Review

CellPress

Status and prospects of DNA barcoding in medically important parasites and vectors

Danielle A. Ondrejicka*, Sean A. Locke*, Kevin Morey, Alex V. Borisenko, and Robert H. Hanner

Biodiversity Institute of Ontario, University of Guelph, 50 Stone Road East, Guelph, Canada N1G 2W1

For over 10 years, DNA barcoding has been used to identify specimens and discern species. Its potential benefits in parasitology were recognized early, but its utility and uptake remain unclear. Here we review studies using DNA barcoding in parasites and vectors affecting humans and find that the technique is accurate (accords with author identifications based on morphology or other markers) in 94–95% of cases, although aspects of DNA barcoding (vouchering, marker implicated) have often been misunderstood. In a newly compiled checklist of parasites, vectors, and hazards, barcodes are available for 43% of all 1403 species and for more than half of 429 species of greater medical importance. This is encouraging coverage that would improve with an active campaign targeting parasites and vectors.

Use of molecular data to distinguish species of parasites and vectors

Over 1 billion people currently suffer from a neglected tropical disease, which in most cases is caused by a parasite [1,2]. Accurate identification of parasites and vectors is key to improving detection and monitoring and to understanding the epidemiological and ecological characteristics of parasitic diseases. However, morphological discrimination of most parasite and many vector species is notoriously difficult. Both parasites and arthropod vectors (see Glossary) are often small and possess strongly dissimilar stages in their life cycles and many lack diagnostic morphological characters (e.g., [3,4]). This complicates both morphological identification and understanding the links between developmental stages found either in different host species or in the environment. For these reasons, molecular data are widely used to complement traditional morphological approaches [3,5].

Because of the wide range of taxa that cause and transmit disease to humans, DNA sequences are commonly used to identify specimens and delineate species, but different markers and genes are often used for different groups of parasites and vectors (e.g., [6]). For example, noncoding spacer regions between ribosomal subunits are often used to differentiate among helminth species, whereas cytochrome *b* is often used for haemosporidians. In some groups, dedicated online resources exist to store, compare, and analyze these data (e.g., http://mbio-serv2.mbioekol.lu.se/Malavi/ [7], http://eupathdb.org [8]). However, the use of these tools and related resources is predicated on *a priori* knowledge of the higher taxonomy of the specimens being identified. This situation begged the development of DNA barcoding, a large-scale, standardized approach to the molecular characterization of biodiversity to aid identification where specialist knowledge may be unavailable. Although several techniques have been referred to as 'barcoding' [9], this term usually refers to the approach proposed by Hebert *et al.* [10,11], currently adopted by International Barcode of Life (iBOL) (http://ibol.org). In most eukaryotes, a DNA barcode is a sequence of approximately 650 nucleotides at the 5′ end of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene from a specimen vouchered in an appropriate collection facility (Box 1). The standardization implicit in barcoding – a single tool applicable to all taxa, sequences linked to physical specimens – is of obvious potential utility in parasitology. Just 10 months after DNA barcoding with COI was first proposed, the case for applying it to parasites and vectors of medical importance was made by Besansky and colleagues [12]. They saw DNA barcoding neither as a displacement of morphology nor as a method that will work in all cases. Instead, barcoding can provide identifications and, in undescribed taxa, a proxy for species-level delineations that are more closely tied to the natural entities that cause and transmit disease [12].

Glossary

Barcode Index Number (BIN): an operational taxonomic unit assigned to a group of similar DNA barcode sequences through the sequence clustering method built into BOLD [44] and used as a proxy for species-level identification.
Hazard: an arthropod that commonly stings, bites, or secretes toxic substances and causes harm.
Morphological voucher: a preserved specimen archived in a collection facility (e.g., a museum). In DNA barcoding, vouchering (the preservation of morphological vouchers) is standard practice for specimens from which DNA barcode sequences were obtained.
Vector: an organism that transmits pathogens.

Corresponding author: Locke, S.A. (salocke@uoguelph.ca, seanlocke@gmail.com).
Keywords: specimen identification; species delineation; pathogens; endoparasites and ectoparasites; vectors; neglected tropical disease.
*These authors contributed equally.

1471-4922/
 http://dx.doi.org/10.1016/j.pt.2014.09.003

Box 1. DNA barcodes and BOLD

DNA barcoding is the use of a short, standardized DNA sequence for specimen identification and species delineation and discovery [10]. In animals, the standard DNA barcode is a 658-bp fragment at the 5′ end of the mitochondrial COI gene, often referred to as the 'Folmer' [96] or 'barcode' region. DNA barcoding is based on the observation that the COI sequence variation between species is typically far greater than that occurring within them. As a result, species can be distinguished based on unique clusters of barcode sequences generated using efficient, scalable algorithms [44,97]. The barcode region is flanked by conserved fragments where broad-spectrum primers (e.g., [96]) can be used to amplify the target marker.

The BOLD system is a free, publicly accessible repository for barcode sequences, particularly COI [98]. Each BOLD record comprises two components linked to a single specimen: the sequence record and the corresponding specimen record. The sequence record contains standard DNA barcode marker information, (i.e., the COI nucleotide sequence) as well as chromatograms, quality scores, primers used, and corresponding data from additional markers. The specimen data record contains taxonomic identification of the specimen from which the tissue was sourced, detailed provenance information, museum collection and voucher catalog numbers, digital specimen images, and personal attributions (e.g., collector, identifier, photographer). For parasites too small to be subsampled, an aliquot of extracted DNA may serve as a molecular voucher and a similar specimen from the same host (paragenophere, *sensu* [99]) can be used as a morphological voucher.

BOLD also offers user-friendly tools for the curation, management, and dissemination of DNA barcode data, such as distance-based modules for statistical analysis (e.g., tree building, minimum distance to specimens in other species), data aggregators (e.g., distribution maps, accumulation curves, virtual datasets), automated GenBank submission, and an online identification engine for querying unknown barcode sequences [98]. The recently introduced BIN algorithm [44] (Box 2) assigns COI sequences to operational taxonomic units, providing a framework for dealing with samples not identified to the species level.

The quality of records deposited in BOLD is automatically evaluated. Those with over 500 bp of COI sequence, trace files, a voucher reference and a full set of provenance information are considered 'barcode compliant', form the core reference library, and are submitted to GenBank with the keyword 'BARCODE' [100]. Sequences of lower quality or lacking information, as well those mined from GenBank, are used to broaden the comparative context of the BOLD identification engine. BOLD data are continually automatically screened for sequence quality and errors, and inconsistencies requiring curatorial input (e.g., misidentifications, contamination) can also be flagged by BOLD users and data managers. Overall, BOLD structure and workflows provide an iterative framework for validating and cross-referencing data across multiple projects, empowering users to collaborate on improving data quality and standardization.

A decade later, barcode coverage was quantitatively evaluated in animals [13,14] and functional groups such as agricultural pests [15]. Fišer Pečnikar and Buzan [16] reviewed applications of DNA barcoding in vectors and parasites but both the coverage and the utility of this technique remain unclear in these organisms, partly because no species checklist exists against which to benchmark progress. Here we address these two gaps by quantifying barcode success in the medical parasitology literature and barcode coverage against a newly compiled checklist of medically important parasites and vectors. While it is not our primary aim to discuss the merits of DNA barcoding, which have been thoroughly debated (reviewed in [14,17–19]), we revisit some of its potential applications in medical parasitology, most of which were foreseen by Besansky *et al.* [12] and speculate how these may fit with rapidly changing sequencing technologies.

