American Journal of Botany

American Journal of Botany 98(3): 415–425. 2011.

BIODIVERSITY ASSESSMENT: STATE-OF-THE-ART TECHNIQUES IN PHYLOGENOMICS AND SPECIES IDENTIFICATION[1]

P. ROXANNE STEELE[2] AND J. CHRIS PIRES

Division of Biological Sciences, 371 B Life Science Center, 1201 Rollins Road, University of Missouri, Columbia, Missouri 65211 USA

- *Premise of the study:* Biodiversity assessment is the first step in protecting the complete range of morphological and genetic diversity of species on Earth, and in reaching the goals of conservation biology. Assessment begins with identifying organisms that make up biological communities and understanding evolutionary histories. Scientific advancements in molecular sequencing can help clarify and provide support for identifications. Massively parallel DNA sequencing technologies are being used to sequence complete genomes of model organisms; however, this resource has not been fully used for species identifications. Animal researchers commonly use one mitochondrial region, and groups of plant scientists have proposed numerous combinations of two or three chloroplast markers as genomic identifiers. Yet, nearly as many studies have reported that the proposed regions are uninformative in some plant groups and at various taxonomic levels.
- *Methods:* We propose a combination of whole (or nearly whole) chloroplast genomes, mitochondrial genes, and nuclear repeat regions for both species identifications and phylogenetic analyses, obtained from a simple total DNA extraction and one run on massively parallel DNA sequencing machines.
- *Key results:* We have recovered both coding and noncoding sequences from multiple genetic sources, providing genomic information for comparisons within and between multiple taxonomic levels.
- *Conclusions:* In combination with morphological and other data, this abundance of genomic information will have a broad range of applications, including not only helping conservation biologists understand ecosystem biodiversity, but also understanding the evolutionary histories of organisms, mending damaged landscapes, and investigating interactions of plants with pollinators and pests.

Key words: citizen science; DNA barcoding; genome; genome survey sequencing; handheld field device; mitochondria; next-generation sequencing; nuclear ribosomal DNA; plastome; whole genome shotgun.

Conservation planning begins with a comprehensive evaluation of regional biodiversity. Knowledge of organisms' life histories, species richness, endemism, rarity, range of morphological and genetic variability, and evolutionary histories are all part of a biodiversity assessment, but the necessary first step is identification of the organisms themselves. Several calls have been made in the last 15 years for systematists to contribute to conservation and biodiversity assessment and to make taxonomic data more accessible to a larger audience (Vane-Wright, 1996; Soltis and Gitzendanner, 1999; Sytsma and Pires, 2001; Cracraft, 2002; Godfray, 2002; Liston, 2003; Hendry et al., 2010). With recent advances in molecular sequencing technologies, there has never been a better time to respond to the call with great effort and enthusiasm. The newest DNA sequencing method, called massively parallel (MP) or next-generation technology, is capable of delivering more data at lower cost and with less presequencing laboratory preparation than traditional Sanger sequencing, which has been the standard for over 30 years. Although MP sequencing has been available only for a few years, equipment manufacturers have already improved the technology substantially, and research scientists have applied the method in a great variety of investigations (e.g., Cronn et al., 2008; Mardis, 2008; Morozova and Marra, 2008; Smith et al., 2008; Wheeler et al., 2008; Harismendy et al., 2009; Valentini et al., 2009; Li et al., 2010; Atherton et al., 2010; Givnish et al., 2010; Rounsley and Last, 2010). A relatively unexplored area of this high-throughput MP sequencing is its application to biodiversity assessment (but see Nock et al., in press), perhaps because of prohibitive costs and lack of technical expertise to process the large quantities of data generated. However, costs are decreasing, and bioinformatics programs are improving.

Our experience with MP sequencing suggests that this method will revolutionize DNA sequencing and phylogenetic efforts. After samples are collected from the field, laboratory work for this type of DNA sequencing includes an extraction of total genomic DNA from <20 mg of leaf tissue or 25 mg of animal tissue and preparation of a sequencing library (Fig. 1). Both steps can be performed with commonly available, prepackaged kits. After MP sequencing, data are available for assembling various regions (using available software) from all the genomes in an organism, including the complete chloroplast genome in plants and the whole mitochondrial genome in insects and other animals. These data can be combined with morphological, ecological, and geographical information to identify the organisms, infer evolutionary histories, and form a complete picture of the biodiversity of any region (Fig. 2). Additionally, this information can be made available on the Internet. This technology is

[1] Manuscript received 10 August 2010; revision accepted 24 January 2011.

The authors thank three anonymous reviewers for valuable comments on the manuscript; W. Allen, M. Bekaert, G. C. Conant, T. Finch, K. L. Hertweck, M. M. Kroll, and E. J. Wheeler for comments on early versions of the manuscript; and members of the University of Missouri DNA Core Facility for informative discussions about the Illumina technology and MP sequencing. The authors acknowledge the following research grants and granting institutions: National Science Foundation (DEB 0829849) and the University of Missouri Research Board.

[2] Author for correspondence (e-mail: roxisteele@msn.com)

doi:10.3732/ajb.1000296

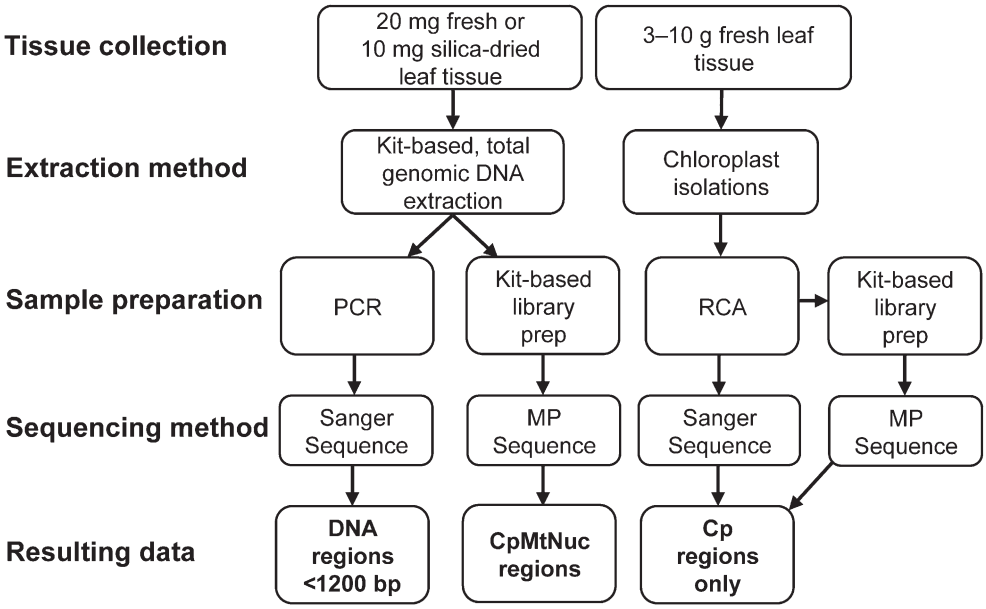

Fig. 1. Comparison of processes for obtaining sequences for individual, short DNA regions, chloroplast, mitochondrial, and nuclear repeat (CpMtNuc) regions, and only chloroplast DNA.

advancing so rapidly that some day there will be desktop sequencing machines and, likely, handheld instruments for field use (first proposed by Paul Hebert, University of Guelph) by both professional scientists and amateurs (citizen scientists) such that they can identify any organism they desire. In this paper, we discuss the current state of biodiversity assessment and organism identification, give a brief overview of various MP sequencing methods by current manufacturers, discuss current applications, and then present the possible applications of MP sequencing to understanding the biodiversity that makes up

