The *Leishmania donovani* species complex: a new insight into taxonomy★

Anna Fernández-Arévalo[a,b,c], Fouad El Baidouri[d], Christophe Ravel[e], Cristina Ballart[a,f], Alba Abras[g], Laurence Lachaud[e], Silvia Tebar[a], Patrick Lami[e], Francine Pratlong[e], Montserrat Gállego[a,f,*], Carme Muñoz[b,c,*]

[a] *Secció de Parasitologia, Departament de Biologia, Sanitat i Medi Ambient, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona, Spain*

[b] *Institut de Recerca Biomèdica Sant Pau, Barcelona, Spain*

[c] *Servei de Microbiologia, Hospital de la Santa Creu i Sant Pau Barcelona, Spain & Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, Spain*

[d] *Department of Botany, University of Hawaii at Manoa, Honolulu, HI 96822, USA*

[e] *National Reference Centre for Leishmaniasis, University Hospital Centre of Montpellier, MiVEGEC, University of Montpellier. Montpellier, France*

[f] *ISGlobal, Hospital Clínic - Universitat de Barcelona, Barcelona, Spain*

[g] *Laboratori d'Ictiologia Genètica, Departament de Biologia, Universitat de Girona, Girona, Spain*

*Corresponding authors.

C. Muñoz: Servei de Microbiologia, Hospital de la Santa Creu i Sant Pau. C/Sant Quintí, 89, 08041 Barcelona, Spain. Tel: +34 935537298, FAX: +34 935537287. *E-mail address*: cmunoz@santpau.cat

M. Gállego: mgallego@ub.edu, Secció de Parasitologia, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona. Av. Joan XXIII, 27-31, 08028 Barcelona, Spain. Tel: +34 934024502, FAX: +34 934024504. *E-mail address*: mgallego@ub.edu

Note: Supplementary data associated with this article.

★Note: Nucleotide sequence data reported in this paper are available in GenBank under accession numbers **MN728685 - MN728796**

Abstract

Among the 20 or so *Leishmania* spp. described as pathogenic for humans, those of the *Leishmania donovani* complex are the exclusive causative agents of systemic and fatal visceral leishmaniasis. Although well studied, the complex is taxonomically controversial, which hampers clinical and epidemiological research. In this work, we analysed 56 *Leishmania* strains previously identified as *L. donovani*, *Leishmania archibaldi* or *Leishmania infantum*, isolated from humans, dogs and sandfly vectors throughout their distribution area. The strains were submitted to biochemical and genetic analyses and the resulting data were compared for congruence. Our results show: i) a partial concordance between biochemical and genetic-based data, ii) very limited genetic variability within the *L. donovani* complex, iii) footprints of frequent genetic exchange along an east-west gradient, marked by a widespread diffusion of alleles across the geographical range, and iv) a large-scale geographical spreading of a few genotypes. From a taxonomic point of view, considering the absence of relevant terminology in existing classes, the *L. donovani* complex could be treated as a single entity.

Keywords: *Leishmania donovani* complex, Taxonomy, MLEE, Sequencing, MALDI-TOF

1. Introduction

Leishmaniasis, a disease caused by protozoa of the *Leishmania* genus, is endemic in 97 countries around the world, with an estimated 700,000 - 1,000,000 new cases and 26,000 to 65,000 deaths per year (World Health Organization, https://www.who.int/leishmaniasis/en/, https://www.who.int/news-room/fact-sheets/detail/leishmaniasis). The two main clinical manifestations described are tegumentary (cutaneous and mucocutaneous) and a visceral form (VL) that is systemic and potentially lethal if untreated (Burza et al., 2018). The clinical presentation and therefore the prognosis, management and treatment of the disease are largely species-dependent (WHO, 2010). Among the 20 *Leishmania* spp. described as pathogenic for humans, those of the *Leishmania donovani* complex are the exclusive causative agents of systemic and fatal VL (WHO, 2010), and they have consequently been the focus of considerable research. However, from the taxonomic standpoint, the complex remains controversial and there are still doubts about its substructure and how clinical and epidemiological data should be linked to these taxa (Maurício, 2018).

From a historical perspective, the etiological agent of kala-azar disease (VL) in India was definitively named, in 1903, as *Leishmania donovani* Laveran & Mensil, 1903 (Ross, 1903). Five years later, based on epidemiological and clinical criteria, Nicolle described a new species causing the disease in the Mediterranean region, mainly in children, and named it *Leishmania infantum* Nicolle, 1908 (Nicolle, 1908). In 1919, Castellani and Chalmers described a variant of *L. donovani* in Sudan, *Leishmania* var. *archibaldi* Castellani & Chalmers, 1919 (Castellani and Chalmers, 1919), and later Cunha and Chagas proposed a new species, *Leishmania chagasi* Cunha & Chagas, 1937, responsible for VL in South America (Cunha and Chagas, 1937). In the following years, the taxonomic status of the different species was questioned many times (Brumpt, 1936; Da Cunha, 1938; Nicoli, 1963; Lainson and Shaw, 1987) and in the 1970s, *L. donovani*, *L. infantum*, *L. archibaldi* and *L. chagasi* were merged into the *Leishmania donovani* species complex (Lainson and Shaw, 1987).

The development of biochemical taxonomic tools in the 1980s (i.e. multilocus enzyme electrophoresis, MLEE) revived discussions on *Leishmania* taxonomy. A distinction between *L. donovani, L. infantum* and *L. archibaldi* was proposed based on the differential electrophoretic mobility of a single isoenzyme, glutamate-oxaloacetate transaminase (GOT) (mobility values of 113, 100 and 110, respectively) (Rioux et al., 1990). However, the relationship between the previously defined taxa and the biochemical characters was not straightforward and some discrepancies were identified. For instance, no enzymatic differences were observed between *L. chagasi* and *L. infantum*, the first considered a synonymous species of *L. infantum* (Rioux et al., 1990). Furthermore, all GOT^{110} strains were considered as *L. archibaldi*, a term originally associated with geographical distribution, regardless of further evidence. Although the taxonomic position of this group was not clear, it was included as a sub-group of *L. donovani* (Rioux et al., 1990; Pratlong et al., 2001; Jamjoom et al., 2004). Later, it was observed that the intermediate mobility of the *L. archibaldi* strains was produced by a heterozygous pattern (Piarroux et al., 1994).

Examples of both synonymous mutations on isoenzyme-coding genes showing identical mobilities and post-translational modifications altering the migration of the enzymes were spotted, jeopardising the phylogenetic value of these markers in the *Leishmania* genus (Maurício et al., 2006). Lastly, genetics-based studies have shown that the genetic clusters within the *L. donovani* complex do not fully correlate with the taxa defined by isoenzyme criteria (Jamjoom et al., 2004; Kuhls et al., 2005; Lukes et al., 2007; El Baidouri et al., 2013). Nowadays the synonymy of *L. chagasi* and *L. archibaldi* with *L. infantum* and *L. donovani*, respectively, is practically accepted by the scientific community, but the correct taxonomic relationship between the two main species, *L. donovani* and *L. infantum*, is subject to debate.

With the aim of clarifying these taxonomic issues and establishing a more functional taxonomy for this important and controversial group of species, we analysed different *Leishmania* strains reflecting the geographical and clinical diversity of the *L. donovani* complex. Firstly, isoenzyme data underlying the current taxonomy (i.e. GOT) were freshly evaluated by MLEE and

104 systematic sequencing of the gene to check for post-translational modifications. Secondly, a global
105 genetic analysis was performed to evaluate the structure of the complex and allele distribution
106 associating the data from the *got* gene with the previous genetic analysis performed by El Baidouri
107 et al. (2013), where no structure was observed for the *L. donovani* complex using multilocus
108 sequence analysis (MLSA), and the *hsp70* gene, the most widely validated species typing marker.
109 Finally, the cultured strains were submitted to mass spectrometry (MALDI-TOF) analysis as an
110 additional approach to achieve a broader perspective on the complex.

111

112 **2. Materials and methods**

113 *2.1.Strains*

114 Sixty-four strains belonging to the *L. donovani* complex were initially selected from those
115 genetically analysed by El Baidouri et al. (2013). Six were no longer available and two were
116 rejected due to misidentification by previous isoenzymatic analysis. The 56 *Leishmania* strains
117 finally included in this study are summarized in Supplementary Table S1. The strains had been
118 previously classified by MLEE as *L. infantum* (*n*=31), *L. donovani* (*n*=20) or *L. archibaldi* (*n*=5).
119 To capture the broadest diversity, the selected strains were isolated from different sand fly species,
120 dogs and humans in 20 different countries in Africa, Europe and Asia. As those were not previously
121 sampled in El Baidouri et al. (2013), South American strains were not included in this study. Thirty-
122 seven zymodemes, polymorphic in 11 enzymes, and 56 of the 64 different *Leishmania* sequence
123 types (LSTs) described in El Baidouri et al. (2013) (from 128 analysed strains) were represented.
124 The strains, preserved at the *Leishmania* French Reference Centre, were cultured in Schneider's
125 medium supplemented with 20% FBS and 1% human urine until the exponential growth stage and
126 then analysed.

127

128 *2.2. Isoenzymatic analysis of the GOT enzyme*

The GOT isoenzymatic analysis was repeated in this study, following the procedure described by Rioux et al. (1990), to confirm the previously reported data. Protein extracts were migrated in starch gels at pH 8.6 using nicotinamide adenine dinucleotide phosphate (NADP) as the substrate and stained using fast blue BB salt dye. When it was necessary to enhance the resolution, isoelectrofocusing (IEF) was performed in acrylamide gels using a pH gradient from 3 to 10 following the methodology described by Piarroux et al. (1994). Gels were prefocused for 15 min at 25 W and samples were run for an additional 15 min at 40 W. To make an inventory of the different alleles, the GOT electrophoretic mobility in gels was systematically compared with that of reference strains. Neighbour joining (NJ) trees were constructed using MEGA-X software (Kumar et al., 2018) from the complete MLEE profile of the strains or GOT results based on a presence/absence matrix.

2.3. Gene sequencing and genetic clustering

DNA was extracted from cultured *Leishmania* strains using a QIAamp DNA Mini Kit (Qiagen, Germany) following the manufacturer's instructions. For *got* and heat-shock protein 70 (*hsp70*) gene sequencing, 1239 bp and 1245 bp fragments were amplified, respectively, as previously described (Maurício et al., 2006; Van der Auwera et al., 2013). Amplicons were enzymatically purified using EXOSAP-IT (Affymetrix USB, USA) and double-stranded DNA was sequenced by the Sanger method at the Scientific and Technologic Centres of the Universitat de Barcelona, Spain. For the *got* locus, additional internal primers, got_Inc (5'-GCATGAGACGGCGCTGAT-3') and got_Fin (5'-TGCGAGATCGAAGATGACATC-3'), were designed. The obtained sequences were edited using Bionumerics software version 7.6.3 (Applied Maths, Belgium) and are available in GenBank under the accession numbers **MN728685 - MN728796**. In addition, for the 56 strains under study, 392 sequences corresponding to seven

153 MLSA loci previously analysed by El Baidouri et al. (2013) were extracted from the GenBank
154 database (see Supplementary Table S2).

155 For the clustering analysis, all nucleotides were duplicated to avoid information loss from
156 ambiguous states (e.g., A to AA or Y to CT). NJ trees were constructed with MEGA-X software
157 using the p-distance method and a bootstrap of 1000 replicates, *got* and *hsp70* genes, the MLSA
158 scheme and *got* allele 823.

159 In parallel, neighbour-nets (NN) were constructed using Splitstree version 4.14.8 software
160 (Huson and Bryant, 2006) from MLSA, *hsp70* and *got* datasets using the uncorrected p method and
161 an equal angles representation. For these analyses, additional sequences from the *L. tropica*
162 complex (*L. tropica* and *Leishmania aethiopica*) and *Leishmania major* complex (*L. major*,
163 *Leishmania turanica*, *Leishmania arabica* and *Leishmania gerbilli*) were downloaded from
164 GenBank (see Supplementary Table S3). MLSA, *hsp70* and *got* NN were then built on a total of
165 132, 90 and 69 strains, respectively.