How useful has DNA barcoding been in medical parasitology?

A search of the literature yielded 83 empirical studies using DNA barcoding in organisms that affect human health, as defined in our checklist of parasites, vectors, and hazards (see below), which are listed in Table S1A in the supplementary material online [two searches: March 2014 – Web of Science, Thomson Reuters, topic string: "DNA barcod*" AND (parasit* OR vector*); July 2014 – Google Scholar search string: parasite OR vector "DNA barcode"].

The use of DNA barcoding is growing in medical parasitology (Figure 1) but critical aspects concerning its scope (Box 1) have received less attention. For example, only 31/83 studies mentioned morphological or DNA vouchers, which in many cases were deposited in university collections of uncertain accessibility and longevity. Thirteen additional studies using 'DNA barcoding' employed downstream regions of COI, other mitochondrial markers, or ribosomal markers (Table S1B in the supplementary material online). DNA barcoding has been unevenly applied among the functional groups in parasitology. Endoparasites were the subject of 18 studies, which were generally of small scope (mean number of species studied = 5.6, median = 2) and most mainly concerned wildlife parasites that infect humans only rarely. Notably absent is any empirical barcoding work on protist parasites. By contrast, the 46 studies of arthropod vectors tended to include data from four times as many species (mean number of species studied = 23.1, median = 5) as studies of endoparasites. Most (39/46) studies of arthropod vectors seem motivated at least in part by medical importance, as they concern mosquitoes (Culicidae), sand flies (Phlebotominae), and black flies (Simuliidae) in regions where these insects transmit major disease.

We evaluated the utility of DNA barcoding in 60 of the 83 studies in which authors identified specimens and discriminated between at least two species that were on the checklist or otherwise identified as medically important. In the remaining studies (seven new-locality records, seven vector blood-meal analyses, two medical diagnoses, two new-species descriptions, and five others), barcode sequences were used to search public databases without independent estimates of accuracy, or data were otherwise not relevant, and barcoding success was therefore not assessed in these studies.

In all but one of the 60 studies in which species were discriminated using DNA barcodes, additional data were used to verify identifications as indicated by reference to morphology (e.g., citation of taxonomic keys, sequencing of museum specimens, character descriptions; 50/60 studies) and by the use of one to five additional molecular markers (29/60 studies). In 52 (87%) of these 60 studies, error rates were less than 10%, meaning that $\geq$90% of DNA barcode groupings matched species considered valid by the authors and that $\geq$90% of specimens were assigned to correct species by barcodes. While informative, this metric does

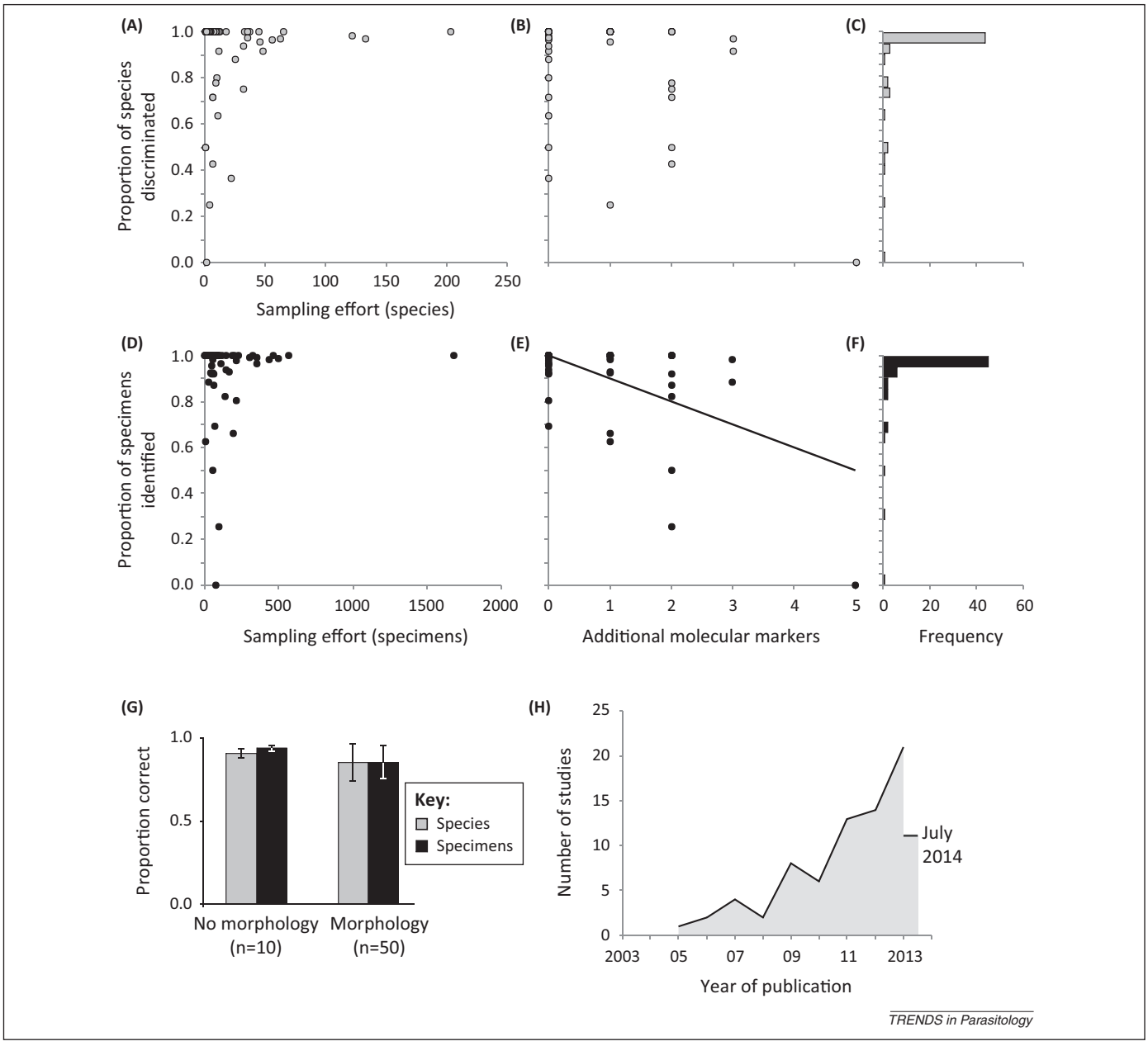

Figure 1. DNA barcoding accuracy and publication rate. The proportion of species correctly distinguished (i.e., matching the conclusions of the study authors) by DNA barcodes is plotted against the number of species sampled **(A)** and the number of molecular markers used in addition to the barcode region of cytochrome oxidase I **(B)** in 60 studies. The frequency of studies with different proportions of correctly identified species is shown in **(C)**. The corresponding plots of the proportions of specimens assigned to the correct species by DNA barcodes are below **(D–F)**. The mean proportions of species correctly distinguished and of specimens assigned to the correct species by DNA barcodes are compared in studies differing by whether they reported using morphology **(G)**; bars, standard error of the mean. The number of studies using DNA barcodes in medically important parasite and vector species is plotted against year of publication **(H)**.

not take into account wide variation in the numbers of specimens and species analyzed. We corrected for this by summing the numbers of species and specimens across studies and calculating proportions of correct identifications from these totals (where concordance with the taxonomic conclusions of the study authors was interpreted as taxonomic accuracy). For example, if 8/10 species in study A and 70/90 species in study B are distinguishable by DNA barcodes, the overall DNA barcode success rate is 0.78. We did not control for the fact that some studies sampled the same species and sequences or for the inclusion of species of no medical importance. The total numbers of species and specimens are therefore somewhat inflated, but proportions of taxonomic accuracy are not. Although studies dealt mainly with closely related species (i.e., congeners), some also included taxonomically distant organisms. Because such comparisons are not strong tests of DNA barcoding as a tool for species discrimination, we excluded specimens and species that were clearly outgroups in phylogenetic analyses. In cases of ambiguity, a conservative approach was taken. For example, it was unclear whether one or two species of *Ascaris* were sampled by Betson *et al.* [20], and the barcode success rate for this study was scored as 0/2 species and 0/78 specimens.

