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

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

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 and for the first time, our findings provide 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 long evolutionary 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 the parasite 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 facilitated by several host-encoded proteins (e.g., kinases and proteases) (14), but the exact mechanisms behind this remain unresolved.

In the latest release of the viral taxonomy 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* and *Blechomonas* spp. (2, 10, 11). Within *Leishmania*, different LRV species have different host ranges: *Leishmaniavirus ichi* (referred below to as "LRV1") associated with *Leishmania* (*Viannia*) spp., *Leishmaniavirus ni* (referred below 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 host-specificity (2, 15) and the relatively low sequence identity between different LRV species (below 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 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 a "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 can be found 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 being genome-wide. Our study addresses this gap 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) 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) 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 ones (Tab. 1; S. Fig. 1; S. Tab. 1). These were subjected to total RNA sequencing to extract both the viral genomes and 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 (25). 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 read depth of 370× – 372×, but with a high number of SNPs (over 600). This suggests infection by at least two LRV1 viruses, a phenomenon previously documented in other *Leishmania*-LRV1 systems (3, 26). 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 67.6% to 99.6% of the LRV1-4 genome (5,284 bp; SRA NC_003601) serving as a reference (S. Tab. 2). 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 101× and 1,093×, with an overall median of 391× (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 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, and LbV) clustered with previously described ones (3,

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

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

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

Total RNA sequencing of the LRV1-positive *Leishmania* isolates allowed us not only to assemble LRV1 genomes, but also to genotype the *Leishmania* isolates themselves based on the transcriptomic data. Similar to our previous whole-genome sequencing studies (27, 28), we compiled a dataset of 115 isolates containing (i) genomic data from 66 isolates, generated for this work and (ii) 49 publicly available sequences from various *Leishmania* spp. that served as reference strains during genotyping and species delimitation. This reference panel contained isolates from the *Leishmania* (*Viannia*) *braziliensis* clade (*Lb* L1: n = 18; *Lb* L2: n = 3; *Lb* L3: n = 2; *Lb* L4: n = 2; *L. peruviana*: n = 2), the *Leishmania* (*V.*) *guyanensis* species complex (*L. guyanensis*: n = 3; *L. panamensis*: n = 14; *L. shawi*: n = 1), *Leishmania* (*V.*) *naiffi* (n = 3), and *Leishmania* (*V.*) *lainsoni* (n = 1) (S. Tab. 6) (27).

As our data was derived from RNA sequencing, the genotyping analyses focused on the coding DNA sequence (CDS) regions. Mapping the reads against the long-read-based M2904 *L. braziliensis* reference genome (29) resulted in a median read depth of 17× (min = 1×; max = 151×) across the entire CDS for all 115 isolates. For the isolates from this study alone, the median CDS coverage was 10.5× (min = 1×; max = 43×). Considering a minimum CDS read depth of 10×, we excluded 31 isolates (30 from this study and one from a previous study) from downstream analyses and filtered out 1,184 individual CDS regions below this threshold in the retained samples. The final genotype dataset comprised 6,299 CDS regions across 84 isolates, yielding 157,649 high-quality SNPs.

179 Based on the differences in the number of SNPs (S. Tab. 8) between the included
180 *Leishmania* (*Viannia*) spp. (Chi-squared = 76.92; df = 10; p = 2.01e-12; pairwise Dunn tests: S.
181 Tab. 9), all isolates, with the exception of IOCL3539 and IOCL3574, could be identified (*L.*
182 *braziliensis* L1: n = 12; *L. guyanensis*: n = 18; *L. naiffi*: n = 4) (Fig. 2a). The latter two isolates
183 harbored distinctively higher numbers of heterozygous SNPs compared to all other isolates (Fig.
184 2a; S. Tab. 8) and showed reticulation in the CDS-wide phylogenetic network, a pattern suggestive
185 of recombination and potential hybrid ancestry (Fig. 2b).

186 However, the alternate allele read depth frequency (ARDF) of these two isolates revealed
187 uncommon profiles. In IOCL3539, the ARDF distribution was unimodal and skewed with a mode
188 around 0.66, compared to the typical symmetric distribution around 0.5 (30) (Fig. 2c). The
189 IOCL3574, on the other hand, showed a distribution that was bimodal with modes near 0.25 and
190 0.77. This contrasted with the typical bimodal ARDF distribution for triploid genomes (modes
191 around 0.33 and 0.66) and was similar to the two outer modes that are typical for tetraploid
192 genomes (0.25 and 0.75) but lacking its typical third mode at 0.5 (30) (Fig. 2d).The CDS-wide shifts
193 in the ARDF indicate that the observed hybrid-like patterns are the result of either the extensive
194 mosaic aneuploidy or mixed infections by two different *Leishmania* spp., rather than true
195 interspecific hybrids (31). Moreover, the ARDF assessment on the chromosome-level revealed that
196 both isolates were devoid of homozygous regions for the alternate allele, a pattern refuting the idea
197 that these isolates are the result of true hybridization even more (S. Fig. 3).

198 **Evidence for phylogenetic congruence across different hierarchical levels of *Leishmania***
199 **and LRV**

200 To assess the extent, to which *Leishmania* and LRV have shared a similar evolutionary
201 trajectory, we performed a matrix regression on the pairwise genetic distances of both symbionts,
202 combined with formal co-phylogenetic tests – global-fit and event-based – to detect overall co-
203 phylogenetic signal and evaluate true phylogenetic congruence, respectively (32). These analyses
204 were based on a dataset including 22 Peruvian *Leishmania* (*V.*) *braziliensis* isolates with their
205 associated LRV1 genomes (3), as well as six Old World *Leishmania* isolates (*L.* (*L.*) *major* (n = 4),
206 *L.* (*S.*) *tarentolae* (n = 1) and *L.* (*S.*) *adleri* (n = 1)) with their respective LRV2 genomes (S. Tab.
207 4,6).

208 On the subgeneric level, we observed a highly significant, strong, and positive correlation
209 between pairwise genetic distances across the 51 *Leishmania*-LRV pairs that were analyzed
210 (MMRR: R^2 = 0.89; F = 10,913.49; p < 0.001) (Fig. 3a). In parallel, our array of global-fit analyses
211 detected a consistently significant co-phylogenetic signal between *Leishmania* and LRV at this level
212 (Fig. 3b; S. Tab. 10). Similarly, the event-based inference revealed strong signatures of
213 phylogenetic congruence, suggesting that the two symbionts share a highly similar phylogenetic
214 history (Fig. 3c). This latter analysis was conducted using nine different event-costs sets (S. Fig.
215 4a; S. Tab. 11), with eight of them yielding significant co-phylogenetic reconciliations, where well-
216 supported co-speciation events predominated the divergence of Old and New World *Leishmania*
217 subgenera with LRV2 and LRV1, respectively (S. Fig. 5).

218 When focusing on the species level (subgenus *L.* (*Viannia*)) and its association with LRV1,
219 we observed a highly significant, yet more moderate, positive correlation between the parasite and
220 viral pairwise genetic distances (MMRR: R^2 = 0.59; F = 1,254; p < 0.001). Interestingly, the
221 pronounced host specificity did not appear to be shaped by spatial separation, as viral genetic

222 distances did not correlate with geographic distances between the isolates (MMRR: R^2 = 0.005; F
223 = 5.28; *p* = 0.135). In parallel, both the global-fit and event-based methods consistently detected
224 strong patterns of co-phylogenetic signal and phylogenetic congruence, respectively (S. Tab.
225 10,12; S. Fig. 6). More specifically, the event-based analysis yielded a diverse set of possible co-
226 phylogenetic histories based under six different event-cost scenarios (S. Fig. 4b; S. Tab. 12).
227 Strikingly, in all cases, the inferred phylogenetic reconciliations were significant and, thus,
228 supporting the presence of strong phylogenetic congruence between the members of the *Viannia*
229 subgenus and LRV1 (S. Fig. 7). Another notable observation was that co-speciation events
230 consistently accounted for species divergence (S. Tab. 12) and were well supported, whereas
231 duplication and transfer events appeared to be more predominant within the separate species
232 partitions (S. Fig. 7).

233 Finally, zooming on the population level of this virus-parasite co-evolutionary history, we
234 focused on *L.* (*V.*) *braziliensis* and *L.* (*V.*) *guyanensis* with their respective LRV1 lineages (too small
235 sample size for *L.* (*V.*) *naiffi*). This also revealed signatures of co-evolution, albeit generally weaker
236 compared to the higher levels and occasionally lacking statistical significance. For instance, we
237 found no significant correlation between the genetic distances of *L. braziliensis* and LRV1 (MMRR:
238 R^2 = 0.027; F = 10.65; *p* = 0.215) whilst a weak, though significant, correlation was observed in the
239 case of *L. guyanensis* (MMRR: R^2 = 0.349; F = 47.68; *p* = 0.002). Surprisingly, both global-fit and
240 event-based analyses revealed significant co-phylogenetic signals and evidence of phylogenetic
241 congruence between each *Leishmania* species and its associated LRV1 lineage (S. Fig. 4c-d, 8-
242 10; S. Tab. 10, 13).

243

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 cophylogenetic history of this epidemiologically 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. 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 interspecific hybridization (*L. guyanensis* × *L. braziliensis* L4 and *L. guyanensis* × *L. naiffi*, respectively) or a mixture of these species. Hybridization (both inter- and intraspecific) is an increasingly recognized phenomenon in parasitic protists (33–35) and *Leishmania* in particular (3, 27–29, 36–39). It 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-, bi- and trimodal patterns, suggesting that the genomic signatures of these isolates are more likely the result of a mixed infection or extensive mosaic aneuploidy, rather than true hybrid signals (30, 31). Moreover, these isolates appeared to be devoid of distinct homozygous regions for the alternate allele and showed heterozygous stretches across the entire genome, contrasting with other recently reported hybrid signatures (27, 31), which, again, strongly suggests that both isolates are mixtures of two *Leishmania* spp. 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 an increased intra-clade diversity within groups previously described, based on whole-genome data (3, 26) and 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. This was supported 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 presumable 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 viruses and their hosts, often based on a few genes. Each *Leishmania* subgenus (except *L.* (*Mundinia*), which harbors a phylogenetically divergent *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, as certain viral clusters are more frequently associated with particular *Leishmania* species, but not in a strictly co-specific manner.

The main goal of the current study was to formally investigate, for the first time, the co-evolutionary history of *Leishmania* spp. and their persistent viruses across the subgeneric, species, and population level. Overall, it revealed clear genomic evidence of virus-parasite co-divergence

288 across these levels. At the population level of *L.* (*V.*) *braziliensis* and *L.* (*V.*) *guyanensis*,
289 phylogenetic congruences decreased (e.g. due to lower signals or loss of statistical significance).
290 At this level, we observed more instances of host-switching, as reported previously for *L.*
291 *braziliensis* (3), relative to the predominance of co-speciation events at the higher hierarchical
292 levels. These results align well with the seemingly mutualistic nature of the virus-parasite
293 symbiosis, where strong phylogenetic congruence has often been reported (43). Conversely, this
294 view of virus-parasite co-divergence is challenged by cases of horizontal transmission of LRV within
295 *Leishmania* (both inter- and intraspecific) (3, 18–23) and to other trypanosomatids, such as
296 *Blechomonas* spp. (17). These horizontal transfers are likely facilitated by the potential of LRV to
297 exploit the parasite's exosomal pathway (44–47). Nevertheless, our findings indicate that host
298 switching is relatively rare and occurs mainly at the intraspecific level, whereas vertical transmission
299 remains the predominant factor shaping the evolutionary trajectory of this symbiosis (2, 10).

