CrossMark

High-throughput sequencing of kDNA amplicons for the analysis of *Leishmania* minicircles and identification of Neotropical species

ARTHUR KOCHER[1,2]*, SOPHIE VALIÈRE[3], ANNE-LAURE BAÑULS[2] *and* JÉRÔME MURIENNE[1]

[1] *Laboratoire Evolution et Diversité Biologique, CNRS, Université Toulouse III Paul Sabatier, ENSFEA, IRD, UMR5174 EDB, Toulouse, France*
[2] *UMR MIVEGEC, IRD 224 – CNRS 5290 – Université de Montpellier, F34394 Montpellier, France*
[3] *GeT–PlaGe, Genotoul, INRA Auzeville, 31326 Castanet-Tolosan, France*

(*Received 25 May 2017; revised 14 September 2017; accepted 3 October 2017; first published online 16 November 2017*)

SUMMARY

Leishmania kinetoplast DNA contains thousands of small circular molecules referred to as kinetoplast DNA (kDNA) minicercles. kDNA minicircles are the preferred targets for sensitive *Leishmania* detection, because they are present in high copy number and contain conserved sequence blocks in which polymerase chain reaction (PCR) primers can be designed. On the other hand, the heterogenic nature of minicircle networks has hampered the use of this peculiar genomic region for strain typing. The characterization of *Leishmania* minicirculomes used to require isolation and cloning steps prior to sequencing. Here, we show that high-throughput sequencing of single minicircle PCR products allows bypassing these laborious laboratory tasks. The 120 bp long minicircle conserved region was amplified by PCR from 18 *Leishmania* strains representative of the major species complexes found in the Neotropics. High-throughput sequencing of PCR products enabled recovering significant numbers of distinct minicircle sequences from each strain, reflecting minicircle class diversity. Minicircle sequence analysis revealed patterns that are congruent with current hypothesis of *Leishmania* relationships. Then, we show that a barcoding-like approach based on minicircle sequence comparisons may allow reliable identifications of *Leishmania* spp. This work opens up promising perspectives for the study of kDNA minicercles and a variety of applications in *Leishmania* research.

Key words: Detection, diagnosis, Illumina, *Leishmania Viannia*, leishmini primers, New World, PCR, sand fly.

INTRODUCTION

Leishmaniases are a group of diseases caused by flagellate parasites of the genus *Leishmania* Ross, 1903 (Kinetoplastida: Trypanosomatidae). They are endemic in approximately 100 countries in Africa, America, Asia and Europe, and cause more than one million cases and several thousand deaths each year (Alvar *et al.* 2012). More than 50 *Leishmania* species have been described to date, among which approximately 20 are known to infect humans (Maroli *et al.* 2013; Akhoundi *et al.* 2016). These may cause various clinical forms (e.g. localized cutaneous, disseminated cutaneous, mucocutaneous or visceral leishmaniasis) and respond differently to treatment (Copeland and Aronson, 2015). Furthermore, each *Leishmania* species is characterized by a specific transmission cycle involving distinct sand fly vectors and reservoir hosts (Lainson and Shaw, 2010; Maroli *et al.* 2013; Ready, 2013). Therefore, in areas where several *Leishmania* species are found in sympatry or when imported cases can occur, reliable identification of strains is crucial for medical management and eco-epidemiological studies.

For more than 40 years, multilocus enzyme electrophoresis (MLEE) has allowed considerable advances in the delineation and classification of *Leishmania* species (Gardener and Howells, 1972; Kreutzer *et al.* 1983; Rioux *et al.* 1990; Thomaz-Soccol *et al.* 1993). Nevertheless, this technique is time-consuming and relies on constraining steps of parasite isolation and culture. Polymerase chain reaction (PCR)-based methods are increasingly used as alternatives to MLEE for *Leishmania* typing because they are cost and time-effective and can be performed on small quantities of material. In particular, restriction fragment length polymorphism (RFLP) analyses (Marfurt *et al.* 2003; Rotureau, 2006; Koarashi *et al.* 2016) and direct comparisons of DNA sequences (Zelazny *et al.* 2005; Marco *et al.* 2006; de Almeida *et al.* 2011) have been widely employed in recent years. Various genomic regions have been targeted for molecular identification of *Leishmania*, in a quest for the best compromise between sensitivity, specificity and taxonomic coverage (Akhoundi *et al.* 2017),

* Corresponding author. Laboratoire EDB, 118 route de Narbonne, 31062 Toulouse CEDEX 9, France. E-mail: arthur.kocher@gmail.com

Parasitology (2018), **145**, 585–594. © Cambridge University Press 2017
doi:10.1017/S0031182017002013

https://doi.org/10.1017/S0031182017002013 Published online by Cambridge University Press

Table 1. List of the strains used in this study

Short code	WHO code	Species	Country	Lab./ref.
2007-24	MHOM/GF/2007/2007–24	*L. amazonensis*	French Guiana	(Simon *et al*. 2010)
LV78	RAT/BR/72/LV78	*L. amazonensis*	Brazil	CRB-L[a]
2006-9	MHOM/GF/2006/2006–9	*L. braziliensis*	French Guiana	(Simon *et al*. 2010)
LC1412	MHOM/PE/91/LCI412	*L. braziliensis*	Peru	(Dujardin *et al*. 1995)
LH699	MHOM/PE/00/LH699	*L. braziliensis*	Peru	(Victoir *et al*. 1998)
LH754	MHOM/PE/89/LH754	*L. braziliensis*	Peru	(Dujardin *et al*. 1995)
M2904	MHOM/BR/75/M2904	*L. braziliensis*	Brazil	CRB-L
2006-3	MHOM/GF/2006/2006–3	*L. guyanensis*	French Guiana	(Simon *et al*. 2010)
2007-1	MHOM/GF/2007/2007–1	*L. guyanensis*	French Guiana	(Simon *et al*. 2010)
M5378	MHOM/BR/78/M5378	*L. guyanensis*	Brazil	CRB-L
HZ2008	MHOM/GF/2008/HZ2008	*L. guyanensis*	French Guiana	CRB-L
TD2008	MHOM/GF/2008/TD2008	*L. guyanensis*	French Guiana	CRB-L
ITMAP263	MHOM/MA/67/ITMAP263	*L. infantum*	Marocco	CRB-L
LEM417	MHOM/DZ/1982/LIPA59	*L. infantum*	Algeria	CRB-L
2006-40	MHOM/GF/2006/2006–40	*L. lainsoni*	French Guiana	(Simon *et al*. 2010)
LC2288	MHOM/PE/91/LC2288	*L. lainsoni*	Peru	(Bastrenta *et al*. 2002)
LEM2204	MDAS/BR/79/M5533	*L. naiffi*	Brazil	CRB-L
LEM5108	MHOM/GF/2005/LEM5108	*L. naiffi*	French Guiana	CRB-L

FG, French Guiana; B, Brazil; P, Peru; M, Marocco; A, Algeria.
[a] Centre de Ressources Biologiques des *Leishmania*, Monptellier, France.

but no real consensus has been reached in this regard.

