We describe *Leishmania* species determination on clinical samples on the basis of partial sequencing of the heat-shock protein 70 gene (*hsp70*), without the need for parasite isolation. The method is especially suited for use in non-endemic infectious disease clinics dealing with relatively few cases on an annual basis, for which no fast high throughput diagnostic tests are needed. We show that the results obtained from this gene are in nearly perfect agreement with those from multilocus enzyme electrophoresis, which is still considered by many clinicians and the World Health Organization (WHO) as the gold standard in *Leishmania* species typing. Currently, 203 sequences are available that cover the entire *hsp70* gene region analysed here, originating from a total of 41 leishmaniasis endemic countries, and representing 15 species and sub-species causing human disease. We also provide a detailed laboratory protocol that includes a step-by-step procedure of the typing methodology, to facilitate implementation in diagnostic laboratories.

**Introduction**

As a result of current human mobility, European infectious disease clinics are occasionally confronted with leishmaniasis patients who got infected in an area endemic for *Leishmania* outside their own country. Typically it concerns tourists, expatriates, military staff, migrants, and relatives visiting friends or family. In many of these centres, the number of such cases seen annually is limited, and investing in the validation of high-throughput methods for discriminating the medically relevant species is therefore too costly. Nevertheless, especially in the case of tegumentary leishmaniasis, knowledge of the aetiological agent is highly relevant, as the disease prognosis and treatment choice depend on it [1-8]. However, one cannot always rely on the known epidemiology in the suspected region of infection. Firstly, because such information is often inaccurate or outdated, and secondly, because the geographic area where the patient got infected may not be known exactly if they resided in different endemic areas or countries. Moreover, even if the exact location of infection and epidemiology are known, different species may circulate sympatrically in a given region. Hence, there is a need for easily applicable, straightforward and standardised species discrimination methods that must above all be accurate, rather than allowing to handle many samples in a high-throughput fashion.

Over the past few years, we have been investing in the use of the heat-shock protein 70 gene (*hsp70*) for discrimination of medically important *Leishmania* species worldwide [9-12]. Initially developed for species discrimination in the New World subgenus *L. (Viannia)* by restriction fragment length polymorphism (RFLP) analysis [13], we have upgraded the specificity and sensitivity of the *hsp70* PCR amplification strategy to suit all *Leishmania* species [12]. In this paper we report on the power of *Leishmania* species typing on the basis of *hsp70* sequences rather than RFLP. The approach is directed specifically towards diagnosis in clinical laboratories dealing with relatively few cases on an annual basis, such as our Institute of Tropical Medicine in Antwerp where an average of 15 patients are diagnosed each year. The method described here was developed in the framework of a European consortium of tropical infectious disease clinics called ‘LeishMan’ (www.leishman.eu), embedded in the European network for tropical medicine and travel health TropNet (www.tropnet.net). LeishMan aims at characterising *Leishmania* parasites using a standardised molecular assay that can be applied for clinical samples, without the need for parasite culture.

Several other single-locus assays have been used for sequence-based species discrimination, such as the mini-exon, the 7SL-RNA, and the ribosomal DNA-ITS1...
[14-17]. We found that hsp70 has some advantages over these (data not shown): it is easily comparable across all Leishmania species worldwide as there is no size variation in the gene [9], it discriminates all relevant species in both subgenera L. (Leishmania) and L. (Viannia), and PCRs have been optimised for direct amplification from clinical samples [10,12]. The gene is arranged as a tandem repeat unit, with almost no sequence variation between the coding sequences of the different copies [18,19]. In this paper we assess the concordance of Leishmania species typing with hsp70 sequences on the one hand, and results obtained from other genetic targets and multilocus enzyme electrophoresis (MLEE) on the other hand.

