Comprehensive genomic analysis of *Trypanosoma rangeli* reveals key insights into the biology and evolution of this non-virulent American mammalian trypanosome

Guilherme Augusto Maia[1,2*], Gabriel Machado Matos[1,3*], Adriana Corrêa da Silva[1,4*], Caibe Alves Pereira[1], Carime Lessa Mansur Pontes[1], Debora Denardin Lückmeyer[1], Rafaela Andrade do Carmo[4], Renato Simões Moreira[5], Thaynara Karoliny de Souza Pereira[1], Amábili de Souza Rosar[1], Theo Cardozo Brascher[6], Guilherme Razzera Maciel[6], José Franco da Silveira Filho[4], Mário Steindel[1], Patrícia Hermes Stoco[1], Glauber Wagner[1], Björn Andersson[7**], Edmundo Carlos Grisard[1**]

** Correspondence: bjorn.andersson@ki.se; edmundo.grisard@ufsc.br.

Department of Cell and Molecular Biology, Karolinska Institutet, P.O. Box 285, SE-171 77, Stockholm, Sweden. Phone: +46 852483987.

Universidade Federal de Santa Catarina, Campus Trindade. Setor F, Bloco G, Sala G809. Trindade, Florianópolis - SC, 88040-900, Brazil. Phone: +55 48 37214551.

The full list of author information is available at the end of the article

* These authors have equally contributed to this work.

ABSTRACT

Background: *Trypanosoma rangeli* is a non-virulent hemoflagellate protozoan parasite that infects mammals, including humans, in Central and South America. It is primarily transmitted through the bites of triatomine bugs and shares an overlapping geographical distribution with *T. cruzi*, as well as triatomine vectors and mammalian hosts, and various shared surface antigens. The life cycle of *T. rangeli* differs from those of other human-infecting trypanosomes, such as *T. cruzi* and *T. brucei*, and the molecular mechanisms underlying host-parasite and host-vector interactions are not well understood, demanding improved molecular and genomic resources.

Results: The use of a hybrid approach to sequence and assemble the *T. rangeli* genome, complemented by transcriptomics and proteomics for functional gene annotation, led to the generation of the near-complete genome sequence of the parasite. Detailed intra- and inter-specific comparative genomics allowed analysis of polymorphisms, genome structure and improved resolution of genes coding for important surface molecules such as Mucins, TASV and GP63.

Conclusions: The improved *T. rangeli* genome assembly, combined with comparative genomics has yielded novel biological insights. These included the first description of a metalloprotease activity, attributed to specific GP63 genes that are absent in *Leishmania* species. In addition, a TASV gene family that is absent in *T. cruzi* was identified, which indicates a possible role in the *T. rangeli* infection process.

Keywords: Trypanosomiasis, Avirulent trypanosome, Comparative genomics, Surface molecules.

1. BACKGROUND

Trypanosomatids are a group of single-celled, mono-flagellated organisms belonging to the class Kinetoplastea. This group includes etiological agents of neglected tropical diseases of humans such as *Trypanosoma cruzi* (Chagas disease or American trypanosomiasis), *Trypanosoma brucei* (African sleeping sickness), and *Leishmania* spp. (Leishmaniasis) [1]. Collectively, these neglected and socially determined tropical diseases collectively affect more than 18 million people worldwide, mostly in the developing countries [2,3].

Trypanosoma rangeli [4], is a mammalian parasite species found in Central and South America, primarily transmitted via bite of triatomine vectors. Unlike *T. brucei* and *T. cruzi*, *T. rangeli* is considered non-virulent to mammalian hosts, despite the genetic and antigenic similarity with *T. cruzi* and its inoculative mode of transmission [5,6]. On the other hand, *T. rangeli* is harmful to insect vectors, particularly within the genus *Rhodnius* [7–9].

Reports of natural mixed infections by *T. rangeli* and *T. cruzi* have been documented in a variety of mammalian hosts and triatomine vectors species throughout Central and South America [4,7,10–12]. However, the prevalence of human infection by *T. rangeli* remains unclear. The sympatric occurrence, along with shared hosts, vectors and several antigens, leads to the possibility of a strong serological cross-reactivity [13–15].

T. rangeli infection biology is understudied. Previously, comparative analyses of isolated genes coding for surface antigens or related to cell infection or evasion of the immune system, partial genomic [5,16], transcriptomic [17], and proteomic datasets [18] have been produced.

Comparative biological and molecular studies have revealed the presence of trans-sialidase, one of the most important and well-studied virulence factors of *T. cruzi*, and the surface glycoprotein 63 (GP63) metalloprotease (also known as Leishmaniolysin), this major surface protease of *Leishmania* spp. plays a crucial role in cell infection and intracellular survival of both *Leishmania* [19], and *T. cruzi* [20].

The *T. rangeli* genome has been estimated to be around 21-24 million base pairs (Mbp) [5,16] in size, sharing over 90% of the core genome with the pathogenic

trypanosomatids. However, these estimates are based on fragmented assemblies containing 3,436 (AM80) and 9,066 (SC58) scaffolds, respectively. While incomplete, these studies revealed that the *T. rangeli* genome is smaller, less repetitive and with shorter telomeric and sub-telomeric regions compared to the *T. cruzi* genome [21–23]. Additionally, it was proposed that the number of *T. rangeli* genes coding for virulence factors or proteins associated with pathogenesis varies between strains from different lineages and differs from those of *T. cruzi* [24,25]. For instance, compared to *T. cruzi*, *T. rangeli* has a similar number of genes coding for GP63, reduced numbers of genes coding for sialidases/trans-sialidases and Mucin-Associated Surface Proteins (MASP), and, interestingly, an increased number of genes coding for Kinetoplastid Membrane Protein 11 (KMP-11) [5,16].

Trypanosomatid genomes contain syntenic regions where genes are organized in directional transcription clusters, separated by strand-switch regions [5,25–28]. Gene families encoding surface proteins are mostly contained within repetitive regions [24,26,29,30]. Recent proposals suggested that regions of the *T. cruzi* genome that are rich in surface protein-coding genes are involved in generating antigenic diversity through gene recombination and conversion events due to the micro-homology of their flanking regions [23]. Additionally, it has been suggested that DNA replication origins in the *T. cruzi* genome are unevenly dispersed across the genome, as indicated by the enrichment of Orc1Cdc6, a component of the pre-replication complex [31].

Among several aspects of trypanosomatid biology and parasitism that remain to be fully understood, the survival and reproduction of *T. rangeli* within mammalian hosts are of particular interest [32]. For this purpose, the understanding of the complete structure and plasticity of the genome is essential for facilitating comparative genomics of virulent and non-virulent trypanosomatids of mammals. We here describe a far more accurate, near-complete genome sequence *of T. rangeli*, as well as a comparative study of other *T. rangeli* strains. Our findings reveal remarkable biological differences between virulent and non-virulent trypanosomatids, pointing out the *T. rangeli* genome structure, intraspecific variability and the repertoire of surface proteins, providing insights into the parasite biology and life cycle.

2. RESULTS

2.1 T. rangeli genome sequencing and assembly

Illumina sequencing of the *T. rangeli* SC58 strain provided 9,264,332 filtered reads (~4.8Gb) and PacBio sequencing provided 361,938 filtered reads (~6.9Gb), respectively, corresponding to approximately 20x and 192x coverage. The resulting assembly is here referred to as version 2 of the *T. rangeli* SC58 genome ("*T. rangeli* SC58 V.2"), and the fragmented assembly previously published by Stoco et al. (2014) as the "SC58 reference genome". A comparison of general assembly metrics for the *T. rangeli* SC58 V.2 genome and the reference genome assembly is shown in Table 1. The *T. rangeli* SC58 V.2 genome is available in Genbank (Bioproject PRJNA1380247).

Table 1. Comparison of general assembly metrics of the *Trypanosoma rangeli* SC58 reference and V.2 genomes and *T. cruzi* (CL Brener Esmeraldo-like and Sylvio X10/1 strains) genomes.

Footnote: * = Data retrieved from the TriTrypDB v.65. Mbp = Million base pairs. Kbp = Thousand base pairs. aa = amino acids. Scaffold N50 indicates the scaffold length where 50% of the entire assembly is contained in contigs equal to or larger than this value. Scaffold L50 is defined as the count of the smallest number of contigs whose length sum makes up half of the genome size. “Number of hypothetical CDS” represents CDS automatically annotated as “Hypothetical protein”. “ND”= Not determined.

Cross-mapping analysis using the MUMmer4 software indicated that 99.52% of the nuclear genomic content previously described in the SC58 reference genome [5] is included in the *T. rangeli* SC58 V.2 genome. Also, 97.71%, 97.25% and 97.52% of the reads from the PIT10, Choachí and R1625 *T. rangeli* strains, respectively, were mapped to the *T. rangeli* SC58 V.2 genome. Unique copies of all 130 universal single-copy orthologs of the phylum Euglenozoa were identified, distributed across 43 different scaffolds of the *T. rangeli* SC58 V.2 genome.

A high level of synteny was observed among the *T. rangeli* genomes, with few structural variants. However, when comparing the *T. rangeli* SC58 V.2 genome assembly to the assemblies of the *T. cruzi* CL Brener Esmeraldo-like and Sylvio X10/1 strains, significant structural diversity was found (Additional file 1: Supplementary Figure S1).

137 The V.2 genome contained tRNA genes for all 20 amino acids and the canonical
138 5.8S, 18S and 28S rRNA genes (Table 1). The V.2 genome contains fewer tRNA genes
139 than the previous reference genome [5], having a significantly higher number of rRNA
140 genes compared to other trypanosomatid genomes (Table 1). Also, a total of 62 Spliced
141 Leader (SL) RNA genes (Mini-Exon) exhibiting a high level of conservancy were identified
142 (Additional file 2: Supplementary Figure S2A) in five distinct scaffolds, ranging from 9 to
143 16 copies/scaffold (Additional file 2: Supplementary Figure S2B).
144 The search for chromosome-specific probe sequences of *T. cruzi* within the
145 scaffolds of the *T. rangeli* SC58 V.2 genome revealed unique mapping for nine out of 11
146 *T. cruzi* probes (Additional file 3: Supplementary Table S1). When comparing the
147 TcCLB.510129.20 chromosomal probe, derived from a 1.38 Kb-long surface membrane
148 protein identified as a single-copy gene in the *T. cruzi* genome, positive BLAST results
149 were obtained with sequences annotated as “surface membrane proteins” on scaffolds 5
150 and 14 of the *T. rangeli* SC58 V.2 genome. These two scaffolds are of 0.73 and 0.45 Mbp,
151 respectively, and exhibited 85.38% similarity to each other. Fourteen segments of the *T.*
152 *cruzi* TcCLB.510129.20 sequence matched these scaffolds, displaying similarities
153 ranging from 78.36 to 93.51% with 100% coverage. However, the regions located
154 upstream and downstream of these matches were distinct, indicating that this gene is
155 present in multiple copies within the *T. rangeli* genome.
156 Probe TcCLB.511041.40 targeting a gene coding for a 1.63 Kb-long hexose
157 transporter gene in *T. cruzi* and showed sequence similarity with several transporter
158 genes distributed in 10 scaffolds in the *T. rangeli* SC58 V.2 genome. This finding supports
159 the observations from Cabrine-Santos et al. (2009) regarding the presence of duplicated
160 genes encoding hexose transporters on different *T. rangeli* chromosomes. These
161 chromosomes range in size from approximately 500 to 650 Kb, indicating the presence
162 of paralogous genes that are specifically unique to *T. rangeli*.
163 Ten out of the 81 scaffolds of *T. rangeli* SC58 V.2 genome contain canonical
164 telomeric repeats close to their extremities, possibly corresponding to chromosomal ends
165 (Additional file 4: Supplementary Table S2). In accordance with the genome topology
166 proposed by Stoco et al. (2014), subtelomeric regions adjacent to the telomeric repeats

167 were found to contain genes from multicopy families such as (Trans-)Sialidase and GP63
168 (Fig. 2).
169 From the 9,454 CDSs identified in the *T. rangeli* SC58 V.2 genome, approximately
170 32.93% encode for hypothetical proteins. This proportion is comparable to that of *T.*
171 *brucei* TREU927 (34.37%), but notably lower than for *T. cruzi*, for example, the
172 Esmeraldo-like haplotype (47.13%) and the Sylvio X10/1 strain (73.81%). The use of a
173 hybrid assembly and meticulous annotation allowed a reduction of hypothetical proteins
174 compared to previous *T. rangeli* datasets, such as AM80 (56.72%) and the SC58
175 reference genome (67.89%). The annotation of the *T. rangeli* SC58 V.2 genome was
176 supported by transcriptomic [17] and proteomic [18] data generated by our group,
177 demonstrating evidence of transcription and expression for 96 (1.02%) of the hypothetical
178 CDSs. The distribution of all predicted and annotated CDSs in the *T. rangeli* SC58 V.2
179 genome, including notable multigenic families and the 110 potential pseudogenes, is
180 illustrated in Fig. 1. Further analysis of genes associated with the RNA interference
181 machinery (RNAi) on the *T. rangeli* SC58 V.2 genome confirmed the non-functionality of
182 this pathway, as described previously [5].
183
184 **Figure 1**: Distribution of multicopy gene families in the scaffolds of the *Trypanosoma*
185 *rangeli* SC58 V.2 genome. Scaffolds were drawn in proportion to their size, indicated in
186 million base pairs (Mbp). Core genes are represented by black boxes, and the
187 hypothetical and multicopy gene families are represented by colored boxes as indicated.
188 The position of the genes above or below the line corresponds to the direction of
189 transcription, and scaffolds containing canonical telomeric repetitions are shown in detail
190 within the black square. The highlighted region represents an enlarged view (0.05 Mb
191 window) of the telomeric regions of scaffolds 22, 29, and 32.
192
193 *2.2 Repetitive genome content*
194 Excepting telomeric repeats, 11.60% of the *T. rangeli* SC58 V.2 genome consists
195 of repetitive sequences, corresponding to approximately 2.8 Mbp, of which 9.22% (~2.2
196 Mbp) comprises structural repetitive elements, while roughly ~2.38% (565,000 bp)
197 includes members of multicopy gene families. Compared to the TriTryp genomes (*T.*

198 *brucei*, *T. cruzi* and *L. major*), *T. rangeli* has the least repetitive genome when compared
199 to *T. cruzi* CL Brener (37.03%) and Sylvio X10/1 (44.86%) and to *T. brucei* TREU927
200 strain (24.73%). This difference is primarily attributed to a lower number of genes from
201 multicopy gene families, fewer structural repetitive elements and shorter telomeric
202 sequences (Fig. 2a).
203
204 **Figure 2**. Comparison of repetitive genomic content in studied trypanosomatid species,
205 discerning structural repetitions from multicopy gene families.
206
207 Footnote: bp = base pairs. (A) Representation of repetitive genomic content in studied trypanosomatid
208 species, in total number of base pairs. (B) Detailed composition of multicopy gene families in the genome
209 of studied trypanosomatid species, relative to the respective genome size of the species. (C) Detailed
210 composition of structural repetitive elements in the genome of studied trypanosomatid species, relative to
211 the respective genome size of the species. "Ref." points to the reference genome of *T. rangeli* (SC58 strain),
212 obtained from TriTrypDB v65. "V.2" indicates the *T. rangeli* SC58 V.2 genome.
213
214 Comparative analysis of the diversity of multicopy gene families in trypanosomatid
215 genomes showed that the *T. brucei* VSG family is the largest, comprising over 65% of the
216 total number of nucleotides among studied multicopy genes, followed by the RHS family
217 that is also present in *T. cruzi* (Fig. 2b). VSG genes have previously been described in
218 the *T. rangeli* AM80 genome, but these genes were not found in the *T. rangeli* SC58
219 reference genome or in the V.2 genome (Fig. 2b).
220 Among these gene families, the (Trans-)Sialidases are highly represented across
221 the studied genomes, varying from 29.37% and 36.66% of repeated genes in the *T. cruzi*
222 CL Brener and Sylvio X10/1 strains, to 41.96% and 42.74% in the *T. rangeli* AM80 and
223 SC58 V.2 genomes, respectively (Fig. 2b). While *T. cruzi* can transfer sialic acid from the
224 host cell to its cellular membrane through a sialyltransferase (trans-sialidase) activity, *T.*
225 *rangeli* sialidases differ regarding the hydrolytic substrates [33] and lacks the transferase
226 activity [34,35]. We therefore referred to *T. rangeli* sialidases as "(Trans-)Sialidase".
227 GP63 is the second largest multicopy gene family in the *T. rangeli* SC58 V.2
228 genome (35.47%) (Fig. 2b). Uniquely for the American trypanosomatids, *T. cruzi* and *T.*

229 *rangeli,* the DGF-1 multicopy gene family constitutes a considerable percentage of the
230 total repetitive content (Fig. 2b).
231 Interspersed repeats substantially contributed to the repetitive content in the *T.*
232 *rangeli* SC58 V.2 genome (65.10%). In total, this is lower than in *T. cruzi* Sylvio X10/1
233 (81.83%), but considerably higher when compared to simple repeats in the SC58 V.2
234 genome (21.42%) (Fig. 2c). These interspersed repeats are composed of sequences that
235 have been found to be present in all trypanosomatid genomes included in this study.
236 Among these sequences, retroelements constitute 15.89% in the *T. rangeli* SC58 V.2
237 genome, which can be compared to 44.39% in the *T. brucei* TREU927 genome (Fig. 2c).
238

239 *2.3 Analysis of orthology*

240 Orthology analysis using 42,969 CDS from *L. amazonensis*, *L. infantum*, *T. brucei*,
241 *T. cruzi*, and the *T. rangeli* SC58 V.2 and reference genomes resulted in 9,111 CDS
242 clusters and demonstrated that 96.56% of all CDSs from these species are orthologous.
243 The orthologous core for these species constituted 4,501 clusters and the distribution of
244 these clusters between species, as well as the species-specific genes (single-CDS
245 clusters), is shown in Fig. 3.
246

247 **Figure 3**. Distribution of clusters formed among the CDS of *Leishmania amazonensis*, *L.*
248 *infantum*, *Trypanosoma brucei*, *T. cruzi*, the reference and the *T. rangeli* SC58 V.2
249 genomes, as well as their respective singletons.
250

251 Footnote: The result of the orthology analysis among the CDS of five species of trypanosomatids is
252 represented: *L. amazonensis* in blue, *L. infantum* in red, *T. brucei* in green, *T. cruzi* in yellow, and *T. rangeli*
253 in brown. The numbers within the Venn diagram represent orthologous and paralogous clusters, which
254 contain the analysed CDS. The singletons are represented as circles outside the Venn diagram.
255

256 Out of the 298 singletons identified in the *T. rangeli* SC58 V.2 genome, 166
257 (55.70%) were annotated as hypothetical proteins. From this subset, only six had a
258 functional annotation through the assignment of GO terms, and one had experimental
259 evidence of expression [18]. *In silico* characterization of this gene indicates that it encodes
260 a protein of approximately 34.76 kDa that is addressed to the cell membrane of the

261 parasite. This protein was predicted to have a transferase activity (GO:0016740),
262 specifically hexapeptidase transferase (IPR:1837) due to the presence of a conserved
263 amino acid site.
264
265 *2.4 Chromosome copy number variation and somy estimation of T. rangeli strains*
266 Whole genome ploidy was estimated by the determination of the proportion of
267 alleles in heterozygous SNP positions in all scaffolds (Fig. 4a). In all four *T. rangeli* strains,
268 a peak at 50% was observed, which is expected for diploid genomes, suggesting mostly
269 disomic strains [29,36].
270 Chromosome Copy Number Variation (CCNV) in each strain was investigated by
271 combining Read Depth Coverage (RDC) and Allele Balance (AB) analyses [29,36]. Here,
272 the RDC analysis was based on the ratio between the mean coverage of each scaffold
273 and the mean coverage of all scaffolds combined (Fig. 4b). Due to the repetitive content
274 of smaller scaffolds and how repetitive regions lead to erroneous somy estimates [36],
275 CCNV estimation was restricted to longer scaffolds (0.15 Mb) (i.e. scaffolds 1 to 49).
276 Despite applying the size cutoff, some scaffolds still contained too few SNPs for reliable
277 AB estimation (Additional file 5: Supplementary Figure S3). Therefore, aneuploidies were
278 only considered in scaffolds where both RDC and AB results were concordant, providing
279 a two-step confirmation. Aneuploidies, in all cases, chromosome numbers >2, were
280 observed in all strains. Our CCNV analysis combining RDC and AB estimations indicated
281 aneuploidies in scaffolds 7, 11, 12, 24, 29, 30, and 47 in the reference assembly *T. rangeli*
282 SC58 V.2 (Fig. 4b, Additional file 5: Supplementary Figure S3). Similarly, aneuploidies
283 were identified in scaffolds 7, 16, 20, 24, 29, 30, 32, 38, 39, 42 and 47 for strain Choachí,
284 in scaffolds 7, 26, 29, 30 and 47 for strain PIT10, and in scaffolds 24, 29, 30, 39, and 47
285 for strain R1625 (Fig. 4b, Additional file 5: Supplementary Figure S3). Interestingly,
286 scaffolds 29, 30, and 47 were shown to have additional copies in all strains (Fig. 4c).
287 Notably, these scaffolds are enriched with genes involved in immune evasion and
288 parasite survival, including sialidases, GPI-anchored surface protein biosynthesis, cAMP
289 signaling, RNA processing, and lipid metabolism (Fig. 4d, Additional file 5: Supplementary
290 Figure S3).
291

Figure 4: *Trypanosoma rangeli* strains are essentially diploid, with aneuploidies in scaffolds enriched in immune evasion genes.

