Chapter 17

DNA BARCODES AND INSECT BIODIVERSITY

Robin M. Floyd, John J. Wilson, and Paul D. N. Hebert
Biodiversity Institute of Ontario and Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

I know this little thing A myriad men will save, O Death, where is thy sting? Thy victory, O Grave?
— Sir Ronald Ross (1857–1932)

Insect Biodiversity: Science and Society, 1st edition. Edited by R. Foottit and P. Adler
 ISBN 978-1-4051-5142-9

About 3500 species of mosquitoes (Diptera: Culicidae) have been described worldwide. In 1897, Ronald Ross, a Scottish physician working in India, discovered that only members of one mosquito genus, *Anopheles*, carry the *Plasmodium* parasite, the single-celled organism that causes malaria in humans. This revelation reflected painstaking efforts, involving the dissection of stomachs from vast numbers of mosquitoes. It was a key breakthrough that paved the way for Ross to demonstrate the life cycle of the parasite in the laboratory, work rewarded by the 1902 Nobel prize in Medicine. Unusually for a scientist, Ross was also a poet, playwright, and novelist; the preceding verse was written in response to this breakthrough in the understanding of malaria (Carey 1995).

Sadly, Ross's hope that this knowledge would quickly allow malarial control proved too optimistic; the disease still causes more than 1 million deaths per year, mainly in tropical Africa and Asia, despite numerous eradication efforts (Greenwood et al. 2005). This situation continues, in part, because the evolutionary dynamics of both *Plasmodium* and its insect vectors are far more complicated than initially realized. Only a limited number of species in the genus *Anopheles* transmit the agents of the human form of malaria. *Anopheles gambiae* (*sensu stricto*), the most important vector of the *Plasmodium* parasite in humans, belongs to a complex of morphologically indistinguishable sibling species that nevertheless differ markedly in their habitat preferences, behavior, and ability to transmit malarial agents (della Torre et al. 2002, Lehmann et al. 2003). Although these species are likely in the midst of speciation (a process expected to result in morphologically cryptic species complexes), they can be readily discriminated on the basis of their ribosomal DNA sequences (Masendu et al. 2004, della Torre et al. 2005, Guelbeogo et al. 2005). Plans are underway to control populations of *A. gambiae* by introducing transgenes, a strategy that will depend on knowledge of gene flow and population dynamics within and among these sibling species (Cohuet et al. 2005, Tripet et al. 2005).

The message from this story is clear: cryptic biological diversity matters. *Anopheles* serves as a pertinent example of the challenges faced by those concerned with biodiversity. Life exists in an immense number of forms, which are often tiny, difficult to study, and even more difficult to discriminate. Yet, this subtle variation can be crucially important; paraphrasing one article on the subject, what we do not know can hurt us (Besansky et al. 2003).

Insects constitute the most diverse group of animals on the planet, with more than 1 million described species (1,004,898; introduction to this volume) and millions more either awaiting description or simply undiscovered (Grimaldi and Engel 2005). They affect human society in myriad ways, both harmful (e.g., disease vectors, crop pests) and helpful (e.g., pollinators, biological control agents). Research of insects has added immensely to our understanding of evolution, ecology, and the genetic control of development. Yet, a fundamental requirement in gaining useful knowledge about any organismal group is the ability to describe, classify, and subsequently identify its member taxa.

Groups, such as insects, present great challenges to the taxonomic enterprise simply because of their diversity. The identification of species by traditional morphological methods is complex and usually requires specialist knowledge. The recognition, description, and naming of new species is more so; yet, the number of undescribed insect species far outweighs the number of taxonomic specialists (Grissell 1999), whose workforce is in decline (Godfray 2002). New approaches are needed to overcome this 'taxonomic impediment' (Weeks and Gaston 1997, Giangrande 2003). These concerns are not purely academic, but have significant practical implications. Agricultural pests cause immense damage. Total annual crop losses due to insect pests in North America have been estimated at US$7.5 billion and far more in the developing world (Yudelman et al. 1998), making it vital to quickly identify destructive species before invasions become uncontrollable. There is also a basic scientific need to describe biological diversity before the destruction of natural habitats by human activity causes the loss of species on a massive scale. We need a rapid way of assembling species catalogs, so that conservation programs can protect those areas of greatest importance before they are lost (Myers et al. 2000).

SPECIES CONCEPTS AND RECOGNITION

Although species have long been considered the basic 'units of biodiversity' (Claridge et al. 1997), and the only 'real' grouping in the taxonomic hierarchy, the issue of how best to delimit species remains controversial. Mayden (1997) listed 22 species concepts that have appeared in the literature (though some are essentially synonymous) and that employed varied criteria

from ecological niches, mate recognition, genetic cohesion, and evolutionary history. These diverse criteria necessarily lead to ambiguity, which can have important implications for studies of biodiversity and conservation, as differing species concepts can produce widely varying estimates of taxon richness (Agapow et al. 2004) Although reproductive isolation is often considered the most important indicator of species status, it is seldom directly tested and fails to address asexual organisms. In practice, most species continue to be recognized by the presence of one or more apparently fixed or nonoverlapping diagnostic differences (Davis and Nixon 1992). For most insect groups, detailed examination of genital morphology has represented the gold standard for species definition for nearly a century, due to the observation of a general phenomenon of rapid and pronounced divergence in the genitalia between species of animals (Eberhard 1985). Actual application of this criterion, however, is hampered by lack of an appropriate methodology to quantify shape variation (Arnqvist 1998) and by questionable homology assessments. All these factors collectively make species identification an extremely specialized and time-consuming science, and even expert taxonomists can have difficulty reaching consensus. Moreover, this reliance on diagnostic characters that are present only in the adult life stage creates a serious constraint on identification, as many specimens lack these characters (Balakrishnan 2005). The life-history stages most commonly intercepted at ports of entry are larvae and pupae (Scheffer et al. 2006), and damage to specimens collected in the field often makes identification difficult or impossible.

Another option exists – species can be diagnosed by the genetic changes that arise between reproductively isolated lineages as a result of genetic drift or selection. The use of DNA sequences to gain information about the taxonomic affinities of an unknown specimen saw its earliest adoption in the least morphologically tractable groups such as viruses and bacteria (Theron and Cloete 2000, Nee 2003). More recently, it has been applied to plants (Chase et al. 2005), to simple metazoan animals such as nematode worms (Floyd et al. 2002), and even to charismatic megafauna such as birds, fish, and mammals (Ward et al. 2005, Clare et al. 2007, Kerr et al. 2007). This approach relies on the use of algorithms enabling DNA-sequence comparison, such as Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990), in conjunction with DNA databases such as GenBank.

Many authors refer to 'operational taxonomic units' (OTUs) when delimiting taxa by purely phenetic or heuristic means (Sokal and Sneath 1963). OTUs may or may not correspond to species in the strict sense, but can be used in instances where speed and ease of application are of more practical importance than theoretical considerations (and if there are reasons to believe that theory and practicality are not directly in conflict). Taxa diagnosed or delimited by phenetic DNA-sequence divergences can be termed 'molecular operational taxonomic units' (MOTUs – Floyd et al. 2002, Blaxter et al. 2005). This approach has become the standard for environmental surveys of bacteria and other microorganisms, which could be seen as a capitulation to necessity, because these groups are virtually impossible to address in any other way (Hagström et al. 2002, Martinez et al. 2004, Hanage et al. 2005). However, correspondence between MOTUs and species can be examined in a number of ways. One approach, tested with Lepidoptera, is the correlation with previously unassociated morphological or ecological traits, for example, host plants and caterpillar phenotypes in *Astraptes fulgerator* (Hebert et al. 2004). Where morphological or ecological information is unavailable, a common situation in many taxonomic studies, congruence with an appropriate nuclear gene is an objective way to delineate interbreeding groups, and has been investigated in tropical beetles (Monaghan et al. 2005) and tachinid flies (Smith et al. 2006). Seven of the nine methods of delimiting species boundaries recently reviewed by Sites and Marshall (2003) require molecular data, which could imply that molecular markers are becoming increasingly important tools applied by taxonomists, possibly due to objectivity, speed, and increased discriminatory capacity.

