Genomic evidence for co-evolution and sporadic host shifts in leishmaniaviruses

Senne Heeren[1,2,3,*], Lilian Motta Cantanhêde[4,5], Khaled Chourabi[4,5], Mayara Cristhine de Oliveira Santana[4,5], Joon Klaps[2], Alexei Y. Kostygov[6,7], Vyacheslav Yurchenko[6], Philippe Lemey[2], Jean-Claude Dujardin[1,3], Frederik Van den Broeck[1,2,*], Elisa Cupolillo[4,5,*]

[1] Department of Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
[2] Department of Microbiology, Immunology and Transplantation, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium
[3] Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium
[4] Leishmaniasis Research Laboratory, Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro 21040360, Brazil
[5] Instituto Nacional de Ciência e Tecnologia de Epidemiologia da Amazônia Ocidental, INCT EpiAmO, Porto Velho 76812100, Brazil
[6] Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava, Czechia
[7] Zoological Institute of Russian Academy of Sciences, St. Petersburg, Russia

*** Correspondence**: Senne Heeren (senne.heeren@kuleuven.be); Frederik Van den Broeck (fvandenbroeck@gmail.com); Elisa Cupolillo (elisa.cupolillo@ioc.fiocruz.br)

FVDB and EC contributed equally as senior author of this manuscript.

Keywords: Co-evolution; Leishmania; Leishmaniavirus; co-phylogenetics

Abstract

The evolutionary relationships between hosts and their symbionts offer valuable insights into the origins, maintenance, and consequences of biological interactions. While co-divergence and host-switching have been extensively explored in systems, such as *Wolbachia*–arthropod symbioses or viruses infecting vertebrates, similar investigations in protistan parasites remain scarce. The *Leishmania*–*Leishmaniavirus* (LRV) association offers a rare opportunity to study co-divergence in a medically relevant symbiotic system, in which the virus modulates the parasitic disease severity in humans. Here, we used total RNA sequencing to capture both the *Leishmania* transcriptome and the LRV genome simultaneously, enabling the first comprehensive investigation of the co-evolutionary history of *Leishmania* and LRV across different hierarchical levels (subgenus, species and population). We found significant positive correlations between the parasite and viral genetic distances at both the subgenus (R^2 = 0.89; F = 10,913.49; $p < 0.001$) and species level (focusing on *L.* (*Viannia*) spp.; R^2 = 0.59; F = 1,254; $p < 0.001$). This was corroborated by additional co-phylogenetic methods (global-fit and event-based), indicating a strong pattern of phylogenetic congruence between *Leishmania* and LRV up to the species level. Only at the population level, focusing on the interactions of *L.* (*V.*) *braziliensis* and *L.* (*V.*) *guyanensis* with LRV1, we found weaker co-phylogenetic signals accompanied by more instances of intraspecific host switching. Overall, our findings provide, for the first time, analytical evidence for the co-evolution between *Leishmania* and LRV, with co-speciation as the predominant process, while also shedding light on how such symbioses were maintained over larger time scales.

Introduction

Inter-species interactions are key to ecosystem functioning (e.g. pollination, dispersal, protection, parasitism, predation). In this regard, viruses also play an important role in our biosphere, where they often cause direct harm to their host's populations, sometimes leading to severe socio-economic crises on regional or global scales. Conversely, viruses may also affect populations indirectly through so-called nesting doll ("matryoshka") infections (1–6). In such cases, viruses have established persistent, sometimes mutualistic-like, relationships with organisms such as pathogenic parasites, where they may alter (and often exacerbate) the disease manifestation in humans. This may complicate treatment of the infection and ultimately hinder the disease control in endemic areas (6–9). Even though efforts are currently being made to assess the impact of such nesting-doll infections, little is known about the underlying evolutionary history between pathogenic parasites and their viruses.

To address this, we investigated the co-evolutionary history of an established virus–parasite symbiotic system comprising *Leishmania* parasites (Trypanosomatidae) and their persistent double-stranded (ds) RNA viruses of the genus *Leishmaniavirus* (LRV) (*Pseudototiviridae*, *Duplornaviricota*) (10, 11). In this system, LRV has been found to mutualistically aid its parasitic host by modulating the innate immune response of the mammalian host through an interaction with Toll-like receptor 3 (7), which, in turn, ensures the inhibition of the Nod-like receptor protein 3 inflammasome in the macrophages (12). This immunological cascade ultimately leads to increased parasite survival, disease progression, exacerbation, and an increased likelihood of treatment failure (8, 9, 12, 13). In exchange, the virus' own metabolism and replication are maintained through the help of several host-encoded proteins (e.g., kinases and proteases) (14), but the exact mechanisms involved remain unknown.

In the latest release by the International Committee on the Taxonomy of Viruses (ICTV), the genus *Leishmaniavirus* comprises six species, which are specifically associated with trypanosomatid parasites, including *Leishmania* spp. and *Blechomonas* spp. (2, 10, 11). Within *Leishmania*, different LRV species have different host ranges: *Leishmaniavirus ichi* (further refered to as "LRV1") associated with *Leishmania* (*Viannia*) spp.; *Leishmaniavirus ni* (further referred to as "LRV2") in part associated with *Leishmania* (*Leishmania*) spp. (except *L.* (*L.*) *aethiopica*) and with *Leishmania* (*Sauroleishmania*) spp.; and *Leishmaniavirus sani* associated with *L.* (*L.*) *aethiopica* (2, 11). Among the LRV species of *Blechomonas*, there is *Leishmaniavirus shi* that is associated with *B. maslovi*; *Leishmaniavirus go* with *B. wendygibsoni*; and *Leishmaniavirus roku* with *B. juanalfonzi* (11).

The current LRV classification, along with reports on the virus' host-specificity (2, 15) and the relatively low sequence identity between different LRV species (< 40% at the nucleotide level between LRV1 and LRV2; (10)), resulted in the general assumption of a long standing co-evolutionary history between the parasites and the viruses. This assumption implies that the LRV acquisition pre-dated the *Leishmania* divergence into different subgenera (at least involving *L.* (*Viannia*), *L.* (*Leishmania*), and *L.* (*Sauroleishmania*)). Widmer and Dooley (1995) were the first to propose this hypothesis based on an analysis of genetic distances between nucleotide sequences, which highlighted the mirroring patterns of the LRV1-LRV2 sequence divergence and the heterogeneity between Old and New World *Leishmania* spp. (16). This view of "strict" co-evolution, established through vertical transmission, has been contested several times by reports of horizontal transfers of LRV2 across Old World *Leishmania* spp., and between *Leishmania* and *Blechomonas* (2, 17–20). Conversely, this appears to be in stark contrast with the situation in Central and South America, where interspecific host switches between LRV1 and *Leishmania* (*Viannia*) spp. seem to

occur less frequently (10, 15, 21, 22). At the intraspecific level however, LRV1 host switching between genetically distinct populations does generally appear to occur at a more frequent pace (3, 23).

While numerous reports and hypotheses regarding the co-evolutionary history of *Leishmania* and LRV exist in the literature (1, 10, 16, 23, 24), a comprehensive evaluation using (co)-phylogenetic methods is lacking. This is because most studies were focused on phylogenetic analyses of either the virus or the parasite separately; moreover, they were often based on several genes or short genomic fragments rather than looking at a genome-wide perspective. Our study addresses this gap, for the first time, by conducting an in-depth assessment of the virus-parasite co-evolutionary history across different hierarchical levels (subgenus, species, and population). The analysis integrates total RNA sequencing of 66 LRV1-positive *Leishmania (Viannia)* isolates accompanied by the whole genome sequencing of six LRV2-positive *L. (Leishmania)* and *L. (Sauroleishmania)* isolates, and 77 publicly available genomes of *L. (Viannia)* spp. (n = 49), LRV1 (n = 22) and LRV2 (n = 6).

Based on the current knowledge of the LRV-*Leishmania* association, we hypothesized to find strong patterns of phylogenetic congruence (i.e. mirroring host-symbiont phylogenies; (25)) at the subgeneric level (i.e., among *L.* (*Viannia*), *L.* (*Leishmania*) and *L.* (*Sauroleishmania*)) and species level (i.e., within *L.* (*Viannia*)). We also expected a more general co-phylogenetic signal (i.e. more closely related hosts tending to associate with more closely related symbionts; (25)) at the population level of a single species. The latter would be characterized by a relatively higher number of host transfers. Broadening our understanding of the (co-)evolutionary history of this epidemiologically important symbiosis will shed new light on how this mutualism has stood the test of time, while also providing insights into the interplay between viruses and (pathogenic) protists in general.

Results

Expanding the known genomic diversity of LRV1 in South America

Starting from 159 *Leishmania* (*Viannia*) spp. isolates, PCR-screening revealed 66 LRV1-positive isolates (Tab. 1; S. Fig. 1; S. Tab. 1). These were subjected to total RNA sequencing to extract both the viral genomes as well as the parasite transcriptomes. Viral genomes were reconstructed based on a de novo assembly approach as implemented in a novel genome reconstruction pipeline for viral sequencing data, Viralmetagenome (26). With this pipeline we were able to reconstruct viral genomes from 49 isolates of which 42 were of high-quality for subsequent (co)-phylogenetic analyses (S. Tab. 2). These 42 isolates originated from four Amazonian states of Brazil (Acre: n = 4; Amazonas: n = 21; Pará: n = 6; Rondônia: n = 10) and one Venezuelan state (Trujillo: n = 1) which were sampled between 1985 and 2017. The remaining seven isolates were not retained for subsequent analyses as these reconstructed genomes were either too short in length (IOCL2398, IOCL3007, IOCL3515, IOCL3654), or harbored a higher number of SNPs when the sequencing reads were remapped to their respective assemblies (IOCL3562, IOCl3563, IOCL3567) (S. Tab. 2). For one isolate in particular, IOCL3567, the pipeline yielded three genomes of 4,667 – 5,185 bp with a median coverage of 370x – 372x, but with a high number of SNPs (>600). This suggests infection by at least two LRV1 variants, a phenomenon previously documented in *Leishmania* carrying that virus (3, 27). However, we were unable to further disentangle these chimeric-like assemblies and discarded this isolate from downstream analyses. The same was done for another 17 isolates for which the assembly pipeline did not yield properly reconstructed LRV1 genomes.

The 42 high-quality assemblies covered 67.5% to 99.5% of the LRV1-4 genome (5,284 bp; SRA NC_003601) serving as a reference. The percentage of sequencing reads mapping back to the assemblies ranged from 0.0002% to 0.027% (median = 0.004%). The median depths of coverage for the assemblies were between 101x and 1,093x, with an overall median of 391x (S. Tab. 2). The analysis of variants revealed a median of 2 heterozygous SNPs (average = 8; min = 0; max = 84) and no homozygous SNPs in any of the samples, demonstrating the reliability of the reconstructed genomes. All 42 genomes resembled the typical arrangement of LRV1, with two overlapping open reading frames coding for a capsid protein (CP; 2,999 bp) and an RNA-dependent RNA polymerase (RDRP; 2,637 bp) which were near-completely covered in most cases (S. Tab. 2).

Phylogenetic reconstruction and calculation of pairwise genetic distances between the 42 assembled genomes revealed 14 highly supported viral clades (Fig. 1a) exhibiting relatively low, within-clade, genetic distances (<0.125 substitutions/site; S. Tab. 3). The clades were named according to their host species with Roman numerals (i.e., LgI-V, LbI-V, LsI, LnI-II) (Fig. 1). Analysis of geographic distribution revealed that, among the studied isolates, those from the Amazonas belonged to eight clades, three of which being unique to this location (LgIII, LnI, LsI) (Fig. 1b). Other clades found in a single state included LbII and LbIV in Acre; LbI and LbIII in Rondônia; and LgVI and LnII in Pará. The remaining clades were found in two or three different neighboring states. The single LRV1 genome of Venezuelan origin belonged to the LgII-clade, which was mainly identified in Amazonas and Acre. However, the current sample size impeded our ability to infer the connection of *Leishmania* and LRV1 between Trujillo and the Western Brazilian states.

Our new panel of LRV1 genomes provides a major expansion of the known genomic diversity of the virus. This becomes clear from the phylogenetic reconstruction, combining our panel with 64 publicly available genomes. Most clades characterized above had not been identified earlier, whereas a few (LgII, LgV, LgVI, LbII, LbV) clustered with previously described ones (3, 27)

158 (Fig. 1c). The concomitant inference of the ancestral geographic origins (Fig. 1c, internal nodes)
159 along with the clear host specificity of LRV1 towards the different *Leishmania* species suggests
160 spatial divergence and subsequent spread of each species-specific clade. One exception to this
161 observed host-specificity is LRV1-Ln-IOCL3574, which clusters with viruses associated with *L.*
162 *guyanensis* (Fig. 1c). The corresponding *Leishmania* isolate was originally typed as *L. naiffi*, but
163 our whole transcriptome data indicated that it is a mixture both genetic material of *L. naiffi* and *L.*
164 *guyanensis* (see below).

165 Phylogenetic reconstructions based on the single-gene alignments (i.e. CP and RDRP)
166 were similar compared to the clade delimitation derived from the whole-genome data (Fig. 1b; S.
167 Fig. 2). Moreover, co-phylogenetic global-fit tests (Robinson-Foulds distance, ParaFit and PACo)
168 revealed a consistently strong co-phylogenetic signal between the different LRV1 phylogenies (i.e.
169 genome-wide, CP and RDRP) (S. Tab. 5; S. Fig. 2). The only striking inconsistency was the internal
170 position of *L. guyanensis*-associated LRV1 from French Guiana (clades: LgV, LgVI, LsI, A*, B*, D*,
171 E* (except M92355.1, origin: Suriname)). Based on both the CP and genome-wide phylogenies,
172 this clade (S. Fig. 2, clade shown in red) appears to share a common ancestor with the other *L.*
173 *guyanensis*-associated LRV1 genomes, whereas it formed a monophyletic group with the *L.*
174 *braziliensis*-associated LRV1 within the RDRP phylogeny. Nevertheless, this inconsistency does
175 not refute the observed LRV1 host-specificity, it merely indicates the uncertainty of the clade
176 divergence (i.e., low bootstrap supports) at these ancestral nodes based on the two single gene
177 alignments (S. Fig. 2). In addition, none of the three alignments showed any indication of
178 recombination, based on the pairwise homoplasy index test (whole genome: *p*-value = 1.0; CP: *p*-
179 value = 0.95; RDRP: *p*-value = 1.0).