Footnote: (A) Ploidy estimation based on the proportion of alleles at heterozygous SNP positions across each scaffold. (B) Aneuploidy analysis using relative depth of coverage (RDC) across the 50 longest scaffolds. Dark blue indicates scaffolds where the ratio of median scaffold coverage to median genome-wide coverage exceeds 1.5. (C) Allelic proportions at heterozygous SNP positions for scaffolds 29, 31, and 47, supporting some estimations. Blue points represent allele proportions along each scaffold; black lines indicate median values within sliding windows. (D) Gene ontology (GO) enrichment analysis of scaffolds exhibiting shared aneuploidy across all strains.

2.5 T. rangeli GP63 family

A total of 76 complete GP63 gene sequences and 40 pseudogenes were identified in *T. rangeli* SC58 V.2, with 26 of the pseudogenes containing an internal stop codon and 14 exhibiting truncations. The complete *Tr*GP63 sequences include the full-length peptidase M8 domain (NCBI CDD code: cl19482) as detailed in Additional file 6 (Supplementary Table S3). Most of these sequences resemble the canonical metzincin class zinc proteases motif 'HEXXHXXGX(~60)H', which features a conserved methionine at the base of the active site, referred to as the Met-turn. This motif was first described for GP63 in *Leishmania* spp. Furthermore, the sequences contain 19 cysteine residues distributed throughout [37] (Additional file 7: Supplementary Figure S4).

*Tr*GP63 gene sequences were classified based on a phylogenetic analysis that included sequences from different kinetoplastids (Fig. 5), along with two sequences from each of the 11 GP63 groups of *T. cruzi* formerly described [20]. This analysis led to the classification of at least 13 distinct GP63 groups in kinetoplastids, and *T. rangeli* GP63 sequences were included in 11 of these groups. Interestingly, group 12 was found to be exclusive to this species. *Tr*GP63 sequences do not cluster with the G0 and G1, which include sequences found in most kinetoplastids, where G0 mainly includes *Leishmania* spp. sequences and monoxenous invertebrate pathogens such as *Crithidia fasciculata* and *Leptomonas pyrrhocoris,* and G1 contains some functional *Leishmania* spp. sequences (represented by LmjF-10-0460) [38,39], sequences from *Bodo saltans*, *Angomonas deanei* and sequences from other *Trypanosoma* species. Thus, *Tr*GP63 did

not group with *T. cruzi* GP63-G1, which intriguingly lacks the Met-turn, and this may reduce its activity. Synteny analysis with the corresponding genomic regions of *T. cruzi* DM28c confirmed the absence of these genes in *T. rangeli*, as the eight upstream and downstream genes are conserved. Except for GP63 group 2, members of all other groups were identified in *T. rangeli* SC58 V.2 as well as in the other three *T. rangeli* strains analysed (Additional file 6: Supplementary Table S3).

Inter-group *Tr*GP63 amino acid sequence identities are approximately 35%, whereas intra-group identities exceed 58%, with G11 being the most variable (Additional file 8: Supplementary Table S4). Four groups have a single representative gene in *T. rangeli* (Groups 3, 5, 7, and 8), while G4 is the largest, with 26 genes. Although *Tr*GP63 grouped with most of the *T. cruzi* groups [20], we did not identify active site residue conservation, except in G6, indicating that groups are formed by general sequence similarity. The classification correlated with the genomic distribution of *Tr*GP63, including the same classification of genes in tandem. All *Tr*GP63 in scaffold 3 belong to group G12, while the 13 *Tr*GP63 in tandem in scaffold 13 were from G4. One GP63 gene from G5, outside the tandem locus, was also identified in this scaffold.

Despite exhibiting a variable presence of both signal peptide and a GPI-anchor signature, most *Tr*GP63 were predicted to be N-glycosylated with an average of 0.62% (± 0.42) residues in the mature protein. The majority of *Tr*GP63 groups were found to be addressed to the parasite cell surface, except members of Groups 2, 6, and 8 that consistently lack a predicted GPI anchor, indicating that these may act extracellularly as free proteins. Some members of groups 3, 4, 9, 10, 11, and 12 showed experimental evidence of protein expression on the parasite surface of epimastigotes and/or trypomastigotes [18], whereas members from groups 4, 9, 10, 11, and 12 were potentially secreted, while members of Groups 2, 6, and 7 were exclusively found in the *T. rangeli* secretome (Additional file 6: Supplementary Table S3), corroborating the hypothesis that these groups have a role as secreted glycoproteins. Interestingly, proteomic evidence of expression from some pseudogenes remains to be addressed.

Figure 5. *Trypanosoma rangeli* displays functional secreted and surface-attached GP63 repertoire and species-specific groups from other Kinetoplastids.

Footnote: (**A**) Orthology analysis of GP63 in the Kinetoplastea class. Maximum-likelihood tree inferred from full-length GP63 amino acid sequences. The branch scale is shown as red straight and dashed lines. Colored labels indicating metalloprotease inferred activity based on ligand cavity scores are shown as green (≥ 0.5), yellow (<0.5 to ≥ 0.4), and red (<0.4), representing active, indeterminate and inactive, respectively. Genes organized in tandem by at least three contiguous GP63-coding genes or pseudogenes are indicated by a black line on top of the scaffold number. In case of signal peptide prediction indicating absence in both analyses, GPI-anchor results are considered not applicable, indicated as blue lines. Detailed description of sequences and values is provided in Supplementary Table 5. (**B**) Structural and *in silico* functional analysis of *T. rangeli* GP63, in comparison to *Tc*GP63-G1 (C4B63_30g122) and *Leishmania major* (PDB: 1LML) as control. 3D Model: Box highlights the region corresponding to the Zinc-binding site, where structural and functional analyses were performed. The colors of the structures were assigned according to their classification: green for the highly promising group, yellow for the moderately promising group, and red for the less promising group according to ligand cavity scores established parameters. Zinc-binding site: Amino acid residues with close interaction with the zinc ion are shown in stick representation. Molecular surface: presence of the S1' pocket is demonstrated by a green arrow or absence by a red blocked arrow. P2Rank analysis: Ligand-binding prediction showing predicted cavity positions by different colors. Boxes demonstrate predicted binding cavity probabilities. Molecular docking: Complexes of metalloproteases with the inhibitor 1,10-phenanthroline. The cyan ligand represents the highest-affinity conformation obtained for control structure, while the white ligands correspond to molecular docking results for other GP63. The boxes present the RMSD values (Å) and ΔG (kcal/mol) of the complexes resulting from molecular docking.

Metalloprotease activity was experimentally detected in both epimastigote and trypomastigote forms of *T. rangeli*, with higher activity observed at pH 5.5. The activity was considerably lower in *T. rangeli* epimastigotes compared to *T. cruzi* epimastigotes and *L. braziliensis* promastigotes (Additional file 9: Supplementary Figure S5). To assess the metalloprotease functionality in each *Tr*GP63 group, 15 sequences were chosen to generate three-dimensional models based on the available crystal structure of GP63 (leishmanolysin) of *L. major* (LmjF-10-0460/PDB: 1LML) [37]. Additionally, the *Tc*GP63-G1 (C4B63_30g122) was modelled for comparative purposes, since it was missing in *T. rangeli*. The models showed pTM scores ranging from 0.57 to 0.90, indicating a high predictive quality (Additional file 10: Supplementary Table S5). The *Tr*GP63 metalloproteases showed a molecular organization similar to the control structure, shown by global RMSD values between 9.11 and 24.79 Å, with lower variation observed when

389 considering only helix B, located at the catalytic site of the metalloprotease. Overall, most
390 of the *Tr*GP63 sequences displayed a conserved active/Zinc-binding site motif. Certain
391 genes showed punctual substitutions or several alterations in the key residues as
392 identified in, for example, TrSC58v2_9356 and TrSC58v2_560. Among the 15 modelled
393 structures, only TrSC58v2_9135 (G4) and TrSC58v2_1328 (G12) exhibited the large
394 active-site cleft (S1' pocket) in the surface analysis (Additional file 11: Supplementary
395 Figure S6). Between antiparallel β-sheets in the N-terminal domain, as part of the active
396 site, we identified a loop region with 'GXXAWA' Motif (Additional file 5: Supplementary
397 Figure S3), that is mimetic of the S-loop, a well-characterized conformation in vertebrate
398 metalloproteases [40,41].
399 Considering that the motif conservation and S1' pocket are important for activity
400 inference, and that other parameters may also be important, the interaction potential of
401 the active sites with ligands was assessed using P2Rank [42], which indicated that seven
402 *Tr*GP63 had a ligand cavity probability above 50%, with TrSC58v2_4795 (G4) and
403 TrSC58v2_2748 (G6) showing values close to the control structure (71%, 74%, and 76%,
404 respectively). These also match the experimentally validated GP63 *T. cruzi* sequences
405 (AY266316, AY266317, AY266318) [43]. These genes were classified as a highly
406 promising group (green structures in Fig. 5b, Additional file 12: Supplementary Figure
407 S6). Further six metalloproteases, including the *Tc*GP63-G1, displayed intermediary
408 values and were classified as a moderately promising group (yellow structures). Finally,
409 three metalloproteases showed values below 40% and were classified as a less
410 promising group (red structures). Molecular docking of the Zinc-binding site region with
411 the inhibitor 1,10-phenanthroline [44] was used to infer the metalloprotease activity. In the
412 *L. major* structure used as control, the selected complex displayed the inhibitor oriented
413 toward the Zn^{2+} ion with a probability of 34% and an affinity of -5.54 kcal/mol. The
414 sequences classified as highly promising showed even higher probability of the inhibitor
415 being faced to the cofactor (67 to 44 %) and low RMSD variations relative to the ligand
416 position in the control structure. *Tc*GP63-G1 presented a behavior similar to that of the
417 highly promising *Tr*GP63, regarding ligand orientation towards Zn^{2+} ion (probability: 56%
418 Docking Affinity: -6.03 kcal/mol), while the structures classified as moderately and less
419 promising did not. Fig. 5b, shows the parameters for the *L. major* structure (control),

420 *Tc*GP63-G1, TrSC58v2_4795 (G4), a highly probable protease that was experimentally
421 identified by two different techniques, and represents the biggest *TrGP63* group, and
422 TrSC58v2_560 (G11) which is unlikely to be an active metalloprotease, despite the high
423 level of expression shown by proteomic analysis in both epimastigotes and
424 trypomastigotes (Fig. 5).
425
426 *2.6 T. rangeli surface Mucins/TASV*
427 A comparative analysis between the previous reference genome sequence and
428 the SC58 V.2 genome revealed an increased number of genes belonging to the mucin-
429 like family in the improved genome sequence, where 15 gene copies were found [5,17].
430 A detailed search for known signatures of mucin-like genes, such as the signal peptide,
431 GPI-anchor signal, abundance of serine and threonine residues that are likely *O*-
432 glycosylated, as well as orthology with members of the *T. cruzi* mucin-like gene family
433 (*Tc*MUC) on the SC58 V.2 genome, resulted in the identification of seven genes (*Tr*MUC1
434 - *Tr*MUC7) that are dispersed in three scaffolds of the SC58 V.2 genome (Additional file
435 12: Supplementary Table S6). *Tr*MUC is slightly more similar to *Tc*MUC type I, sharing
436 an identity of 56.88% ± 22.87, when compared to *Tc*MUC type II (44.34% ± 23.46).
437 Interestingly, a conserved C-terminal region of 22 amino acids is present in members of
438 the *Tr*MUC family, but not in mucins from any other kinetoplastid species present in the
439 TriTryp database (Fig. 6a).
440 A gene that encodes a 33 kDa surface protein was described in the *T. rangeli*
441 proteome as a putative mucin-like [18], here we identified it as actually belonging to the
442 *T. cruzi* Trypomastigote, Alanine, Serine and Valine rich protein glycoprotein family
443 (*Tc*TASV), previously only found in *T. cruzi* [45].
444 Based on TASV canonical features that include post-translational modifications
445 (PTM) such as *N*-glycosylation, phosphorylation, and abundance of other amino acids
446 less common in mucin-like proteins, for instance, aromatic residues (phenylalanine,
447 histidine, tryptophan, tyrosine) and cysteine, the SC58 V.2 genome contains 14 TASV-
448 coding genes, and three TASV pseudogenes, dispersed in six scaffolds. Both
449 glycoprotein families are also present in the Choachí and R1625 strains, while *Tr*MUC
450 was not found in PIT10 assembly (Additional file 12: Supplementary Table S6).

451 A phylogenetic analysis using well described protein sequences of mucin-like
452 families (*Tc*MUC type I and type II), *T. cruzi* small mucin-like family (*Tc*SMUG) [46,47],
453 and *Tc*TASV, where certain members were previously annotated as *Tc*MUC [48] is shown
454 in Fig. 6a.
455
456 **Figure 6.** Overexpression of *Trypanosoma rangeli* glycoproteins homologous to
457 *Trypanosoma cruzi* mucin-like and TASV impaired mammalian infection.
458
459 Footnote: (**A**) Phylogenetic analysis of amino acid sequences of *Tr*MUC, *Tr*TASV-like in comparison to
460 *Tc*MUC I, *Tc*MUC II, *Tc*SMUG and *Tc*TASV. β-tubulin sequences of *Trypanosoma* spp. were used as an
461 outgroup. Bold branches indicate glycoproteins more expressed in *T. cruzi* mammalian forms
462 (trypomastigote or amastigote), and dashed branches show *T. cruzi* replicative forms (epimastigote or
463 amastigote). Bootstrap of 1000 replicates. PTM Percentage is shown internally in the bars considering
464 mature protein size (excluding N- and C- terminal cleavaged sequences). NA stands for 'Non-Applicable'
465 due to a lack of signal peptide. Logo of exclusive *Tr*MUC motif and representation of two hypervariable
466 regions of *Tr*TASV showing conserved aromatic residues. Logo Y-axis: observed information content (in
467 bits), top: residues alignment position; bottom: occupancy values as gap penalty show in blue. Full
468 description of all sequences and analysis is available in Additional file 12: Supplementary Table S6. (**B**)
469 Molecular mass analysis of homologous and heterologous tagged-fused glycoproteins expressed by the *T.*
470 *rangeli* Choachí strain. Western blot assay of 50 µg of total protein extract of epimastigotes. Parasites
471 lineages: Wild-type = non-transfected Choachí strain; *Tc*MUCI = tagged-fused 'clone B4-5'; *Tc*MUCII =
472 tagged-fused 'clone B2-10'; *Tr*MUC3 = tagged-fused *Tr*MUC3; *Tr*TASV-like1 = tagged-fused *Tr*TASV-like1;
473 FlagGPI = synthetic gene not fused to glycoprotein. Top: glycoproteins recognized by anti-FlagTag
474 monoclonal antibody (mAb). Bottom: Loading control recognized by anti-β-tubulin mAb. Molecular weight
475 marker: Bio-Rad Dual Xtra. (**C**) Infection kinetics assessed by bloodstream parasite load in BALB/c mice
476 (n= 5 animals/group). Intraperitoneal infection of $1x10^7$ *in vitro* trypomastigotes. Data are expressed as
477 means (lines) and standard error of mean (shaded regions). Statistical analysis: two-way ANOVA and Post
478 Hoc Bonferroni´s multiple comparisons test (*p≤0.05; ***p≤0.001; ****p≤0.0001). Main biological differences
479 are shown in the graph, and full statistical differences are available in Additional file 13: Supplementary
480 Figure S7.
481
482 An assessment of the biological function of *Tr*MUC and *Tr*TASV-like genes was
483 performed by overexpressing selected homologous genes and by expressing *T. cruzi*
484 type I and II mucins (*Tc*MUC I and *Tc*MUC II) in *T. rangeli,* adopting a similar strategy to
485 that proposed by Cánepa et al. (2012). The expression analysis of *Tr*MUC, *Tr*TASV-like,

486 *Tc*MUC I and *Tc*MUC II by transfected *T. rangeli* trypomastigotes and epimastigotes
487 (Choachí strain) demonstrated an expression of hyperglycosylated molecules as
488 indicated by a comparison of their molecular weight to native proteins (Fig. 6b). Variable
489 glycosylation was observed, ranging from a 5X increment in *Tc*MUC I to 2.5x increment
490 for *Tc*MUC II-expressing parasites (Additional file 12: Supplementary Table S6).

491 Interestingly, overexpression of homologous *Tr*MUC and *Tr*TASV-like resulted in
492 less glycosylated proteins (1.6x and 1.3x, respectively) (Fig. 6b). In addition,
493 overexpression of *Tr*MUC3 by Choachí strain was impaired as indicated by a very faint
494 band around 45 kDa in a Western Blot analysis (Figure 6B), despite all sequences being
495 integrated in the same genome locus under the control of the tubulin promoter and being
496 attached to the parasite surface using the same GPI-anchor sequence.

497 Expression of these glycoproteins by *T. rangeli* altered the parasitemia profile in
498 the mammalian host, as shown in comparative infection assays in BALB/c mice with wild-
499 type and transfected trypomastigotes (Fig. 6c, Additional file 13: Supplementary Figure
500 S7).

501 In contrast with *Tc*MUC I-expressing parasites that showed no difference in
502 parasitemia from wild-type and Flag-GPI expressing parasites used as controls ($p > 0.05$),
503 *T. rangeli* expressing *Tc*MUC II were found earlier and in higher in number in the
504 bloodstream of the experimentally infected mice from the first ($p \leq 0.05$) to the 7th day post-
505 infection (d.p.i.) ($p < 0.001$) in comparison to controls (Fig. 6c). Overexpression of *Tr*MUC
506 and *Tr*TASV-like1 genes in *T. rangeli*, resulted in a moderate decrease in the parasitemia
507 of *Tr*TASV-like1 transfected parasites in mice between 2 ($p < 0.01$) and 4 d.p.i. ($p \leq 0.05$)
508 when compared to the controls. In contrast, overexpression of *Tr*MUC3 by *T. rangeli* was
509 unsuccessful, as shown in Fig. 6b. Mice infected with *Tr*MUC3-transfected parasites
510 showed variable parasitemia between infected animals, and a significantly reduced
511 parasitemia between the 2nd and 4th d.p.i. ($p < 0.0001$), to the 8th d.p.i. ($p < 0.01$) (Fig. 6c,
512 Additional file 13: Supplementary Figure S7). The overall parasitemia in mice observed
513 between *T. rangeli* strains from distinct genetic lineages (10-12 days) was not altered by
514 the expression of any of the studied glycoproteins.

515

516 **3. DISCUSSION**

517 The use of a hybrid deep-sequencing approach (PacBio and Illumina) for a *de novo*
518 assembly of the *T. rangeli* genome resulted in a more contiguous and improved resolved
519 genome sequence, especially for repetitive and complex regions containing multicopy
520 gene families and telomeric sequences. Although a large proportion of the gene content
521 was identified in the previous reference genome [5], a reduction of over 90% of the
522 number of scaffolds was achieved. The use of PacBio long reads has improved the
523 contiguity of some of the complex *T. cruzi* strain genomes by resolving repetitive regions
524 [49]. Among these, strains as Bug2148 [50]; Dm28c and TCC [26]; Brazil and Y [51], and
525 Sylvio X10/1 [23] have achieved a near-complete chromosomal resolution. The additional
526 use of Illumina reads in the present study have further increased the contiguity and the
527 quality of the resulting genome assembly, as well as allowing intraspecific comparative
528 analysis.
529 The identification of 10 possible telomeric regions represents a significant
530 improvement compared to the three telomeric regions previously described in the *T.*
531 *rangeli* SC58 reference genome, since sub-telomeric regions are key sites for the
532 expression of multicopy genes in trypanosomatids. Telomeric regions were not identified
533 at both ends of any of the scaffolds in the *T. rangeli* SC58 V.2 genome. In *T. brucei*, the
534 sub-telomeric region is responsible for the expression of VSG [52,53]. In *T. cruzi*, the
535 regions adjacent to the telomeres contain (Trans-)Sialidases genes, mostly type II (GP82,
536 GP90, GP85, TSA-1, SA85, and ASP-2) [54], DGF-1, and RHS [52]. In *T. rangeli*, we
537 have previously shown that the sub-telomeric region is shorter than that of *T. cruzi*, due
538 to the absence of multicopy genes from the (Trans-)Sialidases, DGF-1, and other gene
539 families [5].
540 Polycistronic transcription of trypanosomatids via RNA polymerase II leads to a
541 pre-mRNA that requires maturation by addition of a 39bp-long spliced leader (SL)
542 sequence at their 5' ends, providing the SL RNA splice donor site and essential motifs for
543 RNA stability and splicing recognition [55,56]. Despite such relevance, the genomic
544 organization and the sequence of SL genes in trypanosomatids vary between species.
545 The sequences of SL genes in *T. cruzi* are conserved but these genes differ widely in
546 copy number and chromosomal location between strains, often appearing on multiple
547 chromosomes [20,22,23]. In *T. brucei*, SL genes are also conserved, but far less

548 dispersed across the genome [57]. Although having a similar structure formed by an exon
549 and an intron of specific lengths as described for *T. cruzi* and *T. brucei*, the *T. rangeli* SL
550 genes are distinctively interspersed with the 5S rRNA genes, a feature conserved among
551 strains of distinct origins [6]. A total of 62 SL gene copies were found across five distinct
552 scaffolds on the *T. rangeli* SC58 V.2 genome, all of them exhibiting the exon/intron
553 structure and being interspersed with the 5S rRNA genes as formerly described.

554 The improved resolution of this *T. rangeli* genome sequence has revealed
555 additional repetitive genomic content, but the *T. rangeli* SC58 V.2 genome still contains
556 a lower number of genes from multicopy gene families and fewer repetitive genomic
557 elements compared to *T. cruzi* and *T. brucei* overall.

