Barcoding in trypanosomes

Rachel Hutchinson¹ and Jamie R. Stevens²

¹School of Biological Sciences, Life Sciences Building, University of Bristol, Tyndall Avenue, Bristol BS8 1TQ, UK and ²Department of Biosciences, College of Life and Environmental Sciences, University of Exeter, Stocker Road, Exeter EX4 4QD, UK

Abstract

Trypanosomes (genus Trypanosoma) are parasites of humans, and wild and domestic mammals, in which they cause several economically and socially important diseases, including sleeping sickness in Africa and Chagas disease in the Americas. Despite the development of numerous molecular diagnostics and increasing awareness of the importance of these neglected parasites, there is currently no universal genetic barcoding marker available for trypanosomes. In this review we provide an overview of the methods used for trypanosome detection and identification, discuss the potential application of different barcoding techniques and examine the requirements of the ‘ideal’ trypanosome genetic barcode. In addition, we explore potential alternative genetic markers for barcoding Trypanosoma species, including an analysis of phylogenetically informative nucleotide changes along the length of the 18S rRNA gene.

Introduction

DNA barcoding: ‘the utilisation of DNA sequences of short standardised gene fragments for quick and accurate determination of the species’ (D’Avila-Levy et al. 2015)

Trypanosoma parasites are flagellated protozoa within the class Kinetoplastida, which is characterized by the presence of a kinetoplast: a mass of mitochondrial ‘kDNA’ (Adl et al. 2012). These parasites cause a wide range of diseases in both humans and animals, and are often transmitted between hosts by insect vectors (Fig. 1). Human diseases caused by parasitic trypanosomes carry a combined health burden of 2.2 million daily adjusted life years and primarily affect people from the poorest demographics in tropical and subtropical climates (Stuart et al. 2008), while in African animals, trypanosomiasis costs the livestock industry over US$ 4.5 billion every year (Yaro et al. 2016). Despite their devastating social and economic impact, these diseases remain widely under-reported; misdiagnosed, unidentified or asymptomatic cases, limited funding and the lack of a universal method for parasite detection and identification make surveillance and monitoring of these parasites difficult (Wastling and Welburn, 2011; Auty et al. 2012a; Stockdale and Newton, 2013; Franco et al. 2014).

Since the development of the first DNA-based identification methods for trypanosomes in the 1980s, the number of molecular detection techniques available (and iterations on these techniques) has increased dramatically; for examples, see the following reviews: Adams and Hamilton (2008); Taberlet et al. (2012). Although they constitute a vast improvement in sensitivity and specificity of diagnosis compared with microscopy methods (Gibson, 2007; Enyaru et al. 2010), the absence of a ‘gold standard’ for the detection and classification of trypanosomes has resulted in a distinct lack of comparable data between surveys (Auty et al. 2012b; Hernández and Ramirez, 2013; D’Avila-Levy et al. 2015). Most molecular techniques are too costly or complex for general use in front-line field diagnostics and, while developments in the transport of blood specimens have allowed samples to be analysed at centralized clinical laboratory facilities, the majority of molecular methods are still confined to research laboratories (Deborggraeve and Büscher, 2010).

Nonetheless, in other areas of biology and medicine, standardized, sequence-based barcoding (Hebert et al. 2003) has provided a sensitive, reliable method for the identification of species across a vast range of taxa and is now used by thousands of researchers worldwide (Coissac et al. 2016). However, despite the growing reference libraries of DNA barcodes for animals, plants and fungi (Ratnasingham and Hebert, 2007), there is currently no universal genetic barcoding marker available for trypanosome species. Accordingly, there is a clear need for a definitive, simple test suitable for the detection of all trypanosomes (Wastling and Welburn, 2011), with sensitivity and specificity sufficient to differentiate between infections at the subspecies level, and usable for known, unknown and mixed infections. This is particularly pertinent from an epidemiological perspective for organisms that are morphologically identical but which require different treatments, such as the two human-infective trypanosomes that cause sleeping sickness (human African trypanosomiasis, HAT), Trypanosoma brucei rhodesiense and Trypanozoon brucei gambiense (Adams and Hamilton, 2008).

