Opinion TRENDS in Genetics Vol.xxx No.x

Full text provided by www.sciencedirect.com ScienceDirect

DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics

Mehrdad Hajibabaei[1], Gregory A.C. Singer[2], Paul D.N. Hebert[1] and Donal A. Hickey[3]

[1] Biodiversity Institute of Ontario, Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada
[2] Human Cancer Genetics Program, The Ohio State University, Columbus, OH 43210, USA
[3] Department of Biology, Concordia University, 7141 Sherbrooke Street, Montreal, Quebec H4B 1R6, Canada

DNA barcoding aims to provide an efficient method for species-level identifications and, as such, will contribute powerfully to taxonomic and biodiversity research. As the number of DNA barcode sequences accumulates, however, these data will also provide a unique 'horizontal' genomics perspective with broad implications. For example, here we compare the goals and methods of DNA barcoding with those of molecular phylogenetics and population genetics, and suggest that DNA barcoding can complement current research in these areas by providing background information that will be helpful in the selection of taxa for further analyses.

Introduction

Because of advances in sequencing and computational technologies, DNA sequences have become the major source of new information for advancing our understanding of evolutionary and genetic relationships. The footprints of comparative sequence analysis are now apparent in almost all areas of the biological sciences, from development to epidemiology [1]. However, two branches of biology have developed the tools and applications employed to assess biological relationships with DNA sequences: molecular phylogenetics, and population genetics. These disciplines focus on different levels of organization. Studies in molecular phylogenetics typically deal with evolutionary relationships among deeper clades, whereas those in population genetics target variation within and among populations of a single species. By comparison, DNA barcoding occupies a middle ground as it seeks comprehensive coverage for species, but focuses on their delineation rather than their relationships (Figure 1).

DNA barcoding is based on the premise that a short standardized sequence can distinguish individuals of a species because genetic variation between species exceeds that within species [2]. Pilot projects have now established the effectiveness of this approach in several large groups of animals, such as birds [3], fish [4], cowries [5], spiders [6], and several arrays of Lepidoptera [7–9]. In addition, DNA barcoding systems are now being established for other groups of organisms, including plants [10], macroalgae [11], fungi [12], protists [13] and bacteria [14]. DNA barcoding datasets are essentially composed of short DNA sequences from several individuals of a large number of species (typically five to ten individuals per species, but these numbers will increase in the future) (Figure 1).

Here, we discuss the role of DNA barcodes in advancing the taxonomic enterprise and its potential to provide a contextual framework for both building phylogenies and for population genetics. In particular, we argue that barcode results can be of high value in aiding the selection of species for more detailed analysis, and demonstrate that DNA barcoding can broaden our understanding of both phylogenetic signal and population-level variation.

The DNA barcoding workflow

Species identification through barcoding is usually achieved by the retrieval of a short DNA sequence – the 'barcode' – from a standard part of the genome (i.e. a specific gene region) from the specimen under investigation. The barcode sequence from each unknown specimen is then compared with a library of reference barcode sequences derived from individuals of known identity (Figure 2). A specimen is identified if its sequence closely matches one in the barcode library. Otherwise, the new record can lead to a novel barcode sequence for a given species (i.e. a new haplotype or geographical variant), or it can suggest the existence of a newly encountered species (see below).

Various gene regions have been employed for species-level biosystematics (Table 1); however, DNA barcoding advocates the adoption of a 'global standard', and a 650-base fragment of the 5′ end of the mitochondrial gene cytochrome *c* oxidase I (COI, *cox1*) [2] has gained designation as the barcode region for animals. This fragment

Glossary

Clade: a group of organisms descended from a common evolutionary ancestor; a branch of the Tree of Life.
Homoplasy: a shared character that was not inherited from a common evolutionary ancestor but, rather, that arose through convergent evolution.
Phylogeny: the evolutionary history of species relationships, often visualized as phylogenetic tree (e.g. the Tree of Life).
Taxon: a group of organisms.
Taxonomy: the science of classification. Biological taxonomy creates a hierarchical classification of biological taxa, usually reflecting evolutionary relationships.

