Genomics 111 (2019) 407–417

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

A comparative *in silico* linear B-cell epitope prediction and characterization for South American and African *Trypanosoma vivax* strains

Rafael Lucas Muniz Guedes[a,b,*], Carla Monadeli Filgueira Rodrigues[c], Nicolas Coatnoan[d], Alain Cosson[d], Fabiano Antonio Cadioli[e], Herakles Antonio Garcia[c], Alexandra Lehmkuhl Gerber[a], Rosangela Zacarias Machado[f], Paola Marcella Camargo Minoprio[d], Marta Maria Geraldes Teixeira[c], Ana Tereza Ribeiro de Vasconcelos[a]

[a] Laboratório Nacional de Computação Científica (LNCC), Av. Getúlio Vargas, 333, Petrópolis, RJ, Brazil
[b] Grupo Hermes Pardini, Setor de Pesquisa e Desenvolvimento, Vespasiano, MG, Brazil
[c] Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, SP 05508-900, Brazil
[d] Trypanosomatids Infectious Processes Laboratory, Department of Infection and Epidemiology, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris, France
[e] Departamento Clínica, Cirurgia e Reprodução Animal, Faculdade de Odontologia e Curso de Medicina Veterinária, Universidade Estadual Paulista – UNESP, Araçatuba, SP, Brazil
[f] Laboratório de Immnoparasitologia, Faculdade de Ciências Agrárias e Veterinárias (FCAV), Universidade Estadual Paulista (UNESP), Campus Jaboticabal, Jaboticabal, SP, Brazil

ARTICLE INFO

Keywords:
Trypanosoma vivax
Linear B-cell epitopes
Bloodstream forms
Bioinformatics
Transcriptomics

ABSTRACT

Trypanosoma vivax is a parasite widespread across Africa and South America. Immunological methods using recombinant antigens have been developed aiming at specific and sensitive detection of infections caused by *T. vivax*. Here, we sequenced for the first time the transcriptome of a virulent *T. vivax* strain (Lins), isolated from an outbreak of severe disease in South America (Brazil) and performed a computational integrated analysis of genome, transcriptome and *in silico* predictions to identify and characterize putative linear B-cell epitopes from African and South American *T. vivax*. A total of 2278, 3936 and 4062 linear B-cell epitopes were respectively characterized for the transcriptomes of *T. vivax* LIEM-176 (Venezuela), *T. vivax* IL1392 (Nigeria) and *T. vivax* Lins (Brazil) and 4684 for the genome of *T. vivax* Y486 (Nigeria). The results presented are a valuable theoretical source that may pave the way for highly sensitive and specific diagnostic tools.

1. Introduction

Parasites of the genus *Trypanosoma* (Trypanosomatidae: Kinetoplastea) are the etiological agents of trypanosomiasis, a group of neglected tropical diseases affecting humans and domestic animals. Animal African Trypanosomiasis (AAT), also known as nagana, is caused by *Trypanosoma vivax*, *Trypanosoma congolense*, and *Trypanosoma brucei brucei*, the two former being the most important pathogens for livestock in Africa [1,2]. *T. vivax* is disseminated across Sub-Saharan Africa and South America, with high prevalence in cattle, sheep, and buffaloes [3–6] and reported in horses, donkeys, and camels [7,8]. Tsetse flies (*Glossina* spp.) are the only vectors responsible for cyclical transmission of *T. vivax*, whereas tabanids and other biting flies are responsible for mechanical transmission in tsetse-free areas in Africa and South America [2,4,8].

T. vivax is difficult to grow in laboratory conditions, with the West African *T. vivax* IL1392 currently being the only strain well-adapted to both mice and culture, enabling experimental infections addressing parasitological, immunological, and pathological aspects [9–11]. Microsatellite and ITS rDNA genotyping revealed that despite being closely related, South American and West African *T. vivax* populations are genetically different, and they largely differ from those circulating in East Africa [12–15]. In general, West African isolates are more pathogenic to livestock than East African *T. vivax*. In South American regions of enzootic stability (Amazonian lowlands, Venezuelan Llanos, and the Brazilian Pantanal), *T. vivax* infections are mostly asymptomatic in cattle and water buffaloes. However, a recent outbreak in a water buffalo herd from an endemic setting in Venezuela [3] and the increasing number of outbreaks throughout non-endemic Brazilian regions with high mortality rates reinforce the role of *T. vivax* as a prevalent agent of nagana in South America. *T. vivax* causes hematological and neurological disorders in dairy cattle, goats, sheep, and horses

* Corresponding author at: Laboratório Nacional de Computação Científica (LNCC), Av. Getúlio Vargas, 333, Petrópolis, RJ, Brazil.
E-mail address: rmguedes@lncc.br (R.L.M. Guedes).

https://doi.org/10.1016/j.ygeno.2018.02.017
Received 5 September 2017; Received in revised form 23 February 2018; Accepted 26 February 2018
Available online 27 February 2018

[16–18]. Experimental infections of calves using *T. vivax* Lins obtained from one outbreak of severe acute disease in Brazilian dairy cattle corroborated the pathogenicity and virulence of this strain reported in field-infected animals [19].