A key point regarding the continued relevance (or otherwise) of sequence-based barcoding, irrespective of the target locus, is the need for such a test to provide information on unknown taxa. This represents a very different requirement from a binary yes/no diagnostic test – generally
When an infected kissing bug takes a blood meal, blood-feeding insect vectors. (A) Salivarian trypanosomes, characterized by development in the foregut of their insect vector, are confined to sub-Saharan Africa across sub-Saharan Africa. (B) Stercorarian trypanosomes, characterized by development in the hindgut of their insect vectors, are mostly non-pathogenic.

Despite the development of a variety of molecular methods for the detection and identification of infectious agents, the usual method for diagnosing trypanosome infections in vertebrate hosts remains the most basic: microscopic examination of sample preparations (Mugasa et al. 2012; Ricciardi and Ndao, 2015). However, this method is time consuming, dependent on operator expertise, unreliable for mixed infections, fails to detect immature infections and, in the case of African trypanosomes, is only useful for distinguishing between parasites to the level of subgenus (Ouma et al. 2000; Gibson, 2009; Enyaru et al. 2010; Auty et al. 2012b; Mugasa et al. 2014).

Early attempts to define the identity of pathogenic trypanosomes relied on a combination of microscopy and the ability, or otherwise, to passage parasites to the level of subgenus (Ouma et al. 2000; Gibson, 2009; Enyaru et al. 2010; Auty et al. 2012b; Mugasa et al. 2014).

For human African trypanosomiasis, microscopic examination of cerebral spinal fluid can be used to determine the stage of disease progression, but the invasive procedure (lumbar puncture) required to collect samples often discourages patients from seeking medical help. A lack of formal training for front-line medical workers, local stigma surrounding diagnosis of sleeping sickness and a delay in patients contacting medical services only exacerbates the problem of surveillance and monitoring of this disease (Mpanya et al. 2012; Acup et al. 2016).

**Timeline of trypanosome detection**

**Old faithful: microscopy**

Despite the development of a variety of molecular methods for the detection and identification of infectious agents, the usual method of infection in the insect vector (Hoare, 1972; Enyaru et al. 2010).

For human African trypanosomiasis, microscopic examination of cerebral spinal fluid can be used to determine the stage of disease progression, but the invasive procedure (lumbar puncture) required to collect samples often discourages patients from seeking medical help. A lack of formal training for front-line medical workers, local stigma surrounding diagnosis of sleeping sickness and a delay in patients contacting medical services only exacerbates the problem of surveillance and monitoring of this disease (Mpanya et al. 2012; Acup et al. 2016).

Fig. 1. Pathogenic trypanosomes of mammals. Trypanosomes are responsible for a number of diseases of both humans and animals. Chagas disease and human African trypanosomiasis (HAT) are considered ‘neglected tropical diseases’ by the World Health Organization and are transmitted between mammalian hosts by blood-feeding insect vectors. (A) Salivarian trypanosomes, characterized by development in the foregut of their insect vector, are confined to sub-Saharan Africa and are spread by the bite of the tsetse fly (Glossina spp.). These African trypanosomes, which include the human-infective T. brucei spp. and the major livestock pathogen T. congolense, cause the wasting diseases sleeping sickness (human African trypanosomiasis, HAT) and Nagana (animal African trypanosomiasis, AAT) across sub-Saharan Africa. (B) Stercorarian trypanosomes, characterized by development in the hindgut of their insect vectors, are mostly non-pathogenic. However, *Trypanosoma cruzi*, transmitted between mammalian hosts by the kissing bug (*Triatoma* spp.), causes Chagas disease, primarily in Latin America. When an infected kissing bug takes a blood meal, *T. cruzi* is passed out in the insect’s feces and is typically deposited near the bite wound. The parasite enters the host when infected feces is spread into the wound, the eyes, mouth or breaks in the skin of the unawares host. (C) Three trypanosome species, *T. evansi*, *T. equiperdum* and *T. vivax*, are the major pathogens of livestock and have become adapted to mechanical transmission; they are now transmitted by a range of biting organisms (and, in the case of *T. equiperdum*, sexual contact) and, having lost the need for their ancestral tsetse fly host, they have spread beyond Africa to become disease agents in many parts of Asia and the Americas.
**Not-so-quick kit: isoenzyme analysis**