Methods

Hsp70 amplification and sequencing

Leishmania hsp70 sequences from 64 cultures and 36 rDNA-PCR-confirmed [20,21] clinical samples were determined on the basis of a single PCR amplicon, i.e. PCR-F in Figure 1. Ca. 50 of these cultures were obtained from the Centre National de Référence des Leishmania (Montpellier, France). Among the clinical samples, 27 were from cutaneous lesions (mostly biopsies), one from a mucocutaneous lesion, two from visceral leishmaniasis patients, and six from an unknown clinical background. In the rare occasions where direct amplification of PCR-F failed from the clinical sample DNA extract, or when an insufficient amount of amplicon was obtained for sequencing, two shorter PCRs were used that together cover the same fragment: PCR-N and PCR-T (Figure 1). These can be run directly on the sample DNA, or alternatively as hemi-nested PCRs using the PCR-F amplicon as first round PCR. A detailed protocol is available from www.itg.be/LeishmaniaHSP70.

All PCRs were performed in 25 µl 1x standard PCR buffer (Qiagen, Hilden, Germany), supplemented by 1 mM MgCl2 and 1x Qiagen Q-solution. Each reaction used 200 µM of each dNTP, 0.8 µM of each PCR primer (Table), and 1U of HotStarTaq Plus DNA polymerase (Qiagen). Up to 2.5 µl of template were used. Cycling conditions were as follows: 5 min at 95 °C denaturation; 35 cycles of 40 sec at 94 °C, 1 min at 61 °C, 2 min at 72 °C; and finally 10 min at 72 °C. For PCR-N and PCR-T, the elongation step was shortened to 1 min at 72 °C.

PCR products were analysed on a 2% agarose gel to check for sufficient and specific amplification, based on the expected product sizes outlined in Figure 1. The fragments were sequenced with primers internal in the PCR fragment (see protocol on www.itg.be/LeishmaniaHSP70). In some strains, a second nucleotide was detected below the main trace signal at some sequence positions. In such cases, IUPAC ambiguity codes [22] were introduced in the sequence whenever the secondary nucleotide showed at least 20% of the intensity of the main peak in each sequence read covering the respective position. The sequences from reference strains were submitted to the European Nucleotide Archive (www.ebi.ac.uk/ena).

Sequence analysis and typing

For the analysis presented in this paper, we compiled
107 available sequences covering the 1,245 bp hsp70 PCR-F fragment (Figure 1) from GenBank (www.ncbi.nlm.nih.gov/genbank, accessed on 25 June 2013). These were aligned with the 100 sequences determined in this study. From the total of 207 sequences, 84 were typed by MLEE, and 54 on the basis of genes other than hsp70. Many MLEE-typed isolates were analysed with genetic methods as well. For 12 sequences, we relied on the species identification as listed in GenBank, where the typing method is not specified. Finally, no typing data were available for the remaining 57 sequences, which included those determined for diagnosis. For the sequences described in the paper by Zhang et al. [23], we did not rely on the GenBank identification, as this was in conflict with data in the paper itself. Aligning was done manually, which was straightforward as no size variation was detected in 205 sequences, while two sequences showed a deletion of three nucleotides, corresponding to one amino acid.

Species delineation was based upon the clustering of aligned sequences in a comparative dendrogram, which was constructed with the freely available software package MEGA5 [24]. Dendrograms were built from the variable sites in the alignment using the neighbour-joining method, with pairwise gap deletion and 2,000 bootstrap replicates. As our aim was to find the most discriminative analysis method rather than to study evolution, we based our dendrograms on p distances, and not on other models such as the popular Kimura 2-parameter method for calculating corrected distances. More details are available from the protocol on www.itg.be/LeishmaniaHSP70.

Results

The final alignment contained 207 sequences from 42 Leishmania-endemic countries, representing 18 species of which 15 are causing human disease (Figure 2 and supplementary dendrogram available at www.itg.be/LeishmaniaHSP70). As further detailed in the Discussion, four GenBank entries contained sequences that did not correspond to the indicated isolate, reducing the number of trustworthy sequences to 203 from 41 countries.