180 **Tab. 1: LRV1 prevalence across all 159 screened *Leishmania* (*Viannia*) spp.**

Species	LRV1 prevalence
L. braziliensis	27.85% (22/79)
L. guyanensis	75.61% (31/41)
L. lainsoni	16.67% (1/6)
L. lindenbergi	0% (0/6)
L. naiffi	61.11% (11/18)
L. panamensis	0% (0/4)
L. shawi	33.33% (1/3)
L. utingensis	0% (0/1)
L. (*Viannia*) *spp.*	0% (0/1)
Total	**41.51% (66/159)**

181

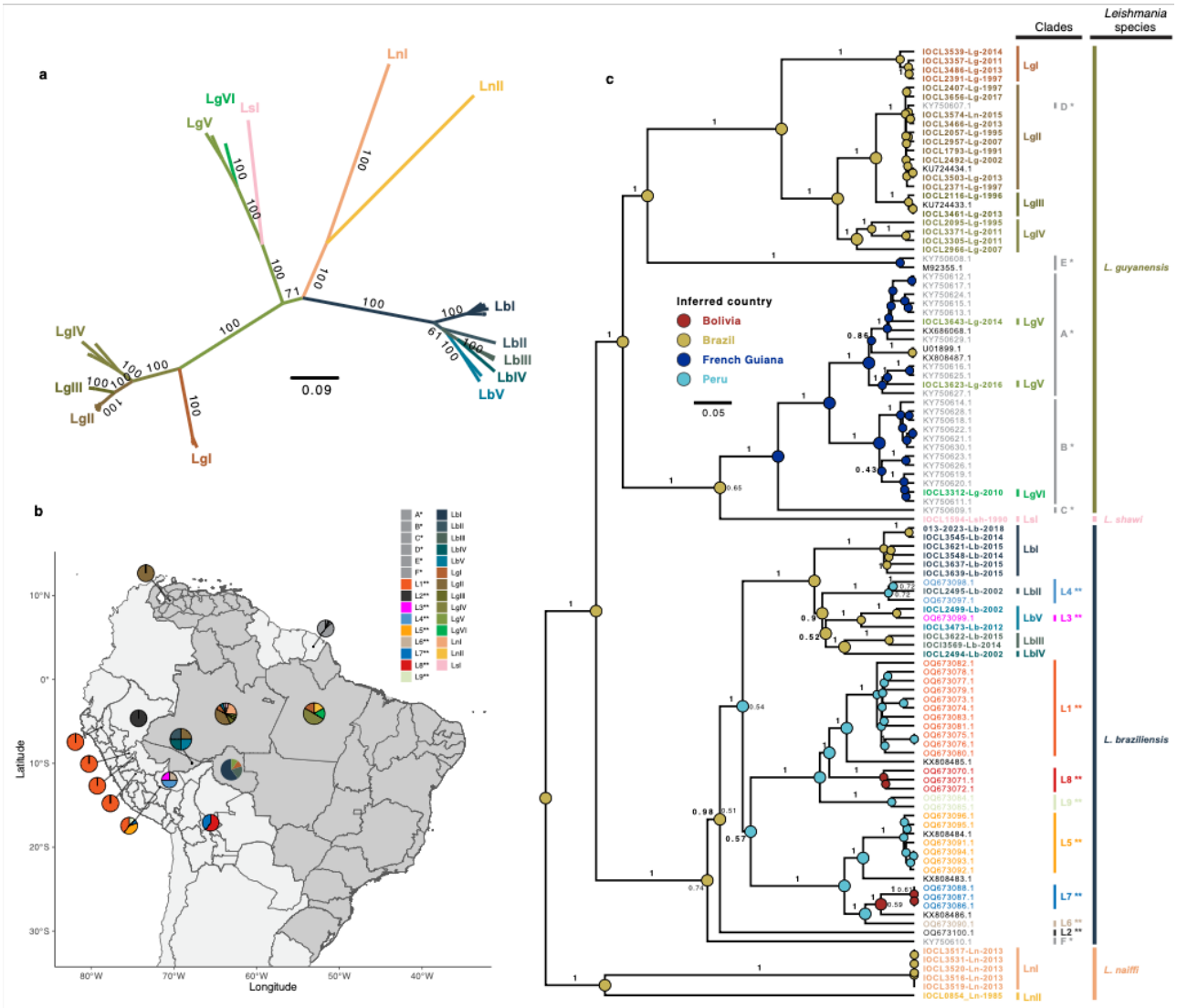


Fig. 1: Lineage diversity and distribution of the LRV1 genomic diversity. a Unrooted maximum likelihood phylogenetic tree based on 42 near-whole genome LRV1 sequences. Branch values represent bootstrap support values based on 10,000 ultra-fast approximated bootstrap replicates. The scale bar depicts the number of substitutions per site. Clades were delineated based on high bootstrap support and low (within-clade) pairwise genetic distances (S. Tab. 3). Colors are set for visual purposes and do not serve any information on the ancestral states (i.e. internal nodes). The abbreviations Lb, Lg, Ln and Ls refer to the *Leishmania* hosts of the LRVs: *L. braziliensis*, *L. guyanensis*, *L. naiffi* and *L. shawi*, respectively. **b** Administrative map of South America showing the geographic origin of the 42 LRV1 genomes (dark gray area) along with geolocalized LRV1 genomes reported previously (3, 27). Administrative country- and state-level data originate from: http://www.diva-gis.org/Data. **c** Midpoint rooted maximum clade credibility tree based on the 42 newly obtained near-whole LRV1 genomes along with 64 publicly available ones. Branch support values (bold faced) represent the posterior probabilities. Nodes are colored according to their inferred geographic origin, with posterior probabilities shown adjacently in regular font. Tip labels are colored according to the viral lineages (from this study or previous ones). Bars beside the tree represent the newly or previously described viral lineages (matching the tip labels) and the associated *Leishmania* species. Single and double asterisks mark the lineages identified by Tirera et al. (2017) (27) and Heeren et al. (2023) (3), respectively.

198 **Transcriptomic diversity of the *Leishmania* (*Viannia*) subgenus**

199 Total RNA sequencing of the LRV1-positive *Leishmania* isolates allowed us, not only to
200 assemble LRV1 genomes, but also to genotype the *Leishmania* isolates themselves based on the
201 transcriptomic data. Similar to our previous whole-genome sequencing studies (28, 29), we
202 compiled a dataset of 115 isolates containing: (i) genomic data from 66 isolates, generated for this
203 work; and (ii) 49 publicly available sequences from various *Leishmania* spp. that served as
204 reference strains during genotyping and species delimitation. This reference panel contained
205 isolates from the *Leishmania* (*Viannia*) *braziliensis* clade (*Lb* L1: n = 18; *Lb* L2: n = 3; *Lb* L3: n = 2;
206 *Lb* L4: n = 2; *L. peruviana*: n = 2), the *Leishmania* (*V.*) *guyanensis* species complex (*L. guyanensis*:
207 n = 3; *L. panamensis*: n = 14; *L. shawi*: n = 1), *Leishmania* (*V.*) *naiffi* (n = 3), and *Leishmania* (*V.*)
208 *lainsoni* (n = 1) (S. Tab. 6) (28).

209 As our data was derived from RNA sequencing, our genotyping analyses focused on the
210 coding DNA sequence (CDS) regions. Mapping the reads against the long-read-based M2904 *L.*
211 *braziliensis* reference genome (30) resulted in a median coverage of 17x (min = 1x; max = 151x)
212 across the entire CDS for all 115 isolates. For the isolates from this study alone, the median CDS
213 coverage was 10.5x (min = 1x; max = 43x). Considering a minimum CDS read depth of 10x, we
214 excluded 31 isolates (30 from this study and one from a previous study) from downstream analyses
215 and filtered out 1,184 individual CDS regions below this threshold in the retained sampels. The final
216 genotype dataset comprised 6,299 CDS regions across 84 isolates, yielding 157,649 high-quality
217 SNPs.

218 Based on the differences in the number of SNPs (S. Tab. 8) between the included
219 *Leishmania* (*Viannia*) spp. (Chi-squared = 76.92; df = 10; p = 2.01e-12; pairwise Dunn tests: S.
220 Tab. 9), all isolates, with the exception of IOCL3539 and IOCL3574, could be identified (*L.*
221 *braziliensis* L1: n = 12; *L. guyanensis*: n = 18; *L. naiffi*: n = 4) (Fig. 2a). The latter two isolates
222 harbored distinctively higher numbers of heterozygous SNPs compared to all other isolates (Fig.
223 2a; S. Tab. 8) and showed reticulation in the CDS-wide phylogenetic network, a pattern suggestive
224 of recombination and potential hybrid ancestry Fig. 2b.

225 However, the alternate allele read depth frequency (ARDF) of these two isolates revealed
226 uncommon profiles. In IOCL3539, the ARDF distribution was unimodal and skewed with a mode
227 around 0.66, compared to the typical symmetric distribution around 0.5 (31) (Fig. 2c). IOCL3574,
228 on the other hand, showed a distribution that was bimodal with modes near 0.2 and 0.8 (in contrast
229 to the typical modes for triploid genomes at 0.33 and 0.66 (31)) (Fig. 2d). These CDS-wide shifts
230 in the ARDF are strong indicators that the observed hybrid-like patterns are the result of either (i)
231 extensive mosaic aneuploidy, or (ii) mixed infections by two different *Leishmania* spp., rather than
232 true interspecific hybrids (32).
233

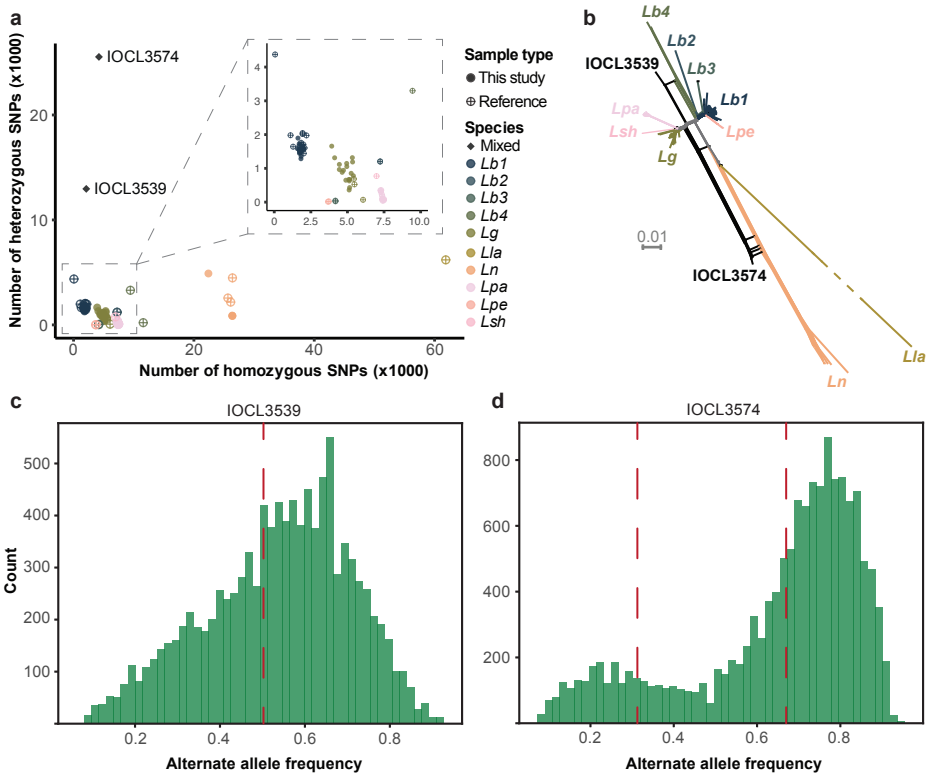


Fig. 2: Species/lineage delimitation of 36 *Leishmania* isolates based on whole transcriptome data (157,649 bi-allelic SNPs). a Quantification of homozygous (x-axis) and heterozygous (y-axis) SNPs for 36 isolates in this study and 48 reference isolates (incl. the *L. braziliensis* clade, the *L. guyanensis* species complex, *L. naiffi* and *L. lainsoni*). **b** Phylogenetic network of all 84 isolates revealing the reticulate pattern caused by IOCL3539 and IOCL3574, indicative of hybrid ancestry. **c-d** Shifted CDS-wide alternate allele frequency distribution plots of IOCL3539 (**c**) and IOCL3574 (**d**). Dashed vertical lines show the typical modalities for diploid (i.e. a unimodal distribution around 0.5) and triploid (i.e. a bimodal distribution around 0.33 and 0.66) genomes (31).

244 **Evidence for phylogenetic congruence across different hierarchical levels of *Leishmania***
245 **and LRV**

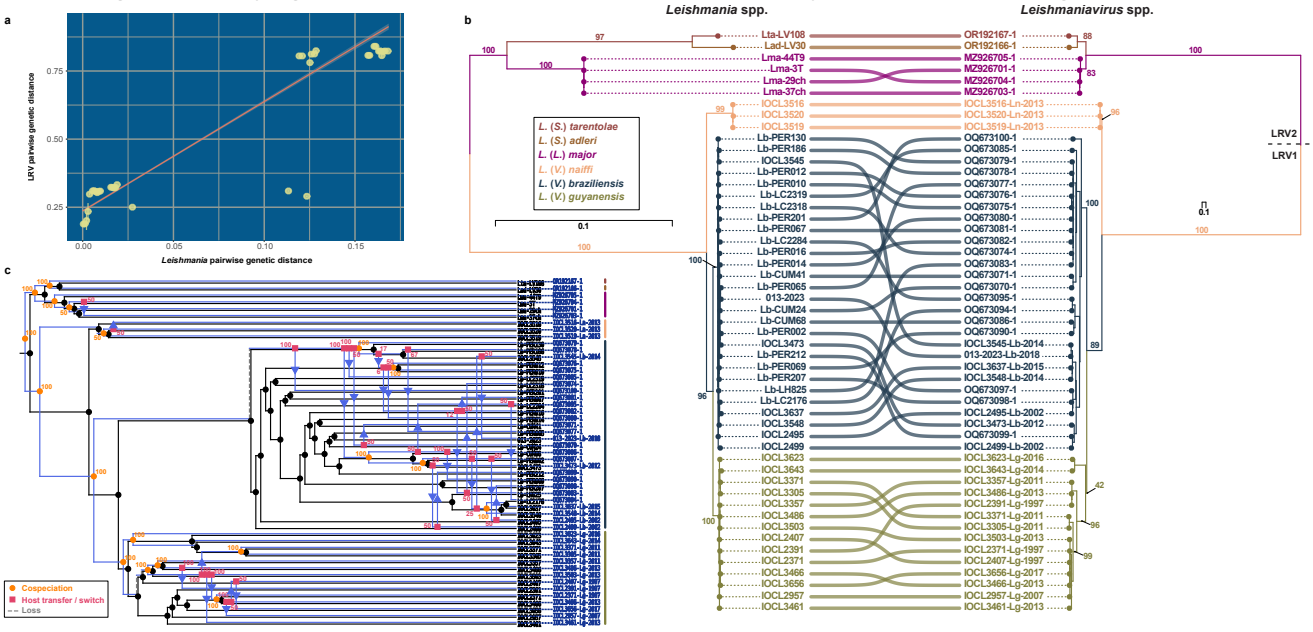
246 To assess the extent to which *Leishmania* and LRV have shared a similar evolutionary
247 trajectory, we performed a matrix regression on the pairwise genetic distances of both symbionts,
248 combined with formal co-phylogenetic tests – global-fit and event-based – to detect overall co-
249 phylogenetic signal and evaluate true phylogenetic congruence, respectively (25). These analyses
250 were based on a dataset including 22 Peruvian *Leishmania* (*V.*) *braziliensis* isolates with their
251 associated LRV1 genomes (3), as well as six Old World *Leishmania* isolates (*L.* (*L.*) *major* (n = 4),
252 *L.* (*S.*) *tarentolae* (n = 1) and *L.* (*S.*) *adleri* (n = 1)) with their respective LRV2 genomes (S. Tab.
253 4,6).

254 On the subgeneric level, we observed a highly significant, strong positive correlation
255 between pairwise genetic distances across the 51 *Leishmania*-LRV pairs that were analyzed
256 (MMRR: R^2 = 0.89; F = 10,913.49; p < 0.001) (Fig. 3a). In parallel, our array of global-fit analyses
257 detected a consistently significant co-phylogenetic signal between *Leishmania* and LRV at this level
258 (Fig. 3b; S. Tab. 10). Similarly, the event-based inference revealed strong signatures of
259 phylogenetic congruence, suggesting that the two symbionts share a highly similar phylogenetic
260 history (Fig. 3c). This latter analysis was conducted using nine different event-costs sets (S. Fig.
261 3a; S. Tab. 11), with eight of them yielding significant co-phylogenetic reconciliations, where well-
262 supported cospeciation events predominated the divergence of Old and New World *Leishmania*
263 subgenera with LRV2 and LRV1, respectively (S. Fig. 4).

264 When focusing on the species level (subgenus *L.* (*Viannia*)) and its association with LRV1,
265 we observed a highly significant, yet more moderate positive correlation between the parasite and
266 viral pairwise genetic distances (MMRR: R^2 = 0.59; F = 1,254; p < 0.001). Interestingly, the
267 pronounced host specificity did not appear to be shaped by spatial separation, as viral genetic
268 distances did not correlate with geographic distances among the isolates (MMRR: R^2 = 0.005; F =
269 5.28; p = 0.135). In parallel, both the global-fit and event-based methods consistently detected
270 strong patterns of co-phylogenetic signal and phylogenetic congruence, respectively (S. Tab.
271 10,12; S. Fig. 5). More specifically, the event-based analysis yielded a diverse set of possible co-
272 phylogenetic histories based under six different event-cost scenarios (S. Fig. 3b; S. Tab. 12).
273 Strikingly, in all cases the inferred phylogenetic reconciliations were significant, and thus supporting
274 the presence of strong phylogenetic congruence between the members of the subgenus *Viannia*
275 and LRV1 (S. Fig. 6). Another notable observation was that co-speciation events consistently
276 accounted for species divergence (S. Tab. 12) and were well supported, whereas duplication and
277 transfer events appeared to more predominant within the separate species partitions (S. Fig. 6).