In the late 1960s, Lanham and Godfrey developed a cellulose column-based method utilizing the differential surface charge between trypanosomes and red blood cells to reliably separate parasites from host blood (Lanham and Godfrey, 1970). With this method, they were able to obtain relatively large-scale, pure preparations of live, undisrupted parasites suitable for subsequent biochemical analysis. At around the same time, Godfrey and colleagues developed a method to characterize trypanosomes using isoenzymes (Kilgour and Godfrey, 1973) and the characterization of many trypanosome species, subspecies and strains quickly followed (e.g. Godfrey and Kilgour, 1976; Miles et al. 1977). Several major isoenzyme-based studies followed and succeeded in defining the species and groupings of epidemiological significance recognized today (e.g. Gashumba et al. 1986; Gibson et al. 1988; Godfrey et al. 1990). Attempts were made subsequently to both streamline the methodology and to optimize the discriminatory power of the enzymes used (e.g. Stevens and Godfrey, 1992; Abderrazak et al. 1993), but ultimately the practical difficulties associated with isolating and preserving parasite enzyme extracts, reproducibility and issues of homoplasy in banding patterns led to the approach being superseded by DNA-based methodologies (e.g. Gibson and Borst, 1986; Hide et al. 1990).

**Quick kit: serological tests**

Antibody-detection tests, such as the card agglutination test and the direct agglutination test, are widely used for the detection of trypanosomes in human hosts (Ricciardi and Ndao, 2015; Lutumba et al. 2016). These tests have excellent field application as they do not require a constant supply of electricity and are cheaper and more rapid than equivalent molecular techniques, although they can vary significantly in their sensitivity and specificity (Ricciardi and Ndao, 2015). Serological tests require relatively large samples and have the potential to yield false-negative results where parasitaemia is low or where antibody production is reduced, such as in immunocompromised patients (Papadopoulos et al. 2004; World Health Organisation, 2013). In addition, positive diagnoses obtained using serological tests nearly always require confirmation by microscopy, as these methods cannot distinguish between active infection and residual antigens from past infection or vaccination (Ulilenberg and Boyt, 1998; Woods, 2013). Misdiagnosis of trypanosome infections remains a major problem, as treatment often carries a significant inherent risk (Barrett and Croft, 2012; Field et al. 2017).

The enzyme-linked immunosorbent assay, offers higher sensitivity than many other serological tests available, but it requires a sophisticated laboratory set-up that has restricted its use for diagnosis in the field (Chappuis et al. 2005).

**The rise of molecular methods**

DNA probes based on non-coding satellite repeats were the first molecular methods sensitive enough for the direct identification of trypanosomes in both host and vector samples without requiring cell cultures (Kukla et al. 1987; Gibson et al. 1988; McNamara et al. 1989). The development of the polymerase chain reaction (PCR) heralded a major advance in the sensitivity of diagnostic techniques; PCR-based methods can identify trypanosomes at the subspecies level, they are suitable for analysis of mixed infections and can be applied to samples where parasite numbers are vanishingly low (Adams and Hamilton, 2008; Gibson, 2009; Matovu et al. 2010). Species-specific PCR-based methods are the most frequently used molecular tests for detection and identification of trypanosomes, but are limited by the number of species for which species-specific primers are available. Critically, these methods only detect known species: they cannot prove the absence of trypanosomes. In addition, screening samples for multiple trypanosome species using species-specific PCR methods requires a panel of probes; this can be expensive, time consuming and limits the number of samples that it is practical to analyse (Gibson, 2009; Adams et al. 2010; De Waal, 2012).