Overall, barcodes provided 'correct' discriminations in 1202/1275 (94%) species and 8553/8985 (95%) specimens. The proportion of species correctly distinguished with DNA barcodes was unrelated to both the number of molecular markers used and the number of species sampled ($P \geq 0.129$) (Figure 1). The proportion of specimens correctly identified was smaller in studies that used more molecular markers (rho = -0.251, $P = 0.024$) and unrelated to the number of specimens analyzed ($P = 0.226$). In studies that used morphology, barcode results matched author conclusions in a smaller proportion of species (mean $_{\text{morphology}} = 85\%$, mean $_{\text{no morphology}} = 91\%$) and specimens (mean $_{\text{morphology}} = 86\%$, mean $_{\text{no morphology}} = 94\%$) than in studies that did not mention morphology, but the differences were not significant (two-tailed t-tests, $P \geq 0.432$; Figure 1).

In several studies there were clear mismatches between barcode results and author conclusions based on morphology and/or other molecular markers. However, even in studies with such problems [21,22], COI provided accurate information (i.e., matched the authors' taxonomic conclusions) in over half of the specimens or species studied. Authors were generally hesitant to accept implications of DNA barcode data in cases of conflict with patterns seen in other molecular markers [23–26]. These cases could be due to nuclear copies of mitochondrial DNA, symbiont DNA, or haplotype sharing among species resulting from incomplete lineage sorting, recent divergence, or hybridization. Although these phenomena have occasionally been observed in relevant taxa [27–30], the overall concordance of DNA barcode results with the conclusions of authors employing independent data suggests that they may be relatively rare.

How many parasites and vectors have been DNA barcoded?

We sought to quantify DNA barcode coverage in parasite and arthropod vector species but found no comprehensive species list against which to perform a gap analysis. We therefore compiled a checklist that we compared with the Barcode of Life Data (BOLD) system (http://www.boldsystems.org), where COI sequences from 2 215 607 individual eukaryotes had been deposited by March 2014, including representatives of over 135 000 species. Initially, our checklist included 1341 species listed in [31–39] and only arthropods among the vectors, ectoparasites, and hazards listed in these sources. Relevant leeches and horsehair worms were subsequently added on the suggestion of an anonymous reviewer (62 species, based on [40,41]). We also added snail hosts of some trematodes (e.g., *Schistosoma*) commonly studied in efforts to reduce transmission. The final list of 1403 species (Table S2 in the supplementary material online) does not include viruses, prions, bacteria, or fungi nor does it include all eukaryotic parasites of humans. Nonetheless, it provides an informative gauge of barcode coverage in eukaryotes that affect human health and serves as a baseline for compiling additional species that should be given priority for barcoding. Nomenclature was validated using the Catalogue of Life (http://www.catalogueoflife.org), the Global Biodiversity Information Facility Checklist Bank (http://tools.gbif.org/namefinder/), and, in some cases, the primary literature.

Table 1. Summary of DNA barcode coverage of medically important species of parasites, vectors, and hazards at March 2014

	Total	Implicated or congeners of species implicated in major disease
Number of species	1403	429
Species with 1–9 DNA barcodes	313	103
Species with >10 DNA barcodes	284	116
Species with >1 public sequence	484	178
Species in which all data were mined from GenBank	217	80
Number of DNA barcode sequences	30 775	15 500

Of the 18 phyla represented in the checklist, the most prevalent were Arthropoda (65% of species), Platyhelminthes (15%), and Nematoda (9%). Ectoparasites comprised 35% of species, vectors 33%, endoparasites 30%, and other hazardous organisms 19% (some arthropod species occur in multiple categories; e.g., both ectoparasite and vector). In total, 30 775 COI barcode sequences >500 nucleotides in length were unevenly distributed among these organisms, with no data available for 57% of species (Table 1 and Figure 2). Barcodes have been obtained from more than ten specimens in 20% of all species. The most intensely sequenced organisms were mosquitoes: *Aedes vexans* (Meigen) (1985 COI sequences); *Culex quinquefasciatus* Say (1686 COI sequences); *Coquillettidia perturbans* (Walker) (1005 COI sequences); and *Aedes trivittatus* (Coquillet) (987 COI sequences). In 36% of the species in which at least one specimen has been barcoded, all data were mined from GenBank. While these sequences represent the barcode region of COI, they typically lack the associated specimen and sequence data necessary to receive annotation with the reserved keyword BARCODE by the International Nucleotide Sequence Database Collaboration (INSDC) (Box 1). Of 372 species in which some sequences were originally published on BOLD, no sequences have yet been made public in nearly a third (113 species).

Many species in the checklist have little medical relevance, for example, having been recorded in humans just once. We assessed barcode coverage in a subset of 429 species of vectors and parasites implicated in malaria and other major diseases (defined as neglected tropical diseases by the World Health Organization (WHO) [1] and Fenwick [2] or considered notable by Liu [42]). Half (50%) of all sequences were obtained from these more medically important species (Table 1), although they represent less than a third of the checklist. Species coverage is also higher in this subset: barcodes are available from 52% of these 'core' species. In just over a third of these cases, all sequences represented in BOLD were mined from GenBank.

As seen in the review of the literature, fewer barcode sequences exist for endoparasites (Figure 3) and sampling intensity is also generally low in this group. For example,