558 In all studied *T. rangeli* strains, the Sialidase and GP63 families have the highest
559 numbers of genes [5,16]. The large amount of genes coding for sialidases in the *T. rangeli*
560 genome might be related to the involvement of this molecule during cell adhesion and
561 invasion processes as described to other trypanosomatids. Thus, it was initially
562 hypothesized that these enzymes would only be in their active states in epimastigote
563 forms in the vector insect [58]. However, it has been shown through experiments using
564 liquid chromatography coupled with mass spectrometry that sialidases are the most
565 abundant proteins identified in both the epimastigote and trypomastigote forms of *T.*
566 *rangeli* [18]

567 The description of VSG-coding genes in the *T. rangeli* AM80 genome by Bradwell
568 et al. (2018) was unexpected, since these surface glycoproteins have only been
569 exclusively described in African trypanosomes, such as *T. brucei*, *T. congolense* and *T.*
570 *vivax*, but not in *T. cruzi* or *T. rangeli* [5,22,59]. No VSG-related genes were found in the
571 assemblies or reads from all sequenced strains in this study (SC58, Choachí, PIT10 and
572 R1625), nor on previous genomic, transcriptomic or proteomic studies of *T. rangeli*
573 strains.

574 The results also show that DGF-1 is present in multiple copies in the genome of
575 all studied species of American trypanosomatids. There is experimental evidence that *T.*
576 *cruzi* trypomastigotes express DGF-1 proteins on the cell surface [60], as computational
577 predictions suggesting an integrin-like activity [60,61]. It has been suggested that DGF-1

578 may play a role in the contaminative transmission of *T. cruzi* and other American
579 trypanosomatids [61].
580 All analysed species contain a significant portion of interspersed repeats and a
581 smaller portion of simple repeats. The number of interspersed repeats identified here may
582 be inflated due to the number of unclassified repeats present in the database used [62].
583 The reliably classified interspersed repeats in the *T. rangeli* SC58 V.2 genome are mostly
584 Long Terminal Repeats (LTR) and Long Interspersed Nuclear Elements (LINEs). The
585 large number of unclassified sequences shows the need for further investigation of the
586 repetitive sequences in trypanosomatids to address their possible biological functions.
587 Simple repeats were previously presumed to be “junk DNA”, but it is now known
588 that these repetitive elements may function as facilitators of evolution, by providing gene-
589 associated sites of genetic variation with minimal genetic load [63]. Indeed, in the latest
590 version of the genome of the Sylvio X10/1 strain of *T. cruzi*, it was shown that simple
591 repeat regions often flank multicopy genes that encode surface proteins [23]. This
592 configuration could allow micro-homology processes to facilitate recombination between
593 repetitive and similar sequences, resulting in new variants of surface molecules, as
594 described in other parasites [23,64,65]. Thus, the resolution and identification of these
595 regions in the *T. rangeli* SC58 V.2 genome allows recombination sites containing
596 virulence factors to be studied in a non-pathogenic trypanosomatid model.
597 The orthology analysis showed a shared core genome in all studied
598 trypanosomatids, which is represented by a cluster of 4,501 orthologous CDS groups.
599 This result is consistent with previous studies indicating that trypanosomatid organisms
600 share a significant portion of their genetic content [29,66]. In comparison, 403 orthologous
601 clusters were exclusively shared between *T. cruzi* and *T. rangeli* in the previous genome
602 study of the SC58 strain, whereas this study identified 453 such orthologous groups. This
603 is not surprising, given the close phylogenetic relationship between these parasites [67].
604 Of the set of 298 unique genes identified in the *T. rangeli* SC58 V.2 genome, 166
605 (55.70%) are annotated as hypothetical proteins. From this subset of, only one has a
606 functional annotation by assignment of GO terms, and experimental evidence of
607 expression from MS/MS data. The *in silico* characterization of this gene indicates that the
608 corresponding protein is located in the plasma membrane of the parasite, with a

609 conserved site with hexapeptide transferase activity and a unique associated peptide.
610 The identification of this CDS is important for studies of genetic markers with
611 biotechnological applications, especially for the differential diagnosis of Chagas disease.
612 It is noteworthy to mention that sequences revealing relevant but partial similarities
613 were annotated as pseudogenes on *T. rangeli* SC58 V.2 genome. These may represent
614 remnants of formerly functional genes that underwent pseudogenization via a number of
615 biological mechanisms, as proposed for *L. major*, *T. brucei*, and *T. cruzi* [20,68].
616 While aneuploidies have been widely reported in trypanosomatids, including *T.*
617 *cruzi* [36,69,70] and *Leishmania* spp. [71–73], their occurrence in *T. rangeli* remains
618 underexplored. Previous studies using pulsed-field gel electrophoresis (PFGE)
619 documented chromosomal size polymorphisms across *T. rangeli* strains [5,74,75], but
620 ploidy and CCNV analyses were not performed. By using our new reference assembly,
621 with improved resolution of repetitive and complex regions, we estimated the somy of *T.*
622 *rangeli* SC58 and three additional strains (Choachí, PIT10, and R1625). The CCNV
623 analyses indicated that these genomes are mostly diploid, with aneuploidies that are
624 shared among all strains and ones that are unique to individual strains.
625 Since some trypanosomatids lack a functional RNA interference pathway [76] and
626 rely on polycistronic transcription, which limits promoter-driven regulation of individual
627 genes, chromosomal duplications can directly influence gene expression by providing
628 additional transcriptional templates [77]. The aneuploidies shared among *T. rangeli*
629 strains in this study were associated with scaffolds enriched with genes involved in
630 immune evasion and parasite survival, including sialidases, GPI-anchored surface protein
631 biosynthesis, cAMP signaling, RNA processing, and lipid metabolism. These functions
632 are likely essential mechanisms to evade both host and vector immune responses. At the
633 same time, the presence of aneuploidies restricted to specific chromosomes in individual
634 strains may reflect selective pressures linked to their geographic distribution and the
635 distinct host- or vector-related environments.
636 The GP63 gene copy number varies greatly among kinetoplastids. *L. major* carries
637 in total six GP63 genes, four of them arranged in tandem (Chr. 10) [78], while *L.*
638 *braziliensis* has 39 copies [79]. *T. brucei* contains about 12 genes of the orthologous
639 Major Surface Protease (MSP) divided into three families (seven MSP-A genes, four

640 MSP-B in tandem, and a single MSP-C gene) [80,81]. In contrast, *T. cruzi* (DM28c strain)
641 has 96 genes and 282 pseudogenes belonging to 11 groups [20]. We found that *T. rangeli*
642 possesses an intermediary number of 76 genes and 40 pseudogenes, in 12 groups, in
643 which G12 is a species-specific group, including four groups that are likely to be
644 catalytically active and experimentally found to be secreted and/or attached to the
645 parasite surface. The G2 genes were not identified in previous studies. Neither the
646 phylogenetic nor the synteny analysis in SC58.V2 support the presence of G1 in *T.*
647 *rangeli*. Additionally, no *T. rangeli* sequence was identified as belonging to G0, the group
648 that includes the majority of *Leishmania* spp. GP63 genes, suggesting a divergence in
649 function and activity.

650 The structural and functional computational analysis of *Tr*GP63 is consistent with
651 the metalloprotease enzyme class, particularly when compared with leishmanolysin and
652 *Tc*GP63-G1. We verified the ability of these proteins to coordinate a Zn^{2+} ion, an essential
653 requirement for catalytic activity [41]. The presence of the S1' pocket was only identified
654 in a few *Tr*GP63 sequences. This pocket is located near the active site and has been
655 extensively studied for its ability to accommodate inhibitory molecules and is also
656 essential for the recognition of hydrophobic amino acids present in oligopeptides [40,82].
657 However, the absence of the active-site cleft cannot be interpreted as evidence of
658 inactivity, as its formation may depend on dynamic structural rearrangements not
659 explored in this study. Moreover, the absence of the S1' pocket may indicate that these
660 metalloproteases interact with distinct sets of ligands that do not rely on this region for
661 recognition or binding. We identified several *Tr*GP63 structures with high functional
662 potential to act as metalloproteases. This included structures with elevated ligand cavity
663 probability values as predicted by P2Rank. Additionally, the set of structures classified as
664 highly promising showed interaction modes with the inhibitor oriented toward the zinc ion,
665 which is the expected interaction pattern for this inhibitor with metalloproteases [44].
666 These findings indicate that the active sites of *Tr*GP63 display a biochemical environment
667 comparable to *L. major,* suggesting enzymatic functionality. Notably, the *T. rangeli*
668 metalloproteases in the highly promising group exhibited stronger functional indicators
669 than those observed for the *Tc*GP63-G4 and G6, which have already been characterized
670 as functionally active. Therefore, *Tr*GP63 may perform equivalent functions and represent

671 suitable candidates for functional substitution to verify them *in vitro* using the *T. cruzi* CL
672 Brener strain [43].
673 The metalloprotease activity of GP63 plays various roles across kinetoplastids,
674 including supporting parasite survival and adaptation to different hosts [43], as well as
675 contributing to basic parasite biology, such as lifecycle progression in *T. brucei* through
676 VSG release along with GPI-specific phospholipase C (GPI-PLC) [81]. In *Leishmania*
677 spp., GP63 is known to act in adhesion, invasion, and immune evasion by interacting with
678 host receptors, modulating complement, degrading signaling pathways to impair immune
679 signaling and vesicular trafficking (phagolysosome biogenesis) and protecting the
680 parasite against oxidative stress [83], via lipid raft–mediated entry, collectively enabling
681 parasite survival inside macrophages [84]. Recently, this has been more clearly defined
682 by identifying the host-target substrates for GP63 [38]. There is a possibility of a non-
683 catalytic role of GP63, including the expression of inactive proteins as adhesins that
684 adhere to the fibronectin-receptor [85]. In addition, GP63 proteins are also expressed by
685 species that do not infect mammals such as *Phytomonas* spp., or even in monoxenic
686 species that develop exclusively in insects such as *Leptomonas* spp., *Herpetomonas*
687 spp., *Angomonas* spp., and *Crithidia* spp. [86]. This suggests a role of GP63 in the
688 defense against the immune system of the invertebrate host, since this metalloprotease
689 is also capable of hydrolyzing antimicrobial peptides produced by insects [86]. Thus,
690 GP63 performs several different roles, which may explain the occurrence of different
691 groups within the same species. The functions of *T. rangeli* GP63 are still largely
692 unknown. A comparative view could suggest a role in macrophage survival since it is
693 known that amastigote-like forms remain viable after internalization by promonocytes [87].
694 It is known that *T. rangeli* encodes sialidases with hydrolytic activity but lacks
695 enzymes with sialic acid transglycosylase activity [88]. Thus, the finding of a GPI-
696 anchored, mucin-like protein of 33 kDa expressed by *T. rangeli* trypomastigotes [18] led
697 us to investigate the biological function of *T. rangeli* mucins as receptors of sialic acid
698 transferred from the host cell glycoconjugates to the *T. cruzi* surface [89].
699 In the *T. rangeli* SC58 V.2 genome, it has been possible to determine that this
700 protein belongs to a previously undetected glycoprotein family in *T. rangeli*: *Tr*TASV-like,
701 consisting of 13 complete genes and three pseudogenes. TASV has previously only been

702 described in *T. cruzi* [48,90]. The *Tc*TASV family is conserved among *T. cruzi* strains
703 with around 40 genes in three subgroups (A, B, and C) [45,48,91]. *Tc*TASV-C is the most
704 studied subgroup of this family due to immunoreactivity in natural and experimental
705 infections [45,91,92]. The proteins are mainly shed through extracellular vesicles by
706 bloodstream trypomastigotes, but also attached to the parasite cell [45,93].
707 Trypomastigotes express *Tc*TASV-C as 60 kDa glycoproteins and, to a lesser degree, as
708 45 kDa proteins by amastigotes and epimastigotes, while the theoretical molecular mass
709 corresponds to ~36 kDa. Phosphorylation and glycosylation of *Tc*TASV-C has been
710 experimentally confirmed [91]. The difference in molecular mass in *T. cruzi*
711 trypomastigotes similar to that exhibited by the Choachí strain overexpressing *Tr*TASV-
712 like1, which strongly indicates that *Tr*TASV-like has the same PTMs as *Tc*TASV.
713 It was possible to show that the overexpression of *Tr*TASV-like1 during *T. rangeli*
714 infection in mice causes a slight decrease in parasitemia at the beginning of the infection.
715 In *T. cruz*i, even more virulent strains exhibit high levels of *Tc*TASV. *Tc*TASV-C delays
716 and decreases parasitemia peaks in the acute mice model [45]. Therefore, new data
717 obtained in *T. rangeli* indicate a possible mechanism of control of parasite number in the
718 bloodstream, possibly through stimulus of phagocytosis and subsequent antigen
719 presentation, which does not necessarily lead to parasite elimination but to intracellular
720 survival instead of exposure in the bloodstream.
721 In the present genome assembly, a more accurate estimation on the copy number
722 variation of multicopy genes was achieved. We showed that *T. rangeli* has only seven
723 genes that encode mucin-like glycoproteins (*Tr*MUC), which is very different from the high
724 copy number in *T. cruzi* (e.g., 700 in the Brazil A4 strain and 797 in Y strain) [51]. A very
725 tight expression control of *the Tr*MUC3 gene product was also observed, even under the
726 control of a vector that enforces their expression (pROCK plasmid). Despite the low
727 expression level, the infection was impacted by the overexpression of *Tr*MUC3 in
728 trypomastigotes, decreasing the parasitemia in BALB/c mice. The expression of *Tr*MUC
729 thus causes the opposite effect compared to *Tc*MUC II in *T. rangeli*, apparently
730 decreasing the presence of parasites in the bloodstream, while TcMUC II causes an
731 increase. While the expression of *Tc*MUC I by the Choachí strain does not affect the
732 number of parasites in the blood, it is not possible to conclude that there is no influence

733 on the parasite load since this glycoprotein is preferentially expressed in *T. cruzi*
734 amastigotes [94,95], a parasite form that is still unexplored in *T. rangeli* during mammalian
735 infection [87], along with the incomplete understanding of the tissue tropism of this
736 avirulent trypanosomatid in mice models [32,96]. Moreover, *T. cruzi* amastigotes do not
737 express *Tc*TS group I (with trans-sialidase activity) [54,97], alike is well-known for *T.*
738 *rangeli* [98,99]. Therefore, the best conditions to express *Tr*MUC might be found during
739 an alleged amastigote phase, playing a role with sialidases in an uncharted way, even in
740 the *T. cruzi,* where these enzymes/substrates are abundant. Further studies, mainly
741 focused on characterizing the *T. rangeli* glycoproteins, are needed to comprehend which
742 glycoproteins and how this parasite can acquire sialic acid from the media. Taking
743 together the low copy number, the tight expression control, and the absence of detection
744 of any peak in the proteome studies so far, we also hypothesize that *Tr*MUC is a gene
745 family in the suppression process, possibly a trace of the evolutionary speciation event.
746

747 **4. CONCLUSIONS**

748 The complexity of the trypanosomatid genomes correlates with their intricate
749 biology, presenting intra and interspecific variability mostly due to regions of highly
750 repetitive content such as sub-telomeric sequences. The combination of PacBio and
751 Illumina reads allowed for an improved version of the *T. rangeli* genome, and was also
752 useful for resolving complex and repetitive regions in the genome of the parasite, which
753 is the smallest and least repetitive among the mammalian infective trypanosomatids.
754 Among these repetitive regions, the results presented here demonstrate that the portion
755 of structural repetitive elements is large compared to the quantity of multicopy genes,
756 which is different from other trypanosomatid genomes studied. We highlight the
757 complexity of the *T. rangeli* multigene families of mucins-TASV and GP63 and their
758 involvement in the biology of the parasite.
759

760 **5. METHODS**

761 *5.1 Parasite culture, DNA extraction and sequencing*

762 Epimastigotes of *T. rangeli* strains Choachí, PIT10, R1625, and SC58 [100]
763 (Additional file 14: Supplementary Table S7) were cultured at 27ºC in liver infusion

764 tryptose (LIT) medium supplemented with 10% fetal calf serum (FCS), as described
765 previously [5]. Total DNA was extracted from exponential growth phase epimastigotes
766 using the ReliaPrep™ gDNA Tissue Miniprep System Kit (Promega). Library generation
767 and sequencing were performed at the National Genomics Infrastructure (NGI) hosted by
768 the Science for Life Laboratory, Stockholm (Sweden). *T. rangeli* DNA from all strains was
769 used to prepare paired-end libraries for Illumina (MiSeq or HiSeq 2500) sequencing.
770 Additionally, a SMRT library for PacBio (RSII) sequencing was prepared only for the SC58
771 strain. Library preparation and DNA sequencing were carried out according to the
772 respective manufacturer's protocol.

773 *T. rangeli* Illumina reads with Q-score mean quality <30 were removed from the
774 analysis, following removal of the adapters from the high-quality reads using the
775 Trimmomatic v.0.39 software [101]. PacBio reads were trimmed based on sequence
776 length (>1Kb) using the FiltLong v.0.2.1 software (https://github.com/rrwick/Filtlong).

777

778 *5.2 Hybrid assembling of T. rangeli SC58 strain genome*

779 Hybrid assembly of the SC58 strain genome was performed *de novo* using high
780 quality PacBio reads by the software CANU v.2.2 [102], following incorporation of Illumina
781 reads into the assembly during the scaffolding and gap-filling steps, using the SSPACE
782 Standard v.3.0 [103] and GapFiller v.1-10 software [104], respectively. Both steps were
783 run through multiple iterations to compensate for base calling errors of PacBio reads.
784 Polishing of the assembled genome was performed using Illumina reads with Q-score
785 mean quality >34 using the Pilon v.1.23 software [105]. Finally, automated allelic content
786 reassignment was performed by the Purge Haplotigs software v.1.1.2 [106].

787 General assembly metrics of *T. rangeli* SC58 V.2 were evaluated using the QUAST
788 v.5.0.2 software [107] and mapping with the SC58 reference genome was performed
789 using the "dnadiff" algorithm of the MUMmer4 v.1.4 software [108].

790 The sequences corresponding to *T. rangeli* kDNA maxicircles and minicircles were
791 removed prior assembling using the BLASTn and BLASTx algorithms from the BLAST+
792 software v.2.9.0 [109], against a database of kDNA sequences from *T. brucei*, *T. cruzi*,
793 and *T. rangeli*.

794 Further validation of the *T. rangeli* SC58 V.2 genome assembly was conducted by
795 mapping 130 universal single-copy orthologs of the Euglenozoa phylum using BUSCO
796 v.5.5.0 software [110], along with *in silico* mapping of *T. cruzi* probe sequences specific
797 to chromosomes 4, 37 and 39 [111] onto the SC58 V.2 genome assembly (Additional file
798 3: Supplementary Table S1). For this later, a high-stringency (E-value >1e-10, coverage
799 >60%, identity >60%) similarity analysis of the *T. cruzi* probes against the *T. rangeli* SC58
800 V.2 genome scaffolds were carried out using the BLASTn and tBLASTx algorithms.
801
802 *5.3 Structural and functional annotation of the T. rangeli SC58 V.2 genome*
803 CDS prediction and automatic annotation were performed using the
804 AnnotaPipeline v1.0 software [112], using default parameters for similarity analysis (E-
805 value cutoff = 1e-5) and peptide identification (Q-value cut-off = 0.05). Gene prediction
806 was performed using AUGUSTUS v.3.3.3 [113] and an in-house training set composed
807 of *T. rangeli* SC58 expressed sequence tags (EST) [17] and deduced amino acid
808 sequences from the reference genome [5]. Automated annotation of CDS was based on
809 sequence similarity analysis using the BLASTp algorithm, matching with the protein
810 sequence dataset from the Swiss-Prot v.2022 [114] and TriTrypDB v.65 [115] databases.
811 CDS showing no similarity were automatically annotated as hypothetical. Search for
812 protein domains and families was performed using InterProScan v.5.61-93.0 [116]
813 allowing assignment of Gene Ontology (GO) classification based on the identified protein
814 domains. Transcriptomic [17] and proteomic [18] datasets generated from epimastigotes
815 and trypomastigotes forms of *T. rangeli* were used to assess evidence of gene
816 transcription by Kallisto v.0.48.0 [117], and evidence of expression by Comet v.2021.01
817 [118] and Percolator v.3.05.0 [119]. A profile for each protein composing the final set of
818 predicted and annotated CDS was generated using FastProtein v.1.0 [120]. Furthermore,
819 CDS separated by less than 100 base pairs, along with adjacent ones sharing similar
820 annotations from the previous analysis, were manually reviewed to assess the potential
821 presence of pseudogenes.
822 Prediction of ribosomal RNA (rRNA) and transfer RNA (tRNA) was based on
823 sequence similarity (Identity >90%) with the Rfam database v.14.2 [121] using the

824 software Infernal v.1.1.2 [122] and validated based on coverage and identity via BLASTn
825 using the RNAcentral database v.20 [123].
826 Analysis of structural repetitive elements was performed using the software
827 RepeatMasker v.4.1.5 [124], disregarding insertions of bacterial origin and using a custom
828 library containing sequences specific to trypanosomatids retrieved from msRepDB [62].
829 The presence of telomeric regions on the extremities (~10Kbp) of each scaffold of
830 the *T. rangeli* SC58 V.2 genome was assessed by search of the canonical telomeric
831 repeat of eukaryotes (TTAGGG)n [1,52], using the Seqtk v.1.4
832 (https://github.com/lh3/seqtk) and the Tidk v.0.2.31 software [125].
833
834 *5.4 Multicopy gene families*
835 Comparative analysis of multicopy gene families (amastins, DGF-1, GP63, KMP-
836 11, mucins, MASP, RHS, sialidase/trans-sialidases, tuzins, and Variant Surface
837 Glycoproteins - VSG) was carried out between the *T. rangeli* (SC58 V.2, SC58 reference
838 genome, AM80), *T. cruzi* (CL Brener Esmeraldo-like and Sylvio X10/1), and *T. brucei*
839 (TREU927) genome.
840 Initially, a search for keywords across CDS annotations of each species genome
841 was performed using the software SeqKit v2.5.1 [126] using the following definitions as
842 queries: *amastin*, *dgf*, *dispersed, dispersed gene family*, *gp63*, *leishmanolysin*, *kmp*,
843 *kinetoplastid*, *kinetoplastid membrane protein*, *masp*, *mucin-associated*, *mucin-*
844 *associated surface protein*, *muc*, *mucin*, *rhs*, *retrotransposon*, *retrotransposon hot spot*,
845 *sialidase*, *vsg*, *surface glycoprotein*, *variant surface glycoprotein*, and *tuzin*. Related
846 terms, such as “MASP” and “Mucin-Associated Surface Proteins”, for example, were
847 considered as belonging to the same multicopy gene family to avoid redundancy. The
848 annotation of each CDS was manually curated to remove false positives and to assign
849 correct description of their products. In this sense, all genes whose automatic annotation
850 were “trans-sialidase” or “leishmaniolysin” were corrected to “sialidase”, since *T. rangeli*
851 lacks trans-sialidase activity [34], and to “GP63”, since members of this multicopy gene
852 family differs from *Leishmania* species, respectively.
853 Furthermore, a similarity analysis was then carried out between the predicted CDS
854 of each multicopy gene family and the corresponding transcript data of trypanosomatid