Leishmania parasites, such as all kinetoplastids, are characterized by a unique mitochondrion containing a remarkably large and dense genome: the kinetoplast. Kinetoplast DNA (kDNA) represents 10–20% of total cellular DNA (Simpson, 1987) and is composed of two types of circular molecules interlinked in a concatenated network: maxicircles and minicircles. Maxicircles are *c*. 20 kb long, present in a few tens of copies, and are analogous to the mitochondrial genome of other eukaryotes. Minicircles are *c*. 800 bp long, present in several thousands of copies. They encode for guide RNAs (gRNAs) involved in the maturation of maxicircle's messenger RNAs through an RNA editing mechanism (Read *et al*. 2015). Due to their extremely high copy number, minicircles are ideal targets for highly sensitive detection of *Leishmania*. Furthermore, the presence of well-conserved sequence blocks (CSBs) in their replication origin (Jensen and Englund, 2012) allows the design of PCR primers with wide taxonomic coverage (Noyes *et al*. 1998; Ceccarelli *et al*. 2014).

On the other hand, the analysis of minicircle sequences for *Leishmania* species identification is difficult. The pool of minicircles (minicirculome) present in each parasitic cell encompasses a set of distinct minicircle versions, referred to as classes, encoding different gRNAs (Hajduk and Ochsenreiter, 2010). Except in the CSBs, high sequence variability is found among minicircle classes (Brewster and Barker, 2002), resulting in difficulties for their analysis (de Oliveira Ramos Pereira and Brandão, 2013). Furthermore, there are redundant gRNAs that encode the same editing information, and the number and identity of minicircle classes vary greatly between species and even between strains of the same species (Simpson, 1997; Gao *et al*. 2001). These features make classical DNA-based identification methods difficult to perform. Thus, in many studies, minicircle-based PCR assays are employed as highly sensitive screening tests for *Leishmania* detection, while other genomic regions are targeted in a second step for the identification of positive samples (Cássia-Pires *et al*. 2014; Richini-Pereira *et al*. 2014; Berzunza-Cruz *et al*. 2015; Pereira Júnior *et al*. 2015).

The characterization of *Leishmania* minicirculomes used to require isolation and cloning steps prior to sequencing (Lee *et al*. 1993; Brewster and Barker, 2002; Telleria *et al*. 2006; Rodrigues *et al*. 2013). Here, we show that high-throughput sequencing of minicircle PCR amplicons may be an efficient alternative in this regard. We then present a barcoding-like approach based on minicircle sequence comparisons, which aims at combining sensitive detection and reliable identification of *Leishmania* spp. The method is evaluated for a use in the Neotropical region, where a high diversity of *Leishmania* species may be found in sympatry (Lainson and Shaw, 2010).

MATERIALS AND METHODS

Leishmania strains and DNA extraction

Eighteen *Leishmania* strains were selected, representative of both *Leishmania* subgenera and of all major species complexes found in the neotropics (Table 1).

https://doi.org/10.1017/S0031182017002013 Published online by Cambridge University Press

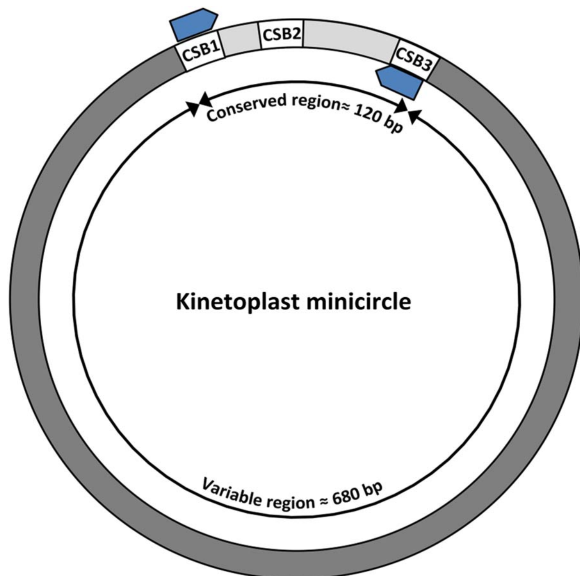

Fig. 1. Structural organization of *Leishmania* kinetoplast minicircles. The three conserved sequence blocks (CSBs) are indicated in white boxes. Blue arrows indicate the binding sites of PCR primers used to amplify the *c*. 120-bp-long fragment within the conserved region.

They belonged to six species (as identified by MLEE or PCR–RFLP; (Simon *et al*. 2010)): *L*. (*Viannia*) *braziliensis*, *L*. (*V*.) *guyanensis*, *L*. (*V*.) *lainsoni*, *L*. (*V*.) *naiffi*, *L*. (*Leishmania*) *amazonensis* and *L*. (*L*.) *infantum*. When possible, we included several strains isolated in distant locations for each *Leishmania* species. *Leishmania* parasites were cultured as previously described (Faye *et al*. 2010). DNA was extracted from culture pellets using the Pure Link Genomic DNA Kit (Life Technologies, Saint-Aubin, France).

PCR amplification

PCR amplification targeted the *c*. 120 pb portion of kDNA minicircle known as the minicircle conserved region. We used PCR primers located within CSB1 and CSB3, similar to previously published ones (Fig. 1; Roque *et al*. 2010; Weirather *et al*. 2011), but further adjusted based on a large set of complete minicircle sequences (leishmini-F: 5′-GGKAGG GGCGTTCTGC-3′; leishmini-R: 5′-STATWTT ACACCAACCCC-3′). Our aim was to ensure that all minicircle classes will be amplified in every *Leishmania* spp.

Amplification was performed in 20 μL mixtures containing 2 μL of DNA template, 10 μL of AmpliTaq Gold PCR Master Mix® (5 U.μL^{-1}; Applied Biosystems, Foster City, CA, USA), 2·5 μL of each primer (5 μM), and nuclease-free water (Promega, Madison, WI, USA). The PCR mixture was denatured at 95 °C (10 min) and followed by 35 cycles of 30 s at 95 °C, 30 s at 60 °C and 15 s at 72 °C, completed at 72 °C for 5 min. To enable the sequencing of multiple PCR products in a single high-throughput sequencing run, tags of eight base pairs with at least five differences between them were added at the 5′ end of each primer (Binladen *et al*. 2007). In order to evaluate the frequency of tag-switching events, which may result in some sequences being assigned to the wrong sample (Schnell *et al*. 2015), we only used one out of two available tag combinations, as suggested by (Esling *et al*. 2015). PCR sensitivity was checked on a serial dilution of *Leishmania* DNA followed by electrophoresis and visualization of PCR product in a 2% agarose gel containing SYBR Safe stain (Invitrogen) under UV light.