Corresponding author: Hajibabaei, M. (mhajibab@uoguelph.ca).
Available online xxxxxx.

www.sciencedirect.com 0168-9525/$ – see front matter © 2007 Elsevier Ltd. All rights reserved. doi:10.1016/j.tig.2007.02.001

Please cite this article in press as: Hajibabaei, M. et al., DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics, Trends Genet. (2007), doi:10.1016/j.tig.2007.02.001

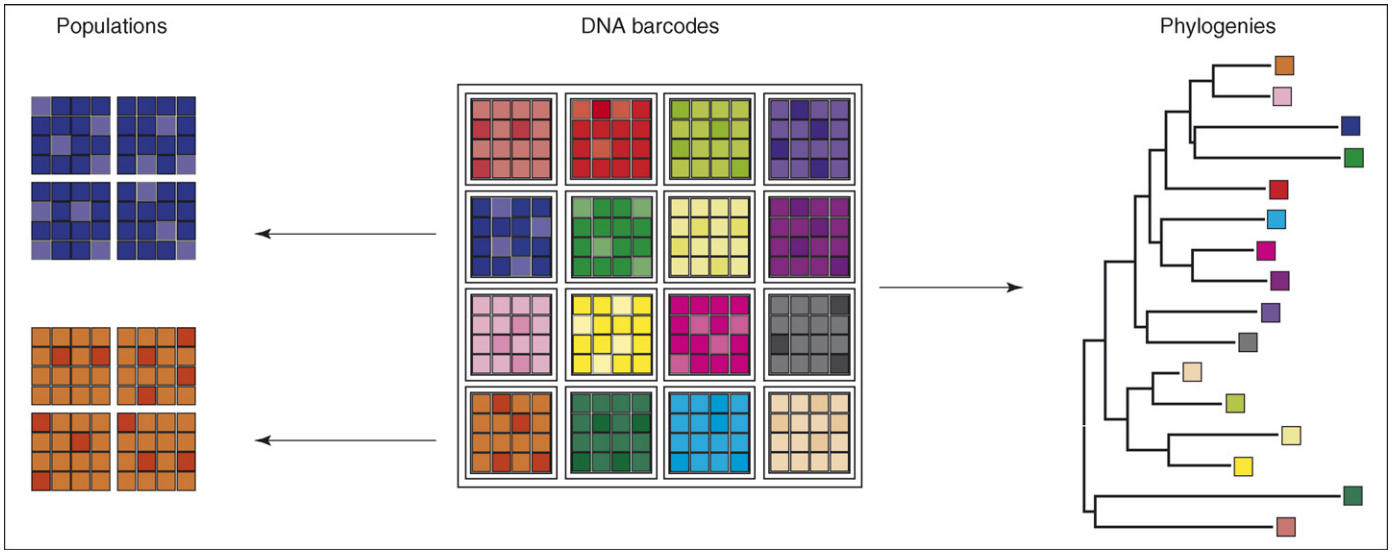

Figure 1. DNA barcodes lie within the gray area between phylogenetics and population genetics. This diagram demonstrates the position of DNA barcode data relative to both population genetics and phylogenetics data. Each small square represents an individual. Different colors are used to represent different species and within-species variation is shown by varying shades of color.

