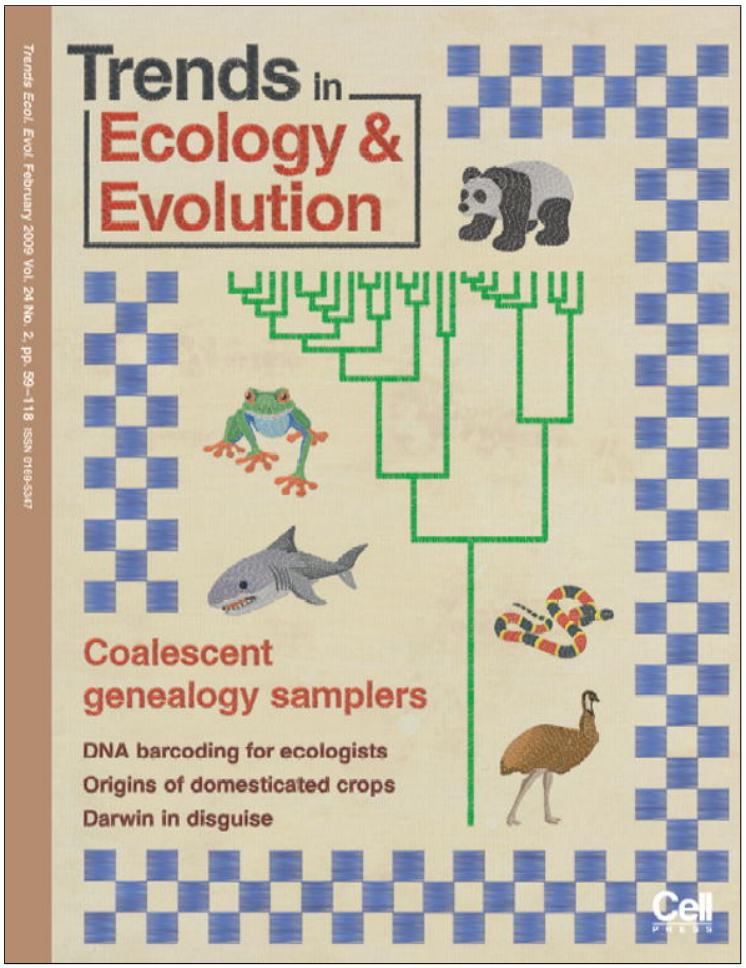

Review

DNA barcoding for ecologists

Alice Valentini[1,2], François Pompanon[1] and Pierre Taberlet[1]

[1] Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France
[2] Dipartimento di Ecologia e Sviluppo Economico Sostenibile, Università della Tuscia, Via S. Giovanni Decollato 1, 01100 Viterbo, Italy

DNA barcoding – taxon identification using a standardized DNA region – has received much attention recently, and is being further developed through an international initiative. We anticipate that DNA barcoding techniques will be increasingly used by ecologists. They will be able to not only identify a single species from a specimen or an organism's remains but also determine the species composition of environmental samples. Short DNA fragments persist in the environment and might allow an assessment of local biodiversity from soil or water. Even DNA-based diet composition can be estimated using fecal samples. Here we review the new avenues offered to ecologists by DNA barcoding, particularly in the context of new sequencing technologies.

A new name for an old concept

The term 'DNA barcoding' is of recent use in the literature [1,2]. It relies on the use of a standardized DNA region as a tag for rapid and accurate species identification [3]. Nevertheless, DNA barcoding is not a new concept. The term 'DNA barcodes' was first used in 1993 [4], in a paper that did not receive very much attention from the scientific community. The concept of species identification using molecular tools is older still [5]. However, the golden age of DNA barcoding began in 2003 [2]. The now well-established Consortium for the Barcode of Life (CBOL, http://barcoding.si.edu), an international initiative supporting the development of DNA barcoding, aims to both promote global standards and coordinate research in DNA barcoding. For animals, the gene region proposed for the standard barcode is a 658 base pair region in the gene encoding the mitochondrial cytochrome *c* oxidase 1 (COI) [2]. For plants the situation is controversial, and many strategies have been proposed, either based on a single chloroplast region [6,7] or on a combination of different regions [8,9].

Taxonomists are not the only potential users of DNA barcoding, as the technique can be useful for scientists from other fields (e.g. forensic science, the biotechnology and food industries and animal diet) [10]. Taxonomists are concerned with DNA barcoding *sensu stricto*. Other scientists will be more interested with DNA barcoding *sensu lato*, that is by DNA-based taxon identification using diverse techniques that can lie outside the CBOL approach (Box 1; see Table S1 in the online supplementary material). The difference between the two approaches derives mainly from different priorities given to the criteria used for designing the molecular markers (Box 1).

Corresponding author: Taberlet, P. (pierre.taberlet@ujf-grenoble.fr).

Here our purpose is to review the main applications of DNA barcoding in ecology, with an emphasis on the new opportunities linked to the recent availability of next-generation DNA sequencers. First, we will present the new tools that make the barcoding approach for ecologists easier. We will then focus on single-species identification, which is the historical fundament of DNA barcoding. Finally, we will discuss the unprecedented potential of DNA barcoding for simultaneous multiple-species identification from a single environmental sample, for biodiversity assessment and for diet analysis from feces.

New tools for new prospects

In the past 20 years the technology of DNA sequencing has greatly improved, from manual sequencing to automated sequencers. A single automated 96-capillary sequencer can provide more than 1000 sequences of 1000 base pairs (bp) per day. Even non-geneticists now have easy access to sequencing via companies that offer this service at a competitive price. Clearly, the development of DNA barcoding is linked to these improvements.

When using the classical sequencing approach via capillary electrophoresis, environmental samples (Box 2) require an additional step of cloning the different amplified DNA fragments into bacteria, followed by sequencing hundreds or thousands of clones to reveal the full complexity of those samples. Such a cloning step is both expensive and time consuming, thus limiting large-scale applications. New DNA sequencing technologies bypassing the cloning step have recently been developed (Box 3), opening the way to applying large-scale DNA barcoding studies to environmental samples (Figure 1).