278 Finally, zooming in towards the population level of this virus-parasite co-evolutionary
279 history, we focused on *L. (V.) braziliensis* and *L. (V.) guyanensis* with their respective LRV1
280 lineages (too small sample size for *L.* (*V.*) *naiffi*). This also revealed signatures of co-evolution,
281 albeit generally weaker compared to the higher levels and occasionally lacking statistical
282 significance. For instance, we found no significant correlation between the genetic distances of *L.*
283 *braziliensis* and LRV1 (MMRR: R^2 = 0.027; F = 10.65; p = 0.215) whilst a weak, though significant,
284 correlation was observed in the case of *L. guyanensis* (MMRR: R^2 = 0.349; F = 47.68; p = 0.002).
285 Surprisingly, both global-fit and event-based analyses revealed significant co-phylogenetic signals
286 and evidence of phylogenetic congruence between each *Leishmania* species and its associated
287 LRV1 lineage (S. Fig. 3c-d, 7-9; S. Tab. 10, 13).

Fig. 3: Co-phylogenetic reconciliation of *Leishmania*-LRV symbiosis at the subgeneric level. a



Regression scatterplot of the pairwise genetic distances of *Leishmania* and LRV (51 pairs). Yellow points represent summarized values in 200 bins along the x-axis. **b** Tanglegram of *Leishmania* and LRV phylogenies. The *Leishmania* phylogeny was inferred from six concatenated gene sequences (S. Tab. 15; see methods). The LRV phylogeny was inferred from concatenated sequences of CP and RDRP (ORF2 and ORF3, respectively). Colors on the tree are set for visual purposes and do not serve any information on the ancestral states. Branch values represent bootstrap support values based on 10,000 ultra-fast approximated bootstrap replicates. The scale bar indicates the number of substitutions per site. **c** Event-based *Leishmania*-LRV co-phylogeny at the subgeneric level (51 tip-pairs). The reconciliation was inferred with the following event-costs: duplication = 1.50; transfer = 1.50; loss = 1.00 (S. Tab. 11; set 03). Support values for the inferred co-evolution events (e.g. co-speciation, host transfer) were estimated from 100 bootstrap replicates. Colored bars adjacent to the tip labels denote the LRV-host species as depicted in panel b.

Discussion

By harnessing total RNA sequencing of cultivated LRV-positive *Leishmania* isolates – enabling the capture of both the parasite transcriptome and (near)-complete viral genome – our study provides the first comprehensive evaluation of the virus-parasite (co)-phylogenetic history of this medically relevant symbiosis across the subgenus, species and population levels.

We first investigated the genomic diversity of each symbiont separately to position the newly sampled genomes within the known phylogenetic context of *Leishmania* and LRV, based on publicly available genomic data. For *Leishmania*, all but two isolates (IOCL3539, IOCL3574) clustered clearly with a single species – either the *L.* (*V.*) *braziliensis* clade, the *L.* (*V.*) *guyanensis* species complex, or *L.* (*V.*) *naiffi* – as judged by homozygous SNP counts and unambiguous phylogenetic placement. Conversely, the reticulate pattern in the phylogenetic network and high heterozygosity in IOCL3539 and IOCL3574 suggested either (i) interspecific hybridization (*L. guyanensis* x *L. braziliensis* L4 and *L. guyanensis* x *L. naiffi*, respectively) or (ii) a mixture of these species. Hybridization (both inter- and intraspecific) is an increasingly recognized phenomenon in parasitic protozoa (33–35) and *Leishmania* in particular (3, 28–30, 36–39). This has important epidemiological implications, making detection and reporting of new hybrids crucial within the context of genomic surveillance (3, 33, 40). Analysis of the alternate allele read depth frequencies at heterozygous sites revealed shifts from the expected uni- and bimodal patterns suggesting that the genomic signatures of these isolates are more likely the result of a mixed infection or extensive mosaic aneuploidy, rather than being true hybrid signals (31, 32). However, these isolates apeared to be devoid of distinct homozygous regions for the alternate allele and showed heterozygous streches across the entire genome – contrasting with other recently reported hybrid signatures (28, 32) – which strongly suggests that both isolates are mixtures of two *Leishmania* species rather than true genomic hybrids.

The phylogenetic reconstruction of the newly assembled LRV1 genomes substantially expanded the known genomic diversity of the virus. This included (i) an increased intra-clade diversity within groups previously described, based on whole-genome data (3, 27); and (ii) the identification of novel clades (e.g., LgI, LgIII, LgIV, LbI, LbIII, LbIV, LsI, LnI, LnII). In concordance with Santana et al. (2023) (15), we observed strict host specificity among LRV1 genomes, independent of sampling location (state/department). This was evidenced by a strong correlation of virus genetic distances with those of their host, but not with geographic distances. Such clear host specificity is already a good indicator of virus-parasite co-divergence at the species level within the *Leishmania* (*Viannia*) subgenus, with the only exceptions being the likely mixed isolates.

Although earlier studies have suggested a long-standing association between *Leishmania* and LRV, potentially involving co-divergence between and within subgenera (1, 10, 16, 23, 24), these hypotheses have not been formally tested. Most previous work focused on separate phylogenetic analyses of virus and host, often based on a few genes. Each *Leishmania* subgenus – except *L.* (*Mundinia*), which harbors a phylogenetically remote *Shilevirus* (*Leishbuviridae*, *Negarnaviricota*) (41, 42) – has its own LRV species. For LRV1, a strong pattern of host specificity and phylogenetic congruence has been observed. In contrast, LRV2 shows a more flexible pattern, with partial specificity: certain viral clusters are more frequently associated with particular *Leishmania* species, but not in a strictly co-specific manner.

345 The main goal of our study was to formally investigate, for the first time, the co-evolutionary
346 history of *Leishmania* spp. and their endosymbiotic viruses across the subgeneric, species and
347 population level. Overall, we revealed clear genomic evidence of virus-parasite co-divergence
348 across these levels. At the population level of *L.* (*V.*) *braziliensis* and *L.* (*V.*) *guyanensis*,
349 phylogenetic congruences decreased (e.g. due to lower signals or loss of statistical significance).
350 Here, we observed more instances of host-switching, as reported previously in *L. braziliensis* (3),
351 relative to the predominance of co-speciation events at the higher hierarchical levels. These results
352 align well with the seemingly mutualistic nature of the virus-parasite symbiosis, where strong
353 phylogenetic congruence has often been observed (43). Conversely, this view of virus-parasite co-
354 divergence is challenged by cases of horizontal transmission of LRV within *Leishmania* (both inter-
355 and intraspecific) (3, 18–23) and to other trypanosomatids, such as *Blechomonas* spp. (17). These
356 horizontal transfers are likely facilitated by the potential of LRV to exploit the parasite's exosomal
357 pathway (44–47). Nevertheless, our findings indicate that host switching is relatively rare and
358 occurs mainly at the intraspecific level, whereas vertical transmission remains the predominant
359 factor shaping the evolutionary trajectory of this symbiosis (2, 10).

360 This study expands our current understanding of the co-evolutionary dynamics of
361 *Leishmania* and LRV, even though it does not elucidate their full evolutionary history. A major
362 limitation is the discrepancy between the amount of paired genomic data between Old and New
363 World regions. Although our analyses revealed strong signatures of phylogenetic congruence at
364 the subgeneric level, these patterns warrant validation by means of larger sample sizes of Old
365 World isolates. Another unresolved question concerns the absence of LRV in the divergent
366 subgenus *Mundinia,* where it is still unknown whether LRV-presence is a plesio- (ancestral) or
367 apomorphic (derived) trait. In other words, LRV could have been either lost in *L.* (*Mundinia*) or
368 gained after its divergence from the most recent common ancestor of the other three subgenera
369 (*Leishmania, Viannia*, and *Sauroleishmania*) (10). Elucidating the broader virus-parasite interplay
370 is also becoming increasingly complex as more LRV species are described and recognized. One
371 additional viral species is known within *Leishmania*, *Leishmaniavirus sani* which is associated to *L.*
372 (*L.*) *aethiopica*, and three divergent species have been described from the flea-parasitic
373 trypanosomatid genus *Blechomonas*: *Leishmaniavirus shi* in *B. maslovi*, *Leishmaniavirus go* in *B.*
374 *wendygibsoni*, and *Leishmaniavirus roku* in *B. juanalfonzi* (11, 17). Moreover, LRV1 is not the sole
375 RNA virus infecting *L.* (*Viannia*) spp. (48). Accounting for this growing co-evolutionary and
376 taxonomic complexity, it is likely that additional, more cryptic, viral diversity remains undetected in
377 these protistan parasites. For each of these virus-parasite interactions, It will therefore be essential
378 to evaluate and monitor the consequences for the mammalian host (incl. humans) infected with
379 such virus-positive parasites.

380 Another point to consider is that even though our co-phylogenetic analyses focused on
381 LRV1-positive *Leishmania* (*Viannia*) isolates, including LRV1-negative strains could yield additional
382 insight into the host-virus evolutionary dynamics. The absence of LRV1 in certain parasite-lineages,
383 e.g., those from the Atlantic region of Brazil (49–51), may reflect historical events of viral loss,
384 resistance to infection, or ecological constraints limiting viral maintenance or transmission.
385 Incorporating these isolates into the host phylogeny could help determine whether LRV1 absence
386 is randomly distributed or associated with specific clades, thereby reinforcing or challenging the
387 current co-phylogenetic patterns. While these isolates cannot directly contribute to host–virus tree
388 congruence analyses due to the lack of viral sequence data, independent analyses of their
389 phylogenetic position and geographic distribution may shed light on the broader context of LRV1
390 evolution and its relationship to the diversification of *Leishmania*.

For this study, we applied a less conventional sequencing approach which proved to be a useful and cost-efficient way to simultanuously sequence the parasite transcriptome and the LRV genome. Even though the resulting phylogenetic patterns were consistent with those derived from standard genome sequencing datasets for both parasite and virus, our approach had some limitations. In particular, it was not always possible to fully reconstruct viral genomes due to the low fraction of LRV reads. Next to that, the transcriptome read depth from the parasite data was sometimes insufficient for making robust phylogenetic and population genomic inferences. These issues could be mitigated either by deeper sequencing, with a potentially unjustified additional cost, or reducing the ribosomal RNA content in the sequencing libraries. Nevertheless, for epidemiologically relevant systems such as *Leishmania*-LRV, a strategy that captures data for both symbionts in parallel represents a meaningful advancement for molecular surveilance.

In conclusion, our total RNA sequencing data allowed us to simultaneously perform phylogenetic analyses on *Leishmania* isolates, via transcriptome-based genotyping, and their respective LRV strains. This effort represents the first explicit reconstruction of the virus-parasite co-evolutionary history, characterized by strong *Leishmania*-LRV co-divergence between and within three parasite subgenera, with the most frequent host-switches occuring at the intraspecific level. Although the virus-parasite interplay has not yet been fully disentangled, this work provides fundamental biological insight in the evolution of the so-called matryoshka infections where – to quote Hartley et al. (2012) – ultimately the host will pay the toll (52).

411 ## Materials and Methods

412 **Parasite culturing & RNA sequencing**

413 *Leishmania* isolates were cultured in vitro, screened for LRV1, and prepared for total RNA
414 sequencing at the Fundação Oswaldo Cruz (Fiocruz). Parasites were grown in liquid culture
415 medium to a concentration of 10^7–10^8 cells/ml, collected through centrifugation (on the third day
416 after the final subculturing step), and preserved in DNA/RNA Shield (Zymo Research). Total RNA
417 was extracted using the PureLink RNA Mini Kit (Invitrogen), following the manufacturer's
418 instructions. All isolates were screened for the presence of LRV1 through PCR using distinct LRV1
419 primers: LRV F–5′- ATGCCTAAGAGTTTGGATTCG-3′ and LRV R–5′-
420 ACAACCAGACGATTGCTGTG-3′ (53). The screening included a positive (*L. guyanensis*
421 MHOM/BR/1975/M4147) and a negative control (*L. braziliensis* MHOM/BR/1975/M2903) (54–56).
422 LRV1-positive isolates (n = 66) were sent for total RNA sequencing without polyA selection or rRNA
423 depletion to make sure the LRV genome would not be lost in the sample. Samples were sequenced
424 on either the NovaSeq (n = 56) or DNBSeq (n = 10) platform at Genewiz (Azenta Life Sciences)
425 and BGI Genomics, respectively. The latter ten samples were part of a pilot study to test sequencing
426 approach for capturing both the LRV genome and the parasite transcriptome. In both cases, paired-
427 end sequences were generated to ~30,000,000 150bp reads.

428 ***Leishmaniavirus* de novo assembly & phylogenetic reconstruction**

429 Prior to the de novo assembly of viral genomes, raw reads from each of the 66 LRV-positive
430 *L. Viannia* sequenced libraries were filtered and trimmed using fastp v.0.23.4 (57), and mapped
431 against the *L. braziliensis* M2904 reference genome using SMALT v.0.7.6 (available at:
432 https://www.sanger.ac.uk/tool/smalt-0/). Only unmapped reads were retained and processed with
433 the novel bioinformatics workflow Viralmetagenome v.0.1.2 (available at: https://github.com/nf-
434 core/viralmetagenome) for de novo reconstructions of viral genomes (26). In summary, with this
435 workflow we combined Megahit and Spades assemblies, followed by a BLAST search to extract
436 contigs matching publicly available LRV1 genomes (58–60). Next, all "raw" assemblies were then
437 iteratively refined by re-mapping reads to the assembled genomes and subsequent variant and
438 consensus calling, making use of BCFtools and iVar (61, 62). The reconstruction of hybrid
439 consensus genomes was skipped (--skip_hybrid_consensus: true). A detailed overview of the
440 parameter settings is provided in a supplementary table (S. Tab. 14). Viralmetagenome was able
441 to resolve a consensus genome for 49 of the 66 LRV1-positive isolates. Additionally, we conducted
442 a post-hoc assessment of the assembly quality which involved re-mapping of the reads against
443 their respective assemblies, and variant calling. In total, 42 assemblies were retained based on the
444 following criteria: length > 3,000bp, read depth > 100x, and low SNP counts (i.e., <100
445 heterozygous and zero homozygous SNPs). Finally, whole-genome sequences, as well as
446 sequences for the Capsid Protein (CP) and the RNA-dependent RNA Polymerase (RDRP), were
447 aligned using the L-INS-I algorithm, as implemented in MAFFT v.7.490 ('--localpair' option) (63).

448 Phylogenetic reconstructions of LRV1 were generated using IQtree2 v2.3.6 (64) over
449 10,000 bootstrap replicates using the ultrafast bootstrap approximation method (65). Substitution
450 models were automaticlly selected by IQtree's ModelFinder algorithm (-m MFP) (66) which selects
451 the best substitution model based on the Bayesian Information Criterion (BIC) after evaluating all
452 models that are implemented in IQtree. Whole genome maximum likelihood phylogenies of the
453 assembled LRV1 genomes were generated with and without a set of 64 publicly available genomes
454 (S. Tab. 4) using GTR+F+I+R4 and GTR+F+I+G4 as respective substitution models. Similarly,
455 gene specific phylogenies, including the 64 publicly available sequences, were inferred using the

456 GTR+F+I+R4 substitution model for both CP and RDRP. For the clade delimitation of the 42 newly
457 assembled genomes, we also calculated the pairwise genetic distances using the 'dist.dna' function
458 of the R-package ape v5.8-1 (67) (model = 'raw').

