MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 1423–1428

NEWS AND VIEWS

OPINION

From barcodes to genomes: extending the concept of DNA barcoding

ERIC COISSAC,*† PETER M. HOLLINGSWORTH,‡ SÉBASTIEN LAVERGNE*† and PIERRE TABERLET*†

CNRS, LECA, F-38000 Grenoble, France; †Univ. Grenoble Alpes[3bis], LECA, F-38000 Grenoble, France; ‡Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, Scotland

Abstract

DNA barcoding has had a major impact on biodiversity science. The elegant simplicity of establishing massive scale databases for a few barcode loci is continuing to change our understanding of species diversity patterns, and continues to enhance human abilities to distinguish among species. Capitalizing on the developments of next generation sequencing technologies and decreasing costs of genome sequencing, there is now the opportunity for the DNA barcoding concept to be extended to new kinds of genomic data. We illustrate the benefits and capacity to do this, and also note the constraints and barriers to overcome before it is truly scalable. We advocate a twin track approach: (i) continuation and acceleration of global efforts to build the DNA barcode reference library of life on earth using standard DNA barcodes *and* (ii) active development and application of extended DNA barcodes using genome skimming to augment the standard barcoding approach.

Keywords: chloroplast DNA, DNA Barcoding, genome skimming, mitochondrial DNA, next-generation DNA sequencing, ribosomal DNA

Received 27 August 2015; revision received 28 December 2015; accepted 19 January 2016

Introduction

DNA barcoding involves sequencing one or a few standard DNA regions to tell the world's species apart. Since its inception in 2003, DNA barcoding has grown into a global research programme involving thousands of researchers whose work has led to the production of millions of barcode sequences. In this opinion paper, we explore the potential for the standard DNA barcoding approach to be complemented and extended using next generation sequencing technologies. Focusing on plants, we highlight the use of shallow-pass shotgun sequencing of genomic DNA in large-scale projects to generate extended barcodes consisting of entire organelle genome and nuclear ribosomal DNAs, along with shallow coverage of single copy nuclear DNA. These extended barcodes are recoverable from herbarium/museum specimens, provide increased phylogenetic signal, and provide a bridge between standard and metabarcoding studies which often use different target regions. They also represent a stepping stone on the continuum between standard barcodes and complete genome sequences, and as sequencing costs decrease, the depth of the skims can increase, resulting in ever increasing data richness. We note the potential for this approach to augment the standard DNA barcoding programme, and although we focus on plants, we highlight its applicability across the three domains of life. We also explore the challenges that arise from incorporating genomic data into high throughput barcoding workflows, particularly the currently higher consumable costs and greatly increased demands on data storage and analytical routines.

Correspondence: Eric Coissac, Fax: +33(0)4 76 51 42 79; E-mail: eric.coissac@inria.fr

Standard DNA barcodes

The DNA barcoding concept proposed by Hebert *et al.* (2003a) represented a major step forward for the DNA-based species identification. The approach harnessed global community efforts to establish large-scale public reference libraries to allow reliable identification of species across vast tracts of life. For animals, the standard barcode is a 648 base pairs (bp) fragment of the mitochondrial gene cytochrome *c* oxidase 1 (COI; Hebert *et al.* 2003b). The use of COI for species identification and discovery has been extremely successful for the animal kingdom, and the BARCODE OF LIFE DATASYSTEMS database (BOLD) contains now more than 4.2 million validated barcodes (http://www.boldsystems.org/index.php/databases; Ratnasingham & Hebert 2007).

In plants, the choice of the standardized barcode(s) has been more complex. The low substitution rates of plant mitochondrial DNA (Wolfe *et al.* 1987) precluded the use of COI. As a consequence, alternative barcoding regions were investigated, leading to selection of two plastid DNA regions, a *c.* 600 bp fragment of the *rbc*L gene, and *c.* 800 bp segment of the *mat*K gene, with the recommendation to complement these using *trn*H-*psb*A (Hollingsworth *et al.* 2009) and the internal transcribed spacers (ITS) of the nuclear ribosomal DNA (Hollingsworth 2011; Hollingsworth *et al.* 2011; Li *et al.* 2011). The same ITS region has also been suggested as the core barcode region for fungi (Schoch *et al.* 2012). Finally, for protists, a two-step barcod-

ing strategy is advocated, with first involving analysis of the V4 region of the 18S ribosomal DNA as a pre-barcode, and then one or several additional barcodes specific to the different protist clades (Pawlowski *et al.* 2012).