More and more sequence data for the accepted barcoding markers are becoming available in public databases (GenBank, http://www.ncbi.nlm.nih.gov; EMBL, http://www.ebi.ac.uk/embl; DDBJ, http://www.ddbj.nig.ac.jp) as sequencing facilities improve. This greatly stimulates the development of species identification via DNA barcoding, and enhances the design of standardized methods by allowing a better design of 'universal' primers. However, the quality of the sequence data in GenBank, EMBL or DDBJ is not always perfect [11], either as a result of sequencing errors, contaminations, sample misidentifications or taxonomic problems. CBOL's recent initiative to build a new database specially dedicated to DNA barcoding will change this situation, and will provide an efficient and accurate tool for species identification (Barcode of Life Data Systems, BOLD, http://www.barcodinglife.org). BOLD has been designed to record not only DNA sequences from several individuals per species (including primer sets,

0169-5347/$ – see front matter doi:10.1016/j.tree.2008.09.011 Available online 25 December 2008

Box 1. DNA barcoding: definitions and ideal properties

DNA barcoding *sensu stricto* corresponds to the identification of the species level using a single standardized DNA fragment. This definition fits with the CBOL view [2].

The definition of DNA barcoding *sensu lato* is much less restrictive. It corresponds to the identification of any taxonomical level using any DNA fragment. The identification of genera or families, from an environmental sample using a very short DNA fragment that has not been recognized as the standardized barcode, can nevertheless be considered as DNA barcoding *sensu lato*. Taxon identification with diagnostic single-nucleotide polymorphisms (SNPs) or with insertions/deletions (indels) can also be considered as DNA barcoding *sensu lato* (see Table S1 in the online supplementary material). The identification of populations within a given species, based either on diagnostic markers or on the differences in marker frequencies among populations, refers to assignment tests [83], and should not be considered as DNA barcoding.

The ideal DNA barcoding system should meet the following criteria [69].

(i) The gene region sequenced should be nearly identical among individuals of the same species, but different between species.
(ii) It should be standardized, with the same DNA region used for different taxonomic groups.
(iii) The target DNA region should contain enough phylogenetic information to easily assign unknown or not yet 'barcoded' species to their taxonomic group (genus, family, etc.).
(iv) It should be extremely robust, with highly conserved priming sites and highly reliable DNA amplifications and sequencing. This is particularly important when using environmental samples, where each extract contains a mixture of many species to be identified at the same time.
(v) The target DNA region should be short enough to allow amplification of degraded DNA. Usually, DNA regions longer than 150 bp are difficult to amplify from degraded DNA.

Thus, the ideal DNA barcoding marker should be variable, standardized, phylogenetically informative, extremely robust and short. Unfortunately, such an ideal DNA marker has not yet been found or, perhaps, does not even exist [49]. As a consequence, taking into account the scientific and technical contexts, the various categories of users (e.g. taxonomists, ecologists, etc.) will not give the same priority to the five criteria listed above [9]. Criteria (i), (ii) and (iii) are the most important for taxonomists (DNA barcoding *sensu stricto*), whereas ecologists working with environmental samples will favor criteria (iv) and (v) (DNA barcoding *sensu lato*) [69].

Box 2. Environmental samples

An environmental sample is a mix of organic and inorganic materials taken from the environment (e.g. soil or feces). It can contain live individuals (i.e. microorganisms or small macroorganisms such as nematodes or springtails) and remains of macroorganisms present around the sampling site. Until now, environmental samples have been used mainly for studying microbial communities, using 16S or the internal transcribed spacer of rDNA as the barcode [43–45,84]. In this case, DNA sequences of several hundreds of base pairs can be retrieved because DNA of good quality is extracted from live microorganisms. Environmental samples should also be useful for characterizing the diversity of macro-organic species (such as plant or animals) in an ecosystem. Here the DNA comes from dead macroorganisms, is in most cases highly degraded, and only short sequences can be amplified; the new sequencing techniques (Box 3), coupled with universal primers that amplify short fragments, now enable the identification of various macro-organic taxa. A few studies have already been carried out using different types of environmental samples:

- water for microbial biodiversity [33] and for identifying an invasive species [40];
- modern soil for comparative analyses of bacterial communities [43–45];
- permafrost for assessing past animal and plant communities [51];
- silty ice at the base of an ice cap for assessing past animal and plant communities at the onset of glacial conditions [52];
- rodents' middens for describing past plant communities [53];
- feces samples for diet analysis [68,70].

electropherogram trace files and translations) but also complete taxonomic information, place and date of collection, and specimen images [12]. All these improvements will probably boost the use of DNA barcoding by ecologists.

Single-species identification

The classical use of morphological traits for species identification has several limitations. They include, for example, the misidentification of a taxon due to the phenotypic plasticity of the trait studied or the existence of cryptic taxa [13]. Moreover, morphological keys are sometimes only effective for a particular life stage or gender, as for example in Diptera, where species identification is mainly based on male genitalia [14]. Thus, a high level of expertise is often required to correctly identify species with the accuracy required in ecological studies. The DNA barcoding approach might currently represent the best solution

Box 3. The next-generation sequencing systems

The current barcoding system (CBOL approach) has been designed to fit with DNA sequencers based on capillary electrophoresis, which yield a typical read length of 500–1000 bp. Recently, next-generation sequencing systems have become available [85,86]. Several new techniques have been implemented, all based on the sequencing of individual DNA molecules (with or without an amplification step), and on a massively parallel approach allowing the simultaneous sequencing of at least hundreds of thousands of molecules (see Table I). All the new sequencers, except for one, produce very short fragments. The only system that allows sequenced fragments longer than 25–35 bp is the 454 GS FLX (Roche) [87], which currently delivers 400 bp fragments. The enormous amount of relatively long sequences produced make this new sequencer suitable for environmental barcoding studies where there is the need to deal with complex samples composed of a mix of many species (e.g. deep sea biodiversity [39] and diet analysis [70]).