The clustering of a representative selection of the hsp70 sequences is depicted in Figure 3. A dendrogram of the complete set of available hsp70 sequences can be found at www.itg.be/LeishmaniaHSP70. The medically relevant clusters indicated in bold in these figures could be easily discriminated, and were supported by bootstrap values between 89 and 99% (Figure 3). These generally coincided with recognised species complexes. Within these complexes, a further distinction was possible, as indicated by the dotted lines. These subdivisions had a lower bootstrap support, between 53 and 71% (Figure 3).

Almost all hsp70 clusters showed a perfect agreement with MLEE-based classifications (isolates identified with ‘M. species’) and typing results from genetic loci different from hsp70 (identified with ‘G. species’). There were nevertheless a few exceptions, which are indicated with * and ** following the taxon designation. Of the 81 isolates typed on the basis of MLEE and from which a trustworthy hsp70 sequence was reported, 76 (94%) grouped in the respective hsp70 cluster. Of the 54 isolates typed on the basis of non-hsp70 genetic loci, 50 (93%) grouped in the respective hsp70 cluster. A few isolates did not group with any known species clade, notably IMON/CN/90/KXG-Y, MHOM/--/94/CRE58, MHOM/PE/--/CU00181, MHOM/PE/95/LQ-8, and MCAN/IR/96/LON-49.

The L. braziliensis isolates separated into two clearly distinct clusters, named L. braziliensis outlier and L. braziliensis complex, which also contained L. peruviana. Even though these two clusters are sister taxa, the bootstrap support was weak (53%). In some dendrograms, the two clusters did not form sister clades, and the outliers rather grouped with L. naiffi (results not shown). One strain, MHOM/PE/--/CU00181, was intermediate between both L. braziliensis clusters.

Table

<table>
<thead>
<tr>
<th>Primer name</th>
<th>PCR</th>
<th>Sequence (5’-3’)</th>
<th>Length</th>
<th>Orientation</th>
<th>Annealing start (5’ of primer)a</th>
<th>Annealing end (3’ of primer)a</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSP70-F25</td>
<td>PCR-F/N</td>
<td>GGACGCCGCGACGATTKCT</td>
<td>19</td>
<td>Sense</td>
<td>480</td>
<td>498</td>
</tr>
<tr>
<td>HSP70-6F</td>
<td>PCR-T</td>
<td>GTGCCAGACGGTGCTGGTGTG</td>
<td>21</td>
<td>Sense</td>
<td>1,000</td>
<td>1,020</td>
</tr>
<tr>
<td>HSP70-R617</td>
<td>PCR-N</td>
<td>CGAAGAGTCTGAGAGAGAGA</td>
<td>22</td>
<td>Antisense</td>
<td>1,072</td>
<td>1,051</td>
</tr>
<tr>
<td>HSP70-R1310</td>
<td>PCR-F/T</td>
<td>CTTGAGTCTGTCAGCAGCCA</td>
<td>22</td>
<td>Antisense</td>
<td>1,765</td>
<td>1,744</td>
</tr>
</tbody>
</table>

*a Annealing position in GenBank entry FR796424 (hsp70 of L. major Friedlin strain).

Primers are listed in order of annealing in the coding sequence of the gene, from 5’-3’ terminus.
Using the here presented hsp70 clustering system, we have so far been able to determine the infecting species in 33 clinical samples presented for diagnosis in our institute, and three from military personnel on mission in Afghanistan. These were from 14 different countries and represented eight Leishmania species (Figure 3 and supplementary dendrogram). The majority (n=27) were from cutaneous lesions (mostly biopsies), one from a mucocutaneous lesion, and two from visceral leishmaniasis patients. From six samples the clinical presentation was not known.