High-throughput sequencing of minicircle amplicons

PCR products obtained from the 18 *Leishmania* strains were pooled and sent to the GeT-PlaGe core facilities of GenoToul (Toulouse, France) for library construction and sequencing on an Illumina Hiseq 3000 platform (Illumina, San Diego, CA, USA). Quality filtering was performed with the Consensus Assessment of Sequence and Variation (CASAVA) pipeline. Sequence data were stored on the NG6 platform (Mariette *et al*. 2012).

Sequencing reads were analysed with OBITOOLS (Boyer *et al*. 2016), as described elsewhere (Kocher *et al*. 2017*a*, 2017*b*). Paired-end reads were aligned and merged, taking into account the Phred quality scores for consensus construction. Merged reads were assigned to a given sample based on the primer tags with two mismatches allowed. Low-quality reads (exhibiting alignment scores <50, containing Ns or shorter than 50 bp) were discarded and then dereplicated in each sample. To limit errors arising from tag-switching events, we discarded all sequences that had coverage below that of the most abundant sequence found with an unexpected tag combination. The bash script used for these bioinformatic steps is available in the Supplementary Material.

Minicircle sequence analysis and use for species identification

Because multiple alignments of minicircle sequences are problematic (de Oliveira Ramos Pereira and Brandão, 2013), we performed pairwise alignments of each sequence pair based on longest common subsequences (Needleman–Wunsch alignment with a score of one for matches and zero for mismatches and gaps). Raw distances were then used to generate a neighbour-joining (NJ) tree with the R package ‘ape’ (Saitou and Nei, 1987; Paradis *et al*. 2004; R Core team, 2014), to visualize clustering patterns among *Leishmania* strains and species. Distance statistics were computed using the R package ‘spider’ (Brown *et al*. 2012). In order to provide an

https://doi.org/10.1017/S0031182017002013 Published online by Cambridge University Press

alternative visualization of the data, we constructed a haplotype network. Minicircle sequences were aligned using muscle v. 3.8.31 (Edgar, 2004) and highly variable regions were removed using Gblocks 0·91b (Castresana, 2000) with default parameters. The network was constructed using PopART (popart.otago.ac.nz) with the minimum spanning network method (Bandelt *et al*. 1999).

Our aim was then to use a barcoding-like strategy to perform *Leishmania* species identifications based on minicircle sequences. For each *Leishmania* strain, several minicircle classes were expected to be amplified and recovered by high-throughput sequencing. Therefore, taxonomic assignments of *Leishmania* spp. required the use of a multilocus approach. However, contrary to multilocus barcoding, where several markers are sequenced separately (Fazekas *et al*. 2008; Dufour *et al*. 2017), we did not have *a priori* knowledge of the identity of each minicircle fragments retrieved. It was thus not possible to perform sequence alignment for each minicircle classes independently. Additionally, the set of minicircle classes may vary between strains, even within the same *Leishmania* species (Simpson, 1997; Gao *et al*. 2001). Thus, there was no guaranty that each minicircle sequence recovered from a strain would find a homologue among those of other conspecific references. Furthermore, some minicircle classes may be species-specific, while others are shared between species (Brewster and Barker, 2002). Therefore, we had to consider that only a fraction of sequences would be informative for taxonomic assignments.

As a consequence, when comparing a set of minicircle sequences from an unidentified *Leishmania* strain to that of reference strains, each may either (i) match closely with one or several homologous references all belonging to the same species (i.e. specific match); (ii) match closely with several homologous references from various species (i.e. unspecific match); or (iii) not match closely with any available reference (i.e. no homologue, Fig. 2). Eventually, the taxonomic assignment of the strain should be based on the subset of minicircles for which specific matches were found.

In practice, for a given unidentified strain, we used the ecotag program (included in the OBITOOLS package) to perform the taxonomic assignment of each minicircle sequence. ecotag includes a 'last common ancestor algorithm' that allows one to distinguish between specific and unspecific matches. The strain was then taxonomically assigned according to the most frequent specific match, after weighting by sequence abundances (Fig. 2). The proportion of specific matches supporting the resulting identification was used as an identification score.

To evaluate the accuracy of the method, we performed a 'leave one out' testing. Each strain (i.e. the full set of minicircle sequences) was consecutively considered unknown and identified using the method described above with the remaining strains as references. Bash and R scripts used to perform the taxonomic assignment of strains and to compute identification scores are available in the Supplementary Material.

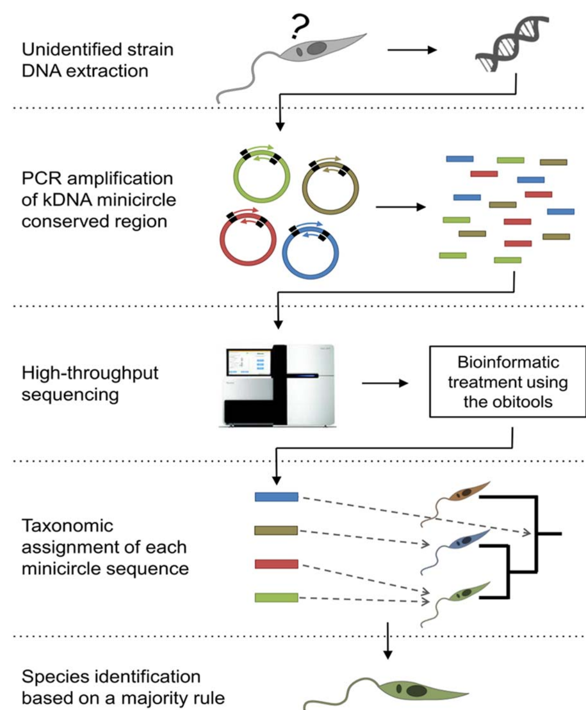

Fig. 2. Schematization of the pipeline used for the identification of *Leishmania* spp. based on high-throughput sequencing of *c*. 80-bp-long kDNA minicircle amplicons.

RESULTS

High-throughput sequencing of minicircle amplicons

PCR amplification of kDNA minicircles allowed the detection of all *Leishmania* species down to 1 fg of template DNA (i.e. below 10^{-2} parasite equivalent; see Supplementary material). High-throughput sequencing of PCR products allowed obtaining a total of 10 347 227 pairs of 150 bp long reads. After paired-end merging, quality filtering removed 0·23% of reads. The 670 890 unique sequences remained after reading dereplication. The most abundant sequence assigned to a non-used tag combination had a coverage of 1172, which was used as a threshold to remove sequences potentially originating from tag-switching events. Remaining sequences had an average read coverage of 4380. Eventually, the number of minicircle sequences retrieved ranged from 10 to 84 depending on the strain, for an average of 37·2 (Table 3; GenBank acc. KY699529–KY700198). Minicircle amplicon length ranged from 80 to 82 bp.

https://doi.org/10.1017/S0031182017002013 Published online by Cambridge University Press