166

167 *2.4. Allele distribution and isolation by distance evaluation*

168 Allele distribution and allelic introgression among the 56 strains were evaluated on
169 concatenated sequences from the seven loci included in the MLSA scheme plus *hsp70* and *got*
170 genes. MHOM/FR/78/LEM75 (BCN 527) was used as a reference strain for subsequent analysis.
171 Being non-contributory for genetic exchange analysis, conserved positions or singletons were not
172 included in the analysis. Genetic isolation by distance was assessed through a Mantel test using the
173 ade4 package of R software version 3.6.0 (R Core Team, www.R-project.org). The matrix of
174 pairwise genetic distances, previously calculated using MEGA-X software with the p-distance
175 method, was compared with a matrix of geographical distances obtained with the Geographic
176 Distance Matrix Generator version 1.2.3 application (Ersts, P.J., American Museum of Natural
177 History, Center for Biodiversity and Conservation, USA) available from

178 http://biodiversityinformatics.amnh.org/open_source/gdmg. Test values range from -1 (negative
179 correlation) to 0 (no correlation) to 1 (positive correlation). To represent identical genotypes found
180 in different countries, 23 additional strains described in El Baidouri et al. (2013) were included.

181

182 *2.5. MALDI-TOF analysis*

183 Sample preparation was done as described by Lachaud et al. (2017). Strain spectra were
184 acquired with an AutoFlex II MALDI TOF/TOF instrument from Bruker Daltonics (USA). Ten
185 replicates of each strain were obtained using Flex Control version 3.4 (Bruker Daltonics) software.
186 Spectra were converted to mzML format using the MSConvertGUI tool (ProteoWizad (Chambers et
187 al., 2012)) and processed and analysed using Bionumerics software version 7.6.3 (Applied Maths).
188 Baseline background noise was initially removed from raw spectra using the rolling disk tool (100
189 points size). Curves were smoothed with a Kaiser window operator (window size of 20 points and
190 beta of 10) and noise was again removed with a second rolling disk (200 points) after being
191 computed with a continuous wavelet transformation (CWT) algorithm (scale 1, noise percentile at
192 50, window size 501 points and minimum noise value of 0.2%). The peak detection threshold was
193 set on a signal to noise ratio of 5. The 10 spectra replicates of each strain were summarized into an
194 average spectrum using the "technical replicate" option and revised manually. Peaks from the
195 average spectra of all strains were matched using a constant tolerance of 5 and linear tolerance of
196 200. Peaks present in at least 10% of strains were selected for further comparative analysis.
197 Identification of groups and clustering analysis were performed by both a principal component
198 analysis (PCA) of quantitative values and by a NJ tree constructed using the Dice similarity
199 coefficient. Correlation between the resemblance of spectra and geographical distances was
200 estimated through a Mantel test using the ade4 R package. Taking into account the
201 absence/presence of the selected peaks, the Jaccard coefficient was used to calculate the matrix of
202 distance between spectra.

2.6. Shimodaira–Hasegawa test

The Shimodaira–Hasegawa (SH) test was used to evaluate the congruence between phylogenetic trees constructed from genetic (*hsp70*, *got*, MLSA), biochemical (MLEE) and mass spectrometry (MALDI-TOF) data. Maximum likelihood (ML) trees were reconstructed with genetic data using the GTR+I+G model with a bootstrap of 100 repetitions using PhyML software version 3.1 (Guindon et al., 2010). The SH test was conducted using PAUP* Phylogenetic analysis using parsimony (*and other methods) version 4.0a165 software (Swofford, D.L., 2003. Sinauer Associates, Sunderland, Massachusetts) for each ML tree versus the trees obtained using the other approaches and 100 additional random trees, to avoid bias in the results. Values obtained were *P* values for a null hypothesis of no difference between trees. Non-significant *P* values (>0.05) indicated no difference between trees, i.e. congruence.

2.7. Average nucleotide identity calculation

Available and comparable genomes from the *L. donovani* complex (*L. donovani* and *L. infantum*, *n*=12), *L. tropica* complex (*L. tropica* and *L. aethiopica*, *n*=5) and *L. major* complex (*L. major*, *L. turanica*, *L. arabica* and *L. gerbilli*, *n*=6) were downloaded from GenBank (see Supplementary Table S4). Average nucleotide identity (ANI) values were obtained by pairwise comparison of genomes inside each complex using the MUMmer at JSpeciesWS online service (Ribocon GmbH) available at http://jspecies.ribohost.com/jspeciesws/ (Richter et al., 2016).

2.8. Data accessibility

Raw data, matrices and command lines used to perform SH and Mantel tests were uploaded to MendeleyData under the DOI: 10.17632/jdyxs9nzg2.1.

3. Results

3.1. Systematic got gene sequencing: full correlation with isoenzyme criteria

For the 56 strains finally included in the study, isoenzymatic analysis of the GOT enzyme confirmed previous results: 31 *L. infantum* (GOT^{100}), 20 *L. donovani* (GOT^{113}) and *L. archibaldi* (GOT^{110}).

got sequence analysis revealed 11 different genotypes due to nine polymorphic positions along the 1239 pb fragment of the full open reading frame. Three of the polymorphic positions implied non-synonymous substitutions (i.e. positions 352, 419 and 823, see Supplementary Table S5). Only the change from aspartic acid to tyrosine, determined by a variation at position 823, was associated with a shift in the electric charge, which altered the protein electrophoretic migration. At that position, there was a full correlation between the nucleotide sequence and the electrophoretic mobility. No post-translational modification was detected. Therefore, strains with GOT^{113} and GOT^{100} mobilities were systematically associated with alleles 823G and 823T, respectively. GOT^{110} was systematically associated with heterozygosity at position 823, where both G and T alleles were present (noted as K according to the International Union of Pure and Applied Chemistry - International Union of Biochemistry (IUPAC-IUB) code). Consequently, clusters defined by alleles 823T, 823G and 823K were considered as reference groups in the present work.

3.2. MALDI-TOF analysis of the L. donovani complex: no evident substructure

MALDI-TOF spectra ranged from 2,000 to 13,000 m/z. Ninety-eight peak classes, mainly found in the zones of 3,000-4,000, 5,000-6,000 and 10,000-11,000 m/z, were considered for strain comparison, with an average of 45 analysed peaks per strain. Most of the peak classes (63) were present in the spectra of less than the 50% of the strains and only 10 peaks were present in all. PCA,

where the three strongest components explained 89.5% of variance in the input data (76.6% PC1(x), 7.08% PC2 (y) and 5.9% PC3(z)), showed a continuous distribution of the strains, with no evident substructure grouping (Fig. 1). The Mantel test gave a positive correlation between distances among strains, based on the presence/absence of the peaks, and geographical distances (r=0.308, *P*=0.001).

3.3. Comparative evaluation of genetic distances within different Leishmania complexes: low genetic variability of the L. donovani complex

Leishmania donovani complex genetic distances were assessed relative to other *Leishmania* complexes, i.e. those of *L. major* and *L. tropica*, by both NN analysis and systematic pairwise comparison of whole-genome sequences. Whatever the dataset used (i.e. an MLSA scheme based on seven genes, partial *hsp70* gene or complete *got* gene sequence), the three NN networks showed comparable structures, with the three complexes clearly separated (Fig. 2). When complex structures were compared, those of *L. major* and *L. tropica* were clearly more diversified (long branches) and structured (less boxes) compared with the *L. donovani* complex. Low genetic variability in the *L. donovani* complex did not appear to be related to the number of strains analysed. The total ratios of genotype number to sequence number were 84% (48/57), 77% (51/66) and 48% (81/168) for *L. tropica*, *L. major* and *L. donovani* complexes, respectively.

For a broader view of the genetic diversity, a pairwise alignment of full-length genomes belonging to *L. donovani*, *L. major* and *L. tropica* complexes was performed (Table 1). The intra-complex ANIs were different between complexes. The high *L. donovani* identity value at the complex level (99.42 ± 0.56) was comparable to values observed for *L. tropica* or *L. major* at the intraspecific level (*L. major* 99.38 ± 0.13, *L. tropica* 98.92 ± 0.24).

3.4. Comparison of isoenzyme-based taxonomic groups with non-isoenzyme-based clustering: partial concordance between approaches

As described above, isoenzyme-based taxonomic groups inside the *L. donovani* complex were correlated with a single nucleotide variation at the *got* gene position 823. In order to evaluate conservation of the three allelic groups defined by other approaches, we compared topologies of NJ phylogenetic trees constructed from isoenzymatic, mass spectrometry and genetic data (Fig. 3). A preliminary visual comparison revealed that the trees obtained using approaches directly related to the allelic position 823 (i.e. *got* and MLEE) were quite similar and supported by moderate bootstrap values (five out of nine were higher than 50%). Conversely, in trees obtained by approaches non-directly related to allelic position 823 (i.e. MLSA, *hsp70* and MALDI-TOF), multiple strains were reshuffled and the result was correlated with lower bootstrap values (all under 50%).

Similarly, the SH test applied to assess tree topology congruence revealed low non-significant values (P>0.05), indicating topological differences between trees (Table 2). Only the *got* tree was congruent with allele 823 and MLEE trees (SH value = 0.22 and SH value = 0.21, respectively). When position 823 was removed from the analysis, congruence between *got* and MLEE decreased to 0.02.

3.5. Evaluation of allele distribution within the L. donovani complex: large-scale allele exchanges

A total of 7161 bp were analysed from the concatenation of seven coding DNA sequences included in an MLSA scheme and *hsp70* and *got* genes. Eighty-four polymorphic positions were detected (65, 10 and nine for MLSA, *hsp70* and *got*, respectively). Being non-contributory for genetic exchange analysis, 48 singletons were hidden. Allele distribution was evaluated at the 36 remaining positions, and the genotype of the BCN 527 strain (MHOM/FR/78/LEM75), a classical reference strain for isoenzyme analysis, was arbitrarily selected as a reference (Fig. 4). This analysis revealed a striking reshuffling of most of the alleles and no substructure was identifiable. Furthermore, when taxonomic position *got* 823 was considered (Supplementary Fig. S1), there were no signs of co-segregation with the other evaluated loci except *got* position 522. Nearly all genotypes were unique and only four identical genotypes were shared by at least two and up to four

302 strains. In addition, heterozygous alleles were observed in at least one strain at 24 out of the 36
303 evaluated positions (67%). It should be highlighted that both parental alleles were systematically
304 found in the studied population except at five positions. Thirty-five of the 56 strains had at least one
305 heterozygous position and 10 strains had more than 10%, and up to 19%, heterozygosity.
306 The existence of isolation by distance was evaluated in the aforementioned 56 strains plus
307 23 additional strains described in El Baidouri et al. (2013) to prevent underrepresentation of
308 identical genotypes from different countries. The Mantel test showed significant positive correlation
309 between geographical and genetic distance matrices (r= 0.393, *P*=0.001). Similarly, it was possible
310 to observe an east-west gradient across the allelic prevalence map (Fig. 5). Additionally, seven
311 genotypes were present in at least two countries and up to 16 different countries from eastern
312 Europe, Africa, the Middle East and Asia. If the most ubiquitous genotype (BCN 527) was removed
313 from the Mantel test analysis, the correlation coefficient increased up to 0.557 (*P*=0.001).
314
315 **4. Discussion**
316 Clinical, geographical and epidemiological criteria have been used for decades to classify
317 *Leishmania* spp. responsible for leishmaniasis. Over time, the *L. donovani* complex was split into
318 different species, most of which have subsequently been considered synonymous and dropped, with
319 only two taxa, *L. donovani* and *L. infantum*, currently remaining (Akhoundi et al., 2016). Broadly,
320 the former includes human-hosted strains that potentially cause post-kala-azar dermal leishmaniasis
321 (PKDL) and the latter includes canid-hosted strains found in the Mediterranean basin (Pratlong et
322 al., 2013). The purpose of this taxonomy is to define pragmatic boxes based on clinical and
323 epidemiological criteria (Maurício, 2018). The isoenzymatic phenotype of the *got* gene is the sole
324 intrinsic criterium apparently consistent with this division (Pratlong et al., 2001).
325 Meanwhile, many exceptions to these clinical, epidemiological and geographical criteria
326 have been described. PKDL cases associated with *L. infantum* have been reported, mostly in HIV-
327 infected subjects (Scaglia et al., 1996; Dereure et al., 2003; Calza et al., 2004; Catorze et al., 2006;

328 Celesia et al., 2014) and vice versa, *L. donovani* strains have been isolated from patients suffering
329 cutaneous leishmaniasis, previously considered typical of *L. infantum* infection (Pratlong et al.,
330 1995; Karunaweera et al., 2003; Antoniou et al., 2008; Elamin et al., 2008), and from animals
331 (Dereure et al., 2003; Mohebali et al., 2004). Several authors have reported PCR positivity or
332 seropositivity in dogs in *L. donovani* foci, including in Bangladesh, Sri Lanka, India, Sudan and
333 Ethiopia, suggesting dogs act as *L. donovani* reservoirs (Rohousova et al., 2015). In turn, non-canid
334 reservoirs have been described for *L. infantum*, including mustelids and felids, and more recently
335 rodents and lagomorphs (Martín-Sánchez et al., 2007; Molina et al., 2012; Jiménez et al., 2014;
336 Millán et al., 2014; Risueño et al., 2018; Galán-Puchades et al., 2019). Contrary to the widely held
337 view, transmission of the different *Leishmania* spp. might only be partially dependent on
338 phlebotomine sand flies (Antoniou et al., 2013; Seblova et al., 2015). Furthermore, there is no
339 indication that sand fly species known to transmit *L. infantum* are not able to transmit *L. donovani*
340 under natural conditions and vice versa. The experimental study by Seblova et al. on *Phlebotomus*
341 *perniciosus* showed a comparable susceptibility of the vector for both *L. donovani* and *L. infantum*
342 (Seblova et al., 2015). In the field, although yet to be confirmed, *Phlebotomus alexandri* is
343 suspected to transmit *L. donovani* in Iran, Iraq and Oman, and is a reported vector of *L. infantum* in
344 China (Azizi et al., 2006; Maroli et al., 2013). From a geographical point of view, *L. infantum* and
345 *L. donovani* overlap in at least China, Cyprus, Israel, Iraq, Sudan and Yemen (Dereure et al., 2003;
346 Mazeris et al., 2010; Pratlong et al., 2013). Hence, the number of exceptions to the classical
347 taxonomy based on clinical, epidemiological and biochemical criteria is growing and they are likely
348 to be underreported due to potential circular reasoning.