459 In addition to regular phylogenetic reconstructions, we also attempted to reconstruct
460 ancestral states across the entire LRV1 phylogeny, focusing on the virus' geographic origin and
461 associated *Leishmania* species. This was achieved by adopting a Bayesian phylogenetic
462 framework making use of the BEAST software package (v.1.10.4) (68) and the BEAGLE likelihood
463 calculation library (69). Consistent with previous analyses (3), no temporal signal was detected in
464 the data. Therefore, all sampling dates were set to zero, after which LRV1's discrete
465 phylogeography was inferred without time calibration, using the default strict molecular clock
466 model). The nucleotide substitution model was set to GTR (General Time Reversible) with
467 estimated base frequencies and the site heterogeneity model set to a Gamma distribution with 4
468 discrete rate categories and accounting for invariable sites (70–72). The tree prior model was set
469 to a speciation model, the Yule process (73), starting the analysis with a random starting tree
470 (default setting). Two discrete traits were added: the country of origin and associated *Leishmania*
471 species. For both traits we selected the symmetric discrete trait substitution model along with the
472 BSSVS option (Bayesian Stochastic Search Variable Selection procedure) in order to reconstruct
473 their ancstral states. Prior and operator settings were kept at their default settings. The MCMC
474 chain was run for ten million states, sampled every 1,000 states and providing a 10% burn-in.
475 Throughout the inference proces, convergence of parameter estimates was evaluated using
476 BEAST's auxiliary program Tracer v.1.7.2 (74). Upon completion of the runs, all effective sample
477 sizes (ESS) of all parameters exceeded the recommended threshold of 200, except for the C-T
478 transition rate in the GTR model, which reached an ESS of 110. Finally, the posterior tree
479 distribution was summarized as a maximum clade credibility (MCC) tree using BEAST's
480 TreeAnnotator.

481 **Transcriptomic genotyping & phylogenetic reconstruction of *Leishmania* (*Viannia*) isolates**

482 Total RNA sequencing reads of the 66 LRV1-positive *Leishmania* (*Viannia*) isolates were
483 mapped against the *L. braziliensis* M2904 v.65 (MHOM/BR/75/M2904) reference genome
484 (available at: https://tritrypdb.org/common/downloads/release-65/LbraziliensisMHOMBR75M2904
485 _2019) using SMALT v0.7.6 (parameters: k = 13, s = 2). Reads with a mapping quality lower than
486 25 were removed using SAMtools v.1.21 ('-q' option) and read depths across all positions were
487 extracted ('samtools depth -a') (61). This reference genome has been shown to work well for
488 genotyping not only *L. braziliensis* isolates, but also other *L.* (*Viannia*) species, resulting in highly
489 accessible genomes for downstream analyses (28-30). Because transcriptomic data were used
490 rather than genomic data, we discarded isolates with low median coverages (<10x) across the
491 coding DNA sequences (CDS) from downstream analyses. Additionally, all CDS regions, in each
492 sample, with a coverage below 10x were omited, ensuring that only transcriptomic regions with
493 sufficient depth were included in the subsequent analyses. Variant calling (i.e., SNPs and INDELs)
494 was caried out using GATK v.4.1.4.1 (75). Here, the initial calling was performed on a per-sample
495 basis using HaplotypeCaller, after which all individual genotype vcf files were combined using
496 CombineGVCFs, and variants were jointly genotyped with GenotypeGVCFs. Next, SNPs and
497 INDELs were separated using SelectVariants after which both panels were subjected to hard
498 filtering with VariantFiltration. For SNPs, the following filters were applied: (i) QD < 2.0, FS > 60.0,
499 MQ < 40.0, MQRankSum < −12.5 and ReadPosRankSum < −8.0, following GATK's
500 recommendations (76); (ii) GQ < 10; and (iii) QUAL < 100. For INDELs, GATK's recommendations
501 were adopted: QD < 2.0, FS > 200.0, or ReadPosRankSum < −20.0 (76).

502 A phylogenetic network was generated using SplitsTree v.4.17.0 (77) based on the
503 pairwise uncorrected *p*-distances calculated across 157,649 bi-allelic SNPs. This network was built
504 using the NeighborNet (78) and EqualAngle (79) algorithms as implemented in SplitsTree. Forty-
505 nine publicly available isolates were included as reference sequences: members of the *L.* (*V.*)
506 *braziliensis* clade, the *L.* (*V.*) *guyanensis* species complex, as well as *L.* (*V.*) *naiffi* and *L.* (*V.*)
507 *lainsoni* (S. Tab. 6). In addition, the alternate allele read depth frequency (ARDF) distributions at
508 heterozygous sites were calculated and visualized to identify whether potential hybrid isolates were
509 the result from a mixed infection. This was done using a custom python script (vcf2freq.py; available
510 at: https://github.com/FreBio/mytools). Differences in SNP counts between the different species
511 were compared using a non-parametric Kruskal-Wallis (stats R-package; (80)) test, followed by
512 pairwise Dunn's tests with additional Benjamini-Hochberg *p*-value correction for multiple testing
513 (FSA R-package; (81)).

514 **Preparing alignments and phylogenies for co-phylogenetic reconciliation**

515 Phylogenetic reconciliations of the LRV-*Leishmania* symbiosis were performed at multiple
516 levels: subgenus (combining Old World and New World species), species (focusing on the South
517 American *Leishmania* (*Viannia*) species), population (focusing on *L.* (*V.*) *braziliensis* and *L.* (*V.*)
518 *guyanensis* with their respective LRV1 lineages) and gene level (reconciling the CP and RDRP
519 gene phylogenies within LRV1).

520 For the subgenus level analysis, we included 22 publicly available Peruvian *L.* (*V.*)
521 *braziliensis* isolates and six Old World *Leishmania* isolates positive for LRV2 (S. Tab. 4,6). Instead
522 of using the entire genome sequence, we concatenated the annotated CP and RDRP genes (i.e.,
523 ORFs 2 and 3 in the LRV genome) and aligned their sequences from LRV1 and LRV2 using the
524 MAFFT v.7.490 as described above (63). The LRV phylogeny was reconstructed using IQtree2
525 with 10,000 ultra-fast bootstrap replicates and the GTR+F+I+R3 substitution model based on the
526 ModelFinder algorithm (64–66). For the maximum likelihood phylogenetic reconstruction of the
527 *Leishmania* phylogeny, which included highly divergent species from different subgenera, a
528 concatenated alignment of six orthologous protein coding genes was used (S. Tab. 15), following
529 Coughlan et al. (2017) (82). The sequences of the six target genes were obtained as follows. Reads
530 were trimmed with fastp v.0.23.4 using similar settings used in Heeren et al. (2023): a minimum
531 base quality (-q) of 30; maximum percentage of unqualified bases per read (-u) to 10; quality-based
532 sliding window trimming front-to-tail (-5) and tail-to-front (-3) with a window size (-W) of 1 and mean
533 quality score (-M) of 30; right-end cutting (--cut-right) using 10 bp windows (--
534 cut_right_window_size) with mean quality score (--cut_right_mean_quality) below 30; and retention
535 of reads between 100 (-l) and 150 bp (-b) (3). Trimmed reads were then assembled de novo using
536 Megahit v.1.2.9 (58). Contigs with six target genes were identified using a blastn (60) search and
537 subsequently extracted. These sequences were combined with 222 publicly available ones
538 (accession numbers listed in S. Tab. 15) and aligned using MAFFT, as described above. Finally, a
539 phylogenetic tree was inferred using IQtree2 with 10,000 ultra-fast bootstrap replicates and
540 TIM2+F+G4 set as optimal substitution model (64–66).

541 For the reconstruction of the parasite phylogenies at the species- and population-level co-
542 phylogenetic analyses, we used concatenated SNP-alignments. Variants were called against the
543 MHOM/BR/75/M2904 reference genome, as described above, resulting in genome-wide alignment
544 containing 97,530 bi-allelic SNPs. From this panel, a phylogenetic tree was inferred using IQtree2
545 with 10,000 ultra-fast bootstrap replicates and the TVM+F+R2 as the most optimal substitution
546 model, correcting for ascertainment bias (+ASC) (64–66). For the population-level analyses of *L.*

547 *braziliensis*–LRV1 and *L. guyanensis*–LRV1, species specific partitions of the inferred *L.* (*Viannia*)
548 phylogeny were used. Finally, for the reconciliation of the LRV1 CP and RDRP genes we made
549 use of the reconstructed phylogenies as described above.

550 **Genetic distance regression**

551 In order to obtain an initial assessment of similarity between the viral and parasite
552 phylogenies, we calculated pairwise genetic distances from the corresponding concatenated
553 sequences using the 'dist.dna' function of the ape R-package (v.5.0) (67) (model = 'K80'). These
554 distances were converted into distance matrices, and analyzed using Multiple Matrix Regression
555 with Randomization (MMRR) (83) where *p*-values were estimated over 1,000 permutations.

556 **Global-fit co-phylogenetic analyses**

557 We tested for the presence of co-phylogenetic signals (i.e., testing whether closely related
558 hosts tend to associate with closely related symbionts) between the parasite and virus phylogenies
559 by making use of three global-fit methods: (i) the Robinson-Foulds distance as implemented in the
560 phytools R-package (84); (ii) ParaFit as implemented in ape (67); and (iii) PACo as implemented
561 in the paco R-package (85). For each method, both virus and parasite phylogenies were
562 downsampled using the 'drop.tip' function of ape (67). This way, only the isolates for which both
563 high quality LRV1 genomes and *L.* (*Viannia*) transcriptome-based genotypes were available. Each
564 approach was performed by running 1,000 permutations for p-value estimation and the Cailliez
565 correction method was used in ParaFit and PACo to correct for negative eigenvalues.

566 **Event-based co-phylogenetic analysis**

567 To test whether *Leishmania* and LRV exhibit signs of phylogenetic congruence, at one or
568 multiple hierarchical levels, we made use of eMPRess, an event-based phylogenetic reconciliation
569 tool that for reconciling phylogenies based on a set of events such as co-speciation, loss,
570 duplication and host-switch (25, 86–88). A key advantage of eMPRess over other event-based
571 reconciliation tools, is its ability to specify different sets of event-costs prior to calculating the
572 reconciliation, allowing the user to evaluate alternative evolutionary scenarios for obtaining more
573 robust inferences (86). This is important as different event-costs schemes can lead to distinct
574 solutions and interpretations. For each level, we visualized the cost-space (S. Fig, 7) and selected
575 multiple sets of event-costs covering most of the cost-space (S. Tab. 11-13). Phylogenetic
576 reconciliations were then calculated for each set, and a median MPR (i.e. a median value of the
577 MPR solution space; MPR = Maximum Parsimony reconciliation) was visualized. Significance was
578 assessed using 1,000 random permutations. Visualizations of the cost-spaces and reconciled
579 phylogenies were generated using the eMPRess GUI, while significance testing was conducted
580 with the CLI version ('*p*-value' function with '--n-samples 1000' option).

581 ## Acknowledgments

582 This work received financial support from the Directie-Generaal Ontwikkelingssamenwerking en
583 Humanitaire Hulp (DGD) (Belgian cooperation). F.V.d.B. and S.H. acknowledge support from the
584 Research Foundation Flanders (grants 1226120N and G092921N). S.H. additionally acknowledges
585 support from KU Leuven's Internal Funds (PDMt1/25/007). E.C. acknowledges support from
586 Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro: CNE -
587 204.165/2024; 269.891/2021; ColBio - 210.285/202; Conselho Nacional de Desenvolvimento
588 Científico e Tecnológico: Research Fellow - 309627/2021-4; PROEP/IOC - 441639/2024-0; INCT-
589 EpiAmo: 465657/2014-1; and the GGBN award: GGI-GGBN-2021-278. L.M.C. acknowledges
590 support from Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro,

Pós-Doutorado; Nota 10 E-26/205.730/2022 and 205.731/2022. E.C., L.M.C, M.C.d.O.S. and K.C. also acknowledge support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Data availability

Sequencing data (i.e., raw sequencing reads and the the reconstructed LRV genomes), that were generated for this study, will be deposited to publicly accessible databases upon publication of the manuscript.

Key-words

- Co-evolution
- *Leishmania*
- *Leishmaniavirus*
- (co-)phylogenetics

604 **References**

605 1. D. Grybchuk, A. Y. Kostygov, D. H. Macedo, C. M. d'Avila-Levy, V. Yurchenko, RNA
606 viruses in trypanosomatid parasites: a historical overview. *Mem. Inst. Oswaldo Cruz* **113**,
607 e170487 (2018).

608 2. D. Klocek, *et al.*, Evolution of RNA viruses in trypanosomatids: new insights from the
609 analysis of Sauroleishmania. *Parasitol. Res.* (2023). https://doi.org/10.1007/s00436-023-07928-x.

610 3. S. Heeren, *et al.*, Diversity and dissemination of viruses in pathogenic protozoa. *Nat.*
611 *Commun.* **14**, 8343 (2023).

612 4. P. Gupta, *et al.*, A parasite odyssey: An RNA virus concealed in Toxoplasma gondii.
613 *Virus Evol.* **10**, veae040 (2024).

614 5. J. Charon, *et al.*, Novel RNA viruses associated with Plasmodium vivax in human malaria
615 and Leucocytozoon parasites in avian disease. *PLoS Pathog.* **15**, e1008216 (2019).

616 6. A. Ibañez-Escribano, *et al.*, Microbial matryoshka: Addressing the relationship between
617 pathogenic flagellated protozoans and their RNA viral endosymbionts (family Totiviridae). *Vet.*
618 *Sci.* **11**, 321 (2024).

619 7. A. Ives, *et al.*, Leishmania RNA virus controls the severity of mucocutaneous
620 leishmaniasis. *Science* **331**, 775–778 (2011).

621 8. V. Adaui, *et al.*, Association of the Endobiont Double-Stranded RNA Virus LRV1 With
622 Treatment Failure for Human Leishmaniasis Caused by Leishmania braziliensis in Peru and
623 Bolivia. *J. Infect. Dis.* **213**, 112–121 (2016).

624 9. E. Bourreau, *et al.*, Presence of Leishmania RNA Virus 1 in Leishmania guyanensis
625 Increases the Risk of First-Line Treatment Failure and Symptomatic Relapse. *J. Infect. Dis.* **213**,
626 105–111 (2016).

627 10. L. M. Cantanhêde, *et al.*, The Maze Pathway of Coevolution: A Critical Review over the
628 Leishmania and Its Endosymbiotic History. *Genes* **12** (2021).

629 11. P. Simmonds, *et al.*, Changes to virus taxonomy and the ICTV Statutes ratified by the
630 International Committee on Taxonomy of Viruses (2024). *Arch. Virol.* **169**, 236 (2024).

631 12. R. V. H. de Carvalho, *et al.*, Leishmania RNA virus exacerbates Leishmaniasis by
632 subverting innate immunity via TLR3-mediated NLRP3 inflammasome inhibition. *Nat. Commun.*
633 **10**, 5273 (2019).

634 13. F. A. Pazmiño, *et al.*, Mucosal leishmaniasis is associated with the Leishmania RNA virus
635 and inappropriate cutaneous leishmaniasis treatment. *PLoS One* **20**, e0317221 (2025).

636 14. B. I. Hillman, A. B. Cohen, "Totiviruses (Totiviridae)" in *Encyclopedia of Virology (Fourth*
637 *Edition)*, D. H. Bamford, M. Zuckerman, Eds. (Academic Press, 2021), pp. 648–657.

638 15. M. C. de O. Santana, K. Chourabi, L. M. Cantanhêde, E. Cupolillo, Exploring Host-
639 Specificity: Untangling the Relationship between Leishmania (Viannia) Species and Its
640 Endosymbiont Leishmania RNA Virus 1. *Microorganisms* **11** (2023).

641 16. G. Widmer, S. Dooley, Phylogenetic analysis of Leishmania RNA virus and Leishmania
642 suggests ancient virus-parasite association. *Nucleic Acids Res.* **23**, 2300–2304 (1995).

643 17. D. Grybchuk, *et al.*, RNA Viruses in Blechomonas (Trypanosomatidae) and Evolution of
644 Leishmaniavirus. *MBio* **9** (2018).

645 18. H. Hajjaran, *et al.*, Detection and molecular identification of leishmania RNA virus (LRV)
646 in Iranian Leishmania species. *Arch. Virol.* **161**, 3385–3390 (2016).

647 19. R. Saberi, *et al.*, Presence and diversity of Leishmania RNA virus in an old zoonotic
648 cutaneous leishmaniasis focus, northeastern Iran: haplotype and phylogenetic based approach.
649 *Int. J. Infect. Dis.* **101**, 6–13 (2020).

650 20. M. Nalçacı, *et al.*, Detection of Leishmania RNA virus 2 in Leishmania species from
651 Turkey. *Trans. R. Soc. Trop. Med. Hyg.* **113**, 410–417 (2019).

652 21. R. Kariyawasam, R. Lau, B. M. Valencia, A. Llanos-Cuentas, A. K. Boggild, Novel
653 detection of Leishmania RNA virus-1 (LRV-1) in clinical isolates of Leishmania Viannia
654 panamensis. *Parasitology* **151**, 151–156 (2024).