Table 2. Minicircle sequence distance statistics by *Leishmania* species

Species	Number of sequences	Mean intraspecific distance (%)	Mean intrastrain distance (%)	Mean interstrain distance (%)
L. amazonensis	62	18·9	18·6	19·2
L. braziliensis	133	10·6	10·7	10·5
L. guyanensis	196	10·4	10·5	10·4
L. infantum	94	12·6	12·6	12·5
L. lainsoni	97	14·7	14·5	14·8
L. naiffi	88	10·3	10·4	10·3

Table 3. Taxonomic identification of the strain using a leave-one-out procedure

Strain	Species	Number of sequences	Proportion of taxonomic assignments (%)[a]							Final identification	Score (%)
			LA	LB	LG	LI	LL	LN	NS		
2007_24	*L. amazonensis*	41	100	0	0	0	0	0	0	*L. amazonensis*	100
LV78	*L. amazonensis*	21	100	0	0	0	0	0	0	*L. amazonensis*	100
2006_9	*L. braziliensis*	26	0	66·4	0	0	0	1·6	32	*L. braziliensis*	97·6
LC1412	*L. braziliensis*	30	0	78·2	2·7	0	0	0	19·1	*L. braziliensis*	96·7
LH699	*L. braziliensis*	37	0	75·5	3	0	0	0	21·5	*L. braziliensis*	96·2
LH754	*L. braziliensis*	23	0	75	1·5	0	0	0	23·5	*L. braziliensis*	98
M2904	*L. braziliensis*	17	0	74·4	3·4	0	0	0	22·2	*L. braziliensis*	95·6
2006_3	*L. guyanensis*	12	0	0	100	0	0	0	0	*L. guyanensis*	100
2007_1	*L. guyanensis*	40	0	1·5	92·7	0	0	0	5·8	*L. guyanensis*	98·4
M5378	*L. guyanensis*	44	0	5·1	61·3	0	0	1·4	32·3	*L. guyanensis*	90·4
HZ2008	*L. guyanensis*	58	0	10·7	81·7	0	0	0	7·5	*L. guyanensis*	88·4
TD2008	*L. guyanensis*	42	0	0	94·3	0	0	0	5·7	*L. guyanensis*	100
ITMAP263	*L. infantum*	10	0	0	0	56	0	0	44	*L. infantum*	100
LEM417	*L. infantum*	84	0	0	0	69·7	0	0	30·3	*L. infantum*	100
LC2288	*L. lainsoni*	51	0	0	0	0	85·8	0	14·2	*L. lainsoni*	100
2006_40	*L. lainsoni*	46	0	0	0	0	100	0	0	*L. lainsoni*	100
LEM2204	*L. naiffi*	50	0	0	0	0	0	93·7	6·3	*L. naiffi*	100
LEM5108	*L. naiffi*	38	0	0	0	0	0	90·8	9·2	*L. naiffi*	100

LA, *L. amazonensis*; LB, *L. braziliensis*; LG, *L. guyanensis*; LI, *L. infantum*; LL, *L. lainsoni*; LN, *L. naiffi*; NS, not specifically assigned.
[a] Proportion of sequences reads assigned to each species.

Minicircle sequence analysis and use for species identification

Based on the pairwise alignment of minicircle sequences, average intraspecific and interspecific raw genetic distances were 11·5 and 26·4%, respectively. The highest mean intraspecific distance was found in *L. amazonensis* (18·9%; Table 2), while the lowest was found in *L. naiffi*, *L. guyanensis* and *L. braziliensis* (10·3, 10·4 and 10·6%, respectively). In the NJ tree, minicircle sequences of each strain did not form single monophyletic clusters, but were rather distributed in several clusters comprising sequences of other strains, generally belonging to the same species (Fig. 3). This pattern was also highlighted by very similar mean intrastrain and interstrain genetic distance within each species (Table 2). At the species level, several patterns were observed. Almost all sequences were grouped in single species-specific clusters for *L. amazonensis*, *L. infantum* and *L. lainsoni*. For the three other species (*L. naiffi*, *L. braziliensis* and *L. guyanensis*), minicircle sequences formed several interleaved clusters. Some of these were well delimited and species-specific, while others included a mix of species (for *L. braziliensis* and *L. guyanensis* especially). After multiple alignments of minicircle sequences and the removal of highly variable sequence blocks, 37 nucleotide positions were selected to construct the minimum spanning network (Fig. 4). This allows an alternative visualization of the previously described pattern. *L. amazonensis* and *L. infantum* did not share any minicircle haplotype with other species (except for one in the case of *L. infantum*) and formed separate clusters. On the contrary, in the *Viannia* subgenus, many minicercles haplotypes were shared between species, especially between *L. guyanensis* and *L. braziliensis*. *L. lainsoni* appeared to share most of its minicircle haplotypes with other species of the