Power and limitations of the current DNA barcodes

The great strengths of DNA barcoding relate to the initial principles of standardization ('agreed' regions of DNA so that joint efforts build a shared global resource), quality control (to ensure the library of DNA sequences is reliable) and minimalism (using one or a few regions of DNA to ensure scalability). Despite some initial criticisms (e.g. Will & Rubinoff 2004), the DNA barcoding concept has been widely accepted and has had a great influence on the scientific community. A search for 'NA barcod*' in the Web of Science on 02 December 2015 produced 12 235 hits, leading to 115 050 citations.

In most animal groups, there is an excellent concordance between barcode sequence clusters and known species (e.g. Meier *et al.* 2006). This concordance between taxonomic frameworks and the shape of sequence discontinuities in COI allows DNA barcoding to function very effectively for both species discovery and species identification (but see Hickerson *et al.* 2006; Elias *et al.* 2007). In plants, the situation is more difficult. In part, this is due to lower variation in plant plastid DNA than animal mtDNA, and, in part, due to a greater propensity for hybridization among related plant species (Hollingsworth *et al.* 2011). There is thus an ongoing drive to find ways of increasing levels of plant barcode discrimination. In addition, despite the intrinsic qualities of the standard barcodes, there is something of a divide in the 'DNA region of choice' between specimen-based barcoding and metabarcoding studies because of methodological constraints. The latter often uses alternative organelle or nuclear rDNA mini-barcodes that are better suited to recovery from degraded DNA and mixed templates (e.g. Baldwin *et al.* 2013; Clarke *et al.* 2014; Deagle *et al.* 2014; Kartzinel *et al.* 2015). Collectively, these challenges mean that the search for improved DNA barcoding protocols is still ongoing.

Genome skimming as a universal 'extended barcode'

Low-coverage shotgun sequencing of genomic DNA

One approach which offers a relatively straightforward mechanism to improve and extend DNA barcodes is genome skimming (Straub *et al.* 2012; Dodsworth 2015). For plants, low-coverage shotgun sequencing of genomic DNA (= genome skimming) based on about one gb of genomic sequences can recover complete sequences of plastid genomes and nuclear ribosomal regions for 48 samples loaded in the same HiSeq 2500 lane. This approach recovers simultaneously all of the different 'standard' plant barcoding regions, while also providing sequence data from many other loci (Besnard *et al.* 2014a), and provide a direct link with all other phylogenetically informative genomic regions (Fig. 1). The rapidly decreasing costs of sequencing consumables is now moving this approach from the realms of small scale application to a handful of samples (e.g. Nock *et al.* 2011; Kane *et al.* 2012), to a feasible methodology scalable to 1000s of samples and hence become a realistic proposition for large-scale barcoding projects. For instance, the 'PhyloAlps project' (https://www.france-genomique.org/spip/spip.php?article112), is producing genome skims consisting of six million 100 bp Illumina reads for 6000 samples covering the whole of the Alpine flora. Likewise, 'NorBol' (the Norwegian initiative for the Barcoding of Life) is implementing genome skimming for 3000 specimens of vascular plants covering the arctic-boreal flora.

The power of the extended barcode

The extended barcode outlined above has the potential to overcome the limitations of some traditional barcodes. First, the genome skimming approach circumvents the need for PCR. For plants, the absence of a PCR stage means that the sequence data is recoverable even when using herbarium specimens containing degraded DNA, enhancing the possibility of recovering data from type specimens. The whole plastid genome has already been assembled using a 100-year old herbarium specimen of an extinct plant species (Besnard *et al.* 2014a). This absence of PCR also has the potential to address recovery problems in groups where universal primers are ineffective (e.g. *matK* from various plant lineages or plastid regions from many parasitic plants).

Second, the additional data greatly increase the phylogenetic signal in the barcoding data set, enabling a single data set to work effectively for species discrimination *and* for assessing phylogenetic relationships. Third, by generating whole plastid genomes and ribosomal sequences, the problem of researchers preferring different loci for some specimen-based and metabarcoding based studies is circumvented, as all relevant loci are routinely recovered. Fourth, the additional sequence data from completely sequenced plastid genomes and ribosomal repeats should lead to some increase in levels of species discrimination in situations where a shortage of variable characters is the limiting factor (Ruhsam *et al.* 2015). Of course, plastid genomes and ribosomal repeats will not address the limitations in discrimination where hybridization, repeated introgressions or recent origins result in plastids/rDNA not matching species boundaries (e.g. Fazekas *et al.* 2008; Percy *et al.* 2014; Ruhsam *et al.* 2015). Finally, the genome skimming approach also generates some low coverage data of single copy nuclear regions. According to the distribution of plant genome sizes extracted from the Kew Plant genome size database (Zonneveld *et al.* 2005), the sequencing effort for an average of six millions 100 bp reads for each species produced by the PhyloAlps project corresponds to a coverage of at least one-fourth of the nuclear genome for half of the genomes (median of the estimated coverage 0.277×). Although the density of these data cur-

