EDTORIAL
Advances and challenges in barcoding of microbes, parasites, and their vectors and reservoirs

SERGE MORAND1,2*

1 CNRS ISEM, CIRAD ASTRE, Université de Montpellier, Montpellier, France
2 Faculty of Veterinary Technology, Kasetsart University, Bangkok, Thailand

(Received 11 April 2018; accepted 26 April 2018)

SUMMARY
DNA barcoding is now a common tool in parasitology and epidemiology, which require good methods for identification not only of parasites and pathogens but vectors and reservoirs. This special issue presents some advances and challenges in barcoding of microbes, parasites, and their vectors and reservoirs. DNA barcoding found new applications in disease ecology, conservation parasitology, environmental parasitology and in paleoparasitology. New technologies such as next-generation sequencing and matrix-assisted laser desorption–ionization time-of-flight have made it now possible to investigate large samples of specimens. By allowing the investigation of parasites at the interface between environment, biodiversity, animal and human health, barcoding and biobanking have important policy outcomes as well as ethics and legal implications. The special issue ‘Advances and challenges in the barcoding of parasites, vectors and reservoirs’ illustrates some recent advances and proposes new avenues for research in barcoding in parasitology.

Key words: DNA barcoding, eDNA, paleoparasitology, next-generation sequencing, MALDI–TOF barcoding, BOLD, R packages, ethics, legal issues.

INTRODUCTION
Waterton et al. (2013) emphasized that ‘Taxonomy entered the 21st century wringing its hands in self-reflective concern at the fragmentation and lack of standing (including funding and new recruitment) of the field’. Parasitology, epidemiology and medical veterinary face similar concerns with, on the one hand, an increase of emerging and re-emerging infectious and parasitic diseases and, on the other, a decrease in taxonomic expertise challenging the need of rapid and accurate identification of pathogens, parasites, vectors and reservoirs.

Facing the loss of expertise in taxonomy, DNA barcoding was proposed by Hebert et al. (2003), as a new system of species identification using a short section of DNA from a standardized region of the genome. In the case of animals, the mitochondrial cytochrome c oxidase 1 gene (CO1) was chosen to establish species delineation and identification. However, this barcode marker soon appeared to be far from universal. In the case of fungi, for example, the most common marker used is the large ribosomal DNA ‘Internal Transcriber Spacer’ (ITS), although this marker does not work for all fungal groups. ITS is an excellent marker to distinguish species of the genus Pneumocystis, which comprises fungal pathogens residing in the pulmonary parenchyma of a wide range of mammals (Danesi et al. 2016; Latinne et al. 2017). In the same manner, ITS appears to be a good candidate in several groups of protists (Wang et al. 2015), such as Trypanosoma species (Desquesnes et al. 2011).

The Barcode of Life (BOLI) promoted DNA barcoding as a way to speed up (and even reinventing taxonomy as emphasized by their promoters) the identification work of traditional taxonomy, with the urgent task to identify all the unknown species before their disappearance (Meier, 2008). BOLI is considered as a tool in support of the Convention on Biological Diversity (CBD). The Consortium for the Barcode of Life (CBOL) was established in 2004 as an international initiative devoted to developing DNA barcoding (http://www.barcodeoflife.org/). The International Barcode of Life project (iBOL) was subsequently launched in 2010 as a research alliance of scientists, technologists and ethicists from 25 nations to construct a DNA barcode reference library (http://ibol.org/phase1/). The Barcode of Life Data Systems database (BOLD) was established as the identification tool for all organism barcodes. BOLD’s infrastructure was initially designed to process and analyse the only CO1, but it can now process multiple genes (http://www.boldsystems.org/). Barcoding needs both
universal barcodes (CO1, ITS) and a high quality of accessible databases (sequences, systematics) (Shen et al. 2013).

BOLI was established at the rise of the genomics era. New throughput technologies [next-generation sequencing (NGS)] and new adapted technologies such as matrix-assisted laser desorption–ionization time-of-flight (MALDI–TOF) mass spectrometry open new avenues (Ilina et al. 2009; Michelet et al. 2014).

Barcoding is now a common tool in parasitology and epidemiology which need good identification assessment not only of parasites and pathogens (Prosser et al. 2013; Ondrejicka et al. 2014) but also vectors (Ruiz-Lopez et al. 2012; Chan et al. 2014; Kumlert et al. 2018) and reservoirs (Galan et al. 2012). Barcoding in parasitology concerns a wide range of organisms from viruses, bacteria, fungi, protists, helminths, arthropods and molluscs, but also vertebrate animals.