DNA BARCODING

In this chapter, we deal with DNA barcoding, the use of short standardized genomic segments as markers for species identification. Just as species differ in morphology, ecology, and behavior, they also differ in their DNA sequences. Hence, at least in principle, a particular gene or gene fragment can be used to identify a given species in much the same way that retail barcodes uniquely identify each consumer product. In practice we would not expect DNA barcoding to work in such a simple manner – real DNA sequences are subject to all the natural complexities of molecular evolution, and can

show considerable variation within species (Mallet and Willmott 2003). They are not systematically 'assigned' to entities one by one as retail barcodes are. Nevertheless, if successful, DNA barcoding promises the ability to automate the identification of specimens by determining the sequence of the barcode region, avoiding the complexities inherent in morphological identifications, and prompting advocates to argue for the establishment of a system that ultimately might be applied to all life (Tautz et al. 2003, Blaxter 2004, Savolainen et al. 2005).

The particular genomic region used as a barcode is an important choice. It must be homologous between the organisms compared and have a rate of evolution fast enough to show variation between closely related species, and it also must have sufficient regions of sequence conservation to allow a limited set of PCR primers to amplify the target gene region from broad sections of the tree of life. The resultant sequence information also must generate a robust alignment so that sequences can be compared. In the animal kingdom, attention has been focused on a ~650 base-pair region near the 5′ end of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene (Hebert et al. 2003a). COI provides an ideal species-identification marker in insects, due to its lack of introns, simple alignment, limited exposure to recombination, and the availability of robust primer sites. Sequence variation in this region generally shows large interspecific, but small intraspecific, divergences, meaning that species frequently form clearly distinguishable clusters on a distance-based or phylogenetic tree. The homogenization of mitochondrial DNA sequences within a species, regardless of population size, is an intriguing phenomenon that has prompted study and speculation as to its evolutionary origin and significance (Bazin et al. 2006). The resulting 'barcoding gap' appears to represent a 'genetic signature' for species (Monaghan et al. 2005). Boundaries signaled by this molecular marker are strongly concordant with species units recognized through past studies of morphological and behavioral characters in a number of specific cases where they have been examined (Hebert et al. 2003a, 2003b).

Important advantages of a sequence-based approach to identification include the digital nature of a DNA sequence, which allows it to be gathered and interpreted objectively. Furthermore, DNA extracts from any life stage of an organism – egg, larva, or adult – or from fragments of dead material will generate a similar identification, whereas traditional identification keys (at least for insects that undergo complete metamorphosis from larval to adult forms) are often dependent on adult features. Additionally, social insects such as ants and termites often exhibit highly divergent caste morphologies that, in some cases, have been diagnosed incorrectly as distinct species – DNA barcoding promises to remove such ambiguities, allowing such forms to be associated (Smith et al. 2005). Sexual dimorphism, too, has long been a source of complications for taxonomists: Janzen et al. (2005) described a case where each sex of the butterfly *Saliana severus* was recorded as a separate species in an inventory, until barcoding revealed that males and females had the same COI sequences, and subsequently led to the recognition of a single, highly sexually dimorphic species.

Extraction and amplification of DNA from insects, including eggs and larvae, presents no technical challenge (Ball and Armstrong 2006). Recent advances in high-throughput DNA sequencing technology (Shendure et al. 2004) and reductions in costs (Hajibabaei et al. 2005) have made the generation of large volumes of DNA data straightforward. Sequences can be produced in the laboratory from a sample within a few hours in a largely automated fashion. While the 'Star Trek' vision of a handheld instant species-identification device (Janzen 2004, Savolainen et al. 2005) remains a speculative (yet attractive) notion for the future, promising advances have been made in reducing the size of the equipment needed to gather barcode data (Blazej et al. 2006). Sequencing is not always a necessary step for rapid identification with DNA, especially when analysis is focused on a small assemblage of closely allied species. Diagnostic restriction-digest enzyme patterns of cytochrome *b* PCR products were used to distinguish *Bombus ruderatus* and *Bombus hortorum*, two cryptic species of bumblebees, one in decline in the UK and illustrated with the same diagram of male genitalia in an identification key (Ellis et al. 2006). A simple PCR assay has been suggested as a molecular diagnostic tool for the swede midge (*Contarinia nasturtii*), an agricultural pest (Frey et al. 2004).

Full exploitation of DNA barcodes for species identification will be possible only after a comprehensive databank linking organisms and their sequences has been assembled (Savolainen et al. 2005). Databases presently have very uneven sequence coverage among taxa. Intensely studied groups and model organisms (e.g., *Drosophila melanogaster*) have many sequences or even entire genomes available. A few genes have been sequenced for many taxa, but the vast majority

of species lack any sequence data (Sanderson et al. 2003). This void raises the possibility that a poor match between a sequence derived from a newly encountered species and an incomplete reference library could result in spurious species diagnoses (Baker et al. 1996). Best BLAST hit, the simplest method of taxonomic assignment, is 'essentially useless' when no relatives have appropriate sequences in GenBank (Tringe and Rubin 2005). Tautz et al. (2003) suggested that an attempt be made to provide a DNA sequence as a component of all future species descriptions; the current barcoding initiatives could go a long way to bridge the gap, at least for major eukaryote groups. Some authors have argued that GenBank is unsuitable for taxonomic purposes due to its failure to include morphological, biogeographical, and ecological information associated with each sequence record (Tautz et al. 2003). However, the concept of 'type sequences' with voucher specimens authenticated by experts on the taxa and with associated taxonomic data is becoming reality. In 2004, the National Centre for Biotechnology Information (NCBI), GenBank's home organization, sealed a partnership with the Consortium for the Barcode of Life, whereby 'barcode standard' DNA sequences with relevant supporting data, including name of the identifier and collection location, can now be archived with the International Nucleotide Sequence Database Collaborative (Hanner 2005, Savolainen et al. 2005), with the keyword 'BARCODE' attached. This approach provides standardization of DNA regions, which previously was lacking and hindering progress in insect molecular systematics (Caterino et al. 2000).

The concept of DNA barcoding has been controversial in the taxonomic community (Moritz and Cicero 2004, Smith 2005). Criticisms of the approach have included questioning whether a single genetic marker has sufficient resolution to discriminate species reliably (Will and Rubinoff 2004, Will et al. 2005); potential problems caused by differing patterns of inheritance between nuclear and mitochondrial genes, which could confound the association between sequence and species (Funk and Omland 2003, Rubinoff 2006); and the feared marginalizing of morphological taxonomy (Lipscomb et al. 2003, Seberg et al. 2003). Other authors have emphasized the benefits of barcoding and DNA-assisted taxonomy in general (Tautz et al. 2002, Blaxter 2004, Hebert and Gregory 2005, Vogler and Monaghan 2006). As with any new concept or methodology in science, barcoding can be judged only by its success in facilitating new research and leading to new and useful knowledge. With the application to insects, and the endeavor to build a systematic database of 'DNA barcodes' linked to data about the species they represent, the barcoding movement has begun to gather real momentum (Hebert et al. 2003a). We, therefore, move to discuss some specific cases in which barcoding has been applied to particular insect groups, and examine how it has advanced our knowledge of biodiversity.