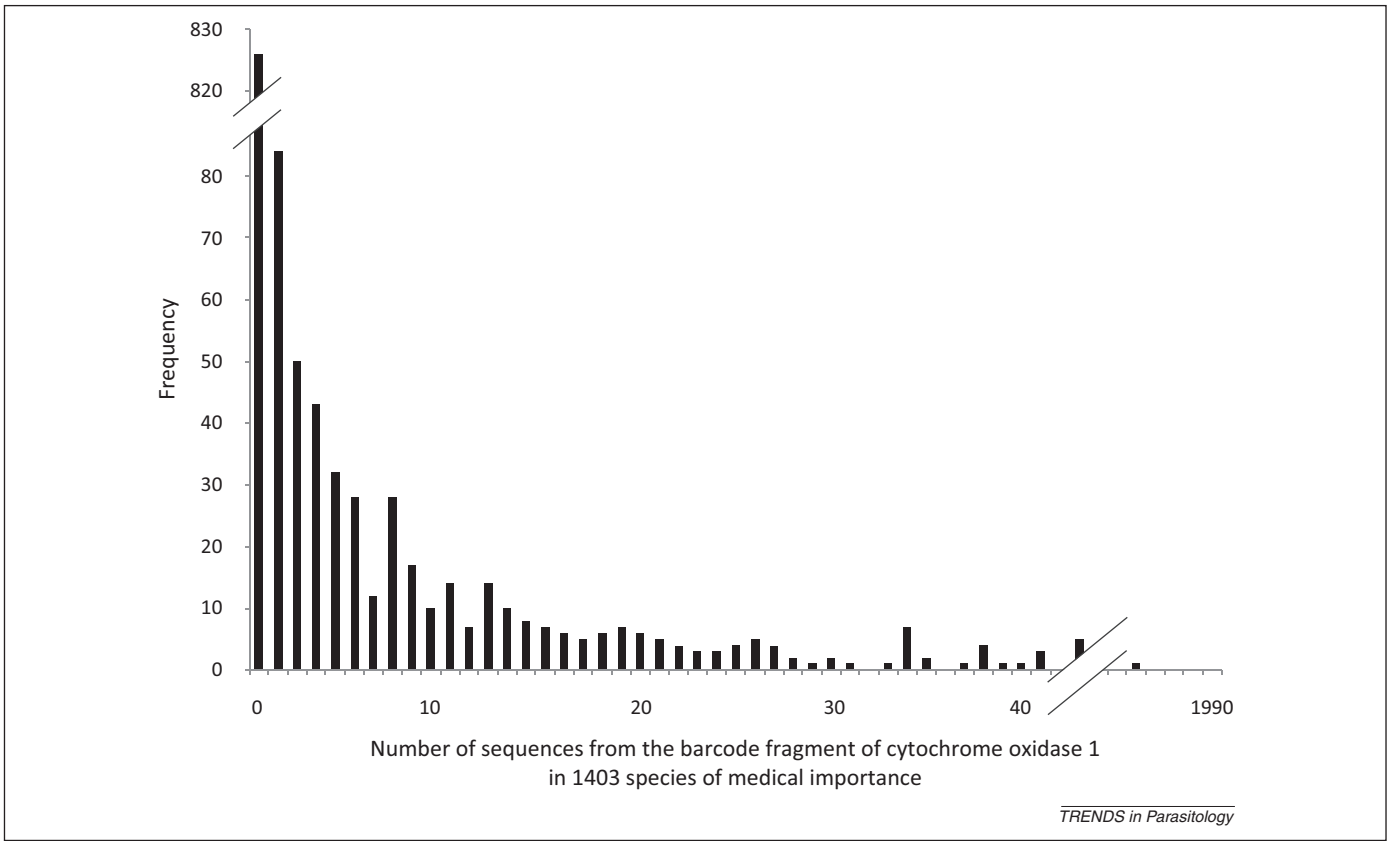

Figure 2. Distribution of 1403 species of animals and apicomplexans associated with disease in humans according to the number of barcode sequences available.

only three species of *Schistosoma* are represented by more than ten sequenced specimens and few barcodes have been obtained from several major soil-transmitted and food-borne helminths [*Ascaris lumbricoides* (L.) (four barcode sequences), *Trichuris trichiura* (L.) (two), *Necator americanus* (Stiles) (three), *Ancylostoma duodenale* (Dubini) (three), *Clonorchis sinensis* (Cobbold) (three), *Fasciola hepatica* (L.) (four)].

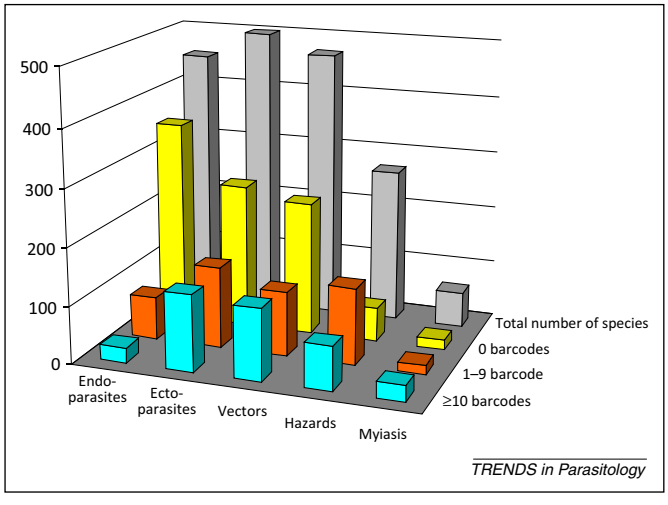

Figure 3. Number of barcode sequences in 1403 medically important animal and apicomplexan species according to functional groups.

How does barcode coverage of parasites and vectors compare with other taxonomic and functional groups?

In 2012, Frewin *et al.* [15] found DNA barcodes for 54% of 1044 agricultural pest species of quarantine significance, which is better than the 43% coverage for 1403 medically important species in early 2014. In our dataset, more species are represented solely by GenBank-mined data (42% of sequenced species) compared with agricultural pests (33%). This is important because GenBank sequences do not always meet standards for barcode compliance (Box 1), limiting their utility in diagnostic applications relying on taxonomically authenticated reference data. Also in 2012, Boykin *et al.* [43] found that 82% of 88 invasive pest species have records in BOLD. Other illuminating comparisons are made with iBOL campaigns focused on various taxa, in which coverage ranges from 30% to 40%. Barcode sequences have been obtained from 4019 of the world's 10 000 bird species, 49 000 of 165 000 Lepidoptera, and over 10 000 of 31 000 teleost species (data from BOLD searches, March 2014).

In light of the small number of species on the checklist of parasites and vectors, and considering that 10 000 new barcodes are added to BOLD each week [44], at first glance 43% coverage may not denote great advancement. It is also low in light of the importance of these organisms to human welfare, and given the appeal by Besansky *et al.* [12]. However, as outlined below, because of logistical and technical obstacles specific to working with parasites and vectors, 43% coverage arguably represents reasonable progress.

A major factor that has limited DNA barcoding of parasites and vectors is that it is framed and funded as biodiversity science, which is not a priority in public health or medicine. Barcoding 'campaigns' (e.g., Fish-BOL) focus on particular taxa [43] rather than polyphyletic functional groups such as parasites and vectors, a constraint that follows naturally from the distribution of taxonomic expertise. Additionally, among some end-users, species-level diagnosis may not be considered relevant. For example, medical clinicians often use the same treatments for most pathogens in the same higher taxon (and often employ formalin fixation for samples, which degrades DNA). When species-level diagnosis is desired, numerous standardized molecular (particularly immunological) approaches are available [45], although these continue to be considered too complex for routine clinical work in endemic areas (e.g., [6,46]). Sampling endoparasites in blood or excreta is also likely to be impeded by the ethical and legal necessity of obtaining informed patient consent (e.g., see [47]). Specimen identification can play an important role in vector-control programs [48], particularly when parasites can be transmitted by ecologically distinctive vectors (e.g., dengue, *Aedes albopictus* (Skuse), *Aedes aegypti* (L.) [49]) that may be subject to different control strategies. Nonetheless, large-scale disease-reduction efforts, including vector-control programs, often focus on strategies [50–52] in which coarse taxonomic information is sufficient. For example, although insecticide-treated bed nets may have distinct effects on *Anopheles* species that differ widely in importance as malaria vectors [53], this knowledge was not necessary to initiate or implement an effective bed-net program or to evaluate key end points [54].

