possible to obtain in such a sequencing experiment. In any case, this proportion should be higher than one per cent for plant and animal DNA, as chloroplast DNA represents around 10% of the total DNA in plants (Bowman 1986), mitochondrial DNA around 1% of the total in animals, and ribosomal nuclear DNA is massively repeated. Currently, a single sequencing lane on the Illumina HiSeq 2000 produces more than 300 million reads, which can probably lead to more than one million informative plant and animal reads, but this depends on the proportion of DNA from archaea, bacteria and fungi. Furthermore, such a shotgun approach would have the big potential advantage to provide the relative proportions among the different types of DNA originating from the different taxonomic groups.

How to meet the need for high-quality taxonomic reference libraries for eDNA metabarcoding? The power of NGS platforms already allows the routine sequencing of whole mitochondrial and whole plastid genomes (Moore *et al.* 2006; Timmermans *et al.* 2010; Nock *et al.* 2011). A relatively limited shotgun sequencing of any eukaryote species can lead to suitable coverage of organelle DNA as well as of ribosomal nuclear DNA. Sequencing several thousands of whole plastid or mitochondrial genomes is now realistic within a single research project. The very large collection of well-curated DNA extracts built by standardized barcoding initiatives can also serve as a resource for this purpose. We can therefore imagine that in the very near future, the number of whole-mitochondria and plastid sequences will dramatically increase, facilitating operational species identification from shotgun sequencing of eDNA, provided that appropriate bioinformatic tools are then available.

Conclusion

Environmental DNA metabarcoding has an enormous potential to boost data acquisition in biodiversity research. At the moment, we are at the very beginning of perfecting this approach, but we anticipate that further developments associated with the impressive recent progress made in DNA sequencing technologies will allow elimination of a DNA amplification step, and that comprehensive taxonomic reference libraries composed of whole organellar genomes and repetitive ribosomal nuclear DNA can be built. In this context, a major challenge will be to develop new bioinformatic pipelines especially designed for exploiting such massive amounts of sequence data in the most efficient way for DNA-based species identification.

Acknowledgements

This work is financially supported by the European Commission under the Sixth Framework Programme (EcoChange project, contract no FP6-036866) and by the Research Council of Norway (BarFrost project, no 191627/V40).

References

Andersen K, Bird KL, Rasmussen M *et al.* (2012) Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, **21**, 1966–1979.

Avila-Arcos MC, Cappellini E, Romero-Navarro JA *et al.* (2011) Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Scientific Reports*, **1**, 74. doi: 10.1038/srep00074.

Bellemain E, Carlsen T, Brochmann C *et al.* (2010) ITS as DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiology*, **10**, 189.

Bienert R, de Danieli S, Miquel C *et al.* (2012) Tracking earthworm communities from soil DNA. *Molecular Ecology*, **21**, 2017–2030.

Binladen J, Gilbert MTP, Bollback JP *et al.* (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, **2**, e197.

Boessenkool S, Epp L, Haile J *et al.* (2012) Blocking human contaminant DNA during PCR allows amplification of rare mammal

species from sedimentary ancient DNA. *Molecular Ecology*, **21**, 1806–1815.

Bowman CM (1986) Copy numbers of chloroplast and nuclear genomes are proportional in mature mesophyll cells of *Triticum* and *Aegilops* species. *Planta*, **167**, 264–274.

Briggs AW, Good JM, Green RE *et al.* (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science*, **325**, 318–321.

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

Cline J, Braman JC, Hogrefe HH (1996) PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, **24**, 3546–3551.

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

Cronin MA, Palmisciano DA, Vyse ER, Cameron DG (1991) Mitochondrial DNA in wildlife forensic science: species identification of tissues. *Wildlife Society Bulletin*, **19**, 94–105.

Curran J, Webster JM (1987) Identification of nematodes using restriction fragment length differences and species-specific DNA probes. *Canadian Journal of Plant Pathology-Revue Canadienne de Phytopathologie*, **9**, 162–166.

Epp L, Boessenkool S, Bellemain E *et al.* (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular Ecology*, **21**, 1821–1833.

Ficetola GF, Coissac E, Zundel S *et al.* (2010) An *in silico* approach for the evaluation of DNA barcodes. *BMC Genomics*, **11**, 434.

Gale KR, Crampton JM (1987) DNA probes for species identification of mosquitos in the *Anopheles gambiae* complex. *Medical and Veterinary Entomology*, **1**, 127–136.

Gibson WC, Dukes P, Gashumba JK (1988) Species-specific DNA probes for the identification of African trypanosomes in tsetse flies. *Parasitology*, **97**, 63–73.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, **11**, 759–769.