Discussion

In general there is good agreement between typing results on the basis of hsp70 and those based on other genes and MLEE, with the following exceptions: (i) L. chagasi isolates could not be distinguished from L. infantum, which agrees with previous studies showing that both are in fact one species, whereby L. chagasi is synonym of South-American L. infantum [25]. (ii) L. archibaldi grouped with L. donovani, in line with the current notion that it is not a separate species [26]. (iii) Two L. infantum isolates, MHOM/CN/94/KXG-LIU and MHOM/CN/93/KXG-XU, were found clustering with L. donovani. According to the authors who published these sequences, however, the species identification is disputable and depends on the technique applied [23]. (iv) The MLEE identified L. donovani isolate MHOM/SU/84/MARZ-KRIM clustered with L. infantum. On the basis of at least eight other genes, this strain was indeed identified as L. infantum (data not shown). (v) Two L. guyanensis strains, MHOM/CO/83/REST147 and MHOM/EC/90/UL.031 were found in the L. panamensis cluster. Although it has been argued that L. panamensis is merely a geographically confined subcluster of L. guyanensis rather than a distinct species [9,27-29], both strains merit a more profound genetic analysis to evaluate their hsp70 classification. (vi) The L. braziliensis isolate MHOM/PE/2001/LH2140 clustered with L. peruviana, even though its sequence is different from those of the other L. peruviana strains. A genome-wide amplified fragment length polymorphism
Figure 3
Dendrogram of selected *Leishmania hsp70* sequences analysed in this study, including for each indicated cluster the most divergent sequences (n=91)
Each taxon is identified as follows:

(i) Identification method if available: G: genetic analysis other than hsp70; M: multilocus enzyme electrophoresis; U: unknown typing method.


(iii) EBI/GenBank accession number if available.

(iv) World Health Organization (WHO) code: Missing data are indicated by .

(v) Clinical samples diagnosed in this study are indicated with + in front of the taxon name, and those from the Institute of Tropical Medicine Antwerp are identified by ITM without WHO code, whereby identical sequences are presented as one taxon. CL following the WHO code indicates that the sample was taken from a cutaneous lesion.

(vi) Strains indicated with * cluster differently compared with other methods, those indicated with ** do not represent the strain as reported in GenBank.

(vii) The dissimilarity scale is presented at the bottom. Bootstrap values higher than 50% from a 2,000 replicate analysis are shown in percentages at the internodes. Clusters strongly supported are indicated in bold, those less supported are indicated by dotted lines. For each recognised cluster, the number of strains in the total of 207 available sequences is given between brackets.

The more disputable species designations are L. infantum, L. panamensis, and L. peruviana, as all these were moderately bootstrap-supported subgroups of the highly robust L. donovani, L. guyanensis, and L. braziliensis complexes, respectively, as previously documented [9,19,26-29,31-35]. In case of doubt, the complex level should be reported rather than the exact species. From a clinical point of view, discriminating L. infantum from L. donovani is not highly relevant, since both species can cause visceral leishmaniasis and treatment is the same [26,36]. Also the discrimination

Figure 3 notes

Dendrogram of selected Leishmania hsp70 sequences analysed in this study, including for each indicated cluster the most divergent sequences (n=91)