300 While this study expands our current understanding of the co-evolutionary dynamics of
301 *Leishmania* and LRV, it does not elucidate their full evolutionary history. A major limitation is the
302 discrepancy between the number of paired genomes in the Old- and New World regions. Although
303 our analyses revealed strong signatures of phylogenetic congruence at the subgeneric level, these
304 patterns warrant validation on a larger sample size of the Old World isolates. Another unresolved
305 question concerns the absence of LRV in the divergent subgenus *Mundinia,* where it is still
306 unknown whether LRV-presence is a plesio- (ancestral) or apomorphic (derived) trait. In other
307 words, LRV could have been either lost in *L.* (*Mundinia*) or gained after its divergence from the
308 most recent common ancestor of the other three subgenera (*Leishmania, Viannia*, and
309 *Sauroleishmania*) (10). Elucidating the broader virus-parasite interplay is also becoming
310 increasingly complex as more LRV species are described and recognized. One additional viral
311 species, *Leishmaniavirus sani,* is associated with *L.* (*L.*) *aethiopica* and three divergent species
312 have been described from the flea-infecting trypanosomatids of the genus *Blechomonas* (11, 17).
313 Moreover, LRV1 is not the sole RNA virus infecting *L.* (*Viannia*) spp. (48). Accounting for this
314 growing co-evolutionary and taxonomic complexity, it is likely that additional, more cryptic, viral
315 diversity remains undetected in these protistan parasites. For each of these virus-parasite
316 interactions, it will, therefore, be essential to evaluate and monitor the consequences for the
317 mammalian hosts, including humans, infected with such virus-positive parasites.

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

329 Our less conventional use of RNA sequencing for, e.g., co-phylogenetic analyses and
330 *Leishmania* genotyping, proved to be a useful and cost-efficient way to simultaneously sequence
331 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 research fields like molecular surveillance.

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 occurring 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).

Materials and Methods

Parasite culturing & RNA sequencing

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

Leishmaniavirus de novo assembly & phylogenetic reconstruction

Prior to the de novo assembly of viral genomes, raw reads from each of the 66 LRV-positive *L. Viannia* sequenced libraries were filtered and trimmed using fastp v.0.23.4 (57), and mapped against the *L. braziliensis* M2904 reference genome using SMALT v.0.7.6 (available at: https://www.sanger.ac.uk/tool/smalt-0/). Only unmapped reads were retained and processed with Viralmetagenome v.0.1.2 (available at: https://github.com/nf-core/viralmetagenome) for de novo reconstructions of viral genomes (25). In summary, with this workflow, we combined Megahit and Spades assemblies, followed by a BLAST search to extract contigs matching publicly available LRV1 genomes (58–60). Next, all "raw" assemblies were iteratively refined by re-mapping reads to the assembled genomes and subsequent variant and consensus calling, making use of BCFtools and iVar (61, 62). The reconstruction of hybrid consensus genomes was skipped (--skip_hybrid_consensus: true). A detailed overview of the parameter settings is provided in a supplementary table (S. Tab. 14). Viralmetagenome was able to resolve a consensus genome for 49 of the 66 LRV1-positive isolates. Additionally, we conducted a post-hoc assessment of the assembly quality which involved re-mapping of the reads against their respective assemblies, and variant calling. In total, 42 assemblies were retained based on the following criteria: length > 3,000bp, read depth > 100×, and low SNP counts (i.e., <100 heterozygous and zero homozygous SNPs). Finally, whole-genome sequences, as well as sequences for the Capsid Protein (CP) and the RNA-dependent RNA Polymerase (RDRP), were aligned using the L-INS-I algorithm, as implemented in MAFFT v.7.490 ('--localpair' option) (63).

Phylogenetic reconstructions of LRV1 were generated using IQtree v.2.3.6 (64) over 10,000 bootstrap replicates using the ultrafast bootstrap approximation method (65). Substitution models were automatically selected by IQtree's ModelFinder algorithm (-m MFP) (66), which selects the best substitution model based on the Bayesian Information Criterion after evaluating all models that are implemented in IQtree. Whole genome maximum likelihood phylogenies of the assembled LRV1 genomes were generated with and without a set of 64 publicly available genomes (S. Tab. 4) using GTR+F+I+R4 and GTR+F+I+G4, as respective substitution models. Similarly,

394 gene specific phylogenies, including the 64 publicly available sequences, were inferred using the
395 GTR+F+I+R4 substitution model for both CP and RDRP. For the clade delimitation of the 42 newly
396 assembled genomes, we also calculated the pairwise genetic distances using the 'dist.dna' function
397 of the R-package ape v.5.8-1 (67) (model = 'raw'). Finally, recombination was tested by means of
398 PHI tests (Pairwise Homoplasy Index), as implemented in SplitsTree v.4.17.0 (68, 69).

399 In addition to regular phylogenetic reconstructions, we also attempted to reconstruct
400 ancestral states across the entire LRV1 phylogeny, focusing on the virus' geographic origin and
401 associated *Leishmania* species. This was achieved by adopting a Bayesian phylogenetic
402 framework using the BEAST v.1.10.4 software package (70) and the BEAGLE likelihood calculation
403 library (71). Consistent with previous analyses (3), no temporal signal was detected in the data.
404 Therefore, all sampling dates were set to zero, after which LRV1's discrete phylogeography was
405 inferred without time calibration, using the default strict molecular clock model. The nucleotide
406 substitution model was set to GTR with estimated base frequencies and the site heterogeneity
407 model set to a Gamma distribution with 4 discrete rate categories and accounting for invariable
408 sites (72–74). The tree prior model was set to a speciation model, the Yule process (75), starting
409 the analysis with a random starting tree (default setting). Two discrete traits were added: the
410 country of origin and associated *Leishmania* species. For both traits, we selected the symmetric
411 discrete trait substitution model along with the BSSVS (Bayesian Stochastic Search Variable
412 Selection procedure) option in order to reconstruct their ancestral states. Prior and operator settings
413 were kept at their default settings. The MCMC chain was run for ten million states, sampled every
414 1,000 states and providing a 10% burn-in. Throughout the inference process, convergence of
415 parameter estimates was evaluated using BEAST's auxiliary program Tracer v.1.7.2 (76). Upon
416 completion of the runs, all effective sample sizes (ESS) of all parameters exceeded the
417 recommended threshold of 200, except for the C-T transition rate in the GTR model, which reached
418 an ESS of 110. Finally, the posterior tree distribution was summarized as a maximum clade
419 credibility (MCC) tree using BEAST's TreeAnnotator.

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

421 Total RNA sequencing reads of the 66 LRV1-positive *Leishmania* (*Viannia*) isolates were
422 mapped against the *L. braziliensis* M2904 (MHOM/BR/75/M2904) reference genome (available at:
423 https://tritrypdb.org/common/downloads/release-65/LbraziliensisMHOMBR75M2904_2019) using
424 SMALT v.0.7.6 (parameters: k = 13, s = 2). Reads with a mapping quality lower than 25 were
425 removed using SAMtools v.1.21 ('-q' option) and read depths across all positions were extracted
426 ('samtools depth -a') (61). This reference genome has been shown to be useful for genotyping not
427 only *L. braziliensis* isolates, but also other *L.* (*Viannia*) species, resulting in highly accessible
428 genomes for downstream analyses (28-30). Because transcriptomic, rather than genomic, data
429 were used, we discarded isolates with low median coverages (<10×) across the coding DNA
430 sequences (CDS) from downstream analyses. Additionally, all CDS regions in every sample, with
431 a coverage below 10× were omitted, ensuring that only transcriptomic regions with sufficient depth
432 were included in the subsequent analyses. Variant calling (i.e., SNPs and INDELs) was caried out
433 using GATK v.4.1.4.1 (77). Here, the initial calling was performed on a per-sample basis using
434 HaplotypeCaller, after which all individual genotype vcf files were combined using CombineGVCFs,
435 and variants were jointly genotyped with GenotypeGVCFs. Next, SNPs and INDELs were
436 separated using SelectVariants, after which both panels were subjected to hard filtering with
437 VariantFiltration. For SNPs, the following filters were applied: (i) QD < 2.0, FS > 60.0, MQ < 40.0,
438 MQRankSum < −12.5 and ReadPosRankSum < −8.0, following GATK's recommendations (78); (ii)

439 GQ < 10; and (iii) QUAL < 100. For INDELs, GATK's recommendations were adopted: QD < 2.0,
440 FS > 200.0, or ReadPosRankSum < −20.0 (78).

441 A phylogenetic network was generated using SplitsTree (68) based on the pairwise
442 uncorrected *p*-distances calculated across 157,649 bi-allelic SNPs. This network was built using
443 the NeighborNet (79) and EqualAngle (80) algorithms, as implemented in SplitsTree. Forty-nine
444 publicly available isolates were included as reference sequences: members of the *L.* (*V.*)
445 *braziliensis* clade, the *L.* (*V.*) *guyanensis* species complex, as well as *L.* (*V.*) *naiffi* and *L.* (*V.*)
446 *lainsoni* (S. Tab. 6). In addition, the alternate allele read depth frequency (ARDF) distributions at
447 heterozygous sites were calculated and visualized to identify, whether potential hybrid isolates were
448 the result from a mixed infection. This was done using a custom python script (vcf2freq.py; available
449 at: https://github.com/FreBio/mytools). Differences in SNP counts between the different species
450 were compared using a non-parametric Kruskal-Wallis (stats R-package (81)) test, followed by
451 pairwise Dunn's tests with additional Benjamini-Hochberg *p*-value correction for multiple testing
452 (FSA R-package (82)).

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

454 Phylogenetic reconciliations of the LRV-*Leishmania* symbiosis were performed at the
455 subgenus (combining Old World and New World species), species (focusing on the South American
456 *Leishmania* (*Viannia*) species), population (focusing on *L.* (*V.*) *braziliensis* and *L.* (*V.*) *guyanensis*
457 with their respective LRV1 lineages), and gene (reconciling the CP and RDRP gene phylogenies
458 within LRV1) levels.

459 For the subgenus level analysis, we included 22 publicly available Peruvian *L.* (*V.*)
460 *braziliensis* isolates and six Old World *Leishmania* isolates positive for LRV2 (S. Tab. 4,6). Instead
461 of using the entire genome sequence, we concatenated the annotated CP and RDRP genes (i.e.,
462 ORFs 2 and 3 in the LRV genome) and aligned their sequences from LRV1 and LRV2 using the
463 MAFFT v.7.490 (63) as described above. The LRV phylogeny was reconstructed using IQtree2
464 with 10,000 ultra-fast bootstrap replicates and the GTR+F+I+R3 substitution model based on the
465 ModelFinder algorithm (64–66). For the maximum likelihood phylogenetic reconstruction of the
466 *Leishmania* phylogeny, which included highly divergent species from different subgenera, a
467 concatenated alignment of six orthologous protein coding genes was used (S. Tab. 15), as in
468 previous publication (83). The sequences of the six target genes were obtained as follows. Reads
469 were trimmed with fastp v.0.23.4 using similar settings used in Heeren et al. (2023): a minimum
470 base quality (-q) of 30; maximum percentage of unqualified bases per read (-u) to 10; quality-based
471 sliding window trimming front-to-tail (-5) and tail-to-front (-3) with a window size (-W) of 1 and mean
472 quality score (-M) of 30; right-end cutting (--cut-right) using 10 bp windows (--
473 cut_right_window_size) with mean quality score (--cut_right_mean_quality) below 30; and retention
474 of reads between 100 (-l) and 150 bp (-b) (3). Trimmed reads were then assembled de novo using
475 Megahit v.1.2.9 (58). Contigs with six target genes were identified using a blastn (60) search and
476 subsequently extracted. These sequences were combined with 222 publicly available ones
477 (accession numbers listed in S. Tab. 15) and aligned using MAFFT, as described above. Finally, a
478 phylogenetic tree was inferred using IQtree with 10,000 ultra-fast bootstrap replicates and
479 TIM2+F+G4 set as optimal substitution model (64–66).

480 For the reconstruction of the parasite phylogenies at the species- and population-level co-
481 phylogenetic analyses, we used concatenated SNP-alignments. Variants were called against the

482 MHOM/BR/75/M2904 reference genome, as described above, resulting in genome-wide alignment
483 containing 97,530 bi-allelic SNPs. From this panel, a phylogenetic tree was inferred using IQtree
484 with 10,000 ultra-fast bootstrap replicates and the TVM+F+R2 as the most optimal substitution
485 model, correcting for ascertainment bias (+ASC) (64–66). For the population-level analyses of *L.*
486 *braziliensis*–LRV1 and *L. guyanensis*–LRV1, species specific partitions of the inferred *L.* (*Viannia*)
487 phylogeny were used. Finally, for the reconciliation of the LRV1 CP and RDRP genes, we made
488 use of the reconstructed phylogenies as described above.