**Generic PCR-based methods and the ‘ideal’ barcode**

Historically, generic PCR methods have been less sensitive than species-specific PCR methods, but allow for multiple trypanosome species to be identified with a single test (Gibson, 2009). Most generic methods, such as restriction fragment length polymorphism PCR (RFLP-PCR) and ribosomal length-based methods, utilize multipurpose primers that target a semi-conserved region of the genome. Identification of an organism is made based on the length of the amplified regions (Adams and Hamilton, 2008). Although these methods each result in a species-specific ‘barcode’, none fulfill the requirements for the ‘ideal’ trypanosome barcode (Box 1).

**Target gene**

The success of a gene as a DNA barcode depends on a number of attributes, which must be considered when selecting gene targets: Is it a multicyclic gene? How conserved is the sequence? How much does it vary across/between taxa/species? Is this level of variation constant across the gene? Some genes have been identified as universal barcodes, and are suitable for vast groups of organisms: the mitochondrial gene cytochrome c oxidase subunit 1 (cox1/COI) is the accepted gold standard for molecular species identification of animals, and equivalents are available for plants and fungi. However, identifying universal barcodes in eukaryotic
groups has proved difficult, not least because the level of genetic variability possible within each species is poorly understood (Enyaru et al. 2010), and consensus is yet to be reached regarding which genes to target and the criteria for delimiting species groups (Pawlowski et al. 2012; Pecník and Buzan, 2014).

Molecular markers have been developed to target a wide range of trypanosome gene regions (Fig. 2A), but few have been the target of barcoding approaches (Fig. 2B). Fluorescent fragment length barcoding (FFLB) has been used to amplify small target regions in both the 18S small subunit ribosomal RNA (rRNA) and the 28S large subunit rRNA (Hamilton et al. 2008; Hamilton et al. 2011; Silva-Iiturrita et al. 2013). This highly sensitive, PCR-based method uses four sets of primers: two target the 18S and are specific to trypanosomes, two target the 28S and are specific to all trypanosomatids (Hamilton et al. 2011). The length of the resulting fragments produces a pattern unique to each species, which can be matched to reference pattern profiles for species identification. FFLB can also detect novel trypanosome species and, although further analysis is needed to identify these novel species, the fragment patterns may provide an indication of phylogenetic relationships (Hamilton et al. 2008). However, there are a limited number of reference profiles available for FFLB, which restricts its use as a trypanosome identification tool at this time (Hamilton et al. 2011; Silva-Iiturrita et al. 2013), and this method cannot be used to discriminate between T. brucei subspecies (Hamilton et al. 2008).

The 18S rRNA gene has long been a popular target for molecular detection methods in protists (D’Avila-Levy et al. 2015). It is a highly expressed multicopy gene, present in all eukaryotes, with an assortment of conserved and variable nucleotide sequences that offer targets for universal primers, whilst still providing a wealth of taxonomic information. As sequence-based molecular methods gained popularity, the 18S rRNA gene succeeded protein-coding genes (e.g. Fernandes et al. 1993; Hashimoto et al. 1995; Adjé et al. 1998) to become the gene of choice for nearly all trypanosome evolutionary analysis (Maslov et al. 1996; Lukes et al. 1997; Haag et al. 1998; Stevens et al. 1998, 1999) and, as a result, has formed the basis of all modern trypanosome taxonomic frameworks (e.g. Hamilton et al. 2007; Lima et al. 2015; Dario et al. 2017). However, while nearly all trypanosome phylogenies have been constructed using 18S rRNA sequences, inadequate signals at certain depths of phylogenetic reconstruction have necessitated the use of additional trypanosome gene markers such as the glyceraldehyde phosphate dehydrogenase (GAPDH) gene. Nonetheless, the framework described using the 18S rRNA has proven robust; other gene markers have complemented and strengthened this framework without fundamentally changing the nature of the basic relationships described based on 18S rRNA data; ultimately, this framework has also been fully supported by whole-genome phylogenetic comparisons (Leonard et al. 2011). In addition, as the 18S rRNA is one of the most widely used markers for trypanosomes, it is well represented in sequence databases such as GenBank (D’Avila-Levy et al. 2015).