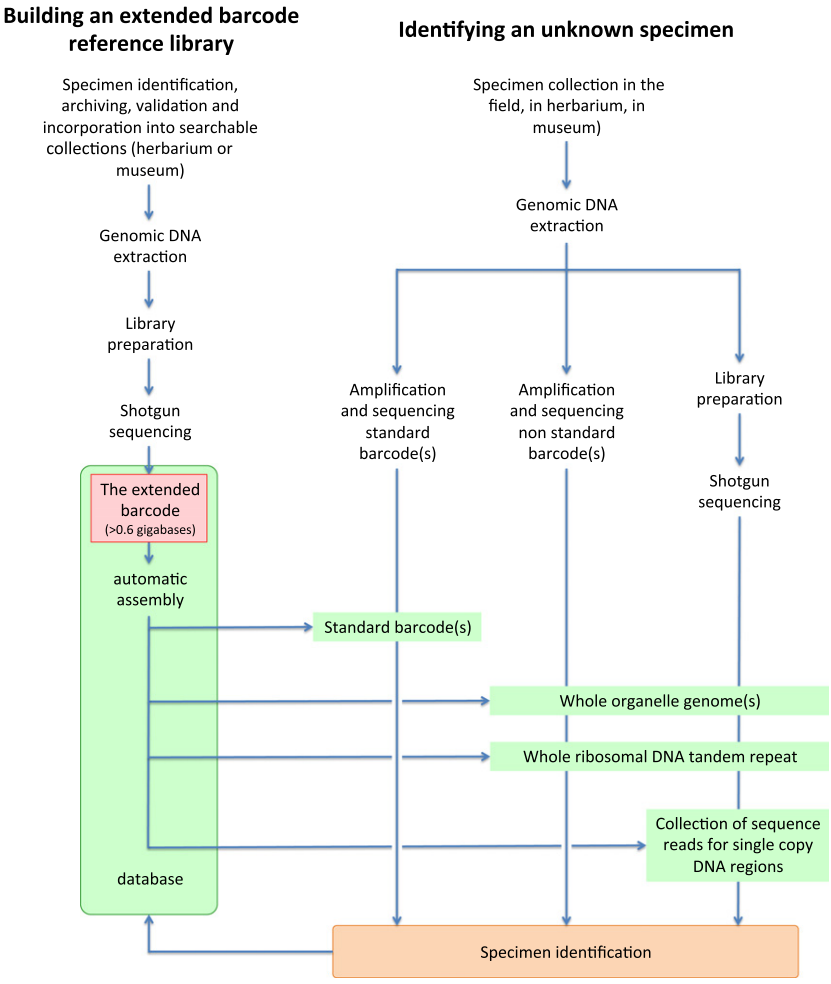

Fig. 1 Overview of the experimental procedures for implementing extended DNA barcoding based on about one gigabase of sequence reads produced by shotgun sequencing of genomic DNA.

rently limits routine recovery of high quality sequences from homologous regions from multiple samples, it does provide initial sequence data which can be used to target further marker development or sequencing of promising loci. Furthermore, these single copy nuclear regions can potentially be used in conjunction with algorithms inspired by Maillet *et al.* (2014) or Fan *et al.* (2015) to generate similarity indices between pairs of nuclear genomes, and to help solve difficult situations involving hybridization and/or recent origins.

Alternative next generation DNA Barcodes

Several genome simplification techniques have the potential to be implemented as alternative species identification tools to tackle the limitations of the standard DNA barcoding method. RAD sequencing (Baird *et al.* 2008) is a commonly used method in population genetics and has already be used to distinguish closely related species (e.g. Hohenlohe *et al.* 2011). RAD sequencing is highly effective at generating sequence data from many thousands of nuclear loci, but the need for taxon-specific optimization precludes its use as a universal barcoding approach.