APPLICATIONS OF BARCODING

Lepidoptera

The Lepidoptera are a diverse and charismatic group of insects that have received significant taxonomic and systematic attention. One might think DNA barcoding has little to offer an order for which bright wing patterns and extensive, previous taxonomic attention suggest a group with a well-resolved species taxonomy. However, this perspective would be overly optimistic; approximately 165,000 species of Lepidoptera have been described, representing about 10% of the roughly 1.5 million known animal species (Wilson 2003). Another 150,000 to 1,250,000 species of Lepidoptera are thought to await description. These species do not all reside in hyperdiverse tropical settings; more than 10,000 species in Australia are still undescribed. Lack of taxonomists, problems with the way species are recognized, and extensive morphological convergence mean that most species are undescribed and numbers can only be estimated, particularly in the tropics.

Lepidoptera have now become the model group for barcoding studies since Hebert et al. (2003a) used North American moths to demonstrate the ability of COI to discriminate among specimens of different species. Since then, research with Lepidoptera has demonstrated the potential of molecular diagnostics and practical applications of DNA barcoding. Barcodes have enabled the linking of the varied life stages of the Lepidoptera, as well as the males and females of sexually dimorphic species (Janzen et al. 2005). This advancement is particularly relevant in the identification of pest and invasive species, as many are intercepted as eggs or larvae (Ball and Armstrong 2006).

One of the most significant potential uses of DNA barcoding lies in facilitating biodiversity surveys. Lepidoptera are a model group in ecology and a 'flagship' group for invertebrate conservation. Macromoths and

butterflies, in particular, have been used to indicate environmental quality (e.g., habitat degradation), to partition habitat diversity, and as indicators of climate change (Scoble 1992). Their role as model organisms in surveys, however, is limited by lack of taxonomic support. Barcoding could provide a new level of efficiency and comparability to ecological surveys, with DNA barcode records enabling more relevant and meaningful correlation of studies carried out by different experts at different locations and times, rather than the use of arbitrary morphological designators such as Noctuid sp. 01.

Work on the Neotropical skipper *Astraptes fulgerator* provides a prime example of the way in which DNA barcoding can aid species discovery, especially when coupled with morphological and ecological studies. Barcoding of 484 specimens from the Area Conservación Guanacaste (ACG) in Costa Rica revealed that the *A. fulgerator* group comprises a complex of sister species, confirming and extending earlier suspicions gained through studies of adult morphology and larval morphologies. Hebert et al. (2004) hypothesized 'ten species in one' based on COI divergences and association with caterpillar morphology and food plant. Brower (2006) reanalyzed the original DNA dataset under a different framework and also concluded that the *Astraptes fulgerator* sample contained multiple species but was critical of the methods used in the original study. However, other investigators who reanalyzed the same dataset supported the conclusion of 10 taxa (Nielsen and Matz 2006). The ideal framework for the use of barcodes in species delineation requires further research. This example, however, demonstrates the power of the DNA sequences themselves, once submitted to publicly available databases as unambiguous digital data immune from subjective assessment and open to repeated analysis and testing of the species and phylogenetic hypotheses generated.

Barcoding studies on the lepidopteran fauna from one region (the ACG) of Costa Rica are now well underway. Hajibabaei et al. (2006) sequenced more than 4000 individuals from 521 species belonging to 3 families (Hesperiidae, Saturniidae, and Sphingidae) and found that 97.9% of the individuals could be identified to species based on COI divergence patterns. The expanded ACG project now has the goal of barcoding every species of butterfly and moth in the preserve (about 9600 species) within 3 years, representing the first large-scale regional barcoding project. Because parallel initiatives (www.lepbarcoding.org) seek to barcode all Lepidoptera from two continents (North America, Australia), and all species in two families (Saturniidae and Sphingidae), the Lepidoptera are poised to become the first 'barcode-complete' order of insects. The realization of this goal will not only bring many advances in barcode data collection and analysis, but also provide a newly detailed framework for species delineation and research in molecular evolution.

Diptera

The flies (Diptera) constitute another hyperdiverse insect order, with around 150,000 described species (Grimaldi and Engel 2005, Beutel and Pohl 2006). Among insects, their members have the greatest negative impact on human health and livestock, with groups such as mosquitoes and tse tse acting as vectors of the agents for several major diseases including malaria, sleeping sickness, and filariasis (Yeates and Wiegmann 2005). Even before the establishment of DNA barcoding per se, many molecular diagnostic tools had been applied to the identification of mosquito species, including allozyme electrophoresis (Green et al. 1992), DNA hybridization (Beebe et al. 1996), and restriction fragment length polymorphism (RFLP) (Fanello et al. 2002). Sequencing-based approaches have also been used extensively, albeit mainly focusing on genes other than COI (e.g. Kent et al. 2004, Marrelli et al. 2005, Michel et al. 2005). However, a number of recent studies have shown that the standard COI barcode marker also serves effectively for species-level discrimination in surveys of Canadian (Cywinska et al. 2006) and Indian (Kumar et al. 2007) mosquitoes. Foley et al. (2007) constructed a molecular phylogeny of the Australian *Anopheles annulipes* species complex based on four different loci, both nuclear and mitochondrial (COI, COII, ITS2, and EF-1α). Despite using a shorter fragment of COI (258 bp) than the standard barcode region (658 bp), it was found in this study that 11 of the 17 sibling species (65%) had unique COI sequences, and the authors concluded that 'DNA barcoding holds some promise for diagnosing species within the Annulipes Complex, and perhaps for other anophelines'.

One of the earliest applications of a DNA-based approach to species identification involved fly species important to forensic science. Blow flies (Calliphoridae) and flesh flies (Sarcophagidae) lay eggs on corpses shortly after death. Because each species has a timeframe for development from egg to adult, the

particular life stage associated with a corpse can provide key evidence in determining time of death (postmortem interval, or PMI) (Smith 1986, Catts and Haskell 1990). However, because different species of flies have different development rates, accurate species identifications are necessary to make an accurate estimate of the PMI. Because only adults can be placed reliably to species, maggots previously had to be collected and reared to adults, constituting a significant time delay to the process (Nelson et al. 2007). Forensic entomologists were quick to realize the potential of DNA-based methods to distinguish species from any life stage and from dead, preserved material. As a result, extensive literature now exists, detailing how DNA sequences (mainly COI) can accurately discriminate fly species of forensic importance (Sperling et al. 1994, Malgorn and Coquoz 1999, Vincent et al. 2000, Wallman and Donnellan 2001, Wells and Sperling 2001, Wells et al. 2001).