Another popular molecular marker is the GAPDH gene. Few if any gaps are required for alignment of trypanosome GAPDH sequences, and sequences are shorter than those of 18S rRNA; sequencing this ‘housekeeping’ gene can be more economical, but provides a complementary depth of phylogenetic information in trypanosomes (Hamilton et al. 2004; Adams and Hamilton, 2008). GAPDH genes are relatively conserved and are therefore useful for resolving deep phylogenetic relationships (Hamilton et al. 2004). However, in order to determine close relationships, GAPDH must be used in conjunction with another barcoding marker; GAPDH has been used successfully with the 18S rRNA for trypanosome identification, and has proven suitable for novel species and mixed infections (e.g. Hamilton et al. 2008; Barbosa et al. 2016).

Internal transcribed spacer (ITS) regions have been widely used for barcoding in some organisms, e.g. fungi (Pawlowski et al. 2012); however, while they have long been utilized for the detection of trypanosomes (Desquesnes and Davila, 2002; Adams et al. 2008; Desquesnes et al. 2011; Hernández and Stevens, 2013), whilst cluster analysis has been used to delimit species when targeting the V7-V8 regions of the 18S rRNA gene (Lima et al. 2015) and regions of SL RNA (Vatyak et al. 2010). Definitions: SS, SS rRNA gene insertions; 185, 18S small subunit ribosomal RNA (rRNA); 205, editosome protein complex; 285, 28S large subunit rRNA; 70 kDa, 70 kDa heat shock protein; GAPDH, glyceraldehyde phosphate dehydrogenase; ITS, internal transcribed spacer; SL RNA, spliced leader RNA; SRA, serum resistance associated gene. Meta analysis method: A literature search of the Web of Science (Clarivate Statistics, 2017) database was conducted in September 2017 to identify gene regions used for the identification or classification of trypanosomatids in the past 25 years. Subsequent searches were conducted to quantify the number of papers published between 1993 and 2017 that utilized each identified gene region (full Boolean search criteria available in Supplementary Information).

Fig. 2. (A) Gene regions commonly used for the identification of trypanosomes. (B) Gene regions that have also been the target of barcoding techniques for trypanosomes. The 18S and 28S rRNA regions have both been targeted using fluorescent fragment length barcoding (FFLB) (Hamilton et al. 2008, 2011; Silva-Iiturrita et al. 2013), whilst cluster analysis has been used to delimit species when targeting the V7-V8 regions of the 18S rRNA gene (Lima et al. 2015) and regions of SL RNA (Vatyak et al. 2010). Definitions: SS, SS rRNA gene insertions; 185, 18S small subunit ribosomal RNA (rRNA); 205, editosome protein complex; 285, 28S large subunit rRNA; 70 kDa, 70 kDa heat shock protein; GAPDH, glyceraldehyde phosphate dehydrogenase; ITS, internal transcribed spacer; SL RNA, spliced leader RNA; SRA, serum resistance associated gene. Meta analysis method: A literature search of the Web of Science (Clarivate Statistics, 2017) database was conducted in September 2017 to identify gene regions used for the identification or classification of trypanosomatids in the past 25 years. Subsequent searches were conducted to quantify the number of papers published between 1993 and 2017 that utilized each identified gene region (full Boolean search criteria available in Supplementary Information).
Ramírez, 2013), they have not yet been used specifically for the barcoding of different trypanosome species. Identification of species depends on the length of the amplified fragments of ribosomal RNA produced via PCR using primers complementary to conserved regions of the 18S, 28S and 5·8S rRNA genes matching all species of interest. This means species determination is possible for mixed infections, except in cases where the amplicon length is similar between species or there is intra-species variation (Adams and Hamilton, 2008; Hamilton et al. 2008; Gibson, 2009). Another constraint of the ITS region, as with all mitochondrial genes as targets for barcoding, is its relatively low copy number (100–200 repeats), compared with that of satellite DNA (10 000–20 000 repeats), which can limit the sensitivity of tests (Desquesnes and Davila, 2002).