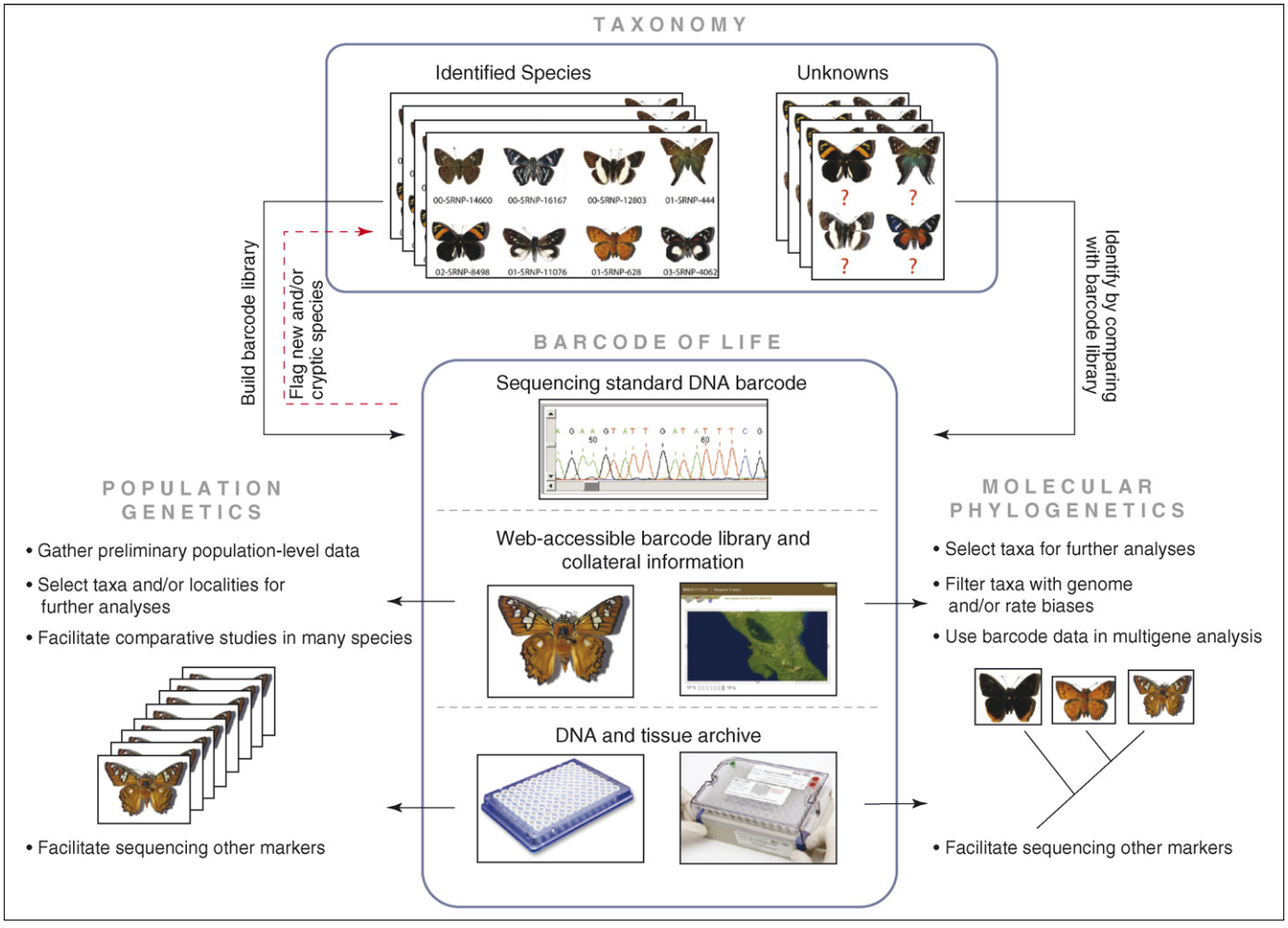

Figure 2. Major components of the Barcode of Life projects and their contribution to taxonomy, reconstruction of molecular phylogenies and population genetics investigations. This diagram shows how DNA barcoding libraries can support the conventional taxonomic workflow by high-throughput identification of unknown specimens and by helping to draw attention to new and cryptic species. Barcode sequences and collateral data for each specimen are accessible through a global online data base (e.g. BOLD: http://www.barcodinglife.org). This information can be useful in other contexts, such as phylogenetics (Tree of Life projects) and population-level studies. In addition, archival DNA and tissue specimens collected in barcoding projects provide an excellent resource for other investigations. Butterfly images are taken from the database of Daniel Janzen and Winnie Hallwachs (http://janzen.sas.upenn.edu/).