Leafmining flies (family Agromyzidae) are economically important agricultural pests whose periodic population outbreaks are capable of destroying entire crops, particularly potatoes (Shepard et al. 1998). They are also a group for which considerable information on species limits is available (Scheffer and Wiegmann 2000). COI sequences were generated from 258 individuals belonging to three species of invasive leafminers in the Phillipines: *Liriomyza huidobrensis*, *L. trifolii*, and *L. sativae* (Scheffer et al. 2006). As is commonly observed in introduced or invasive populations, fewer mitochondrial haplotypes were found than in the endemic ranges of these species, and those seen were often highly divergent even within a species. This pattern is due to population bottlenecks that tend to occur during introduction (Nei et al. 1975), an effect that is particularly relevant for a marker such as mitochondrial DNA, which is both haploid and maternally inherited. Sequence analysis was able to place all specimens in the correct morphospecies as currently diagnosed. This study also illustrated some of the complexities that the barcoding approach must take into account. Certain mitochondrial sequences in both the *L. trifolii* and *L. sativae* groups were sufficiently divergent that they might suggest new, cryptic species, but no data other than COI-sequence divergence supported this conclusion. Based on existing knowledge of this group, these species are expected to contain highly divergent mitochondrial lineages; therefore, depending on which reference sequences were used, barcoding might overestimate the number of species present. Future research, however, possibly will reveal that these divergent lineages do represent distinct species. Although some ambiguity might be associated with barcoding in complex cases, these cases normally should be possible to resolve by combining COI data with information from other sources, such as morphology, behavior, or complementary DNA regions.

Insect parasitoids are not only a major component of global biodiversity, but also have significant demographic effects on their host species. Parasitoids also conceal a large diversity of morphologically cryptic species, distinguished by strong host specificity (Godfray 1994). Flies of the family Tachinidae are endoparasitoids of other insects, often lepidopteran larvae. A recent study of the tachinid genus *Belvosia* from northwestern Costa Rica examined their diversity by rearing specimens from wild-caught caterpillars, recording their morphology, and sequencing their COI genes (Smith et al. 2006). DNA sequences were able not only to discriminate 17 known host-specific species of the genus *Belvosia*, but also raised the number of species to 32 by revealing that 3 species, each believed to be host generalists, were complexes of highly host-specific cryptic species. Again, this study illustrates the power of DNA barcoding to reveal unknown diversity in morphologically difficult groups.

Finally, the dipteran family Chironomidae (nonbiting midges) is a species-rich group whose freshwater larval stages are often used in environmental monitoring. However, the connection of larval stages to known species (whose descriptions are mainly based on adult morphology) is a difficult challenge; but DNA barcoding has helped to address this problem in recent studies (Ekrem et al. 2007, Pfenninger et al. 2007). The former paper nevertheless cautions us that in order to use barcoding as a tool to identify unknown individuals by their COI sequence, a comprehensive library of known sequences is necessary for such identifications to be reliable.

Coleoptera

In a famous, though possibly apocryphal, incident, when geneticist J. B. S. Haldane was asked what the study of nature revealed about the mind of God, he answered: 'an inordinate fondness for beetles'. One out of every five animals on the planet is thought to be a beetle. As a consequence, the Coleoptera

represent a group where the taxonomic enterprise has been overwhelmed by diversity. Although 350,000 species of beetles have been described (including many economically important pest species), as many as 5–8 million might exist in total. With so many unknown species, major barcoding research with beetles has focused on the use of DNA-based methods in species discovery and delineation.

Monaghan et al. (2005) used the 3′ end of COI and the nuclear gene 28S rRNA to identify clusters of beetles in dung beetles of the genus *Canthon* and water beetles of the family Hydrophilidae. An exact match of nuclear genotypes and mitochondrial clusters suggested that the mtDNA groupings were not misleading due to introgression, and the clusters likely correlated with previously described or undescribed species. The results indicated that COI provides a largely accurate picture of species boundaries in these two beetle groups and provides validation for its use in species discovery.

Barcoding research with water beetles was continued by Monaghan et al. (2006), using the genus *Copelatus* from Fiji. Four DNA markers (three mitochondrial regions: COI, cytochrome b, and 16S rRNA, and the nuclear histone 3 gene) were sequenced for 118 specimens from 20 islands. This effort was seen as a particularly challenging test case for barcoding because many lineages on oceanic islands have undergone rapid 'radiations', resulting in large numbers of recently diverged species with complex gene histories. Beetle taxa were clustered using the concatenated DNA sequences and separately with traditional morphological methods (i.e., male genital morphology). Although the clustering pattern was largely incongruent using the two approaches, the authors concluded that if the morphological approach had been followed with a Linnaean system of naming, it would have formalized, at best, a partial taxonomic resolution, with limited evolutionary understanding of lineage diversification (Monaghan et al. 2006). The morphological approach is time intensive and requires specialized knowledge of character differences associated with species-level classification. Subsequent identification of the 'species', using morphology, thus can be problematic due to ambiguous descriptions and difficulties obtaining the type specimens, the situation encountered with the five formerly described species of *Copelatus* from Fiji. The sequencing approach, combined with phylogenetic analysis, provides an extensive summary of evolutionary history, and once sequences have been submitted to databases (in this case, EMBL), the data are easily accessible and subsequent repeat analyses can be done by anyone. Monaghan et al. (2006) suggest DNA sequences themselves could constitute a system of taxonomic grouping and communication without need of a formal Linnaean classification. The study suggests DNA sequencing could make the task of global species classification achievable when standard morphological methods are inadequate or too time consuming.

Hymenoptera

With about 125,000 described species, the Hymenoptera are the fourth largest insect order after Coleoptera, Lepidoptera, and Diptera (Grimaldi and Engel 2005, Beutel and Pohl 2006). Given the number of cryptic species suspected to exist, its true species richness might even surpass the 'big three' (Grissell 1999).

Ants (Formicidae) constitute the major component of arthropod biomass in many of the world's ecosystems. They are important in nutrient recycling, and their activities within soil create widely varying nutrient microhabitats influencing plant succession, growth, and distribution (Hölldobler and Wilson 1990). In Madagascar, the ant fauna represents a hyperdiverse group, currently estimated to include about 1000 species, of which 96% are thought to be endemic. However, only 25% of this estimated total has been described, presenting a major obstacle to studying their biogeography, conservation status, and roles in ecosystem processes. A recent case study (Smith et al. 2005) examined the question of whether DNA barcoding could act as an effective surrogate for morphological species identifications. A total of 280 specimens from four localities were collected and independently identified to morphospecies and sequenced for COI. The specimens were classified both into MOTUs based on their sequence data and morphospecies based on their morphological traits, allowing the two methods to be directly compared. Additionally, two different sequence-divergence thresholds (2% and 3%) were tested for MOTU assignment.

Although instances of incongruities occurred between the molecular and morphological taxon assignments, strong correlations, nevertheless, existed between the two. A total of 90 morphospecies, 117 3% MOTU and 126 2% MOTU were found. Morphological species designations, therefore, tended

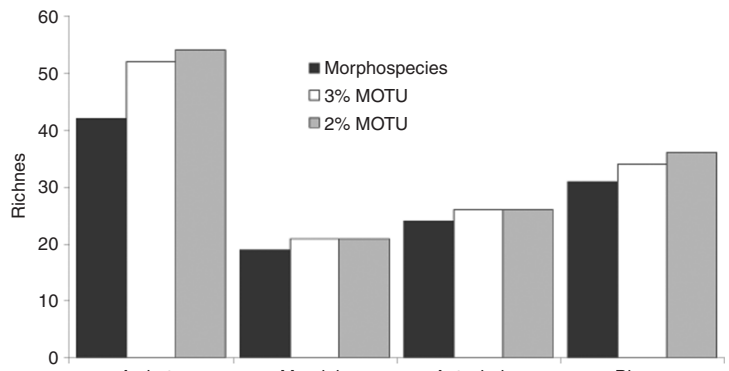

Fig. 17.1 Estimates of taxonomic richness (based on both MOTU and morphospecies) from a survey of ants across four sites in Madagascar (from Smith et al. 2005, used with permission).

to lump specimens that were split by the molecular approach. As in many cases, molecular markers detect cryptic taxa that are difficult or impossible to detect by morphology alone. These cryptic taxa may or may not correspond to true 'species', but they form a starting point for further investigation of their status. After examining the patterns of taxon richness across the four sites, no significant differences occurred between the data shown by MOTUs and morphospecies (Fig. 17.1). Additionally, whether MOTUs were defined on the basis of 2% or 3% divergence altered only the absolute number of taxa delineated, and did not make a significant difference to the overall patterns of diversity observed. This finding is important because it suggests that MOTUs can be used as effective surrogates for traditional species – although they will not necessarily delineate exactly the same taxonomic groupings, they will identify the same general patterns such as the most versus least diverse sites. Studies that delineate taxa by DNA-sequence arrays alone would enable surveys across much larger geographical regions and taxonomic groups than would be possible if the slow and laborious process of morphological identification were required, without losing resolution or information content.