Table I. Next-generation sequencing systems and their characteristics

Sequencer	Company	Fragment length per read	Number of reads	Total output	Time per run
Genetic Analyzer/Solexa [79]	Illumina	25–50 bases	60 000 000 reads per run	2 Gb per run	6.5 days
SOLiD DNA Sequencer [80]	Applied Biosystems	25–35 bases	85 000 000 reads per run	3 Gb per run	6 days
Heliscope [81]	Helicos	25–50 bases	100 000 000 bases per hr	2 Gb bases per day	Not relevant
454 GS FLX [82]	Roche Diagnostics	400 bases	2 000 000 reads per run	1 Gb per run	8 h

Abbreviations: Gb, gigabases.

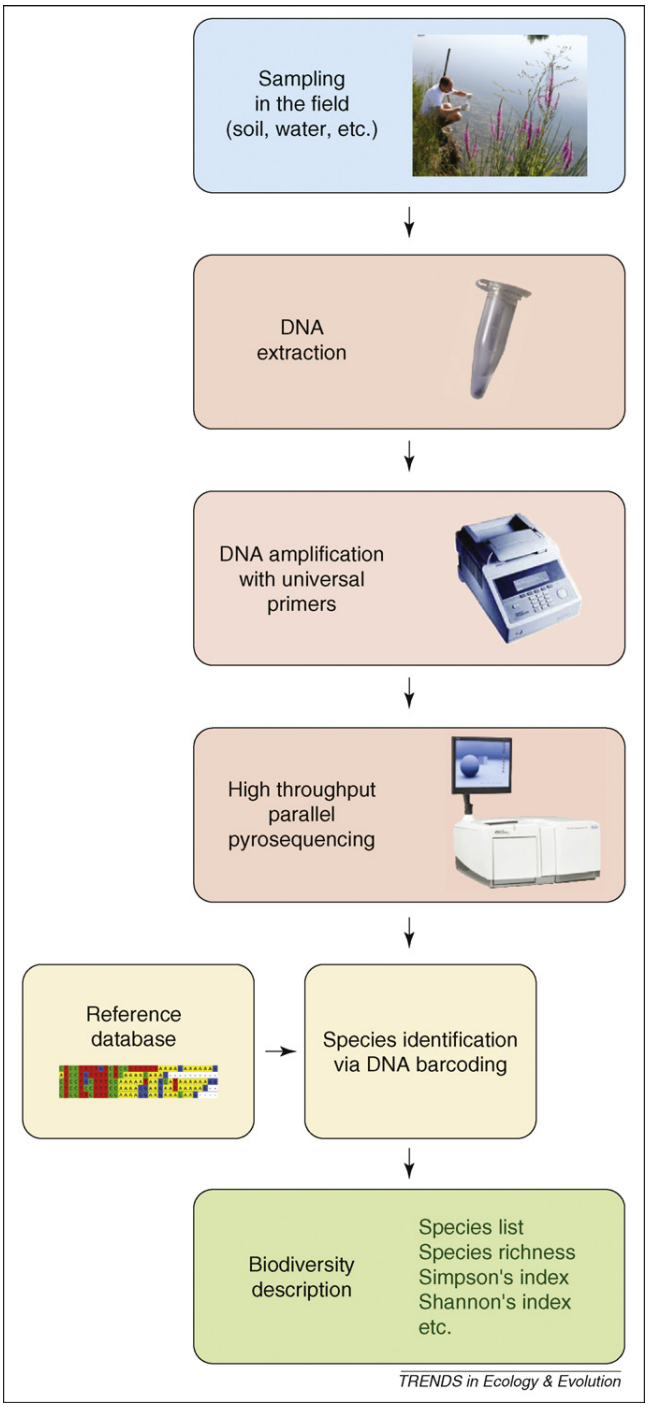

Figure 1. Methodology for analyzing biodiversity from environmental samples based on next-generation DNA sequencers. After collecting environmental samples in the field, extracting DNA and amplifying with universal primers that target very short DNA fragments (less than 150 base pairs), hundreds or thousands of amplified DNA molecules are sequenced using next-generation sequencers (Box 3). Using a reference DNA database, the taxa these sequences come from are identified and used to estimate different biodiversity parameters.

for identifying species when their morphology is of limited use [15], even if DNA barcoding itself also presents some limitations (see e.g. Ref. [16] and the section Limitations and Perspectives below). The following examples illustrate some situations where DNA barcoding has greatly helped ecologists.

Nematodes play an important role in ecosystems. They can contribute to nitrogen mineralization and distribution of biomass within plants in soil ecosystems [17]. Despite this important role, their identification is still extremely challenging. Due to these difficulties, their biodiversity is greatly underestimated. For example, it is estimated that global marine nematode species richness exceeds one million, while only a few thousand of them are described [18]. For a better assessment of nematode biodiversity, ecologists studying this group should take the opportunity to use DNA barcoding in both marine and terrestrial environments [1,19,20].

In addition, ecologists can take advantage of DNA tools when only hair, feces or urine left behind by animals are available for species identification. Such an approach is now widely used, and is particularly useful for detecting the presence of elusive or endangered species. For example, the recent wolf range expansion in France and Switzerland has been traced back using mainly feces and hair samples [21]. Another study demonstrated that feces from the sympatric Amur leopard (*Panthera pardus orientalis*) and Siberian tiger (*P. tigris altaica*) can be easily distinguished based on mitochondrial DNA polymorphism, enabling the possibility to record the respective presence of these two endangered subspecies in the field [22].