Please cite this article in press as: Hajibabaei, M. et al., DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics, Trends Genet. (2007), doi:10.1016/j.tig.2007.02.001

Table 1. Common species-level molecular markers

Gene[a]	Genomic location	Number of sequences			
		Animals	Plants	Protists	Fungi
COI-barcode[b]	Mitochondria	195 777	520	1931	410
16S-rDNA	Mitochondria	41 381	221	2059	285
cytb	Mitochondria	88 324	165	1920	1084
ITS1-rDNA	Nucleus	12 175	57 693	68 839	56 675
ITS2-rDNA	Nucleus	13 923	58 065	67 332	56 349
18S-rDNA	Nucleus	21 063	17 121	32 290	33 327
rbcL	Plastid	NA[c]	30 663	37 328	NA

[a]Gene abbreviations: COI, cytochrome *c* oxidase I; *cytb*, cytochrome *b*; ITS, internal transcribed spacer; *rbcL*, large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase.
[b]COI-barcode statistics are retrieved from Barcode of Life Data systems (http://www.barcodinglife.org). Statistics for other loci are retrieved from GenBank.
[c]NA, not applicable.

size has been selected so that a reliable sequence read can be obtained by a single sequence pass in conventional cycle-sequencing platforms. Shorter fragments of COI have also been shown to be effective for the identification of specimens with degraded DNA, however, where a 650-base sequence is not easily obtainable [15]. In addition, the usability and robustness of COI in a standard high-throughput barcoding analysis have been extensively assessed [16].

Other researchers have suggested that alternate loci might also serve as a basis for species identification. For example, 18S rDNA has been used for the identification of soil nematodes and other small organisms in an approach known as 'DNA taxonomy' [17]. This approach differs from DNA barcoding in that it does not aim to link the genetic entities recognised through sequence analysis with Linnaean species. As such, it is most useful for groups of organisms that lack detailed taxonomic systems. Alternate markers have also been used where COI sequences have not been produced robustly or are shown to be divergent within species [18] or as further molecular evidence in the discovery of cryptic species [19]. Moreover, in some groups such as plants, COI (and mitochondrial genomes at large) do not evolve rapidly enough to provide species-level resolution, and alternative markers are being pursued [10,20].

Several studies have demonstrated the effectiveness of DNA barcoding in different animal groups [2–4,7,8,19,21]. These projects have shown that >95% of species possess unique COI barcode sequences; thus species-level identifications are regularly attained. The earliest barcode studies received some criticism, mainly owing to their limited taxonomic and geographical breadth [22]; however, more recent studies have addressed these issues by targeting species-rich groups (i.e. those containing many closely related species) in tropical settings [8], and by comprehensive analyses of all the species in a given taxonomic assemblage [5]. Momentum has further been aided by establishment of the Consortium for the Barcode of Life (CBOL, http://barcoding.si.edu) – an international alliance of research organizations that support the development of DNA barcoding as an international standard for species identification [23] – and by development of the Barcode of Life Data Systems (http://www.barcodinglife.org) – a global online data management system for DNA barcodes [24].

Barcoding projects typically involve gathering specimens of a given taxonomic group (identified by conventional taxonomic methods such as morphology; see below), cataloging them together with collateral data such as photographs and locality information, and assembling the barcode library (i.e. a 650-base segment of the COI gene) [16]. The analysis of DNA barcoding data is usually performed by a clustering method, such as distance-based neighbor-joining (NJ) [25], and by evaluating genetic distances within and between species (e.g. [8]). More complex methodologies for data analysis are under development, including statistical tests for species assignment [26,27], and character-based clustering methods [28].