REFERENCE SPECIMENS AND OPEN DATABASES

The tasks are to collect and curate specimens, to obtain barcode records from these specimens, to use (or to be built) the informatics platform to store these records and to enable their use by a large community. Following collection and barcoding of the specimens, two imperatives in barcoding practices remain: (i) to reference DNA barcoding to voucher specimens (and collection) and (ii) to link data in open databases.

By definition, a morphological voucher is a preserved specimen archived in a collection facility such as a museum. In DNA barcoding, the preservation of morphological vouchers is a standard practice for specimens from which DNA barcode sequences were obtained. At the beginning of BOLI, CO1 barcoding was mostly used to link established Linnaean taxonomy with curated voucher specimens in museums. BOLI should be seen as a classification system and not as a taxonomic system (Vogler and Monahan, 2006). With more accessible technologies in barcoding, it remains even more imperative to keep voucher specimens, which are representative of individual organisms identified using current technology and taxonomic classification. Collection facilities and biobanks for tissues or even living organisms (e.g. bacteria and protists) are then complementary to parasitology barcoding initiatives.

The BOLD was established in 2005 as a repository platform of DNA barcodes for all eukaryotic life. The latest version of BOLD was released in 2015 (http://v4.boldsystems.org/) and now hosts more than 6 million barcodes from more than 270,000 species (including animals, plants and fungi). Barcode sequences are catalogued in GenBank (http://www.ncbi.nlm.nih.gov/genbank). Linking barcode to voucher specimen, an important process in barcoding, should operate through accurate and updated taxonomic classification like the Catalogue of Life (http://www.catalogueoflife.org/), which helps at resolving inaccurate identification and/or change in the systematics of the organisms in consideration.

Investigation of BOLD showed that parasitic organisms are far from all being barcoded and/or that their barcodes if existing were not archived in BOLD. For example, approximately 1300 species of Acanthocephala have been morphologically described (Poulin and Morand, 2004; Garcia-Varela and Pérez-Ponce de León, 2015), but only 38 species (< 3%) have their barcodes recorded in BOLD. Similarly, there are 663 species of Platyhelminthes with barcodes in BOLD, whereas there are currently around 30,000 known species (Caira and Littlewood, 2013) (a little more than 2% are barcoded).

Another concern is the geographic localization of barcode specimens. Adding accurate geo-localization will enable the barcode specimens to be geo-referenced and processes to other international databases such as Global Biodiversity Information Facility (GBIF) (https://www.gbif.org/).

ADVANCES AND NEW APPLICATIONS IN BARCODING

Development of the barcoding approach was enhanced by new throughput technologies such as NGS. The breakthrough of integration of genomic data has been acknowledged in ecological genetics (Shafer et al. 2016) and found high relevance in epidemiology and public health. In microbiology, rapid NGS of whole-genome sequencing (WGS) associated with bioinformatic pipelines found increasing applications from microbial taxonomy to public health surveillance of pathogens (Allard, 2016).

Environmental genomics is a growing domain studying molecular components, DNA and RNA in (meta)genomes and (meta)transcriptomes, in environmental samples (Taberlet et al. 2012; Joly and Faure, 2015), with wide applications in biodiversity, monitoring and conservation biology (Stat et al. 2017). Environmental DNA (eDNA) as named in biodiversity screening has found recent applications in parasitology (Bass et al. 2015). eDNA may help at detecting free-living stages of parasites (eggs, cysts, larvae) in environmental samples collected in water or soil surveys or within intermediate hosts (Huver et al. 2015).

Among the available new technologies, MALDI–TOF mass spectrometry starts to be widely used as a new tool for barcoding (Sandrin et al. 2013; Rothen et al. 2016; Yssouf et al. 2016; Diarra et al. 2017), although this technique should be based on
accurate species identification both morphologically and genetically.

Advances can be observed in a more friendly and user-oriented access to the different databases thanks to the freeware statistical programming language R (R Core Team, 2018, www.R-project.org/). Among several available packages, one can cite ‘bold’ (Chamberlain, 2017) developed by BOLD, which offers functions to search sequences and specimens and download trace files or ‘BarcodingR’ (Zhang et al. 2017), which provides a comprehensive implementation of species identification methods. Packages have been also developed in R for creating and analysing DNA barcodes such as ‘DNAbarcodes’ (Buschmann, 2017), which finds utilities for manipulating large datasets obtained by NGS.

NEW APPLICATIONS OF BARCODING

A first new application of DNA barcoding concerns the identification of ancient parasite DNA (Côté and Le Bailly, 2018; Wood, 2018) irrigating a growing interest in paleogenomics and paleoparasitology.

In ecological parasitology, barcoding allows to follow vectors, the parasites they carry, but also the feeding activities of these vectors if they are blood feeders (i.e. biting arthropods). Applications then extend to use blood-feeding arthropods as vertebrate samplers (Kocher et al. 2017; Muturi et al. 2011), or as ‘flying syringes’ to collect blood samples and detect parasites in them (Bitome-Essonzo et al. 2017).