349 To date, with a few exceptions, taxonomic studies of the *L. donovani* complex have
350 generally been based on the analysis of a low number of strains, often from concrete geographical
351 areas, or discriminatory genetic markers, producing unrepresentative and contradictory results
352 (Jamjoom et al., 2004; Mauricio et al., 2006; Lukes et al., 2007; Zemanová et al., 2007; Kuhls et al.,
353 2008; Gouzelou et al., 2013; Baleela et al., 2014; Banu et al., 2019). As we show in this work, the

resolving power of genetic analysis of the *L. donovani* complex is hampered by very low genetic diversity compared with other complexes, and especially by the strong genetic flow across its range.

Today, population structure and changes over time are assessed by whole genome analysis (Imamura et al., 2016; Rai et al., 2017). Depending on the studied foci and the geographical scale, various competing scenarios have been proposed. For example, in the Indian subcontinent, Imamura et al. (2016) reported an extremely low level of overall genetic diversity in a study of 204 parasite isolates obtained from confirmed VL patients (Imamura et al., 2016). The six closely related monophyletic groups detected by whole genome analysis were present throughout the sampling window (from 2002 to 2011) and area (both India and Nepal), indicating the weak genetic structure of the population across this large distribution area (eight of the 204 samples were clearly hybrid strains and no recombination was detected elsewhere).

A different scenario was described in a smaller focus in Ethiopia, where differences in ecology and sand fly vectors had been reported (Zackay et al., 2018). A genome-wide comparison of *L. donovani* strains revealed two distinct genetic clusters. The correlation between genotypes and geographical origin (northern versus. southern Ethiopia) was almost perfect (for 38/41 isolates), demonstrating that the populations are delineated by geography/ecology rather than clinical presentations. A third interesting example of *L. donovani* complex genetic structure was described by Rogers et al. (2014) in southeastern Turkey (Rogers et al., 2014). In this focus, *Leishmania* strains were hybrids descended from a single outcrossing between a parent similar to *L. infantum* strains described in Spain and another related to *L. donovani* strains, producing a genome-wide mosaic of ancestry from the two parents. As pointed out by Rogers et al., the cross was probably detected because the genetic backgrounds were sufficiently divergent.

The different scenarios and population genetic structures described above are not mutually exclusive and are compatible with our data. As shown in Figs. 4 and 5, and by the high Mantel test values, allele frequency was roughly correlated with the geographic distance along an east-west gradient (e.g. allelic positions 258 and 4287 in the MLSA analysis or position 370 in *hsp70*).

However, the most striking fact revealed by these data is a genetic mixing at most of the allelic positions. For example, allele C at position 1479 was present in strains isolated from China, India, Israel, Yemen, Turkey, Ukraine, eastern Africa and countries around the Mediterranean basin, whereas allele T was present in eastern Africa and countries around the Mediterranean basin. In addition, heterozygous alleles (*i.e.* C/T) were detected in strains from the Mediterranean basin and eastern Africa. Similarly, differences in the spectra obtained by MALDI-TOF analysis were correlated with geographic distances; however, no isolated groups of strains were identifiable.

Although no convincing chronology for the evolution is available, this genetic patchwork is probably the result of intensive genetic exchange over time. In our data, more than two-thirds of the allelic positions analysed were heterozygous in at least one strain, and remarkably both parental alleles were almost systematically found in the studied population. In addition, 62% of the strains had at least one heterozygous position and heterozygosity in 10 strains was higher than 10% and up to 19%. The high prevalence of heterozygous strains throughout the geographical range is congruent with the hypothesis of frequent allelic exchanges and genetic hybridization events within the population. It might also indicate that the heterozygous positions are stable over time, possibly selected for a functional role.

A key additional observation is the presence of identical or near-identical genotypes spread across the geographical range. For example, closely related genotypes BCN 527, 948 and 906 (equivalent to genotypes LST0001, LST0092 and LST0135 from El Baidouri et al. (2013)) were isolated not only all around the Mediterranean basin but also in central Asia, the Middle East, eastern and central Africa, Arabia and China. This result is supported by biochemical analysis of the ubiquitous distribution of zymodeme MON-1 (Kuhls et al., 2008; Pratlong et al., 2013). Furthermore, although strains from America were not included in our study, previous research has proved the presence of those ubiquitous genotypes in the continent. The low diversity found among American strains from the *L. donovani* complex, together with a full similarity with southwestern

405 European strains, lead Kuhls et al. (2011) to conclude that a recent importation had occurred (Kuhls
406 et al., 2011).

407 These widespread genotypes have adapted to various ecological conditions, and the mobility
408 of the reservoirs (i.e. humans and dogs) is probably a contributing factor. However, it is difficult to
409 make a convincing hypothesis to explain this pattern over several decades (see below).

410 Overall, the description of the *L. donovani* complex genetic structure involves
411 heterogeneous and interlinked scenarios. There is i) a continuum of the different genotypes across
412 the geographical range of the population; ii) a roughly east-west allele frequency gradient; iii) a
413 large amount of genetic exchange footprints; iv) the existence of ecotypes (i.e. distinct geographic
414 varieties genotypically adapted to specific environmental conditions) and local sub-structures, and
415 v) a large-scale geographical spread of a few genotypes.

416 Another question concerns evolution over time. The use of phylogenetic models has allowed
417 evolutionary histories of the *L. donovani* complex over 150 years to be proposed. For example,
418 Imamura et al. (2016) reported a divergence of the Indian population in the mid-19th century, which
419 matches historical large-scale outbreaks, while another divergence coincides with the end of the
420 dichlorodiphenyltrichloroethane (DDT) spraying campaign in the 1960s (Imamura et al., 2016).
421 However, on a shorter time-scale, the detection of rapid population change might be more
422 challenging. Baleela et al. (2014) analysed 124 *Leishmania* strains isolated between 1954 and 2006
423 in Sudan and their dataset revealed multiple examples of temporally separated strains which were
424 closely related genetically, at least on a time-scale of a few dozen years. As noted above, in our data
425 there are no signs of a recent evolution of the population structure pattern across the *L. donovani*
426 complex geographical range. In addition, from a temporal point of view, the analysed strains were
427 isolated over a 50 year period (1955 to 2005) and there is no evidence of a fast or recent
428 modification of the genotypes. For example, closely related genotypes BCN 988 and BCN 912 (two
429 allelic divergences among 7170 allelic positions analysed) were isolated in 1955 in Kenya and in
430 1990 in Ethiopia, respectively, and other similar examples are included in Fig. 4.

In conclusion, although there are obvious differences between foci, including in clinical presentations, reservoirs, vectors, etc., there is no genetic evidence for separating the two species within the *L. donovani* complex (at least if using the phylogenetic definition of species) (Staley, 2009; Maurício, 2018). The term 'sub-species' is too vague to take into account the network of genetic exchanges and introgressions between them. The term 'ecotype', which refers to a genetically distinct geographic population within a species adapted to specific environmental conditions, should be rejected for the same reason. Due to this absence of relevant terminology in existing classes, it would probably be more useful to analyse *L. donovani* as a single entity, as recently proposed by Maurício (2018).

Although the speciation process within the *L. donovani* complex can create reproductive isolation barriers between diverging populations, this might be countered by the increasing mobility of the reservoirs (i.e. humans and dogs). In addition, advanced sequencing technologies have shed light on genome-wide divergence between closely related species, revealing heterogeneous divergence patterns between entities during the speciation process, including in genomic regions resistant to introgression (Cruickshank and Hahn, 2014). Work on these islands of differentiation could help to develop a better understanding of the *L. donovani* complex structure and evolution.

Acknowledgements

The ISGlobal team is supported by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, Catalonia, Spain) (2017 SGR 00924), the Tropical Disease Cooperative Research Network (RICET, Spain) (RD12/0018/0010), and the Spanish Ministry of Science, Innovation and Universities through the "Centro de Excelencia Severo Ochoa 2019-2023" Program (CEX2018-000806-S). ISGlobal is a member of the Centres de Recerca de Catalunya (CERCA) Programme, Government of Catalonia (Spain). This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Akhoundi, M., Kuhls, K., Cannet, A., Votýpka, J., Marty, P., Delaunay, P., Sereno, D., 2016. A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. PLoS Negl. Trop. Dis. 10, e0004349. https://doi.org/10.1371/journal.pntd.0004349.

Antoniou, M., Haralambous, C., Mazeris, A., Pratlong, F., Dedet, J.P., Soteriadou, K., 2008. *Leishmania donovani* leishmaniasis in Cyprus. Lancet Infect. Dis. 8, 6–7. https://doi.org/10.1016/S1473-3099(07)70297-9.

Antoniou, M., Gramiccia, M., Molina, R., Dvorak, V., Volf, P., 2013. The role of indigenous phlebotomine sandflies and mammals in the spreading of leishmaniasis agents in the Mediterranean region. Euro Surveill. 18, 20540. https://doi.org/10.2807/1560-7917.ES2013.18.30.20540.

Azizi, K., Rassi, Y., Javadian, E., Motazedian, M.H., Rafizadeh, S., Yaghoobi Ershadi, M.R., Mohebali, M., 2006. *Phlebotomus* (*Paraphlebotomus*) *alexandri*: a probable vector of *Leishmania infantum* in Iran. Ann. Trop. Med. Parasitol. 100, 63–68. https://doi.org/10.1179/136485906X78454.

Baleela, R., Llewellyn, M.S., Fitzpatrick, S., Kuhls, K., Schönian, G., Miles, M.A., Maurício, I.L., 2014. *Leishmania donovani* populations in Eastern Sudan: temporal structuring and a link between human and canine transmission. Parasit. Vectors. 7, 496. https://doi.org/10.1186/s13071-014-0496-4.

Banu, S.S., Meyer, W., Ferreira-Paim, K., Wang, Q., Kuhls, K., Cupolillo, E., Schönian, G., Lee, R., 2019. A novel multilocus sequence typing scheme identifying genetic diversity amongst *Leishmania donovani* isolates from a genetically homogeneous population in the Indian subcontinent. Int. J. Parasitol. 49, 555–567. https://doi.org/10.1016/j.ijpara.2019.02.010.

Brumpt, E., 1936. Précis de Parasitologie. Masson et Cie, Paris.

Burza, S., Croft, S.L., Boelaert, M., 2018. Leishmaniasis. Lancet. 392, 951–970. https://doi.org/10.1016/S0140-6736(18)31204-2.

Calza, L., D'Antuono, A., Marinacci, G., Manfredi, R., Colangeli, V., Passarini, B., Orioli, R., Varoli, O., Chiodo, F., 2004. Disseminated cutaneous leishmaniasis after visceral disease in a patient with AIDS. J. Am. Acad. Dermatol. 50, 461–465. https://doi.org/10.1016/j.jaad.2003.10.005.

Castellani, A., Chalmers, A., 1919. Manual of Tropical Medicine, 3rd ed, Balliere, Tindall and Cox, London.

Catorze, G., Alberto, J., Afonso, A., Vieira, R., Cortes, S., Campino, L., 2006. Co-infection *Leishmania infantum*/VIH: Lésions cutanées après traitement d'une leishmaniose viscérale. Ann. Dermatol. Venereol. 133, 39–42. https://doi.org/10.1016/S0151-9638(06)70841-9.