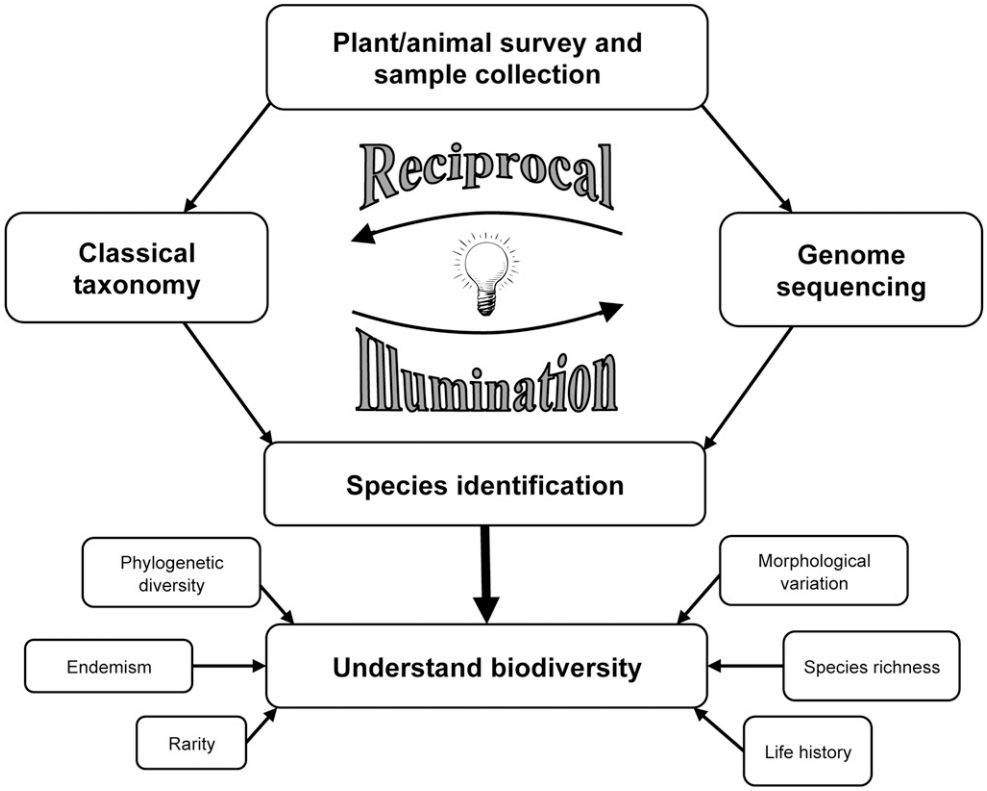

Fig. 2. Biodiversity assessment. Conservation planning begins with understanding the biodiversity of a geographical region, and this starts with identifying the organisms, based on a combination of classical taxonomic techniques and genomic sequencing CpMtNuc data.

Earth's ecosystems. We conclude with a comment on the feasibility of even small laboratories conducting the methods described herein, and an outlook on the future of species identification, given advancements in technology and online resources.

BIODIVERSITY ASSESSMENT

Biodiversity assessments typically consist of identifying species and evaluating one or more metrics: (1) species richness (number of species), (2) evenness (ratio of various species), and/or (3) character diversity (phenotypic difference) (Purvis and Hector, 2000). Because it is the simplest way to describe community and regional diversity (Magurran, 1988), the most common of these metrics is species richness. (For in-depth discussions of the pros and cons of measuring species richness, see Myers et al., 2000; Purvis and Hector, 2000; Nilsson et al., 2001; Gotelli and Colwell, 2001; Funk and Richardson, 2002.) Biodiversity inventories of most geographic regions are not comprehensive, and many scientists believe that identifying all species in a particular landscape is nearly impossible (Nilsson et al., 2001). Because of this, a few species are often selected as "indicator" taxa, which act as surrogates for the diversity of those taxa that are not inventoried. To direct conservation planning, this practice is often combined with complementarity, whereby the land is divided into a grid, species distribution is assessed, and then complementary sets of grids containing each taxon at least once are selected for conservation (Van Jaarsveld et al., 1998). In other words, given a set of areas to potentially conserve, select the area (or areas) with the greatest species richness. However, there are many problems with these techniques, such as uninformative surrogate species (Van Jaarsveld et al., 1998) and sampling issues (Gotelli and Colwell, 2001). The greater the number of species that can be identified and counted without using surrogates, the better.

Additional metrics being used to evaluate biodiversity include phylogenetic diversity (PD), first described by Faith (1992), and more recently discussed by Magnuson-Ford et al. (2010). Phylogenetic diversity estimates evolutionary relationships and reveals genetic distances between samples in a data set. Species and geographical regions are selected to preserve the greatest combination of evolutionary history. Several studies have used this technique to successfully evaluate a set of species (e.g., Beenaerts et al., 2010), including a suggestion to combine PD assessment with endemism (Moøers and Redding, 2009). The method has seen limited use in evaluating geographic regions (but see Forest et al., 2007; Donoghue, 2008), perhaps because of the difficulties in identifying and sequencing useful molecular markers for many varied samples. More recent applications have incorporated additional measures such as extinction probability (Faith, 2008), rarity (Rosauer et al., 2009), and abundance (e.g., Cadotte et al., 2010). Applications of phylogeny to biodiversity have broadened into studies of community assembly and ecology (Donoghue, 2008) and into ecosystem functioning research (King, 2009).

Regardless of which biodiversity metric is used, species identification is central to the measurement. It is the primary area in which new systematics tools can make the greatest contribution to conservation and biodiversity assessment. As discussed by Hendry et al. (2010), phylogeneticists and evolutionary biologists can improve biodiversity science, conservation, and policy by applying their investigative techniques to discovering and documenting biodiversity, understanding the causes of diversification, and evaluating evolutionary responses to human disturbances. In fact, these authors initiated a new project called bioGENESIS, in which evolutionary biologists can brainstorm ways to make practical contributions in understanding and reducing the loss of biodiversity (Hendry et al., 2010).

The call for systematists to play a role in biodiversity assessment is not new. In addition to training parataxonomists (field-trained, biodiversity specialists who gather and organize specimens; Janzen, 2004), setting up easily accessible reference collections, and making conservation and environmental evaluations, Vane-Wright (1996) requested user-friendly species identification systems. Soltis and Gitzendanner (1999) identified four areas in which systematists could contribute to conservation of rare plant species: (1) solidifying species concepts, (2) identifying lineages worthy of conservation, (3) setting conservation priorities, and (4) evaluating the effects of hybridization on the biology and conservation of rare species. And more recently, Godfray (2002) has challenged the taxonomic community to completely rethink its strategies for assembling and distributing information about species classification and nomenclature. Indeed, the newly developed, web-based CATE (Creating a Taxonomic e-Science) project promotes community-facilitated revisions such that it may one day be the single source of authoritative information about taxa for both the scientific community and amateurs (Clark et al., 2009). Using state-of-the-art technologies, the procedures described next can help build these online databases and will allow systematists to make important contributions to conservation planning and biodiversity inventories.

ORGANISM IDENTIFICATION

Organism identification is essential to many disciplines. Here, we highlight some of those applications, discuss the importance and limits of identification based on both classical taxonomic techniques and on DNA sequencing, and argue for an integrated approach to organism identification (Fig. 2). In this paper, we refer to the fundamental units of biodiversity as species, because it is the term most used by scientists and the public. Indeed, conservation planners and government agencies would have difficulty protecting ecosystems and writing effective laws if species boundaries were not defined (Primack, 2008).

Species identification is important in many industries and fields of study. Restoration ecologists must accurately identify native plant species suitable for rebuilding damaged ecosystems (Guerrant et al., 2004). Forensic scientists use plant traces to aid in resolution of legal issues, and a correctly identified plant along with its anatomy and ecological requirements can serve as botanical evidence about a crime scene or the whereabouts of a suspect or victim (Lane et al., 1990). Law enforcement and customs agents must be able to efficiently and accurately identify plant species to prevent trafficking of rare plants such as orchids (CITES, 2010). Harvesters of wild plants for food and medicinal uses must be certain that the right plant species are collected before distribution to the public. Ecologists need to identify organisms of study at different life stages to understand their life histories. And finally, conservation biologists must be able to correctly identify plant species for many reasons, including (1) fighting nonnative, invasive species by documenting the scope of the problem and helping to raise awareness, (2) reseeding restoration areas with appropriate species, (3) protecting native and/or threatened species by

preserving all life in an ecosystem, and (4) understanding biological diversity (CPC, 2010).