DNA barcoding and taxonomy

Although its role in identifying specimens to a species level is an important aid for taxonomic workflow (Figure 2), barcoding is no replacement for comprehensive taxonomic analysis. For example, when an unknown specimen does not return a close match to existing records in the barcode library, the barcode sequence does not qualify the unknown specimen for designation as a new species. Instead, such specimens are flagged for thorough taxonomic analysis. When viewed in the context of the traditional taxonomic framework – which is usually much slower than barcoding analysis – this flagging of atypical specimens has much potential to aid species discovery [8]. In poorly studied taxonomic groups, DNA barcoding can be performed before conventional taxonomic work to quickly sort specimens into genetically divergent groups. For example, Smith *et al.* [29] used DNA barcoding to rapidly assess the biodiversity of ants in Madagascar and it is now a routine element of a large-scale biodiversity inventory on these organisms.

DNA barcoding has also been used in well-studied groups such as Lepidoptera [9]. For example, barcoding is now used routinely to understand the biodiversity of the caterpillar fauna in northwestern Costa Rica. This project, which was started over 25 years ago by Janzen, Hallwachs and colleagues [9,30,31], has, over the past three years, employed barcoding to create a reference sequence library for more than 25 000 specimens representing >2000 species of moths and butterflies and their parasitoids [7–9,19]. This library is now used to speed up the sorting and identification of specimens in concert with the traditional taxonomic workflow.

Although the task of identifying and describing new species is ultimately achieved through comprehensive taxonomic work, DNA barcodes can significantly facilitate this process. The conventional taxonomic workflow, which usually requires the collection of morphological and ecological data, can vary for different taxonomic assemblages (i.e. taxonomic identification of birds and fish require

Please cite this article in press as: Hajibabaei, M. et al., DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics, Trends Genet. (2007), doi:10.1016/j.tig.2007.02.001

different methods and skills), whereas barcode analysis can be applied in a more or less standardized way across large domains of life (i.e. all animal taxa).

DNA barcoding and molecular phylogenetics

The molecular phylogenetics workflow

The incorporation of sequence data has revolutionized phylogenetics over the past two decades [32] and an increasing number of large-scale projects are underway to resolve different branches of the Tree of Life (http://tolweb.org). A typical molecular phylogenetics project involves a primary decision in relation to the target group for analysis (e.g. family), the assembly of representative taxa, the acquisition of sequence information, and the construction of phylogenetic trees by using optimality criteria such as Maximum Likelihood, Maximum Parsimony, or Bayesian analysis. Although the details of these procedures have been discussed extensively elsewhere [33], it is important to emphasize that care must be exercised in the selection of both loci and representative taxa to optimize the recovery of a strongly supported phylogenetic tree. Until recently, gene selection was largely restricted to easy-to-sequence and universal targets such as ribosomal genes. However, recent technological advances mean that much more strategic thinking is possible. Consequently, most recent phylogenetic analyses use sequence information from multiple loci (covering several kilobases), often from different genomic compartments (i.e. nucleus, mitochondrion and chloroplast) to enhance resolution at different taxonomic levels [34] and to avoid gene-specific biases [35]. With the advance of whole-genome sequencing projects, some researchers have even advocated the use of entire genomic sequences for phylogenetic inference [36].

Another important consideration, taxon sampling, has received rather less attention than the choice of loci. It is generally recognised that increasing the number of taxa aids recovery of the correct phylogeny (see Glossary) by reducing branch lengths and homoplasy (see Glossary), both factors that can produce misleading phylogenies [37,38]. In fact, some argue that adding additional taxa is more valuable than adding more genes to improve the resolution of a phylogeny [39,40]. Although there is no doubt that adding more taxa is helpful, it is also important to select these taxa with care, given that computational factors can limit the number of taxa that can be analyzed. Because of this, researchers have relied on heuristics and simplified analytical methods when dealing with phylogenies with large number of taxa (i.e. hundreds of species). The examination of multiple individuals from a single species, or the analysis of large sets of closely allied species, remains uncommon.