The kinetoplast is a modified mitochondrion unique to kinetoplastid protists and kinetoplast DNA (kDNA) minicircles have been successfully used in PCR assays for the identification of a number of Trypanosoma species. The high copy number of these minicircles – several thousand per cell – lends itself to highly sensitive diagnostics. However, high levels of nucleotide polymorphism between repeats of kDNA fragments make these genes unsuitable for sequence alignment (De Oliveira Ramos Pereira and Brandão, 2013). Only very short regions (100–200 base pairs) of kDNA minicircles are conserved and for some trypanosomes, such as T. brucei, there is only one of these regions per minicircle (Jensen and Englund, 2012). Low levels of conserved sequences in kDNA make it difficult to develop universal primers and limit the depth of phylogenetic information that can be elucidated from these sequences.

Spliced leader RNA (SL RNA) or ‘mini-exon donor RNA’ is another feature unique to kinetoplastid protists and has also been used as a target for barcoding (Rodrigues et al. 2010; Lima et al. 2015). The SL RNA genes are arranged as tandem repeats, with each repeat comprising many repeat units with regions of differing variability (Rodrigues et al. 2010). The conserved regions are convenient for primer targeting, whilst the more variable intergenic regions permit distinction between closely related trypanosomes (Westenberger et al. 2004). However, there are no primers currently available that are applicable to all trypanosomes (D’Avila-Levy et al. 2015), and the high mutation rate of intergenic regions makes it difficult to compare sequences across the full spectrum of trypanosomes or to define any meaningful phylogeny beyond closely related taxa (Gibson et al. 2000). Previous attempts to use SL RNA barcodes for trypanosomes delimited species using an arbitrary level of sequence similarity (90%) (Votýpka et al. 2010). However, this threshold is insufficient for discriminating between closely related Trypanosoma species that share up to 98% similarity in their SL transcripts (Gibson et al. 2000).

A significant (and pragmatic) consideration when choosing a target gene for barcoding is the availability of sequences. Protists are poorly represented in sequence libraries and comprise just over 2% of the sequences currently in GenBank (National Center for Biotechnology Information (NCBI), 2017), despite constituting the majority of samples in environmental surveys (Del Campo et al. 2015). In addition, the availability of Trypanosoma species is further skewed towards human-infective species and those infecting important agricultural species, such as cattle, which are over-represented in sequence databases relative to other trypanosomes, including parasites of insects and plants (D’Avila-Levy et al. 2015).

Whilst a bias towards medically important parasites is understandable, the paucity of genomic data from other Trypanosoma is a continuing impediment to our understanding of the evolutionary history and intricate phylogenetic relationships within this diverse group of parasites.

Genes or genes?

As the number of genes scrutinized for their barcoding potential has increased, it has become apparent that no test amplifying a single fragment has the differential power necessary to fully and reliably resolve the phylogeny of all trypanosomes (Hamilton et al. 2007; Adams and Hamilton, 2008; Pompanon and Samadi, 2015). Barcoding methods that utilize multiple loci have the advantage of additional power and accuracy (Mallo and Posada, 2016), and nested strategies that utilize ‘a universal pre-barcode’ and a ‘group specific’ barcode have been proposed by the Protist Working Group (ProWG) as alternative methods to resolve interspecies relationships (Pawłowski et al. 2012). In addition, we anticipate that the increasing ease and ever reducing costs of genome-wide SNP discovery in non-model organisms will lead to major advances in the use of SNP chip-based diagnostics in the near future.

Optimizing fragment length

In the past, target fragment length has been limited by the technology available. When molecular methods were first introduced, sequencing was only possible up to a few hundred base pairs. However, with the growth of Next-Generation Sequencing, the cost of sequencing has decreased by a factor of $10^4$ in the last 10 years (Hayden, 2014; Van Nannenwegen et al. 2016).