Celesia, B.M., Cacopardo, B., Massimino, D., Gussio, M., Tosto, S., Nunnari, G., Pinzone, M.R., 2014. Atypical presentation of PKDL due to *Leishmania infantum* in an HIV-infected patient with relapsing visceral leishmaniasis. Case Rep. Infect. Dis. 2014, 370286. https://doi.org/10.1155/2014/370286.

Chambers, M.C., Maclean, B., Burke, R., Amode, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T.A., Brusniak, M.Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S.L., Nuwaysir, L.M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E.W., Moritz, R.L., Katz, J.E., Agus, D.B., MacCoss, M., Tabb, D.L., Mallick, P., 2012. A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. 30, 918–920. https://doi.org/10.1038/nbt.2377.

Cruickshank, T.E., Hahn, M.W., 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23, 3133–3157. https://doi.org/10.1111/mec.12796.

Cunha, A.M., Chagas, E., 1937. Nova especie de protozoario do genero *Leishmania* pathogenica para o homen: *Leishmania chagasi* n.sp. Nota prévia. Hospital (Rio J). XI, 148–152.

509 Cunha, A.M, 1938. A agglutinaçao e o diagnostico differencial das leishmanias. Bras. Med. 38,
510 849–855.
511 Dereure, J., El-Safi, S.H., Bucheton, B., Boni, M., Kheir, M.M., Davoust, B., Pratlong, F., Feugier,
512 E., Lambert, M., Dessein, A., Dedet, J.P., 2003. Visceral leishmaniasis in eastern Sudan:
513 parasite identification in humans and dogs; host-parasite relationships. Microbes Infect. 5,
514 1103–1108. https://doi.org/10.1016/j.micinf.2003.07.003
515 El Baidouri, F., Diancourt, L., Berry, V., Chevenet, F., Pratlong, F., Marty, P., Ravel, C., 2013.
516 Genetic structure and evolution of the *Leishmania* genus in Africa and Eurasia: what does
517 MLSA tell us. PLoS Negl. Trop. Dis. 7, e2255. https://doi.org/10.1371/journal.pntd.0002255.
518 Elamin, E.M., Guizani, I., Guerbouj, S., Gramiccia, M., El Hassan, A.M., Di Muccio, T., Taha,
519 M.A., Mukhtar, M.M., 2008. Identification of *Leishmania donovani* as a cause of cutaneous
520 leishmaniasis in Sudan. Trans. R. Soc. Trop. Med. Hyg. 102, 54–57.
521 https://doi.org/10.1016/j.trstmh.2007.10.005.
522 Galán-Puchades, M.T., Gómez-Samblás, M., Suárez-Morán, J.M., Osuna, A., Sanxis-Furió, J.,
523 Pascual, J., Bueno-Marí, R., Franco, S., Peracho, V., Montalvo, T., Fuentes, M. V., 2019.
524 Leishmaniasis in Norway rats in sewers, Barcelona, Spain. Emerg. Infect. Dis. 25, 1222–1224.
525 https://doi.org/10.3201/eid2506.181027.
526 Gouzelou, E., Haralambous, C., Antoniou, M., Christodoulou, V., Martinkovi, F., Živičnjak, T.,
527 Smirlis, D., Pratlong, F., Dedet, J.P., Özbel, Y., Toz, S.Ö., Presber, W., Schönian, G.,
528 Soteriadou, K., 2013. Genetic diversity and structure in *Leishmania infantum* populations from
529 southeastern Europe revealed by microsatellite analysis. Parasit. Vectors 6, 342.
530 https://doi.org/10.1186/1756-3305-6-342.
531 Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New
532 algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
533 performance of PhyML 3.0. Syst. Biol. 59, 307–321. https://doi.org/10.1093/sysbio/syq010.
534 Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol.

Biol. Evol. 23, 254–267. https://doi.org/10.1093/molbev/msj030.

Imamura, H., Downing, T., Van den Broeck, F., Sanders, M.J., Rijal, S., Sundar, S., Mannaert, A., Vanaerschot, M., Berg, M., De Muylder, G., Dumetz, F., Cuypers, B., Maes, I., Domagalska, M., Decuypere, S., Rai, K., Uranw, S., Bhattarai, N.R., Khanal, B., Prajapati, V.K., Sharma, S., Stark, O., Schönian, G., De Koning, H.P., Settimo, L., Vanhollebeke, B., Roy, S., Ostyn, B., Boelaert, M., Maes, L., Berriman, M., Dujardin, J.C., Cotton, J.A., 2016. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. Elife 5, e12613. https://doi.org/10.7554/eLife.12613.

Jamjoom, M.B., Ashford, R.W., Bates, P.A., Chance, M.L., Kemp, S.J., Watts, P.C., Noyes, H.A., 2004. *Leishmania donovani* is the only cause of visceral leishmaniasis in East Africa; previous descriptions of *L. infantum* and *"L. archibaldi"* from this region are a consequence of convergent evolution in the isoenzyme data. Parasitology. 129, 399–409. https://doi.org/10.1017/S0031182004005955.

Jiménez, M., González, E., Martín-Martín, I., Hernández, S., Molina, R., 2014. Could wild rabbits (*Oryctolagus cuniculus*) be reservoirs for *Leishmania infantum* in the focus of Madrid, Spain? Vet. Parasitol. 202, 296–300. https://doi.org/10.1016/j.vetpar.2014.03.027.

Karunaweera, N.D., Pratlong, F., Siriwardane, H.V.Y.D., Ihalamulla, R.L., Dedet, J.P., 2003. Sri Lankan cutaneous leishmaniasis is caused by *Leishmania donovani* zymodeme MON-37. Trans. R. Soc. Trop. Med. Hyg. 97, 380–381. https://doi.org/10.1016/S0035-9203(03)90061-7.

Kuhls, K., Maurício, I.L., Pratlong, F., Presber, W., Schönian, G., 2005. Analysis of ribosomal DNA internal transcribed spacer sequences of the *Leishmania donovani* complex. Microbes Infect. 7, 1224–1234. https://doi.org/10.1016/j.micinf.2005.04.009.

Kuhls, K., Chicharro, C., Cañavate, C., Cortes, S., Campino, L., Haralambous, C., Soteriadou, K., Pratlong, F., Dedet, J.P., Maurício, I., Miles, M., Schaar, M., Ochsenreither, S., Radtke, O.A., Schönian, G., 2008. Differentiation and gene flow among European populations of *Leishmania infantum* MON-1. PLoS Negl. Trop. Dis. 2, e261.

https://doi.org/10.1371/journal.pntd.0000261.

Kuhls, K., Alam, M.Z., Cupolillo, E., Ferreira, G.E.M., Maurício, I.L., 2011. Comparative microsatellite typing of New World *Leishmania infantum* reveals low heterogeneity among populations and its recent Old World origin. PLoS Negl. Trop. Dis. 5, e1155. https://doi.org/10.1371/journal.pntd.0001155.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X : Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35,1547−1549. https://doi:10.1093/molbev/msy096.

Lachaud, L., Fernández-Arévalo, A., Normand, A.C., Lami, P., Nabet, C., Donnadieu, J.L., Piarroux, M., Djenad, F., Cassagne, C., Ravel, C., Tebar, S., Llovet, T., Blanchet, D., Demar, M., Harrat, Z., Aoun, K., Bastien, P., Muñoz, C., Gállego, M., Piarroux, R., 2017. Identification of *Leishmania* by matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry using a free web-based application and a dedicated mass-spectral library. J. Clin. Microbiol. 55, 2924–2933. https://doi.org/10.1128/JCM.00845-17.

Lainson, R., Shaw, J.J., 1987. Evolution, classification and geographical distrbution. In: Peters, W., Killick-Kendrick, R. (Eds.), The leishmaniases in biology and medicine. Academic Press, London, pp. 1−120.

Lukes, J., Maurício, I.L., Schönian, G., Dujardin, J.C., Soteriadou, K., Dedet, J.P., Kuhls, K., Tintaya, K.W.Q., Jirků, M., Chocholová, E., Haralambous, C., Pratlong, F., Oborník, M., Horák, A., Ayala, F.J., Miles, M.A., 2007. Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. Proc. Natl. Acad. Sci. U. S. A. 104, 9375–9380. https://doi.org/10.1073/pnas.0703678104.

Maroli, M., Feliciangeli, M.D., Bichaud, L., Charrel, R.N., Gradoni, L., 2013. Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. Med. Vet. Entomol. 27, 123–147. https://doi.org/10.1111/j.1365-2915.2012.01034.x.

Martín-Sánchez, J., Acedo, C., Muñoz-Pérez, M., Pesson, B., Marchal, O., Morillas-Márquez, F.,

2007. Infection by *Leishmania infantum* in cats: epidemiological study in Spain. Vet. Parasitol. 145, 267–273. https://doi.org/10.1016/j.vetpar.2006.11.005.

Maurício, I.L., Yeo, M., Baghaei, M., Doto, D., Pratlong, F., Zemanova, E., Dedet, J.P., Lukes, J., Miles, M.A., 2006. Towards multilocus sequence typing of the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD). Int. J. Parasitol. 36, 757–769. https://doi.org/10.1016/j.ijpara.2006.03.006.

Maurício, I.L., 2018. *Leishmania* taxonomy. In: Bruschi, F., Gradoni, L. (Eds.), The leishmaniases: old neglected tropical diseases. Springer, pp. 15–30.

Mazeris, A., Soteriadou, K., Dedet, J.P., Haralambous, C., Tsatsaris, A., Moschandreas, J., Messaritakis, I., Christodoulou, V., Papadopoulos, B., Ivović, V., Pratlong, F., Loucaides, F., Antoniou, M., 2010. Leishmaniases and the Cyprus paradox. Am. J. Trop. Med. Hyg. 82, 441–448. https://doi.org/10.4269/ajtmh.2010.09-0282.

Millán, J., Ferroglio, E., Solano-Gallego, L., 2014. Role of wildlife in the epidemiology of *Leishmania infantum* infection in Europe. Parasitol. Res. 113, 2005–2014. https://doi.org/10.1007/s00436-014-3929-2.

Mohebali, M., Javadian, E., Yaghoobi-Ershadi, M.R., Akhavan, A.A., Hajjaran, H., Abaei, M.R., 2004. Characterization of *Leishmania* infection in rodents from endemic areas of the Islamic Republic of Iran. East. Mediterr. Health J. 10, 591–599.

Molina, R., Jiménez, M.I., Cruz, I., Iriso, A., Martín-Martín, I., Sevillano, O., Melero, S., Bernal, J., 2012. The hare (*Lepus granatensis*) as potential sylvatic reservoir of *Leishmania infantum* in Spain. Vet. Parasitol. 190, 268–271. https://doi.org/10.1016/j.vetpar.2012.05.006.

Nicoli, R., 1963. Le genre *Leishmania* R. Ross, 1903. Bull. Soc. Pathol. Exot. 56, 408–416.

Nicolle, C., 1908. Su trois cas d'infection splénique infantile a corps de Leishman observés en Tunisie. Arch. Inst. Pasteur Tunis. 3–26.

Piarroux, R., Trouvé, V., Pratlong, F., Martini, A., Lambert, M., Rioux, J.A., 1994. The use of isoelectric focusing on polyacrylamide gel for the enzymatic analysis of 'Old World'

Leishmania species. Trans. R. Soc. Trop. Med. Hyg. 88, 475–478. https://doi.org/10.1016/0035-9203(94)90439-1.

Pratlong, F., Bastien, P., Perello, R., Lami, P., Dedet, J.P., 1995. Human cutaneous leishmaniasis caused by *Leishmania donovani sensu stricto* in Yemen. Trans. R. Soc. Trop. Med. Hyg. 89, 398–399. https://doi.org/10.1016/0035-9203(95)90025-x.

Pratlong, F., Dereure, J., Bucheton, B., El-Safi, S., Dessein, A., Lanotte, G., Dedet, J.P., 2001. Sudan: the possile original focus of visceral leishmaniasis. Parasitology 122 (6), 599–605. https://doi.org/doi: 10.1017/S0031182001007867

Pratlong, F., Lami, P., Ravel, C., Balard, Y., Dereure, J., Serres, G., El Baidouri, F., Dedet, J.P., 2013. Geographical distribution and epidemiological features of Old World *Leishmania infantum* and *Leishmania donovani* foci, based on the isoenzyme analysis of 2277 strains. Parasitology 140, 423–434. https://doi.org/10.1017/S0031182012001825.