Taken all evidence together, of the 135 trustworthy sequences for which either MLEE or independent genetic species identification was done, 130 (96.3%) grouped with the correct species in the hsp70 sequence dendrogram; two (1.5%) did not group with any species; and three (2.2%) were assigned to the correct species complex, but the wrong species. Since we started routine species typing on the basis of hsp70 sequences, we could type 33 clinical samples that were sent to our clinic for diagnosis, along with three samples sent to us by other institutes (accessions HF586387, HF586390, HF586391). In the same period, amplification failed from two samples with an extremely low parasite load. We provide a detailed protocol and sequence reference set on the website www.itg.be/LeishmaniaHSP70, which outlines a step-by-step guideline of the PCRs, sequencing, and interpretation. We acknowledge that implementing sequence analysis in a routine diagnostic laboratory may be difficult in some settings and that the entire analysis may take a few days. Nevertheless, in our hands the method proved highly convenient, and in view of the few samples diagnosed per year, more cost-effective than validating a high-throughput system with a simple readout. Alternatively, sequencing could provide a clear identification in case other assays fail.
between *L. guyanensis* and *L. panamensis* is not a priority in clinical practice [36]. Separating *L. braziliensis* from *L. peruviana* is considered more relevant, because *L. braziliensis* potentially causes mucocutaneous complications, while *L. peruviana* generally does not [34]. As no markers are currently available that discriminate strains that do from those that do not cause mucocutaneous leishmaniasis, identification at the species level is the only option. Both MLEE and genetic analyses have revealed that *L. peruviana* is a subcluster in the *L. braziliensis* complex, but discrimination is impaired by the fact that many parasites of this complex seem to have a composite genotype carrying signatures of both species [19,32-35]. The situation is further complicated by the fact that occasionally, *L. peruviana* can cause mucocutaneous disease [34]. Isolates belonging to the *L. braziliensis* outlier group have been isolated from mucous lesions as well (data not shown), but whether these were primary or secondary infections is not known. Two other tightly linked species in the hsp70 dendrogram are *L. tropica* and *L. aethiopica*. Although the *L. tropica* isolates cover the entire endemic region, from Morocco to eastern Africa, the Middle-East, and India, they form a clearly separated recognisable group. The same applies to separating *L. mexicana* from *L. amazonensis*, even though the latter could not be distinguished from *L. garnhami*.

One may wonder why some of the above species in the recognised larger complexes seem less clearly defined by sequencing than by single-nucleotide polymorphism (SNP) assays such as species-specific PCRs or RFLP analysis. The reason is that these assays use a point mutation in the genome of the parasite, which is either present or absent, thereby allowing a binary discrimination. When using sequences, much more information is provided from many polymorphisms and is sometimes contradictory. Typing based on sequencing can therefore be more difficult, but it is more reliable as it uses more data. An accidental mutation may lead to erroneous conclusions in a SNP-based assay, while this is less likely when analysing entire sequences.

The current complete set of trustworthy sequences that can be used for typing amounts to 203, representing 15 species of human medical importance, and originating from 41 endemic countries. This reference set is updated continuously for further improvement of the geographic and genetic coverage, to ensure an adequate representation of the existing inter- and intra-species variability. Some species are over-represented from some regions (such as in Peru), but that does not interfere with the typing outcome. It is of crucial importance to base species typing upon sequences that have been quality-checked. In practice, BLAST searches are often used for identification purposes, on the basis of hsp70 sequences found in public databases such as GenBank. We have found several instances where the species designation reported in these databases was incorrect. For example, in entries JX021425 up to JX021443 and JX970993 up to JX970996, several erroneous *L. donovani* sequences are reported, even disagreeing with the species assignment as listed in the related publication [23]. Two entries were here shown related to *Leptomonas* rather than *Leishmania* (Figure 3 and supplementary dendrogram). Two entries were determined from the *L. infantum* type strain MHOM/FR/78/LEM75: Y08020 and X85798. Both sequences clearly grouped with *L. tropica*, unlike all genuine *L. infantum* sequences in our analysis. This illustrates that one should be extremely careful when using sequences that have not been quality-controlled for species typing by comparison with other sequences from the same species, as this could result in incorrect typing outcomes, with potential adverse consequences for the patient.