But is bigger always better?

Should we strive for barcode fragments with a length at the ever-increasing limit of our sequencing ability? Here, there is a significant trade-off to consider: optimal sequence length of the target region is highly dependent on the user’s requirements. Shorter fragments result in higher sensitivity tests, favourable for analysis of degraded DNA from field samples. In diagnostic or clinical situations, for example, where the objective is to discriminate between the two human-infective subspecies of T. brucei, a shorter fragment is likely to provide all the required information. However, it is only with longer fragments that we can infer robust phylogenetic information at the subspecies level (Pompanon and Samadi, 2015); recreating the evolutionary history of a collection of poorly known or newly discovered species is likely to call for a very long target region, though this is, of course, a very different task than routine, high-throughput barcoding of large numbers of specimens.

An alternative future for diagnostics? Isothermal techniques

The use of isothermal amplification molecular methods, such as loop-mediated isothermal amplification and nucleic acid sequence-based amplification are becoming increasingly popular for the detection of trypanosomes as they offer simple, rapid and cheap alternatives to traditional PCR-based methods (Mugasa et al. 2014; Besuschio et al. 2017; Rivero et al. 2017). Isothermal tests involve a single reaction in a single tube incubated at a constant temperature; therefore, these techniques do not require the expensive thermocycling equipment that is necessary for PCR (Matovu et al. 2010; Wastling and Welburn, 2011). The simplicity, sensitivity and low cost of isothermal techniques make them strong candidates for the application of molecular methods in field diagnostics in resource-poor areas (Laohasinnarong, 2011; Ricciardi and Ndao, 2015). However, a number of additional costs must be considered when evaluating the suitability of these methods for field diagnostics, including the need for six primers, heating and maintaining samples at 65 °C, and expensive dyes for visualization of results (Enyaru et al. 2010; Wastling and Welburn, 2011). In addition, the ability of these tests to amplify extremely small amounts of
DNA mean that they are highly prone to contamination. Developing simplified ‘kit’ forms of these techniques, and refining those already available, may yield promising alternatives to sequence-based barcoding for clinical purposes (Mugasa et al. 2014).

**Discussion**

**Towards a spectrum of similarity: an alternative approach to barcoding**

Rather than identifying species by the length of their amplified fragments, we propose the adoption of a technique that identifies species, within a defined group, by the level of concordance across a selected gene, e.g. 18S RNA, or partial gene (Fig. 3). Sequence differences between a cohort of species are tracked along a specified gene, highlighting regions rich with phylogenetic variety. Species can then be identified by the degree of similarity across the selected region(s), for example, see Stevens and Wall (2001). The resulting spectrum of similarity can provide a valuable tool for understanding the relative level of sequence differentiation of any putative species, as their place in the spectrum will provide clues as to their phylogenetic placement.

Such an approach offers several benefits, including (as with any barcoding approach) the adoption of a standardized marker (or set of markers) and the ability to compare findings across studies, together with the practical benefits of being able to utilize a limited number of standardized primers. In the ‘sliding window’ approach proposed by Stevens and Wall (2001), the use of a given molecular marker in conjunction with a particular group of taxa allows the gene region (to be adopted for subsequent barcoding) to be selected based on the degree of phylogenetic resolution delivered by the particular sequence positions used within the target gene. More recently, Hadziavdic et al. (2014) undertook a much broader study along similar lines, screening for variation across more than 500,000 eukaryote 18S rRNA sequences (see also Pawlowski et al. (2012) for a review of the potential role of the V4 region of 18S rRNA as a candidate universal barcoding marker). Such approaches go a long way towards fulfilling the requirements for marker selection as set out in Box 1. To date, however, while several studies have focused on the use of the V7–V8 sub-region of 18S rRNA (e.g. Smith et al. 2008; Averis et al. 2009), citing its phylogenetic informativeness (but, see Hamilton and Stevens, 2011), this approach remains to be systematically applied across the full 18S rRNA gene in trypanosomes.