Rai, K., Bhattarai, N.R., Vanaerschot, M., Imamura, H., Gebru, G., Khanal, B., Rijal, S., Boelaert, M., Pal, C., Karki, P., Dujardin, J.C., Van der Auwera, G., 2017. Single locus genotyping to track *Leishmania donovani* in the Indian subcontinent: application in Nepal. PLoS Negl. Trop. Dis. 11, e0005420. https://doi.org/10.1371/journal.pntd.0005420.

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., Peplies, J., 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics. 32, 929–931. https://doi.org/10.1093/bioinformatics/btv681.

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. Ann. Parasitol. Hum. Comp. 65, 111–125. https://doi.org/10.1051/parasite/1990653111.

Risueño, J., Ortuño, M., Pérez-Cutillas, P., Goyena, E., Maia, C., Cortes, S., Campino, L., Bernal, L.J., Muñoz, C., Arcenillas, I., Martínez-Rondán, F.J., Gonzálvez, M., Collantes, F., Ortiz, J., Martínez-Carrasco, C., Berriatua, E., 2018. Epidemiological and genetic studies suggest a common *Leishmania infantum* transmission cycle in wildlife, dogs and humans associated to

vector abundance in Southeast Spain. Vet. Parasitol. 259, 61–67. https://doi.org/10.1016/j.vetpar.2018.05.012.

Rogers, M.B., Downing, T., Smith, B.A., Imamura, H., Sanders, M., Svobodova, M., Volf, P., Berriman, M., Cotton, J.A., Smith, D.F., 2014. Genomic confirmation of hybridisation and recent inbreeding in a vector-isolated *Leishmania* population. PLoS Genet. 10, e1004092. https://doi.org/10.1371/journal.pgen.1004092.

Rohousova, I., Talmi-Frank, D., Kostalova, T., Polanska, N., Lestinova, T., Kassahun, A., Yasur-Landau, D., Maia, C., King, R., Votypka, J., Jaffe, C.L., Warburg, A., Hailu, A., Volf, P., Baneth, G., 2015. Exposure to *Leishmania* spp. and sand flies in domestic animals in northwestern Ethiopia. Parasit. Vectors. 8, 360. https://doi.org/10.1186/s13071-015-0976-1.

Ross, R., 1903. Note on the bodies recently described by Leishman and Donovan. Br. Med. J. 2, 1261–1262. https://doi.org/10.1136/bmj.2.2237.1261.

Scaglia, M., Malfitano, A., Douville, H., Sacchi, P., Gatti, S., Gradoni, L., Gramiccia, M., 1996. Dermonodular and visceral leishmaniasis due to *Leishmania infantum* with a new isoenzyme pattern: report of a case involving a patient with AIDS. Clin. Infect. Dis. 22, 376–377. https://doi.org/10.1093/clinids/22.2.376.

Seblova, V., Myskova, J., Hlavacova, J., Votypka, J., Antoniou, M., Volf, P., 2015. Natural hybrid of *Leishmania infantum*/*L. donovani*: development in *Phlebotomus tobbi*, *P. perniciosus* and *Lutzomyia longipalpis* and comparison with non-hybrid strains differing in tissue tropism. Parasit. Vectors. 8, 605. https://doi.org/10.1186/s13071-015-1217-3.

Staley, J.T., 2009. Universal species concept: pipe dream or a step toward unifying biology? J. Ind. Microbiol. Biotechnol. 36, 1331–1336. https://doi.org/10.1007/s10295-009-0642-8.

Van der Auwera, G., Maes, I., De Doncker, S., Ravel, C., Cnops, L., Van Esbroeck, M., Van Gompel, A., Clerinx, J., Dujardin, J.C., 2013. Heat-shock protein 70 gene sequencing for *Leishmania* species typing in European tropical infectious disease clinics. Euro Surveill. 18, 20543. https://doi.org/10.2807/1560-7917.es2013.18.30.20543.

World Health Organization, 2010. Control of the leishmaniasis: report of a meeting of the WHO Expert Commitee on the Control of Leishmaniases, Geneva, 22-26 March 2010. Technical report series, 949. Available from: https://apps.who.int/iris/handle/10665/44412.

Zackay, A., Cotton, J.A., Sanders, M., Hailu, A., Nasereddin, A., Warburg, A., Jaffe, C.L., 2018. Genome wide comparison of Ethiopian *Leishmania donovani* strains reveals differences potentially related to parasite survival. PLOS Genet. 14, e1007133. https://doi.org/10.1371/journal.pgen.1007133.

Zemanová, E., Jirků, M., Maurício, I.L., Horák, A., Miles, M.A., Lukes, J., 2007. The *Leishmania donovani* complex: genotypes of five metabolic enzymes (ICD, ME, MPI, G6PDH, and FH), new targets for multilocus sequence typing. Int. J. Parasitol. 37, 149–60. https://doi.org/10.1016/j.ijpara.2006.08.008.

678 **Legends to Figures**

679

680 **Fig. 1.** Principal component analysis (PCA) of the *Leishmania donovani* complex MALDI-TOF data.
681 The PCA is colour-coded according to the allele 823 variations of the glutamate-oxaloacetate
682 transaminase gene.

683

684 **Fig. 2**. Genetic distance evaluation of multilocus sequence analysis (MLSA), heat-shock protein 70
685 (*hsp70*) and glutamate-oxaloacetate transaminase (*got*) genes by neighbour-nets (NN) in three
686 different *Leishmania* complexes.. The *Leishmania major* complex gathers *L. major*, *Leishmania*
687 *turanica*, *Leishmania arabica* and *Leishmania gerbilli* taxa. The *Leishmania tropica* complex gathers
688 *L. tropica*, *Leishmania killicki* and *Leishmania aethiopica* taxa. The *Leishmania donovani* complex
689 gathers *L. donovani*, *Leishmania infantum* and *Leishmania archibaldi* taxa. At the right side of each
690 NN, a box shows an amplification of the *L. donovani* complex net regions . The nodes of the
691 amplification are colour-coded according to the allele 823 variations of the *got* gene.

692

693 **Fig. 3.** Tree topology comparison among different approaches. Neighbour joining trees of the
694 *Leishmania donovani* complex based on genetic data, multilocus enzyme electrophoresis (MLEE)
695 and MALDI-TOF were analysed. Colour coding is based on allele 823 variations of the glutamate-
696 oxaloacetate transaminase (*got*) gene. Bootstrap value was added for remarkable nodes. Distance
697 values for truncated outgroup branches appear bracketed. *hsp70*, heat-shock protein 70 gene; MSLA,
698 multilocus sequence analysis concatenated genes.

699

700 **Fig. 4.** Allele sharing between the *Leishmania donovani* complex genotypes. Fifty-six analysed
701 genotypes (singletons not included) were compared and ordered according to a neighbour joining tree
702 analysis (represented on the left side and colour-coded according to allele 823 variations of the
703 glutamate-oxaloacetate transaminase (*got*) gene). Country and year of isolation are indicated on the

right side. Genotypes are colour-coded according to BCN 527 (MHOM/FR/78/LEM75) alleles (in bold) as a reference. M.B., Mediterranean Basin; E.A., East Africa; I.S., Indian subcontinent; M.E., Middle East and S.E., South Europe. *hsp70*, heat-shock protein 70 gene; MSLA, multilocus sequence analysis genes.

Fig. 5. Geographical allele distribution of the *Leishmania donovani* complex. Allelic frequencies of each strain using BCN 527 (MHOM/FR/78/LEM75) as the reference are represented by pie charts. Genotypes are colour-coded according to BCN 527 alleles. Strains without geographical information (i.e. BCN 969 and BCN 989) are not represented. Stars indicate added strains analysed in El Baidouri et al. (2013). Map from the NASA Worldview Snapshots application.

Legends to Supplementary figures

Supplementary Fig. S1. Allele sharing between *Leishmania donovani* complex genotypes organized by 823 allele variants of the glutamate-oxaloacetate transaminase (*got*) gene. Fifty-six analysed genotypes (singletons not included) were compared and ordered according to allele 823 variations of the *got* gene. Genotypes are colour-coded using BCN 527 alleles (in bold) as a reference. *hsp70*, heat-shock protein 70 gene; MSLA, multilocus sequence analysis genes.

Table 1. Comparison of whole genome variability between different *Leishmania* complexes and taxa.

Taxa	Genomes	ANI (%)
Leishmania donovani complex	*n* = 12	99.42 ± 0.56
Leishmania tropica complex	*n* = 5	97.55 ± 1.62
Leishmania major complex	*n* = 6	96.50 ± 3.015
L. donovani	*n* = 8	99.57 ± 0.42
Leishmania infantum	*n* = 4	99.80 ± 0.17
Leishmania tropica	*n* = 3	98.92 ± 0.24
Leishmania aethiopica	*n* = 2	99.12
L. major	*n* = 3	99.38 ± 0.13
Leishmania turanica	*n* = 1	NA
Leishmania arabica	*n* = 1	NA
Leishmania gerbilli	*n* = 1	NA

Interval of average nucleotide identity (ANI) values obtained by genome comparison within each taxon. Pairwise comparison was performed with MUMmer using the JSpeciesWS web application on complete *Leishmania* genomes available in GenBank (listed in Supplementary Table S4).

NA, not applicable.

732 **Table 2.** Shimodaira-Hasegawa test for tree congruence.

Tree topology	***got***		***got w/o 823***		**MLSA**		***hsp70***	
	Diff -ln L	SH-test *P* values	Diff -ln L	SH-test *P* values	Diff -ln L	SH-test *P* values	Diff -ln L	SH-test *P* values
Allele *got*-823	57.0450	**0.22**	61.9450	0.02	824.4488	0.00	93.4464	0.00
MLEE	55.0232	**0.21**	58.6352	0.02	543.4564	0.00	61.9529	0.02
got	Best	NA	8.1739	**0.62**	765.0142	0.00	78.6692	0.00
got w/o 823	20.6977	**0.55**	Best	NA	857.5685	0.00	83.1619	0.00
MLSA	142.3409	0.00	90.1543	0.01	Best	NA	55.0976	**0.07**
hsp70	166.5193	0.00	106.4447	0.00	706.1710	0.00	Best	NA
MALDI	142.8053	0.00	100.8120	0.00	642.4023	0.00	68.0326	0.00

733 Pairwise likelihood-score differences (Diff -ln L) between topologies of the trees listed in the first
734 column and the glutamate-oxaloacetate transaminase (*got*) gene, *got* gene without position 823 (i.e.
735 got w/o 823), multilocus sequence analysis (MLSA, i.e. concatenation of seven loci 03.0980-
736 31.2610) and heat-shock protein 70 (*hsp70*) gene trees. Shimodaira-Hasegawa (SH) *P* values were
737 considered significant when <0.05. In bold, non-significant values indicate no significant difference
738 between trees, i.e. congruence.