A related suggestion to the genome skim concept we have outlined is that of Li *et al.* (2015). This proposes using genome skims to increase the density and phylogenetic coverage of complete plastid genome sequences, and to use this to develop 'specific barcodes' – e.g. plastid loci selected for having optimal variation for individual clades of plants. At one level, this approach has some common threads with our proposal (shotgun sequencing to produce plastid genome sequences). However, it differs fundamentally in that it essentially argues for establishing multiple sets of clade-specific barcodes. This may be practical for monographic studies, but this 'taxon-specific approach' will be time consuming to apply in floristic/environmental sample sets, and it steps away from the principle of barcoding: establishing a database centred on standardized loci.

A more widely applicable method is the use of capture probes (Peñalba *et al.* 2014; Nicholls *et al.* 2015). Probe sets are available for the entire plastid genomes of eudicots and monocots and offer a cost-effective approach for obtaining complete organelle genomes via targeted enrichment (Stull *et al.* 2013). Probe sets are being further developed for nuclear markers, and the potential exists for this type of

approach to offer a powerful extended barcode. The rate limiting step remains the development of probe sets with extremely wide phylogenetic coverage applicable across the different domains of life.

Challenges for massive upscaling of the extended barcode

The widespread adoption of genome skimming as an extended barcode will be dependent on the efficacy of its implementation at a very large scale, and the cost implications for consumables, informatics, computational power and data storage. In this respect, although the size of the Phyloalps and NorBol genome skimming projects described above are small compared to the 4.5 million COI sequences present in BOLD, they represent critically important large-scale pilot studies addressing the challenges of upscaling genome skimming.

At current market rates, the consumables cost for sequencing one gigabases is about $80, but before the sequencing step, specific adaptors must be ligated onto each side of the fragmented genomic DNA (= library preparation). The cost of building the library is still relatively high. In the PhyloAlps project, the outsourced market rate for the cost of sequencing, including library preparation is approximately $200 per sample. The development of fully automatized library preparation, either using robots or a microfluidic approach (Kim *et al.* 2013) has the potential to significantly decrease these library preparation costs.

At the bioinformatic level, a large and complex database and an automated workflow must be designed to process and manage this amount of data. The system must include (i) the automation of the assembly of the standard barcodes (*rbc*L, *mat*K, ITS, etc.), (ii) the automation of the assembly of organellar DNA as well as nuclear ribosomal tandem repeats (iii) the automated annotation of the different assembled fragments, (iv) the removal of potential contaminant sequence reads (i.e. reads from DNA of fungi or bacteria co-extracted with the DNA of the target species), (v) the estimation of the sequencing coverage for single copy genes, (vi) the extraction of reads corresponding to single-copy genes such as the Conserved Ortholog Set II (COSII) markers for plants (http://www.sgn.cornell.edu/markers/cosii_markers.pl), (vii) the automated identification of an unknown specimen from either a small shotgun sequencing, or from any DNA fragment and (viii) ultimately, the automated identification of the different organisms from shotgun sequencing of environmental DNA (Taberlet *et al.* 2012). All these bioinformatic developments will probably require the collaboration of several research teams over a few years to design such efficient database and workflow. Some of the needed tools already exist (e.g. Wyman *et al.* 2004; Liu *et al.* 2012; Hahn *et al.* 2013) or are in development. For example, the Organelle Assembler (http://metabarcoding.org/org.asm) developed for the PhyloAlps project is able to automate de novo assemblies of plastid genomes as single circular contigs in less than half an hour on a single core, from c. 70% of the 3000 genome skims already produced by the PhyloAlps project. A final, but important step is to ensure the presence of adequate data storage facilities, (*c.* 1 gb per sample in the PhyloAlps project).

With these challenges in mind, we expect that the extended barcode approach via genome skimming will be initially implemented by well-resourced labs on projects of a few thousand samples. In parallel, conventional plant and animal DNA barcoding will continue to be routinely used. Fortunately (i) all of the extended barcode data sets recover standard barcodes providing 'approach overlap' and (ii) the well-identified specimens that have been used for the standard barcodes can be re-sequenced to produce the extended barcode when costs and practicalities permit.

One nontechnical challenge that remains (and is somewhat difficult to quantify) is whether a move from one or a few barcode loci to genome skimming will result in increased difficulties in obtaining permits or material transfer agreements (MTAs). The restricted bio-functional information in standard barcode sequences has enabled the barcoding enterprise to obtain permission to proceed under MTAs that constrain sequence analysis to the barcode genes. Even shallow-pass genome skims are likely to result in production of data on genes of functional importance, which may lead to greater sensitivities in granting permits for the export of samples or data to third-party countries.