855 species on TriTrypDB. This step was performed by the BLASTx algorithm of the software
856 BLAST+ v.2.14.1 [109], considering an E-value cutoff of 1e-5. Results from this analysis
857 were only considered if the corresponding match presented a coverage of over 30%, at
858 least 50% identity and 50% positivity. Hence, redundancy was removed once more, to
859 ensure no sequence duplicates between the dataset obtained from the similarity analysis
860 and the recovered annotated CDS. Finally, the total number of nucleotides for each
861 multicopy gene family was obtained using the “stats" algorithm of the software SeqKit.
862
863 *5.4.1 T. rangeli GP63*
864 Sequences identified by BLAST as GP63 underwent further examination using the
865 NCBI's Conserved Domain Database (CDD v3.21). Sequences shorter than 200 amino
866 acids and classified as “incomplete”, “frameshift” or “interrupted Peptidase M8
867 Superfamily (cl19482)” were considered pseudogenes. Fragmented CDS with CDD e-
868 value > $1x10^{-32}$ for the Peptidase M8 Superfamily were artificially combined with upstream
869 or downstream coding sequences that contain complementary domains in the most
870 repetitive nucleotides at the break points, primarily occurring in glycine or proline
871 residues. The identification of the zinc-binding and active site was accomplished by
872 recognizing the “ZnMc astacin-like” domain (cd04280) and conducting a manual search
873 using Jalview version 2.11.4.1 [127]. The characterization of the signal peptide,
874 glycosylphosphatidylinositol (GPI) anchoring, and N-glycosylation was performed through
875 the use of SignalP v.6.0 [128], DeepTMHMM v.1.0 [129], NetGPI v.1.1 [130], NetN-Glyc
876 v.1.0 predictors, respectively.
877 Analysis of orthology of GP63 genes among trypanosomatids used a dataset of
878 156 sequences from *T. cruzi* strains Dm28c (21) [20] and CL Brener (5) [43] , *L. infantum*
879 JPCM5 strain (7), *L. major* Friedlin strain (6) [38,131,132], *T. brucei* TREU927 strain (17)
880 [78]. In addition, sequences from *Angomonas deanei* strain Carvalho ATCC PRA-265 (5),
881 *Leptomonas pyrrhocoris* strain H10 (5), *Crithidia fasciculata* - strain Cf-Cl (5), and the
882 free-living kinetoplastid *Bodo saltans* strain Lake Konstanz (8) were obtained by BLASTp
883 search in TriTrypDB using the *L. infantum* GP63 entries (LinJ.10.0500 and LinJ.10.0520)
884 as query [78,79]. Phylogenetic analysis was carried out using the workflow provided by
885 NGPhylogeny.fr [133] with default parameters of each included tool. Data was then

886 integrated using the Interactive Tree of Life (iTOL) v.7.2 online tool [134]. Identity matrix
887 analysis was performed using Clustal Omega (EMBL-EBI Job Dispatcher [135]) and SIAS
888 tool (http://imed.med.ucm.es/Tools/sias.html). Active site motif was describe using
889 ScanProsite tool [136].

890 Metalloprotease activity assays were carried out using a total of 1×10^8 parasites,
891 which included *Leishmania braziliensis* (MHOM/BR/75/M2904 strain promastigotes), *T.*
892 *cruzi* (Y strain epimastigotes), and *T. rangeli* (Choachí strain epimastigotes and
893 trypomastigotes). The parasites were lysed in a buffer containing 1% Triton X-100 and 10
894 mM Tris-HCl at pH 6.8. SDS-PAGE gels were prepared with 0.2% gelatin, and the
895 samples were neither heated nor reduced prior to analysis. Following electrophoresis, the
896 gels were washed twice with a 0.1 M sodium acetate buffer (pH 5.5) that contained 2.5%
897 Triton X-100. Subsequently, the gels were incubated for 48 hours at 37 °C in a buffer
898 consisting of 0.1M sodium acetate and 1mM DTT at pH 5.5. In some cases, 10mM 1,10-
899 phenanthroline (Sigma-Aldrich) was included in the incubation. Finally, the gels were
900 stained with Coomassie Brilliant Blue R-250. To evaluate the impact of pH, additional
901 experiments were conducted by incubating the gels at 37°C for 48 hours in reaction
902 buffers with varying pH values: 3.5, 5.5, 7.0, and 9.0.

903 Building on the proteomic evidence of GP63 expression in *T. rangeli* [18], we
904 conducted further analyses of the raw MS/MS data from the surface-enriched protein
905 fraction (SEP) of *T. rangeli* SC58 strain for detailed identification of GP63 expression.
906 This analysis employed MSFragger [137] and, also incorporated unpublished, in-house
907 generated secretome datasets from *T. rangeli*.

908 The final dataset comprises all 11 GP63 sequences, supplemented with
909 contaminant proteins from cRAP (https://thegp_m.org/crap/index.html), including porcine
910 trypsin, *Saccharomyces* spp. alcohol dehydrogenase, and human keratin. Decoy
911 database searches were performed automatically. MSFragger parameters specified
912 tryptic peptides allowing one missed cleavage, carbamidomethylation of cysteine,
913 methionine oxidation as a variable modification, and precursor and fragment tolerances
914 of 10 and 5 p.p.m., respectively. Protein identification and scoring were conducted using
915 the ProteinProphet algorithm within Philosopher [138]. Only proteins supported by at least

916 two unique peptides, a false discovery rate (FDR) below 0.1%, and peptide/protein
917 probabilities exceeding 95% and 99%, respectively, were considered as valid.
918
919 *5.4.2 Structural and functional analysis of T. rangeli metalloproteases*
920 To investigate the presence of metalloprotease activity in *T. rangeli*, we conducted
921 *in silico* structural and functional analyses of these proteins. Initially, we generated 3D
922 models of *T. rangeli* metalloproteases using AlphaFold 3 [139]. For comparative
923 purposes, we also obtained the model of the group 1 metalloprotease from *T. cruzi*
924 (TriTrypDB: C4B63_30g122). All models were generated using the default parameters of
925 AlphaFold 3, as outlined in its user guide. Quality assessment of the model was carried
926 out using the confidence scores provided by the program, including pLDDT (per-residue
927 confidence) and pTM (predicted template modelling). Specifically, pTM values above 0.5
928 were considered indicative of adequate modelling, while values above 0.8 reflected high
929 predictive quality [139]. All generated models were visualized, analysed, and figures were
930 created using PyMOL v.2.5.0 (https://www.pymol.org/).
931 For the structural and functional analyses of metalloproteases, leishmanolysin - a
932 metalloprotease from *L. major* (PDB: 1LML) - was utilized as the control structure [37].
933 The structural analysis focused on the identification of the active site or zinc-binding motif
934 (HEXXHXXG[~60]H) located in helix B, which is a conserved characteristic of
935 metalloproteases, was evaluated [41,44]. Surface representations of the models were
936 created to identify the S1' pocket, a common feature of these proteins extensively studied
937 for its importance in interactions with various inhibitors [140]. The loop region preceding
938 helix B, located between antiparallel β-sheets in the N-terminal domain, was examined
939 due to its presence in the active site of the metalloproteases. Furthermore, we assessed
940 the structural similarities between the metalloprotease models of *T. rangeli*, *T. cruzi* and
941 the control *L. major* models through structural alignments. This process was followed by
942 RMSD (Root Mean Square Deviation) calculations for both the overall protein and for
943 helix B specifically, considering only the alpha carbons in each case.
944 To evaluate the interaction potential of the modeled metalloprotease active sites
945 with ligands, the proteins were submitted to the P2Rank web server [42]. P2Rank is a
946 machine-learning tool that identifies potential binding sites by analysing surface features,

947 shape, and electrostatic potential. The server provides a ranked list of potential binding
948 sites, derived from the P2Rank score and the ligand cavity probability. The
949 metalloprotease models were categorized into three groups based on the ligand cavity
950 probability value: (i) highly promising, when probability is ≥ 50%; (ii) moderately
951 promising, when it falls between < 50% and ≥ 40%; and (iii) less promising, when it is <
952 40%.

953 The metalloprotease models underwent molecular docking with 1,10-
954 phenanthroline, an established inhibitor of metalloproteinases [38,44]. This compound
955 typically interacts with the active site of these enzymes, specifically targeting the Zn^{2+} ion,
956 which serves as an indicator of a catalytic environment comparable to that of
957 metalloproteinases. Docking simulations were carried out using AutoDock Vina [141]. An
958 optimized grid box was strategically positioned over the active site to generate nine poses
959 for each protein, while all other parameters adhered to the program's default settings.
960 Initially, the control structure was docked with 1,10-phenanthroline, and the pose with the
961 lowest affinity value and the most favourable orientation toward the Zn^{2+} ion was selected
962 as the reference model. Subsequently, docking was conducted with the metalloproteases
963 from *T. rangeli* and *T. cruzi*, focusing on poses that displayed the highest similarity to the
964 control complex. To accomplish this, the proteins were aligned, and the RMSD between
965 the ligands was calculated, with the pose showing the lowest RMSD chosen for further
966 analysis.

967

968 *5.4.3 T. rangeli mucin-like glycoproteins*

969 Identification of *T. rangeli* mucin-like glycoproteins was carried out by BLASTn
970 using the first 93 nt of *T. rangeli* mucin-likes TRSC58_06796 (TriTrypDB) and KC544919
971 (GenBank) genes described in the reference genome [5] and proteome [18], respectively.
972 Also, the presence of N- and C-terminal domains of *T. cruzi* mucin-like genes signatures
973 [142,143] was assessed. Hits that did not match to a CDS automatically annotated were
974 evaluated by analysis ORFs obtained from 3 Kbp flanking regions of identified hits using
975 the GetORF tool (https://www.bioinformatics.nl/cgi-bin/emboss/getorf), considering the
976 following parameters: 150 nt minimum length; and only Met as start codon. Only
977 sequences encoding for valid signal peptides according to SignalP v.6.0 predictor [128]

978 or DeepTMHMM v.1.0 [129] were considered as putative *T. rangeli* mucins (*Tr*MUC). The
979 presence of GPI-anchor, *N*- and *O*-glycosylation sites, in addition to phosphorylation
980 (ATM, CKI, CKII, CaM-II, DNAPK, EGFR, GSK3, INSR, PKA, PKB, PKC, PKG, RSK,
981 SRC, cdc2, cdk5, and p38MAPK) recognition sites was evaluated using NetGPI v.1.1
982 [130], NetN-Glyc v.1.0, NetO-glyc v.4.0 [144], and NetPhos v.3.1 [145] servers,
983 respectively. Prediction of amino acid abundance and molecular weight was performed
984 by Pepstats tool available at EMBL-EBI Job Dispatcher [135]. The conservative regions
985 were evaluated by Gblocks [146] applying Clustal Omega [147] alignments as inputs, as
986 well as to obtain logos representations through the Skylign tool [148], and overall similarity
987 analysis performed by SIAS tool. Phylogenetic analysis based on maximum-likelihood
988 tree inferred amino acid sequences identified from the *T. rangeli* SC58 V.2 genome, as
989 well as members of the *T. cruzi* mucin-like (MUC) and Trypomastigote Alanine-Serine-
990 Valine (TASV) glycoprotein families retrieved from TryTripDB and Genbank databases
991 was conducted as described for GP63 analysis.
992
993 *5.4.4 Biological assessment of T. rangeli mucins and TASV*
994 We designed a synthetic gene (purchased from Biomatik) hereby called 'FlagGPI'
995 encoding a 3xFlagTag® (Sigma-Aldrich) sequence, followed by the mNeonGreen
996 fluorescent protein gene [149], and a *T. rangeli* GPI-anchor sequence (KC544919-
997 GenBank) under control of *T. cruzi* gp82 3' untranslated region (UTR). The GFP gene
998 present in the integrative plasmid pROCK-Neo [150] was replaced by the FlagGPI gene
999 through *Xba*I and *Xho*I restriction enzyme cleavage, generating the FlagGPI-pROCK
1000 vector.
1001 The *T. rangeli* (Choachí strain) genes MUCg3 and TASV-like1, and the *T. cruzi* (Y
1002 strain) genes MUC I (clone B4-5), and MUC II (clone B2-10), were PCR amplified from
1003 genomic DNA. The amplicons, that excluded the GPI-anchor addition sequence, were gel
1004 purified using the QIAquick Purification kit (Qiagen), cloned in pGEM T-Easy vectors
1005 (Promega), sequenced (ABI3500 - Applied Biosystems), and subcloned in the FlagGPI-
1006 pROCK vector, resulting in tagged-fused constructs. A complete list of primers and
1007 restriction enzymes is provided in Additional file 15: Supplementary Table S8.

1008 Epimastigotes of *T. rangeli* (Choachí strain) were transfected with 10 µg of *Not*I-
1009 linearized FlagGPI-pROCK vector containing the distinct tagged-fused constructs or with
1010 the FlagGPI synthetic gene, using the Nucleofactor® II (Lonza) U-33 program. The
1011 transfected parasites were cultivated in LIT+NNN and selection was performed by
1012 increasing concentrations of geneticin (G418) up to 500 µg/mL. Plasmid integration into
1013 the parasite β-tubulin locus was confirmed by PCR using a set of primers designed to
1014 anneal exclusively in the genomic β-tubulin region upstream of the homologous
1015 recombination site and the HX1 5' UTR present particularly in FlagGPI-pROCK vector
1016 (Additional file 15: Supplementary Table S8). Once the integration into the parasite
1017 genome was confirmed, parasites were cultivated in the absence of selective drugs.
1018 To assess the expression of the FlagGPI-fused glycoproteins, total protein extracts
1019 of *T. rangeli* epimastigotes and trypomastigotes from each transfected lineage and wild-
1020 type parasites were obtained by lysis in proper buffer (50 mM NaCl, 200 mM Tris-HCl pH
1021 8.0, 1% (v/v) Triton X-100), supplemented with a protease inhibitor cocktail (Sigma-
1022 Aldrich). A total of 50 µg of each extract were resolved in 12% SDS-Page gels and then
1023 wet-transferred to Hybond® PVDF membranes (Amersham). Membranes were incubated
1024 with primary antibodies (mouse anti-FlagTag M2 or rabbit anti-β-tubulin), then with HRP
1025 anti-IgG conjugates (anti-mouse or anti-rabbit) (Sigma-Aldrich). The expression analysis
1026 was periodically assessed.
1027 Comparative *in vivo* assessment of the infectivity of all transfected lineages was
1028 performed using 6 weeks old female BALB/c mice, maintained in ventilated cages under
1029 a 12-hour light/dark cycle and having water and food *ad libitum*. A total $1x10^7$ *in vitro*
1030 differentiated *T. rangeli* trypomastigotes were injected by intraperitoneal (i.p.) route and
1031 parasitemia was assessed daily by the Pizzi-Brener method from 6h post-infection [151].
1032 *In vivo* experiments were approved by the UFSC Animal Ethics Committee (Protocol
1033 9923170516). Graph and statistical analysis were performed using GraphPad Prism
1034 v.8.4.0 software. The expression of *T. rangeli* mucins and TASV was also evaluated by
1035 search on *T. rangeli* MS/MS proteomic data as performed for GP63 (Item 2.4.1.).
1036
1037 *5.5 Analysis of orthologous groups*

1038 Orthology analysis was performed with OrthoFinder v.2.3.11 [152] using the
1039 predicted proteins from *T. rangeli* SC58 V.2 genome and proteins from *T. brucei*
1040 (TREU927 strain), *T. cruzi* (CL Brener Esmeraldo-like and Sylvio X10/1 strains), *T. rangeli*
1041 (SC58 reference strain), *Leishmania amazonensis* (MHOM/BR/71973/M2269 strain) and
1042 *Leishmania infantum* (JPCM5 strain), obtained from TriTrypDB version 65 [115], and also
1043 *T. rangeli* (AM80 strain) retrieved from GenBank (Access code: GCF_003719475.1).
1044
1045 *5.6 Comparative intra-specific genomics of T. rangeli strains*
1046 Illumina paired-end libraries of distinct *T. rangeli* strains (SC58, Choachí, PIT10,
1047 and R1625) were quality-filtered using Trimmomatic v0.40 [101] and the filtered reads
1048 were mapped against the *T. rangeli* SC58 V.2 genome using the Burrows-Wheeler
1049 Aligner v. 0.7.17 [153].
1050 For SNP calling, the mapping files were sorted with Samtools v.1.17 [154], PCR-
1051 duplicates were marked with Picard v.2.27.5 (http://broadinstitute.github.io/picard/), and
1052 SNPs were identified using the Genome Analysis Toolkit v.3.8
1053 (https://gatk.broadinstitute.org/hc/en-us) using the "HaplotypeCaller" algorithm with a
1054 minimum quality value of 30 and a minimum depth of coverage of 10. A strict mapping
1055 quality filter was applied to remove the influence of low mapping quality in repeated
1056 regions and any SNP with mapping quality <50 was removed using BCFtools v.1.9 [154].
1057 CCNV and ploidy estimation were performed using a combination of RDC and AB
1058 methods as previously described [36]. Briefly, the estimation of CCNV was based on the
1059 ratio between the mean coverage in each *T. rangeli* SC58 V.2 scaffold and the mean
1060 coverage of all scaffolds in the assembly. If the ratio between the median chromosome
1061 coverage and the median genome coverage was approximately one, the chromosome
1062 had the same copy as the genome overall, while fluctuations in this ratio (<0.75 or >1.25)
1063 were considered as putative aneuploidies. Heterozygous SNPs with a mapping quality
1064 higher than 50 were selected for AB analyses to confirm aneuploidies and to estimate
1065 chromosome somy and whole genome ploidy. Whole genome ploidy was estimated by
1066 analysing the proportion of the alleles in heterozygous SNP positions of all scaffolds. To
1067 confirm the RDC somy estimations, the proportion of each allele with heterozygous SNPs
1068 per chromosome position was plotted for the 82 scaffolds. All RDC and AB graphs were

generated in R with ggplot2 (available at https://ggplot2.tidyverse.org), pheatmap (available at https://CRAN.R-project.org/package=pheatmap), and vcfR [155] packages. Gene Ontology Enrichment analyses in CCNV regions were performed using TriTryp tools (https://tritrypdb.org/tritrypdb/app/) based on orthologs with the *T. cruzi* reference CL Brener Esmeraldo-like.

ADDITIONAL FILES

Additional file 1: Supplementary Figure S1. Comparative alignment of the assemblies of *Trypanosoma rangeli* SC58 V.2 genome with other *T. rangeli* and *T. cruzi* genomes.

Footnote: Dot-plot representation of the alignment of the *T. rangeli* SC58 V.2 genome assembly with the reference genome (A), *T. rangeli* Choachí (B), PIT10 strain (C) and R1625 (D) strains genomes, *T. cruzi* CL Brener Esmeraldo-like (E), and Sylvio X10/1 strains genomes using comparative dot-plots generated by D-Genies (Cabanettes et al., 2018).

Additional file 2: Supplementary Figure S2. *Trypanosoma rangeli* five loci of the Spliced Leader RNA genes interspersed with 5S RNA genes.

Footnote: (**A**) Alignment of 62 *T. rangeli* Spliced-Leader (SL) RNA genes highlight the exon region (light green), intron (light red), and splicing site at position 39. (**B**) Five loci show SL genes interspersed with 5S RNA genes, varying from 9 to 16 SL gene copies.

Additional file 3: Supplementary Table S1. Mapping of *Trypanosoma cruzi* (CL Brener strain) chromosome-specific probes onto the *Trypanosoma rangeli* SC58 V.2 genome.

Footnote: *Access codes of *T. cruzi* genes in the TriTrypDB v.65 database. **Genomic location and “P” and “S” haplotypes of *T. cruzi* CL Brener genome (Weatherly; Boehlke; Tarleton, 2009).

Additional file 4: Supplementary Table S2. Telomeric regions identified in the *Trypanosoma rangeli* SC58 V.2 genome scaffolds.

Footnote: bp = base pairs.

1101

1102 **Additional file 5: Supplementary Figure S3 -** Allelic proportions at heterozygous SNP
1103 positions for the 81 scaffolds of each *Trypanosoma rangeli* strain (SC58, Choachí, PIT10,
1104 R1625). Blue points represent allele proportions along each scaffold; black lines indicate
1105 median values within sliding windows.

1106

1107 **Additional file 6: Supplementary Table S3**. Summary of GP63 genes of *T. rangeli* and
1108 other Kinetoplastids species are presented in the phylogenetic analysis.

1109

1110 Footnote: 'GPI-Anchored' in red= sequence forecasted as GPI-anchored, however biologically not possible
1111 if signal peptide prediction is correct. 'BNP' = biologically not possible; SP = Signal Peptide; GLOB =
1112 globular; TM = transmembrane, ----- = presence of gap in active site region; XXxxX = Non-conserved amino
1113 acids in zinc-binding site/active site (other than HExxH). NE: Non-evaluated. Peptidase_M8 Superfamily
1114 (NCBI CDD code: cl19482).