739 NA, not applicable; MLEE, multilocus enzyme electrophoresis.

740

741

Figure 1 colour

Click here to access/download;Figure;IJPara20_068_R2_Fig_1_COL_CMunoz.tif

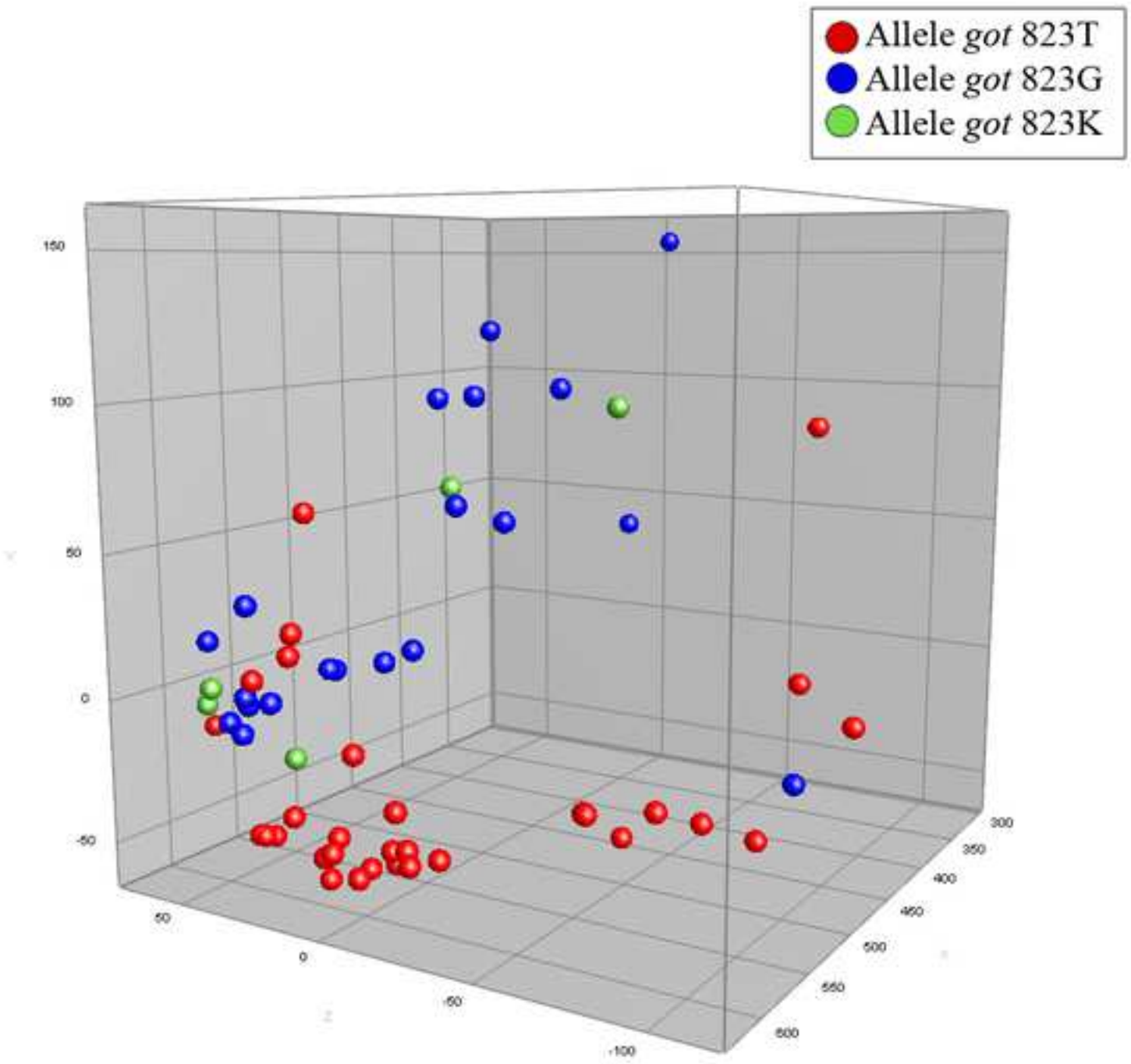


Figure 1 black and white

Click here to access/download;Figure;IJPara20_068_R2_Fig_1_BW_CMunoz.tif

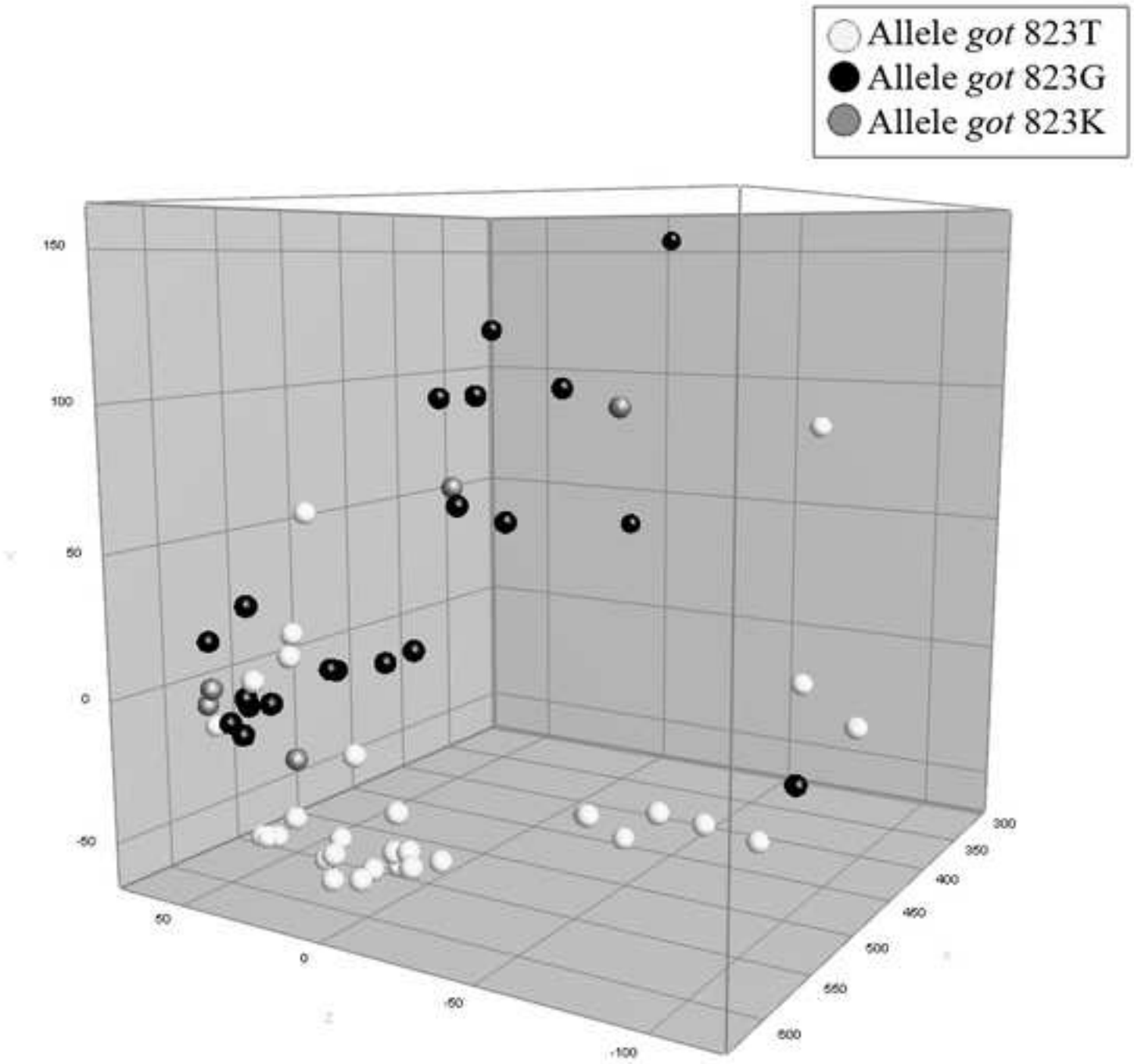


Figure 2 colour

Click here to access/download;Figure;IJPara20_068_R2_Fig_2_COL_CMUno

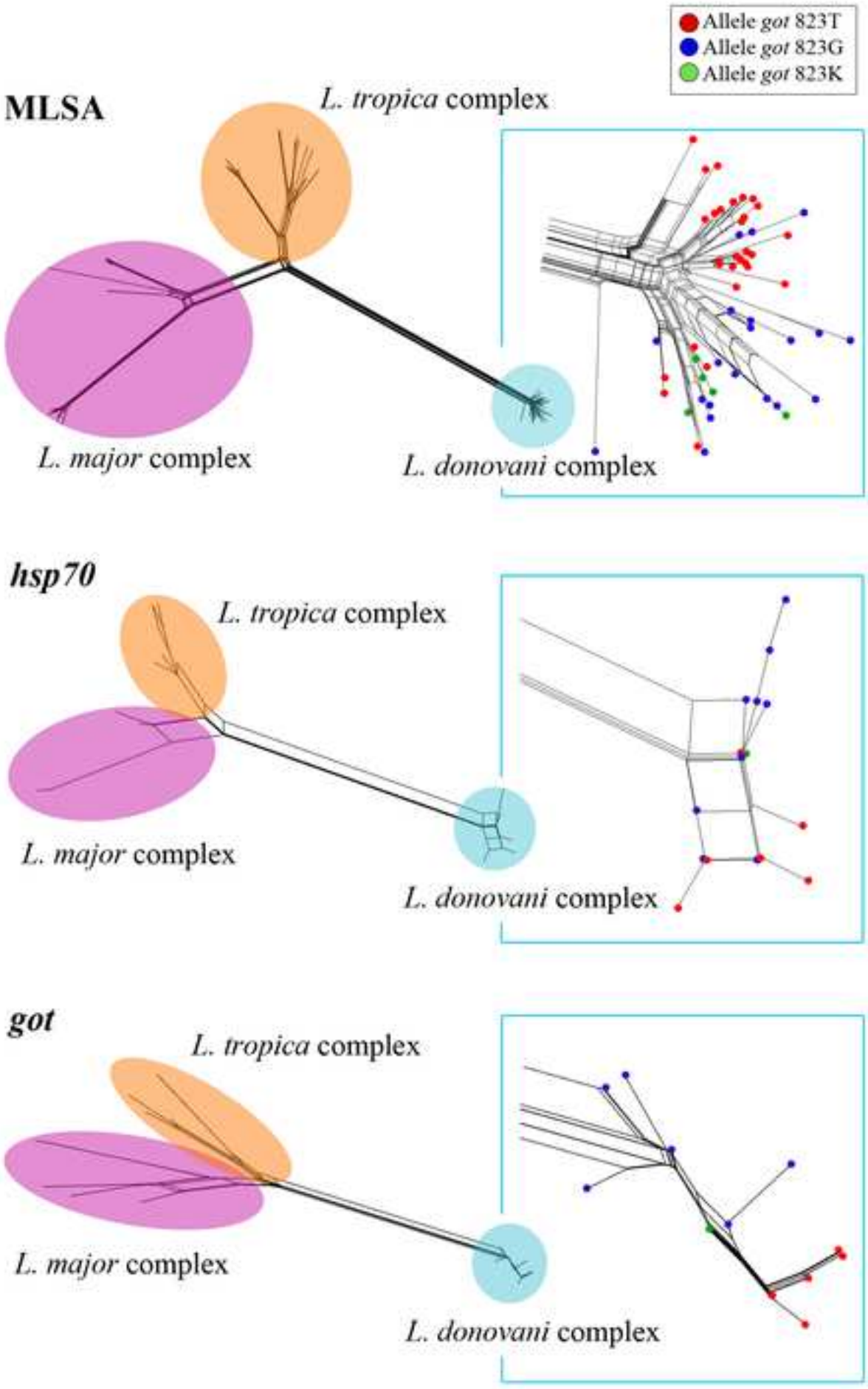


Figure 2 black and white

Click here to access/download;Figure;IJPara20_068_R2_Fig_2_BW_CMunoz.