489 **Genetic distance regression**

490 In order to obtain an initial assessment of similarity between the viral and parasite
491 phylogenies, we calculated pairwise genetic distances from the corresponding concatenated
492 sequences using the 'dist.dna' function of the ape R-package v.5.0 (67) (model = 'K80'). These
493 distances were converted into distance matrices and analyzed using Multiple Matrix Regression
494 with Randomization (MMRR) (84) where *p*-values were estimated over 1,000 permutations.

495 **Global-fit co-phylogenetic analyses**

496 The presence of co-phylogenetic signals (i.e., testing whether closely related hosts tend to
497 associate with closely related symbionts) between the parasite and virus phylogenies by making
498 use of three global-fit methods was tested via (i) the Robinson-Foulds distance as implemented in
499 the phytools R-package v.2.4-4 (85), (ii) ParaFit as implemented in ape (67), and (iii) PACo as
500 implemented in the paco R-package v.0.4.2 (86). For each method, both virus and parasite
501 phylogenies were downsampled using the 'drop.tip' function of ape. This way, only the isolates, for
502 which both high quality LRV1 genomes and *L.* (*Viannia*) transcriptome-based genotypes were
503 available. Each approach was performed by running 1,000 permutations for *p*-value estimation and
504 the Cailliez correction method was used in ParaFit and PACo to correct for negative eigenvalues.

505 **Event-based co-phylogenetic analysis**

506 To test whether *Leishmania* and LRV exhibit signs of phylogenetic congruence, at one or
507 multiple hierarchical levels, we made use of eMPRess, an event-based phylogenetic reconciliation
508 tool used for reconciling phylogenies based on a set of events such as co-speciation, loss,
509 duplication, and host-switch (32, 87–89). A key advantage of eMPRess over other event-based
510 reconciliation tools is its ability to specify different sets of event-costs prior to calculating the
511 reconciliation, allowing for the evaluation of alternative evolutionary scenarios for obtaining more
512 robust inferences (87). This is important as different event-costs schemes can lead to distinct
513 solutions and interpretations. For each level, we visualized the cost-space (S. Fig. 4) and selected
514 multiple sets of event-costs covering most of the cost-space (S. Tab. 11-13). Phylogenetic
515 reconciliations were then calculated for each set, and a median MPR (i.e., a median value of the
516 MPR solution space; MPR = Maximum Parsimony reconciliation) was visualized. Significance was
517 assessed using 1,000 random permutations. Visualizations of the cost-spaces and reconciled
518 phylogenies were generated using the eMPRess GUI, while significance testing was conducted
519 with the CLI version ('*p*-value' function with '--n-samples 1000' option).
520

Acknowledgments

This work received financial support from the Directie-Generaal Ontwikkelingssamenwerking en Humanitaire Hulp (DGD) (Belgian cooperation). F.V.d.B. and S.H. acknowledge support from the Research Foundation Flanders (grants 1226120N and G092921N). S.H. additionally acknowledges support from KU Leuven's Internal Funds (PDMt1/25/007). E.C. acknowledges support from Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro: CNE - 204.165/2024; 269.891/2021; ColBio - 210.285/2021; Conselho Nacional de Desenvolvimento Científico e Tecnológico: Research Fellow - 309627/2021-4; PROEP/IOC - 441639/2024-0; INCT-EpiAmo: 465657/2014-1; and the GGBN award: GGI-GGBN-2021-278. L.M.C. acknowledges 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. V.Y. and A.Y.K. acknowledge support from the European Union through the Operational Programme "Just Transition" and the Czech Ministry of Environment (CZ.10.03.01/00/22_003/0000003 LERCO), as well as the Czech Science Foundation (grant 24-10009S).

538 **References**

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

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

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

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

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

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

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

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

558 9. E. Bourreau, *et al.*, Presence of Leishmania RNA Virus 1 in *Leishmania guyanensis*
559 Increases the Risk of First-Line Treatment Failure and Symptomatic Relapse. *J. Infect. Dis.* **213**,
560 105–111 (2016).

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

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

565 12. R. V. H. de Carvalho, *et al.*, *Leishmania* RNA virus exacerbates Leishmaniasis by
566 subverting innate immunity via TLR3-mediated NLRP3 inflammasome inhibition. *Nat. Commun.*
567 **10**, 5273 (2019).

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

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

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

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

577

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

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

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

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

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

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

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

595 24. S. M. Scheffter, Y. T. Ro, I. K. Chung, J. L. Patterson, The complete sequence of
596 *Leishmania* RNA virus LRV2-1, a virus of an Old World parasite strain. *Virology* **212**, 84–90
597 (1995).

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

600 26. S. Tirera, *et al.*, Unraveling the genetic diversity and phylogeny of *Leishmania* RNA virus
601 1 strains of infected *Leishmania* isolates circulating in French Guiana. *PLoS Negl. Trop. Dis.* **11**,
602 e0005764 (2017).

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

605 28. S. Heeren, *et al.*, Evolutionary genomics of *Leishmania braziliensis* across the
606 neotropical realm. *Commun. Biol.* **7**, 1587 (2024).

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

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

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

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

617

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

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

623 35. A. Izeta-Alberdi, C. N. Ibarra-Cerdeña, D. A. Moo-Llanes, J. M. Ramsey, Geographical,
624 landscape and host associations of *Trypanosoma cruzi* DTUs and lineages. *Parasit. Vectors* **9**,
625 631 (2016).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

656

657 50. L. de O. R. Pereira, *et al.*, Severity of tegumentary leishmaniasis is not exclusively
658 associated with *Leishmania* RNA virus 1 infection in Brazil. *Mem. Inst. Oswaldo Cruz* **108**, 665–
659 667 (2013).

660 51. D. H. Macedo, *et al.*, Low frequency of LRV1 in *Leishmania braziliensis* strains isolated
661 from typical and atypical lesions in the State of Minas Gerais, Brazil. *Mol. Biochem. Parasitol.*
662 **210**, 50–54 (2016).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

693 66. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin,
694 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–
695 589 (2017).

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

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

700 69. T. C. Bruen, H. Philippe, D. Bryant, A simple and robust statistical test for detecting the
701 presence of recombination. Genetics 172, 2665–2681 (2006).

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

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

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

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

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

712 75. T. Gernhard, The conditioned reconstructed process. *J. Theor. Biol.* **253**, 769–778
713 (2008).

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

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

718 78. G. A. Van der Auwera, *et al.*, From FastQ data to high confidence variant calls: the
719 Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–
720 11.10.33 (2013).

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

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

726 81. R Core Team, R: A Language and Environment for Statistical Computing. (2024).
727 Available at: https://www.R-project.org/.

728 82. D. H. Ogle, J. C. Doll, A. P. Wheeler, A. Dinno, FSA: Simple fisheries stock assessment
729 methods. R package version 0.9. 4. (2023). Available at: https://CRAN.R-
730 project.org/package=FSA.

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

733

84. 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).

85. L. J. Revell, phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ*, **12**, e16505 (2024).

86. Balbuena, J. A., Míguez-Lozano, R., & Blasco-Costa, I., PACo: A novel Procrustes application to cophylogenetic analysis. PloS One, **8**(4), e61048 (2013).

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

88. 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.

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

Figures and Tables

Pages 23 to 26.

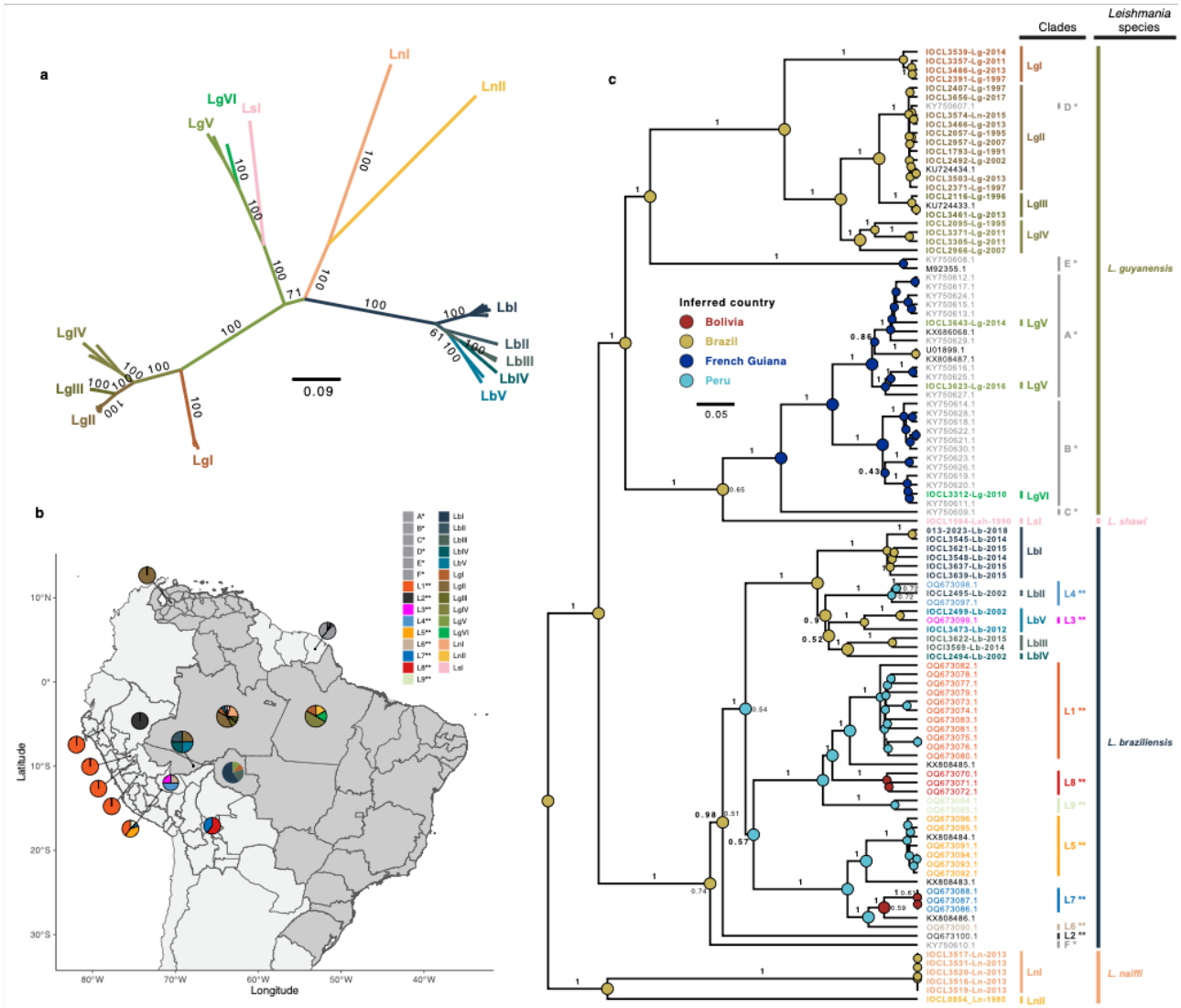


Fig. 1: Lineage diversity and distribution of the LRV1 genomic diversity. a Unrooted maximum likelihood phylogenetic tree based on 42 near-whole genome LRV1 sequences. Branch values represent bootstrap support values based on 10,000 ultra-fast approximated bootstrap replicates. The scale bar depicts 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.