Collembola

Springtails (Collembola) are not true insects but are basal members of the same superclass, Hexapoda (Grimaldi and Engel 2005). Phylogenetically, they are potentially important in unraveling the relationships of higher taxa (Mayhew 2002). They are among the most diverse and numerically abundant of all soil arthropods (Petersen and Luxton 1982) and have the widest distribution of any hexapod group, occurring throughout the world, including Antarctica. There are about 7000 known species, but many more likely remain undiscovered (Hopkin 1997). In particular, the Arctic regions appear to have a vast uncataloged diversity (Danks 1981). As is typical for such groups, a great diversity combined with difficulty of identification and a global lack of taxonomic specialists creates a severe impediment to understanding their diversity.

Springtails show all the hallmarks of a group for which DNA barcoding could prove highly informative. A recent study (Hogg and Hebert 2004) tested whether COI was able to resolve species differences among a set of Collembola sampled from the Canadian Arctic. In all cases – 19 species in 13 genera – COI sequences were able to discriminate species, with between-species divergences above 8% in all cases and within-species divergences generally below 1%. The single exception to this pattern was that several individuals identified as *Folsomia quadrioculata* showed divergences of up to 13%, likely representing a case of an undescribed and morphologically cryptic sister species, which is a well-known phenomenon among Collembola (Stevens and Hogg 2003).

Ephemeroptera

Mayflies (Ephemeroptera) are an insect order whose larval stages develop in freshwater habitats. They are important in aquatic research, particularly in

biomonitoring of water quality: the particular species composition of mayfly and other insect larvae are useful indicators of chemical pollution in rivers (Lenat and Resh 2001). Identifications are often problematic, however, as frequently only larvae are available, and species-level identification keys normally depend on adult features. A DNA-based system allowing identification from any life stage, therefore, would be highly beneficial.

One recent study applied the standard COI barcoding method to a test set of Ephemeroptera specimens (Ball et al. 2005). Sequences were generated from 150 individuals – initially 80 reference specimens that were used to create a profile matching sequences to named species, followed by a further 70 specimens that were used to test whether the correct species assignments could be made on the basis of their COI sequences. All but one of the 70 test specimens were correctly identified, with a mean sequence divergence within species of 1% and mean divergence among congeneric species an order of magnitude greater (18%). The sole exception was an individual identified morphologically as *Maccaffertium modestum*, which showed deep genetic divergence from other *M. modestum* specimens, again suggesting an undescribed sister species.

CONCLUSIONS

Taxonomy is the framework by which we name and classify biological diversity into the groupings used in all areas of biology. As a component of modern systematic science, taxonomy seeks to recognize natural evolutionary groupings – those that are monophyletic. This effort has led to countless revisions as different systematists uncover new data or interpret character distributions in different ways, but despite years of attention, the monophyly of many insect groups remains questionable. Even in intensively studied groups, the evolutionary history is unresolved and the Linnaean hierarchy only adds to the confusion. DNA barcodes represent data points able to be integrated into the traditional Linnaean system (Dayrat 2005), yet at the same time independent from it, for accumulating ecological, geographical, morphological, and other data about organisms. Once submitted to online databases, nucleotide sequences represent a freely available taxonomic resource that allows species recognition to be accomplished in a uniform manner by nonexperts. Barcoding has the potential to become a universal communication tool in a way that complicated and often incomprehensible morphological descriptions cannot be, especially in developing countries where the majority of biodiversity resides (Agosti 2003).

Barcoding need not be restricted to a single gene region. From the point of view of both economics (sequencing is still relatively expensive, though costs are dropping yearly) and simplicity of use, a system based on a single sequencing 'read' per specimen could be established. In some instances, this single marker will fail to discriminate taxa, most often when dealing with recently diverged sister species, which are the most difficult to discriminate in any system (Mallet and Willmott 2003, Hickerson et al. 2006, Meier et al. 2006, Whitworth et al. 2007). In such problematic cases, a single sequence will narrow the options to a small number of closely related taxa, and additional sequence or other data can be added to provide species-level resolution. One of the most important scientific outcomes of large-scale barcoding initiatives will be the production of a library of genomic DNA extracts from archived voucher specimens, which can serve as a basis for numerous future lines of research besides the generation of the initial barcode sequence. Commentators who have criticized barcoding on the basis of its costs (Cameron et al. 2006) have generally ignored such collateral benefits of the research.

Sequence information is easy to obtain, unambiguous, and makes species identification possible by nonspecialists unfamiliar with the intricacies of morphology. MOTU, good species or not, depending on the species concept applied, nevertheless, can be a suitable surrogate for identifying units of diversity in biodiversity studies. This approach enables users to obtain the information much faster than with the traditional morphological taxonomic process, making surveys scalable across much larger taxonomic groupings and wider geographical regions (Smith et al. 2005). Yet, the appeal of barcoding stems not only from speed and operationality; it reflects the increasingly held view that DNA-sequence analysis is as appropriate a mechanism for recognizing and delimiting evolutionary units as morphological comparisons. Although it does not automatically follow from this premise that a barcoding system based on single-gene comparisons will always delimit species-level groups, the studies we have cited offer evidence that it often will be the case.

Support for a large-scale barcoding initiative has grown rapidly; in 2004, the Consortium for the Barcode

of Life (CBOL) was established to act as a central organizing body for the barcoding effort (www.barcoding.si.edu). Based at the Smithsonian Institution in Washington, CBOL represents an international collaboration of more than 120 organizations, including many prominent museums. National organizations, such as the Canadian Barcode of Life Network (Dooh and Hebert 2005), have established specimen supply chains and centralized facilities for the generation and analysis of sequence data, including such tools as the Barcode of Life Data Systems (BOLD; www.barcodinglife.org), a central repository for barcode records in conjunction with various analytical tools (Ratnasingham and Hebert 2007). The present total of all insect barcode sequences deposited in BOLD currently stands at 206,434 (including those gathered from GenBank), from 26,262 different species, but most are in the process of final taxonomic validation. As a simple illustration of diversity within insect COI sequences, Fig. 17.2 shows a neighbor-joining tree of sequences for a selection of 4675 validated and published records.