Classical identification—Plant species are typically described by taxonomists, and they are differentiated from other species based predominantly on morphological characters. Other criteria are used in taxonomic keys to delineate species such as habitat, ecological niche, life history traits, and geographic distribution. Identification of species based on morphology requires taxonomic expertise and the presence of multiple characteristics at a certain life stage of the organism. This can make identification difficult when characters such as flower color, flower shape, and leaf morphology change over the life of an individual (e.g., Steele et al., 2010). Furthermore, classical taxonomic methods alone often cannot be used to determine species from seeds or plant fragments that may be present in animal dung, on an herbarium specimen, or at a crime scene. Consequently, an identification strategy is necessary for augmenting classical taxonomic techniques.

DNA barcoding—In recent years, systematists have begun to consider the variation in DNA sequences among species as characters to define the species. However, monographs, taxonomic revisions, keys to identification, and descriptions of new species are still primarily based on morphological characteristics (but see Steele, 2010 and Fig. 3). Molecular data will not (and should not) replace species distinctions based on physical characteristics; however, they can provide supporting evidence for and contribute to making taxonomic decisions (as demonstrated by the process in Fig. 2).

Since it was brought to mainstream attention by Hebert et al. in 2003, DNA barcoding has gained much attention, both positive and negative (for examples of extensive discussions on this topic, see: Godfray, 2002; Tautz et al., 2002; Blaxter, 2003; Hebert et al., 2003; Lipscomb et al., 2003; Seberg et al., 2003; Tautz et al., 2003; Moritz and Cicero, 2004; Will and Rubinoff, 2004; Chase et al., 2005; DeSalle et al., 2005; Hebert and Gregory, 2005; Marshall, 2005; Savolainen et al., 2005; Will et al., 2005; Cowan et al., 2006; Rubinoff et al., 2006; Hajibabaei et al., 2007; Sass et al., 2007; Vogler and Monaghan, 2007; Kress and Erickson, 2008; Stoeckle and Hebert, 2008; Fazekas et al., 2008, 2009; Gonzalez et al., 2009; Packer et al., 2009; Spooner, 2009; Vernooy et al., 2010). DNA barcoding is a method of identifying organisms using standardized portions of their genome. In addition to the contributions that DNA barcoding has made and will likely continue to make to scientific investigations, the technique has also found mainstream public use in what is sometimes called citizen science. A few of the more popular applications are described in several Internet articles, such as identifying endangered tuna species in consumer products (Lowenstein et al., 2009), exposing fake ferns in the international plant trade (e! Science News, 2010a), and detecting the movement of insect pests (e! Science News, 2010b).

In 2003, Hebert et al. suggested a mitochondrial gene, cytochrome c oxidase I (COI), as a DNA barcode for global identification of animals. Although it has not always worked optimally (Meier et al., 2006), COI has been tested in several insect and other animal groups with much success (e.g., Blaxter et al., 2004; Hebert et al., 2003, 2004a, b; Smith et al., 2005, Hajibabaei et al., 2006; Witt et al., 2006). The task of finding an equivalent, suitable DNA barcode for plants, however, has proven to be much more difficult and controversial.

Because the mitochondrial genome in plants typically has very low levels of variability, most attention has focused on the chloroplast genome with the addition, in some cases, of the internal transcribed spacers (ITS) of the nuclear ribosomal genes. Plant scientists generally agree that multiple chloroplast regions must be used together for a DNA barcode for plants, and several working groups have proposed various combinations for identifying flowering plants (ITS + *trnH-psbA*, Kress et al., 2005; portion of *matK*, Lahaye et al., 2008), land plants (ITS + *rbcL*, Chase et al., 2005; *rpoC1* + *rpoB* + *matK* or *rpoC1* + *matK* + *psbA-trnH*, Chase et al., 2007; portion of *rbcL* + *trnH-psbA*, Kress and Erickson, 2007; *rbcL* + *matK*, CBOL, 2009), Amazonian trees (morphology plus various combinations of *rbcL, rpoC1, rpoB, matK, ycf5, trnL, psbA-trnH*, and ITS Gonzalez et al., 2009), and a 50-ha forest in Panama (*rbcL* + *matK* + *trnH-psbA*, Kress et al., 2009). Each of these groups suggested a different combination of regions as DNA barcodes for plants. Further, a multitude of studies has since shown that

1. Leaves simple, trilobed or 5-lobed, sometimes deeply, but never trifoliolate 2
   2. Chloroplast DNA region *rpoB-trnC* matches *P. racemosa* reference sequence and geographic distribution in Colombia or Venezuela ..... 2. *P. racemosa*
   2. Chloroplast DNA regions *rpoB-trn*C and/or *psbZ-trnM* match *P. umbrosa* reference sequences and geographic distribution in Venezuela, Brazil, or Lesser Antilles ..... 5. *P. umbrosa*
1. Leaves bifid, trifoliolate, or pedate, with immature leaves often simple and lobed, or sometimes a combination of shapes 3
   3. Stem surface light green with darker green splotches, leaves trifoliolate, pedate, or seldom simple and trilobed, large lobes on outside of side leaflets, and chloroplast DNA region *ndhF-rpL32* matches *P. umbrosa* reference sequence, but not matching regions *rpoB-trnC* and/or *psbZ-trnM* with *P. umbrosa* reference sequences ..... 6. *P. warscewiczii*
   3. Stem surface solid green, leaves bifid, trifoliolate, pedate, with immature leaves often simple and lobed, or sometimes a combination of complexities 4
      4. Leaf margins with veins extending beyond the margins of the leaves, chloroplast DNA region *ndhC-trnV* matches *P. pedata* reference sequence, and geographic distribution in Greater Antilles or rarely Costa Rica. ..... 1. *P. pedata*
      4. Leaf margins seldom with veins extending beyond the margins of the leaves 5
         5. Leaves trifoliolate, or rarely pedate, and main vein in side leaflets strongly off-center, typically with small lobes on outside base of side leaflets, chloroplast DNA region *3rps16-trnQ* matches *P. ternata* reference sequence, and geographic distribution in Peru, Brazil, Bolivia, or Paraguay ..... 3. *P. ternata*
         5. Leaves simple and elliptic, 2-lobed, trilobed, trifoliolate, or mixed complexities; main vein in side leaflets not strongly off-center, typically without lobes on outside base of side leaflets, chloroplast DNA regions *ndhC-trnV, ndhF-rpL32*, and/or *3rps16-trnQ* match *P. triphylla* reference sequences, and geographic distribution in Mexico, Central America, or South America ..... 4. *P. triphylla*

Fig. 3. Taxonomic key to species of *Psiguria* (Cucurbitaceae) without staminate flowers, reproduced from Steele (2010). GenBank accessions for DNA regions are listed in Steele et al. (2010).

these and other marker combinations differ in their ability to amplify, produce clean sequences, or discriminate species, and some have failed completely in certain plant groups (e.g., Sass et al., 2007; Fazekas et al., 2008, 2009; Gonzalez et al., 2009; Spooner, 2009).

Despite some failures, these gene combinations will continue to be successful in some plant groups and for some investigations. However, as noted by Fazekas et al. (2008), the poor results obtained by some plant scientists indicate that DNA barcoding systems that include only a few chloroplast markers may indicate a limit to their ability to differentiate species. For those applications in which a close approximation of the species or identification to family or genus of the organism is sufficient, combinations of two to three of the genes mentioned will often be adequate. However, other challenges such as the design of broadly useful primers and variation in PCR success (discussed later) may hinder the present DNA barcoding methods. Fazekas et al. (2008) further conclude that regardless of the region or regions chosen as plant DNA barcodes, some species may be better distinguished by other regions. Because this gene-by-gene approach has not yielded a universal barcode despite the effort expended, an alternative approach is needed that has the power to detect variation at all taxonomic levels and to distinguish hybrids. Ideally, the alternative would be as simple and as economical as current DNA barcodes, but it would include information from both organellar DNA and the nuclear genome.