Identifying species through DNA barcoding is also helpful for understanding interspecies interactions [23,24]. For example, the barcoding approach has already shown that the existence of cryptic species could mask the specialization of a parasite to a single host. Several morphospecies of tachinid parasitoid flies thought to be generalist actually corresponded to many different cryptic species that were less generalist or even specialist [25,26]. In the same way, three species of Cerambycidae insects (*Palame* spp.) feeding on trees of the Lecythidaceae family yielded three new more specialized cryptic species with different host and season preferences [27].

Furthermore, DNA barcoding can be advantageous for monitoring illegal trade in animal byproducts. When such products are sold, identification through morphological characteristics might no longer be possible. Sometimes only hairs are available for species identification, and it is very difficult or even impossible to visually determine whether a hair came from an endangered or a legally sold species. DNA methodology has been successfully implemented to identify Eurasian badger (*Meles meles*) hairs in expensive shaving brushes [28] and Tibetan antelope (*Pantholops hodgsonii*) in shahtoosh, a luxury shawl [29]. Many other examples can be found in the food industry (see e.g. Ref. [30]). For instance, scientists using this technique revealed that 23% of black caviar samples purchased in the USA were labeled with an incorrect species name, and in some cases the commercial species (Russian sturgeon, *Acipenser gueldenstaedtii*) was replaced with an endangered one (ship sturgeon, *A. nudiventris*) [31].

In the field of biosecurity, the reliable and fast identification of exotic species is fundamental. In many cases, particularly for insects, a pest at the egg or larval stage might not be recognizable without DNA identification. For example, stem borer larvae of the genus *Busseola* occurring in Ethiopia on sugarcane were identified using COI

sequences [32]. The North American bullfrog (*Rana catesbeiana*) is one of the world's worst invasive species, preying upon and competing with local species. Recently, a novel approach allowed the detection of this invasive frog in southwest France, using water samples from ponds and targeting a very short mitochondrial DNA fragment [33]. To summarize, there are many situations where DNA-based single-species identification has the potential to help solve important ecological questions.

Biodiversity assessment

DNA barcoding might also play an important role in biodiversity assessment, both for present and for past animal and plant communities.

Current biodiversity

Even if morphological identification of a species is possible, DNA barcoding might enhance biodiversity inventories by being faster and cheaper, and by overcoming the taxonomic impediment [34,35]. It could allow biodiversity assessment through the identification of taxa from the traces of DNA present in environmental samples such as soil or water (Figure 1). The cost for identifying a sample via barcoding, including the processes of DNA extraction, DNA amplification, purification of the PCR product and sequencing using capillary electrophoresis, has been estimated to range from $2.5 to $8 per sample, depending on laboratory facilities and consumable equipment [36–38]. A technician in DNA barcoding could replace dozens of taxonomists for routine identification, allowing taxonomists to concentrate on identifying reference specimens for establishing reliable databases. The use of DNA barcoding will not be necessary for assessing the biodiversity of well-known ecosystems such as Alpine prairies. However, in ecosystems showing high species richness, such as those in tropical environments, it is unrealistic, within a limited time period, to identify all animals and plants by morphology alone. The biodiversity of environments with low accessibility can also be estimated with this new technique, as demonstrated by the study of the microbial biodiversity in deep sea [39]. Large-scale studies also become possible, because the barcoding approach allows the simultaneous identification of most species from a given biotope (e.g. the Global Ocean Sampling Expedition [40], http://www.jcvi.org/cms/research/projects/gos).

The use of classical biodiversity indices, such as species richness, Simpson's index and Shannon's index [41], could be complemented by new indices developed to exploit the information contained in whole sequence sets obtained from a single environment. Thus, the estimation of biodiversity indices can be based on molecular operational taxonomic units (MOTU) detected using the barcoding approach [1,42], where the relative abundances of each type of DNA sequence (MOTU) replace the classical relative abundance of each species estimated from the number of individuals. This approach is now common in environmental microbiology for estimating different diversity indices (species richness, Shannon's or Simpson's indices; e.g. Refs [43–45]), but presents some bias when the number of species is very large (>1000) [46]. To our knowledge, the use of MOTUs for estimating biodiversity indices has not yet been evaluated for plants or animals using environmental samples such as water or soil. Potential bias can be generated either by the occurrence of a single MOTU for different species, leading to an underestimation of the biodiversity (caused by low resolution of the barcoding marker), or by the occurrence of many MOTUs for a single species, leading to an overestimation of the biodiversity (caused by intraspecific polymorphism of the barcoding marker) [47–50]. With the recent development of parallel pyrosequencing, this type of approach might also be evaluated for plants and animals in the near future.

Paleoecology

Generally, reconstructing the ecosystems of the past from fossil data is very difficult because of the nature of the samples and their low preservation. In most cases, morphological identification to the species level is very difficult or even impossible. In this case, molecular tools and DNA barcoding could help scientists to successfully describe past plant and animal communities, and to reconstruct past environments.

Only a few sites are suitable for DNA-based paleoecology, where molecules are well preserved in dry or cold environments. Past communities have been analyzed from samples collected in Siberian permafrost sediment from the Pleistocene and Holocene periods, revealing a change in plant composition between these two periods, and identifying eight different species of mammals, including mammoth (*Mammuthus primigenius*), musk ox (*Ovibos moschatus*), reindeer (*Rangifer tarandus*) and lemming (*Lemmus lemmus*) [51]. The same type of study has been recently carried out on samples taken from 450 000-year-old silty ice extracted from the bottom of the Greenland ice cap. The results showed that southern Greenland was covered by a forest at that time, composed of trees of the genera *Picea*, *Pinus* and *Alnus* as in the forests found in southern Scandinavia today [52].

Rodent middens represent another interesting source of information for studying past communities. Middens are a mix of plant macrofossils, pollen, rodent feces, bones and insects coming from an area within a radius of ~100 m from the den and agglomerated by the rodent's urine salts. Using animal mitochondrial genes and chloroplast genes, it is possible to identify the rodent species and to deduce the nature of their environment from the flora and fauna present. DNA analyses carried out on 11 700-year-old middens from the Atacama Desert in Chile [53] revealed that this past environment was more productive, more diverse and much more humid, with a fivefold decrease in precipitation since that time.