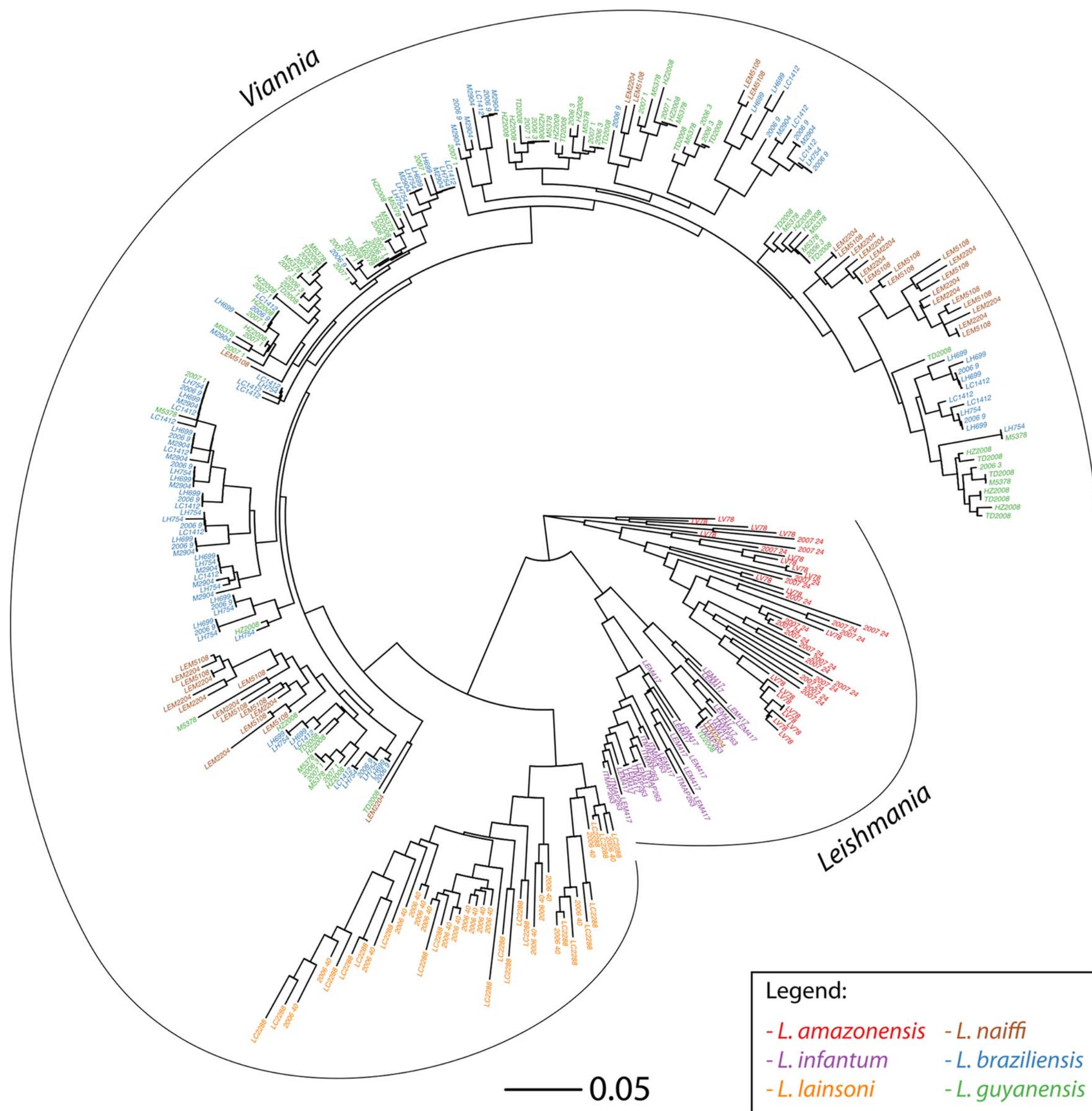

Fig. 3. Neighbour-joining tree based on pairwise alignments of *c*. 80-bp-long kDNA minicircle fragments. Only the 20 most abundant sequences are included for each strain to allow visualization. The colours refer to *Leishmania* species. Each sequence is referenced with its strain ID.

Viannia subgenus, while the majority of *L. naiffi* haplotypes were monospecific.

The leave-one-out procedure showed that, using the method presented here, all strains would have been correctly identified if considered unknown (Table 3). On average, for each strain, 84·8% of minicircle sequences matched on species-specific references, and these were indicative of the correct *Leishmania* species in 97·85% of cases. Most strains (10/18) were identified with an identification score of 100%, including all *L. amazonensis*, *L. infantum*, *L. lainsoni* and *L. naiffi* strains. Some minicircle sequences of *L. braziliensis* were identified as *L. guyanensis* and *vice versa*, which resulted in identification scores below 100%. However, most *L. braziliensis and L. guyanensis* were identified with a score above 95%. The most ambiguous identifications were that of strains M5378 and HZ2008 (scores = 90·4 and 88·4%, respectively), but these were still correctly identified as *L. guyanensis*.

DISCUSSION

The kDNA minicircle conserved region is known to be an ideal target for the detection of all *Leishmania* spp. with high sensitivity (Akhoundi *et al*. 2017). This was further confirmed by our results, since PCR amplification was successful for the six *Leishmania* species belonging to both *Viannia* and *Leishmania* subgenera, down to an extremely low DNA concentration. On the other hand, we would like to stress that other non-*Leishmania* kinetoplastids could probably be detected with the same PCR assay due to the high conservation of CSBs,

https://doi.org/10.1017/S0031182017002013 Published online by Cambridge University Press

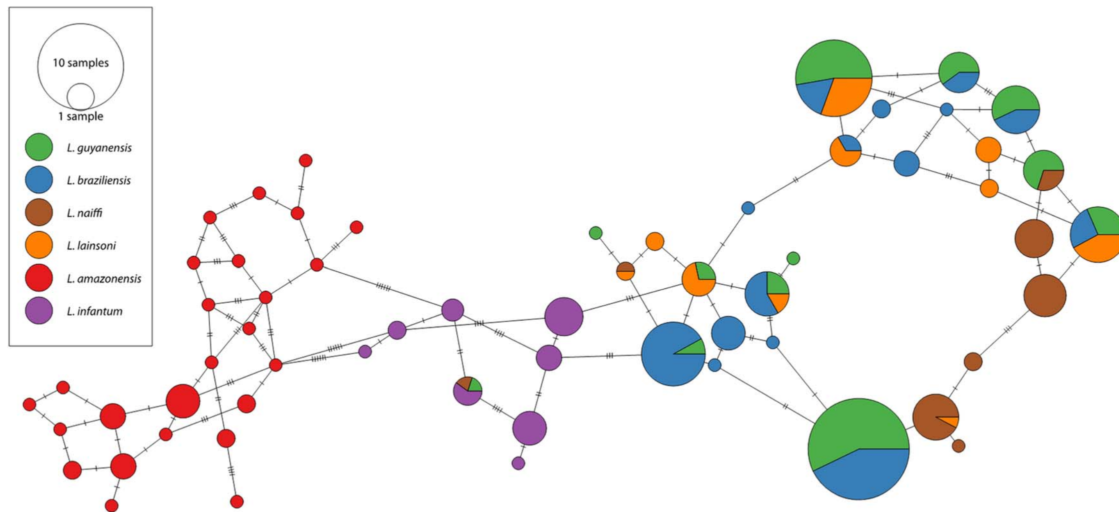

Fig. 4. Minimum spanning network of minicircle sequences after removal of highly variable regions.

which should not be used for the specific diagnosis without confirmation by sequencing.

Despite the usefulness of kDNA minicircles for *Leishmania* detection, classical DNA-based identification methods may hardly be applicable for these peculiar genomic molecules. One of the reasons is that several distinct minicircle classes coexist in each parasitic cell. Even if all minicircles can be amplified simultaneously with the same PCR primers located in the CSBs, the comprehensive characterization of minicircle population used to require isolation and cloning of PCR products prior to sequencing (Lee *et al*. 1993; Brewster and Barker, 2002; Telleria *et al*. 2006; Rodrigues *et al*. 2013). High-throughput sequencing allows bypassing these laborious tasks and will likely contribute to fulfil important methodological gaps in kinetoplastid genomics. By sequencing single PCR products, we were able to easily retrieve significant numbers of distinct minicircle amplicons from each strain, reflecting the diversity of minicircle classes.

Another difficulty comes with the variability in the number and identity of minicircle classes found among strains, even within the same species. Therefore, we used a barcoding-like strategy that does not rely on *a priori* knowledge of minicircle homology. Our results show that, using our method, every strain would have been correctly identified, even when conspecific reference strains were isolated from very distant geographic locations. This highlights the robustness of the approach for the identification of all major species complexes found in the New World. Other genetic markers can provide such level of taxonomic resolution (Akhoundi *et al*. 2017), but thanks to the very high copy number of kDNA minicercles, the present method may be employed in a single assay for sensitive *Leishmania* detection and *Leishmania* spp. identification. The inclusion of additional closely related *Leishmania* species would be necessary to assess whether the method is reliable beyond the species complex level. Further effort is also needed to extend the present work to a wider range of parasites, including Old World *Leishmania* spp. as well as other Trypanosomatidae. In the future, the development of similar strategies targeting the minicircle variable region may allow the typing of strains at a finer resolution. Indeed, kDNA minicircle fingerprinting has long been known as a valuable tool for describing and tracking *Leishmania* genetic diversity, with a discriminatory power superior to that of MLEE (Angelici *et al*. 1989; Botilde *et al*. 2006).