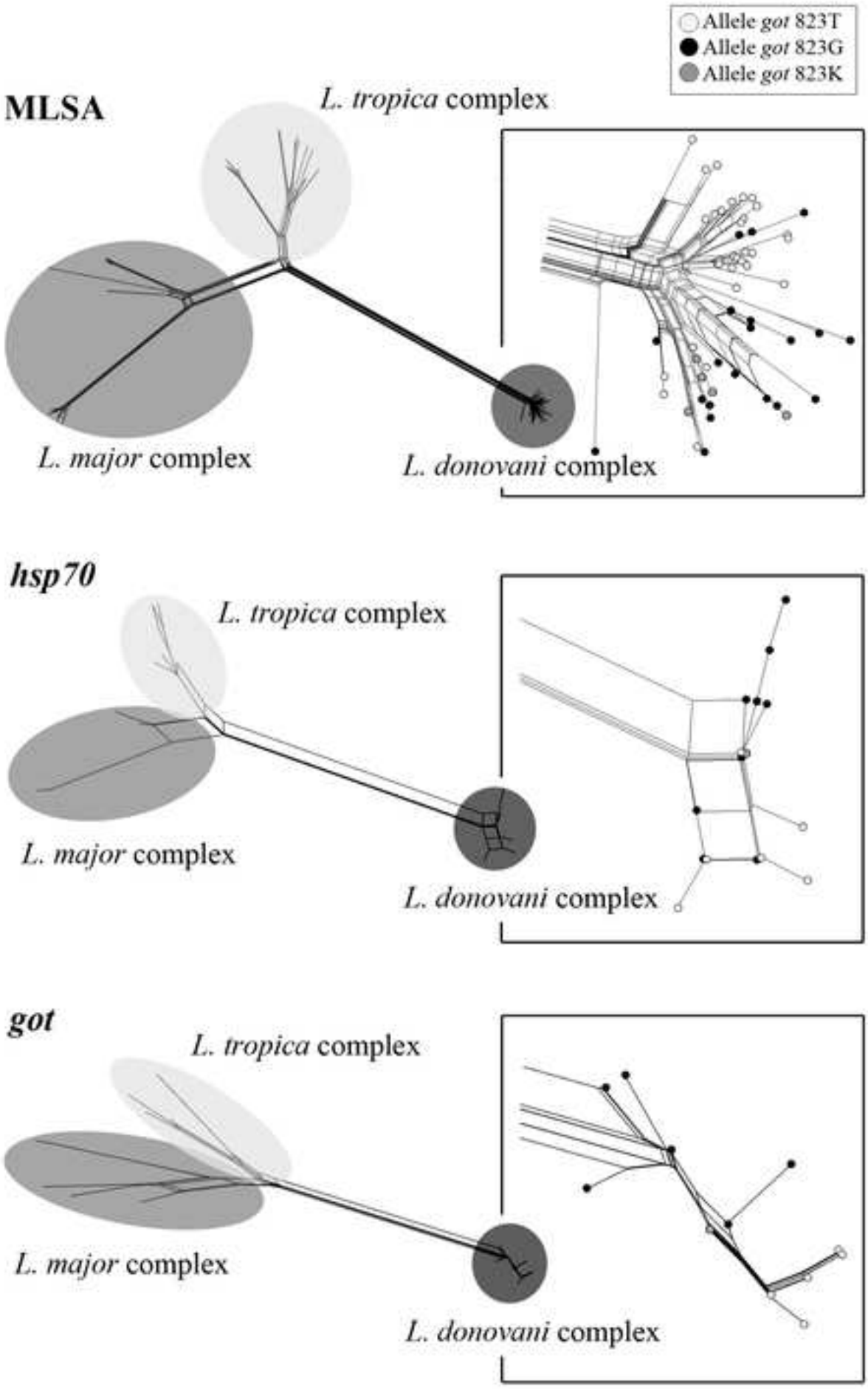


Figure 3 colour

Click here to access/download;Figure;IJPara20_068_R2_Fig_3_COL_CMunoz.tif

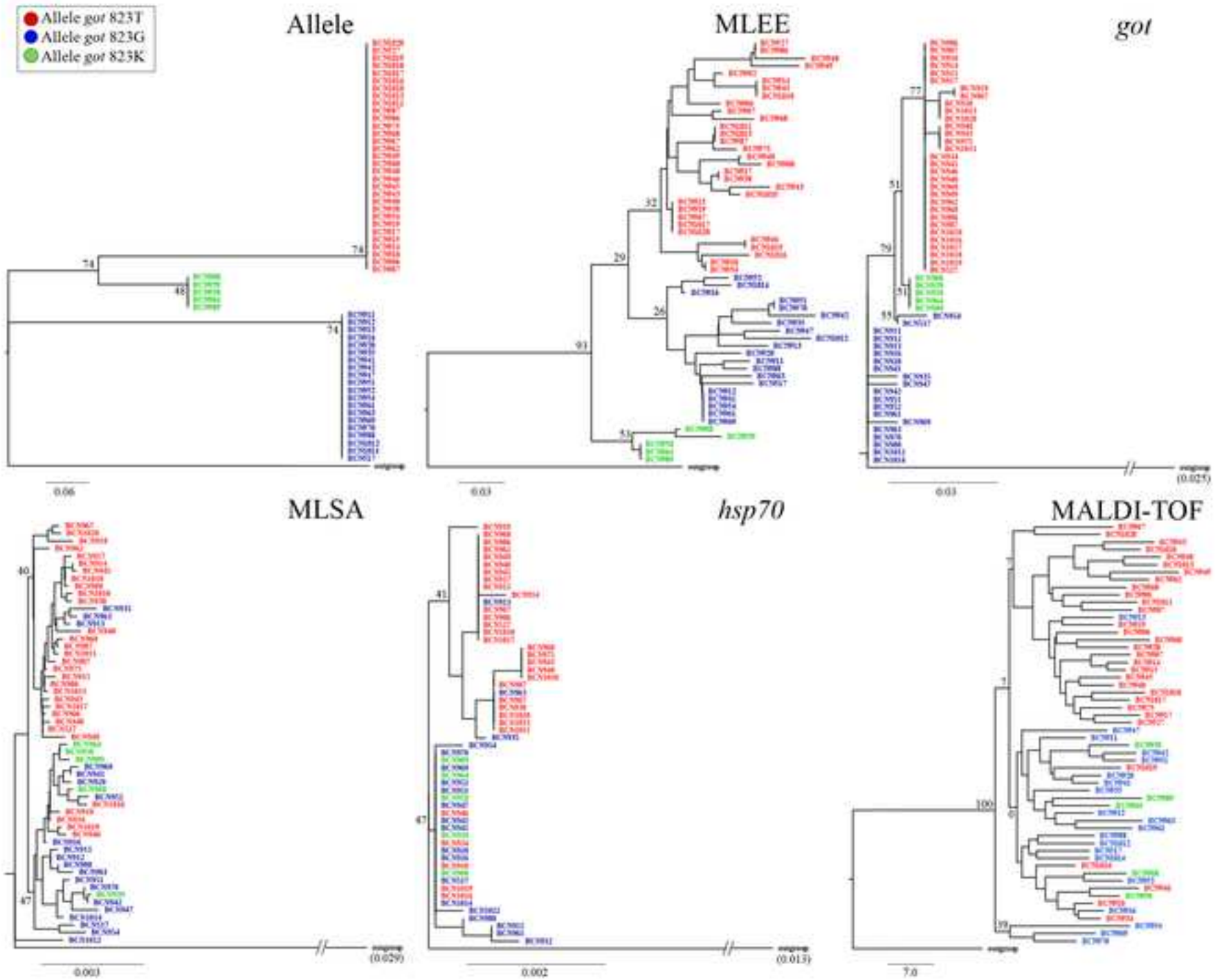


Figure 3 black and white

Click here to access/download;Figure;IJPara20_068_R2_Fig_3_BW_CMunoz.tif

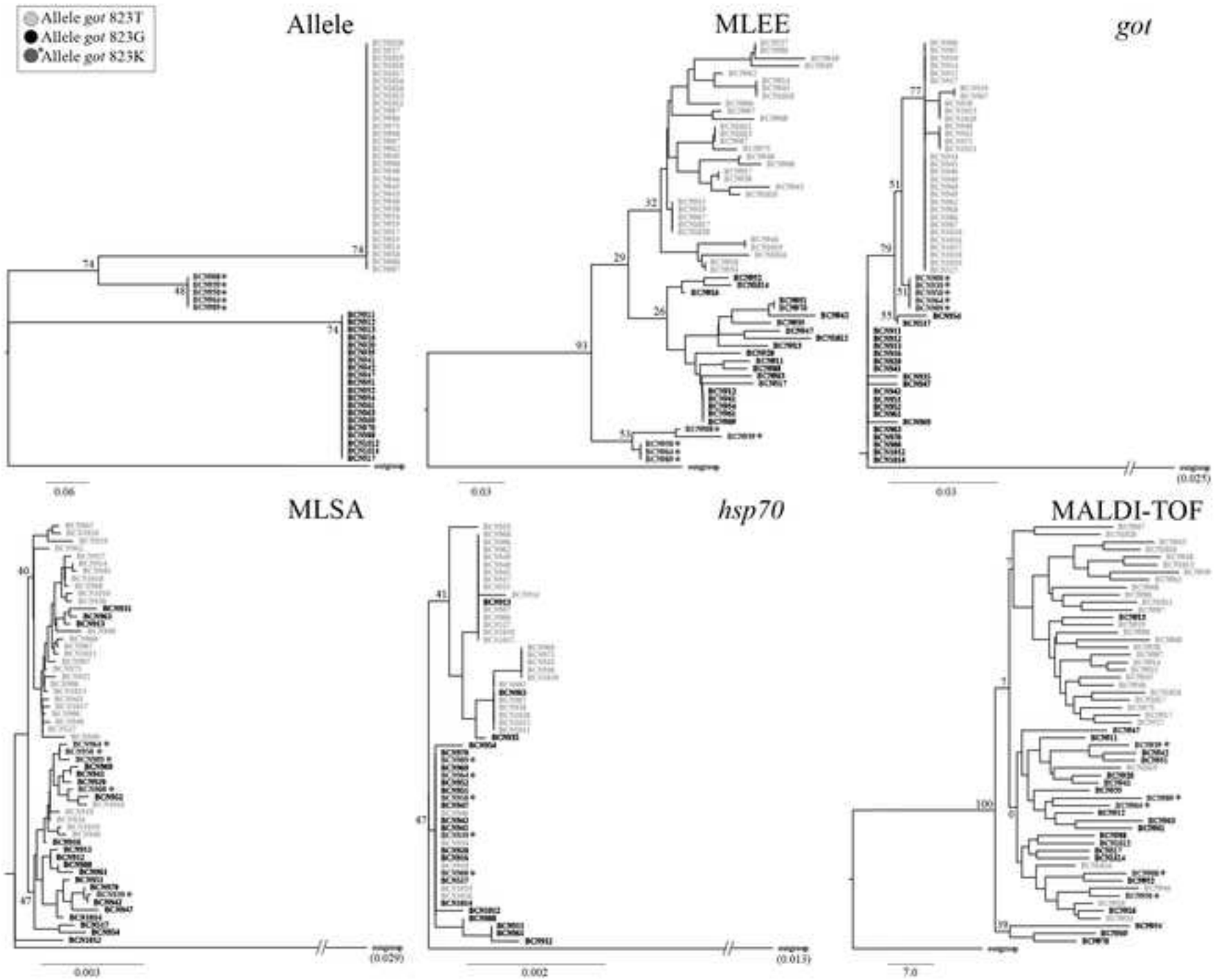


Figure 4 colour

Click here to access/download;Figure;IJPara20_068_R2_Fig_4_COL_CMunoz.tif

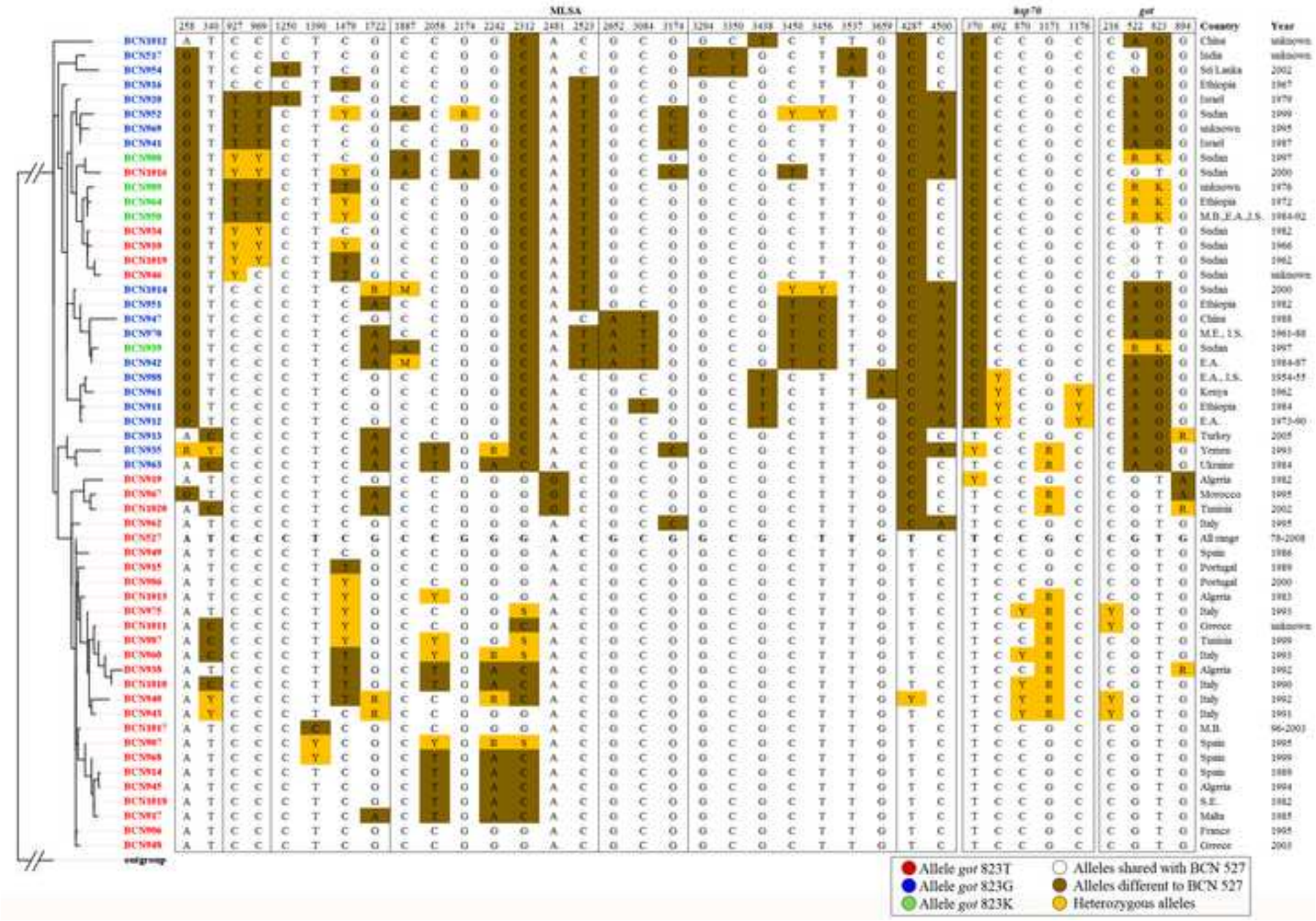


Figure 4 black and white

Click here to access/download;Figure;IJPara20_068_R2_Fig_4_BW_CMunoz.tif