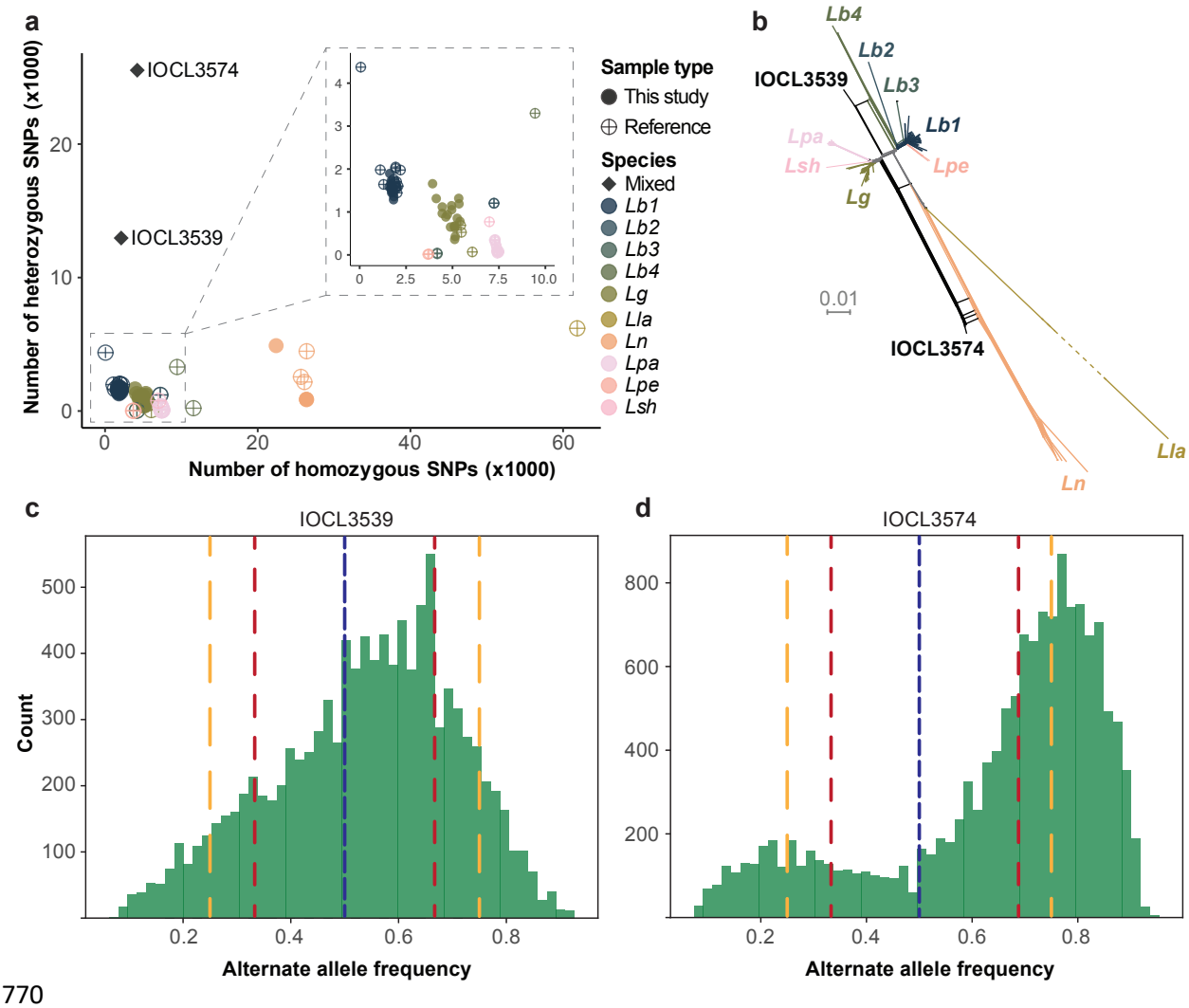


Fig. 2: Species/lineage delimitation of 36 *Leishmania* isolates based on whole transcriptome data (157,649 bi-allelic SNPs). a Quantification of homozygous (x-axis) and heterozygous (y-axis) SNPs for 36 isolates in this study and 48 reference isolates (incl. the *L. braziliensis* clade, the *L. guyanensis* 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 read depth frequency distribution plots of IOCL3539 (**c**) and IOCL3574 (**d**). Dashed vertical lines show the typical modalities for diploid (i.e., a unimodal distribution with a mode around 0.5; blue) and triploid (i.e., a bimodal distribution with modes around 0.33 and 0.66 in red) and tetraploid (i.e., the outer two modes around 0.25 and 0.75 in yellow, and the mode in-between at 0.5 in blue) genomes (31).

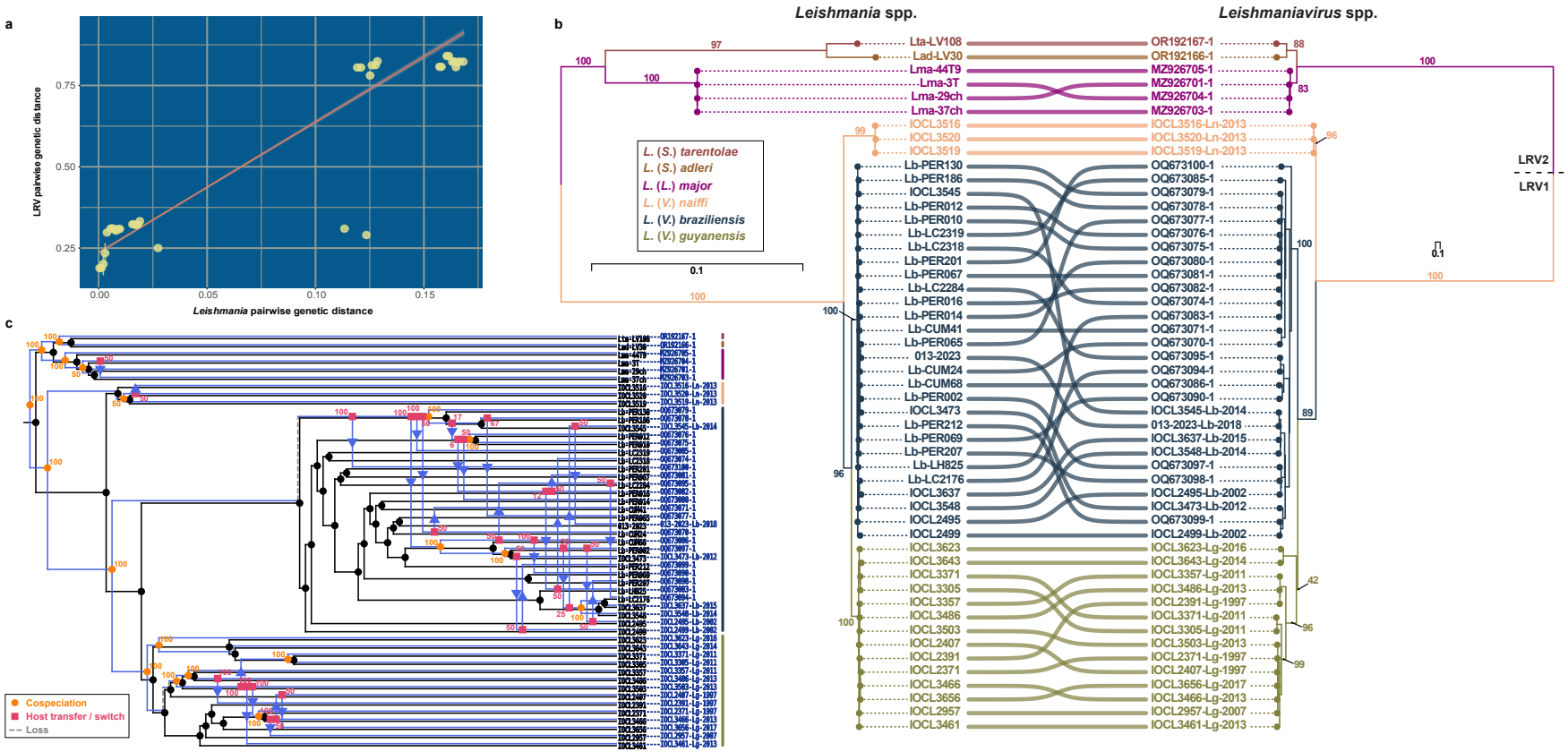


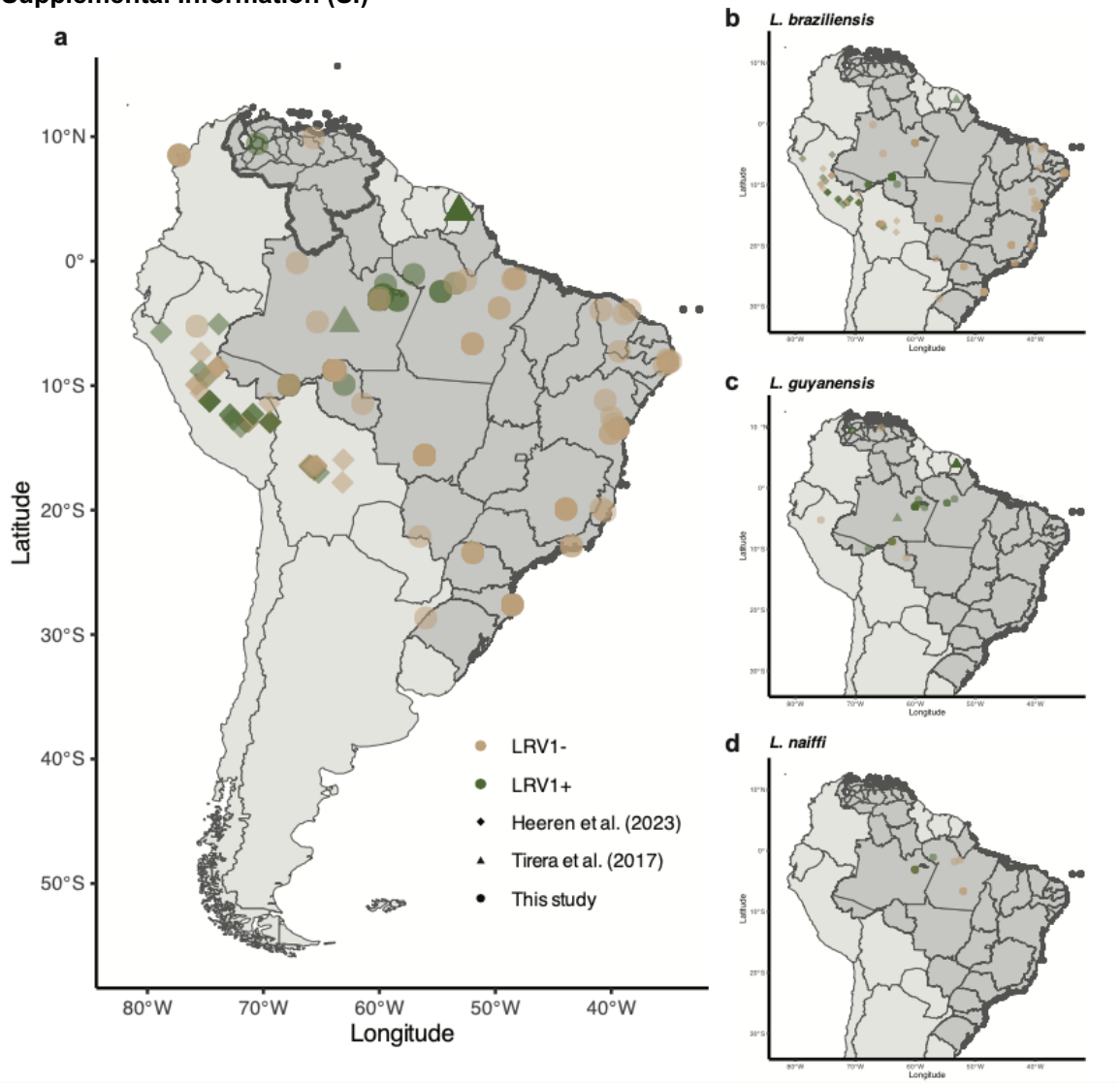
Fig. 3: Co-phylogenetic reconciliation of *Leishmania*-LRV symbiosis at the subgeneric level. a Regression scatterplot of the pairwise genetic distances of *Leishmania* and LRV (51 pairs). Yellow points represent summarized values in 200 bins along the x-axis. **b** Tanglegram of *Leishmania* and LRV 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.