As with all other assays based on the analysis of a single genomic locus, it is assumed that the relationship between the sequences mirror the relationship between the parasites. A first requirement to meet this objective is to avoid the use of paralogous sequences. This poses no problem in the case of hsp70 because, although this gene is part of a gene family [18], the primers used in our protocol specifically amplify only one of the family members. Another problem is presented by the occasional inter-species hybrids that have been reported [19,34,37,40], and that are not necessarily evidenced in all genes. Nevertheless, such hybrids do not necessarily go undetected when looking at single genomic loci. For instance, isolate MHOM/PE/2006/ CU00181, by AFLP clearly identified as a hybrid between *L. braziliensis* and *L. braziliensis* outliers [19], also holds an intermediate position in hsp70 sequences (Figure 3 and supplementary dendrogram). On the contrary, MHOM/PE/2003/LH2538, also shown to be a hybrid between these two clusters, grouped with *L. braziliensis*. As this isolate derived a much smaller proportion of its genome from the *L. braziliensis* outliers, such classification is however not problematic. In natural *L. donovani–L. aethiopica* hybrids, both species alleles were present in all genes investigated, including hsp70 [38]. In an *L. infantum–L. major* hybrid, both genomes were present [39], hence enabling to type the parasite based on a single gene assay. Theoretically, the chance of detecting inter-species recombinants increases as more loci are analysed, such as in multilocus-microsatellite, -sequence and -enzyme electrophoresis assays, but this also raises the cost and time of species typing. Given all currently available evidence on the potential of hsp70 to detect reported inter-species hybrids, and as such hybrids are rare, we consider this a negligible setback of using the hsp70 single-locus assay for routine species typing. Nevertheless, additional more variable genes may be able to perform better in discriminating within the complexes, but this would probably require a separate approach for each complex or subgenus [32,41,42].

Ultimately, the use of single-locus sequencing for species discrimination could be substituted by whole-genome sequencing [43]. With this method becoming
cheaper, it may soon be the standard in clinical studies. Comparison of whole-genome information could reveal clinically relevant intra-species differences, and has the highest chance of detecting recombination events. Such typing methodology could even make abstraction of the classical concept of typing at the taxonomic levels of species and species complexes [44]. It could open up a whole new era of relating strains on the basis of a selection of genes relevant for disease progression and treatment options, rather than based upon species definitions that may at times not correlate with clinical outcome. Nevertheless, whole-genome sequencing seems at present miles away from being implemented in everyday clinical practice, not only because of the complexity of data analysis, but also because it is complicated by the presence of human DNA contamination in clinical samples and related ethical issues. In the meantime there is a need for standardised methods to identify *Leishmania* strains. We advocate that *hsp70* has this potential: the gene is easily amplified, it can be analysed by sequencing in high-resource settings, and by simpler methods such as RFLP in limited-resource endemic areas [12,45]. If a global database of *hsp70* sequences from endemic regions were to be established, new sequences found in imported leishmaniasis could immediately be related to documented parasites, with clinical information on the patients from whom they were isolated. Such analysis could even be independent of currently used species boundaries, and provide adequate links based on genetic similarity irrespective of the species.

**Conclusion**

We present in this paper a complete validation and globally applicable standardised protocol for the use of *hsp70* sequences in *Leishmania* typing. As this validation includes a detailed comparison with other species identification methods currently used in various laboratories, we feel that implementation of the here presented typing strategy in a clinical diagnostic laboratory should be straight-forward, and could entail the validation of only the sequencing process itself rather than the actual species assignment, in view of this report. We intend to further promote our strategy, to identify additional strains with linked clinical information, and to establish a global database of circulating *Leishmania* parasites.

**Acknowledgements**

The authors would like to thank Ingrid Felger (Swiss Tropical and Public Health Institute, Basel, Switzerland), Aldert Bart (Academic Medical Centre, Amsterdam, The Netherlands), Mark Bailey (Army Medical Directorate, Camberley, UK), and Ricardo Leonart (INDICASAT, Panama City, Panama) for providing DNA from strains and clinical samples. We acknowledge the constructive input from all colleagues of the LeishMan consortium. We appreciate the help of Eva De Clercq (Université Catholique de Louvain, Louvain-la-Neuve, Belgium) for the construction of Figure 2. Salary support of GVDA was provided by the Belgian Directorate General for Development cooperation (third framework agreement with ITM).

**References**

16. Stevenson LG, Fedorko DP, Zelazny AM. An enhanced method for the identification of *Leishmania* spp. using real-time polymerase chain reaction and sequence analysis of the 7SL