655 22. A. A. Bonilla, *et al.*, Epidemiology and Genetic Characterization of Leishmania RNA Virus
656 in Leishmania (Viannia) spp. Isolates from Cutaneous Leishmaniasis Endemic Areas in Panama.
657 *Microorganisms* **12**, 1317 (2024).

658 23. A. Y. Kostygov, *et al.*, Analyses of Leishmania-LRV Co-Phylogenetic Patterns and
659 Evolutionary Variability of Viral Proteins. *Viruses* **13** (2021).

660 24. S. M. Scheffter, Y. T. Ro, I. K. Chung, J. L. Patterson, The complete sequence of
661 Leishmania RNA virus LRV2-1, a virus of an Old World parasite strain. *Virology* **212**, 84–90
662 (1995).

663 25. B. Perez-Lamarque, H. Morlon, Distinguishing cophylogenetic signal from phylogenetic
664 congruence clarifies the interplay between evolutionary history and species interactions. *Syst.*
665 *Biol.* (2024). https://doi.org/10.1093/sysbio/syae013.

666 26. J. Klaps, P. Lemey, nf-core community, L. E. Kafetzopoulou, Nf-core/viralmetagenome: A
667 novel pipeline for untargeted viral genome reconstruction. *bioRxiv* 2025.06.27.661954 (2025).

668 27. S. Tirera, *et al.*, Unraveling the genetic diversity and phylogeny of Leishmania RNA virus
669 1 strains of infected Leishmania isolates circulating in French Guiana. *PLoS Negl. Trop. Dis.* **11**,
670 e0005764 (2017).

671 28. F. Van den Broeck, *et al.*, Genome Analysis of Triploid Hybrid Leishmania Parasite from
672 the Neotropics. *Emerg. Infect. Dis.* **29**, 1076–1078 (2023).

673 29. S. Heeren, *et al.*, Evolutionary genomics of Leishmania braziliensis across the
674 neotropical realm. *Commun. Biol.* **7**, 1587 (2024).

675 30. F. Van den Broeck, *et al.*, Ecological divergence and hybridization of Neotropical
676 Leishmania parasites. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 25159–25168 (2020).

677 31. M. B. Rogers, *et al.*, Chromosome and gene copy number variation allow major structural
678 change between species and strains of Leishmania. *Genome Res.* **21**, 2129–2142 (2011).

679 32. E. Tihon, H. Imamura, J.-C. Dujardin, J. Van Den Abbeele, F. Van den Broeck, Discovery
680 and genomic analyses of hybridization between divergent lineages of Trypanosoma congolense,
681 causative agent of Animal African Trypanosomiasis. *Mol. Ecol.* **26**, 6524–6538 (2017).

682 33. K. C. King, R. B. Stelkens, J. P. Webster, D. F. Smith, M. A. Brockhurst, Hybridization in
683 Parasites: Consequences for Adaptive Evolution, Pathogenesis, and Public Health in a Changing
684 World. *PLoS Pathog.* **11**, e1005098 (2015).

685 34. W. Gibson, *et al.*, Genetic recombination between human and animal parasites creates
686 novel strains of human pathogen. *PLoS Negl. Trop. Dis.* **9**, e0003665 (2015).

687 35. A. Izeta-Alberdi, C. N. Ibarra-Cerdeña, D. A. Moo-Llanes, J. M. Ramsey, Geographical,
688 landscape and host associations of Trypanosoma cruzi DTUs and lineages. *Parasit. Vectors* **9**,
689 631 (2016).

690 36. N. S. Akopyants, *et al.*, Demonstration of genetic exchange during cyclical development
691 of Leishmania in the sand fly vector. *Science* **324**, 265–268 (2009).

692 37. M. B. Rogers, *et al.*, Genomic confirmation of hybridisation and recent inbreeding in a
693 vector-isolated Leishmania population. *PLoS Genet.* **10**, e1004092 (2014).

694 38. E. Inbar, *et al.*, Whole genome sequencing of experimental hybrids supports meiosis-like
695 sexual recombination in Leishmania. *PLoS Genet.* **15**, 1–28 (2019).

696 39. A. Hadermann, *et al.*, Genome diversity of Leishmania aethiopica. *Front. Cell. Infect.*
697 *Microbiol.* **13** (2023).

698 40. M. A. Domagalska, J.-C. Dujardin, Next-generation molecular surveillance of TriTryp
699 diseases. *Trends Parasitol.* **36**, 356–367 (2020).

700 41. D. Grybchuk, *et al.*, The First Non-LRV RNA Virus in Leishmania. *Viruses* **12** (2020).

701 42. H. R. Hughes, *et al.*, Summary of taxonomy changes ratified by the International
702 Committee on Taxonomy of Viruses (ICTV) from the Animal dsRNA and ssRNA(-) Viruses
703 Subcommittee, 2025. *J. Gen. Virol.* **106** (2025).

704 43. A. Hayward, R. Poulin, S. Nakagawa, A broadscale analysis of host-symbiont
705 cophylogeny reveals the drivers of phylogenetic congruence. *Ecol. Lett.* **24**, 1681–1696 (2021).

706 44. V. D. Atayde, *et al.*, Exploitation of the Leishmania exosomal pathway by Leishmania
707 RNA virus 1. *Nat Microbiol* **4**, 714–723 (2019).

708 45. A. Lafleur, M. Olivier, Viral endosymbiotic infection of protozoan parasites: How it
709 influences the development of cutaneous leishmaniasis. *PLoS Pathog.* **18**, e1010910 (2022).

710 46. M. Olivier, D. S. Zamboni, Leishmania Viannia guyanensis, LRV1 virus and extracellular
711 vesicles: a dangerous trio influencing the faith of immune response during muco-cutaneous
712 leishmaniasis. *Curr. Opin. Immunol.* **66**, 108–113 (2020).

713 47. G. Dong, V. Wagner, A. Minguez-Menendez, C. Fernandez-Prada, M. Olivier,
714 Extracellular vesicles and leishmaniasis: Current knowledge and promising avenues for future
715 development. *Mol. Immunol.* **135**, 73–83 (2021).

716 48. A. Y. Kostygov, *et al.*, A novel strain of Leishmania braziliensis harbors not a toti- but a
717 bunyavirus. *PLoS Negl. Trop. Dis.* **18**, e0012767 (2024).

718 49. L. Guilbride, P. J. Myler, K. Stuart, Distribution and sequence divergence of LRV1 viruses
719 among different Leishmania species. *Mol. Biochem. Parasitol.* **54**, 101–104 (1992).

720 50. L. de O. R. Pereira, *et al.*, Severity of tegumentary leishmaniasis is not exclusively
721 associated with Leishmania RNA virus 1 infection in Brazil. *Mem. Inst. Oswaldo Cruz* **108**, 665–
722 667 (2013).

723 51. D. H. Macedo, *et al.*, Low frequency of LRV1 in Leishmania braziliensis strains isolated
724 from typical and atypical lesions in the State of Minas Gerais, Brazil. *Mol. Biochem. Parasitol.*
725 **210**, 50–54 (2016).

726 52. M.-A. Hartley, C. Ronet, H. Zangger, S. M. Beverley, N. Fasel, Leishmania RNA virus:
727 when the host pays the toll. *Front. Cell. Infect. Microbiol.* **2**, 99 (2012).

728 53. L. M. Cantanhêde, *et al.*, Further evidence of an association between the presence of
729 Leishmania RNA virus 1 and the mucosal manifestations in tegumentary leishmaniasis patients.
730 *PLoS Negl. Trop. Dis.* **9**, e0004079 (2015).

731 54. A. Zakharova, *et al.*, Leishmania guyanensis M4147 as a new LRV1-bearing model
732 parasite: Phosphatidate phosphatase 2-like protein controls cell cycle progression and
733 intracellular lipid content. *PLoS Negl. Trop. Dis.* **16**, e0010510 (2022).

734 55. P. I. Tarr, *et al.*, LR1: a candidate RNA virus of Leishmania. *Proc. Natl. Acad. Sci. U. S.*
735 *A.* **85**, 9572–9575 (1988).

736 56. P. Ranque, "Etude morphologique et biologique de quelques Trypanosomatides récoltés
737 au Sénégal," Université de Aix-Marseilles. (1973).

738 57. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor.
739 *Bioinformatics* **34**, i884–i890 (2018).

740 58. D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, MEGAHIT: an ultra-fast single-node
741 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
742 *Bioinformatics* **31**, 1674–1676 (2015).

743 59. A. Prjibelski, D. Antipov, D. Meleshko, A. Lapidus, A. Korobeynikov, Using SPAdes DE
744 Novo Assembler. *Curr. Protoc. Bioinformatics* **70**, e102 (2020).

745 60. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment
746 search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

747 61. P. Danecek, *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).

748 62. N. D. Grubaugh, *et al.*, An amplicon-based sequencing framework for accurately
749 measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20**, 8 (2019).

750 63. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7:
751 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

752 64. B. Q. Minh, *et al.*, IQ-TREE 2: New models and efficient methods for phylogenetic
753 inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

754 65. B. Q. Minh, M. A. T. Nguyen, A. von Haeseler, Ultrafast approximation for phylogenetic
755 bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).

756 66. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin,
757 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–
758 589 (2017).

759 67. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and
760 evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2018).

761 68. M. A. Suchard, *et al.*, Bayesian phylogenetic and phylodynamic data integration using
762 BEAST 1.10. *Virus Evol* **4**, vey016 (2018).

763 69. D. L. Ayres, *et al.*, BEAGLE: an application programming interface and high-performance
764 computing library for statistical phylogenetics. *Syst. Biol.* **61**, 170–173 (2012).

765 70. S. Tavaré, Some probabilistic and statistical problems in the analysis of DNA sequences.
766 *Lecture of mathematics for life science* (1986).

767

71. Z. Yang, Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).

72. P. Lemey, A. Rambaut, A. J. Drummond, M. A. Suchard, Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* **5**, e1000520 (2009).

73. T. Gernhard, The conditioned reconstructed process. *J. Theor. Biol.* **253**, 769–778 (2008).

74. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

75. G. A. Van der Auwera, B. D. OConnor, *Genomics in the Cloud Using Docker, GATK, and WDL in Terra (First edition)* (O'Reilly Media, 2020).

76. G. A. Van der Auwera, *et al.*, From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).

77. D. H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).

78. D. Bryant, V. Moulton, NeighborNet: An Agglomerative Method for the Construction of Planar Phylogenetic Networks in *Algorithms in Bioinformatics*, (Springer Berlin Heidelberg, 2002), pp. 375–391.

79. A. W. M. Dress, D. H. Huson, Constructing splits graphs. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **1**, 109–115 (2004).

80. R Core Team, R: A Language and Environment for Statistical Computing. [Preprint] (2024). Available at: https://www.R-project.org/.

81. D. H. Ogle, J. C. Doll, A. P. Wheeler, A. Dinno, FSA: Simple fisheries stock assessment methods. R package version 0.9. 4. [Preprint] (2023).

82. S. Coughlan, *et al.*, The genome of Leishmania adleri from a mammalian host highlights chromosome fission in Sauroleishmania. *Sci. Rep.* **7**, 43747 (2017).

83. I. J. Wang, Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* **67**, 3403–3411 (2013).

84. L. J. Revell, phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* [Preprint] (2024). Available at: http://dx.doi.org/10.7717/peerj.16505.

85. J. A. Balbuena, T. Poisot, M. Hutchinson, F. C. S. P. O. B. et al 2013 <https://doi.org/10.1371/journal.pone.0061048>, paco: Procrustes Application to Cophylogenetic Analysis. [Preprint] (2020). Available at: https://CRAN.R-project.org/package=paco.

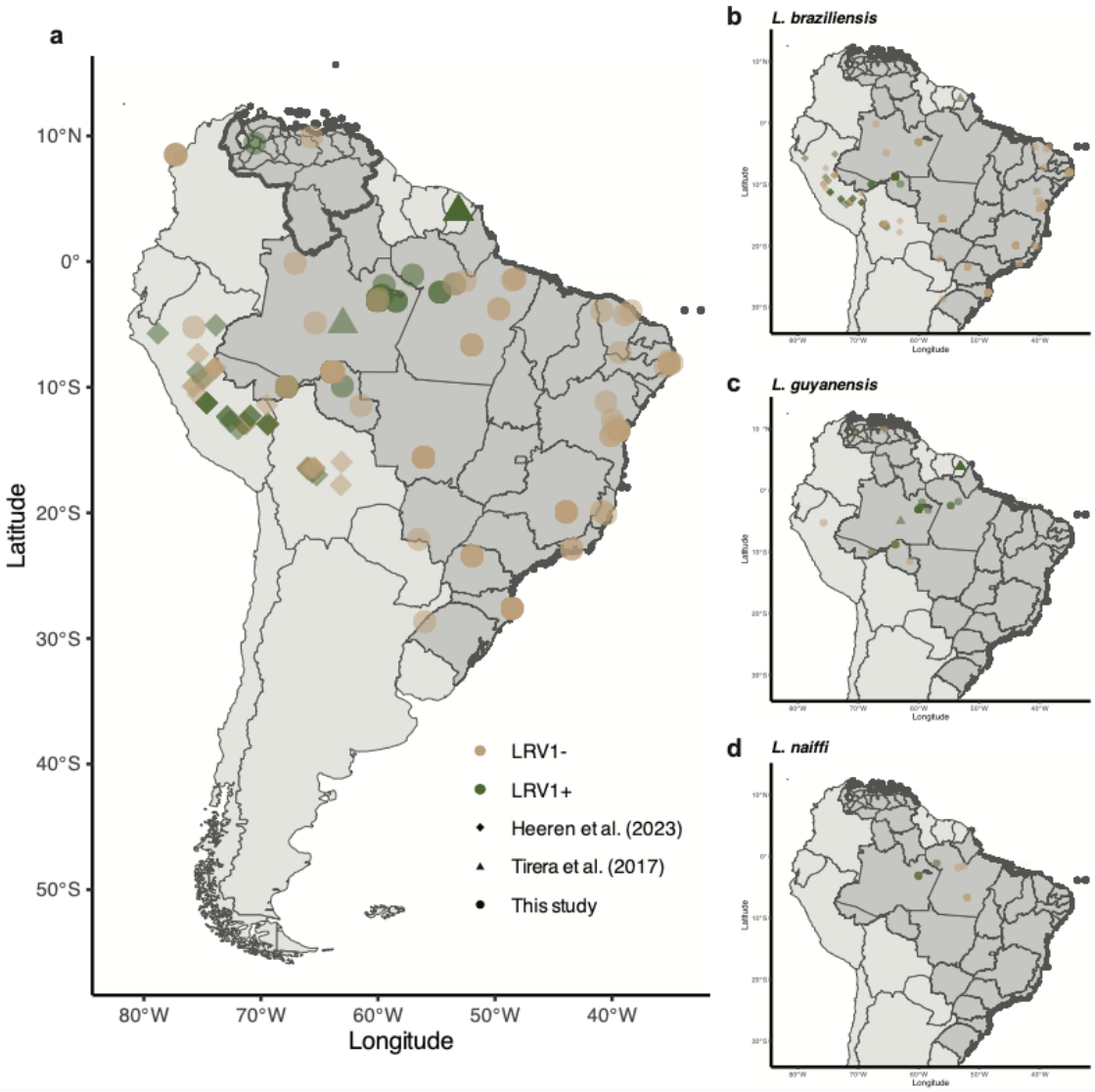
86. S. Santichaivekin, *et al.*, eMPRess: a systematic cophylogeny reconciliation tool. *Bioinformatics* **37**, 2481–2482 (2021).

87. W. Dismukes, M. P. Braga, D. H. Hembry, T. A. Heath, M. J. Landis, Cophylogenetic Methods to Untangle the Evolutionary History of Ecological Interactions. *Annu. Rev. Ecol. Evol. Syst.* (2022). https://doi.org/10.1146/annurev-ecolsys-102320-112823.