1 **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)**

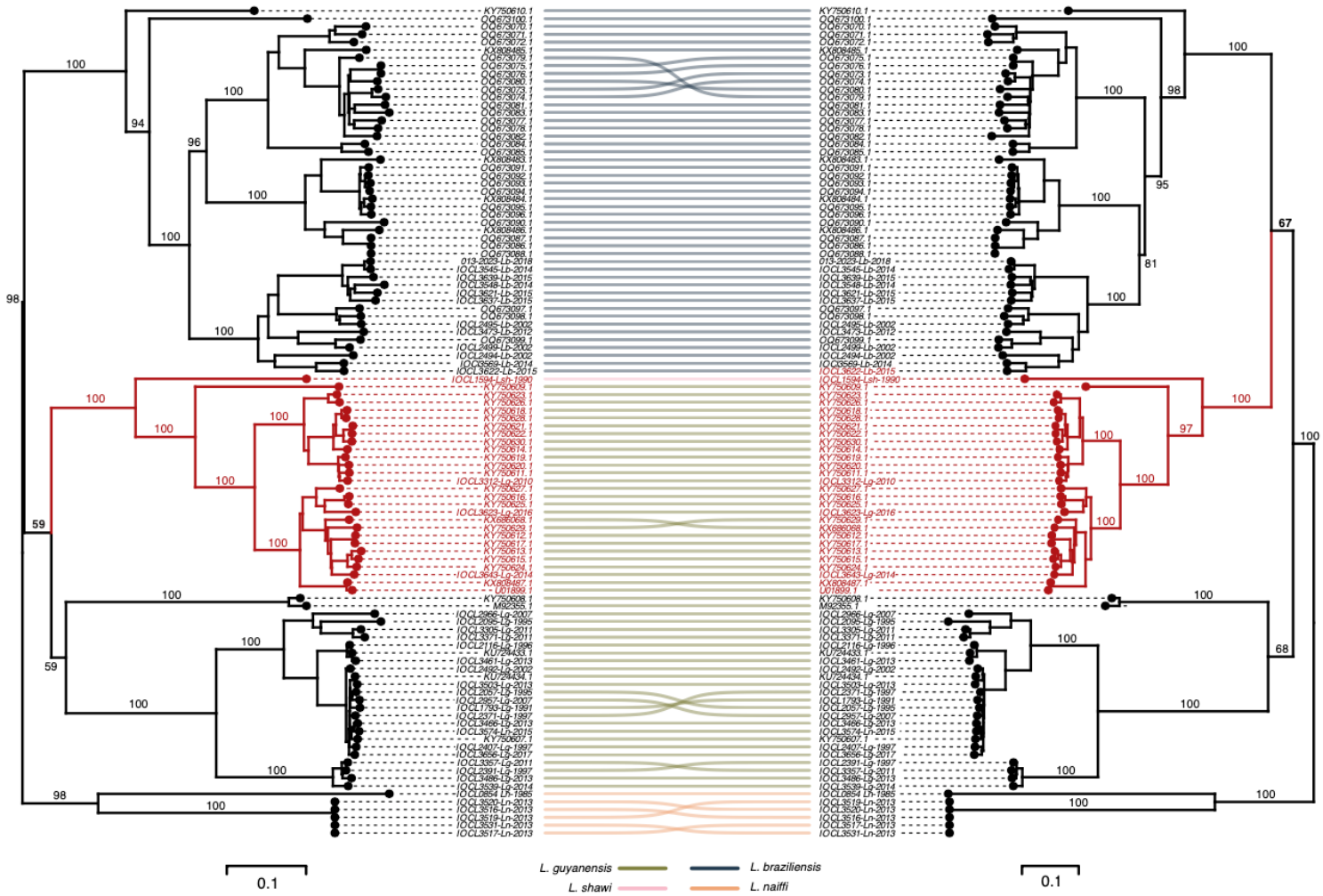
2
3

4 **Supplemental Information (SI)**

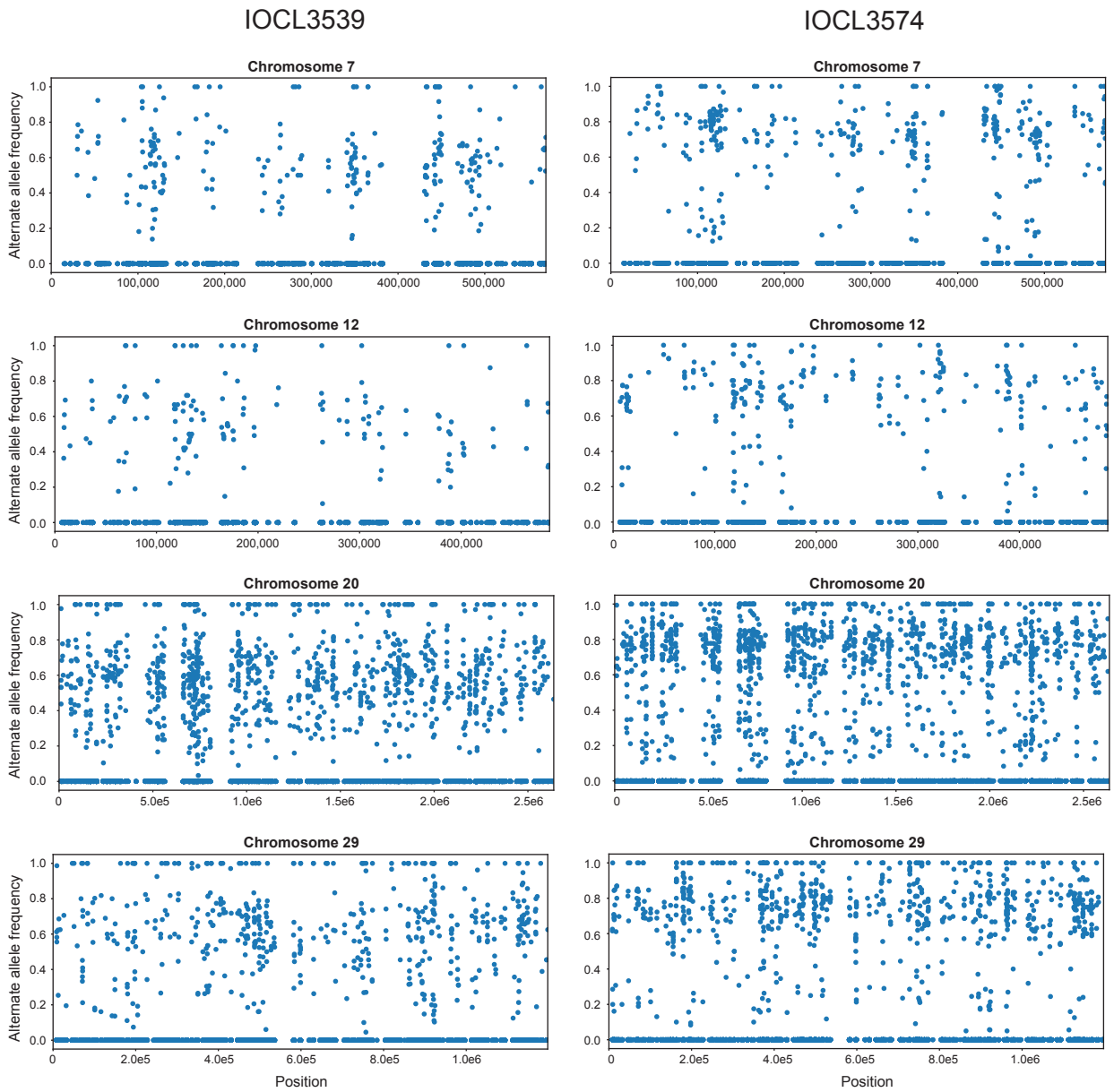


5

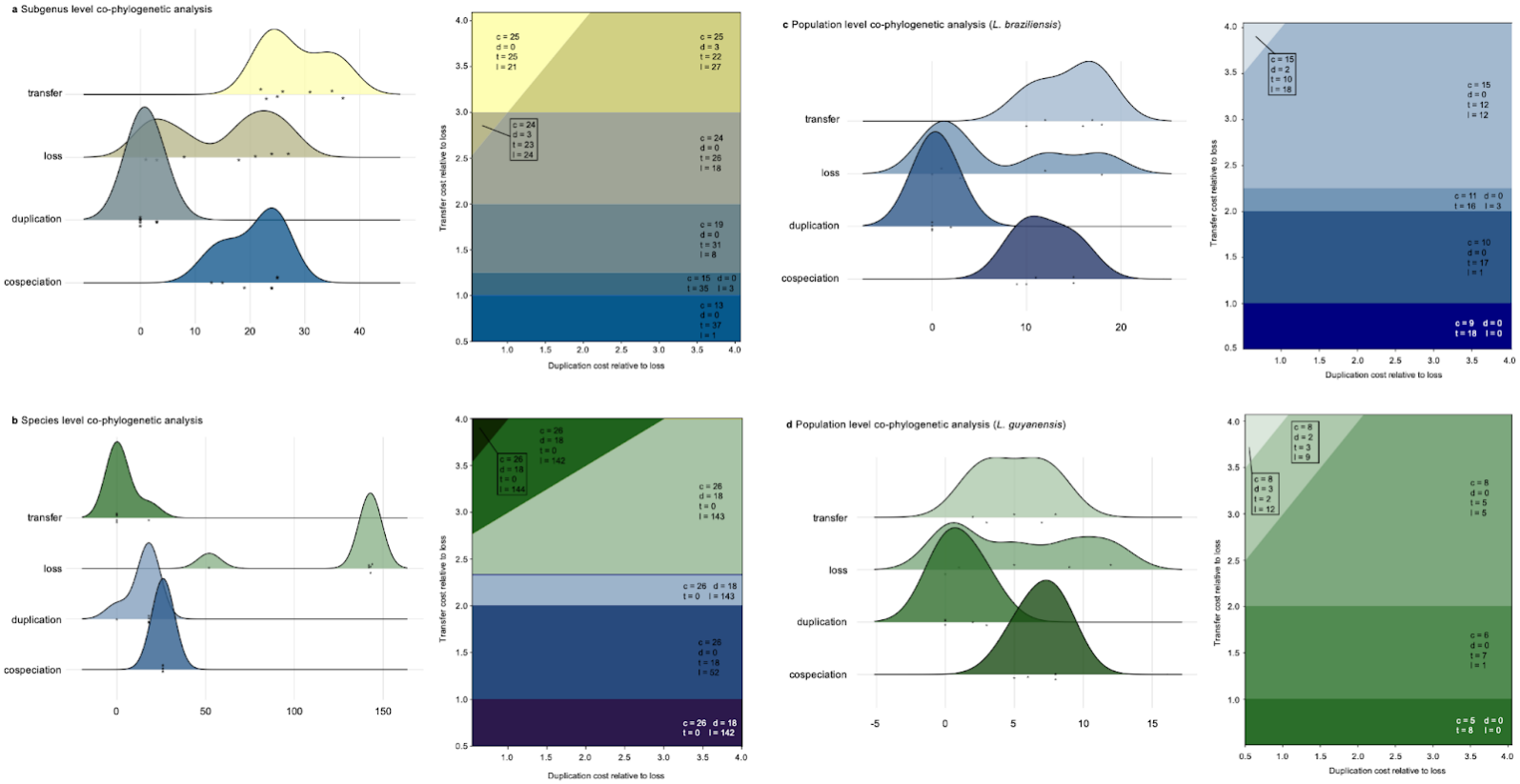
6 **S. Fig. 1.** Spatial distribution of LRV1-positive and negative *Leishmania* isolates. **a** Spatial distribution of all
7 isolates used in this study along with geolocalized *Leishmania* isolates from Heeren et al. (2023) (3) and Tirera
8 et al. (2017) (24). **b-d** Spatial distribution of specific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