One limitation of the approach is that it may be unable to identify *Leishmania* hybrids (Bañuls *et al*. 1997; Ravel *et al*. 2006), or to detect the presence of several *Leishmania* spp. in the same sample [e.g. in the case of a coinfection, or when analysing pools of sand flies (Kocher *et al*. 2017*b*)]. This is because, even in one single *Leishmania* strain, minicircle sequences may give conflicting taxonomic signals (i.e. some sequences being identified as belonging to a given species and other being identified as belonging to another). We showed that considering only the majority taxonomic signal allowed robust species identifications. However, mixed infections or hybrid strains could hardly be distinguished from the intrastrain mixed signal. The constitution of more comprehensive and local minicircle reference databases will probably narrow the spectrum of species-specific reference minicircle sequences, which could help resolve this issue.

Another drawback of the method is that high-throughput sequencing remains relatively costly unless large batches of samples are analysed, and that several days or weeks may be required for laboratory work to bioinformatic treatment of the data. Such an approach is therefore not adapted for all applications yet. In particular, the present assay

https://doi.org/10.1017/S0031182017002013 Published online by Cambridge University Press

is clearly not adapted for routine diagnostic. However, it can already be useful for one-time studies involving the analysis of large number of samples. Furthermore, given the rapid evolution of high-throughput sequencing technologies and their increasing accessibility, this type of approach could be used more widely in a near future.

Our analyses revealed patterns of sequence clustering that are remarkably congruent with the current hypothesis of *Leishmania* relationships and timings of speciation (Harkins *et al*. 2016). In particular, the observation of the NJ tree and minimum spanning network seems to indicate that the presence of shared or homologue minicircle classes between species tend to decrease when the given species are more distantly related. More extensive studies based on similar high-throughput sequencing approaches may provide precious insights into the evolutionary aspects of kinetoplast minicircle networks. Additionally, the method presented here could be adapted to acquire knowledge in fundamental biological processes of these parasites. Minicircle analyses in *Trypanosoma brucei* and *T. cruzi* allowed to evidence clear signatures of intraspecific recombination (Gibson *et al*. 1997). Population structure in *Leishmania* still needs to be further detailed, especially as to the occurrence of genetic exchanges (Rougeron *et al*. 2017). High-throughput sequencing of minicircle amplicons may represent a powerful tool in this regard.

Concluding remarks

In this study, we showed the efficiency of high-throughput sequencing to recover various minicircle classes simultaneously amplified by PCR. Despite the known difficulties for minicircle sequence analysis, we developed a barcoding-like strategy that proved to be robust for all major *Leishmania* species complexes found in the New World, while taking advantage of minicircle properties for ultra-sensitive PCR detection. Although the method has several limitations as a diagnostic tool, this pioneering work opens up avenues for the study of kDNA minicercles. In the context of important undergoing technological progress, high-throughput sequencing of kDNA minicircles could soon represent a useful tool for a variety of applications in *Leishmania* research.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0031182017002013

ACKNOWLEDGMENTS

We would like to thank Cécile Cassan for her help in the laboratory work. *Leishmania* strains were provided by the Biological Resource Center for *Leishmania* (Montpellier, France), the Parasitology and Mycology Laboratory of Cayenne Hospital (French Guiana), the Unit of Molecular Parasitology (Institute of Tropical Medicine, Antwerp, Belgium) and the Alexander von Humboldt Institute (Universidad Peruana Cayetano Heredia, Lima, Peru).

FINANCIAL SUPPORT

This work was supported by 'Investissement d'Avenir' grants managed by Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01; TULIP: ANR-10-LABX-41, ANR-11-IDEX-0002-02; IAEL: ANR-06-SEST-20) as well as the Inkermann fund hosted at the Fondation de France.

REFERENCES

Akhoundi, M., Downing, T., Votýpka, J., Kuhls, K., Lukeš, J., Cannet, A., Ravel, C., Marty, P., Delaunay, P., Kasbari, M., Granouillac, B., Gradoni, L. and Sereno, D. (2017). *Leishmania* infections: molecular targets and diagnosis. *Molecular Aspects of Medicine* **57**, 1–29.

Akhoundi, M., Kuhls, K., Cannet, A., Votýpka, J., Marty, P., Delaunay, P. and Sereno, D. (2016). A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. *PLoS Neglected Tropical Diseases* **10**, e0004349.

Alvar, J., Vélez, I. D., Bern, C., Herrero, M., Desjeux, P., Cano, J., Jannin, J. and Boer, M. and the WHO Leishmaniasis Control Team (2012). Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE* **7**, e35671.

Angelici, M. C., Gramiccia, M. and Gradoni, L. (1989). Study on genetic polymorphism of *Leishmania infantum* through the analysis of restriction enzyme digestion patterns of kinetoplast DNA. *Parasitology* **99**, 301–309.

Bandelt, H.-J., Forster, P. and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37–48.

Bañuls, A.-L., Guerrini, F., Pont, F. L., Barrera, C., Espinel, I., Guderian, R., Echeverria, R. and Tibayrenc, M. (1997). Evidence for hybridization by multilocus enzyme electrophoresis and random amplified polymorphic DNA between *Leishmania braziliensis* and *Leishmania panamensis/guyanensis* in Ecuador. *Journal of Eukaryotic Microbiology* **44**, 408–411.

Bastrenta, B., Buitrago, R., Vargas, F., Le Pont, F., Torrez, M., Flores, M., Mita, N. and Brenière, S. F. (2002). First evidence of transmission of *Leishmania* (*Viannia*) *lainsoni* in a Sub Andean region of Bolivia. *Acta Tropica* **83**, 249–253.

Berzunza-Cruz, M., Rodríguez-Moreno, Á, Gutiérrez-Granados, G., González-Salazar, C., Stephens, C. R., Hidalgo-Mihart, M., Marina, C. F., Rebollar-Téllez, E. A., Bailón-Martínez, D., Balcells, C. D., Ibarra-Cerdeña, C. N., Sánchez-Cordero, V. and Becker, I. (2015). *Leishmania* (*L.*) *mexicana* infected bats in Mexico: novel potential reservoirs. *PLoS Neglected Tropical Diseases* **9**, e0003438.

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R. and Willerslev, E. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* **2**, e197.