Application to all domains of life

The challenges associated with plant DNA barcoding and phylogeny were a key motivating factor for the PhyloAlps and NorBol projects to move from standard DNA barcodes to extended DNA barcodes. Nevertheless, genome skimming is in principle more widely applicable to all domains of life. Genome or metagenome skimming to obtain complete mitochondrial DNA sequences is being applied in diverse set of animal taxa ranging from nematodes (Besnard *et al.* 2014b), to insects (Linard *et al.* 2015) and bighorn sheep (Miller *et al.* 2012). As of 2 December 2015, Google Scholar reports 168 hits including 97 references for 2015 (search terms: 'genome skimming'). Furthermore, with the availability of single-cell sequencing (see reviews in Lasken & McLean 2014; Liang *et al.* 2014), the extended barcode even becomes possible for unicellular eukaryotes, for bacteria, and for archaea. It can also be implemented for composite organisms such as lichens or corals, with single-cell sequencing for the algae to allow the deconstruction of the different components of these composite organisms. As a consequence, the extended barcode based on shotgun sequencing of genomic DNA has the potential of being universally applied across all domains of life.

The wider genomic context

The ultimate goal of DNA barcoding – is to tell all of the world's species apart. This involves sample sizes of many millions. Given this scale of the task – the crux issue is allocating the minimal sequencing effort per sample to achieve the

goal. Where one or a few barcoding loci will achieve this task, there is no need to sequence more. However, where full species discrimination is not achievable, where other applications are also of interest (phylogenetics, population genetics), or where the barcode loci are not suited for recovery from degraded DNAs, more data are appropriate. Ultimately, this additional data can extend further and further to entire genomes. This is becoming more and more feasible, as costs continue to fall and for example, a project to sequence 10K bird species genomes has just been announced (http://b10k.genomics.cn). Yet on the massive scale of 'life on earth', there remains an inevitable and an unavoidable resource trade-off between the depth of sequence coverage and breadth of the sample set. The genome skimming approach can thus be considered useful as a scalable and moveable bridge between standard barcodes and genome sequencing. The depth of the skim can be determined by the data required and resources available, and when data needs increase (and costs decrease), samples can be re-sequenced at increasing depths.

This potential for increase in sequencing depth is particularly important in plants. In this study, we have outlined various practical reasons – beyond species discrimination – why it is desirable to use genome skimming which can recover complete plastid genomes and rDNA sequences as an extended plant barcode. However, these organelle genomes and rDNA sequences will not deliver universal species discrimination (Hollingsworth *et al.* 2011). To achieve this in many plant genera, and indeed other 'difficult' groups, approaches which enable routine access to hundreds of nuclear loci in phylogenetically disparate sample are likely to be required. As noted earlier, shallow-pass nuclear data generated during genome skims can provide relevant information for marker development, and there is also some intrinsic signal in the data itself. Further studies are needed to establish in which groups, which level of coverage, would give sufficient single copy nuclear data to provide appreciable species discrimination gains, but these additional data represent a promising line of enquiry. And as sequencing costs fall, and read length increase, the richness of nuclear data from genome skims will continue to grow as a resource for species discrimination and other biological applications.

Acknowledgements

SL received support from the European Research Council under the European Community's Seven Framework Programme FP7/2007–2013 Grant Agreement no. 281422 (TEEMBIO) to Wilfried Thuiller. PMH received support from the Scottish Government's Rural and Environment Science and Analytical Services Division (RESAS). The sequencing of the PhyloAlps project is funded by the France Genomique 'Grands investissements d'Avenir' and the Genoscope (Evry).

References

Baird NA, Etter PD, Atwood TS, *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Baldwin DS, Colloff MJ, Rees GN *et al.* (2013) Impacts of inundation and drought on eukaryote biodiversity in semi-arid floodplain soils. *Molecular Ecology*, **22**, 1746–1758.

Besnard G, Christin P-A, Malé P-JG *et al.* (2014a) From museums to genomics: old herbarium specimens shed light on a C3 to C4 transition. *Journal of Experimental Botany*, **65**, 6711–6721.

Besnard G, Jühling F, Chapuis É *et al.* (2014b) Fast assembly of the mitochondrial genome of a plant parasitic nematode (*Meloidogyne graminicola*) using next generation sequencing. *Comptes Rendus Biologies*, **337**, 295–301.