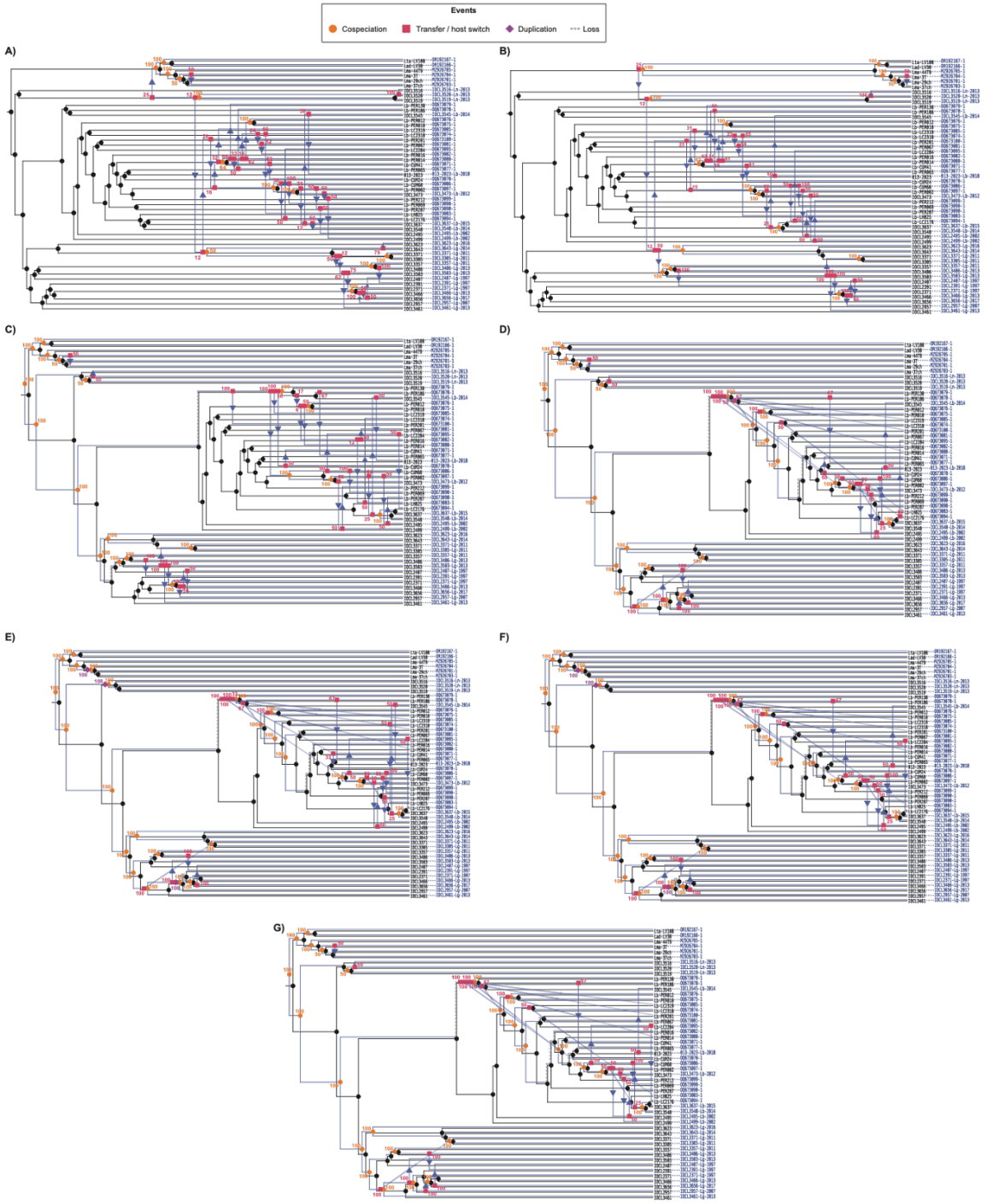
S. Fig. 2: Tanglegram of the LRV1 capsid protein (CP) and RNA-dependent RNA polymerase (RDRP) gene phylogenies. The clade colored in red occupies different positions in the CP and RDRP phylogenies. Branch values represent bootstrap support values based on 10,000 ultra-fast approximated bootstrap replicates. 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).



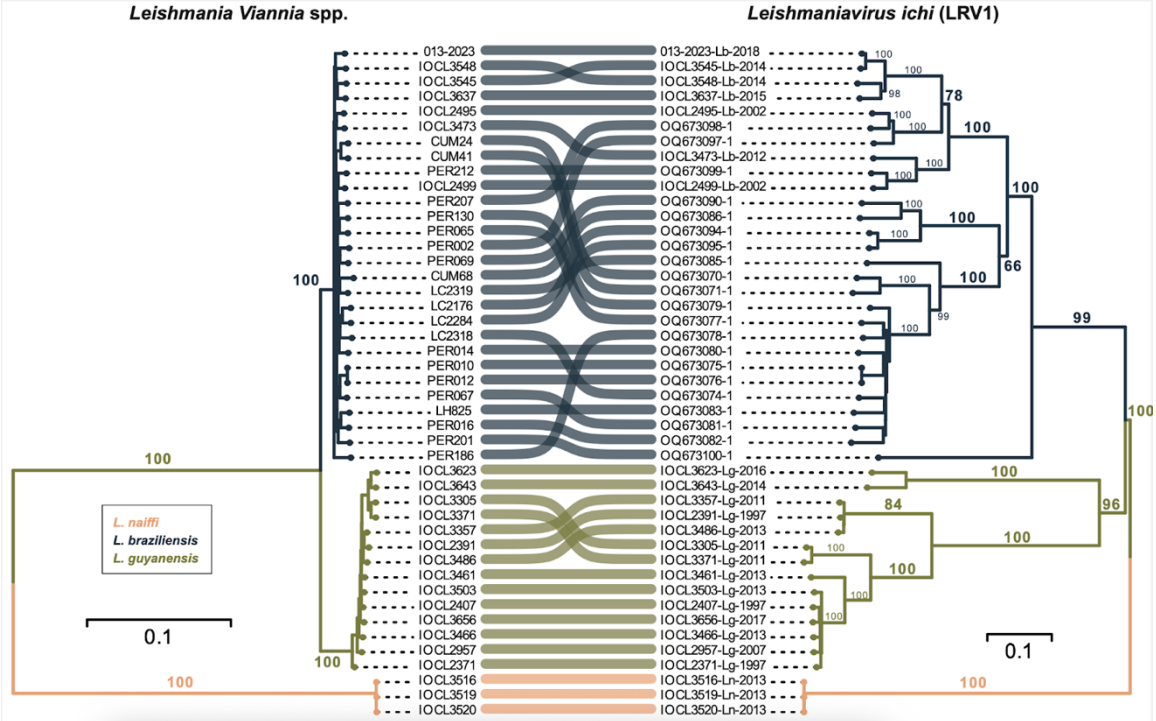
S. Fig. 3: Chromosome-level read depth frequency distribution plots for the alternate allele for IOCL3539 (**left**) and IOCL3574 (**right**). X-axis indicates the position of each along the chromosome, while the y-axis denotes the read depth frequency of the alternate allele that is present in that SNP. Only chromosomes 7, 12, 20 and 29 are shown to exemplify the heterozygous nature of both isolates which are both devoid of homozygous stretches across their genomes.



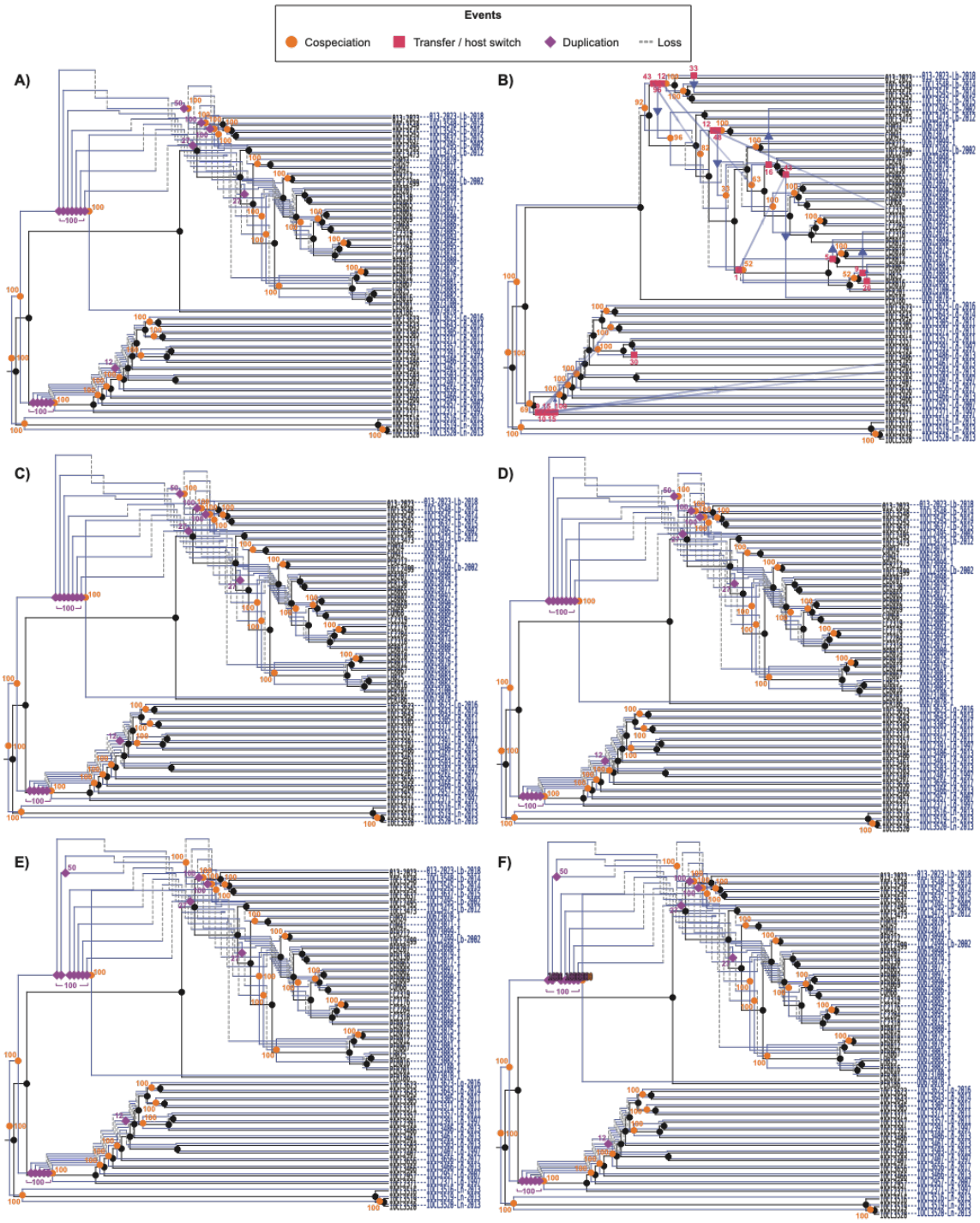
S. Fig. 4: Event-cost distributions and cost spaces from event-based co-phylogenetic analyses of the *Leishmania*–LRV symbiosis across the different hierarchical levels. Per panel, the left-side plot depicts the density of the number of events that have been calculated over all the tested cost-sets; (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*.