Botilde, Y., Laurent, T., Quispe Tintaya, W., Chicharro, C., Cañavate, C., Cruz, I., Kuhls, K., Schönian, G. and Dujardin, J.-C. (2006). Comparison of molecular markers for strain typing of *Leishmania infantum*. *Infection, Genetics and Evolution* **6**, 440–446.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P. and Coissac, E. (2016). Obitools: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources* **16**, 176–182.

Brewster, S. and Barker, D. C. (2002). Analysis of minicircle classes in *Leishmania* (*Viannia*) species. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **96**, S55–S63.

Brown, S. D., Collins, R. A., Boyer, S., Lefort, M.-C., Malumbres-Olarte, J., Vink, C. J. and Cruickshank, R. H. (2012). Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources* **12**, 562–565.

Cássia-Pires, R., Boité, M. C., D'Andrea, P. S., Herrera, H. M., Cupolillo, E., Jansen, A. M. and Roque, A. L. R. (2014). Distinct

https://doi.org/10.1017/S0031182017002013 Published online by Cambridge University Press

Leishmania species infecting wild caviomorph rodents (Rodentia: Hystricognathi) from Brazil. *PLoS Neglected Tropical Diseases* **8**, e3389.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**, 540–552.

Ceccarelli, M., Galluzzi, L., Migliazzo, A. and Magnani, M. (2014). Detection and characterization of *Leishmania* (*Leishmania*) and *Leishmania* (*Viannia*) by SYBR green-based real-time PCR and high resolution melt analysis targeting kinetoplast minicircle DNA. *PLoS ONE* **9**, e88845.

Copeland, N. K. and Aronson, N. E. (2015). Leishmaniasis: treatment updates and clinical practice guidelines review. *Current Opinion in Infectious Diseases* **28**, 426–437.

de Almeida, M. E., Steurer, F. J., Koru, O., Herwaldt, B. L., Pieniazek, N. J. and da Silva, A. J. (2011). Identification of *Leishmania* spp. by molecular amplification and DNA sequencing analysis of a fragment of rRNA internal transcribed spacer 2. *Journal of Clinical Microbiology* **49**, 3143–3149.

de Oliveira Ramos Pereira, L. and Brandão, A. (2013). An analysis of trypanosomatids kDNA minicircle by absolute dinucleotide frequency. *Parasitology International* **62**, 397–403.

Dufour, P., Pons, J.-M., Collinson, J. M., Gernigon, J., Dies, J. I., Sourrouille, P. and Crochet, P.-A. (2017). Multilocus barcoding confirms the occurrence of elegant terns in Western Europe. *Journal of Ornithology* **158**, 351–361.

Dujardin, J.-C., Bañuls, A.-L., Llanos-Cuentas, A., Alvarez, E., De Doncker, S., Jacquet, D., Le Ray, D., Arevalo, J. and Tibayrenc, M. (1995). Putative *Leishmania* hybrids in the Eastern Andean valley of Huanuco, Peru. *Acta Tropica* **59**, 293–307.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797.

Esling, P., Lejzerowicz, F. and Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research* **43**, 2513–2524.

Faye, B., Bañuls, A. L., Bucheton, B., Dione, M. M., Bassanganam, O., Hide, M., Dereure, J., Choisy, M., Ndiaye, J. L., Konaté, O., Claire, M., Senghor, M. W., Faye, M. N., Sy, I., Niang, A. A., Molez, J. F., Victoir, K., Marty, P., Delaunay, P., Knecht, R., Mellul, S., Diedhiou, S. and Gaye, O. (2010). Canine visceral leishmaniasis caused by *Leishmania infantum* in Senegal: risk of emergence in humans? *Microbes and Infection* **12**, 1219–1225.

Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., Percy, D. M., Hajibabaei, M. and Barrett, S. C. H. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE **3**, e2802.

Gao, G., Kapushoc, S. T., Simpson, A. M., Thiemann, O. H. and Simpson, L. (2001). Guide RNAs of the recently isolated LEM125 strain of *Leishmania tarentolae*: an unexpected complexity. *RNA* **7**, 1335–1347.

Gardener, P. J. and Howells, R. E. (1972). Isoenzyme variation in leishmanial parasites. *Journal of Protozoology* **19**, 47.

Gibson, W., Crow, M. and Kearns, J. (1997). Kinetoplast DNA minicircles are inherited from both parents in genetic crosses of *Trypanosoma brucei*. *Parasitology Research* **83**, 483–488.

Hajduk, S. and Ochsenreiter, T. (2010). RNA editing in kinetoplastids. *RNA Biology* **7**, 229–236.

Harkins, K. M., Schwartz, R. S., Cartwright, R. A. and Stone, A. C. (2016). Phylogenomic reconstruction supports supercontinent origins for *Leishmania*. *Infection, Genetics and Evolution* **38**, 101–109.

Jensen, R. E. and Englund, P. T. (2012). Network news: the replication of kinetoplast DNA. *Annual Review of Microbiology* **66**, 473–491.

Koarashi, Y., Cáceres, A. G., Saca, F. M. Z., Flores, E. E. P., Trujillo, A. C., Alvares, J. L. A., Yoshimatsu, K., Arikawa, J., Katakura, K., Hashiguchi, Y. and Kato, H. (2016). Identification of causative *Leishmania* species in Giemsa-stained smears prepared from patients with cutaneous leishmaniasis in Peru using PCR–RFLP. *Acta Tropica* **158**, 83–87.

Kocher, A., Thoisy, B. D., Catzeflis, F., Huguin, M., Valiere, S., Zinger, L., Bañuls, A.-L. and Murienne, J. (2017*a*). Evaluation of short mitochondrial metabarcodes for the identification of amazonian mammals. *Methods in Ecology and Evolution* **8**, 1276–1283.

Kocher, A., Gantier, J.-C., Gaborit, P., Zinger, L., Holota, H., Valiere, S., Dusfour, I., Girod, R., Bañuls, A.-L. and Murienne, J. (2017*b*) Vector soup: high-throughput identification of Neotropical phlebotomine sand flies using metabarcoding. *Molecular Ecology Resources* **17**, 172–182.

Kreutzer, R. D., Semko, M. E., Hendricks, L. D. and Wright, N. (1983). Identification of *Leishmania* spp. by multiple isozyme analysis. *The American Journal of Tropical Medicine and Hygiene* **32**, 703–715.

Lainson, R. and Shaw, J. J. (2010). New world leishmaniasis. *Topley and Wilson's Microbiology and Microbial Infections*. doi: 10.1002/9780470688618.taw0182.

Lee, S.-T., Tarn, C. and Chang, K.-P. (1993). Characterization of the switch of kinetoplast DNA minicircle dominance during development and reversion of drug resistance in *Leishmania*. *Molecular and Biochemical Parasitology* **58**, 187–203.