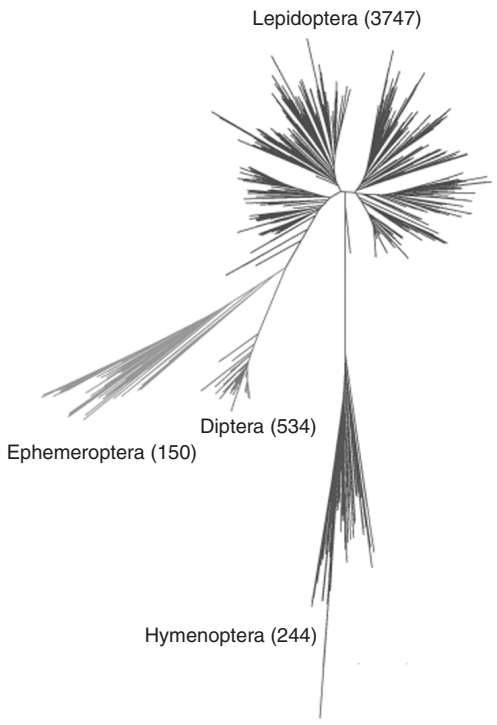

Fig. 17.2 A neighbor-joining tree, based on Kimura 2-parameter distance, of 4675 cytochrome *c* oxidase I sequences from across the class Insecta.

To describe approximately 1.5 million species, using traditional approaches, has taken taxonomy two centuries. DNA-assisted species discovery has the potential to rapidly accelerate this process, an advantage that cannot be ignored in the light of the current biodiversity crisis affecting our planet (Eldredge 1992). By allowing more rapid detection and monitoring of agricultural pests and disease vectors, the pragmatic significance of barcoding can hardly be over emphasized. Perhaps more importantly in the long term, barcoding promises a blossoming of 'bioliteracy' (Janzen 2004) by shifting the accessibility of taxonomic knowledge from the realm of the specialist into the wider public domain. DNA sequencing technology is still relatively expensive and hence accessible only to well-funded labs, mainly in the developed world. Like any emerging technology, it is expected to become cheaper, faster, and simpler in the future, as has been the case with personal computers, GPS units, and mobile phones. One can envisage a time when a handheld DNA barcoding device allows any curious child to scan some interesting organism and gain immediate access to a library of information – not only the organism's name, but also its biology, ecology, conservation status, and more. Human beings preserve those things we value, and we can only value those things we perceive; for biological diversity to become something valued by all, it must be made visible and understandable in all its complexity – a goal for which barcoding can play a significant role in making reality.

ACKNOWLEDGMENTS

This contribution was supported through funding to the Canadian Barcode of Life Network from Genome Canada (through the Ontario Genomics Institute), NSERC, and other sponsors listed at www.bolnet.ca. We are grateful to Gregory Downs, Robert Dooh, and Sujeevan Ratnasingham for IT/bioinformatics support. We thank Robert Hanner, M. Alex Smith, and two anonymous reviewers for constructive comments on drafts of the manuscript; M.A. Smith also kindly provided the data used in Fig. 17.1.

REFERENCES

Agapow, P. M., O. R. Bininda-Emonds, K. A. Crandall, J. L. Gittleman, G. M. Mace, J. C. Marshall, and A. Purvis.

2004. The impact of species concept on biodiversity studies. *Quarterly Review of Biology* 79: 161–179.

Agosti, D. 2003. Encyclopaedia of life: should species description equal gene sequence? *Trends in Ecology and Evolution* 18: 273–273.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Arnqvist, G. 1998. Comparative evidence for the evolution of genitalia by sexual selection. *Nature* 393: 784–786.

Baker, C. S., F. Cipriano, and S. R. Palumbi. 1996. Molecular genetic identification of whale and dolphin products from commercial markets in Korea and Japan. *Molecular Ecology* 5: 671–685.

Balakrishnan, R. 2005. Species concepts, species boundaries and species identification: a view from the tropics. *Systematic Biology* 54: 689–693.

Ball, S. L. and K. F. Armstrong. 2006. DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera : Lymantriidae). *Canadian Journal of Forest Research* 36: 337–350.

Ball, S. L., P. D. N. Hebert, S. K. Burian, and J. M. Webb. 2005. Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *Journal of the North American Benthological Society* 24: 508–524.

Bazin, E., S. Glemin, and N. Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science* 312: 570–572.

Beebe, N. W., D. H. Foley, R. D. Cooper, J. H. Bryan, and A. Saul. 1996. DNA probes for the *Anopheles punctulatus* complex. *American Journal of Tropical Medicine and Hygiene* 54: 395–398.

Besansky, N. J., D. W. Severson, and M. T. Ferdig. 2003. DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. *Trends in Parasitology* 19: 545–546.

Beutel, R. G. and H. Pohl. 2006. Endopterygote systematics – where do we stand and what is the goal (Hexapoda, Arthropoda)? *Systematic Entomology* 31: 202–219.

Blaxter, M., J. Mann, T. Chapman, F. Thomas, C. Whitton, R. Floyd, and E. Abebe. 2005. Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society. B. Biological Sciences* 360: 1935–1943.

Blaxter, M. L. 2004. The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society. B. Biological Sciences* 359: 669–679.

Blazej, R. G., P. Kumaresan, and R. A. Mathies. 2006. Microfabricated bioprocessor for integrated nanoliter-scale Sanger DNA sequencing. *Proceedings of the National Academy of Sciences USA* 103: 7240–7245.

Brower, A. V. Z. 2006. Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). *Systematics and Biodiversity* 4: 127–132.

Cameron, S., D. Rubinoff and K. Will. 2006. Who will actually use DNA barcoding and what will it cost? *Systematic Biology* 55: 844–847.

Carey, J. 1995. *The Faber Book of Science*. Faber and Faber, London.

Caterino, M. S., S. Cho, and F. A. H. Sperling. 2000. The current state of insect molecular systematics: A thriving Tower of Babel. *Annual Review of Entomology* 45: 1–54.

Catts, E. P. and N. H. Haskell. 1990. *Entomology and Death: A Procedural Guide*. Joyce's Print Shop, Clemson, SC.

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. B. Biological Sciences* 360: 1889–1895.

Clare, E. L., B. K. Lim, M. D. Engstrom, J. L. Eger, and P. D. N. Hebert. 2007. DNA barcoding of Neotropical bats: species identification and discovery within Guyana. *Molecular Ecology Notes* 7: 184–190.

Claridge, M. F., H. A. Dawah, and M. R. Wilson (eds). 1997. *Species: The Units of Biodiversity*. Chapman and Hall, London.

Cohuet, A., I. Dia, F. Simard, M. Raymond, F. Rousset, C. Antonio-Nkondjio, P. H. Awono-Ambene, C. S. Wondji, and D. Fontenille. 2005. Gene flow between chromosomal forms of the malaria vector *Anopheles funestus* in Cameroon, Central Africa, and its relevance in malaria fighting. *Genetics* 169: 301–311.

Cywinska, A., F. F. Hunter, and P. D. N. Hebert. 2006. Identifying Canadian mosquito species through DNA barcodes. *Medical and Veterinary Entomology* 20: 413–424.

Danks, H. V. 1981. *Arctic Arthropods : A Review of Systematics and Ecology with Particular Reference to the North American Fauna*. Entomological Society of Canada, Ottawa.

Davis, J. I. and K. C. Nixon. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology* 41: 421–435.

Dayrat, B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society* 85: 407–415.

della Torre, A., Z. J. Tu, and V. Petrarca. 2005. On the distribution and genetic differentiation of *Anopheles gambiae* s.s. molecular forms. *Insect Biochemistry and Molecular Biology* 35: 755–769.

della Torre, A., C. Costantini, N. J. Besansky, A. Caccone, V. Petrarca, J. R. Powell, and M. Coluzzi. 2002. Speciation within *Anopheles gambiae* – the glass is half full. *Science* 298: 115–117.

Dooh, R. and P. D. N. Hebert. 2005. The Canadian Barcode of Life Network. http://www.bolnet.ca.

Eberhard, W. G. 1985 *Sexual Selection and Animal Genitalia*. Harvard University Press, Cambridge, Massachusetts.