MASSIVELY PARALLEL SEQUENCING: METHODS

The first decade of this century saw the development and introduction of several novel methods for high-throughput DNA sequencing (without cloning of DNA fragments into bacterial vectors), such as the 454-FLX (Roche Diagnostics Corp., Basel, Switzerland), SOLiD (Applied Biosystems by Life Technologies Corp., Carlsbad, California, USA), and Genome Analyzer (Illumina, San Diego, California). These massively parallel (MP) sequencing technologies were called next-generation by most researchers; however, with the recent launch of even newer technologies, sometimes termed next-next generation, later generation, 3rd generation or now generation, we suggest the "generation" terminology has outlived its usefulness. Since their introduction, the various MP sequencing machines have continued to improve, offering increasing quantities (and quality) of sequence at ever-lower costs. A review of the technological details of MP sequencing and differences between the various manufacturers are discussed in several other publications (e.g., Margulies et al., 2005; Bentley, 2006; Church, 2006; Hutchison, 2007; Shaffer, 2007; Bentley et al., 2008; Mardis, 2008; Rothberg and Leamon, 2008; Rusk and Kiermer, 2008; Schuster, 2008; Shendure and Ji, 2008; von Bubnoff, 2008; Ansorge, 2009; Lister et al., 2009; Pettersson et al., 2009; Metzker, 2010). Here, we review template preparation, sequencing, and data analysis for MP sequencing.

Template preparation—Extracting total genomic DNA using a kit is much faster than isolating chloroplast genomes, and the further preparation of samples for MP sequencing (making libraries) is simpler than that for Sanger sequencing used in current DNA barcoding methods. The following is a discussion about the differences between these methods for both DNA extraction and sequencing preparation (Fig. 1).

For sequencing an entire chloroplast genome using Sanger sequencing, chloroplasts must first be isolated. In fact, some researchers using MP sequencing prepare samples this way; however, it is quite complex and unnecessary. Chloroplast genomes (and, potentially, proplastids and other nonchloroplast genomes) are obtained by first isolating the organelles, lysing their membranes, and then multiplying the genomes using rolling circular amplification (RCA) following the procedures of Palmer (1986), Jansen et al., (2005), or others. Rolling circular amplification products are then sent to large central laboratories for sequencing, and several weeks later, electronic files of reads (sequence fragments) are obtained that are ready for assembly. The results obtained will include only chloroplast genome sequences. In contrast, using a kit such as Qiagen's DNeasy Plant Mini (Germantown, Maryland, USA), total genomic DNA from several different samples can be extracted each day, utilizing small, bench-top microcentrifuges, even in small laboratories. The resulting DNA can be tested for chloroplast content using real-time PCR (rt-PCR) with 1 μL of the extraction. However, to get accurate results, it is necessary to be extremely precise when measuring DNA quantities and dilution volumes, and many samples contain molecules that confound these measurements and the results. Therefore, this test is not highly accurate but can still be informative, giving some indication of the percent chloroplast DNA in the sample.

For traditional polymerase chain reaction (PCR) and Sanger sequencing of a few genes, once DNA is extracted, forward and reverse primers may need to be designed for each region to be sequenced and for each taxonomic group of interest. This is not a trivial task. PCR is conducted, and amplicons are visualized on agarose gels. If no amplicons are detected, the PCR reaction ingredients or ratios, thermal cycler routine, primers, or a combination of these in multiple iterations must be adjusted in an attempt to successfully amplify the amplicon in all samples. Finally, after obtaining amplicons, the products are cleaned, and then sequencing reactions are sent for Sanger sequencing. The typical products are reliable forward and reverse reads of the gene region. This procedure must be repeated for every gene region up to ca. 1200 bp long.

With MP sequencing, first a library is made from total genomic DNA. The library is sent to a core sequencing facility, and reads are received that will nearly always assemble into a complete chloroplast genome for plants, various mitochondrial genes (whole mitochondrial genomes for animals/insects), and nuclear ribosomal repeat regions. Sequencing libraries may also be made from the RCA product of chloroplast isolations, but the resulting reads will not include sequences from the mitochondrial or nuclear genomes. Making libraries for Illumina sequencing requires two kit-based procedures. (We focus on this manufacturer/technology because it is the one with which we have the most experience. Similar template/library preparation procedures are necessary for other MP sequencers.) A multistep kit procedure (using a kit such as NEB #E6000L from New England Biolabs, Ipswich, Massachusetts, USA) is conducted to make a solution of particular-sized fragments of all DNA in the organism. Steps include shearing the DNA (by sonication or chemical treatment), repairing ends, preparing fragments for adapter ligation, ligating adapters, selecting the appropriate fragment size from an agarose gel, and then enriching the product with a PCR. Other than shearing time in the first step, no adjustments are required in the preparation of different samples. The adapters, and therefore PCR primers, are standard and the same for each template, so no additional input or design

is required for each sample. In sum, while traditional DNA barcoding may require fine tuning of the PCR process for each taxonomic group, MP sequencing library preparation uses a universal protocol.

Sequencing—Since 454 Life Sciences first introduced its technology in 2005 (Margulies et al., 2005), the top three manufacturers of MP sequencers (Roche Diagnostics, Applied Biosystems by Life Technologies, and Illumina) have continually improved their systems for the number of reads per run and output read lengths. Early sequencing read lengths from 454 Life Sciences, Illumina/Solexa, and Applied Biosystems/SOLiD were 250 bp, 25–35 bp, and 25–35 bp, respectively (Mardis, 2008). On 1 June 2010, the 454 Life Sciences website indicated that their FLX sequencer could sequence "more than 1 million high-quality reads per run and read lengths of 400 bases," and they anticipated increasing that to nearly 1000 bases in the year 2010 (Roche, 2010). On this same date, Illumina's website reported "75+ bp reads for a total of > 20 Gb of paired-end data per run" (Illumina, 2010). In 2010, on an Illumina sequencing machine, our laboratory tested new chemistry and software that resulted in 120-bp reads (unpublished data). Applied Biosystems' SOLiD technology is now reaching up to 75-bp read lengths (Applied Biosystems, 2010). With longer reads, coverage of the genomes will be greater and confidence in final sequences higher. As these outputs improve, the resulting data will be more informative and more useful for genome sequencing and organismal identification.

Data analysis—The greatest challenges facing those researchers wanting to use MP technology for sequencing genomes are related to bioinformatics—processing and analyzing the data. The read files are quite large, and assembly programs are in their infancy. Desktop computers or access to servers with large amounts of storage and memory are required for the initial steps. Once compiled into a handful of longer contig sequences using a de novo assembler such as Velvet (Zerbino and Birney, 2008) or Mira (Chevreux et al., 1999) or reference-based assemblers such as Yasra (Ratan, 2009) sequences can be manipulated and visualized in traditional sequence-manipulating programs such as Geneious (Biomatters, Aukland, New Zealand) or Sequencher (Gene Codes Corp., Ann Arbor, Michigan, USA). Then genes can be annotated, and exons, introns, rRNAs, and tRNAs can be identified. DOGMA (Wyman et al., 2004) can be used for annotating organellar genomes and preparing them for publication. Additional bioinformatics challenges are reviewed by Pop and Salzberg (2008), and various de novo assembly programs are reviewed by Miller et al. (2010).

As MP sequencing and its applications have grown, members of the bioinformatics community have been developing computer programs to help biologists analyze the massive amounts of data produced. In addition to those programs already listed, biologists and computer scientists have been working together to write computer programs that can be run on a desktop computer with minimal specialized expertise. For example, biologists at the University of Missouri and computer scientists at the University of Missouri Informatics Institute have been writing scripts that will ease the burden of analyzing large amounts of data (A. Tegge, C. Hudson, B. Pang, and N. Shao, unpublished data). These programs will perform three important functions: (1) assembling sequence reads into larger consensus contigs, (2) identifying and annotating genes, and (3) aligning coding and noncoding sequences for phylogenetic analyses. The various computer programs may be used in a variety of ways, depending on the investigation. The programs are being written along with instruction manuals such that anyone familiar with using a computer will be able to use them. Eventually, a web-based interface could be developed.