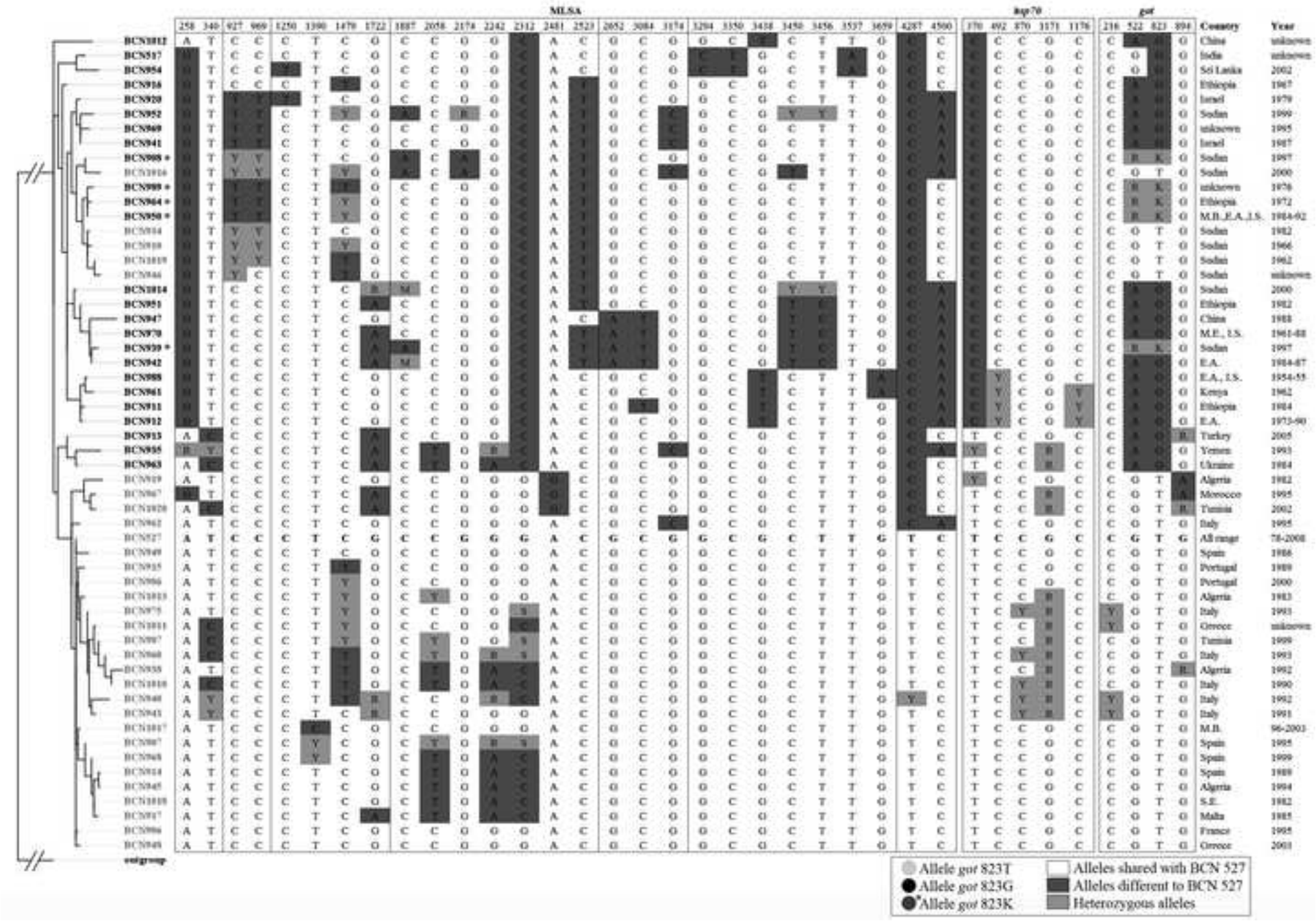


Figure 5 colour

Click here to access/download;Figure;IJPara20_068_R2_Fig_5_COL_CMunoz.tif

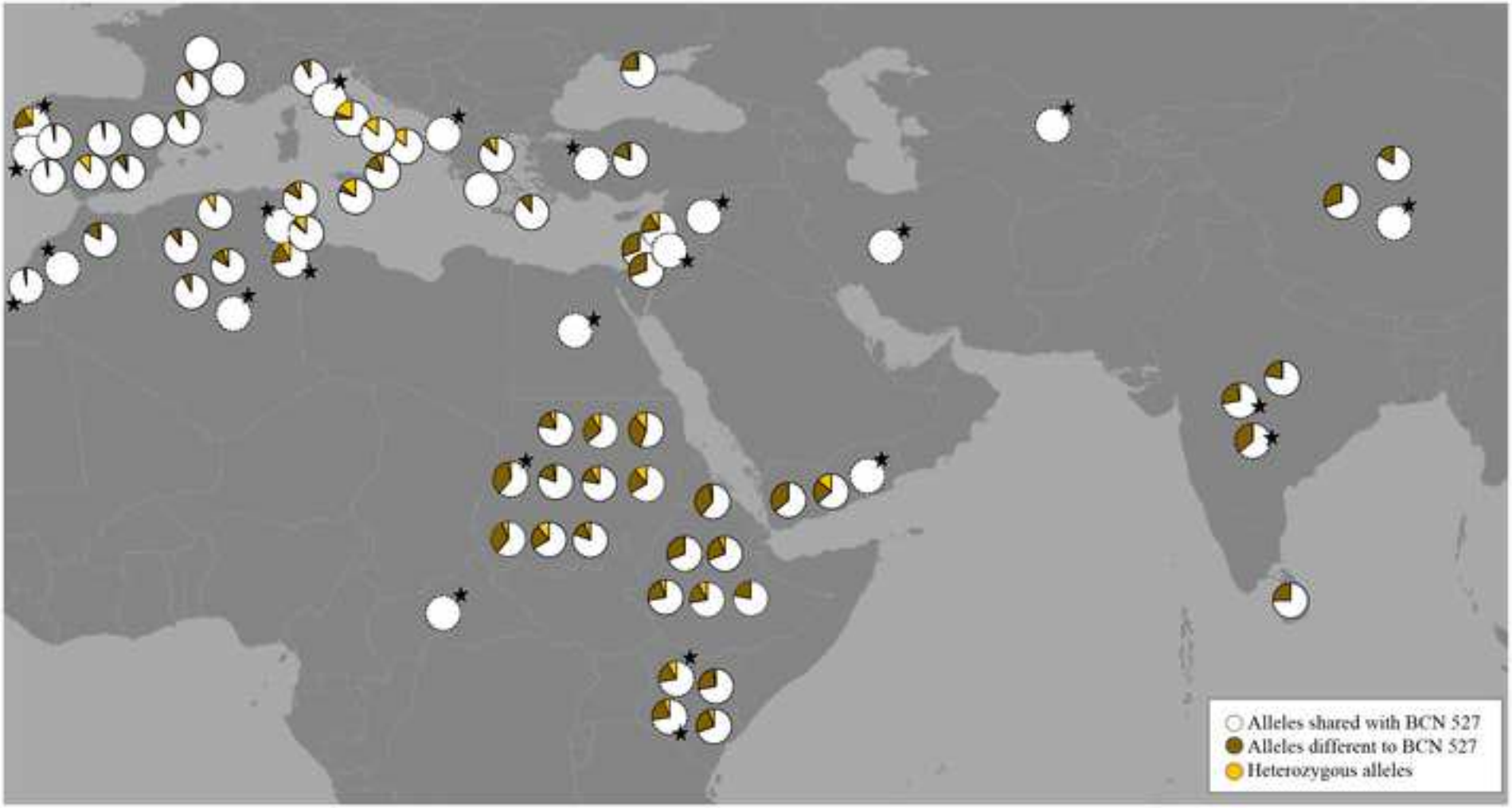
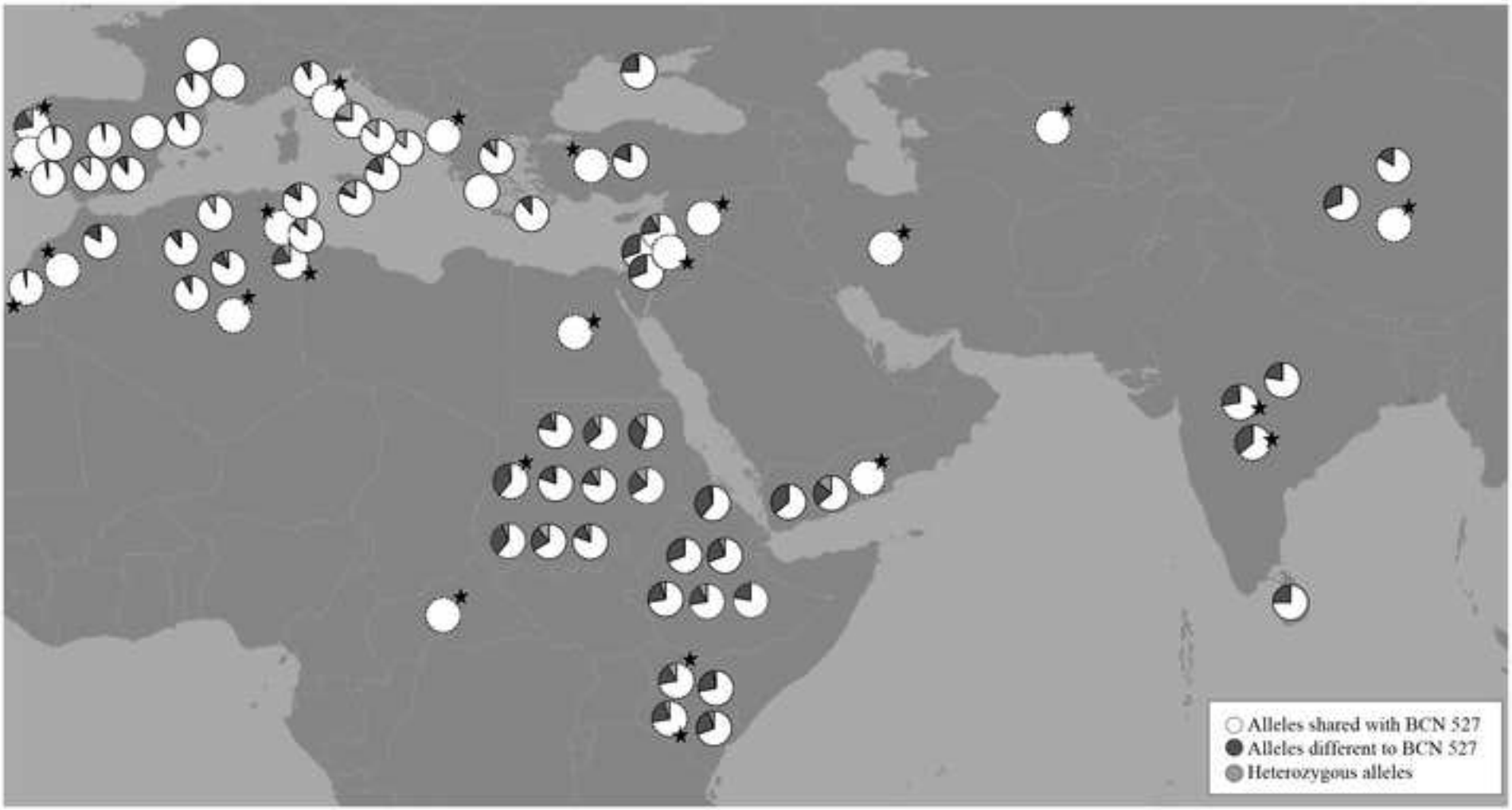


Figure 5 black and white

Click here to access/download;Figure;IJPara20_068_R2_Fig_5_BW_CMunoz.tif



Supplementary Table S1

Supplementary Table S2

Supplementary Table S3

Supplementary Table S4

Supplementary Table S5

Supplementary Figure S1

Click here to access/download

Multi-media supplement

IJPara20_068 Supplementary Figure S1.pptx