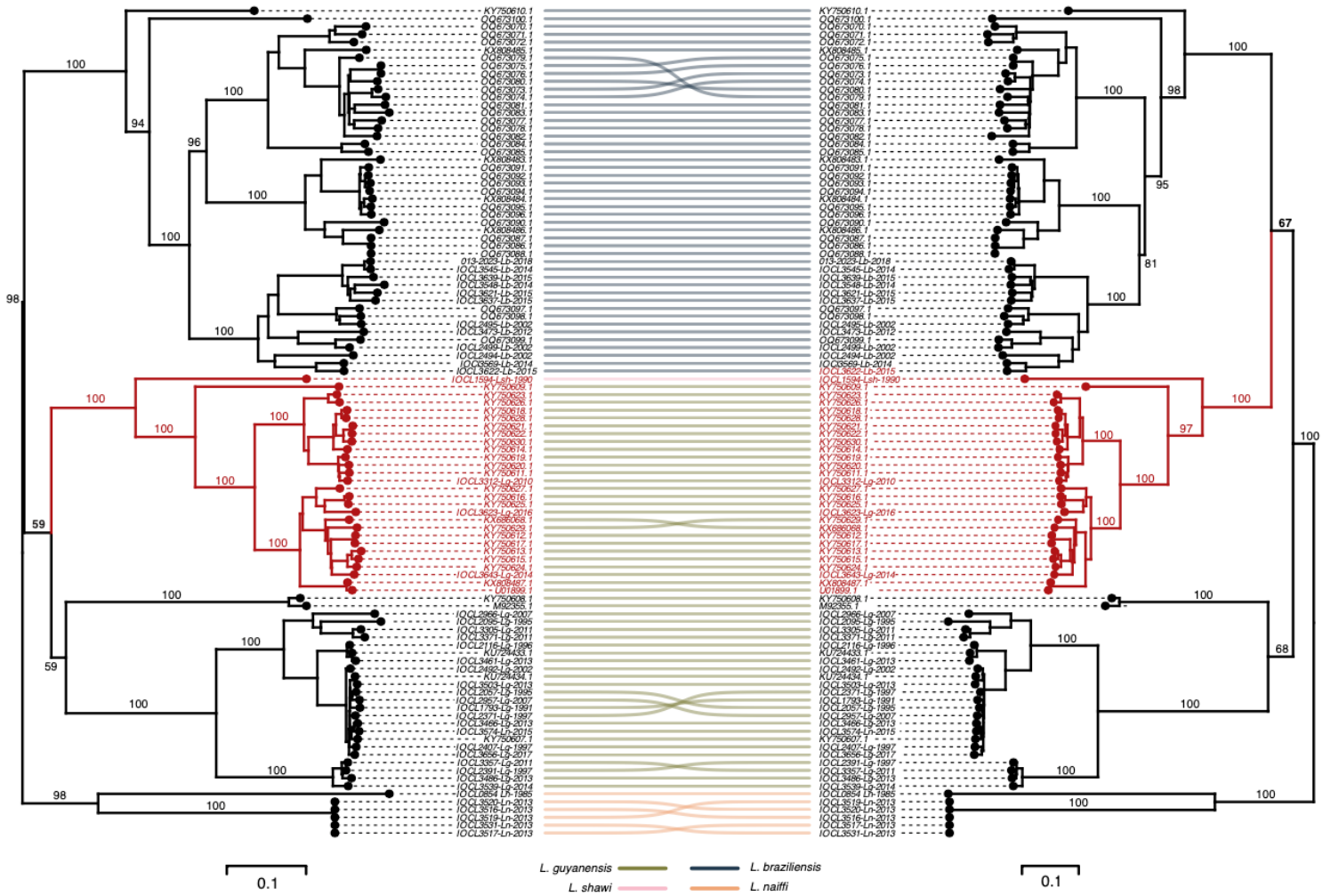
809 88. D. M. de Vienne, *et al.*, Cospeciation vs host-shift speciation: methods for testing,
810 evidence from natural associations and relation to coevolution. *New Phytol.* **198**, 347–385 (2013).

811

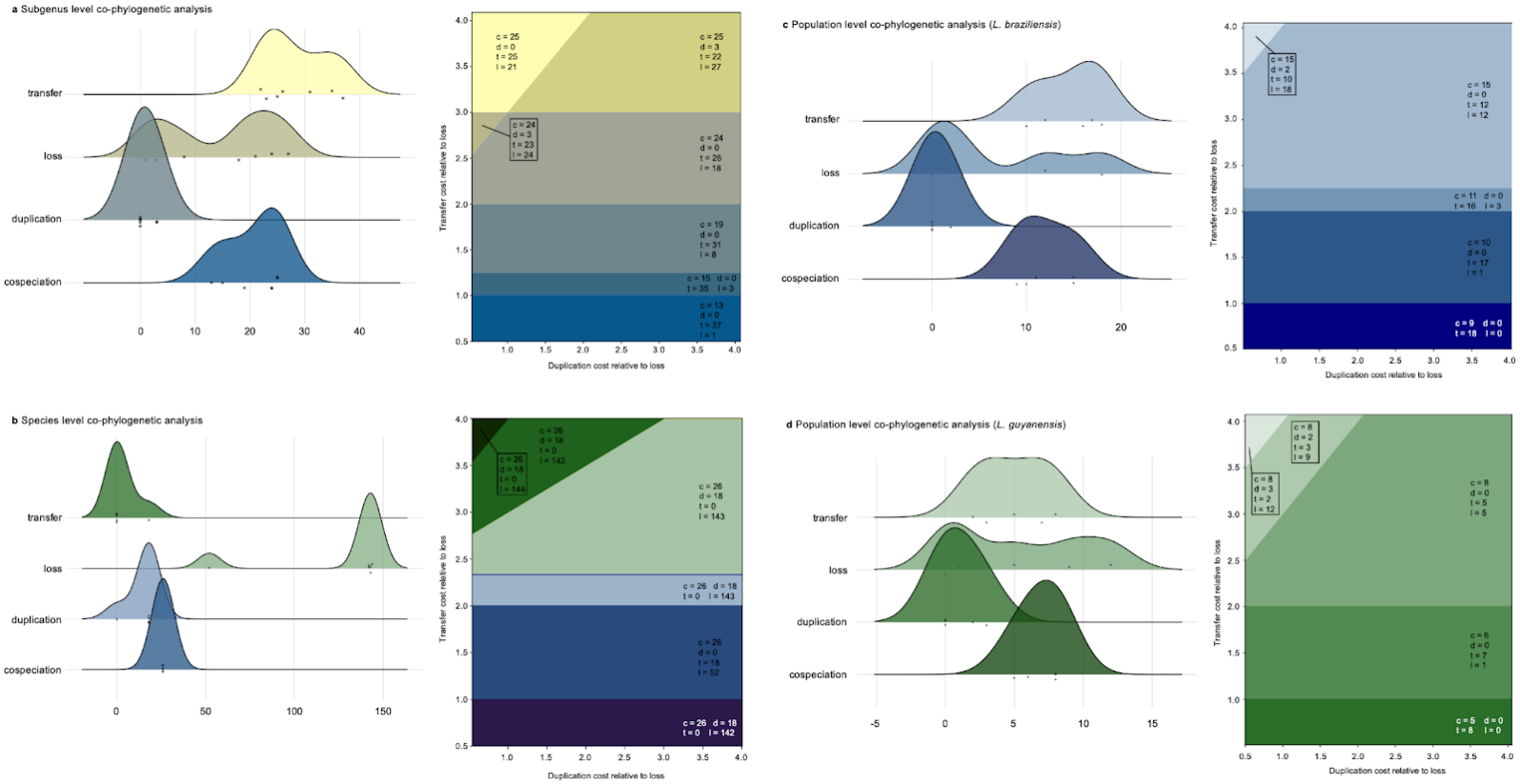
Supplementary Figures



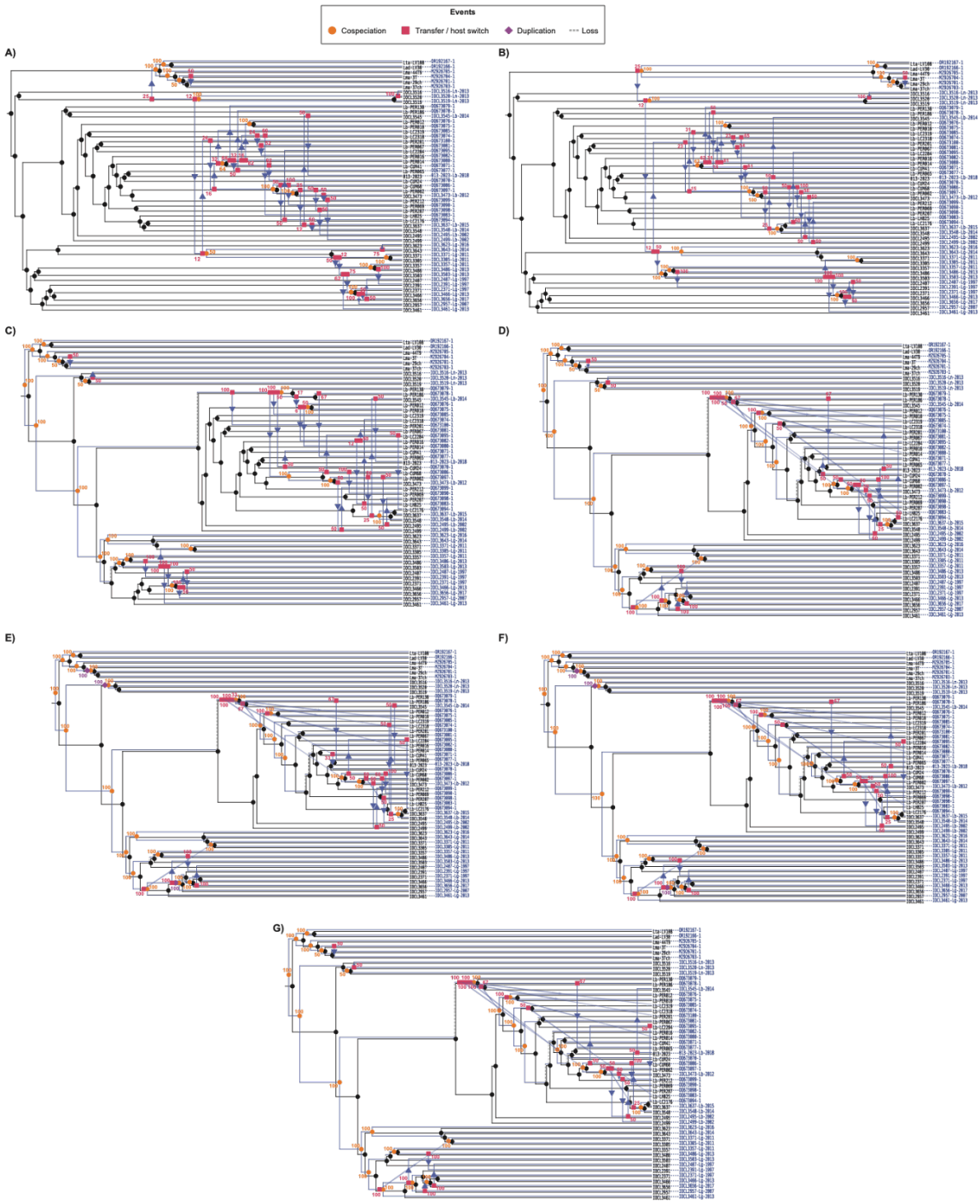
S. Fig. 1. Spatial distribution of LRV1-positive and negative *Leishmania* isolates. **a** Spatial distribution of all isolates used in this study along with geolocalized *Leishmania* isolates from Heeren et al. (2023) (3) and Tirera et al. (2017) (24). **b-d** Spatial distribution of specific *Leishmania* species: *L. braziliensis* (**b**), *L. guyanensis* (**c**) and *L. naiffi* (**d**). Administrative country- and state-level data originate from: http://www.diva-gis.org/Data.



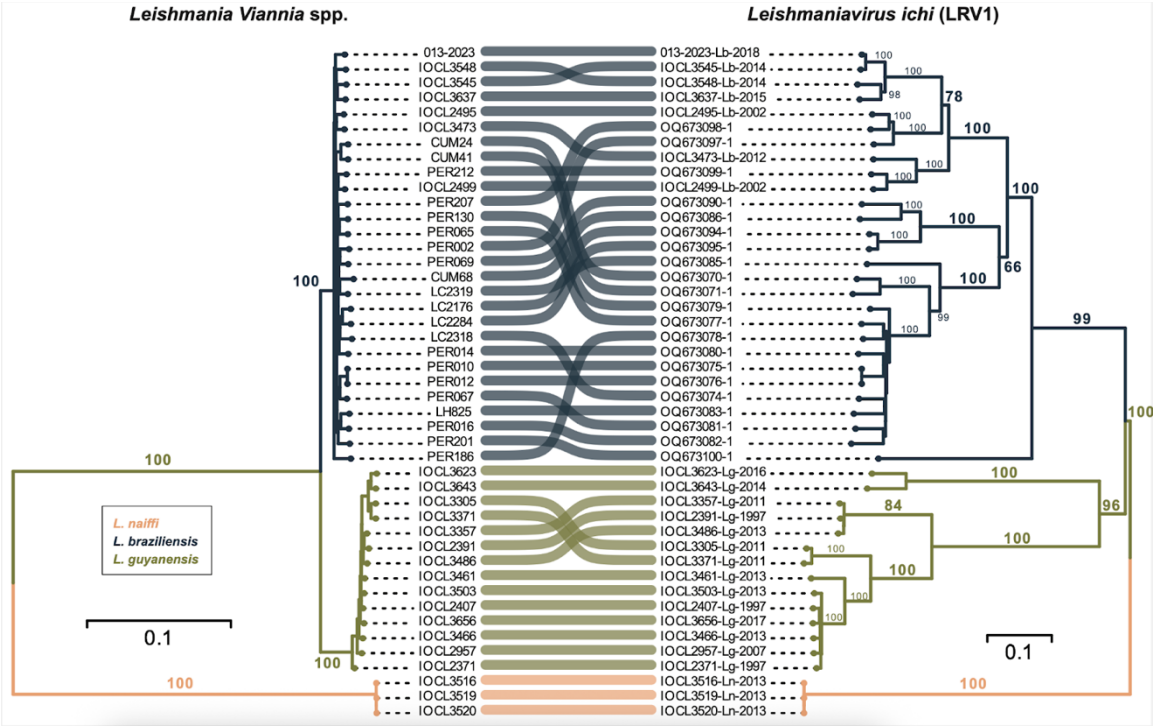
S. Fig. 2: Tanglegram of the LRV1 capsid protein (CP) and RNA-dependent RNA polymerase (RDRP) gene phylogenies. The clade colored in red occupies different positions in the CP and RDRP phylogenies. Branch values represent bootstrap support values based on 10,000 ultra-fast approximated bootstrap replicates. The scale bar depicts the number of substitutions per site. Links between the trees are colored according to the identified clades of the 42 newly assembled LRV1 genomes (Figure 1; Suppl Figure 1).