Costs—Another limitation to the widespread use of MP technologies is the current cost of sequencing. When compared to Sanger technology for sequencing a few genes, the price of sequencing entire organellar genomes plus nuclear repeat regions using MP sequencing is still higher; however, the amount of data generated is also several times higher. A current estimate is that sequencing one chloroplast region costs ca. US$25–30 per sample and sequencing chloroplast–mitochondrial–nuclear repeat (CpMtNuc) regions costs ca. US$200–300, which translates into 10 times the cost for more than 100 times the information. The costs for sequencing must be weighed against the amount and value of the data generated, along with comparisons of time required to prepare samples for sequencing and time to analyze the data. The cost for MP sequencing is continuously falling as technologies improve, but the decision about which method to use will depend on the specific investigation and its goals. If a researcher is working on those plant groups in which universal primers successfully amplify the desired regions and if there is variation between samples at the taxonomic level being studied, then traditional PCR methods may suffice in both phylogenetics and species identification investigations. However, if the DNA barcoding process does not generate sufficient data or sequence variation, MP sequencing may be preferable. MP sequencing also gives data for additional studies such as chloroplast biology (e.g., rearrangement, gene loss or gain, whole plastid genome rates of evolution)

MASSIVELY PARALLEL SEQUENCING: APPLICATIONS

Initially, the scientific community was slow to accept MP sequencing systems because they were unproven and more expensive than the classical Sanger sequencing technology used since its inception (Sanger et al., 1977). Today, the MP approach seems to have finally gained general acceptance as the number of applications has greatly increased over the last few years, costs are decreasing, and future applications appear to be numerous. In Table 1, applications of the gene-by-gene approach to sequencing using traditional PCR are compared to those using MP sequencing.

Current applications—One of the earliest uses of MP technologies was to sequence the whole nuclear genome of a human, James D. Watson (Wheeler et al., 2008), and more recently that of the giant panda (Li et al., 2010). Ecologists are using MP sequencing in a variety of investigations, including metagenomics, which evaluates genetic diversity in soil or water samples (e.g., Valentini et al., 2009; Yahara et al., 2010). Several applications have been addressed in genetics, such as mutation discovery, regulatory protein binding, discovering noncoding RNAs (Mardis, 2008), and targeted sequencing of candidate genes (Harismendy et al., 2009) as well as functional genomics, such as gene expression profiling, protein coding gene annotation, detection of aberrant transcription events (Morozova and Marra, 2008), and characterizing a mutant genome (Smith et al., 2008). MP sequencing is also finding applications in plant biology (e.g., Rounsley and Last, 2010), conservation genetics

TABLE 1. Comparison of the applications of the gene-by-gene sequencing approach to the massively parallel sequencing of CpMtNuc regions.

Topic of investigation	Gene-by-gene Cp + ITS	CpMtNuc
ID to family or genus	+	+
ID to species	±	+
Species ID support from multiple lines of evidence	±	+
Phylogenetics at various taxonomic levels	±	+
Identify hybrids	±	±
Use in taxonomic keys	+	+

Notes: CpMtNuc = chloroplast, mitochondrial, and nuclear repeat; ID = identification; + = yes; ± = sometimes.

(Avise, 2010), and assessment of genetic diversity of functional traits (Yahara et al., 2010).

Potential applications to biodiversity and beyond—The DNA regions obtained from MP sequencing—partial/whole chloroplast genomes, partial/whole mitochondrial genomes, and nuclear ribosomal and other repeat regions (collectively called CpMtNuc regions)—have the potential to contribute a wealth of useful information to scientific research and inquiry. Both as a species identification tool and as a tool for understanding evolutionary history and relationships between species, CpMtNuc regions can aid in a great number of investigations including but not limited to restoration ecology, forensic science, agricultural science, law enforcement and border patrol, identifying areas of endemism, behavioral ecology, selecting the right biological control agent, bioprospecting, assessing the chemical components of closely related species, and conservation biology.

The development of tools that are useful for these applications will begin with a database of information, built by the systematics community. As species are identified by taxonomists, CpMtNuc regions may be sequenced for each species and made available online along with information such as morphology, ecological characteristics, and geographic distribution, similar to or in conjunction with the CATE database (Clark et al., 2009). As the information is gathered, MP sequencing technologies will improve, costs will drop, and bioinformatics tools will become more readily available, making it easier for even amateurs to use these data as the process becomes more automated. It will take many years to build a database such as this and will require efforts by expert taxonomists, biologists, and phylogeneticists. However, as mentioned before, this is a great time to start this process.

Although some researchers argue that species identifications and phylogenies should be conducted separately (Chase et al., 2005; Rubinoff et al., 2006), CpMtNuc regions can be used for both. CpMtNuc regions include both coding and noncoding regions in conserved and highly variable locations in the genome; therefore, comparisons can be made at higher and lower taxonomic levels with different portions of the data. Applications to both types of investigation are discussed here.

Species identification—Some scientists have come to believe that the best prospect for a consistent, reliable identification system for all species lies in DNA (e.g., Hebert et al., 2003). However, to develop a complete picture of any species, these data should be combined with morphological, geographical, and ecological information in species descriptions and taxonomic keys (Figs. 2, 3). The way to tie these data together is to have DNA sequences published and available for multiple individuals across the morphological and geographical range of a species that contains a vouchered specimen as identified by a taxonomist. Not only will this procedure provide molecular support for species identifications, but it will also make available multiple DNA sequences for each species for applied scientific investigations. It should also encourage the training of more taxonomic experts. The challenge, as has been typical throughout taxonomic history, will be to define boundaries between species. Molecular evidence from MP sequencing should aid this process. The ultimate goal is to provide tools for both scientists and amateurs (though the tools may be different) for understanding biodiversity and distinguishing species.

A species description is typically based on character measurements from the range of collected samples from multiple populations, even though only one sample is designated as the type specimen. In similar fashion, multiple samples of a putative species must be sequenced to measure genomic variation. The range of molecular variation across the collections will represent the tolerable genomic differences within a species. How is that range of acceptable sequence divergence determined? By the investigating scientist, just as it has always been at the discretion of the expert taxonomist to investigate, understand, and describe the morphological range of variation within a species. Scientists will likely never agree upon a universal sequence divergence threshold to distinguish all species, just as species have varying levels of morphological differences between taxonomic groups. Therefore, just as with classical delineations decided by the taxonomist working on the group, the systematist will be the one who assesses genomic differences, aligns them with phylogenetic, morphological, geographical, and other evidence, and decides where to draw the line between species (Fig. 2).

CpMtNuc regions may be incorporated into taxonomic keys. In a recent taxonomic revision by Steele (2010) of a genus of tropical vines, chloroplast DNA regions were included in a key to species (reproduced here in Fig. 3). Male flowers in this genus provide the best morphological variation between species, but they are not always present because, although these plants are monoecious, the carpellate and staminate flowers are temporally and spatially separated by great distances (Steele, 2010). As a consequence, the more traditional morphological key is only useful if male flowers are present. The second key (Fig. 3) utilizes a combination of leaf characteristics, geographical information, and a specific chloroplast DNA region for each species identified from multiple collections from multiple populations (Steele, 2010). These DNA barcodes are available in the GenBank database so that future scientific investigations in the genus can positively identify a species by comparing the appropriate DNA sequence to those in the online database. CpMtNuc regions resulting from MP sequencing can be used in the same way and can be linked with collections through herbarium and museum databases.

As discussed, efforts to identify DNA barcodes for all plants have resulted in several possible groupings of two or three chloroplast regions and, in some cases, nuclear ITS. These combinations have proven difficult to apply globally across plant clades, and most do not include nuclear regions that are necessary for assessing hybridization (Spooner, 2009). CpMtNuc regions provide an alternative option for the creation of an Internet database of DNA sequences (Table 1), and because these data include nuclear regions, they may be able to inform systematists about hybridization. As MP sequencing improves, and as mainstream technology is developed, the Internet database can

be incorporated into hand-held devices, which can be used by scientists and amateurs for species discovery in the pursuit of comprehensive biodiversity assessment.