The deep evolutionary divergence among endoparasites also presents obstacles for a standardized molecular approach. Broad-spectrum primers for the barcode fragment are lacking in several groups, a major impediment for high-volume barcoding [55]. Recently developed primers with high efficacy should aid barcoding progress in parasitic nematodes [56]. In the Platyhelminthes, a two-step process was proposed, with truly universal 18S rDNA primers serving to narrow the identity of unknown specimens sufficiently to select more taxon-specific primers to be used for COI [57]. In protists, 18S screening has been suggested as a tool not only for primer selection but also to indicate which of a range of still-undefined barcode targets would be used for species discrimination [58]. This is particularly necessary in taxa that lack mitochondria, such as *Giardia*, *Entamoeba*, *Blastocystis*, and *Cryptosporidium*, and COI may also prove problematic in taxa with high numbers of nuclear copies of mitochondrial DNA (e.g., *Toxoplasma* [59]). Paradoxically, however, the primer 'problem' can also be a boon. It is often difficult to obtain endoparasite samples free of host tissue [3] and similar problems occur when analyzing blood meals in hematophagous arthropods [60–63]. Some level of primer specificity is therefore desirable to avoid co-amplification of parasite and host DNA.

In light of these difficulties, 43% coverage is fair progress, particularly given that species sampling has proceeded without oversight or coordination. (A relevant campaign initiated in 2009, HealthBOL, awaits funding; T.R. Gregory, personal communication.) For example, higher barcode coverage in agricultural pests is likely to reflect the large number of relevant Lepidoptera [15], which are the focus of a major barcoding campaign (http://www.lepbarcoding.org), rather than a concerted effort to safeguard food supplies. Indeed, a critical component of such a campaign is the existence of a species checklist such as the one provided here (Table S2 in the supplementary material online). Finally, one further comparison suggests that medical importance may have more than doubled the likelihood that a species has been DNA barcoded. If species sampling were random, checklist coverage should be close to 3.5% [13] or 15% [14], which are the estimated proportions of described metazoan species that were DNA barcoded in 2012.

Prospects for the future

Although DNA barcoding (or any form of molecular diagnostics) may not offer convenient tools for clinical diagnosis at present (but see [64,65]), its potential utility in epidemiological studies and vector control is clearer. Climate change, urbanization, and globalized trade and transportation are altering the distribution of both vectors and parasites and understanding how these factors influence human disease transmission and morbidity is important [66]. For example, the range and incidence of several unrelated pathogens transmitted by ticks are increasing in Europe and North America [67–70]. In these and other emerging zoonoses, it is challenging to attribute changes in disease dynamics to specific causes, which can include vector species introductions, changes in local vector competence, changes in population density in both humans and vectors associated with urbanization and sprawl, altered irrigation and damming practices, and local and global human transportation networks [71–74]. DNA barcodes can provide specific and potentially useful information in such cases, for example, through early detection of introduced vector species (Table S1 in the supplementary material online) or by mapping changes in vector distribution [75] subsequent to deforestation [76]. Accurate identification is necessary to understand the ecology of arthropod vectors in breeding habitats, which is regaining momentum as a target for mosquito control after long neglect in favor of indoor spraying and insecticide-treated bed nets [51,77–79]. Several studies have used DNA barcoding to identify blood-meal sources from disease vectors (Table S1 in the supplementary material online) (reviewed in [80]). Such data can be critical for refining models of local transmission risks; for example, by distinguishing difficult-to-identify rodent reservoirs of hantavirus, Chagas disease, and leishmaniasis in Brazil [81] (see also [82] for an example with reservoirs of avian influenza).

In our view, a standardized, single-locus approach provides a useful tool for specimen identification, initial delineation of species, and highlighting taxonomic problems in parasites and vectors. While no single molecular marker will differ uniformly among the diverse organisms that cause and transmit disease in humans, and all markers will fail in some groups, the 'strategy' of using different markers tailored to distinguish species in different groups quickly leads to an impractical situation for routine

Box 2. Taxonomic discrepancies illuminated by DNA barcodes

The coherence of taxonomic names associated with barcode sequences in species of parasites, vectors, and hazards was assessed in 13 609 public records obtained by searching for checklist species in the BOLD system. Barcode sequences were assigned to putative species based on BINs assigned by BOLD. Sequences from checklist species formed 934 BIN clusters, of which 723 contained more than one sequence. In most (471/723) cases, a single species name was used to identify organisms in each BIN; however, 252 BINs were identified using more than one species name. Some of this discordance may be biologically accurate; that is, arising from haplotype sharing among species resulting from incomplete lineage sorting, hybridization, or recent divergence. For example, BIN clusters do not capture some recently proposed changes in *Plasmodium* systematics. In the latter case, a BIN dominated by 128 sequences of *Plasmodium falciparum* Welch is discordant because it includes 14 sequences of *Plasmodium* sp. gorilla clade G1 [102], for which the name *Plasmodium praefalciparum* was recently proposed by Rayner *et al.* [101] based on the specificity of this lineage for gorillas. The near identity of sequences of human *P. falciparum* and *P. praefalciparum* also occurs in apicoplast and nuclear genes and in complete mitochondrial genomes [102]. In other words, arguably, the failure of barcodes to distinguish between *P. falciparum* and *P. praefalciparum* reflects the ecological basis of this newly proposed species rather than insufficiency of information in the DNA barcode region or shortcomings of the BIN algorithm. Much of the inconsistency in the 252 discordant BINs is likely to represent nomenclatural error, as BINs are 90% consistent with taxonomy in other groups [44]. In the *Anopheles funestus* Giles species complex, for example, species recently recognized (based on COI and other markers [103,104]) are accorded different BINs. The BIN algorithm flags the recently described human schistosome *Schistosoma guineensis* (Pagès *et al.*) as distinct from *Schistosoma intercalatum* (Fisher) and correctly classifies misidentified isolates of the former (cf. [30]; see Table S1 in the supplementary material online). By automatically flagging discordance in a standardized marker, BOLD provides a useful starting point for distinguishing between database problems (misidentifications or inconsistent nomenclature) and biological phenomena (incipient speciation, hybridization, incomplete lineage sorting).

identification. To select optimal markers (or sets of markers), it becomes necessary to identify specimens at a higher taxon, which is often impossible for nonspecialists. The importance of standardization is illustrated by misidentified *Plasmodium* sequences in GenBank, which Valkiūnas *et al.* [5] could note only because the same marker had been sequenced at sufficient length in all isolates. This difficulty clearly plagues species in our checklist (Box 2), in which different names have been used for many clusters of similar COI sequences. As noted by Besansky *et al.* [12], DNA barcoding does not preclude the use of other markers (indeed, BOLD is equipped to host such data; Box 1), but the added benefit should be weighed against added effort in the context of a standardized approach. The initial use of a single marker (specifically, COI) can be seen as an optimal choice in a trade-off between effort and taxonomic confidence (Figure 4). Additional markers will increase taxonomic confidence, but the high accuracy initially obtained with barcodes suggests that the increase will generally be marginal. By contrast, the small increase in confidence with more markers requires a substantial increase in effort [83], particularly in the context of Sanger sequencing and if identifications need to be made rapidly [43,84].