CHALLENGES IN BARCODING

Several pitfalls can occur in DNA barcoding linked to the events that have contributed to the evolutionary history of the species into consideration.

First, DNA barcoding based on mitochondrial genes (COI) may overestimate the number of species due to the presence of pseudogenes (Song et al. 2008). The removal of nuclear mitochondrial pseudogenes requires a careful examination of sequences.

Hybridization is the second issue in barcoding. Introggression of mitochondrial DNA due to hybridization and/or incomplete lineage sorting of mitochondrial DNA haplotypes may lead to misidentification as reported in several groups of mammals such as rodents and bats (Nesi et al. 2011; Pages et al. 2013; Ermakov et al. 2016), which are important reservoirs of diseases.

POLICY RELEVANCE OF BARCODING

The 2010 Nagoya Protocol identified key common goals between BOLI and the CBD, among others: to promote the capacity building in species identification and discovery, and to support CBD with respect to biodiversity targets (i.e. Aichi targets), national biodiversity strategies and action plans, monitoring, indicators and assessments, and invasive alien species (Vernooy et al. 2010). In Mars 2018, the GBIF has formalized collaboration with the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) (https://www.gbif.org/news/2gg|1Fqrxr4im4oSM eMmA2/gbif-formalizes-collaboration-with-biodiversity-assessment-platform). The Memorandum of Understanding says formalizes continuing cooperation between GBIF and IPBES since the platform stands that GBIF will help IPBES to identify and access biodiversity datasets relevant to IPBES assessments and indicators and, using knowledge gaps, to identify and prioritize mobilization of new data through GBIF. The ongoing process will obviously favour the interoperability of major international databases (BOLD, GBIF, GenBank), although it still leaves open questions concerning access and sharing.

In the sector of public or animal health, the development of eDNA for pathogen discovery in the environment may have relevance for policy. As emphasized by Bass et al. (2015) ‘the detection of apparently specific genomic material from a politically important (listed) pathogen in an environmental matrix relates to the universally applied principles of ’infection’ and ’disease’ detection and reporting according to the World Organization for Animal Health (Office International des Epizooties, OIE) (www.oie.int)’ (see also Stentiford et al. 2014). Similar concerns for public health are to take into consideration when eDNA studies and metabarcoding of environmental samples (water, soil, vectors) can detect important human pathogens or parasites. (HealthBold http://www.healthbol.org/).

ETHICS, LEGAL ISSUES

The CBD has changed the old taxonomic practices for collecting and archiving specimens in museum collections (Lajaunie et al. 2014). More stringent regulation is now applying following the implementation of the Nagoya Protocol and the Access and Benefit Sharing (ABS) of biodiversity and particularly for the health sector (Lajaunie and Mazzega, 2016). IBOL took in charge this new international issue of access, collect and curate of materials (specimens, genes, data) and made recommendations to the ABS to facilitate access to biodiversity samples for pure ‘non-commercial’ research (using distinct Material Transfer Agreements and arrangements for Prior Informed Consent) and to improve access to provider countries of information generated by the scientific use of their biodiversity and genetic resources. A MoU was signed in Nagoya at COP10 between the iBOL Board Chair and the CBD Executive Secretary.
Table 1. Presentation of the articles of the special issue ‘Advances and challenges in barcoding of microbes, parasites, and their vectors and reservoirs’