A molecular approach has also been used for studying the diet of extinct animals. The diet of the ground sloth *Nothrotheriops shastensis* was studied using the large subunit of the ribulose-bisphosphate carboxylase (*rbc*L) chloroplast gene from five coprolites in Gypsum Cave, Nevada. These coprolites were dated approximately to 11 000, 20 000 and 28 500 years BP. Additionally, it was possible to study the vegetation changes caused by different climate conditions in those three periods. According to the vegetation identified, it appears that the climate is

dryer now than ~11 000 years ago [54]. Another interesting study dealt with the analysis of the last meals of a Neolithic glacier mummy [55], which DNA sequences revealed to be red deer (*Cervus elaphus*) and alpine ibex (*Capra ibex*). Thus, the association of ancient DNA with the barcoding concept offers unique opportunities to reconstruct past environments.

Diet analysis

The study of food webs and their dynamics is fundamental to understand how the feeding habits of different species can affect the community. Thus, diet analysis of the animal species present in a given environment can improve our understanding of the functioning of the ecosystem as a whole (e.g. Ref. [56]). Furthermore, the study of feeding ecology becomes crucial when it concerns endangered species. A precise knowledge of the diet of these species can identify key environmental resources for designing reliable conservation strategies (e.g. Refs [57,58]). DNA barcoding makes it possible to establish the diet of an individual from its feces or stomach contents. This is helpful when the food is not identifiable by morphological criteria, such as in liquid feeders, for example spiders [59]. This technique also provides valuable information when the diet cannot be deduced by observing the eating behavior, as in the case of krill-eating diatoms [60], giant squid (*Architeuthis* spp.) in the sea abyss [61] or deep sea invertebrates (e.g. two amphipods and one bivalve species; see Ref. [62]). Most of the studies that use DNA markers for diet analysis are based on carnivorous animals, for example insects [63,64], fin whale (*Balaenoptera physalus*) and Adelie penguin (*Pygoscelis adeliae*) [65]. Fewer studies have been carried on herbivorous animals such as western gorillas (*Gorilla gorilla*) and black and white colobus monkeys (*Colobus guereza*) [66].

There are two different strategies when using molecular tools for diet analysis: the use of group-specific primers [67] and the use of universal primers. In general, the use of specific primers requires an *a priori* knowledge of the animal's diet. This is not possible in most cases, and makes the universal approach more appropriate. Both group-specific and universal primers were implemented when analyzing the diet of the Macaroni penguin (*Eudyptes chrysolophus*) using feces as a source of DNA [68]. The results obtained with five different sets of specific primers were similar to those involving universal 16S rDNA primers and subsequent cloning of the PCR products. This clearly demonstrates the relevance of using universal primers for diet analysis. Recently, the possible use of a very short fragment (10–140 bp) of the intron of the gene encoding chloroplast tRNA for leucine, codon UAA (*trn*L [UAA] intron), as a target for plant barcoding in situations where only degraded DNA is available has been proposed [69]. The *trn*L approach is suitable for diet studies of herbivorous animals because the primers are universal (for gymnosperms and angiosperms) and because it works for feces that contain degraded DNA. This barcoding system, combined with high-throughput parallel pyrosequencing, was successfully used to amplify and analyze the diet of mammals (Figure 2), birds, mollusks and insects [70]. Thus, DNA barcoding might become a precious tool for field ecologists interested in the diet of their favorite model species.

Limitations and perspectives

The main limitation of the barcoding approach comes from its single-locus identification system. Even if several regions from these organelle DNAs are sequenced, this is still a single-locus approach because the different genes of mitochondrial or chloroplast DNA are linked. It is well known that identical mitochondrial or chloroplast DNA sequences can be present in different related species due to introgression, or due to incomplete lineage sorting since the time of speciation [71]. Furthermore, nuclear copies of fragments of mitochondrial or chloroplast DNA are common and can be preferentially amplified in some circumstances [72], leading to potential identification errors. Finally, heteroplasmy can also confuse the identification system [73]. These potential problems have been extensively discussed (e.g. Ref. [16]). Additionally, the use of a divergence threshold for distinguishing intra- versus interspecific sequence variation [2] suffers from severe statistical problems, and can seriously compromise species identification [74]. Up to 17% of species misidentification has been observed in cypraeid marine gasteropods if the reference database is not comprehensive, that is does not contain all the species of the group under study [47]. Another statistical difficulty is caused by the low number of individuals analyzed per species, which does not allow a precise estimation of the intraspecific variation compared with interspecific variation [48,49].

Another limitation of the current DNA barcoding approach *sensu stricto* lies in the length of the sequences used, usually >500 bp [2], which prevents the amplification of degraded DNA. Unfortunately, many potential DNA barcoding applications can only be based on degraded DNA. This is the case for all environmental samples where the target is DNA from dead animals or dead parts of plants. It is usually difficult to amplify DNA fragments longer than 150 bp from such samples [75]. As a consequence, there is a need for shorter barcoding markers [69,76,77]. We can imagine that in the near future, species identification could be carried out by the combined analysis of several short universal barcoding markers. Furthermore, when analyzing environmental samples using parallel pyrosequencing (Figure 1), it would be tempting to interpret the results in a quantitative way, based on the number of DNA molecules assigned to the different MOTUs [70]. However, this attractive perspective still requires an empirical validation.