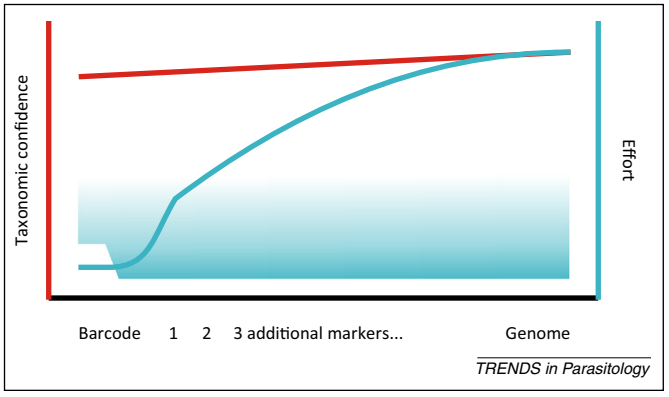

Figure 4. Trade-off between taxonomic confidence and effort, as a function of molecular data. Taxonomic confidence (the likelihood that species will be correctly distinguished and specimens correctly identified) increases with increasing molecular data (red line). However, the high accuracy obtained with DNA barcodes (94–95% in the studies reviewed here; see Figure 1C,F) suggests that the increase will generally be marginal. Sanger sequencing of each additional marker results in large increases in effort (reagents, time, and analysis) (blue line). The blue line also rises over 'barcode' to reflect, for example, the primer development and optimization of extraction techniques that are necessary for some taxa. The shaded blue area represents effort associated with next-generation sequencing (NGS) and its haziness reflects the uncertainty stemming from the rapid pace of change in this field. The initial acquisition of NGS capacity is costly in terms of both expense and analytic capacity. Thereafter, the cost per sequence decreases to below that of Sanger sequencing, with sequences from genomes costing little in principal. However, as discussed in the text (see section on 'Prospects for the future'), this does not necessarily translate into decreased cost per specimen and species.

With next-generation sequencing (NGS) technologies, the cost of obtaining each sequence read from a DNA template has fallen by orders of magnitude, leading to the notion that the use of multiple markers for molecular species identification is an inevitable and positive development [17]. Nonetheless, the start-up investment (infrastructure, training), cost per run (rather than per base sequenced), and rapid pace of technological change (when to invest and which platform to invest in) of NGS present significant obstacles for many [85], and multiple marker analysis is vastly more computationally intensive [83,86] and therefore less scalable. Wide use of Sanger sequencing is likely to continue for some time, as illustrated by its use in all of the studies in Figure 1H (cf. [17]). Another obstacle to the wide uptake of NGS in molecular species identification is discrepancy between the scale of the information needed and that produced by the technology. Eukaryotic species circumscription and specimen identification require far less than a complete genomic portrait of a specimen, but that is the scale of data produced in each experimental unit of a NGS run. The cost per specimen is low only if specimens are pooled, which usually requires unique molecular tagging of the DNA of large numbers of specimens in each run, a step that is technologically challenging and, in some platforms, costly. Data storage and analysis remain a major bottleneck in NGS [87] and, for this reason, minimalism and standardization also remain desirable attributes of a standardized NGS species-identification protocol. For example, one way to view the data problem is as a trade-off between deeply sequencing each specimen's DNA and more shallowly sequencing that of more specimens, with the aim of optimizing the return on sequence quality and species-level information content [87,88]. The DNA barcode generally provides a high level

of accuracy (Figure 1) and, if targeted in NGS, COI paralogs, heteroplasmy, and contamination can be detected directly [89], which will further increase the initial taxonomic confidence. One relevant, promising NGS application is the inventory of species inhabiting environmental media through their residual DNA; barcoding has been used in several such studies (reviewed in [89–91]). Such approaches could be employed to screen soil and water for suites of both metazoan vectors and pathogens, instead of assays that target single species (e.g., [92,93]).

The increased volume of species, specimens, and sequences emerging with NGS makes the need for computational efficiency and a standardized library of reference sequences acute [83,94]. Ideally, such a library should be dominated by sequences from well-curated specimens with the strongest possible links to existing names rather than indiscriminate collections [14,95]. However, the small numbers of specimens and sequences obtained from focused collections in type hosts and localities are generally more conducive to Sanger sequencing than NGS.

After 11 years, we are less than halfway 'there' in barcoding medically important parasites and vectors (i.e., barcodes exist for 43% of checklist species), but the clear utility of the approach and spate of recent work (Figure 1) are grounds for hope and the parasite and vector checklist compiled here can help focus future efforts.

Acknowledgments

This study was supported by funding to R.H.H. from the OMAF Emergency Management Program and to Paul D.N. Hebert from NSERC and from the government of Canada through Genome Canada and the Ontario Genomics Institute in support of the iBOL project. Nick Daunt helped assemble the checklist and Muhammad Ashfaq, Andrew Frewin, and Shadi Shokralla provided useful suggestions. The authors acknowledge Megan Milton, James Robertson, and the BOLD informatics team for technical support, as well as helpful suggestions from Rhiannon Macrae and three anonymous reviewers.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.pt.2014.09.003.