1115

1116 **Additional file 7: Supplementary Figure S4.** *Trypanosoma rangeli* Surface
1117 Glycoprotein 63 (GP63) presents the main classical leishmanolysin motifs.

1118

1119 Footnote: Logo representation of residue conservation as weighted counts of Zinc-binding site. Logo Y-axis
1120 = observed information content (in bits), top = residues position in the alignment; bottom = occupancy
1121 values.

1122

1123 **Additional file 8: Supplementary Table S4**. Amino Acid Percent Identity Matrix of
1124 *Tr*GP63 complete genes.

1125

1126 Footnote: Colors highlight the *Tr*GP63 groups categorization as shown in the header of the matrix.

1127

1128 **Additional file 9: Supplementary Figure S5.** *Trypanosoma rangeli* exhibits
1129 metalloprotease activity, and it is biochemically inhibited by chelating agents.

1130

1131 Footnote: (**A**) Total protein extracts from epimastigotes (E) and culture-derived trypomastigotes (T) of *T.*
1132 *rangeli* (10^8 cells) were analysed by SDS-PAGE (10%) using a 0.2% gelatin-containing gel, without heating
1133 or reducing the samples before electrophoresis. Gels were incubated at 37 °C for 48 h in reaction buffer,

without (-) or with (+) specific inhibitor 1,10-phenanthroline. Proteolytic activity is visualized as clear bands against the Coomassie brilliant blue-stained gelatin background. (**B**) Total protein extracts from axenic forms of *T. rangeli* Choachí strain (Tr), *T. cruzi* Y strain (Tc), and promastigotes of *L. braziliensis* MHOM/BR/75/M2904 strain (10^8 cells) were analysed in gels incubated with reaction buffers at different pH conditions.

Additional file 10: Supplementary Table S5. Values from the structural and *in silico* functional analyses of *Trypanosoma rangeli* GP63 metalloproteases.

Footnote: Metalloprotease from *L. major* (PDB: 1LML) was used as a control, and a single-sequence from *T. cruzi* group 1 (TriTrypDB:C4B63_30g122) was included for comparative analysis. Colors indicate metalloprotease inferred activity based on ligand cavity scores: Green for highly promising (≥ 0.5), yellow for moderately promising (<0.5 to ≥ 0.4), and red for moderately promising (<0.4). pTM = predicted template modelling. RMSD = root mean square deviation.

Additional file 11: Supplementary Figure S6. Structure of *Trypanosoma rangeli* GP63 metalloproteases from distinct groups.

Footnote: **(A)** Structural and *in silico* functional analysis of *T. rangeli* GP63 classified as a highly promising group regarding inferred activity. (**B**) *Tr*GP63 is classified as a moderately promising group. (**C**) *Tr*GP63 was predicted as a less promising group. 3D Model: Three-dimensional models demonstrating the region corresponding to the active/Zinc-binding site (Highlighted by a box), where structural and functional analyses were performed. Zinc-binding site: Visualization of the conserved amino acid residues are shown in stick representation. Molecular surface: representation of the proteins indicating the presence or absence of the S1' pocket representing the position which interacts with inhibitors. Presence of S1' pocket is indicated by a green arrow, and absence by a red arrow. P2Rank analysis: Ligand-cavity prediction shown by colours demonstrates different predicted cavity positions. Predicted binding cavity probability is shown in the boxes. Molecular docking analysis: Complexes of metalloproteases with the inhibitor 1,10-phenanthroline. The cyan ligand represents the highest-affinity conformation obtained for the control structure, while the white ligands correspond to molecular docking results for the other GP63. Blue indicates nitrogen in both representations. The boxes present the RMSD values (Å) and ΔG (kcal/mol) of the complexes resulting from molecular docking.

Additional file 12: Supplementary Table S6. Description of *Tr*MUC and *Tr*TASV genes and predicted glycoprotein characteristics.

1169

1170 Footnote: Yellow background= *T. rangeli* glycoproteins selected for biological characterization. “MW”=
1171 Molecular Weight marker, “WB”= Western Blot. *excluded mature FlagGPI construct 30.39 kDa. #GPI-
1172 anchor cleavage did not occur in the synthetic gene due to a lack of signal peptide. Aromatic residues (Phe,
1173 His, Trp, Tyr). NP = Non-provided by DeepTMHMM Signal peptide predictor.

1174

1175 **Additional file 13: Supplementary Figure S7:** Comparative parasitemia between wild-
1176 type *Trypanosoma rangeli* and transfectants expressing T*r*MUC, *Tr*TASV-like1, or
1177 *Tc*MUC genes.

1178

1179 Footnote: Graphs expressed as means and standard error of mean by lines and shaded regions,
1180 respectively. Statistical analysis: Normal distribution of data evaluated by Kolmogorov-Smirnov test
1181 (α=0.05), followed by two-way ANOVA and Post Hoc Bonferroni´s multiple comparisons test. Differences
1182 between groups per time point are shown in the graphs. Unique statistical symbols demonstrate the same
1183 statistical difference in all time points between 1 to 10 d.p.i. (ns $p>0.05$; *$p\leq0.05$; **$p\leq0.01$;*** $p\leq0.001$; ****
1184 $p\leq0.0001$). The overall evaluation between the transfected parasite lineages, regardless of time point, is
1185 provided in the detailed table.

1186

1187 **Additional file 14: Supplementary Table S7**. *Trypanosoma rangeli* strains used in this
1188 study, showing their original hosts and geographical origins, as well as their distinct
1189 genetic classifications based on kDNA and ribosomal markers.

1190

1191 Footnote: *kDNA-based classification as proposed by Vallejo et al. (2002). **Ribosomal markers-based (SSU and
1192 ITS) classification as proposed by Maia da Silva et al. (2004). ND= Not determined.

1193

1194 **Additional file 15: Supplementary Table S8:** Primers used in PCR amplification and
1195 cloning of *Trypanosoma rangeli* and *Trypanosoma cruzi* mucins into the FlagGPI-pROCK
1196 vector and on assessment of genomic integration.

1197

1198 Footnote: 'F' and 'R' stand for Forward and Reverse primers, respectively. Restriction sites are highlighted in
1199 bold/italics at the 5'-end of each primer sequence. *= Amplification product generated with primer HX1 R for integration
1200 evaluation by PCR.

1201

ABBREVIATIONS

AB: Allele Balance; Bp: Base pairs; CCNV: Chromosome Copy Number Variation; CDS: Coding sequence; DGF-1: Disperse Gene Family 1; GO: Gene Onthology; GP: Glycoprotein; GPI: Glycosylphosphatidylinositol; Kb: Kilobases; Kbp: Thousand base pairs; kcal/mol: Kilocalories per mole; kDa: Kilodaltons; KMP: Kinetoplastid Membrane Protein; mAb: Monoclonal antibody; MASP: Mucin-Associated Surface Proteins; Mb: Megabase; Mbp: Million base-pairs; pTM: Predicted template modelling; RDC: Read Depth Coverage; RHS: Retrotransposon Hot Spot; RMSD: Root Mean Square Deviation; RNAi: RNA interference machinery; SNP: Single Nucleotide Polymorphism; TASV: Trypomastigote, Alanine, Serine and Valine rich protein glycoprotein family; VSG: Variable Surface Glycoprotein.

ACKNOWLEDGMENTS

The authors would like to thank the Laboratório de Estudos em Biologia (LAMEB/UFSC, Brazil) and the Science for Life Laboratory (KI, Stockholm) for technical support. Authors are indebted to Dr Ariel Silber (USP, Brazil) for critical reading and valuable comments on this manuscript.

AUTHOR'S CONTRIBUTIONS

GAM, GMM and ACS: Equally contributed to Formal analysis, Investigation, Methodology, Validation, Visualization and Writing – original draft.

CAP, CLMP, DDL, RAC, RSM and TKSP: Formal analysis, Investigation, Methodology and Validation. ASR: Formal analysis and Investigation.

TCB and GRM: Formal analysis, Investigation, Methodology, Validation and Writing – original draft.

JFSF, MS: Data curation, Formal analysis, Investigation, Resources, Supervision, Validation and Writing – original draft.

PHS and GW: Data curation, Formal analysis, Investigation, Methodology, Validation, Resources, Supervision and Writing – original draft.

BA and ECG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation and Writing – original draft.

FUNDING

This work was supported by grants from the following Brazilian Government Agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (Grants 477385/2012-5 and 308616/2015-4), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (Grants CAPES/PRINT 88887.311315/2018-00 and CAPES/STINT 23038.003782/2014–11), Financiadora de Estudos e Projetos - FINEP (Grant 0847/07); and from the Swedish Research Council - SRC (Grant 2021-05667). GAM, ACS, CLMP, DDL, ASR, TCB, and RAC were recipients of CNPq or CAPES scholarships.

AVAILABILITY OF DATA AND MATERIALS

Data generated during the current study is available in the GenBank database (Bioproject PRJNA1380247). Further datasets or subsets used and/or analysed during the current study are available from the corresponding authors on reasonable request. DNA transfer from UFSC to KI was carried out exclusively for sequencing purposes and was registered in the Brazilian National System for the Management of Genetic Heritage and Associated Traditional Knowledge (SISGEN A206000).

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

COMPETING INTERESTS

The authors declare no competing interests. The funders had no role in the study design, data analysis, or the decision to publish.

AUTHOR DETAILS

1- Departamento de Microbiologia, Imunologia e Parasitologia, Universidade Federal de Santa Catarina (UFSC), Florianópolis - SC, Brazil.

2- Department of Medical Biochemistry and Biophysics, Karolinska Institutet (KI), Stockholm, Sweden.
3- Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet (KI), Stockholm, Sweden.
4- Departamento de Microbiologia, Imunobiologia e Parasitologia, Universidade Federal de São Paulo (UNIFESP), São Paulo - SP, Brazil.
5- Instituto Federal de Santa Catarina (IFSC), Gaspar - SC, Brazil.
6- Departamento de Bioquímica, Universidade Federal de Santa Catarina (UFSC), Florianópolis - SC, Brazil.
7- Department of Cell and Molecular Biology, Karolinska Institutet (KI), Stockholm, Sweden.

REFERENCES

1. Horn D, Spence C, Ingram AK. Telomere maintenance and length regulation in Trypanosoma brucei. EMBO J. 2000;19:2332–9. https://doi.org/10.1093/emboj/19.10.2332
2. WHO WHO. Chagas disease (also known as American trypanosomiasis) [Internet]. 2025 [cited 2025 Aug 31]. https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis). Accessed 31 Aug 2025
3. WHO WHO. Leishmaniasis [Internet]. 2023 [cited 2025 Aug 31]. https://www.who.int/news-room/fact-sheets/detail/leishmaniasis. Accessed 31 Aug 2025
4. Urdaneta-Morales S, Tejero F. Trypanosoma (Herpetosoma) rangeli Tejera, 1920: intracellular amastigote stages of reproduction in white mice. Rev Inst Med Trop Sao Paulo [Internet]. Instituto de Medicina Tropical de São Paulo; 1986;28:166–9. https://doi.org/10.1590/S0036-46651986000300005
5. Stoco PH, Wagner G, Talavera-Lopez C, Gerber A, Zaha A, Thompson CE, et al. Genome of the Avirulent Human-Infective Trypanosome—Trypanosoma rangeli. PLoS Negl Trop Dis. 2014;8:e3176. https://doi.org/10.1371/journal.pntd.0003176
6. Grisard EC, Steindel M, Guarneri AA, Eger-Mangrich I, Campbell DA, Romanha AJ. Characterization of Trypanosoma rangeli Strains Isolated in Central and South America: an Overview. Mem Inst Oswaldo Cruz. 1999;94:203–9. https://doi.org/10.1590/S0074-02761999000200015
7. Guhl F, Vallejo GA. Trypanosoma (Herpetosoma) rangeli Tejera, 1920: an updated review. Mem Inst Oswaldo Cruz. 2003;98:435–42. https://doi.org/10.1590/S0074-02762003000400001

8. Dario MA, Moratelli R, Schwabl P, Jansen AM, Llewellyn MS. Small subunit ribosomal metabarcoding reveals extraordinary trypanosomatid diversity in Brazilian bats. PLoS Negl Trop Dis. 2017;11:e0005790. https://doi.org/10.1371/journal.pntd.0005790

9. Duarte da Silva B, Guarneri AA. *Trypanosoma rangeli* infection impairs reproductive success of *Rhodnius prolixus*. Parasitology. 2023;150:42–8. https://doi.org/10.1017/S0031182022001470

10. Bayão T de S, Cupertino M do C, Mayers NAJ, Siqueira-Batista R. A systematic review of the diagnostic aspects and use of Trypanosoma rangeli as an immunogen for Trypanosoma cruzi infection. Rev Soc Bras Med Trop. 2020;53. https://doi.org/10.1590/0037-8682-0608-2019

11. Dario MA, Pavan MG, Rodrigues MS, Lisboa CV, Kluyber D, Desbiez ALJ, et al. Trypanosoma rangeli Genetic, Mammalian Hosts, and Geographical Diversity from Five Brazilian Biomes. Pathogens. 2021;10:736. https://doi.org/10.3390/pathogens10060736

12. Cruz-Saavedra L, Ospina C, Gutiérrez SA, Jaimes-Dueñez J, Cantillo-Barraza O, Hernández C, et al. Exploring Trypanosoma cruzi transmission dynamics in an acute Chagas disease outbreak using next-generation sequencing. Parasit Vectors. 2024;17:395. https://doi.org/10.1186/s13071-024-06445-9

13. de Moraes MH, Guarneri AA, Girardi FP, Rodrigues JB, Eger I, Tyler KM, et al. Different serological cross-reactivity of Trypanosoma rangeli forms in Trypanosoma cruzi-infected patients sera. Parasit Vectors. 2008;1:20. https://doi.org/10.1186/1756-3305-1-20

14. Rodrigues ES, Santos GQ, Silva MV da, Barros JHS, Bernardo AR, Diniz RL, et al. Chagas Immunochromatographic Rapid Test in the Serological Diagnosis of Trypanosoma cruzi Infection in Wild and Domestic Canids. Front Cell Infect Microbiol. 2022;12. https://doi.org/10.3389/fcimb.2022.835383

15. Kann S, Concha G, Weinreich F, Hahn A, Rückert C, Kalinowski J, et al. Comparative Assessment of Two Commercial Real-Time PCR Assays for the Diagnosis of Trypanosoma cruzi DNA in Serum. Microorganisms. 2023;11:901. https://doi.org/10.3390/microorganisms11040901

16. Bradwell KR, Koparde VN, Matveyev A V, Serrano MG, Alves JMP, Parikh H, et al. Genomic comparison of Trypanosoma conorhini and Trypanosoma rangeli to Trypanosoma cruzi strains of high and low virulence. BMC Genomics. 2018;19:770. https://doi.org/10.1186/s12864-018-5112-0

17. Grisard EC, Stoco PH, Wagner G, Sincero TCM, Rotava G, Rodrigues JB, et al. Transcriptomic analyses of the avirulent protozoan parasite Trypanosoma rangeli. Mol Biochem Parasitol. 2010;174:18–25. https://doi.org/10.1016/j.molbiopara.2010.06.008

18. Wagner G, Yamanaka LE, Moura H, Lückemeyer DD, Schlindwein AD, Stoco PH, et al. The Trypanosoma rangeli trypomastigote surfaceome reveals novel proteins and

1338 targets for specific diagnosis. J Proteomics. 2013/03/08. 2013;82:52–63.
1339 https://doi.org/10.1016/j.jprot.2013.02.011
1340 19. McGwire BS, Chang K-P, Engman DM. Migration through the Extracellular Matrix by
1341 the Parasitic Protozoan *Leishmania* Is Enhanced by Surface Metalloprotease gp63.
1342 Infect Immun. 2003;71:1008–10. https://doi.org/10.1128/IAI.71.2.1008-1010.2003
1343 20. Berná L, Chiribao ML, Pita S, Alvarez-Valin F, Parodi-Talice A. Exploring the
1344 genomic landscape of the GP63 family in Trypanosoma cruzi: Evolutionary dynamics
1345 and functional peculiarities. PLoS Negl Trop Dis. 2025;19:e0012950.
1346 https://doi.org/10.1371/journal.pntd.0012950
1347 21. Chiurillo MA, Peralta A, Ramírez JL. Comparative study of Trypanosoma rangeli
1348 and Trypanosoma cruzi telomeres. Mol Biochem Parasitol. 2002;120:305–8.
1349 https://doi.org/10.1016/S0166-6851(02)00005-1
1350 22. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran A-N, et al.
1351 The Genome Sequence of *Trypanosoma cruzi* , Etiologic Agent of Chagas Disease.
1352 Science (1979). 2005;309:409–15. https://doi.org/10.1126/science.1112631
1353 23. Talavera-López C, Messenger LA, Lewis MD, Yeo M, Reis-Cunha JL, Matos GM, et
1354 al. Repeat-Driven Generation of Antigenic Diversity in a Major Human
1355 Pathogen, Trypanosoma cruzi. Front Cell Infect Microbiol. 2021;11.
1356 https://doi.org/10.3389/fcimb.2021.614665
1357 24. Herreros-Cabello A, Callejas-Hernández F, Gironès N, Fresno M. Trypanosoma
1358 Cruzi Genome: Organization, Multi-Gene Families, Transcription, and Biological
1359 Implications. Genes (Basel). 2020;11:1196. https://doi.org/10.3390/genes11101196
1360 25. Reis-Cunha JL, Coqueiro-dos-Santos A, Pimenta-Carvalho SA, Marques LP,
1361 Rodrigues-Luiz GF, Baptista RP, et al. Accessing the Variability of Multicopy Genes in
1362 Complex Genomes using Unassembled Next-Generation Sequencing Reads: The Case
1363 of Trypanosoma cruzi Multigene Families. mBio. 2022;13.
1364 https://doi.org/10.1128/mbio.02319-22
1365 26. Berná L, Rodriguez M, Chiribao ML, Parodi-Talice A, Pita S, Rijo G, et al.
1366 Expanding an expanded genome: long-read sequencing of Trypanosoma cruzi. Microb
1367 Genom. 2018;4. https://doi.org/10.1099/mgen.0.000177
1368 27. Ghedin E, Bringaud F, Peterson J, Myler P, Berriman M, Ivens A, et al. Gene
1369 synteny and evolution of genome architecture in trypanosomatids. Mol Biochem
1370 Parasitol. 2004;134:183–91. https://doi.org/10.1016/j.molbiopara.2003.11.012
1371 28. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, et al.
1372 Comparative genomics of trypanosomatid parasitic protozoa. Science. 2005;309:404–9.
1373 https://doi.org/10.1126/science.1112181
1374 29. Reis-Cunha JL, Bartholomeu DC. Trypanosoma cruzi Genome Assemblies:
1375 Challenges and Milestones of Assembling a Highly Repetitive and Complex Genome.
1376 2019. p. 1–22. https://doi.org/10.1007/978-1-4939-9148-8_1

30. Müller LSM, Cosentino RO, Förstner KU, Guizetti J, Wedel C, Kaplan N, et al. Genome organization and DNA accessibility control antigenic variation in trypanosomes. Nature. 2018;563:121–5. https://doi.org/10.1038/s41586-018-0619-8
31. Vitarelli M de O, Franco TA, Pires D da S, Lima ARJ, Viala VL, Kraus AJ, et al. Integrating high-throughput analysis to create an atlas of replication origins in *Trypanosoma cruzi* in the context of genome structure and variability. mBio. 2024;15. https://doi.org/10.1128/mbio.00319-24
32. Ferreira L de L, Araújo FF de, Martinelli PM, Teixeira-Carvalho A, Alves-Silva J, Guarneri AA. New features on the survival of human-infective Trypanosoma rangeli in a murine model: Parasite accumulation is observed in lymphoid organs. PLoS Negl Trop Dis. 2020;14:e0009015. https://doi.org/10.1371/journal.pntd.0009015
33. Medina-Acosta E, Franco AM, Jansen AM, Sampol M, Nevés N, Pontes-de-Carvalho L, et al. Trans-sialidase and sialidase activities discriminate between morphologically indistinguishable trypanosomatids. Eur J Biochem. 1994;225:333–9. https://doi.org/10.1111/j.1432-1033.1994.00333.x
34. Pontes-de-Carvalho LC, Tomlinson S, Nussenzweig V. Trypanosoma rangeli sialidase lacks trans-sialidase activity. Mol Biochem Parasitol. 1993;62:19–25. https://doi.org/10.1016/0166-6851(93)90173-U
35. Buschiazzo A. Structural basis of sialyltransferase activity in trypanosomal sialidases. EMBO J. 2000;19:16–24. https://doi.org/10.1093/emboj/19.1.16
36. Matos GM, Lewis MD, Talavera-López C, Yeo M, Grisard EC, Messenger LA, et al. Microevolution of Trypanosoma cruzi reveals hybridization and clonal mechanisms driving rapid genome diversification. Elife. 2022;11. https://doi.org/10.7554/eLife.75237
37. Schlagenhauf E, Etges R, Metcalf P. The crystal structure of the Leishmania major surface proteinase leishmanolysin (gp63). Structure. 1998;6:1035–46. https://doi.org/10.1016/S0969-2126(98)00104-X
38. Guay-Vincent M-M, Matte C, Berthiaume A-M, Olivier M, Jaramillo M, Descoteaux A. Revisiting Leishmania GP63 host cell targets reveals a limited spectrum of substrates. PLoS Pathog. 2022;18:e1010640. https://doi.org/10.1371/journal.ppat.1010640
39. Joshi PB, Kelly BL, Kamhawi S, Sacks DL, McMaster WR. Targeted gene deletion in Leishmania major identifies leishmanolysin (GP63) as a virulence factor. Mol Biochem Parasitol. 2002;120:33–40. https://doi.org/10.1016/S0166-6851(01)00432-7
40. Yao C, Chen B, Kang Z, Liu Y, Qi X, Wang Q. Binding, selectivity and sequence recognition of matrix metalloproteinase-2 to oligopeptides. J Biomol Struct Dyn. 2020;38:275–82. https://doi.org/10.1080/07391102.2019.1578694
41. Varghese A, Chaturvedi SS, Fields GB, Karabencheva-Christova TG. A synergy between the catalytic and structural Zn(II) ions and the enzyme and substrate dynamics underlies the structure–function relationships of matrix metalloproteinase collagenolysis.