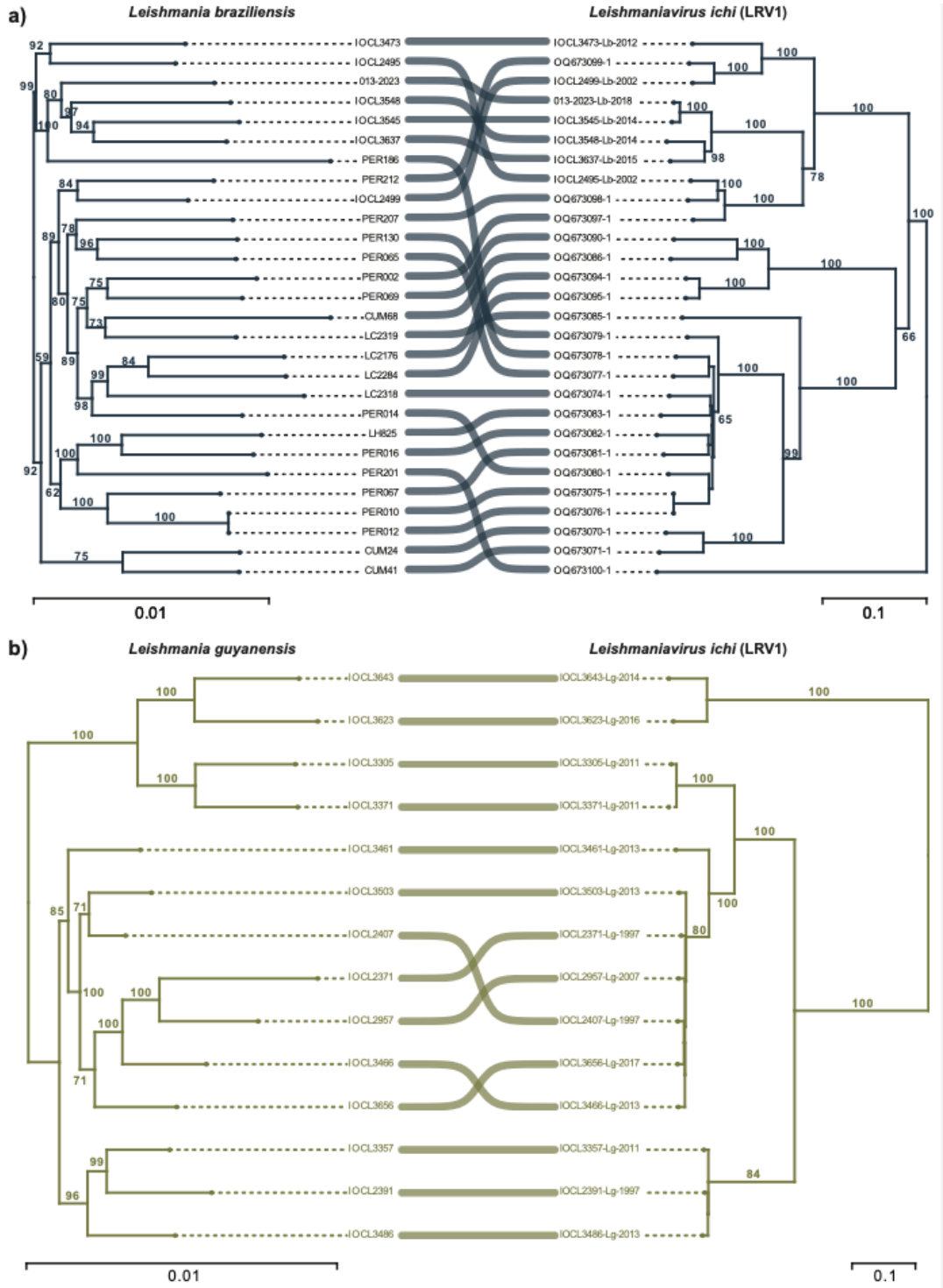
S. Fig. 5: Event-based co-phylogenies of the *Leishmania* divergence, across different subgenera (*L.* (*Viannia*), *L.* (*Leishmania*), *L.* (*Sauroleishmania*)), and LRV species (LRV1, LRV2). Each reconciliation was calculated with different sets of event-costs (S. Fig. 3a; S. Tab. 11): **A)** set 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.



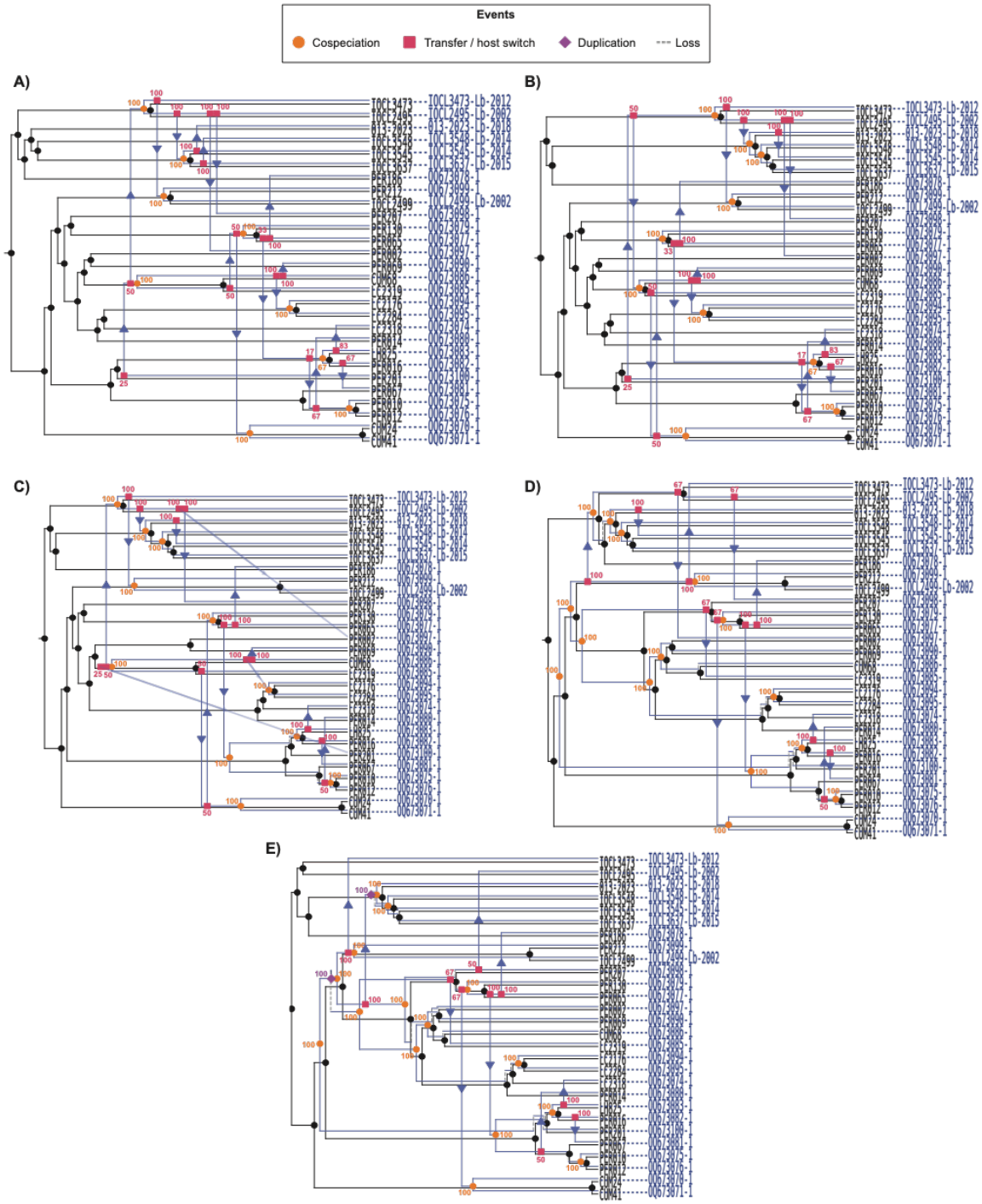
S. Fig. 6: Tanglegram of the *Leishmania Viannia* subgenus and LRV1 phylogenies based on 45 LRV1-positive *L. Viannia* isolates. The *Leishmania* phylogeny was inferred from an alignment of 97,530 concatenated SNPs, corrected for ascertainment bias. The LRV1 phylogeny was inferred from a concatenation 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.



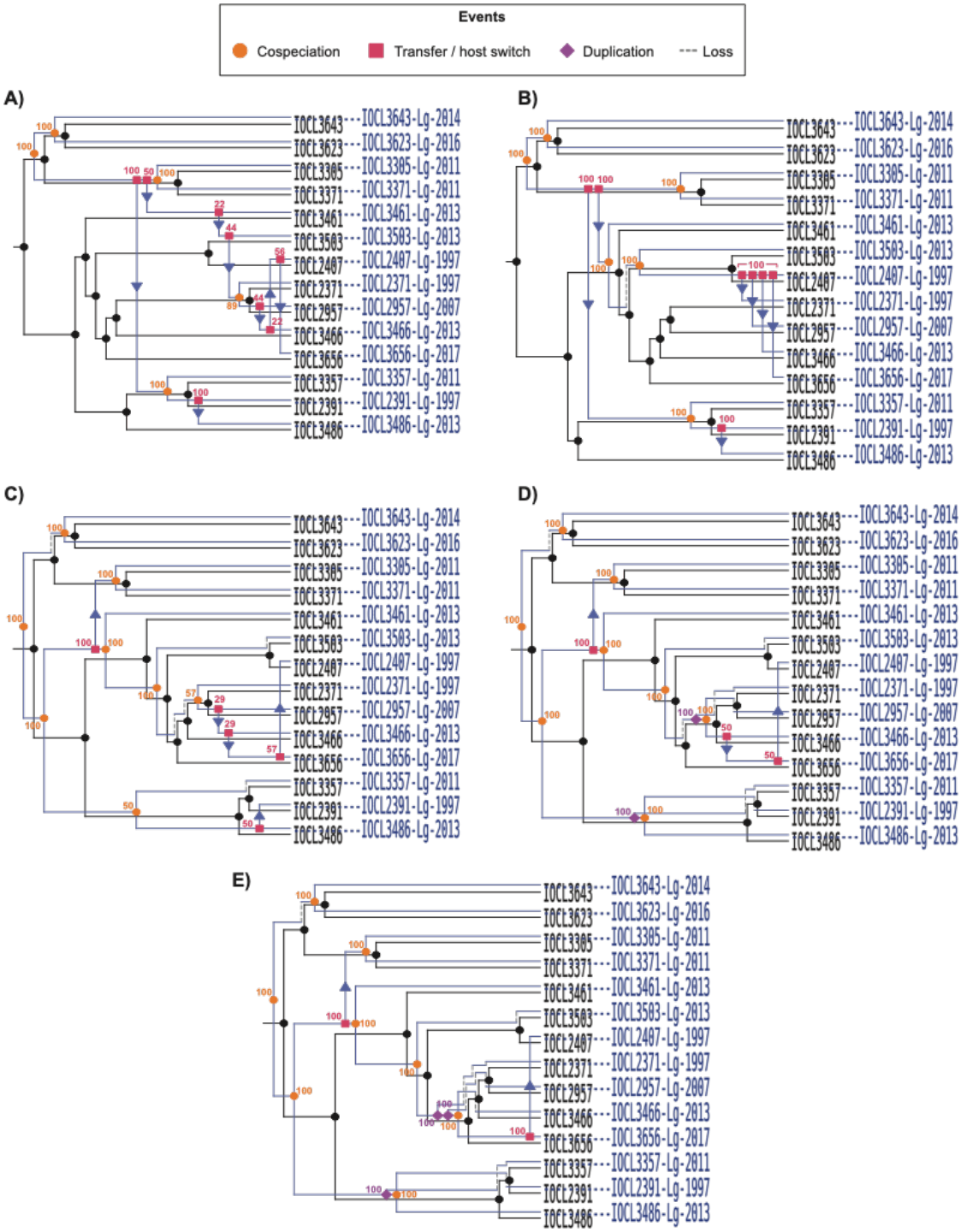
S. Fig. 7: Event-based co-phylogenies of *Leishmania Viannia* and LRV1. Each reconciliation was calculated using different sets of event-costs (S. Fig. 3b;S. Tab. 12): **A)** set 1, **B)** set 2, **C)** set 3, **D)** set 4, **E)** set 5, **F)** set 6. Support values for the inferred events were estimated from 100 bootstrap replicates.