From barcodes to phylogenies

While barcode libraries have similarities to molecular phylogenetic data (both are sequence information from assemblages of species), DNA barcodes do not usually have sufficient phylogenetic signal to resolve evolutionary relationships, especially at deeper levels [41]. Although barcode sequences have been analyzed mainly by using phylogenetic tree reconstruction methods such as NJ, these barcode-based trees should not be interpreted as phylogenetic trees [41]. DNA barcoding projects can aid the construction of phylogenies by aiding the selection of taxa (Figures 1,2), however. Because DNA barcodes are used both to identify species and to draw attention to overlooked and new species, they can help identify candidate exemplar taxa for a comprehensive phylogenetic study (Figure 1). Barcode of Life projects create a perfect taxonomic sampling environment for conducting phylogenetic studies on different branches of the Tree of Life (Figure 2); and, as mentioned earlier, this sampling has been shown to be a key factor in obtaining a robust phylogeny [39,40]. Consequently, phylogenies that are constructed on top of barcode libraries, in a given taxonomic group, are less likely to be influenced by insufficient taxon sampling. Additionally, barcoding aids in pinpointing and subsequent replacement of taxa with attributes – such as exceptionally elevated rates of evolution or nucleotide compositional biases [33] – that can mislead the recovery of phylogenetic trees.

Barcode sequence data can also provide a shared genomic cornerstone for the variable repertoire of genes that can be used to build the phylogenetic tree. It can be used as a link between the deeper branches of the tree to its shallow, species-level branches.

DNA barcoding and population genetics

The population genetics workflow

Traditional analytical approaches, such as patterns of allozyme or restriction enzyme polymorphisms, have now largely been replaced by sequence-based analyses. However, the selection of an appropriate marker system for a population genetics survey requires careful consideration of issues such as sensitivity for the questions being asked and practical measures for obtaining the information (i.e. ease of amplification by PCR). Because mitochondrial DNA markers are haploid and uniparentally inherited, they are frequent targets for analysis and have made a particularly strong contribution to population-level studies [42]. However, multigene (multilocus) analyses have recently gained credibility because they are less sensitive to specific gene genealogies that might not correctly reflect population history [43]. More recently, with the availability of high-throughput sequencing technology, fine-scale sequence analysis methods, such as single-nucleotide polymorphism (SNP), have given a new dimension to population-level studies [44]. Large-scale analysis of these sequence-based markers are now underway in the framework of projects such as HapMap (http://www.hapmap.org), which aims to identify the genetic variation associated with human diseases [45].

Although a vast number of genetic markers have been used to study population structure, the number of species studied is far fewer. Typical population genetics studies examine variation within populations of a single species, and this sort of information has been successfully applied to geographical studies of populations, to investigate issues such as migration and genetic drift. However, few taxonomic groups have been studied extensively with DNA sequence data. For example, a GenBank search suggests that human, chimpanzee and a few other model organisms such as mouse and *Drosophila melanogaster* account for

Please cite this article in press as: Hajibabaei, M. et al., DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics, Trends Genet. (2007), doi:10.1016/j.tig.2007.02.001

more than half of all sequence-based population genetics studies (data not shown).

From barcodes to population analysis

Although the typical sequence information gathered for DNA barcoding is not sufficient to rigorously address population-level questions [22,46], it can provide an early insight into the patterning of genomic diversity within a species. Because barcoding typically targets a large number of species, it can be a powerful tool to facilitate comparative studies of genetic diversity in different species or ecological settings (Figure 2). Models of population genetics, such as coalescent-based models, have been proposed for the assignment of individuals to species in DNA barcode analysis [26,27]. It would be interesting to investigate further the usefulness of barcode data (and the effect of sample size) in the study of genetic diversity of species. Conventional measures such as haplotype diversity or Watterson's theta [47] can potentially provide useful information in this regard. Moreover, with the accumulation of large amount of barcode sequences and collateral information such as ecological and geographical information, it is important to investigate the type of questions that can be addressed through barcode data. For example, habitat continuity and genetic structure in species, within a taxonomic assemblage or geographical region, can potentially be studied by expanding the sampling of both individuals and genetic markers.