References

1 WHO (2010) *Working to Overcome the Global Impact of Neglected Tropical Diseases: First WHO Report on Neglected Tropical Diseases*, WHO
2 Fenwick, A. (2012) The global burden of neglected tropical diseases. *Public Health* 126, 233–236
3 Perkins, S.L. *et al.* (2011) Do molecules matter more than morphology? Promises and pitfalls in parasites. *Parasitology* 138, 1664–1674
4 Reuben, R. (1994) Illustrated keys to species of *Culex* (*Culex*) associated with Japanese encephalitis in Southeast Asia (Diptera: Culicidae). *Mosq. Syst.* 26, 75–96
5 Valkiūnas, G. *et al.* (2008) Parasite misidentifications in GenBank: how to minimize their number? *Trends Parasitol.* 24, 247–248
6 Wong, S.S. *et al.* (2014) Molecular diagnosis in clinical parasitology: when and why? *Exp. Biol. Med. (Maywood)* http://dx.doi.org/10.1177/1535370214523880
7 Bensch, S. *et al.* (2009) MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome *b* lineages. *Mol. Ecol. Resour.* 9, 1353–1358
8 Aurrecoechea, C. *et al.* (2010) EuPathDB: a portal to eukaryotic pathogen databases. *Nucleic Acids Res.* 38, D415–D419
9 Valentini, A. *et al.* (2009) DNA barcoding for ecologists. *Trends Ecol. Evol.* 24, 110–117
10 Hebert, P.D.N. *et al.* (2003) Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270, 313–321
11 Hebert, P.D.N. *et al.* (2003) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proc. Biol. Sci.* 270 (Suppl. 1), S96–S99
12 Besansky, N.J. *et al.* (2003) DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. *Trends Parasitol.* 19, 545–546
13 Kwong, S. *et al.* (2012) An update on DNA barcoding: low species coverage and numerous unidentified sequences. *Cladistics* 28, 639–644
14 Kvist, S. (2013) Barcoding in the dark? A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. *Mol. Phylogenet. Evol.* 69, 39–45
15 Frewin, A.C. *et al.* (2013) DNA barcoding for plant protection: applications and summary of available data for arthropod pests. *CAB Rev.* 8, 1–13
16 Fišer Pečnikar, Ž. and Buzan, E.V. (2014) 20 Years since the introduction of DNA barcoding: from theory to application. *J. Appl. Genet.* 55, 43–52
17 Taylor, H.R. and Harris, W.E. (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Mol. Ecol. Resour.* 12, 377–388
18 Teletchea, F. (2010) After 7 years and 1000 citations: comparative assessment of the DNA barcoding and the DNA taxonomy proposals for taxonomists and non-taxonomists. *Mitochondrial DNA* 21, 206–226
19 Frézal, L. and Leblois, R. (2008) Four years of DNA barcoding: current advances and prospects. *Infect. Genet. Evol.* 8, 727–736
20 Betson, M. *et al.* (2011) A molecular epidemiological investigation of *Ascaris* on Unguja, Zanzibar using isoenyzme analysis, DNA barcoding and microsatellite DNA profiling. *Trans. R. Soc. Trop. Med. Hyg.* 105, 370–379
21 Foster, P.G. *et al.* (2013) Phylogenetic analysis and DNA-based species confirmation in *Anopheles* (*Nyssorhynchus*). *PLoS ONE* 8, e54063
22 Laurito, M. *et al.* (2013) COI barcode versus morphological identification of *Culex* (*Culex*) (Diptera: Culicidae) species: a case study using samples from Argentina and Brazil. *Mem. Inst. Oswaldo Cruz* 108 (Suppl. 1), 110–122
23 Cohnstaedt, L.W. *et al.* (2011) Phylogenetics of the phlebotomine sand fly group *Verrucarum* (Diptera: Psychodidae: *Lutzomyia*). *Am. J. Trop. Med. Hyg.* 84, 913–922
24 Zapata, S. *et al.* (2012) A study of a population of *Nyssomyia trapidoi* (Diptera: Psychodidae) caught on the Pacific coast of Ecuador. *Parasit. Vectors* 5, 144
25 Gómez, G. *et al.* (2013) Wing geometric morphometrics and molecular assessment of members in the *Albitarsis* complex from Colombia. *Mol. Ecol. Resour.* 13, 1082–1092
26 Loaiza, J.R. *et al.* (2013) Novel genetic diversity within *Anopheles punctimacula* s.l.: phylogenetic discrepancy between the barcode cytochrome *c* oxidase I (COI) gene and the rDNA second internal transcribed spacer (ITS2). *Acta Trop.* 128, 61–69
27 Brabec, J. *et al.* (2012) Substitution saturation and nuclear paralogs of commonly employed phylogenetic markers in the Caryophyllidea, an unusual group of non-segmented tapeworms (Platyhelminthes). *Int. J. Parasitol.* 42, 259–267
28 Huyse, T. *et al.* (2013) Hybridisation between the two major African schistosome species of humans. *Int. J. Parasitol.* 43, 687–689
29 Huyse, T. *et al.* (2009) Bidirectional introgressive hybridization between a cattle and human schistosome species. *PLoS Pathog.* 5, e1000571
30 Webster, B.L. *et al.* (2006) A revision of the interrelationships of *Schistosoma* including the recently described *Schistosoma guineensis*. *Int. J. Parasitol.* 36, 947–955
31 Ashford, R. and Crewe, W. (2003) *An Annotated Checklist of the Protozoa, Helminths and Arthropods for which We are Home*, CRC Press
32 Eldridge, B.F. and Edman, J.D. (2000) *Medical Entomology: A Textbook on Public Health and Veterinary Problems Caused by Arthropods*, Kluwer Academic