We anticipate that ecologists will increasingly turn to the DNA barcoding approach, because in many circumstances it represents the only easy way to identify species. This trend will be further enhanced by the availability of reliable databases now under construction (e.g. BOLD). However, as suggested by the diverse DNA-based taxon identification used until now, ecologists will probably continue to develop diverse approaches in addition to the standardized DNA barcoding. There are some circumstances where the number of species to distinguish is limited, and where the sequencing of long DNA fragments is not necessary (see e.g. Ref. [22]). Another advantage of

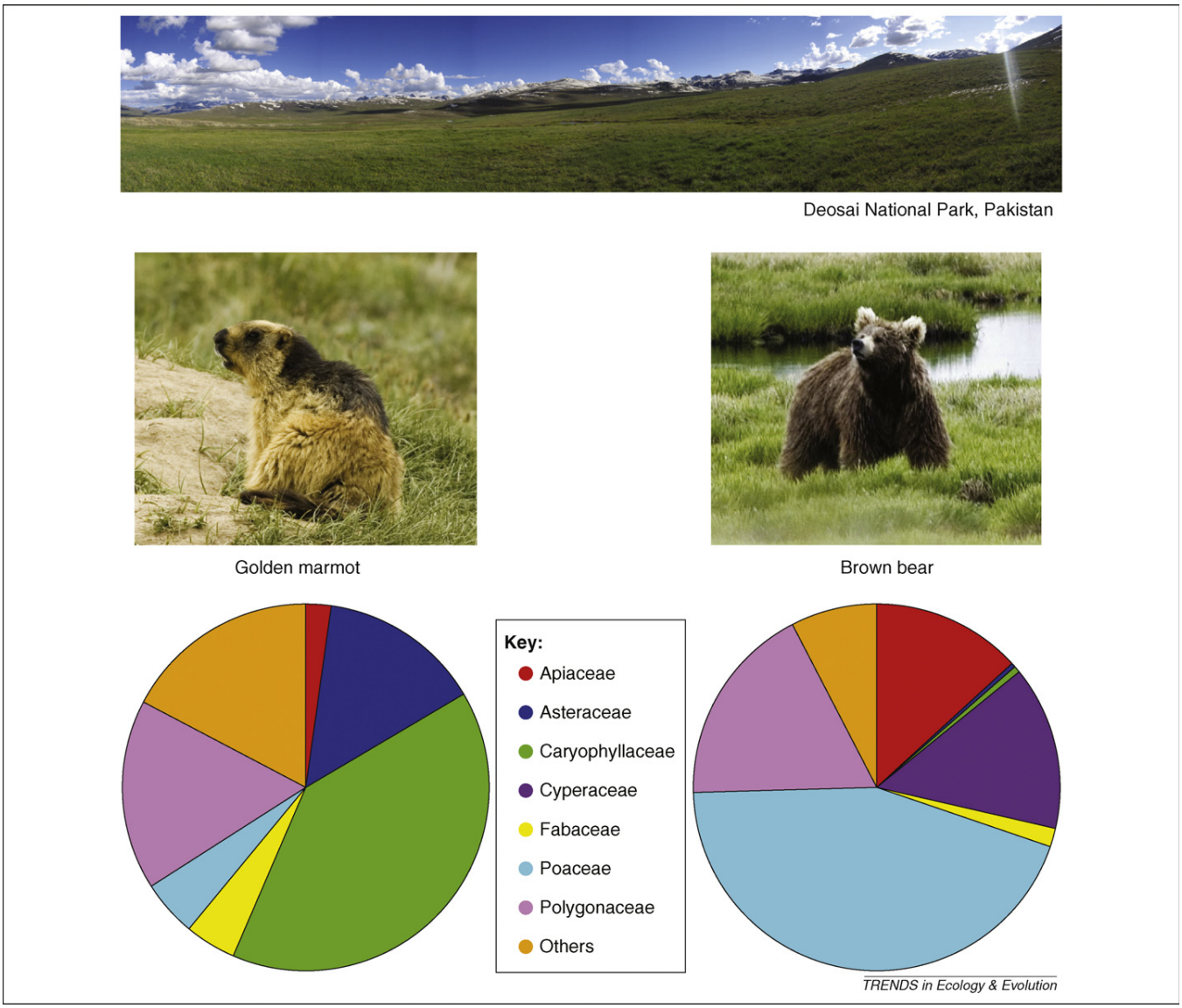

Figure 2. Example of feces analysis for estimating diet composition using a DNA barcoding approach (from Ref. [70]). Comparison of brown bear (*Ursus arctos*) and golden marmot (*Marmota caudata*) plant diets in the Himalayan environment. Twelve feces of each species were collected in the Deosai National Park (Pakistan). After DNA extraction and amplification with universal primers targeting a short fragment of the chloroplast *trn*L (UAA) intron [69], the amplicons were analyzed on the 454 GS FLX sequencer. More than 2000 DNA sequences were obtained per feces. The plant taxa eaten by both species were identified by comparison with available reference sequences. The results show that the diets of the two species are different: bears prefer Poaceae, whereas marmots prefer Caryophyllaceae.

the DNA barcoding approach is that the basic data – the sequences – are not prone to subjectivity and can be reanalyzed in the future in accordance with improvements in taxonomic knowledge.

The possibility of analyzing complex mixtures of DNA fragments, either by DNA hybridization on microarrays [30,78] or in association with the new sequencing technologies (Box 3), opens new horizons for ecologists. We anticipate that DNA-based biodiversity assessment using environmental samples will be implemented for plants and animals, as is already the usual approach for microorganisms. DNA-based taxon identification techniques will soon fully integrate the ecologist's toolbox.

Acknowledgements
We thank four anonymous reviewers for constructive comments on the manuscript. We also thank Claude Miaud and Usman Ghani for kindly providing the photographs of environmental sampling (Figure 1) and of golden marmot and brown bear (Figure 2), respectively.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tree.2008.09.011.