1416 JBIC Journal of Biological Inorganic Chemistry. 2021;26:583–97.
1417 https://doi.org/10.1007/s00775-021-01876-6
1418 42. Jendele L, Krivak R, Skoda P, Novotny M, Hoksza D. PrankWeb: a web server for
1419 ligand binding site prediction and visualization. Nucleic Acids Res. 2019;47:W345–9.
1420 https://doi.org/10.1093/nar/gkz424
1421 43. Cuevas IC, Cazzulo JJ, Sánchez DO. gp63 Homologues in *Trypanosoma cruzi* :
1422 Surface Antigens with Metalloprotease Activity and a Possible Role in Host Cell
1423 Infection. Infect Immun. 2003;71:5739–49. https://doi.org/10.1128/IAI.71.10.5739-
1424 5749.2003
1425 44. Nam KH. Structural analysis of metal chelation of the metalloproteinase thermolysin
1426 by 1,10-phenanthroline. J Inorg Biochem. 2021;215:111319.
1427 https://doi.org/10.1016/j.jinorgbio.2020.111319
1428 45. Caeiro LD, Alba-Soto CD, Rizzi M, Solana ME, Rodriguez G, Chidichimo AM, et al.
1429 The protein family TcTASV-C is a novel Trypanosoma cruzi virulence factor secreted in
1430 extracellular vesicles by trypomastigotes and highly expressed in bloodstream forms.
1431 PLoS Negl Trop Dis. 2018;12:1–26. https://doi.org/10.1371/journal.pntd.0006475
1432 46. Canepa GE, Mesias AC, Yu H, Chen X, Buscaglia CA, Cánepa GE, et al. Structural
1433 features affecting trafficking, processing, and secretion of Trypanosoma cruzi mucins. J
1434 Biol Chem. 2012/06/19. 2012;287:26365–76. https://doi.org/10.1074/jbc.M112.354696
1435 47. Di Noia JM, D'Orso I, Sánchez DO, Frasch ACC. AU-rich Elements in the 3′-
1436 Untranslated Region of a New Mucin-type Gene Family of Trypanosoma cruzi Confers
1437 mRNA Instability and Modulates Translation Efficiency. Journal of Biological Chemistry.
1438 2000;275:10218–27. https://doi.org/10.1074/jbc.275.14.10218
1439 48. García EA, Ziliani M, Agüero F, Bernabó G, Sánchez DO, Tekiel V. TcTASV: A
1440 novel protein family in Trypanosoma cruzi identified from a subtractive trypomastigote
1441 cDNA library. PLoS Negl Trop Dis. 2010;4. https://doi.org/10.1371/journal.pntd.0000841
1442 49. Díaz-Viraqué F, Pita S, Greif G, de Souza R de CM, Iraola G, Robello C. Nanopore
1443 Sequencing Significantly Improves Genome Assembly of the Protozoan Parasite
1444 Trypanosoma cruzi. Genome Biol Evol. 2019;11:1952–7.
1445 https://doi.org/10.1093/gbe/evz129
1446 50. Callejas-Hernández F, Gironès N, Fresno M. Genome Sequence of *Trypanosoma*
1447 *cruzi* Strain Bug2148. Genome Announc. 2018;6.
1448 https://doi.org/10.1128/genomeA.01497-17
1449 51. Wang W, Peng D, Baptista RP, Li Y, Kissinger JC, Tarleton RL. Strain-specific
1450 genome evolution in Trypanosoma cruzi, the agent of Chagas disease. PLoS Pathog.
1451 2021;17:e1009254. https://doi.org/10.1371/journal.ppat.1009254
1452 52. Ramirez JL. An Evolutionary View of Trypanosoma Cruzi Telomeres. Front Cell
1453 Infect Microbiol. 2020;9. https://doi.org/10.3389/fcimb.2019.00439

1454 53. Li B. Keeping Balance Between Genetic Stability and Plasticity at the Telomere and
1455 Subtelomere of Trypanosoma brucei. Front Cell Dev Biol. 2021;9.
1456 https://doi.org/10.3389/fcell.2021.699639
1457 54. Freitas LM, dos Santos SL, Rodrigues-Luiz GF, Mendes TAO, Rodrigues TS,
1458 Gazzinelli RT, et al. Genomic Analyses, Gene Expression and Antigenic Profile of the
1459 Trans-Sialidase Superfamily of Trypanosoma cruzi Reveal an Undetected Level of
1460 Complexity. PLoS One. 2011;6:e25914. https://doi.org/10.1371/journal.pone.0025914
1461 55. Borst P. Discontinuous Transcription and Antigenic Variation In Trypanosomes.
1462 Annu Rev Biochem. 1986;55:701–32.
1463 https://doi.org/10.1146/annurev.bi.55.070186.003413
1464 56. Mayer MG, Santos MG dos, Silva MFL da, Floeter-Winter LM. Footprints of a
1465 trypanosomatid RNA world: pre-small subunit rRNA processing by spliced leader
1466 addition trans-splicing. Mem Inst Oswaldo Cruz. 2012;107:522–31.
1467 https://doi.org/10.1590/S0074-02762012000400013
1468 57. Parsons M, Nelson RG, Agabian N. The trypanosome spliced leader small RNA
1469 gene family: stage-specific modification of one of several similar dispersed genes.
1470 Nucleic Acids Res. 1986;14:1703–17. https://doi.org/10.1093/nar/14.4.1703
1471 58. Añez-Rojas N, Peralta A, Crisante G, Rojas A, Añez N, Ramírez JL, et al.
1472 Trypanosoma rangeli expresses a gene of the group II trans-sialidase superfamily. Mol
1473 Biochem Parasitol. 2005;142:133–6. https://doi.org/10.1016/j.molbiopara.2005.03.012
1474 59. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et
1475 al. The Genome of the African Trypanosome *Trypanosoma brucei*. Science (1979).
1476 2005;309:416–22. https://doi.org/10.1126/science.1112642
1477 60. Kawashita SY, da Silva C V., Mortara RA, Burleigh BA, Briones MRS. Homology,
1478 paralogy and function of DGF-1, a highly dispersed Trypanosoma cruzi specific gene
1479 family and its implications for information entropy of its encoded proteins. Mol Biochem
1480 Parasitol. 2009;165:19–31. https://doi.org/10.1016/j.molbiopara.2008.12.010
1481 61. Ramírez JL. The Elusive Trypanosoma cruzi Disperse Gene Protein Family (DGF-
1482 1). Pathogens. 2023;12:292. https://doi.org/10.3390/pathogens12020292
1483 62. Liao X, Hu K, Salhi A, Zou Y, Wang J, Gao X. msRepDB: a comprehensive
1484 repetitive sequence database of over 80 000 species. Nucleic Acids Res.
1485 2022;50:D236–45. https://doi.org/10.1093/nar/gkab1089
1486 63. KASHI Y, KING D. Simple sequence repeats as advantageous mutators in
1487 evolution. Trends in Genetics. 2006;22:253–9. https://doi.org/10.1016/j.tig.2006.03.005
1488 64. Hall JPJ, Wang H, Barry JD. Mosaic VSGs and the Scale of Trypanosoma brucei
1489 Antigenic Variation. PLoS Pathog. 2013;9:e1003502.
1490 https://doi.org/10.1371/journal.ppat.1003502
1491 65. Claessens A, Hamilton WL, Kekre M, Otto TD, Faizullabhoy A, Rayner JC, et al.
1492 Generation of Antigenic Diversity in Plasmodium falciparum by Structured

1493 Rearrangement of Var Genes During Mitosis. PLoS Genet. 2014;10:e1004812.
1494 https://doi.org/10.1371/journal.pgen.1004812
1495 66. Morel CM, Acharya T, Broun D, Dangi A, Elias C, Ganguly NK, et al. Health
1496 Innovation Networks to Help Developing Countries Address Neglected Diseases.
1497 Science (1979). 2005;309:401–4. https://doi.org/10.1126/science.1115538
1498 67. Maia da Silva F, Marcili A, Lima L, Cavazzana M, Ortiz PA, Campaner M, et al.
1499 Trypanosoma rangeli isolates of bats from Central Brazil: Genotyping and phylogenetic
1500 analysis enable description of a new lineage using spliced-leader gene sequences. Acta
1501 Trop. 2009;109:199–207. https://doi.org/10.1016/j.actatropica.2008.11.005
1502 68. Abrahim M, Machado E, Alvarez-Valín F, de Miranda AB, Catanho M. Uncovering
1503 Pseudogenes and Intergenic Protein-coding Sequences in TriTryps' Genomes. Genome
1504 Biol Evol. 2022;14. https://doi.org/10.1093/gbe/evac142
1505 69. Reis-Cunha JL, Valdivia HO, Bartholomeu DC. Gene and Chromosomal Copy
1506 Number Variations as an Adaptive Mechanism Towards a Parasitic Lifestyle in
1507 Trypanosomatids. Curr Genomics. 2018;19.
1508 https://doi.org/10.2174/1389202918666170911161311
1509 70. Schwabl P, Imamura H, Van den Broeck F, Costales JA, Maiguashca-Sánchez J,
1510 Miles MA, et al. Meiotic sex in Chagas disease parasite Trypanosoma cruzi. Nat
1511 Commun. 2019;10:3972. https://doi.org/10.1038/s41467-019-11771-z
1512 71. Bussotti G, Gouzelou E, Côrtes Boité M, Kherachi I, Harrat Z, Eddaikra N, et al.
1513 *Leishmania* Genome Dynamics during Environmental Adaptation Reveal Strain-Specific
1514 Differences in Gene Copy Number Variation, Karyotype Instability, and Telomeric
1515 Amplification. mBio. 2018;9. https://doi.org/10.1128/mBio.01399-18
1516 72. Negreira GH, de Groote R, Van Giel D, Monsieurs P, Maes I, de Muylder G, et al.
1517 The adaptive roles of aneuploidy and polyclonality in *Leishmania* in response to
1518 environmental stress. EMBO Rep. 2023;24. https://doi.org/10.15252/embr.202357413
1519 73. Grünebast J, Clos J. Leishmania: Responding to environmental signals and
1520 challenges without regulated transcription. Comput Struct Biotechnol J. 2020;18:4016–
1521 23. https://doi.org/10.1016/j.csbj.2020.11.058
1522 74. Cabrine-Santos M, Ferreira KAM, Tosi LRO, Lages-Silva E, Ramírez LE, Pedrosa
1523 AL. Karyotype variability in KP1(+) and KP1(−) strains of Trypanosoma rangeli isolated
1524 in Brazil and Colombia. Acta Trop. 2009;110:57–64.
1525 https://doi.org/10.1016/j.actatropica.2009.01.004
1526 75. Henriksson J, Solari A, Rydåker M, Sousa OE, Pettersson U. Karyotype variability in
1527 *Trypanosoma rangeli*. Parasitology. 1996;112:385–91.
1528 https://doi.org/10.1017/S0031182000066610
1529 76. Lye L-F, Owens K, Shi H, Murta SMF, Vieira AC, Turco SJ, et al. Retention and
1530 Loss of RNA Interference Pathways in Trypanosomatid Protozoans. PLoS Pathog.
1531 2010;6:e1001161. https://doi.org/10.1371/journal.ppat.1001161

1532 77. Reis-Cunha JL, Pimenta-Carvalho SA, Almeida L V., Coqueiro-dos-Santos A,
1533 Marques CA, Black JA, et al. Ancestral aneuploidy and stable chromosomal duplication
1534 resulting in differential genome structure and gene expression control in trypanosomatid
1535 parasites. Genome Res. 2024; https://doi.org/10.1101/gr.278550.123
1536 78. Valdivia HO, Scholte LLS, Oliveira G, Gabaldón T, Bartholomeu DC. The
1537 Leishmania metaphylome: a comprehensive survey of Leishmania protein phylogenetic
1538 relationships. BMC Genomics. 2015;16:887. https://doi.org/10.1186/s12864-015-2091-2
1539 79. Castro Neto AL, Brito ANALM, Rezende AM, Magalhães FB, de Melo Neto OP. In
1540 silico characterization of multiple genes encoding the GP63 virulence protein from
1541 Leishmania braziliensis: identification of sources of variation and putative roles in
1542 immune evasion. BMC Genomics. 2019;20:118. https://doi.org/10.1186/s12864-019-
1543 5465-z
1544 80. LaCount DJ, Gruszynski AE, Grandgenett PM, Bangs JD, Donelson JE. Expression
1545 and Function of the Trypanosoma brucei Major Surface Protease (GP63) Genes.
1546 Journal of Biological Chemistry. 2003;278:24658–64.
1547 https://doi.org/10.1074/jbc.M301451200
1548 81. Grandgenett PM, Otsu K, Wilson HR, Wilson ME, Donelson JE. A Function for a
1549 Specific Zinc Metalloprotease of African Trypanosomes. PLoS Pathog. 2007;3:e150.
1550 https://doi.org/10.1371/journal.ppat.0030150
1551 82. Choi JY, Chung E. Molecular Dynamics Simulations of Matrix Metalloproteinase 13
1552 and the Analysis of the Specificity Loop and the S1′−Site. Int J Mol Sci. 2023;24:10577.
1553 https://doi.org/10.3390/ijms241310577
1554 83. Matheoud D, Moradin N, Bellemare-Pelletier A, Shio MT, Hong WJ, Olivier M, et al.
1555 Leishmania Evades Host Immunity by Inhibiting Antigen Cross-Presentation through
1556 Direct Cleavage of the SNARE VAMP8. Cell Host Microbe. 2013;14:15–25.
1557 https://doi.org/10.1016/j.chom.2013.06.003
1558 84. Contreras I, Gómez MA, Nguyen O, Shio MT, McMaster RW, Olivier M. Leishmania-
1559 Induced Inactivation of the Macrophage Transcription Factor AP-1 Is Mediated by the
1560 Parasite Metalloprotease GP63. PLoS Pathog. 2010;6:e1001148.
1561 https://doi.org/10.1371/journal.ppat.1001148
1562 85. Brittingham A, Chen G, McGwire BS, Chang K-P, Mosser DM. Interaction of
1563 *Leishmania* gp63 with Cellular Receptors for Fibronectin. Infect Immun. 1999;67:4477–
1564 84. https://doi.org/10.1128/IAI.67.9.4477-4484.1999
1565 86. Santos ALS, Branquinha MH, D'Avila-Levy CM. The ubiquitous gp63-like
1566 metalloprotease from lower trypanosomatids: in the search for a function. An Acad Bras
1567 Cienc. 2006;78:687–714. https://doi.org/10.1590/S0001-37652006000400006
1568 87. Osorio Y, Travi BL, Palma GI, Saravia NG. Infectivity of Trypanosoma rangeli in a
1569 promonocytic mammalian cell line. Journal of Parasitology. 1995;81:687–93.
1570 https://doi.org/10.2307/3283955

1571 88. Bueren-Calabuig JA, Pierdominici-Sottile G, Roitberg AE. Unraveling the differences
1572 of the hydrolytic activity of Trypanosoma cruzi trans-sialidase and Trypanosoma rangeli
1573 sialidase: a quantum mechanics-molecular mechanics modeling study. J Phys Chem B.
1574 2014/05/13. 2014;118:5807–16. https://doi.org/10.1021/jp412294r
1575 89. Burle-Caldas G de A, dos Santos NSA, de Castro JT, Mugge FLB, Grazielle-Silva V,
1576 Oliveira AER, et al. Disruption of Active Trans-Sialidase Genes Impairs Egress from
1577 Mammalian Host Cells and Generates Highly Attenuated Trypanosoma cruzi Parasites.
1578 mBio. 2022;13. https://doi.org/10.1128/mbio.03478-21
1579 90. Caeiro LD, Masip YE, Rizzi M, Rodríguez ME, Pueblas Castro C, Sánchez DO, et
1580 al. The Trypanosoma cruzi TcTASV-C protein subfamily administrated with U-Omp19
1581 promotes a protective response against a lethal challenge in mice. Vaccine.
1582 2020;38:7645–53. https://doi.org/10.1016/j.vaccine.2020.10.006
1583 91. Bernabó G, Levy G, Ziliani M, Caeiro LD, Sánchez DO, Tekiel V. TcTASV-C, a
1584 Protein Family in Trypanosoma cruzi that Is Predominantly Trypomastigote-Stage
1585 Specific and Secreted to the Medium. PLoS One. 2013;8:1–10.
1586 https://doi.org/10.1371/journal.pone.0071192
1587 92. Floridia-Yapur N, Monje Rumi M, Ragone P, Lauthier J, Tomasini N, Alberti
1588 D'Amato A, et al. The TcTASV proteins are novel promising antigens to detect active
1589 *Trypanosoma cruzi* infection in dogs. Parasitology. 2016;143:1382–9.
1590 https://doi.org/10.1017/S0031182016000822
1591 93. Queiroz RML, Charneau S, Bastos IMD, Santana JM, Sousa M V., Roepstorff P, et
1592 al. Cell Surface Proteome Analysis of Human-Hosted *Trypanosoma cruzi* Life Stages. J
1593 Proteome Res. 2014;13:3530–41. https://doi.org/10.1021/pr401120y
1594 94. Acosta-Serrano A. The mucin-like glycoprotein super-family of Trypanosoma cruzi:
1595 structure and biological roles. Mol Biochem Parasitol. 2001;114:143–50.
1596 https://doi.org/10.1016/S0166-6851(01)00245-6
1597 95. Campo VA, Buscaglia CA, Di Noia JM, Frasch ACC. Immunocharacterization of the
1598 mucin-type proteins from the intracellular stage of Trypanosoma cruzi. Microbes Infect.
1599 2006;8:401–9. https://doi.org/10.1016/j.micinf.2005.07.008
1600 96. Santos LF, Rocha F de S, Lorenzo MG, Guarneri AA. Revisiting the development of
1601 Trypanosoma rangeli in the vertebrate host. Mem Inst Oswaldo Cruz. 2024;119.
1602 https://doi.org/10.1590/0074-02760240138
1603 97. Frevert U, Schenkman S, Nussenzweig V. Stage-specific expression and
1604 intracellular shedding of the cell surface trans-sialidase of Trypanosoma cruzi. Infect
1605 Immun. 1992;60:2349–60. https://doi.org/10.1128/iai.60.6.2349-2360.1992
1606 98. Buschiazzo A, Campetella O, Frasch ACC. Trypanosoma rangeli sialidase: Cloning,
1607 expression and similarity to T. cruzi trans-sialidase. Glycobiology. 1998/02/10.
1608 1997;7:1167–73. https://doi.org/10.1093/glycob/7.8.1167

1609 99. Jers C, Michalak M, Larsen DM, Kepp KP, Li H, Guo Y, et al. Rational design of a
1610 new Trypanosoma rangeli trans-sialidase for efficient sialylation of glycans. PLoS One.
1611 2014/01/10. 2014;9:e83902. https://doi.org/10.1371/journal.pone.0083902
1612 100. Vallejo GA, Guhl F, Carranza JC, Lozano LE, Sánchez JL, Jaramillo JC, et al.
1613 kDNA markers define two major Trypanosoma rangeli lineages in Latin-America. Acta
1614 Trop. 2002;81:77–82. https://doi.org/10.1016/S0001-706X(01)00186-3
1615 101. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
1616 sequence data. Bioinformatics. 2014;30:2114–20.
1617 https://doi.org/10.1093/bioinformatics/btu170
1618 102. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu:
1619 scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat
1620 separation. Genome Res. 2017;27:722–36. https://doi.org/10.1101/gr.215087.116
1621 103. Boetzer M, Henkel C V., Jansen HJ, Butler D, Pirovano W. Scaffolding pre-
1622 assembled contigs using SSPACE. Bioinformatics. 2011;27:578–9.
1623 https://doi.org/10.1093/bioinformatics/btq683
1624 104. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome
1625 Biol. 2012;13:R56. https://doi.org/10.1186/gb-2012-13-6-r56
1626 105. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An
1627 Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly
1628 Improvement. PLoS One. 2014;9:e112963.
1629 https://doi.org/10.1371/journal.pone.0112963
1630 106. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig
1631 reassignment for third-gen diploid genome assemblies. BMC Bioinformatics.
1632 2018;19:460. https://doi.org/10.1186/s12859-018-2485-7
1633 107. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for
1634 genome assemblies. Bioinformatics. 2013;29:1072–5.
1635 https://doi.org/10.1093/bioinformatics/btt086
1636 108. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4:
1637 A fast and versatile genome alignment system. PLoS Comput Biol. 2018;14:e1005944.
1638 https://doi.org/10.1371/journal.pcbi.1005944
1639 109. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al.
1640 BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
1641 https://doi.org/10.1186/1471-2105-10-421
1642 110. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: Assessing Genomic
1643 Data Quality and Beyond. Curr Protoc. 2021;1. https://doi.org/10.1002/cpz1.323
1644 111. Weatherly DB, Boehlke C, Tarleton RL. Chromosome level assembly of the hybrid
1645 Trypanosoma cruzi genome. BMC Genomics. 2009;10:255.
1646 https://doi.org/10.1186/1471-2164-10-255