<table>
<thead>
<tr>
<th>Authors</th>
<th>Domains</th>
<th>Type</th>
<th>Organisms</th>
<th>Methodologies</th>
<th>Discussion findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aivelo and Medlar (2018)</td>
<td>Environmental sampling, conservation medicine</td>
<td>Review</td>
<td>Helminths</td>
<td>Metabarcoding</td>
<td>Limitations of metabarcoding</td>
</tr>
<tr>
<td>Beebe (2018)</td>
<td>Evolution, epidemiology</td>
<td>Review</td>
<td>Arthropods (mosquitoes)</td>
<td>DNA barcodes</td>
<td>Limitations due to incomplete lineage sorting and introgression events</td>
</tr>
<tr>
<td>Boon et al. (2018)</td>
<td>Evolution, epidemiology</td>
<td>Original article</td>
<td>Schistosoma (Platyhelminthes)</td>
<td>DNA barcodes (ITS)</td>
<td>Limitations due to hybridization</td>
</tr>
<tr>
<td>Côté and Le Bailly (2018)</td>
<td>Paleoparasitology, paleogenomics</td>
<td>Review</td>
<td>Human parasites (platyhelminthes, protists)</td>
<td>DNA barcodes</td>
<td>New techniques (target-enrichment capture methods based on DNA hybridization), paleogenomics and health</td>
</tr>
<tr>
<td>Guernier et al. (2018)</td>
<td>Environmental sampling, screening</td>
<td>Review</td>
<td>Leptospira (bacteria)</td>
<td>Metabarcoding</td>
<td>Core genome MLST (cgMLST)</td>
</tr>
<tr>
<td>Hutchinson and Stevens (2018)</td>
<td>Evolution, screening</td>
<td>Review</td>
<td>Trypanosomes (protists)</td>
<td>DNA barcodes</td>
<td>Limitations interest of new techniques (LAMP, NASBA)</td>
</tr>
<tr>
<td>Kocher et al. (2018)</td>
<td>Screening, epidemiology</td>
<td>Original article</td>
<td>Leishmania (protists)</td>
<td>kDNA minicircles</td>
<td>Limitation in diagnostic tools</td>
</tr>
<tr>
<td>Kosoy et al. (2018)</td>
<td>Screening, epidemiology</td>
<td>Review</td>
<td>Bartonella (bacteria)</td>
<td>DNA barcodes</td>
<td>Extension to other bacteria</td>
</tr>
<tr>
<td>Lajaunie and Ho (2018)</td>
<td>Legal and ethical issues</td>
<td>Original article</td>
<td>All</td>
<td></td>
<td>Big Data, MTA</td>
</tr>
<tr>
<td>Laroche et al. (2018)</td>
<td>Identification, surveillance</td>
<td>Original article</td>
<td>Triomine (arthropod)</td>
<td>MALDI-TOF MS</td>
<td>New tool for vector identification</td>
</tr>
<tr>
<td>Nebbak et al. (2018)</td>
<td>Identification, surveillance</td>
<td>Original article</td>
<td>Mosquitoes (arthropod)</td>
<td>MALDI-TOF MS</td>
<td>New tool for vector identification</td>
</tr>
<tr>
<td>Šlapeta (2018)</td>
<td>Identification, surveillance</td>
<td>Review</td>
<td>Cryptosporidium (protist)</td>
<td>SSU rDNA barcodes</td>
<td></td>
</tr>
<tr>
<td>Wood (2018)</td>
<td>Paleoparasitology</td>
<td>Review</td>
<td>Fungi, protists, platyhelminthes, arthropods</td>
<td>18s rDNA metabarcoding</td>
<td>Metabarcoding applications</td>
</tr>
</tbody>
</table>

**THE SPECIAL ISSUE**

The special issue ‘Advances and challenges in the barcoding of parasites, vectors and reservoirs’ aims at illustrating some recent advances and new research avenues of barcoding in parasitology (Table 1). All types of organisms were covered with bacteria (Guernier et al. 2018; Kosoy et al. 2018), protists (Hutchinson and Stevens, 2018; Kocher et al. 2018; Šlapeta, 2018), platyhelminths (Aivelo and Medlar, 2018, Boon et al., 2018), arthropod vectors (Beebe, 2018; Laroche et al. 2018; Nebbak et al. 2018) emphasizing new technologies such as metabarcoding (Aivelo and Medlar, 2018) and MALDI-TOF MS (Laroche et al. 2018; Nebbak et al. 2018), but also target-enrichment capture methods based on DNA hybridization in paleoparasitology (Côté and Le Bailly, 2018), LAMP and NASBA in protistology (Hutchinson and Stevens, 2018), or core genome MLST in bacteriology (Guernier et al. 2018).

Applications of metabarcoding in paleoparasitology were reviewed (Côté and Le Bailly, 2018; Wood, 2018), opening new ways to investigate the health of ancient communities of humans, domesticated animals and wildlife.

The limitations of barcoding were illustrated in the case of *Schistosoma* (Boon et al. 2018) and mosquitoes (Beebe, 2018). Although mitochondrial DNA barcode (CO1) shows utility in discriminating cryptic/sibling species, its use can be problematic when incomplete lineage sorting and introgression events can lead to indistinguishable COI sequences.

Finally, and as emphasized above, there are important ethical and legal issues of barcoding and biobanking. These issues are comprehensively addressed by Lajaunie and Ho (2018), who provided guidelines for implementing barcode research in parasitology.

**ACKNOWLEDGEMENTS**

I thank Professor John Ellis for his great help in preparing this special issue.

**FINANCIAL SUPPORT**

This study is supported by the French ANR Future HeathSEA (ANR-17-CE35-0003).


Paggi, M., Bazin, E., Galan, M., Chaval, Y., Claude, J., Herbreteau, V., Michaux, J., Piry, S., Morand, S. and Cosson, J. F.