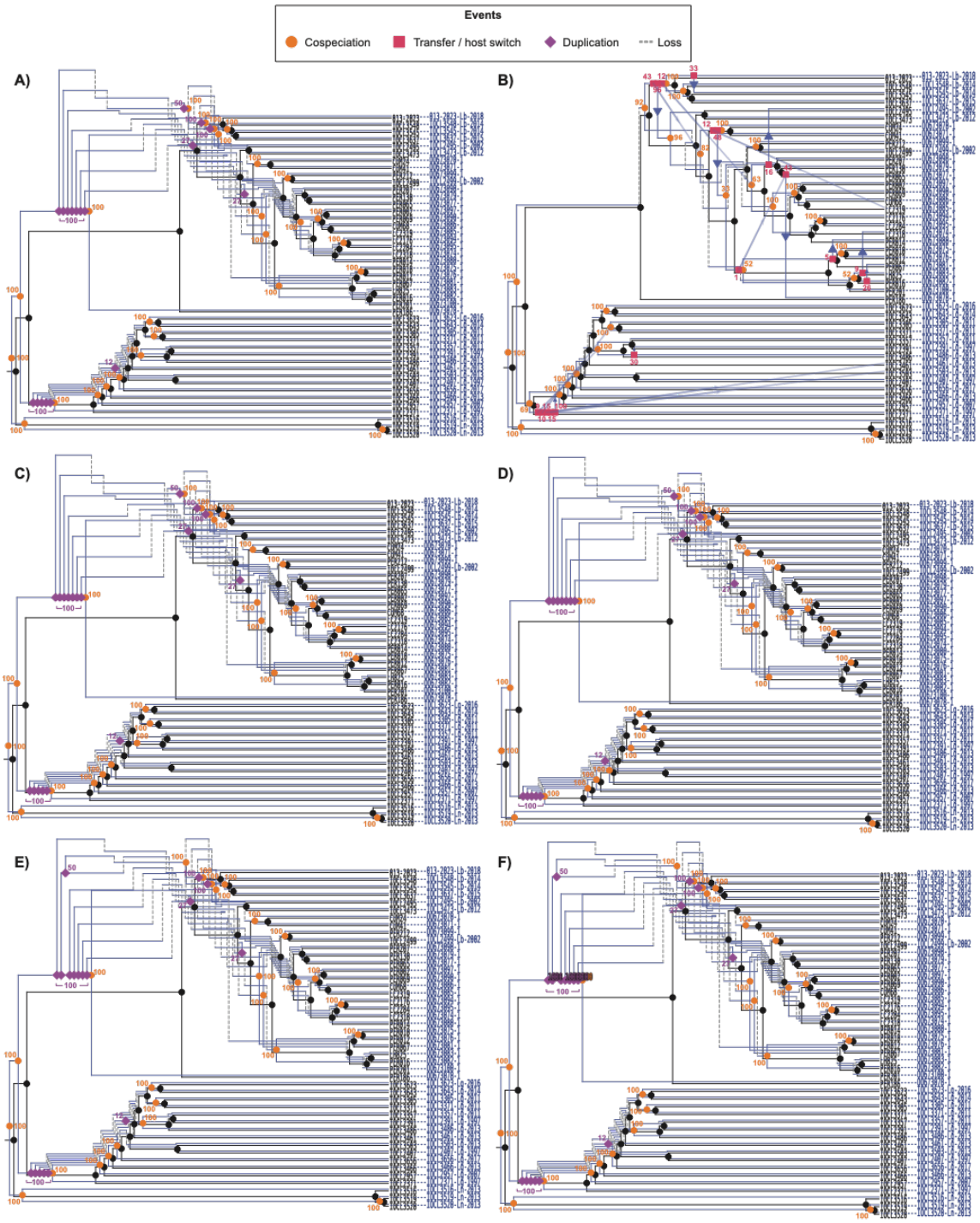
S. Fig. 3: Event-cost distributions and cost spaces from event-based co-phylogenetic analyses of the *Leishmania*–LRV symbiosis across the different hierarchical levels. Per panel, the left-side plot depicts the density of the number of events that have been calculated over all the tested cost-sets; (ii) the right-side plot indicates the costspace as calculated by eMPRess. Each colored zone corresponds to the same number of events and reconciliation solutions. **a** Subgenus level. **b** Species level (within *L.* (*Viannia*)). Note the high number of inferred loss events in this analysis, resulting from an adjustmen of the default loss cost from 1 to 0 to enable co-phylogenetic reconciliations in eMPRess. **c** Population level for *L.* (*V.*) *braziliensis*. **d** Population level for *L.* (*V.*) *guyanensis*.



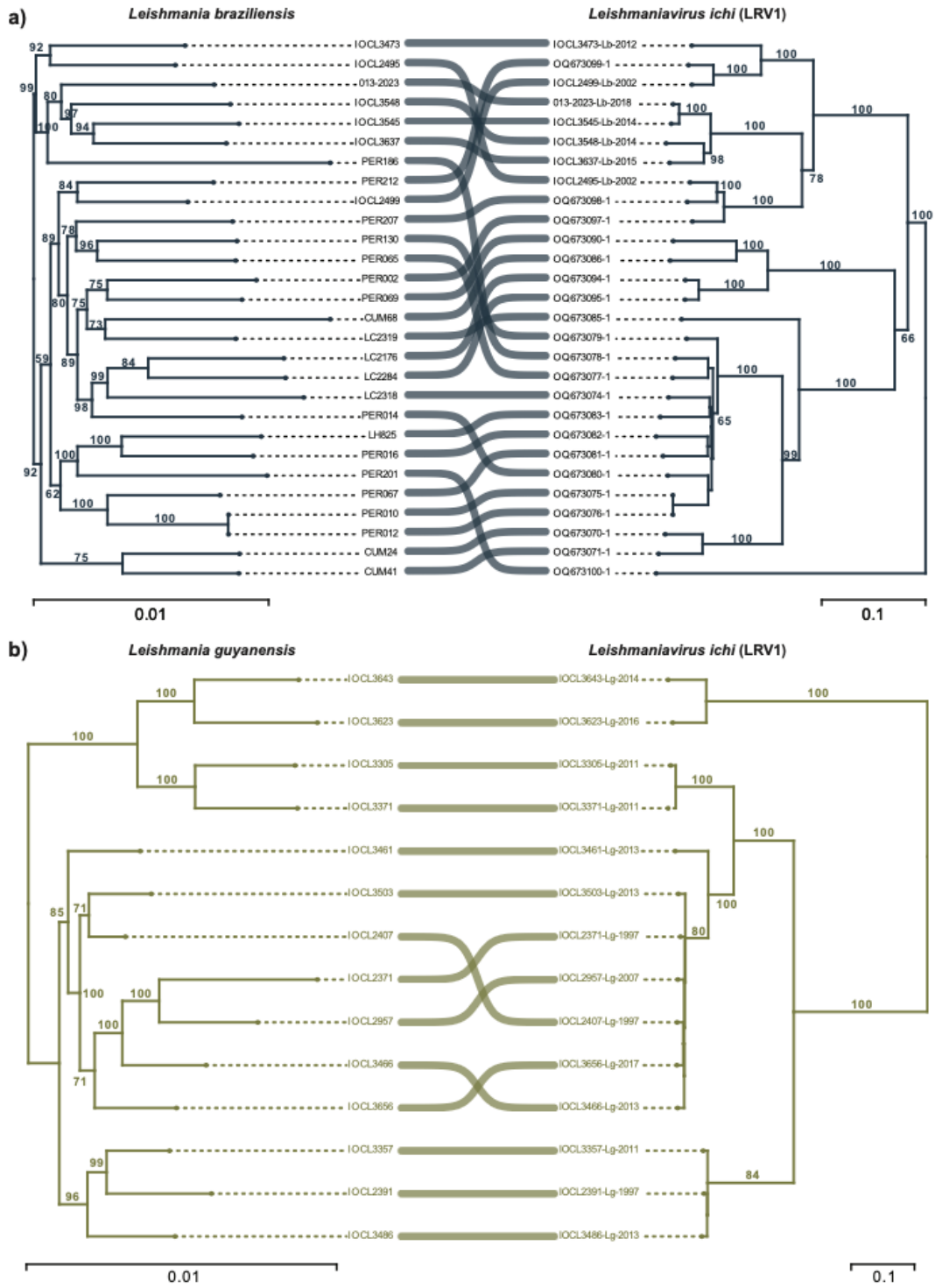
S. Fig. 4: Event-based co-phylogenies of the *Leishmania* divergence, across different subgenera (*L.* (*Viannia*), *L.* (*Leishmania*), *L.* (*Sauroleishmania*)), and LRV species (LRV1, LRV2). Each reconciliation was calculated with different sets of event-costs (S. Fig. 3a; S. Tab. 11): **A)** set 1, **B)** set 2, **C)** set 3, **D)** set 4, **E)** set 5, **F)** set 6, **G)** set 7. Support values for the inferred events were based on 100 bootstrap replicates.



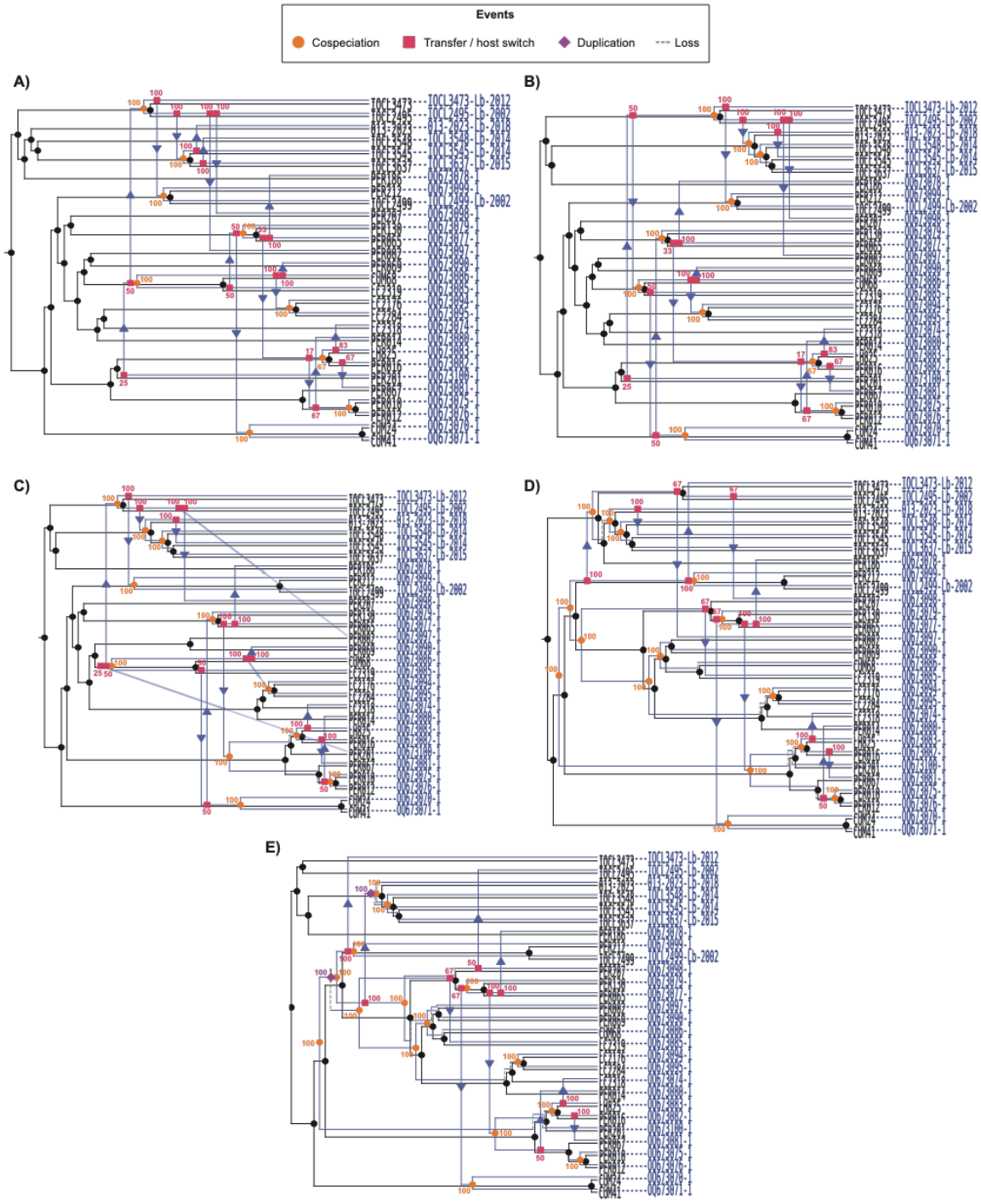
S. Fig. 5: Tanglegram of the *Leishmania Viannia* subgenus and LRV1 phylogenies based on 45 LRV1-positive *L. Viannia* isolates. The *Leishmania* phylogeny was inferred from an alignment of 97,530 concatenated SNPs, corrected for ascertainment bias. The LRV1 phylogeny was inferred from a concatenation of the two protein-coding gene sequences, CP and RDRP (ORF2 and 3, respectively). Colors on the tree are set for visual purposes and do not serve any information on the ancestral states (i.e. internal nodes). Branch values represent bootstrap support values from 10,000 ultra-fast approximated bootstrap replicates. The scale bar depicts the number of substitutions per site.



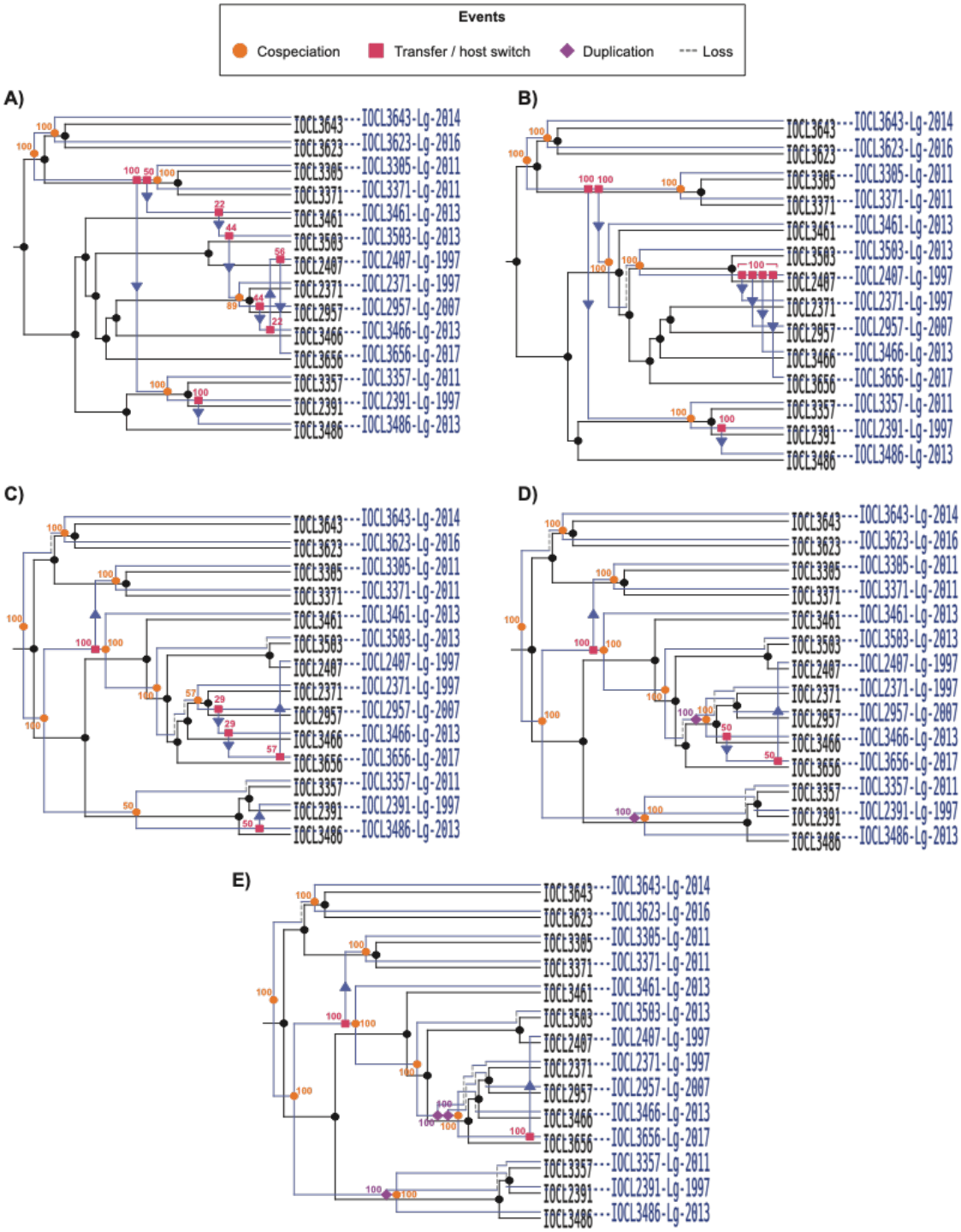
S. Fig. 6: Event-based co-phylogenies of *Leishmania Viannia* and LRV1. Each reconciliation was calculated using different sets of event-costs (S. Fig. 3b;S. Tab. 12): **A)** set 1, **B)** set 2, **C)** set 3, **D)** set 4, **E)** set 5, **F)** set 6. Support values for the inferred events were estimated from 100 bootstrap replicates.