112. Maia GA, Filho VB, Kawagoe EK, Teixeira Soratto TA, Moreira RS, Grisard EC, et al. AnnotaPipeline: An integrated tool to annotate eukaryotic proteins using multi-omics data. Front Genet. 2022;13. https://doi.org/10.3389/fgene.2022.1020100
113. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics. 2003;19:ii215–25. https://doi.org/10.1093/bioinformatics/btg1080
114. Bateman A, Martin M-J, Orchard S, Magrane M, Agivetova R, Ahmad S, et al. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49:D480–9. https://doi.org/10.1093/nar/gkaa1100
115. Amos B, Aurrecoechea C, Barba M, Barreto A, Basenko EY, Bażant W, et al. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. Nucleic Acids Res. 2022;50:D898–911. https://doi.org/10.1093/nar/gkab929
116. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40. https://doi.org/10.1093/bioinformatics/btu031
117. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525–7. https://doi.org/10.1038/nbt.3519
118. Eng JK, Jahan TA, Hoopmann MR. Comet: An open-source <scp>MS</scp> / <scp>MS</scp> sequence database search tool. Proteomics. 2013;13:22–4. https://doi.org/10.1002/pmic.201200439
119. The M, MacCoss MJ, Noble WS, Käll L. Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. J Am Soc Mass Spectrom. 2016;27:1719–27. https://doi.org/10.1007/s13361-016-1460-7
120. Moreira RS, Benetti Filho V, Maia GA, Soratto TAT, Kawagoe EK, Russi BC, et al. FastProtein—an automated software for *in silico* proteomic analysis. PeerJ. 2024;12:e18309. https://doi.org/10.7717/peerj.18309
121. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 2018;46:D335–42. https://doi.org/10.1093/nar/gkx1038
122. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–5. https://doi.org/10.1093/bioinformatics/btt509
123. Sweeney BA, Petrov AI, Burkov B, Finn RD, Bateman A, Szymanski M, et al. RNAcentral: a hub of information for non-coding RNA sequences. Nucleic Acids Res. 2019;47:D221–9. https://doi.org/10.1093/nar/gky1034
124. Smit AFA, Hubley R, Green P. RepeatMasker [Internet]. 2025 [cited 2025 Aug 18]. https://www.repeatmasker.org/. Accessed 18 Aug 2025
125. Brown MR, Manuel Gonzalez de La Rosa P, Blaxter M. tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. Bioinformatics. 2025;41. https://doi.org/10.1093/bioinformatics/btaf049

126. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One. 2016;11:e0163962. https://doi.org/10.1371/journal.pone.0163962
127. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25:1189–91. https://doi.org/10.1093/bioinformatics/btp033
128. Nielsen H, Teufel F, Brunak S, von Heijne G. SignalP: The Evolution of a Web Server. 2024. p. 331–67. https://doi.org/10.1007/978-1-0716-4007-4_17
129. Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H, et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. 2022. https://doi.org/10.1101/2022.04.08.487609
130. Gíslason MH, Nielsen H, Almagro Armenteros JJ, Johansen AR. Prediction of GPI-anchored proteins with pointer neural networks. Curr Res Biotechnol. 2021;3:6–13. https://doi.org/10.1016/j.crbiot.2021.01.001
131. Button LL, Russell DG, Klein HL, Medina-Acosta E, Karess RE, McMaster WR. Genes encoding the major surface glycoprotein in Leishmania are tandemly linked at a single chromosomal locus and are constitutively transcribed. Mol Biochem Parasitol. 1989;32:271–83. https://doi.org/10.1016/0166-6851(89)90076-5
132. Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, et al. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. Genome Res. 2011;21:2129–42. https://doi.org/10.1101/gr.122945.111
133. Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S, et al. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. Nucleic Acids Res. 2019;47:W260–5. https://doi.org/10.1093/nar/gkz303
134. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res [Internet]. 2019;47:W256–9. https://doi.org/10.1093/nar/gkz239
135. Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, et al. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. Nucleic Acids Res. 2024;52:W521–5. https://doi.org/10.1093/nar/gkae241
136. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. Nucleic Acids Res. 2012;41:D344–7. https://doi.org/10.1093/nar/gks1067
137. Kong AT, Leprevost F V, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. Nat Methods. 2017;14:513–20. https://doi.org/10.1038/nmeth.4256
138. da Veiga Leprevost F, Haynes SE, Avtonomov DM, Chang H-Y, Shanmugam AK, Mellacheruvu D, et al. Philosopher: a versatile toolkit for shotgun proteomics data analysis. Nat Methods. 2020;17:869–70. https://doi.org/10.1038/s41592-020-0912-y

1726 139. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate
1727 structure prediction of biomolecular interactions with AlphaFold 3. Nature.
1728 2024;630:493–500. https://doi.org/10.1038/s41586-024-07487-w
1729 140. Devel L, Czarny B, Beau F, Georgiadis D, Stura E, Dive V. Third generation of
1730 matrix metalloprotease inhibitors: Gain in selectivity by targeting the depth of the S1′
1731 cavity. Biochimie. 2010;92:1501–8. https://doi.org/10.1016/j.biochi.2010.07.017
1732 141. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking
1733 with a new scoring function, efficient optimization, and multithreading. J Comput Chem.
1734 2010;31:455–61. https://doi.org/10.1002/jcc.21334
1735 142. Di Noia JM, D'Orso I, Aslund L, Sánchez DO, Frasch AC. The Trypanosoma cruzi
1736 mucin family is transcribed from hundreds of genes having hypervariable regions. J Biol
1737 Chem. 1998/06/06. 1998;273:10843–50. https://doi.org/10.1074/jbc.273.18.10843
1738 143. Bartholomeu DC, Cerqueira GC, Leão ACA, daRocha WD, Pais FS, Macedo C, et
1739 al. Genomic organization and expression profile of the mucin-associated surface protein
1740 (masp) family of the human pathogen Trypanosoma cruzi. Nucleic Acids Res.
1741 2009;37:3407–17. https://doi.org/10.1093/nar/gkp172
1742 144. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB,
1743 Schjoldager KT, et al. Precision mapping of the human O-GalNAc glycoproteome
1744 through SimpleCell technology. EMBO J. 2013/04/16. 2013;32:1478–88.
1745 https://doi.org/10.1038/emboj.2013.79
1746 145. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of
1747 eukaryotic protein phosphorylation sites. J Mol Biol. 1999;294:1351–62.
1748 https://doi.org/10.1006/jmbi.1999.3310
1749 146. Castresana J. Selection of conserved blocks from multiple alignments for their use
1750 in phylogenetic analysis. Mol Biol Evol. 2000/03/31. 2000;17:540–52.
1751 https://doi.org/10.1093/oxfordjournals.molbev.a026334
1752 147. Sievers F, Higgins DG. Clustal omega. Curr Protoc Bioinformatics. 2014/12/17.
1753 2014;48:3 13 1-16. https://doi.org/10.1002/0471250953.bi0313s48
1754 148. Wheeler TJ, Clements J, Finn RD. Skylign: a tool for creating informative,
1755 interactive logos representing sequence alignments and profile hidden Markov models.
1756 BMC Bioinformatics. 2014;15:7. https://doi.org/10.1186/1471-2105-15-7
1757 149. Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, et al. A bright
1758 monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nat
1759 Methods [Internet]. 2013;10:407–9. https://doi.org/10.1038/nmeth.2413
1760 150. DaRocha WD, Silva RA, Bartholomeu DC, Pires SF, Freitas JM, Macedo AM, et al.
1761 Expression of exogenous genes in Trypanosoma cruzi : improving vectors and
1762 electroporation protocols. Parasitol Res. 2004;92:113–20.
1763 https://doi.org/10.1007/s00436-003-1004-5

151. de Araújo-Jorge TC, de Castro SL. Doença de chagas: manual para experimentação animal. Doença de chagas: manual para experimentação animal. 2000. https://doi.org/10.7476/9788575413937

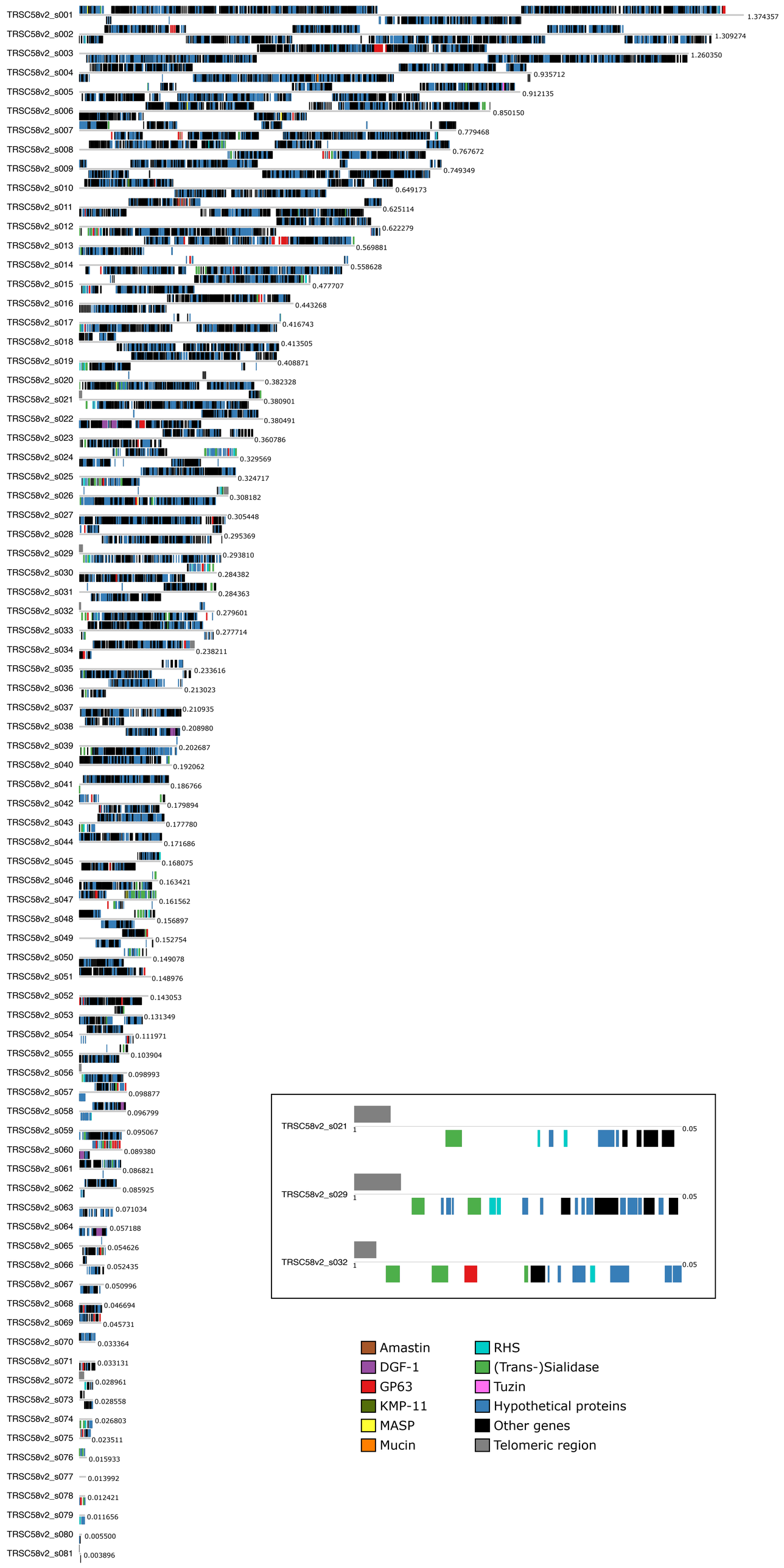
152. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238. https://doi.org/10.1186/s13059-019-1832-y

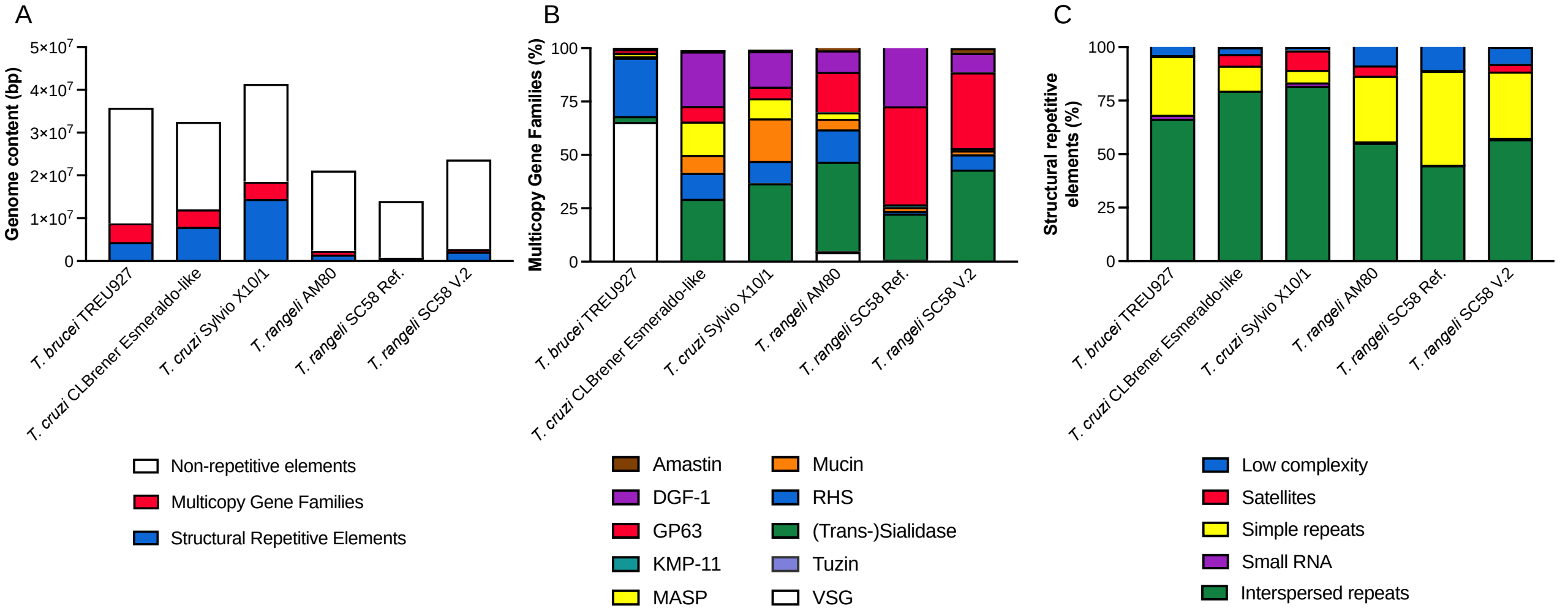
153. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324

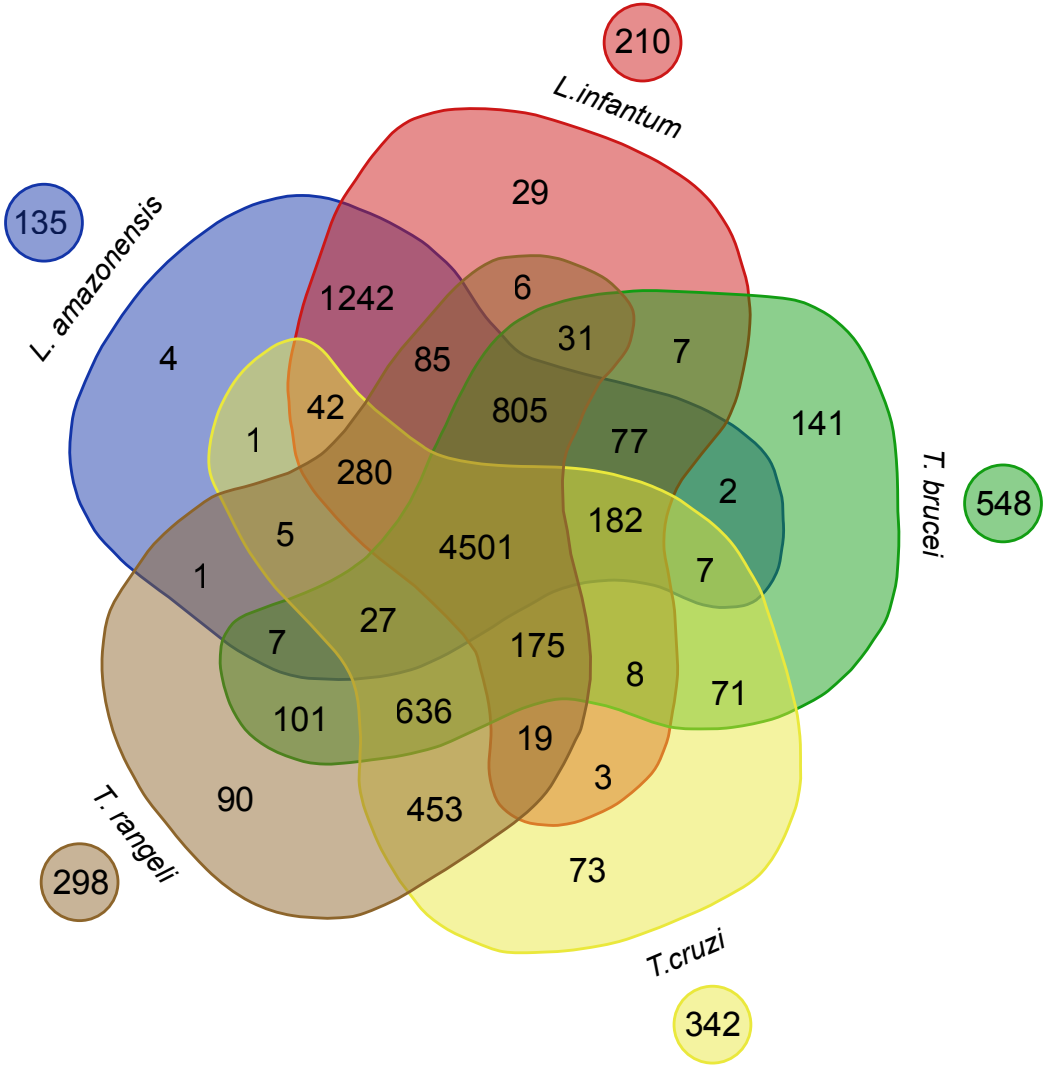
154. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352

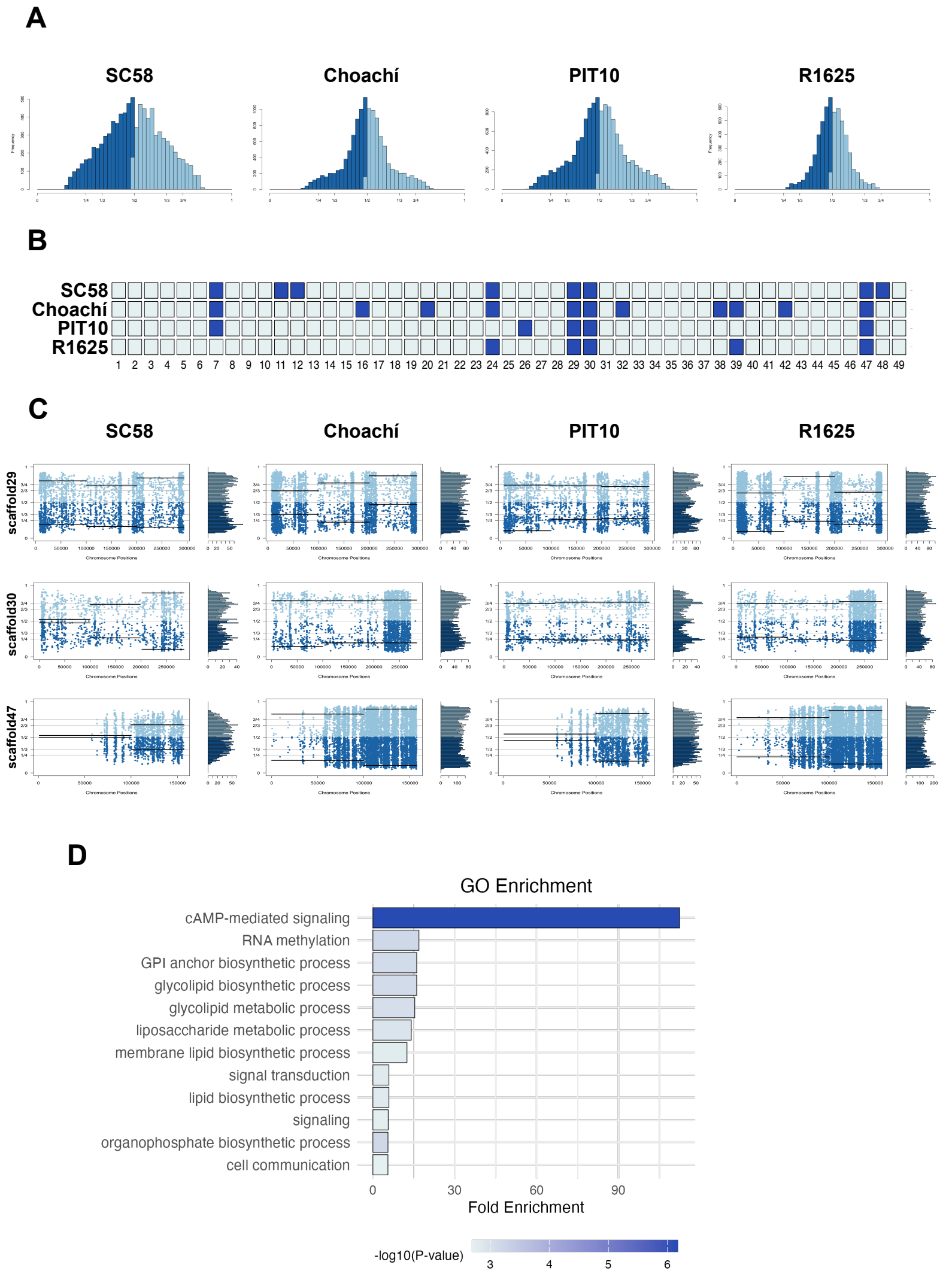
155. Knaus BJ, Grünwald NJ. <scp>vcfr</scp> : a package to manipulate and visualize variant call format data in R. Mol Ecol Resour. 2017;17:44–53. https://doi.org/10.1111/1755-0998.12549