One exception to the use of CpMtNuc regions may be with older or degraded herbarium specimens. The possibility of using these methods still needs to be tested, and this is an instance in which traditional DNA barcoding may be a good option. However, using MP sequencing methods, we have successfully sequenced the whole chloroplast genome of a plant sample from a 16-yr-old herbarium specimen (unpublished data).

Phylogenetics and conservation—CpMtNuc sequences can be used for estimating phylogenetic relationships at various taxonomic levels. They can contribute to investigations of gene and character trait evolution, whole chloroplast or mitochondrial genome evolution, and biogeographic studies. For all research projects utilizing partial/whole chloroplast genomes, partial/whole mitochondrial genomes, or nuclear ribosomal and other repeat sequences, the use of MP technologies to sequence CpMtNuc regions may provide a more efficient alternative to traditional methods. Having whole chloroplast or mitochondrial genomes for even one or two representatives from each clade in a phylogenetic study can provide a plethora of information about variation in the group. These data can help identify the most variable regions of these genomes and can lead to more efficient primer design and PCR practices when only a few genes are desired for the questions being studied.

Some of the most important conservation questions can be addressed with phylogenetics. Understanding genetic, morphological, and phylogenetic diversity in potential reserve areas can help conservation planners make decisions about the most important regions on which to concentrate available funds. It is not enough to calculate species richness; the evolutionary relationships between species must be understood (Hendry et al., 2010). Knowledge about the biological communities and the species that make them up can give scientists evidence that helps policy makers make good, logical choices about nature preserves and urban planning. Furthermore, phylogenetic studies can help scientists identify areas that contain actively speciating groups, identify the origin of and track the spread of invasive species, understand the causes of diversification, and evaluate evolutionary responses to human disturbances (Soltis and Gitzendanner, 1999; Hendry et al., 2010).

Anticipating the future—MP sequencing machines are quickly becoming miniaturized and are moving from DNA core facilities to individual laboratories. In November 2009, Roche (maker of the 454-FLX sequencer) announced their GS Junior System—a desktop printer-sized sequencer—that, coupled with a computer also supplied by Roche, can do MP sequencing and assembly of smaller genomes (such as bacterial, viral, or small fungal genomes). It is conceivable that a single handheld field device will be designed that can assess the DNA, incorporate morphology, ecology, and geography through interactive keys and graphics, and provide species identification of any biological sample. Indeed, many modern cell phones already have cameras and GPS capabilities that may soon have applications for connecting to specimen databases and electronic keys. Until a handheld field device is available, researchers can collect samples in silica gel, conduct the two kit-based laboratory procedures described earlier, and have their samples sequenced by a sequencing center. In fact, any researcher with access to a microcentrifuge and a computer with Internet access can perform the methods described here. Fast, accurate species identification will be facilitated by the development of reference genomes across the tree of life and online resources for data analysis. MP sequencing and therefore CpMtNuc regions are within reach of any researcher with the foresight to apply this technology to their investigation. Systematists can make significant contributions to conservation efforts by applying this technology in biodiversity assessments and by providing reliable and efficient identifications of species to conservation planners and policy makers.

LITERATURE CITED

Ansorge, W. J. 2009. Next-generation DNA sequencing techniques. *New Biotechnology* 25: 195–203.

Applied Biosystems. 2010. 550 series SOLiD sequencers [online]. Life Technologies, Carlsbad, California, USA. Website http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems.html [accessed 1 June 2010].

Atherton, R. A., B. J. McComish, L. D. Shepherd, L. A. Berry, N. W. Albert, and P. J. Lockhart. 2010 . Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform. *Plant Methods* 6: 22. doi:10.1186/1746-4811-6-22

Avise, J. C. 2010. Perspective: Conservation genetics enters the genomics era. *Conservation Genetics* 11: 665–669.

Beenaerts, N., R. Pethiyagoda, P. K. L. Ng, D. C. J. Yeo, G. J. Bex, M. M. Bahir, and T. Artois. 2010. Phylogenetic diversity of Sri Lankan freshwater crabs and its implications for conservation. *Molecular Ecology* 19: 183–196.

Bentley, D. R. 2006. Whole-genome re-sequencing. *Current Opinion in Genetics & Development* 16: 545–552.

Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456: 53–59.

Blaxter, M. 2003. Molecular systematics: Counting angels with DNA. *Nature* 421: 122–124.

Blaxter, M., B. Elsworth, and J. Daub. 2004. DNA taxonomy of a neglected animal phylum: An unexpected diversity of tardigrades. *Proceedings of the Royal Society of London, B, Biological Sciences* 271: S189–S192.

Cadotte, M. W., T. Jonathan Davies, J. Regetz, S. W. Kembel, E. Cleland, and T. H. Oakley. 2010. Phylogenetic diversity metrics for ecological communities: Integrating species richness, abundance and evolutionary history. *Ecology Letters* 13: 96–105.

CBOL. [Consortium for Bar Code of Life Plant Working Group: P. Hollingsworth, L. L. Forrest, J. L. Spouge, M. Hajibabaei, S. Ratnasingham, M. Van Der Bank, M. W. Chase, et al.] 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, USA* 106: 12794–12797.

Chase, M. W., R. S. Cowan, P. M. Hollingsworth, C. Van Den Berg, S. Madriñán, G. Petersen, O. Seberg, et al. 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon* 56: 295–299.

Chase, M. W., N. Salamin, M. Wilkinson, J. M. Dunwell, R. P. Kesanakurthi, N. Haidar, and V. Savolainen. 2005. Land plants and DNA barcodes: Short-term and long-term goals. *Philosophical Transactions of the Royal Society of London, B*, *Biological Sciences* 360: 1889–1895.

Chevreux, B., T. Wetter, and S. Suhai. 1999. Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics* 99: 45–56.

Church, G. M. 2006. Genomes for all. *Scientific American* 294: 46–54.

CITES. 2010. Convention on International Trade in Endangered Species of Wild Fauna and Flora [online]. CITES, Châtelaine, Switzerland. Website http://www.cites.org [accessed 22 January 2010].

CLARK, B. R., H. C. J. GODFRAY, I. J. KITCHING, S. J. MAYO, AND M. J. SCOBLE. 2009. Taxonomy as an eScience. *Philosophical Transactions of the Royal Society of London, A, Mathematical, Physical and Engineering Sciences* 367: 953–966.

COWAN, R. S., M. W. CHASE, W. J. KRESS, AND V. SAVOLAINEN. 2006. 300,000 Species to identify: Problems, progress, and prospects in DNA barcoding of land plants. *Taxon* 55: 611–616.

CPC. 2010. Center for Plant Conservation. Center for Plant Conservation, St. Louis, Missouri, USA. Website http://www.centerforplantconservation.org [accessed 22 January 2010].

CRACRAFT, J. 2002. The seven great questions of systematic biology: An essential foundation for conservation and the sustainable use of biodiversity. *Annals of the Missouri Botanical Garden* 89: 127–144.

CRONN, R., A. LISTON, M. PARKS, D. S. GERNANDT, R. SHEN, AND T. MOCKLER. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research* 36: e122.

DESALLE, R., M. G. EGAN, AND M. SIDDALL. 2005. The unholy trinity: Taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 360: 1905–1916.

DONOGHUE, M. J. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences, USA* 105: 11549–11555.

e! SCIENCE NEWS. 2010a. DNA barcoding exposes fake ferns in international plant trade. Website http://esciencenews.com/articles/2010/05/04/dna.barcoding.exposes.fake.ferns.international.plant.trade [accessed 22 June 2010].

e! SCIENCE NEWS. 2010b. Hidden habits and movements of insect pests revealed by DNA barcoding. Website http://esciencenews.com/articles/2010/03/09/hidden.habits.and.movements.insect.pests.revealed.dna.barcoding [accessed 22 June 2010].

FAITH, D. P. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61: 1–10.

FAITH, D. P. 2008. Threatened species and the potential loss of phylogenetic diversity: Conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis. *Conservation Biology* 22: 1461–1470.