S. Fig. 8: 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.

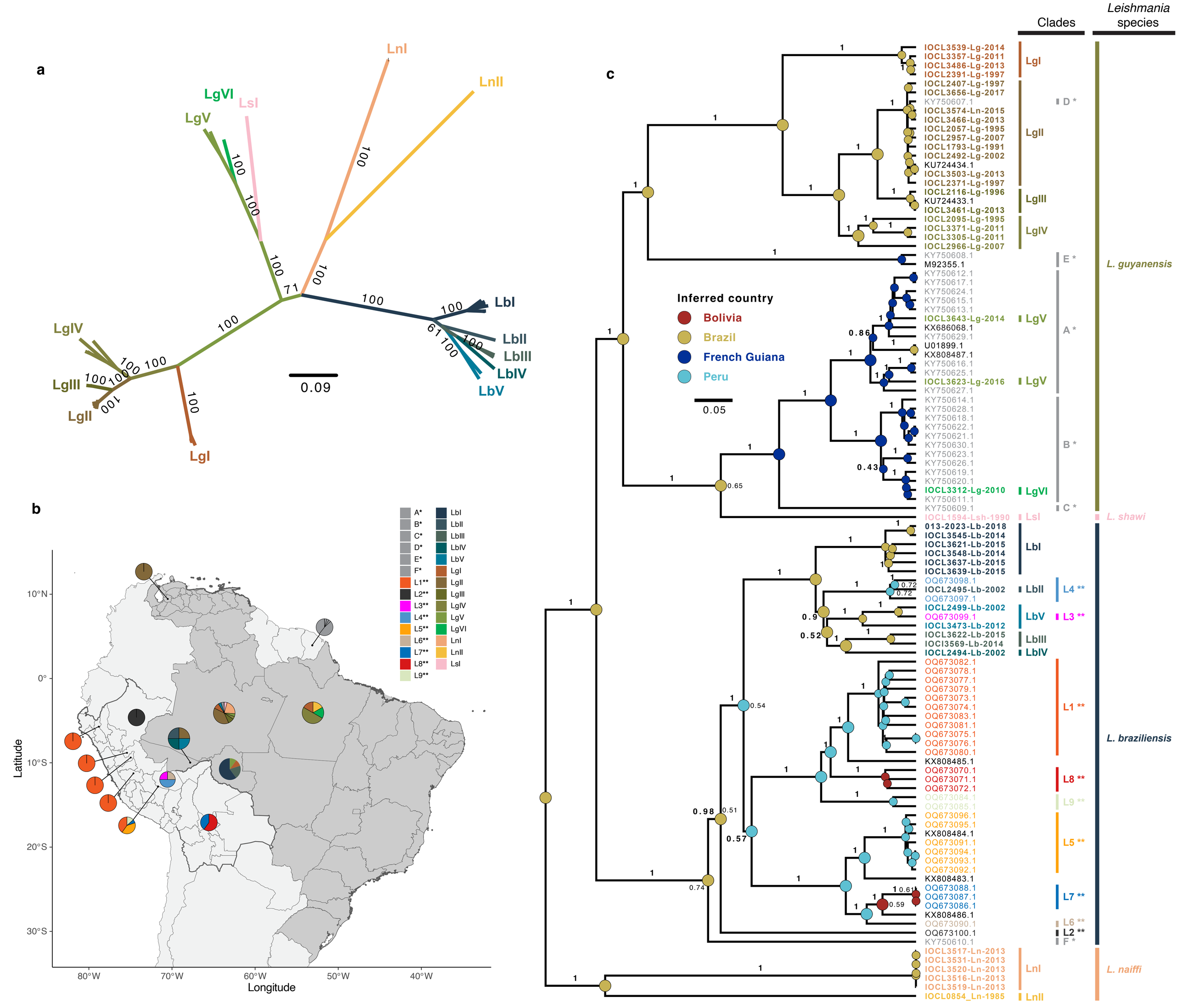


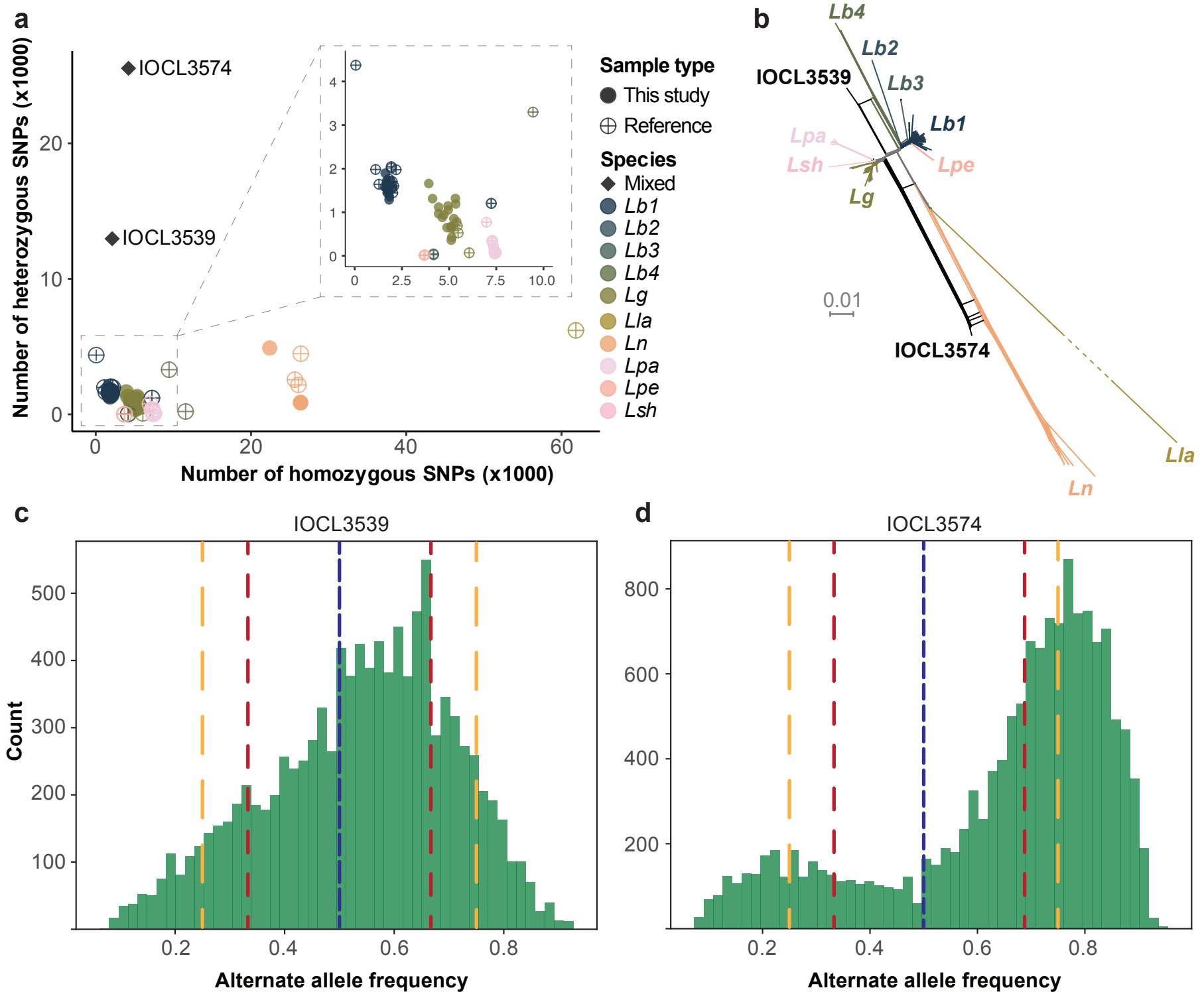
S. Fig. 9: Event-based co-phylogenies of *Leishmania* (*V.*) *braziliensis* and LRV1. Each reconciliation was calculated from different sets of event-costs (S. Fig. 3c; S. Tab. 13): **A)** set 1, **B)** set 2, **C)** set 3, **D)** set 4, **E)** set 5. Support values for the inferred events were estimated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

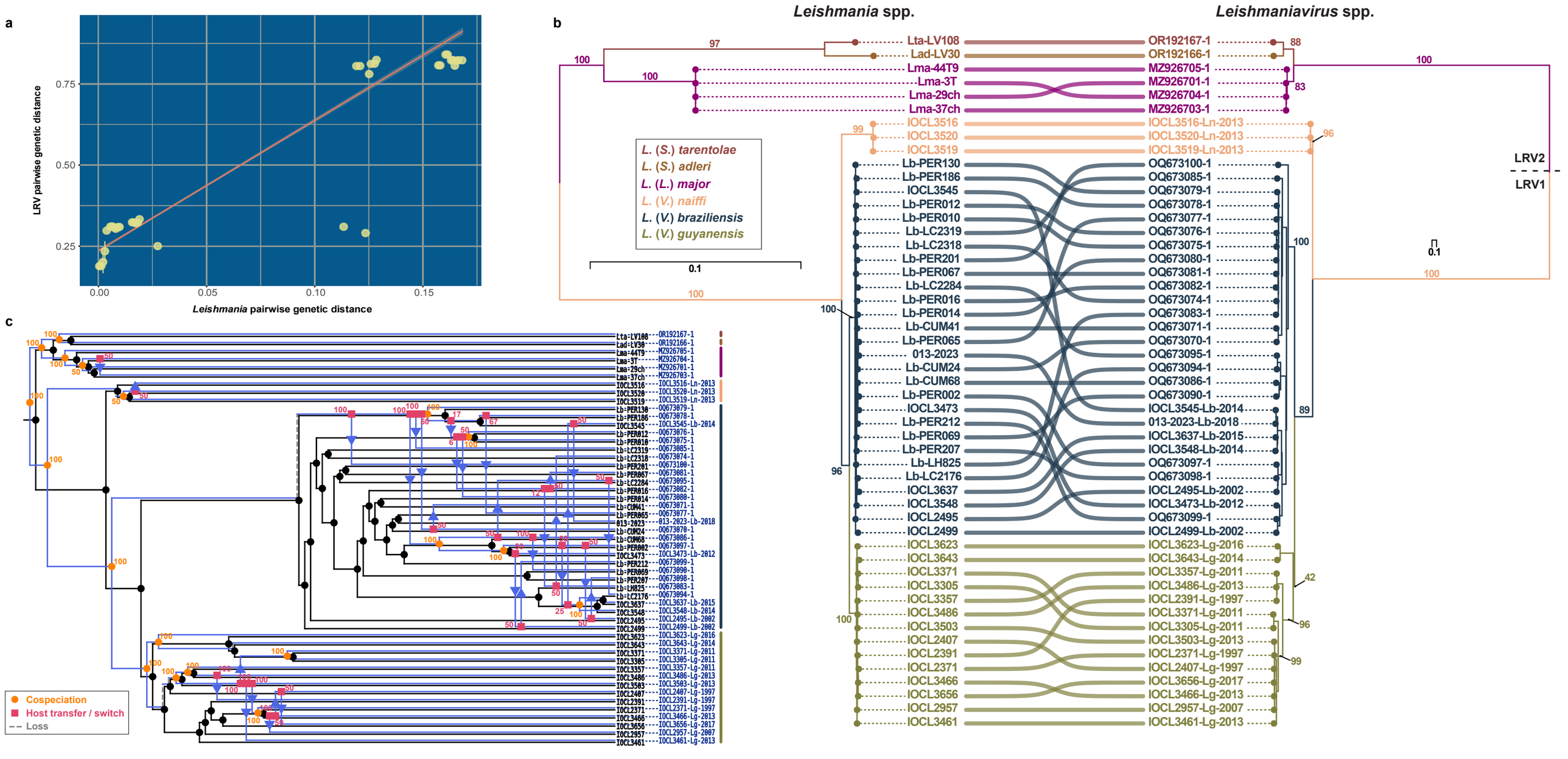


S. Fig. 10: Event-based co-phylogenies of *Leishmania* (*V.*) *guyanensis* and LRV1. Each reconciliation was calculated using different sets of event-costs (S. Fig. 3d; S. Tab. 13): **A)** set 1, **B)** set 2, **C)** set 3, **D)** set 4, **E)** set 5. Support values for the inferred events were estimated from 100 bootstrap replicates.

Supplemental Tables: See excel file.


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