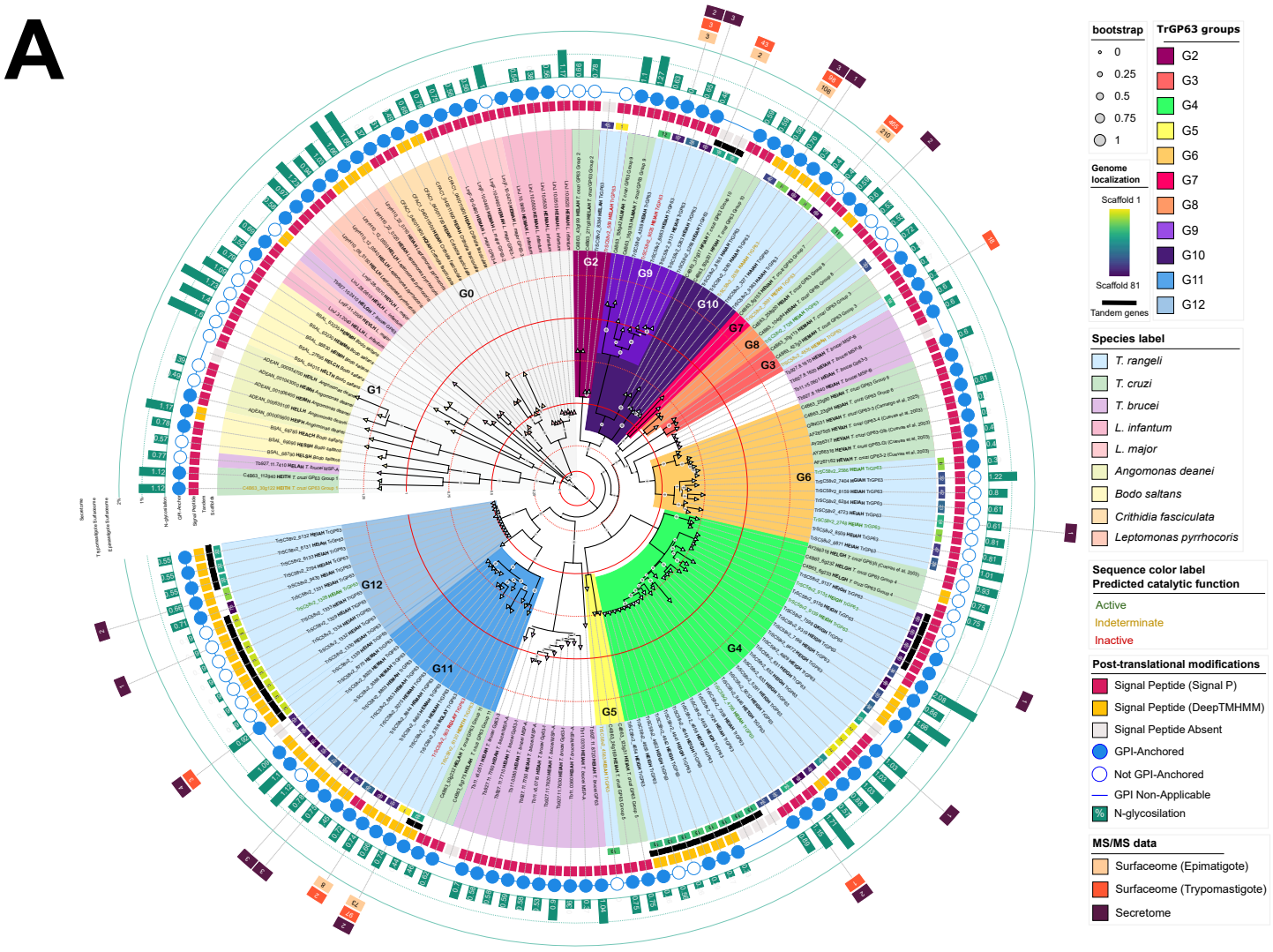
Metrics	*T. rangeli*		*T. cruzi*	
	SC58 V.2	**SC58 Reference***	**CL Brener Esmeraldo-like***	**Sylvio X10/1***
Genome size (Mbp)	23.68	14.02	32.53	41.38
Number of scaffolds	81	7,433	41	47
G+C content (%)	53.30	53.26	50.38	51.57
Scaffolds > 50 Kbp	67	0	41	47
Longest scaffold (Mbp)	1.37	0.25	2.37	3.12
Scaffold N50	558,628	2,203	870,934	1,006,492
Scaffold L50	14	2,096	12	14
***N* bases / 100 Kbp**	0	21.31	20,604.70	3,242.42
Number of rRNA	58	16	9	ND
Number of tRNA	73	120	55	65
Total number of CDS	9,203	7,475	9,039	18,643
Number of hypothetical CDS	3,067	5,075	4,260	13,760
Mean CDS size (aa)	505	460	493	178
CDS < 100 aa	299	410	207	5,573

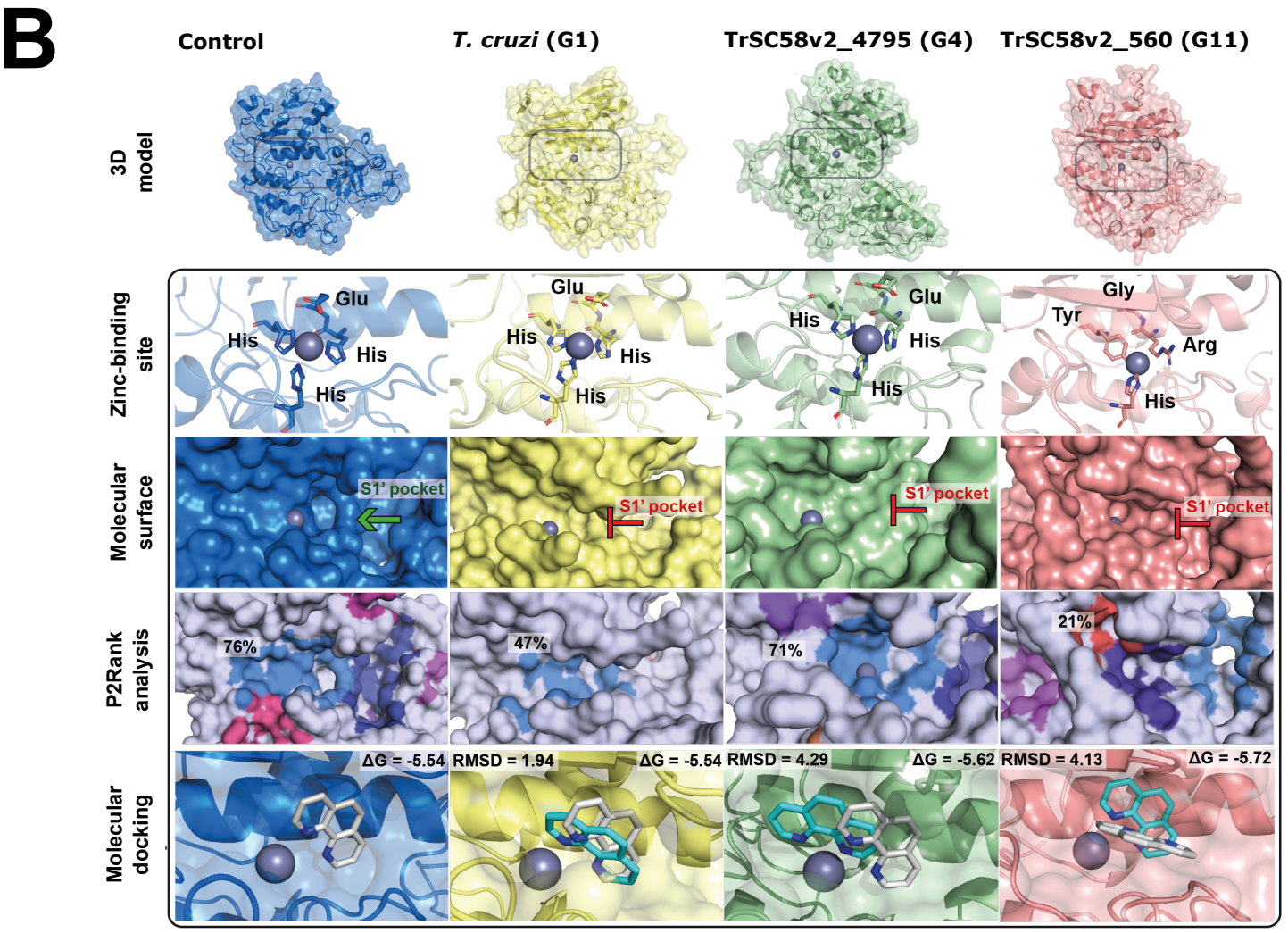

TRSC58v2_s001 1.374357
TRSC58v2_s002 1.309274
TRSC58v2_s003 1.260350
TRSC58v2_s004 0.935712
TRSC58v2_s005 0.912135
TRSC58v2_s006 0.850150
TRSC58v2_s007 0.779468
TRSC58v2_s008 0.767672
TRSC58v2_s009 0.749349
TRSC58v2_s010 0.649173
TRSC58v2_s011 0.625114
TRSC58v2_s012 0.622279
TRSC58v2_s013 0.569881
TRSC58v2_s014 0.558628
TRSC58v2_s015 0.477707
TRSC58v2_s016 0.443268
TRSC58v2_s017 0.416743
TRSC58v2_s018 0.413505
TRSC58v2_s019 0.408871
TRSC58v2_s020 0.382328
TRSC58v2_s021 0.380901
TRSC58v2_s022 0.380491
TRSC58v2_s023 0.360786
TRSC58v2_s024 0.329569
TRSC58v2_s025 0.324717
TRSC58v2_s026 0.308182
TRSC58v2_s027 0.305448
TRSC58v2_s028 0.295369
TRSC58v2_s029 0.293810
TRSC58v2_s030 0.284382
TRSC58v2_s031 0.284363
TRSC58v2_s032 0.279601
TRSC58v2_s033 0.277714
TRSC58v2_s034 0.238211
TRSC58v2_s035 0.233616
TRSC58v2_s036 0.213023
TRSC58v2_s037 0.210935
TRSC58v2_s038 0.208980
TRSC58v2_s039 0.202687
TRSC58v2_s040 0.192062
TRSC58v2_s041 0.186766
TRSC58v2_s042 0.179894
TRSC58v2_s043 0.177780
TRSC58v2_s044 0.171686
TRSC58v2_s045 0.168075
TRSC58v2_s046 0.163421
TRSC58v2_s047 0.161562
TRSC58v2_s048 0.156897
TRSC58v2_s049 0.152754
TRSC58v2_s050 0.149078
TRSC58v2_s051 0.148976
TRSC58v2_s052 0.143053
TRSC58v2_s053 0.131349
TRSC58v2_s054 0.111971
TRSC58v2_s055 0.103904
TRSC58v2_s056 0.098993
TRSC58v2_s057 0.098877
TRSC58v2_s058 0.096799
TRSC58v2_s059 0.095067
TRSC58v2_s060 0.089380
TRSC58v2_s061 0.086821
TRSC58v2_s062 0.085925
TRSC58v2_s063 0.071034
TRSC58v2_s064 0.057188
TRSC58v2_s065 0.054626
TRSC58v2_s066 0.052435
TRSC58v2_s067 0.050996
TRSC58v2_s068 0.046694
TRSC58v2_s069 0.045731
TRSC58v2_s070 0.033364
TRSC58v2_s071 0.033131
TRSC58v2_s072 0.028961
TRSC58v2_s073 0.028558
TRSC58v2_s074 0.026803
TRSC58v2_s075 0.023511
TRSC58v2_s076 0.015933
TRSC58v2_s077 0.013992
TRSC58v2_s078 0.012421
TRSC58v2_s079 0.011656
TRSC58v2_s080 0.005500
TRSC58v2_s081 0.003896
TRSC58v2_s021
1
0.05
TRSC58v2_s029
1
0.05
TRSC58v2_s032
1
0.05
Amastin
DGF-1
GP63
KMP-11
MASP
Mucin
RHS
(Trans-)Sialidase
Tuzin
Hypothetical proteins
Other genes
Telomeric region

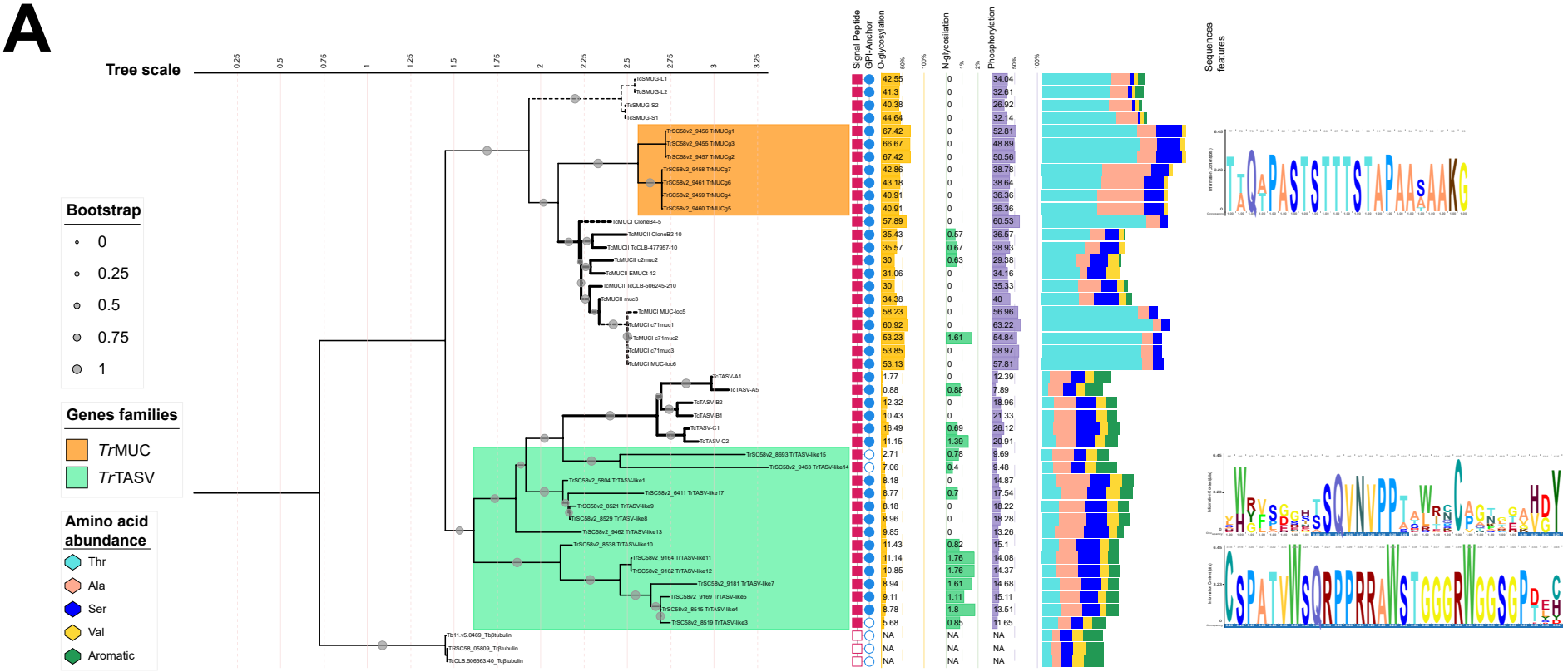

A
Genome content (bp)
5×10^7
4×10^7
3×10^7
2×10^7
1×10^7
0
T. brucei TREU927
T. cruzi CLBrener Esmeraldo-like
T. cruzi Sylvio X10/1
T. rangeli AM80
T. rangeli SC58 Ref.
T. rangeli SC58 V.2
B
Multicopy Gene Families (%)
100
75
50
25
0
C
Structural repetitive elements (%)
Non-repetitive elements
Multicopy Gene Families
Structural Repetitive Elements
Amastin
DGF-1
GP63
KMP-11
MASP
Mucin
RHS
(Trans-)Sialidase
Tuzin
VSG
Low complexity
Satellites
Simple repeats
Small RNA
Interspersed repeats

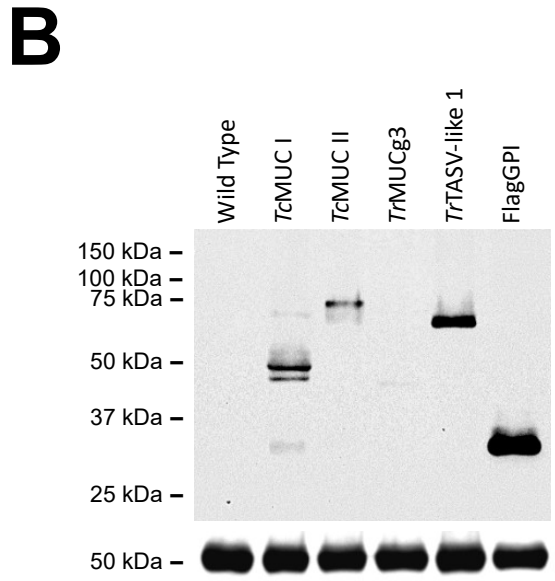

210
L.infantum
135
L. amazonensis
548
T. brucei
298
T. rangeli
342
T.cruzi
29
4
1242
6
85
31
7
42
805
141
1
77
280
2
5
182
4501
1
7
27
7
175
8
71
101
636
19
3
90
453
73

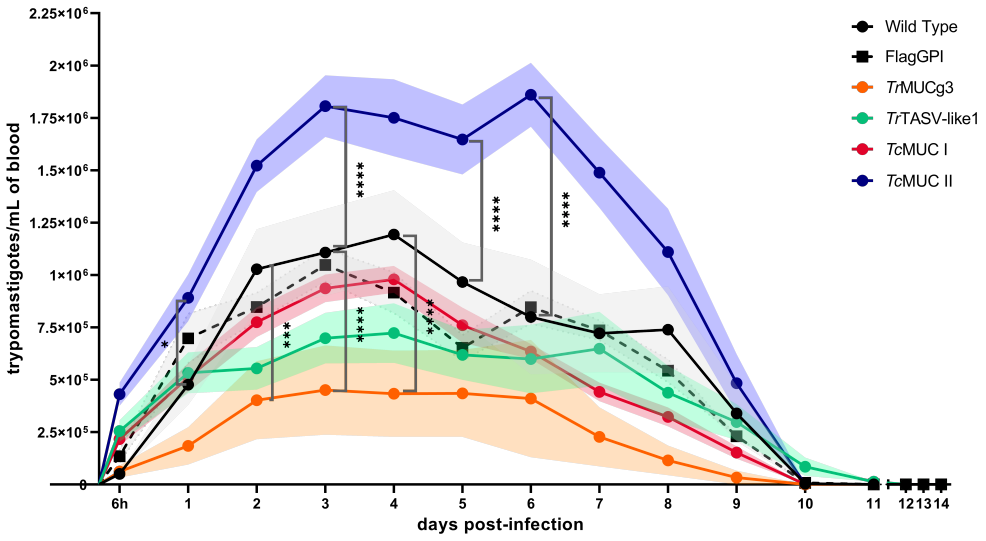

A
SC58
Choachí
PIT10
R1625
B
SC58
Choachí
PIT10
R1625
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49
C
SC58
Choachí
PIT10
R1625
scaffold29
scaffold30
scaffold47
Chromosome Positions
D
GO Enrichment
cAMP-mediated signaling
RNA methylation
GPI anchor biosynthetic process
glycolipid biosynthetic process
glycolipid metabolic process
liposaccharide metabolic process
membrane lipid biosynthetic process
signal transduction
lipid biosynthetic process
signaling
organophosphate biosynthetic process
cell communication
0
30
60
90
Fold Enrichment
-log10(P-value)
3
4
5
6


A
bootstrap
0
0.25
0.5
0.75
1
TrGP63 groups
G2
G3
G4
G5
G6
G7
G8
G9
G10
G11
G12
Genome localization
Scaffold 1
Scaffold 81
Tandem genes
Species label
T. rangeli
T. cruzi
T. brucei
L. infantum
L. major
Angomonas deanei
Bodo saltans
Crithidia fasciculata
Leptomonas pyrrhocoris
Sequence color label
Predicted catalytic function
Active
Indeterminate
Inactive
Post-translational modifications
Signal Peptide (Signal P)
Signal Peptide (DeepTMHMM)
Signal Peptide Absent
GPI-Anchored
Not GPI-Anchored
GPI Non-Applicable
N-glycosilation
MS/MS data
Surfaceome (Epimatigote)
Surfaceome (Trypomastigote)
Secretome
G0
G1
G2
G3
G4
G5
G6
G7
G8
G9
G10
G11
G12


B
Control
T. cruzi (G1)
TrSC58v2_4795 (G4)
TrSC58v2_560 (G11)
3D model
Zinc-binding site
Molecular surface
P2Rank analysis
Molecular docking
Glu
His
His
His
Glu
His
His
His
Glu
His
His
His
Gly
Tyr
Arg
His
S1' pocket
S1' pocket
S1' pocket
S1' pocket
76%
47%
71%
21%
ΔG = -5.54
RMSD = 1.94
ΔG = -5.54
RMSD = 4.29
ΔG = -5.62
RMSD = 4.13
ΔG = -5.72


A
Tree scale
Bootstrap
0
0.25
0.5
0.75
1
Genes families
TrMUC
TrTASV
Amino acid abundance
Thr
Ala
Ser
Val
Aromatic
Signal Peptide
GPI-Anchor
O-glycosylation
N-glycosilation
Phosphorylation
Sequences features


B
Wild Type
TcMUC I
TcMUC II
TrMUCg3
TrTASV-like 1
FlagGPI
150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
50 kDa


C
Wild Type
FlagGPI
TrMUCg3
TrTASV-like1
TcMUC I
TcMUC II
trypomastigotes/mL of blood
days post-infection
2.25×10⁶
2×10⁶
1.75×10⁶
1.5×10⁶
1.25×10⁶
1×10⁶
7.5×10⁵
5×10⁵
2.5×10⁵
0
6h
1
2
3
4
5
6
7
8
9
10
11
12
13
14