The strategic value of DNA and tissue archiving

DNA barcoding requires the assembly of tissue samples and the subsequent isolation and archiving of genomic DNA. These archived samples can act as useful resource for phylogenetics, population genetics and phylogeographic studies. Sampling, vouchering and DNA isolation constitute a substantial cost for any molecular study. Recent technologies such as whole-genome amplification [48], or sequencing by synthesis [49], which employ small amounts of genetic material for analysis, can potentially facilitate even broader genomics applications for the genetic material gathered in DNA barcode studies.

Concluding remarks

In summary, DNA barcoding is poised to contribute to taxonomic research and to population genetics and phylogenetics. In taxonomy, DNA barcoding can be used for routine identification of specimens; and it can also flag atypical specimens for comprehensive taxonomic investigation. In phylogenetic studies, DNA barcoding can be a starting point for optimal selection of taxa, and barcode sequences can be added to the sequence dataset for phylogenetic analysis. In population genetics investigations, DNA barcodes can provide a first signal of the extent and nature of population divergences and will facilitate comparative studies of population diversity in many species.

On the basis of recent developments, we expect that the barcode databases will grow rapidly – some facilities are already processing >100 000 specimens per year. Consequently, the International Nucleotide Sequence Database (INSD: GenBank, EMBL, and DDBJ) has adopted a unique keyword identifier (BARCODE) to recognise standard barcode sequences specified by the scientific community (i.e. CBOL).

The introduction of DNA barcoding is a natural addition to the post-genomic era, in which whole-genome sequencing has provided a vast amount of sequence information from a limited number of species. DNA barcodes can help expand our knowledge by exploring many more species rapidly and inexpensively. The results obtained from DNA barcoding studies can then help us identify species that are good targets for more detailed genetic analyses.

Acknowledgements

Funding for this study was provided by Genome Canada through the Ontario Genomics Institute, and by the Gordon and Betty Moore Foundation. We thank the reviewers for their comments on an earlier version of this manuscript.

References

1 Tibayrenc, M. (2005) Bridging the gap between molecular epidemiologists and evolutionists. *Trends Microbiol.* 13, 575–580
2 Hebert, P.D.N. *et al.* (2003) Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B. Biol. Sci.* 270, 313–321
3 Hebert, P.D.N. *et al.* (2004) Identification of birds through DNA barcodes. *PLoS Biol.* 2, e312
4 Ward, R.D. *et al.* (2005) DNA barcoding Australia's fish species. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1847–1857
5 Meyer, C.P. and Paulay, G. (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.* 3, e422
6 Barrett, R.D.H. and Hebert, P.D.N. (2005) Identifying spiders through DNA barcodes. *Can. J. Zool.* 83, 481–491
7 Hebert, P.D.N. *et al.* (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14812–14817
8 Hajibabaei, M. *et al.* (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. U. S. A.* 103, 968–971
9 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
10 Kress, W.J. *et al.* (2005) Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8369–8374
11 Saunders, G.W. (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1879–1888
12 Summerbell, R.C. *et al.* (2005) Microcoding: the second step in DNA barcoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1897–1903
13 Scicluna, S.M. *et al.* (2006) DNA barcoding of blastocystis. *Protist* 157, 77–85
14 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
15 Hajibabaei, M. *et al.* (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Mol. Ecol. Notes* 6, 959–964
16 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
17 Blaxter, M.L. (2004) The promise of a DNA taxonomy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 669–679
18 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
19 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
20 Rubinoff, D. *et al.* (2006) Are plant DNA barcodes a search for the Holy Grail? *Trends Ecol. Evol.* 21, 1–2
21 Hebert, P.D.N. *et al.* (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B. Biol. Sci.* 270 (Suppl. 1), S96–S99
22 Moritz, C. and Cicero, C. (2004) DNA barcoding: promise and pitfalls. *PLoS Biol.* 2, e312