FAZEKAS, A. J., K. S. BURGESS, P. R. KESANAKURTI, S. W. GRAHAM, S. G. NEWMASTER, B. C. HUSBAND, D. M. PERCY, ET AL. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* 3: e2802.

FAZEKAS, A. J., P. R. KESANAKURTI, K. S. BURGESS, D. M. PERCY, S. W. GRAHAM, S. C. H. BARRETT, S. G. NEWMASTER, ET AL. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources* 9: 130–139.

FOREST, F., R. GRENYER, M. ROUGET, T. J. DAVIES, R. M. COWLING, D. P. FAITH, A. BALMFORD, ET AL. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445: 757–760.

FUNK, V. A., AND K. S. RICHARDSON. 2002. Systematic data in biodiversity studies: Use it or lose it. *Systematic Biology* 51: 303–316.

GIVNISH, T. J., M. AMES, J. R. MCNEAL, M. R. MCKIN, P. R. STEELE, C. W. DEPAMPHILIS, S. W. GRAHAM, ET AL. 2010. Assembling the tree of the monocotyledons: Plastome sequence phylogeny and evolution of Poales. *Annals of the Missouri Botanical Garden* 97: 584–616. doi: 10.3417/2010023

GODFRAY, H. C. J. 2002. Challenges for taxonomy. *Nature* 417: 17–19.

GONZALEZ, M. A., C. BARALOTO, J. ENGEL, S. A. MORI, P. PÉTRONELLI, B. RIÉRA, A. ROGER, C. THÉBAUD, AND J. CHAVE. 2009. Identification of Amazonian trees with DNA barcodes. *PLoS ONE* 4: e7483.

GOTELLI, N. J., AND R. K. COLWELL. 2001. Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4: 379–391.

GUERRANT, E. O. JR., K. HAVENS, AND M. MAUNDER. 2004. Ex situ plant conservation: Supporting species survival in the wild, 400. Island Press, Washington, D.C., USA.

HAJIBABAEI, M., G. A. C. SINGER, P. D. N. HEBERT, AND D. A. HICKEY. 2007. DNA barcoding: How it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics* 23: 167–172.

HAJIBABAEI, M., M. A. SMITH, D. H. JANZEN, J. J. RODRIGUEZ, J. B. WHITFIELD, AND P. D. N. HEBERT. 2006. A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes* 6: 959–964.

HARISMENDY, O., P. C. NG, R. L. STRAUSBERG, X. WANG, T. B. STOCKWELL, K. Y. BEESON, N. J. SCHORK, ET AL. 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology* 10: R32.

HEBERT, P. D. N., A. CYWINSKA, S. L. BALL, AND J. R. DEWAARD. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, B, Biological Sciences* 270: 313–321.

HEBERT, P. D. N., AND T. R. GREGORY. 2005. The promise of DNA barcoding for taxonomy. *Systematic Biology* 54: 852–859.

HEBERT, P. D. N., E. H. PENTON, J. M. BURNS, D. H. JANZEN, AND W. HALLWACHS. 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences, USA* 101: 14812–14817.

HEBERT, P. D. N., M. Y. STOECKLE, T. S. ZEMLAK, AND C. M. FRANCIS. 2004b. Identification of birds through DNA barcodes. *PLoS Biology* 2: e312.

HENDRY, A. P., L. G. LOHMANN, E. CONTI, J. CRACRAFT, K. A. CRANDALL, D. P. FAITH, C. HÄUSER, ET AL. 2010. Evolutionary biology in biodiversity science, conservation, and policy: A call to action. *Evolution* 64: 1517–1528.

HUTCHISON, C. A. III. 2007. DNA sequencing: Bench to bedside and beyond. *Nucleic Acids Research* 35: 6227–6237.

ILLUMINA. 2010. Sequencing technology [online]. Illumina, San Diego, California, USA. Website http://www.illumina.com/technology/sequencing_technology.ilmn [accessed 1 June 2010].

JANSEN, R. K., L. A. RAUBESON, J. L. BOORE, C. W. DEPAMPHILIS, T. W. CHUMLEY, R. C. HABERLE, S. K. WYMAN, ET AL. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology* 395: 348–384.

JANZEN, D. H. 2004. Setting up tropical biodiversity for conservation through non-damaging use: Participation by parataxonomists. *Journal of Applied Ecology* 41: 181–187.

KING, I. 2009. The need for the incorporation of phylogeny in the measurement of biological diversity, with special reference to ecosystem functioning research. *BioEssays* 31: 107–116.

KRESS, W. J., AND D. L. ERICKSON. 2007. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the noncoding *trnH-psbA* spacer region. *PLoS ONE* 2: e508.

KRESS, W. J., AND D. L. ERICKSON. 2008. DNA barcodes: Genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences, USA* 105: 2761–2762.

KRESS, W. J., D. L. ERICKSON, F. A. JONES, N. G. SWENSON, R. PEREZ, O. SANJUR, AND E. BERMINGHAM. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences, USA* 106: 18621–18626.

KRESS, W. J., K. J. WURDACK, E. A. ZIMMER, L. A. WEIGT, AND D. H. JANZEN. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences, USA* 102: 8369–8374.

LAHAYE, R., M. VAN DER BANK, D. BOGARIN, J. WARNER, F. PUPULIN, G. GIGOT, O. MAURIN, ET AL. 2008. DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences, USA* 105: 2923–2928.

LANE, M. A., L. C. ANDERSON, T. M. BARKLEY, J. H. BOCK, E. M. GIFFORD, D. W. HALL, D. O. NORRIS, ET AL. 1990. Forensic botany: Plants, perpetrators, pests, poisons, and pot. *BioScience* 40: 34–39.

LI, R., W. FAN, G. TIAN, H. ZHU, L. HE, J. CAI, Q. HUANG, ET AL. 2010. The sequence and *de novo* assembly of the giant panda genome. *Nature* 463: 311–317.

LIPSCOMB, D., N. PLATNICK, AND Q. WHEELER. 2003. The intellectual content of taxonomy: A comment on DNA taxonomy. *Trends in Ecology & Evolution* 18: 65–66.

LISTER, R., B. D. GREGORY, AND J. R. ECKER. 2009. Next is now: New technologies for sequencing of genomes, transcriptomes, and beyond. *Current Opinion in Plant Biology* 12: 107–118.

LISTON, A. 2003. Genomics, bioinformatics, and plant systematics. *Israel Journal of Plant Sciences* 51: S-89–S-94.

LOWENSTEIN, J. H., G. AMATO, AND S.-O. KOLOKOTRONIS. 2009. Unknowingly consuming endangered tuna. Science Daily, 20 November 2009, website http://www.sciencedaily.com/releases/2009/11/091119135636.htm [accessed 22 June 2010].

MAGNUSON-FORD, K., A. MOOERS, S. RIOUX PAQUETTE, AND M. STEEL. 2010. Comparing strategies to preserve evolutionary diversity. *Journal of Theoretical Biology* 266: 107–116.

MAGURRAN, A. E. 1988. Ecological diversity and its measurement. Princeton University Press, Princeton, New Jersey, USA.

MARDIS, E. R. 2008. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* 9: 387–402.

MARGULIES, M., M. EGHOLM, W. E. ALTMAN, S. ATTIYA, J. S. BADER, L. A. BEMBEN, J. BERKA, ET AL. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.

MARSHALL, E. 2005. Will DNA bar codes breathe life into classification? *Science* 307: 1037.

MEIER, R., K. SHIYANG, G. VAIDYA, AND P. K. L. NG. 2006. DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology* 55: 715–728.

METZKER, M. L. 2010. Sequencing technologies—The next generation. *Nature Reviews Genetics* 11: 31–46.

MILLER, J. R., S. KOREN, AND G. SUTTON. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* 95: 315–327.

MOØERS, A. O., AND D. W. REDDING. 2009. Where the rare species are. *Molecular Ecology* 18: 3955–3957.