Clarke LJ, Soubrier J, Weyrich LS, Cooper A (2014) Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, **14**, 1160–1170.

Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the COI marker: not a perfect match. *Biology Letters*, **10**, UNSP 20140562.

Dodsworth S (2015) Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science*, **20**, 525–527.

Elias M, Hill RI, Willmott KR *et al.* (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B-Biological Sciences*, **274**, 2881–2889.

Fan H, Ives AR, Surget-Groba Y, Cannon CH (2015) An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data. *BMC Genomics*, **16**, 522.

Fazekas AJ, Burgess KS, Kesanakurti PR *et al.* (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE*, **3**, e2802.

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads–a baiting and iterative mapping approach. *Nucleic Acids Research*, **41**, e129.

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

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**(Suppl. 1), S96–S99.

Hickerson MJ, Meyer CP, Moritz C (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, **55**, 729–739.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**(Suppl. 1), 117–122.

Hollingsworth PM (2011) Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 19451–19452.

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

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. *PLoS ONE*, **6**, e19254.

Kane N, Sveinsson S, Dempewolf H *et al.* (2012) Ultra-barcoding in cacao (Theobroma spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany*, **99**, 320–329.

Kartzinel TR, Chen PA, Coverdale TC *et al.* (2015) DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 8019–8024.

Kim H, Jebrail MJ, Sinha A *et al.* (2013) A microfluidic DNA library preparation platform for next-generation sequencing. *PLoS ONE*, **8**, e68988.

Lasken RS, McLean JS (2014) Recent advances in genomic DNA sequencing of microbial species from single cells. *Nature Reviews Genetics*, **15**, 577–584.

Li DZ, Gao LM, Li HT, Wang H, Ge XJ (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 19641–19646.

Li XW, Yang Y, Henry RJ *et al.* (2015) Plant DNA barcoding: from gene to genome. *Biological Reviews of the Cambridge Philosophical Society*, **90**, 157–166.

Liang J, Cai W, Sun Z (2014) Single-cell sequencing technologies: current and future. *Journal of Genetics and Genomics = Yi chuan xue bao*, **41**, 513–528.

Linard B, Crampton-Platt A, Gillett CPDT, Timmermans MJTN, Vogler AP (2015) Metagenome skimming of insect specimen pools: potential for comparative genomics. *Genome Biology and Evolution*, **7**, 1474–1489.

Liu C, Shi L, Zhu Y *et al.* (2012) CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics*, **13**, 715.

Maillet N, Collet G, Vannier T, Lavenier D, Peterlongo P (2014) Commet: Comparing and combining multiple metagenomic datasets. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2014, pp. 94–98. ieeexplore.ieee.org.

Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.

Miller JM, Malenfant RM, Moore SS, Coltman DW (2012) Short reads, circular genome: skimming solid sequence to construct the bighorn sheep mitochondrial genome. *The Journal of Heredity*, **103**, 140–146.

Nicholls JA, Pennington RT, Koenen EJM *et al.* (2015) Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science*, **6**, 710.

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

Pawlowski J, Audic S, Adl S *et al.* (2012) CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biology*, **10**, e1001419.

Peñalba JV, Smith LL, Tonione MA *et al.* (2014) Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Molecular Ecology Resources*, **14**, 1000–1010.

Percy DM, Argus GW, Cronk QC *et al.* (2014) Understanding the spectacular failure of DNA barcoding in willows (*Salix*): does this result from a trans-specific selective sweep? *Molecular Ecology*, **23**, 4737–4756.

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Ruhsam M, Rai HS, Mathews S *et al.* (2015) Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in Araucaria? *Molecular Ecology Resources*, **15**, 1067–1078.

Schoch CL, Seifert KA, Huhndorf S *et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 6241–6246.

Straub SCK, Parks M, Weitemier K *et al.* (2012) Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *American Journal of Botany*, **99**, 349–364.

Stull GW, Moore MJ, Mandala VS *et al.* (2013) A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. *Applications in Plant Sciences*, **1**, 1200497.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**, 2045–2050.

Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics-the International Journal of the Willi Hennig Society*, **20**, 47–55.

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 9054–9058.

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, **20**, 3252–3255.

Zonneveld BJM, Leitch IJ, Bennett MD (2005) First nuclear DNA amounts in more than 300 angiosperms. *Annals of Botany*, **96**, 229–244.

E.C., P.H., S.L. and P.T. all contributed equally to developing the ideas presented here and writing the paper.

---

doi: 10.1111/mec.13549