**Can barcoding be all things to all people?**

The ideal barcode from a gene region that yields enough sequence variation to capture the vast diversity of trypanosomes may provide a level of discrimination sufficient for diagnostic and identification purposes. However, it is questionable whether the same barcode could also provide enough variation to fully capture the phylogenetic relationships or complex evolutionary history of such a diverse group of organisms. In cases where genetic functionality is the key interest, barcoding is likely to be of little use. In the field, adequate preservation methods would be required to maintain the integrity of DNA from samples in order to apply any barcoding method successfully (Reeves et al. 2016).

The development of a perfect and truly universal barcode, based on a single primer pair, may be not only unattainable but also impractical. Different avenues of research have different requirements, in terms of both the techniques they use and the information required/acquired. A geneticist studying the evolution of trypanosomes needs a way to detect intricate relationships over a range of evolutionary timescales (from, for example, the (putatively) most ancient to most recent: Simpson et al. 2004; Flegontov et al. 2013; Hamilton et al. 2004; Stevens & Rambaut, 2001; Haag et al. 1998; Lima et al. 2015; Balmer et al. 2011; Messenger et al. 2012), and it may be that a suite of gene markers is required to provide sufficient detail at all levels of phylogenetic depth. Conversely, for a clinician diagnosing patients in a resource-poor community, the nuances of an organism’s evolutionary history are all but irrelevant. Identification of the parasite often determines treatment, so in this case the sensitivity and specificity of a diagnostic test becomes the overriding priority.

The range of requirements for the detection and identification of trypanosomes must be considered when selecting gene targets for barcoding, and the benefits of each molecular marker weighed against its limitations. For example, SL RNA is an ideal marker for detection of parasites in field samples, as this region is not present...
Concluding remarks and future directions

At present, molecular methods are mostly used only in sophisticated research laboratories, and there is a concern that new techniques are ‘merely another addition to an ever-expanding toolbox of molecular assays for research’ (Wastling and Welburn, 2011), rather than having any clinical diagnostic utility. And, whilst there has been a drive to develop and refine new molecular diagnostics, the sensitivity of existing techniques may be greatly improved if more research was conducted on initial stages, such as sample preparation and DNA extraction (Dunlop et al. 2014). However, recent developments in molecular methods for trypanosome identification have succeeded in unveiling a number of previously unidentified species (Adams et al. 2010; Hutchinson and Gibson, 2015) and may offer new opportunities for the identification of novel hybrids (Koffi et al. 2015; Tihon et al. 2017) and the epidemiological tracking of trypanosome strains spread by the movement of host cattle (Févre et al. 2005). Nonetheless, a lack of comparable data between parasite surveys makes it difficult to draw any firm conclusions regarding species prevalence, and the full extent of trypanosome diversity remains unknown at this time (Adams et al. 2010; D’Avila-Levy et al. 2015). Priority should be given to the establishment of a standardized barcoding protocol for the detection and identification of trypanosomes (matching as close as possible the criteria given in Box 1). A standard barcoding protocol with requirement-dependant refinements is likely to be the closest we can ever come to obtaining a truly universal barcode for trypanosomes.

Data statement

The data research supporting this publication can be publicly accessed at the Parasitology journal website as Supplementary Information and/or by contacting the corresponding author.

Supplementary Material

The supplementary material for this article can be found at
https://doi.org/10.1017/S0031182017002049.

Author ORCIDs. R. Hutchinson: orcid ID 0000-0003-0336-9594; J. R. Stevens: orcid ID: 0000-0002-1317-6721.

Financial support. R. H. is funded by a UK Biotechnology and Biological Sciences Research Council PhD studentship (grant No. BB/M009122/1) held by Professor Wendy Gibson (Bristol) and J. R. S., as part of the South West Doctoral Training Partnership (SWBio DTP).

References


flagellates in the high-throughput DNA sequencing era. Memórias Do Instituto Oswaldo Cruz 110, 956–965.