MORITZ, C., AND C. CICERO. 2004. DNA barcoding: Promise and pitfalls. *PLoS Biology* 2: e354.

MOROZOVA, O., AND M. A. MARRA. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92: 255–264.

MYERS, N., R. A. MITTERMELER, C. G. MITTERMELER, G. A. B. DA FONSECA, AND J. KENT. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.

NILSSON, S. G., J. HEDIN, AND M. NIKLASSON. 2001. Biodiversity and its assessment in boreal and nemoral forests. *Scandinavian Journal of Forest Research* 16 (supplement 3): 10–26.

NOCK, C. J., D. L. E. WATERS, M. A. EDWARDS, S. G. BOWEN, N. RICE, G. M. CORDEIRO, AND R. J. HENRY. In press. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal*.

PACKER, L., J. C. GRIXTI, R. E. ROUGHLEY, AND R. HANNER. 2009. The status of taxonomy in Canada and the impact of DNA barcoding. *Canadian Journal of Zoology* 87: 1097–1110.

PALMER, J. D. 1986. Isolation and structural analysis of chloroplast DNA. *Methods in Enzymology* 118: 167–186.

PETTERSSON, E., J. LUNDEBERG, AND A. AHMADIAN. 2009. Generations of sequencing technologies. *Genomics* 93: 105–111.

POP, M., AND S. L. SALZBERG. 2008. Bioinformatics challenges of new sequencing technology. *Trends in Genetics* 24: 142–149.

PRIMACK, R. B. 2008. A primer of conservation biology, 4th ed. Sinauer, Sunderland, Massachusetts, USA.

PURVIS, A., AND A. HECTOR. 2000. Getting the measure of biodiversity. *Nature* 405: 212–219.

RATAN, A. 2009. Assembly algorithms for next generation sequence data. Ph.D. dissertation, Pennsylvania State University, University Park, Pennsylvania, USA.

ROCHE. 2010. 454 Sequencing system portfolio [online]. Roche Diagnostics, Basel, Switzerland. Website http://454.com/products-solutions/454-sequencing-system-portfolio.asp [accessed 1 June 2010].

ROSAUER, D., S. W. LAFFAN, M. D. CRISP, S. C. DONNELLAN, AND L. G. COOK. 2009. Phylogenetic endemism: A new approach for identifying geographical concentrations of evolutionary history. *Molecular Ecology* 18: 4061–4072.

ROTHBERG, J. M., AND J. H. LEAMON. 2008. The development and impact of 454 sequencing. *Nature Biotechnology* 26: 1117–1124.

ROUNSLEY, S. D., AND R. L. LAST. 2010. Shotguns and SNPs: How fast and cheap sequencing is revolutionizing plant biology. *Plant Journal* 61: 922–927.

RUBINOFF, D., S. CAMERON, AND K. WILL. 2006. Are plant DNA barcodes a search for the Holy Grail? *Trends in Ecology & Evolution* 21: 1–2.

RUSK, N., AND V. KIERMER. 2008. Primer: Sequencing—The next generation. *Nature Methods* 5: 15.

SANGER, F., S. NICKLEN, AND A. R. COULSON. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, USA* 74: 5463–5467.

SASS, C., D. P. LITTLE, D. WM. STEVENSON, AND C. D. SPECHT. 2007. DNA barcoding in the Cycadales: Testing the potential of proposed barcoding markers for species identification of cycads. *PLoS ONE* 2: e1154.

SAVOLAINEN, V., R. S. COWAN, A. P. VOGLER, G. K. RODERICK, AND R. LANE. 2005. Towards writing the encyclopaedia of life: An introduction to DNA barcoding. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 360: 1805–1811.

SCHUSTER, S. C. 2008. Next-generation sequencing transforms today's biology. *Nature Methods* 5: 16–18.

SEBERG, O., C. J. HUMPHRIES, S. KNAPP, D. W. STEVENSON, G. PETERSEN, N. SCHARFF, AND N. M. ANDERSEN. 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends in Ecology & Evolution* 18: 63–65.

SHAFFER, C. 2007. Next-generation sequencing outpaces expectations. *Nature Biotechnology* 25: 149.

SHENDURE, J., AND H. JI. 2008. Next-generation DNA sequencing. *Nature Biotechnology* 26: 1135–1145.

SMITH, D. R., A. R. QUINLAN, H. E. PECKHAM, K. MAKOWSKY, W. TAO, B. WOOLF, L. SHEN, ET AL. 2008. Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Research* 18: 1638–1642.

SMITH, M. A., B. L. FISHER, AND P. D. N. HEBERT. 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: The ants of Madagascar. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 360: 1825–1834.

SOLTIS, P. S., AND M. A. GITZENDANNER. 1999. Molecular systematics and the conservation of rare species. *Conservation Biology* 13: 471–483.

SPOONER, D. M. 2009. DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *American Journal of Botany* 96: 1177–1189.

STEELE, P. R. 2010. Taxonomic revision of the neotropical genus *Psiguria* (Cucurbitaceae). *Systematic Botany* 35: 341–357.

STEELE, P. R., L. M. FRIAR, L. E. GILBERT, AND R. K. JANSEN. 2010. Molecular systematics of the neotropical genus *Psiguria* (Cucurbitaceae): Implications for phylogeny and species identification. *American Journal of Botany* 97: 156–173.

STOECKLE, M. Y., AND P. D. N. HEBERT. 2008. Barcode of life. *Scientific American* 299: 82–88.

SYTSMA, K. J., AND J. C. PIRES. 2001. Plant systematics in the next 50 years: Re-mapping the new frontier. *Taxon* 50: 713–732.

TAUTZ, D., P. ARCTANDER, A. MINELLI, R. H. THOMAS, AND A. P. VOGLER. 2002. DNA points the way ahead in taxonomy. *Nature* 418: 479.

TAUTZ, D., P. ARCTANDER, A. MINELLI, R. H. THOMAS, AND A. P. VOGLER. 2003. A plea for DNA taxonomy. *Trends in Ecology & Evolution* 18: 70–74.

VALENTINI, A., F. POMPANON, AND P. TABERLET. 2009. DNA barcoding for ecologists. *Trends in Ecology & Evolution* 24: 110–117.

VAN JAARSVELD, A. S., S. FREITAG, S. L. CHOWN, C. MULLER, S. KOCH, H. HULL, C. BELLAMY, ET AL. 1998. Biodiversity assessment and conservation strategies. *Science* 279: 2106–2108.

VANE-WRIGHT, R. I. 1996. Systematics and the conservation of biological diversity. *Annals of the Missouri Botanical Garden* 83: 47–57.

VERNOOY, R., E. HARIBABU, M. RUIZ MULLER, J. H. VOGEL, P. D. N. HEBERT, D. E. SCHINDEL, J. SHIMURA, AND G. A. C. SINGER. 2010. Barcoding life to conserve biological diversity: Beyond the taxonomic imperative. *PLoS Biology* 8: e1000417.

VOGLER, A. P., AND M. T. MONAGHAN. 2007. Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research* 45: 1–10.

VON BUBNOFF, A. 2008. Next-generation sequencing: The race is on. *Cell* 132: 721–723.

WHEELER, D. A., M. SRINIVASAN, M. EGHOLM, Y. SHEN, L. CHEN, A. MCGUIRE, W. HE, ET AL. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452: 872–876.

WILL, K. W., B. D. MISHLER, AND Q. D. WHEELER. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology* 54: 844–851.

WILL, K. W., AND D. RUBINOFF. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20: 47–55.

WITT, J. D. S., D. L. THRELOFF, AND P. D. N. HEBERT. 2006. DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: Implications for desert spring conservation. *Molecular Ecology* 15: 3073–3082.

WYMAN, S. K., R. K. JANSEN, AND J. L. BOORE. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.

YAHARA, T., M. DONOGHUE, R. ZARDOYA, D. P. FAITH, AND J. CRACRAFT. 2010. Genetic diversity assessments in the century of genome science. *Current Opinion in Environmental Sustainability* 2: 43–49.

ZERBINO, D. R., AND E. BIRNEY. 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn. *Genome Research* 18: 821–829.