Haile J, Froese DG, MacPhee RDE *et al.* (2009) Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 22352–22357.

Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One*, **6**, e17497.

Hansen AJ, Mitchel DL, Wiuf C *et al.* (2006) Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics*, **173**, 1175–1179.

Hauge XY, Litt M (1993) A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. *Human Molecular Genetics*, **2**, 411–415.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identification through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 313–321.

Hodges E, Xuan Z, Balija V *et al.* (2007) Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*, **39**, 1522–1527.

Hodges E, Rooks M, Xuan ZY *et al.* (2009) Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nature Protocols*, **4**, 960–974.

Hollingsworth PM, Forrest LL, Spouge JL *et al.* (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12794–12797.

Jørgensen T, Haile J, Möller P *et al.* (2012a) A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational. *Molecular Ecology*, **21**, 1989–2003.

Jørgensen T, Kjær KH, Haile J *et al.* (2012b) Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA Meta-barcoding. *Molecular Ecology*, 1980–1988.

Kocher TD, Thomas WK, Meyer A *et al.* (1989) Dynamics of mitochondrial DNA evolution in animals : amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 6196–6200.

Litt M, Hauge X, Sharma V (1993) Shadow bands seen when typing polymorphic dinucleotide repeats: some causes and cures. *BioTechniques*, **15**, 280–284.

Moore MJ, Dhingra A, Soltis PS *et al.* (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology*, **6**, 17.

Mullis KB, Faloona FA (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalysed chain reaction. *Methods in Enzymology*, **155**, 335–350.

Murray V, Monchawin C, England PR (1993) The determination of the sequences present in the shadow bands of a dinucleotide repeat PCR. *Nucleic Acids Research*, **21**, 2395–2398.

Nock CJ, Waters DLE, Edwards MA *et al.* (2011) Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal*, **9**, 328–333.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology & Evolution*, **24**, 386–393.

Pompanon F, Deagle BE, Symondson WOC *et al.* (2012) Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology*, **21**, 1931–1950.

Porazinska DL, Giblin-Davis RM, Esquivel A *et al.* (2010) Ecometagenetics confirms high tropical rainforest nematode diversity. *Molecular Ecology*, **19**, 5521–5530.

Riaz T, Shehzad W, Viari A *et al.* (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, **39**, e145.

Rieseberg LH, Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*, **5**, 65–84.

Riesenfeld CS, Schloss PD, Handelsman J (2004) Metagenomics: genomic analysis of microbial communities. *Annual Review of Genetics*, **38**, 525–552.

Rubinoff D, Cameron S, Will K (2006) A genomic perspective on the shortcomings of mitochondrial DNA for ''barcoding'' identification. *Journal of Heredity*, **97**, 581–594.

Saiki RK, Gelfand DH, Stoffel S *et al.* (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.

Shehzad W, Riaz T, Nawaz MA *et al.* (2012) Carnivore diet analysis based on next generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, **21**, 1951–1965.

Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**, 1035–1045.

Sønstebø JH, Gielly L, Brysting A *et al.* (2010) A minimalist DNA barcoding approach for reconstructing past Arctic vegetation and climate. *Molecular Ecology Resources*, **10**, 1009–1018.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.

Taberlet P, Coissac E, Pompanon F *et al.* (2007) Power and limitations of the chloroplast *trn*L(UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, **35**, e14.

Teletchea F, Bernillon J, Duffraisse M, Laudet V, Hanni C (2008) Molecular identification of vertebrate species by oligonucleotide microarray in food and forensic samples. *Journal of Applied Ecology*, **45**, 967–975.

Timmermans MJTN, Dodsworth S, Culverwell CL *et al.* (2010) Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic Acids Research*, **38**, e197.

Tringe SG, von Mering C, Kobayashi A *et al.* (2005) Comparative metagenomics of microbial communities. *Science*, **308**, 554–557.

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

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844–851.

Willerslev E, Hansen AJ, Binladen J *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791–795.

Willerslev E, Hansen AJ, Poinar HN (2004) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology & Evolution*, **19**, 141–147.

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

Yoccoz NG, Bråthen KA, Gielly L *et al.* (2012) DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, 10.1111/j.1365-294X.2012.05545.x.

Since a few years, all authors are strongly involved in developing the DNA metabarcoding approach, focusing on the bioinformatics aspects (E.C.), on the analysis of ancient samples (E.W., C.B.), on soil analysis (P.T.), and on diet studies (F.P. and P.T.).

---

doi: 10.1111/j.1365-294X.2012.05470.x