References

1 Floyd, R. *et al.* (2002) Molecular barcodes for soil nematode identification. *Mol. Ecol.* 11, 839–850
2 Hebert, P.D.N. *et al.* (2003) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proc. Biol. Sci.* 270, S96–S99
3 Hebert, P.D.N. and Gregory, T.R. (2005) The promise of DNA barcoding for taxonomy. *Syst. Biol.* 54, 852–859
4 Arnot, D.E. *et al.* (1993) Digital codes from hypervariable tandemly repeated DNA sequences in the *Plasmodium falciparum* circumsporozoite gene can genetically barcode isolates. *Mol. Biochem. Parasitol.* 61, 15–24
5 Kangethe, E.K. *et al.* (1982) Identification of the species origin of fresh meat using an enzyme-linked immunosorbent-assay procedure. *Meat Sci.* 7, 229–240

6 Kress, W.J. *et al.* (2005) Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8369–8374
7 Lahaye, R. *et al.* (2008) DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. U. S. A.* 105, 2923–2928
8 Kress, W.J. and Erickson, D.L. (2007) A two-locus global DNA barcode for land plants: the coding *rbc*L gene complements the non-coding *trn*H-*psb*A spacer region. *PLoS ONE* 2, e508
9 Chase, M.W. *et al.* (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon* 56, 295–299
10 Chase, M.W. *et al.* (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1889–1895
11 Harris, D.J. (2003) Can you bank on GenBank? *Trends Ecol. Evol.* 18, 317–319
12 Ratnasingham, S. and Hebert, P.D.N. (2007) BOLD: the barcode of life data system (www.barcodinglife.org). *Mol. Ecol. Notes* 7, 355–364
13 Knowlton, N. (1993) Sibling species in the sea. *Annu. Rev. Ecol. Syst.* 24, 189–216
14 Hennig, W. (1976) Anthomyiidae. In *Die Fliegen der Palaearktischen Region* (Lindner, E., ed.), pp. 329–376, Schweizerbart
15 Ahrens, D. *et al.* (2007) DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). *Mol. Phylogenet. Evol.* 44, 436–449
16 Rubinoff, D. (2006) Utility of mitochondrial DNA barcodes in species conservation. *Conserv. Biol.* 20, 1026–1033
17 Neher, D.A. (2001) Role of nematodes in soil health and their use as indicators. *J. Nematol.* 33, 161–168
18 Lambshead, P. (2004) Marine nematode biodiversity. In *Nematology: Advances and Perspectives* (Chen, Z.X. *et al.*, eds), pp. 436–467, CABI
19 Bhadury, P. *et al.* (2006) Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes. *Mar. Ecol. Prog. Ser.* 320, 1–9
20 Powers, T. (2004) Nematode molecular diagnostics: from bands to barcodes. *Annu. Rev. Phytopathol.* 42, 367–383
21 Valière, N. *et al.* (2003) Long distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. *Anim. Conserv.* 6, 83–92
22 Sugimoto, T. *et al.* (2006) Species and sex identification from faecal samples of sympatric carnivores, Amur leopard and Siberian tiger, in the Russian Far East. *Conserv. Genet.* 7, 799–802
23 Pfenninger, M. *et al.* (2007) Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic *Chironomus* (Diptera) species. *Mol. Ecol.* 16, 1957–1968
24 Tedersoo, L. *et al.* (2007) Ectomycorrhizal fungi of the Seychelles: diversity patterns and host shifts from the native *Vateriopsis seychellarum* (Dipterocarpaceae) and *Intsia bijuga* (Caesalpiniaceae) to the introduced *Eucalyptus robusta* (Myrtaceae), but not *Pinus caribea* (Pinaceae). *New Phytol.* 175, 321–333
25 Smith, M.A. *et al.* (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proc. Natl. Acad. Sci. U. S. A.* 103, 3657–3662
26 Smith, M.A. *et al.* (2007) DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4967–4972
27 Berkov, A. (2002) The impact of redefined species limits in Palame (Coleoptera: Cerambycidae: Lamiinae: Acanthocinini) on assessments of host, seasonal, and stratum specificity. *Biol. J. Linn. Soc.* 76, 195–209
28 Domingo-Roura, X. *et al.* (2006) Badger hair in shaving brushes comes from protected Eurasian badgers. *Biol. Conserv.* 128, 425–430
29 Lee, J.C.I. *et al.* (2006) DNA profiling of shahtoosh. *Electrophoresis* 27, 3359–3362
30 Teletchea, F. *et al.* (2008) Molecular identification of vertebrate species by oligonucleotide microarray in food and forensic samples. *J. Appl. Ecol.* 45, 967–975
31 Birstein, V.J. *et al.* (2000) Polyphyly of mtDNA lineages in the Russian sturgeon, *Acipenser gueldenstaedtii*: forensic and evolutionary implications. *Conserv. Genet.* 1, 81–88
32 Assefa, Y. *et al.* (2007) DNA identification of *Busseola* (Lepidoptera: Noctuidae) larvae in Ethiopian sugarcane. *Afr. Entomol.* 15, 375–379
33 Ficetola, G.F. *et al.* (2008) Species detection using environmental DNA from water samples. *Biol. Lett.* 4, 423–425
34 Gaston, K.J. and O'Neill, M.A. (2004) Automated species identification: why not? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 655–667
35 Gotelli, N.J. (2004) A taxonomic wish-list for community ecology. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 585–597
36 Hajibabaei, M. *et al.* (2005) Critical factors for assembling a high volume of DNA barcodes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1959–1967
37 Cameron, S. *et al.* (2006) Who will actually use DNA barcoding and what will it cost? *Syst. Biol.* 55, 844–847
38 Janzen, D.H. *et al.* (2005) Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1835–1845
39 Sogin, M.L. *et al.* (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U. S. A.* 103, 12115–12120
40 Rusch, D.B. *et al.* (2007) The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* 5, e77
41 Margurran, A.E. (2004) *Measuring Biological Diversity*. Blackwell Science
42 Blaxter, M. *et al.* (2005) Defining operational taxonomic units using DNA barcode data. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1935–1943
43 Herrera, A. *et al.* (2007) Species richness and phylogenetic diversity comparisons of soil microbial communities affected by nickel-mining and revegetation efforts in New Caledonia. *Eur. J. Soil Biol.* 43, 130–139
44 Oline, D.K. (2006) Phylogenetic comparisons of bacterial communities from serpentine and nonserpentine soils. *Appl. Environ. Microbiol.* 72, 6965–6971
45 Gomez-Alvarez, V. *et al.* (2007) Comparative bacterial diversity in recent Hawaiian volcanic deposits of different ages. *FEMS Microbiol. Ecol.* 60, 60–73
46 Blackwood, C.B. *et al.* (2007) Interpreting ecological diversity indices applied to terminal restriction fragment length polymorphism data: insights from simulated microbial communities. *Appl. Environ. Microbiol.* 73, 5276–5283
47 Meyer, C.P. and Paulay, G. (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.* 3, e422
48 Matz, M.V. and Nielsen, R. (2005) A likelihood ratio test for species membership based on DNA sequence data. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1969–1974
49 Nielsen, R. and Matz, M. (2006) Statistical approaches for DNA barcoding. *Syst. Biol.* 55, 162–169
50 Hickerson, M.J. *et al.* (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. *Syst. Biol.* 55, 729–739
51 Willerslev, E. *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300, 791–795
52 Willerslev, E. *et al.* (2007) Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* 317, 111–114
53 Kuch, M. *et al.* (2002) Molecular analysis of a 11 700-year-old rodent midden from the Atacama Desert, Chile. *Mol. Ecol.* 11, 913–924
54 Hofreiter, M. *et al.* (2000) A molecular analysis of ground sloth diet through the last glaciation. *Mol. Ecol.* 9, 1975–1984
55 Rollo, F. *et al.* (2002) Otzi's last meals: DNA analysis of the intestinal content of the Neolithic glacier mummy from the Alps. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12594–12599
56 Duffy, J.E. *et al.* (2007) The functional role of biodiversity in ecosystems: incorporating trophic complexity. *Ecol. Lett.* 10, 522–538
57 Marrero, P. *et al.* (2004) Diet of the endemic Madeira Laurel Pigeon Columba trocaz in agricultural and forest areas: implications for conservation. *Bird Conserv. Int.* 14, 165–172
58 Cristobal-Azkarate, J. and Arroyo-Rodriguez, V. (2007) Diet and activity pattern of howler monkeys (*Alouatta palliata*) in Los Tuxtlas, Mexico: effects of habitat fragmentation and implications for conservation. *Am. J. Primatol.* 69, 1013–1029
59 Agusti, N. *et al.* (2003) Collembola as alternative prey sustaining spiders in arable ecosystems: prey detection within predators using molecular markers. *Mol. Ecol.* 12, 3467–3475
60 Passmore, A.J. *et al.* (2006) DNA as a dietary biomarker in Antarctic krill, *Euphausia superba*. *Mar. Biotechnol.* 8, 686–696
61 Deagle, B.E. *et al.* (2005) Genetic screening for prey in the gut contents from a giant squid (Architeuthis sp.). *J. Hered.* 96, 417–423