Please cite this article in press as: Hajibabaei, M. et al., DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics, Trends Genet. (2007), doi:10.1016/j.tig.2007.02.001

23 Marshall, E. (2005) Taxonomy. Will DNA bar codes breathe life into classification? *Science* 307, 1037
24 Ratnasingham, S. and Hebert, P.D.N. (2007) The Barcode of Life Data Systems (www.barcodinglife.org). *Mol. Ecol. Notes* DOI:10.1111/j.1471-8286.2007.01678.x (www.blackwell-synergy.com)
25 Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425
26 Nielsen, R. and Matz, M. (2006) Statistical approaches for DNA barcoding. *Syst. Biol.* 55, 162–169
27 Abdo, Z. and Golding, G.B. (2007) A step toward barcoding life: a model-based, decision theoretic method to assign genes to pre-existing species groups. *Systematic Biology* 56, 1–13
28 DeSalle, R. *et al.* (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1905–1916
29 Smith, M.A. *et al.* (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1825–1834
30 Janzen, D.H. (2004) Setting up tropical biodiversity for conservation through non-damaging use: participation by parataxonomists. *J. Appl. Ecol.* 41, 181–187
31 Janzen, D.H. (2003) Spatio-temporal dynamics and resource use in the canopy. In *Arthropods of Tropical Forests* (Basset, Y., Novotny, V., Miller, S.E. and Kitching, R.L., eds), pp. 369–379, Cambridge University Press
32 Pagel, M. (1999) Inferring the historical patterns of biological evolution. *Nature* 401, 877–884
33 Felsenstein, J. (2004) *Inferring Phylogenies,* Sinauer Associates Inc.
34 Hajibabaei, M. *et al.* (2006) Seed plant phylogeny: gnetophytes are derived conifers and a sister group to Pinaceae. *Mol. Phylogenet. Evol.* 40, 208–217
35 Foster, P.G. and Hickey, D.A. (1999) Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J. Mol. Evol.* 48, 284–290
36 Murphy, W.J. *et al.* (2004) Mammalian phylogenomics comes of age. *Trends Genet.* 20, 631–639
37 Huelsenbeck, J.P. (1995) The performance of phylogenetic methods in simulation. *Syst. Biol.* 44, 17–48
38 Kolaczkowski, B. and Thornton, J.W. (2004) Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature* 431, 980–984
39 Zwickl, D.J. and Hillis, D.M. (2002) Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* 51, 588–598
40 Pollock, D.D. *et al.* (2002) Increased taxon sampling is advantageous for phylogenetic inference. *Syst. Biol.* 51, 664–671
41 Hajibabaei, M. *et al.* (2006) Benchmarking DNA barcodes: an assessment using available primate sequences. *Genome* 49, 851–854
42 Avise, J.C. (2004) *Molecular Markers, Natural History and Evolution,* Chapman and Hall
43 Rosenberg, N.A. and Nordborg, M. (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* 3, 380–390
44 Brumfield, R.T. *et al.* (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol. Evol.* 18, 249–256
45 International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437, 1299–1320
46 Bazin, E. *et al.* (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* 312, 570–572
47 Li, W.H. (1997) DNA polymorphism in populations. In *Molecular Evolution* (Li, W.H., ed.), pp. 237–267, Sinauer Associates Inc.
48 Hawkins, T.L. *et al.* (2002) Whole genome amplification – applications and advances. *Curr. Opin. Biotechnol.* 13, 65–67
49 Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380

Please cite this article in press as: Hajibabaei, M. et al., DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics, Trends Genet. (2007), doi:10.1016/j.tig.2007.02.001