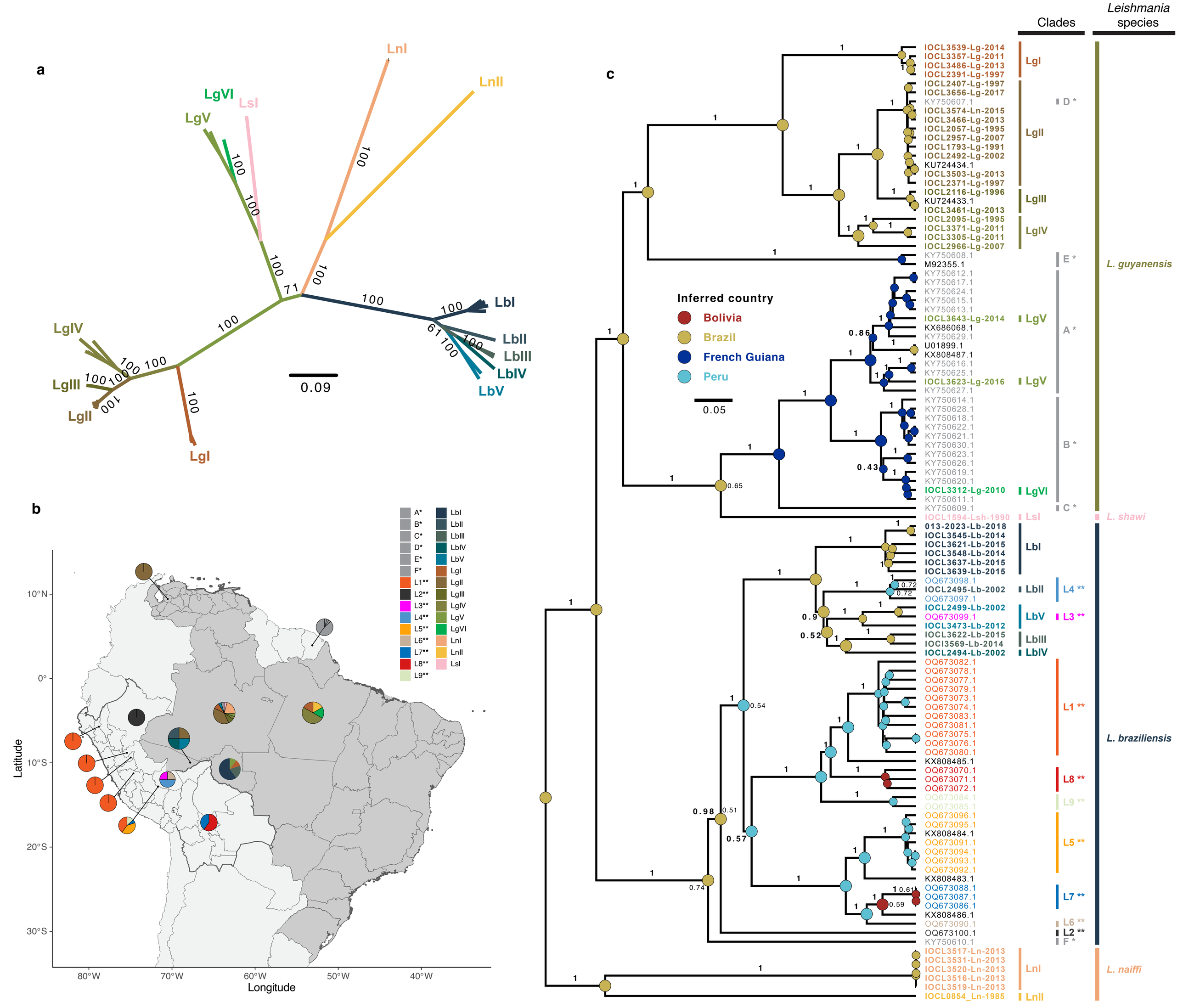
S. Fig. 7: Species-specific tanglegrams of *Leishmania* (*V.*) *braziliensis* (**a**) and *L.* (*V.*) *guyanensis* (**b**) and their associated LRV1 lineages. The *Leishmania* phylogenies represent species-specific partitions of the *L.* (*Viannia*) SNP-based phylogeny (Suppl. Figure 5). Similarly, the LRV1 phylogenies represent specific partitions of the LRV1 phylogeny (Suppl. Figure 5). Branch values represent bootstrap support values from 10,000 ultra-fast approximated bootstrap replicates. The scale bar depicts the number of substitutions per site.

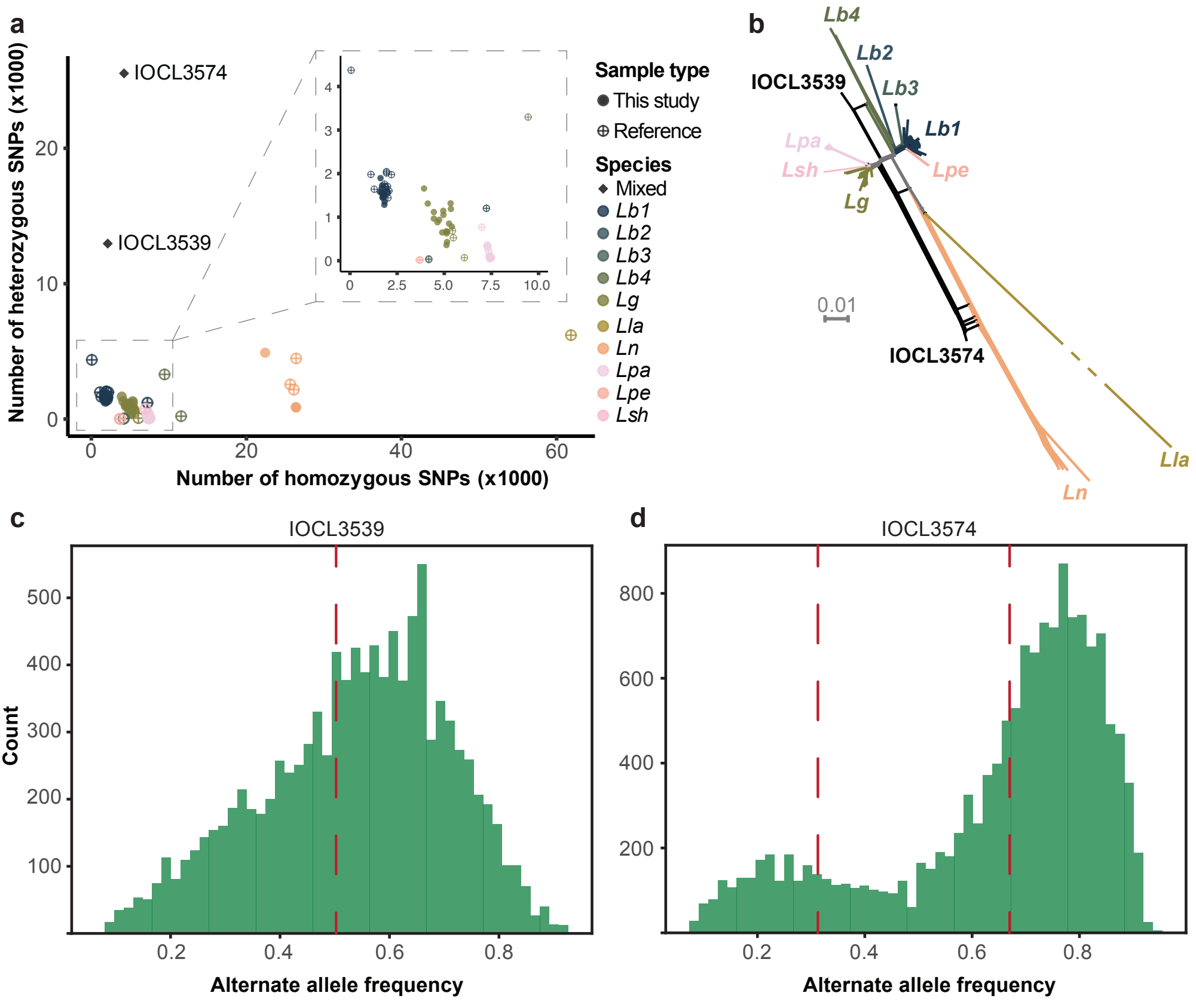


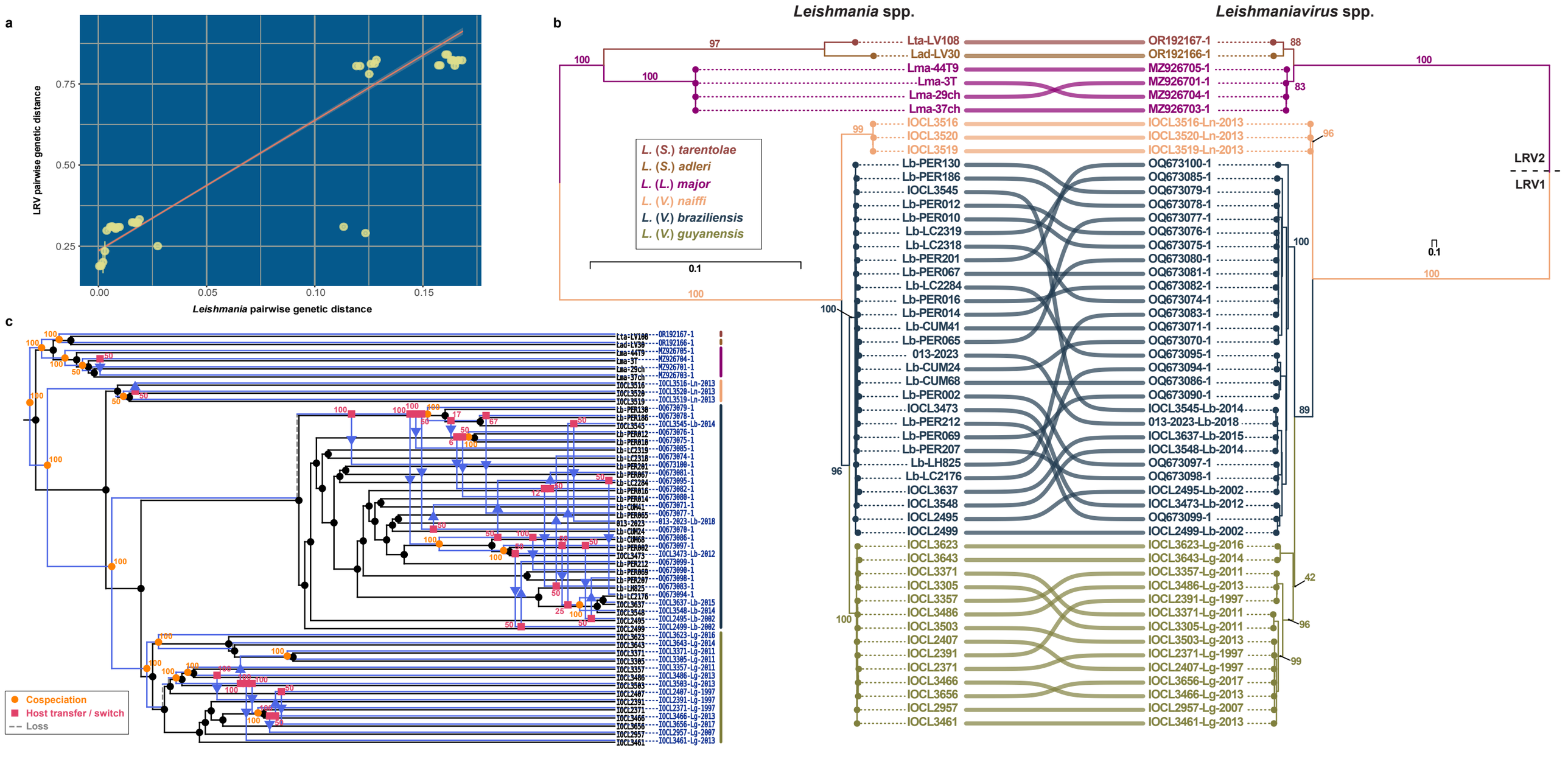
S. Fig. 8: Event-based co-phylogenies of *Leishmania* (*V.*) *braziliensis* and LRV1. Each reconciliation was calculated from different sets of event-costs (S. Fig. 3c; S. Tab. 13): **A)** set 1, **B)** set 2, **C)** set 3, **D)** set 4, **E)** set 5. Support values for the inferred events were estimated from 100 bootstrap replicates.



S. Fig. 9: Event-based co-phylogenies of *Leishmania* (*V.*) *guyanensis* and LRV1. Each reconciliation was calculated using different sets of event-costs (S. Fig. 3d; S. Tab. 13): **A)** set 1, **B)** set 2, **C)** set 3, **D)** set 4, **E)** set 5. Support values for the inferred events were estimated from 100 bootstrap replicates.


a
LnI
LnII
LgVI
LsI
LgV
LgIV
LgIII
LgII
LgI
LbI
LbII
LbIII
LbIV
LbV
0.09
b
A*
B*
C*
D*
E*
F*
L1**
L2**
L3**
L4**
L5**
L6**
L7**
L8**
L9**
LbI
LbII
LbIII
LbIV
LbV
LgI
LgII
LgIII
LgIV
LgV
LgVI
LnI
LnII
LsI
10°N
0°
10°S
20°S
30°S
80°W
70°W
60°W
50°W
40°W
Latitude
Longitude
c
Clades
Leishmania species
Inferred country
Bolivia
Brazil
French Guiana
Peru
0.05
L. guyanensis
L. shawi
L. braziliensis
L. naiffi


a
IOCL3574
IOCL3539
Number of heterozygous SNPs (x1000)
Number of homozygous SNPs (x1000)
Sample type
This study
Reference
Species
Mixed
Lb1
Lb2
Lb3
Lb4
Lg
Lla
Ln
Lpa
Lpe
Lsh
b
Lb4
Lb2
Lb3
Lb1
IOCL3539
Lpa
Lsh
Lg
Lpe
0.01
IOCL3574
Lla
Ln
c
IOCL3539
Count
Alternate allele frequency
d
IOCL3574
Alternate allele frequency


a
LRV pairwise genetic distance
0.75
0.50
0.25
0.00
0.05
0.10
0.15
Leishmania pairwise genetic distance
b
Leishmania spp.
Leishmaniavirus spp.
L. (S.) tarentolae
L. (S.) adleri
L. (L.) major
L. (V.) naiffi
L. (V.) braziliensis
L. (V.) guyanensis
0.1
Lta-LV108
Lad-LV30
Lma-44T9
Lma-3T
Lma-29ch
Lma-37ch
IOCL3516
IOCL3520
IOCL3519
Lb-PER130
Lb-PER186
IOCL3545
Lb-PER012
Lb-PER010
Lb-LC2319
Lb-LC2318
Lb-PER201
Lb-PER067
Lb-LC2284
Lb-PER016
Lb-PER014
Lb-CUM41
Lb-PER065
013-2023
Lb-CUM24
Lb-CUM68
Lb-PER002
IOCL3473
Lb-PER212
Lb-PER069
Lb-PER207
Lb-LH825
Lb-LC2176
IOCL3637
IOCL3548
IOCL2495
IOCL2499
IOCL3623
IOCL3643
IOCL3371
IOCL3305
IOCL3357
IOCL3486
IOCL3503
IOCL2407
IOCL2391
IOCL2371
IOCL3466
IOCL3656
IOCL2957
IOCL3461
OR192167-1
OR192166-1
MZ926705-1
MZ926701-1
MZ926704-1
MZ926703-1
IOCL3516-Ln-2013
IOCL3520-Ln-2013
IOCL3519-Ln-2013
OQ673100-1
OQ673085-1
OQ673079-1
OQ673078-1
OQ673077-1
OQ673076-1
OQ673075-1
OQ673080-1
OQ673081-1
OQ673082-1
OQ673074-1
OQ673083-1
OQ673071-1
OQ673070-1
OQ673095-1
OQ673094-1
OQ673086-1
OQ673090-1
IOCL3545-Lb-2014
013-2023-Lb-2018
IOCL3637-Lb-2015
IOCL3548-Lb-2014
OQ673097-1
OQ673098-1
IOCL2495-Lb-2002
IOCL3473-Lb-2012
OQ673099-1
IOCL2499-Lb-2002
IOCL3623-Lg-2016
IOCL3643-Lg-2014
IOCL3357-Lg-2011
IOCL3486-Lg-2013
IOCL2391-Lg-1997
IOCL3371-Lg-2011
IOCL3305-Lg-2011
IOCL3503-Lg-2013
IOCL2371-Lg-1997
IOCL2407-Lg-1997
IOCL3656-Lg-2017
IOCL3466-Lg-2013
IOCL2957-Lg-2007
IOCL3461-Lg-2013
LRV2
LRV1
c
Cospeciation
Host transfer / switch
Loss