Marco, J. D., Uezato, H., Mimori, T., Barroso, P. A., Korenaga, M., Nonaka, S., Basombrío, M. A., Taranto, N. J. and Hashiguchi, Y. (2006). Are cytochrome B gene sequencing and polymorphism-specific polymerase chain reaction as reliable as multilocus enzyme electrophoresis for identifying *Leishmania* spp. from Argentina? *The American Journal of Tropical Medicine and Hygiene* **75**, 256–260.

Marfurt, J., Nasereddin, A., Niederwieser, I., Jaffe, C. L., Beck, H.-P. and Felger, I. (2003). Identification and differentiation of *Leishmania* species in clinical samples by PCR amplification of the miniexon sequence and subsequent restriction fragment length polymorphism analysis. *Journal of Clinical Microbiology* **41**, 3147–3153.

Mariette, J., Escudié, F., Allias, N., Salin, G., Noirot, C., Thomas, S. and Klopp, C. (2012). NG6: integrated next generation sequencing storage and processing environment. *BMC Genomics* **13**, 462.

Maroli, M., Feliciangeli, M. D., Bichaud, L., Charrel, R. N. and Gradoni, L. (2013). Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. *Medical and Veterinary Entomology* **27**, 123–147.

Noyes, H. A., Reyburn, H., Bailey, J. W. and Smith, D. (1998). A nested-PCR-based schizodeme method for identifying *Leishmania* kinetoplast minicircle classes directly from clinical samples and its application to the study of the epidemiology of *Leishmania tropica* in Pakistan. *Journal of Clinical Microbiology* **36**, 2877–2881.

Paradis, E., Claude, J. and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* (*Oxford, England*) **20**, 289–290.

Pereira Júnior, A. M., Teles, C. B. G., de Azevedo dos Santos, A. P., de Souza Rodrigues, M., Marialva, E. F., Pessoa, F. A. C. and Medeiros, J. F. (2015). Ecological aspects and molecular detection of *Leishmania* DNA Ross (Kinetoplastida: Trypanosomatidae) in Phlebotomine sandflies (Diptera: Psychodidae) in terra firme and várzea environments in the Middle Solimões region, Amazonas State, Brazil. *Parasites & Vectors* **8**, 180.

R Core team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ravel, C., Cortes, S., Pratlong, F., Morio, F., Dedet, J.-P. and Campino, L. (2006). First report of genetic hybrids between two very divergent *Leishmania* species: *Leishmania infantum* and *Leishmania major*. *International Journal for Parasitology* **36**, 1383–1388.

Read, L. K., Lukeš, J. and Hashimi, H. (2015). Trypanosome RNA editing: the complexity of getting U in and taking U out. *Wiley Interdisciplinary Reviews: RNA* **7**, 33–51.

Ready, P. D. (2013). Biology of phlebotomine sand flies as vectors of disease agents. *Annual Review of Entomology* **58**, 227–250.

Richini-Pereira, V. B., Marson, P. M., Hayasaka, E. Y., Victoria, C., da Silva, R. C. and Langoni, H. (2014). Molecular detection of *Leishmania* spp. in road-killed wild mammals in the Central Western area of the state of São Paulo, Brazil. *Journal of Venomous Animals and Toxins Including Tropical Diseases* **20**, 1–7.

Rioux, J. A., Lanotte, G., Serres, E., Pratlong, F., Bastien, P., Perieres, J. (1990). Taxonomy of *Leishmania*. Use of isoenzymes. Suggestions for a new classification. *Annales de Parasitologie Humaine et Comparee* **65**, 111–125.

Rodrigues, E. H. G., da Silva Soares, F. C., Werkhäuser, R. P., de Brito, M. E. F., Fernandes, O., Abath, F. G. C. and Brandão, A. (2013). The compositional landscape of minicircle sequences isolated from active lesions and scars of American cutaneous leishmaniasis. *Parasites & Vectors* **6**, 1.

Roque, A. L. R., Cupolillo, E., Marchevsky, R. S. and Jansen, A. M. (2010). *Thrichomys laurentius* (Rodentia; Echimyidae) as a putative reservoir of *Leishmania infantum* and *L. braziliensis*: patterns of experimental infection. *PLoS Neglected Tropical Diseases* **4**, e589.

Rotureau, B. (2006). Ecology of the *Leishmania* species in the Guianan ecoregion complex. *The American Journal of Tropical Medicine and Hygiene* **74**, 81–96.

Rougeron, V., De Meeûs, T. and Bañuls, A.-L. (2017). Reproduction in *Leishmania*: a focus on genetic exchange. *Infection, Genetics and Evolution* **50**, 128–132.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

Schnell, I. B., Bohmann, K. and Gilbert, M. T. P. (2015). Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources* **15**, 1289–1303.

Simon, S., Veron, V. and Carme, B. (2010). *Leishmania* spp. Identification by polymerase chain reaction–restriction fragment length

https://doi.org/10.1017/S0031182017002013 Published online by Cambridge University Press

polymorphism analysis and its applications in French Guiana. *Diagnostic Microbiology and Infectious Disease* **66**, 175–180.
Simpson, L. (1987). The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution. *Annual Review of Microbiology* **41**, 363–380.
Simpson, L. (1997). The genomic organization of guide RNA genes in kinetoplastid protozoa: several conundrums and their solutions. *Molecular and Biochemical Parasitology* **86**, 133–141.
Telleria, J., Lafay, B., Virreira, M., Barnabé, C., Tibayrenc, M. and Svoboda, M. (2006). *Trypanosoma cruzi*: sequence analysis of the variable region of kinetoplast minicircles. *Experimental Parasitology* **114**, 279–288.
Thomaz-Soccol, V., Lanotte, G., Rioux, J. A., Pratlong, F., Martini-Dumas, A. and Serres, E. (1993). Monophyletic origin of the genus *Leishmania* Ross, 1903. *Annales de Parasitologie Humaine et Comparée* **68**, 107–107.
Victoir, K., Bañuls, A. L., Arevalo, J., Llanos-Cuentas, A., Hamers, R., Noël, S., Doncker, S. D., Ray, D. L., Tibayrenc, M. and Dujardin, J. C. (1998). The gp63 gene locus, a target for genetic characterization of *Leishmania* belonging to subgenus Viannia. *Parasitology* **117**, 1–13.
Weirather, J. L., Jeronimo, S. M., Gautam, S., Sundar, S., Kang, M., Kurtz, M. A., Haque, R., Schriefer, A., Talhari, S., Carvalho, E. M., Donelson, J. E. and Wilson, M. E. (2011). Serial quantitative PCR assay for detection, species discrimination, and quantification of *Leishmania* spp. in human samples. *Journal of Clinical Microbiology* **49**, 3892–3904.
Zelazny, A. M., Fedorko, D. P., Li, L., Neva, F. A. and Fischer, S. H. (2005). Evaluation of 7sl RNA gene sequences for the identification of *Leishmania* spp. *The American Journal of Tropical Medicine and Hygiene* **72**, 415–420.

https://doi.org/10.1017/S0031182017002013 Published online by Cambridge University Press