Eldredge, N. (ed). 1992. *Systematics, Ecology, and the Biodiversity Crisis*. Columbia University Press, New York.

Ellis, J. S., M. E. Knight, C. Carvell, and D. Goulson. 2006. Cryptic species identification: a simple diagnostic tool for

discriminating between two problematic bumblebee species. *Molecular Ecology Notes* 6: 540–542.

Ekrem, T., E. Willassen, and E. Stur. 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution* 43: 530–542.

Fanello, C., F. Santolamazza, and A. della Torre. 2002. Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Medical and Veterinary Entomology* 16: 461–464.

Floyd, R., E. Abebe, A. Papert, and M. Blaxter. 2002. Molecular barcodes for soil nematode identification. *Molecular Ecology* 11: 839–850.

Foley, D. H., R. C. Wilkerson, R. D. Cooper, M. E. Volovsek, and J. H. Bryan. 2007. A molecular phylogeny of *Anopheles annulipes* (Diptera: Culicidae) sensu lato: the most species-rich anopheline complex. *Molecular Phylogenetics and Evolution* 43: 283–297.

Frey, J. E., B. Frey, and R. Baur. 2004. Molecular identification of the swede midge (Diptera: Cecidomyiidae). *Canadian Entomologist* 136: 771–780.

Funk, D. J. and K. E. Omland. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics* 34: 397–423.

Giangrande, A. 2003. Biodiversity, conservation, and the 'taxonomic impediment'. *Aquatic Conservation: Marine and Freshwater Ecosystems* 13: 451–459.

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

Godfray, H. C. J. 1994. *Parasitoids: Behavioral and Evolutionary Ecology*. Princeton University Press, Princeton, New Jersey.

Green, C. A., L. E. Munstermann, S. G. Tan, S. Panyim, and V. Baimai. 1992. Population genetic evidence for species-A, species-B, species-C and species-D of the *Anopheles dirus* complex in Thailand and enzyme electromorphs for their identification. *Medical and Veterinary Entomology* 6: 29–36.

Greenwood, B. M., K. Bojang, C. J. Whitty, and G. A. Targett. 2005. Malaria. *Lancet* 365: 1487–1498.

Grimaldi, D. A. and M. S. Engel. 2005. *Evolution of the Insects*. Cambridge University Press, Cambridge.

Grissell, E. E. 1999. Hymenopteran biodiversity: some alien notions. *American Entomologist* 45: 235–244.

Guelbeogo, W. M., O. Grushko, D. Boccolini, P. A. Ouedraogo, N. J. Besansky, N. F. Sagnon, and C. Costantini. 2005. Chromosomal evidence of incipient speciation in the Afrotropical malaria mosquito *Anopheles funestus*. *Medical and Veterinary Entomology* 19: 458–469.

Hagström, A., T. Pommier, F. Rohwer, K. Simu, W. Stolte, D. Svensson, and U. L. Zweifel. 2002. Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. *Applied and Environmental Microbiology* 68: 3628–3633.

Hajibabaei, M., D. H. Janzen, J. M. Burns, W. Hallwachs, and P. D. N. Hebert. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences USA* 103: 968–971.

Hajibabaei, M., J. R. de Waard, N. V. Ivanova, S. Ratnasingham, R. T. Dooh, S. L. Kirk, P. M. Mackie, and P. D. N. Hebert. 2005. Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society. B. Biological Sciences* 360: 1959–1967.

Hanage, W., C. Fraser, and B. Spratt. 2005. Fuzzy species among recombinogenic bacteria. *BMC Biology* 3: 6.

Hanner, R. 2005. Proposed Standards for BARCODE Records in INSDC (BRIs). Consortium for the Barcode of Life. http://www.barcoding.si.edu/PDF/DWG_data_standards-Final.pdf.

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. 2004. 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., A. Cywinska, S. L. Ball, and J. R. deWaard. 2003a. Biological identifications through DNA barcodes. *Proceedings of the Royal Society. B. Biological Sciences* 270: 313–321.

Hebert, P. D. N., S. Ratnasingham, and J. R. deWaard. 2003b. Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society. B. Biological Sciences* 270 (Supplement): S96–S99.

Hickerson, M. J., C. P. Meyer, and C. Moritz. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology* 55: 729–739.

Hogg, I. D. and P. D. N. Hebert. 2004. Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. *Canadian Journal of Zoology* 82: 749–754.

Hölldobler, B. and E. O. Wilson. 1990. *The Ants*. Belknap Press of Harvard University Press, Cambridge, Massachusetts.

Hopkin, S. P. 1997. *Biology of the Springtails (Insecta, Collembola)*. Oxford University Press, Oxford.

Janzen, D. 2004. Now is the time. *Philosophical Transactions of the Royal Society. B. Biological Sciences* 359: 731–732.

Janzen, D. H., M. Hajibabaei, J. M. Burns, W. Hallwachs, E. Remigio, and P. D. N. Hebert. 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philosophical Transactions of the Royal Society. B. Biological Sciences* 360: 1835–1845.

Kent, R. J., A. J. West, and D. E. Norris. 2004. Molecular differentiation of colonized human malaria vectors by 28S ribosomal DNA polymorphisms. *American Journal of Tropical Medicine and Hygiene* 71: 514–517.

Kerr, K. C. R., M. Y. Stoeckle, C. J. Dove, L. A. Weigt, C. M. Francis, and P. D. N. Hebert. 2007. Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes* 7: 535–543.

Kumar, N. P., A. R. Rajavel, R. Natarajan, and P. Jambulingam. 2007. DNA barcodes can distinguish species of Indian mosquitoes (Diptera: Culicidae). *Journal of Medical Entomology* 44: 1–7.

Lehmann, T., M. Licht, N. Elissa, B. T. A. Maega, J. M. Chimumbwa, F. T. Watsenga, C. S. Wondji, F. Simard, and W. A. Hawley. 2003. Population structure of *Anopheles gambiae* in Africa. *Journal of Heredity* 94: 133–147.

Lenat, D. R. and V. H. Resh. 2001. Taxonomy and stream ecology – the benefits of genus- and species-level identifications. *Journal of the North American Benthological Society* 20: 287–298.

Lipscomb, D., N. Platnick, and Q. Wheeler. 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends in Ecology and Evolution* 18: 65–66.

Malgorn, Y. and R. Coquoz. 1999. DNA typing for identification of some species of Calliphoridae. An interest in forensic entomology. *Forensic Science International* 102: 111–119.

Mallet, J. and K. Willmott. 2003. Taxonomy: renaissance or Tower of Babel? *Trends in Ecology and Evolution* 18: 57–59.

Marrelli, M. T., L. M. Floeter-Winter, R. S. Malafronte, W. P. Tadei, R. Lourenco-de-Oliveira, C. Flores-Mendoza, and O. Marinotti. 2005. Amazonian malaria vector anopheline relationships interpreted from ITS2 rDNA sequences. *Medical and Veterinary Entomology* 19: 208–218.

Martinez, J., L. Martinez, M. Rosenblueth, J. Silva, and E. Martinez-Romero. 2004. How are gene sequence analyses modifying bacterial taxonomy? The case of *Klebsiella*. *International Microbiology* 7: 261–268.

Masendu, H. T., R. H. Hunt, J. Govere, B. D. Brooke, T. S. Awolola, and M. Coetzee. 2004. The sympatric occurrence of two molecular forms of the malaria vector *Anopheles gambiae* Giles sensu stricto in Kanyemba, in the Zambezi Valley, Zimbabwe. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 98: 393–396.