33 Goddard, J. (2003) *Physician's Guide to Arthropods of Medical Importance*. (4th edn), CRC Press
34 Marquardt, W.C. (2005) *Biology of Disease Vectors*, Elsevier
35 Mullen, G.R. and Durden, L.A. (2009) *Medical and Veterinary Entomology*, Academic Press
36 Service, M. (2012) *Medical Entomology for Students*. (5th edn), Cambridge University Press
37 Taylor, L.H. *et al.* (2001) Risk factors for human disease emergence. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 356, 983–989
38 US Environmental Protection Agency (2008) *Contaminant Candidate List 3 Microbes: Screening to the PCCL*, Office of Water
39 Woolhouse, M.E.J. and Gowtage-Sequeria, S. (2005) Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* 11, 1842–1847
40 Goddard, J. (2006) Arthropods, tongue worms, leeches, and arthropod-borne diseases. In *Tropical Infectious Diseases* (2nd edn) (Guerrant, R.L. *et al.*, eds), pp. 1370–1385, Elsevier
41 Phillips, A.J. *et al.* (2010) *Tyrannobdella rex* n. gen. n. sp. and the evolutionary origins of mucosal leech infestations. *PLoS ONE* 5, e10057
42 Liu, D. (2013) Introductory remarks. In *Detection of Human Parasitic Pathogens* (Liu, D., ed.), pp. 1–12, Taylor & Francis
43 Boykin, L.M. *et al.* (2012) Species delimitation and global biosecurity. *Evol. Bioinform. Online* 8, 1–37
44 Ratnasingham, S. and Hebert, P.D.N. (2013) A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. *PLoS ONE* 8, e66213
45 John, D.T. (2006) *Markell and Voge's Medical Parasitology*. (9th edn), Elsevier
46 Fürst, T. *et al.* (2012) Global burden of human food-borne trematodiasis: a systematic review and meta-analysis. *Lancet Infect. Dis.* 12, 210–221
47 Truog, R.D. *et al.* (2012) Research ethics. Paying patients for their tissue: the legacy of Henrietta Lacks. *Science* 337, 37–38
48 Walker, K. and Lynch, M. (2007) Contributions of *Anopheles* larval control to malaria suppression in tropical Africa: review of achievements and potential. *Med. Vet. Entomol.* 21, 2–21
49 Higa, Y. (2011) Dengue vectors and their spatial distribution. *Trop. Med. Health* 39, 17–27
50 Molyneux, D.H. *et al.* (2005) "Rapid-impact interventions": how a policy of integrated control for Africa's neglected tropical diseases could benefit the poor. *PLoS Med.* 2, e336
51 Ferguson, H.M. *et al.* (2010) Ecology: a prerequisite for malaria elimination and eradication. *PLoS Med.* 7, e1000303
52 Kabatereine, N.B. *et al.* (2010) How to (or not to) integrate vertical programmes for the control of major neglected tropical diseases in sub-Saharan Africa. *PLoS Negl. Trop. Dis.* 4, e755
53 Bayoh, M. *et al.* (2010) *Anopheles gambiae*: historical population decline associated with regional distribution of insecticide-treated bed nets in western Nyanza Province. *Malar. J.* 9, 62
54 Choi, H.W. *et al.* (1995) The effectiveness of insecticide-impregnated bed nets in reducing cases of malaria infection: a meta-analysis of published results. *Am. J. Trop. Med. Hyg.* 52, 377–382
55 Hajibabaei, M. *et al.* (2005) Critical factors for assembling a high volume of DNA barcodes. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 360, 1959–1967
56 Prosser, S.W.J. *et al.* (2013) Advancing nematode barcoding: a primer cocktail for the cytochrome *c* oxidase subunit I gene from vertebrate parasitic nematodes. *Mol. Ecol. Resour.* 13, 1108–1115
57 Moszczynska, A. *et al.* (2009) Development of primers for the mitochondrial cytochrome *c* oxidase I gene in digenetic trematodes (Platyhelminthes) illustrates the challenge of barcoding parasitic helminths. *Mol. Ecol. Resour.* 9 (Suppl. s1), 75–82
58 Pawlowski, J. *et al.* (2012) CBOL Protist Working Group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* 10, e1001419
59 Gjerde, B. (2013) Characterisation of full-length mitochondrial copies and partial nuclear copies (numts) of the cytochrome *b* and cytochrome *c* oxidase subunit I genes of *Toxoplasma gondii*, *Neospora caninum, Hammondia heydorni and Hammondia triffittae* (Apicomplexa: Sarcocystidae). *Parasitol. Res.* 112, 1493–1511
60 Townzen, J.S. *et al.* (2008) Identification of mosquito bloodmeals using mitochondrial cytochrome oxidase subunit I and cytochrome *b* gene sequences. *Med. Vet. Entomol.* 22, 386–393
61 Alcaide, M. *et al.* (2009) Disentangling vector-borne transmission networks: a universal DNA barcoding method to identify vertebrate hosts from arthropod bloodmeals. *PLoS ONE* 4, e7092
62 Gariepy, T.D. *et al.* (2012) Identifying the last supper: utility of the DNA barcode library for bloodmeal identification in ticks. *Mol. Ecol. Resour.* 12, 646–652
63 Muturi, C.N. *et al.* (2011) Tracking the feeding patterns of tsetse flies (*Glossina* genus) by analysis of bloodmeals using mitochondrial cytochrome genes. *PLoS ONE* 6, e17284
64 Rukke, B.A. *et al.* (2014) Confirming *Hypoderma tarandi* (Diptera: Oestridae) human ophthalmomyiasis by larval DNA barcoding. *Acta Parasitol.* 59, 301–304
65 Fukuda, M. *et al.* (2011) Zoonotic onchocerciasis in Hiroshima, Japan, and molecular analysis of a paraffin section of the agent for a reliable identification. *Parasite* 18, 185–188
66 WHO (2004) *Global Strategic Framework for Integrated Vector Management*, WHO
67 Colwell *et al.* (2011) Vector-borne parasitic zoonoses: emerging scenarios and new perspectives. *Vet. Parasitol.* 182, 14–21
68 Randolph, S.E. and Rogers, D.J. (2010) The arrival, establishment and spread of exotic diseases: patterns and predictions. *Nat. Rev. Microbiol.* 8, 361–371
69 Petney, T.N. (2011) Progress in parasitology. In *Parasitology Research Monographs* (Vol. 2) (Mehlhorn, H., ed.), In pp. 283–296, Springer–Verlag
70 Tuite, A.R. *et al.* (2013) Effect of latitude on the rate of change in incidence of Lyme disease in the United States. *CMAJ Open* 1, E43–E47
71 Harrus, S. and Baneth, G. (2005) Drivers for the emergence and re-emergence of vector-borne protozoal and bacterial diseases. *Int. J. Parasitol.* 35, 1309–1318
72 Tatem, A.J. *et al.* (2006) Global traffic and disease vector dispersal. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6242–6247
73 Stoddard, S.T. *et al.* (2009) The role of human movement in the transmission of vector-borne pathogens. *PLoS Negl. Trop. Dis.* 3, e481
74 Tabachnick, W.J. (2010) Challenges in predicting climate and environmental effects on vector-borne disease episystems in a changing world. *J. Exp. Biol.* 213, 946–954
75 Ashfaq, M. *et al.* (2014) Analyzing mosquito (Diptera: Culicidae) diversity in Pakistan by DNA barcoding. *PLoS ONE* 9, e97268
76 Abad-Franch, F. *et al.* (2009) Ecology, evolution, and the long-term surveillance of vector-borne Chagas disease: a multi-scale appraisal of the tribe Rhodniini (Triatominae). *Acta Trop.* 110, 159–177
77 Killeen, G.F. *et al.* (2002) Advantages of larval control for African malaria vectors: low mobility and behavioural responsiveness of immature mosquito stages allow high effective coverage. *Malar. J.* 1, 8
78 Impoinvil, D.E. *et al.* (2007) Comparison of mosquito control programs in seven urban sites in Africa, the Middle East, and the Americas. *Health Policy* 83, 196–212
79 Manguin, S. *et al.* (2010) Review on global co-transmission of human *Plasmodium* species and *Wuchereria bancrofti* by *Anopheles* mosquitoes. *Infect. Genet. Evol.* 10, 159–177
80 Kent, R.J. (2009) Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. *Mol. Ecol. Resour.* 9, 4–18
81 Müller, L. *et al.* (2013) DNA barcoding of sigmodontine rodents: identifying wildlife reservoirs of zoonoses. *PLoS ONE* 8, e80282
82 Lee, D-H. *et al.* (2010) DNA barcoding techniques for avian influenza virus surveillance in migratory bird habitats. *J. Wildl. Dis.* 46, 649–654
83 Collins, R.A. and Cruickshank, R.H. (2014) Known knowns, known unknowns, unknown unknowns and unknown knowns in DNA barcoding: a comment on Dowton *et al*. *Syst. Biol.* http://dx.doi.org/10.1093/sysbio/syu060
84 Ivanova, N.V. *et al.* (2009) Express barcodes: racing from specimen to identification. *Mol. Ecol. Resour.* 9 (Suppl. s1), 35–41
85 Glenn, T.C. (2011) Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* 11, 759–769
86 Dowton, M. *et al.* (2014) A preliminary framework for DNA barcoding, incorporating the multispecies coalescent. *Syst. Biol.* 63, 639–644
87 McCormack, J.E. *et al.* (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phylogenet. Evol.* 66, 526–538

88 Caporaso, J.G. *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624
89 Shokralla, S. *et al.* (2014) Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Mol. Ecol. Resour.* 14, 892–901
90 Hajibabaei, M. *et al.* (2012) Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecol.* 12, 28
91 Yu, D. (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol. Evol.* 3, 613–623
92 Akande, I.S. *et al.* (2012) Polymerase chain reaction (PCR) investigations of prepatent *Schistosoma haematobium* cercariae incidence in five water bodies, South West, Nigeria. *Afr. J. Med. Med. Sci.* 41 (Suppl.), 75–80
93 Macuhova, K. *et al.* (2013) Contamination, distribution and pathogenicity of *Toxocara canis* and *T. cati* eggs from sandpits in Tokyo, Japan. *J. Helminthol.* 87, 271–276
94 Taberlet, P. *et al.* (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21, 2045–2050
95 Kvist, S. *et al.* (2010) Barcoding, types and the *Hirudo* files: using information content to critically evaluate the identity of DNA barcodes. *Mitochondrial DNA* 21, 198–205
96 Folmer, O. *et al.* (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3, 294–299
97 Puillandre, N. *et al.* (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol. Ecol.* 21, 1864–1877
98 Ratnasingham, S. and Hebert, P.D.N. (2007) BOLD: the Barcode of Life Data system (http://www.barcodinglife.org). *Mol. Ecol. Notes* 7, 355–364
99 Pleijel, F. *et al.* (2008) Phylogenies without roots? A plea for the use of vouchers in molecular phylogenetic studies. *Mol. Phylogenet. Evol.* 48, 369–371
100 Hanner, R. (2009) *Data Standards for BARCODE Records in INSDC* (BRIs), Consortium for the Barcode of Life
101 Rayner, J.C. *et al.* (2011) A plethora of *Plasmodium* species in wild apes: a source of human infection? *Trends Parasitol.* 27, 222–229
102 Liu, W. *et al.* (2010) Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* 467, 420–425
103 Choi, K.S. *et al.* (2012) Population genetic structure of the major malaria vector *Anopheles funestus* s.s. and allied species in southern Africa. *Parasit. Vectors* 5, 283
104 Coetzee, M. and Koekemoer, L.L. (2013) Molecular systematics and insecticide resistance in the major African malaria vector *Anopheles funestus*. *Annu. Rev. Entomol.* 58, 393–412