62 Blankenship, L.E. and Yayanos, A.A. (2005) Universal primers and PCR of gut contents to study marine invertebrate diets. *Mol. Ecol.* 14, 891–899
63 Pons, J. (2006) DNA-based identification of preys from non-destructive, total DNA extractions of predators using arthropod universal primers. *Mol. Ecol. Notes* 6, 623–626
64 Symondson, W.O.C. (2002) Molecular identification of prey in predator diets. *Mol. Ecol.* 11, 627–641
65 Jarman, S.N. *et al.* (2004) Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. *Mol. Ecol.* 13, 1313–1322
66 Bradley, B.J. *et al.* (2007) Plant DNA sequences from feces: potential means for assessing diets of wild primates. *Am. J. Primatol.* 69, 699–705
67 Nystrom, J. *et al.* (2006) Effect of local prey availability on gyrfalcon diet: DNA analysis on ptarmigan remains at nest sitesm. *J. Zool. (Lond.)* 269, 57–64
68 Deagle, B.E. *et al.* (2007) Studying seabird diet through genetic analysis of feces: a case study on Macaroni penguins (*Eudyptes chrysolophus*). *PLoS ONE* 2, e831
69 Taberlet, P. *et al.* (2007) Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 35, e14
70 Valentini, A. *et al.* New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trn*L approach. *Mol. Ecol. Res*. (in press)
71 Ballard, J.W.O. and Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Mol. Ecol.* 13, 729–744
72 Zhang, D.X. and Hewitt, G.M. (1996) Highly conserved nuclear copies of the mitochondrial control region in the desert locust *Schistocerca gregaria*: some implications for population studies. *Mol. Ecol.* 5, 295–300
73 Kmiec, B. *et al.* (2006) Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Curr. Genet*. 50, 149–159
74 Vences, M. *et al.* (2005) Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1859–1868
75 Deagle, B.E. *et al.* (2006) Quantification of damage in DNA recovered from highly degraded samples – a case study on DNA in faeces. *Front. Zool.* 3, 11
76 Hajibabaei, M. *et al.* (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Mol. Ecol. Notes* 6, 959–964
77 Meusnier, I. *et al.* (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* 9, 214
78 Hajibabaei, M. *et al.* (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet*. 23, 167–172
79 Illumina. http://www.solexa.com
80 Applied Biosystems. http://solid.appliedbiosystems.com
81 Helicos. http://helicosbio.com
82 Roche Diagnostics. https://www.roche-applied-science.com/sis/sequencing/flx/index.jsp
83 Manel, S. *et al.* (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends Ecol. Evol.* 20, 136–142
84 Zinger, L. *et al.* (2008) CE-SSCP and CE-FLA, simple and high-throughput alternatives for fungal diversity studies. *J. Microbiol. Methods* 72, 42–53
85 Hudson, M.E. (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Res.* 8, 3–17
86 Schuster, S.C. (2008) Next-generation sequencing transforms today's biology. *Nat. Methods* 5, 16–18
87 Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380