Mayden, R. L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. Pp. 381–424. *In* M. F. Claridge, H. A. Dawah, and M. R. Wilson (eds). *Species: The Units of Biodiversity*. Chapman and Hall, London.

Mayhew, P. J. 2002. Shifts in hexapod diversification and what Haldane could have said. *Proceedings of the Royal Society. B. Biological Sciences* 269: 969–974.

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

Michel, A. P., W. M. Guelbeogo, O. Grushko, B. J. Schemerhorn, M. Kern, M. B. Willard, N. Sagnon, C. Costantini, and N. J. Besansky. 2005. Molecular differentiation between chromosomally defined incipient species of *Anopheles funestus*. *Insect Molecular Biology* 14: 375–387.

Monaghan, M. T., M. Balke, J. Pons, and A. P. Vogler. 2006. Beyond barcodes: complex DNA taxonomy of a South Pacific Island radiation. *Proceedings of the Royal Society. B. Biological Sciences* 273: 887–893.

Monaghan, M. T., M. Balke, T. R. Gregory, and A. P. Vogler. 2005. DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philosophical Transactions of the Royal Society. B. Biological Sciences* 360: 1925–1933.

Moritz, C. and C. Cicero. 2004 DNA barcoding: promise and pitfalls. *PLoS Biology* 2: e354.

Myers N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.

Nee, S. 2003. Unveiling prokaryotic diversity. *Trends in Ecology and Evolution* 18: 62–63.

Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29: 1–10.

Nielsen, R. and M. Matz. 2006. Statistical approaches for DNA barcoding. *Systematic Biology* 55: 162–169.

Nelson, L. A., J. F. Wallman, and M. Dowton. 2007. Using COI barcodes to identify forensically and medically important blowflies. *Medical and Veterinary Entomology* 21: 44–52.

Petersen, H. and M. Luxton. 1982. A comparative analysis of soil fauna populations and their role in decomposition processes. *Oikos* 39: 287–388.

Pfenninger, M., C. Nowak, C. Kley, D. Steinke, and B. Streit. 2007. Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic *Chironomus* (Diptera) species. *Molecular Ecology* 16: 1957–1968.

Ratnasingham, S. and P. D. N. Hebert. 2007. The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes* 7: 355–364.

Rubinoff, D. 2006. Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology* 20: 1026–1033.

Sanderson, M. J., A. C. Driskell, R. H. Ree, O. Eulenstein, and S. Langley. 2003. Obtaining maximal concatenated phylogenetic data sets from large sequence databases. *Molecular Biology and Evolution* 20: 1036–1042.

Savolainen, V., R. S. Cowan, A. P. Vogler, G. K. Roderick, and R. Lane. 2005. Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society. B. Biological Sciences* 360: 1805–1811.

Scheffer, S. J., M. L. Lewis, and R. C. Joshi. 2006. DNA barcoding applied to invasive leafminers (Diptera: Agromyzidae) in the Philippines. *Annals of the Entomological Society of America* 99: 204–210.

Scheffer, S. J. and B. M. Wiegmann. 2000. Molecular phylogenetics of the holly leafminers (Diptera: Agromyzidae: *Phytomyza*): species limits, speciation, and dietary specialization. *Molecular Phylogenetics and Evolution* 17: 244–255.

Scoble, M. J. 1992. *Lepidoptera: Form, Function, and Diversity*. Oxford University Press, Oxford.

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 and Evolution* 18: 63–65.

Shendure, J., R. D. Mitra, C. Varma, and G. M. Church. 2004. Advanced sequencing technologies: methods and goals. *Nature Reviews Genetics* 5: 335–344.

Shepard, B. M., Samsudin, and A. R. Braun. 1998. Seasonal incidence of *Liriomyza huidobrensis* (Diptera : Agromyzidae) and its parasitoids on vegetables in Indonesia. *International Journal of Pest Management* 44: 43–47.

Sites, J., W. Jack, and J. C. Marshall. 2003. Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology and Evolution* 18: 462–470.

Smith, K. G. V. 1986. *Manual of Forensic Entomology*. British Museum (Natural History), London.

Smith, M. A., N. E. Woodley, D. H. Janzen, W. Hallwachs, and P. D. N. Hebert. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences USA* 103: 3657–3662.

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. B. Biological Sciences* 360: 1825–1834.

Smith, V. S. 2005. DNA barcoding: perspectives from a ''Partnerships for Enhancing Expertise in Taxonomy'' (PEET) debate. *Systematic Biology* 54: 841–844.

Sokal, R. R. and P. H. A. Sneath. 1963. *Principles of Numerical Taxonomy*. W. H. Freeman and Co, San Francisco, California.

Sperling, F. A., G. S. Anderson, and D. A. Hickey. 1994. A DNA-based approach to the identification of insect species used for postmortem interval estimation. *Journal of Forensic Sciences* 39: 418–427.

Stevens, M. I. and I. D. Hogg. 2003. Long-term isolation and recent range expansion from glacial refugia revealed for the endemic springtail *Gomphiocephalus hodgsoni* from Victoria Land, Antarctica. *Molecular Ecology* 12: 2357–2369.

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 and Evolution* 18: 70–74.

Theron, J. and T. E. Cloete. 2000. Molecular techniques for determining microbial diversity and community structure in natural environments. *Critical Reviews in Microbiology* 26: 37–57.

Tringe, S. G. and E. M. Rubin. 2005. Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics* 6: 805–814.

Tripet, F., G. Dolo, and G. C. Lanzaro. 2005. Multilevel analyses of genetic differentiation in *Anopheles gambiae* s.s. reveal patterns of gene flow important for malaria-fighting mosquito projects. *Genetics* 169: 313–324.

Vincent, S., J. M. Vian, and M. P. Carlotti. 2000. Partial sequencing of the cytochrome oxydase b subunit gene I: a tool for the identification of European species of blow flies for postmortem interval estimation. *Journal of Forensic Sciences* 45: 820–823.

Vogler, A. and M. T. Monaghan. 2006. Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research,* published online, doi: 10.1111/j.1439–0469.2006.00384.x.

Wallman, J. F. and S. C. Donnellan. 2001. The utility of mitochondrial DNA sequences for the identification of forensically important blowflies (Diptera: Calliphoridae) in southeastern Australia. *Forensic Science International* 120: 60–67.

Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. N. Hebert. 2005. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society. B. Biological Sciences* 360: 1847–1857.

Weeks, P. J. D. and K. J. Gaston. 1997. Image analysis, neural networks, and the taxonomic impediment to biodiversity studies. *Biodiversity and Conservation* 6: 263–274.

Wells, J. D., T. Pape, and F. A. Sperling. 2001. DNA-based identification and molecular systematics of forensically important Sarcophagidae (Diptera). *Journal of Forensic Sciences* 46: 1098–1102.

Wells, J. D. and F. A. Sperling. 2001. DNA-based identification of forensically important Chrysomyinae (Diptera: Calliphoridae). *Forensic Science International* 120: 110–115.

Whitworth, T. L., R. D. Dawson, H. Magalon, and E. Baudry. 2007. DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proceedings of the Royal Society. B. Biological Sciences* 274: 1731–1739.

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

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.

Wilson, E. O. 2003. The encyclopedia of life. *Trends in Ecology and Evolution* 18: 77–80.

Yeates, D. K. and B. M. Wiegmann. 2005. *Evolutionary Biology of Flies*. Columbia University Press, New York.

Yudelman, M., A. Ratta, and D. F. Nygaard. 1998. *Pest Management and Food Production: Looking to the Future*. International Food Policy Research Institute, Washington, DC.