In general, *T. vivax* is an extracellular trypanosome species, remaining mainly in the host's bloodstream, intravascular spaces, and connective tissues. However, recent studies demonstrated *T. vivax* invading the central nervous system during highly parasitemic stages in natural and experimental infections [16,20–22]. Like all other African trypanosome agents of AAT, the presence of the VSG (Variant Surface Glycoprotein) coat is fundamental for parasite evasion from the host immune response [23]. *T. vivax*, considered the most basal species of the clade constituted by trypanosome agents of AAT, displays reduced VSG repertoire compared with other trypanosomes [24,25]. A wide transcriptomic study revealed the presence of many exclusive proteins expressed in the cell-surface of *T. vivax* bloodstream forms, which are then promising candidate targets for vaccine and diagnostic approaches [11].

Diagnosis of trypanosomiasis can be performed using parasitological, molecular, and serological approaches. PCR methods [26,27] permit specific and sensitive diagnosis of *T. vivax* infection, even in animals showing extremely low parasitemias [28,29]. Serological methods enable the diagnosis of chronically infected animals, and are highly appropriate for epidemiological studies because they can be automatized and adapted to many devices suitable for large-scale field studies. However, all African trypanosomes share many antigens, and the use of crude extracts of parasites, and even purified native antigens, for immunological diagnosis revealed high cross-reactivity among all species, including *T. vivax*. Immune-diagnostic assays based on recombinant antigens also showed cross-reactivity and exhibit low sensitivity [30–34]. Antigen-capture methods have also been developed for *T. vivax* diagnosis, with previously tested methods lacking both high sensitivity and specificity [30,35]. The recently developed easy to use pen-side lateral flow test (LFT) device based on an recombinant invariant surface glycoprotein (ISG) specific of *T. vivax* (selected by proteomic approaches) improved the sensitivity and specificity (92% and 89.8%, respectively) of *T. vivax* serological diagnosis. Tests using sera from experimentally infected animals suggested that the selected ISG antigen do not routinely cross-react with *T. congolense,* a very prevalent trypanosome in cattle often mixed with *T. vivax* across Africa, and it is suitable for epidemiological surveys [36]. High species-specificity is desirable to undoubtedly distinguish *T. vivax* from other ungulate trypanosomes. A previously developed LFT based on recombinant GM6 antigen [31] exhibit lower sensitivity and cross-reacts with sera from cattle infected with multiple trypanosome species, thus being useful for pan-diagnosis of AAT. Therefore, improved serological diagnostic tests are required to detect *T. vivax* with high specificity and sensitivity permitting early and appropriated treatment of acutely (many times lethal) and chronically (progressive wasting and reproductive disorders) infected animals. The development of highly sensitive and specific method is crucial for successful strategies aiming control and tracking of *T. vivax* outbreaks and dispersal, as sometimes a single *T. vivax*-infected animal lacking patent parasitemia and showing very low levels of antibodies specific to *T. vivax,* can trigger outbreaks of high mortality. *T. vivax* is the only African trypanosome agent of AAT that can disperse out of Sub-Saharan Africa, in regions free of tsetse flies.. Consequently, the identification of species-specific linear B-cell epitopes, expressed by blood forms of African and South American strains of *T. vivax*, is urgently required for the development of more specific and sensitive immune-diagnosis of *T. vivax* infections.

T. vivax can be experimentally expanded in ruminant models such as calves, sheep, and goats. Despite the availability of culture and murine models for the *T. vivax* IL1392 strain [9,10], the production of large amounts of *T. vivax* antigens remains a very difficult task. Recombinant proteins or synthetic peptides have been valuable for species-specific serodiagnosis and serotyping (identification of infraspecific genotypes) as recently shown for *T. cruzi* [29,37,38].

The use of recombinant protein or peptide technology allowed for a great improvement over native parasite protein extracts, resulting in lower cross-reactivity and better standardization of diagnostic methods. Nowadays, designing synthetic peptides is a faster option to test the antigenicity of hundreds of peptides in a single experiment [29,39]. Bioinformatics tools can help to select potential linear B-cell epitopes expressed and/or secreted by bloodstream forms of trypanosomes that can be used to improve immunological diagnosis. The main goal of this study was to identify and characterize the best theoretical epitope candidates for the development of specific and sensitive immunological diagnosis of *T. vivax*. With this purpose, we characterized through whole transcriptome surveys the repertoires of *T. vivax*-specific linear B-cell epitopes shared by American and African isolates, as well as epitopes isolate-specific useful for genotyping.

2. Materials and methods

2.1. Trypanosoma vivax Lins samples preparation and sequencing

RNA was obtained from blood of two goats experimentally infected with the strain *T. vivax* Lins recovered from an outbreak of cattle trypanosomiasis in the municipality of Lins, state of São Paulo, Brazil [17]. The experimental infection and all animal procedures were performed at the Universidade Estadual Paulista, Jaboticabal, state of São Paulo, Brazil, as previously reported [17]. When parasitaemia reached $\sim 3 \times 10^6$ trypanosomes/ml, parasites were purified using Percoll gradient as previously described [40]. RNA preparations were obtained using Dynabeads® mRNA DIRECT™ Kit (Thermo Fisher Scientific).

The RNA-seq was performed preparing a double-stranded cDNA library with about 80 ng of RNA using the TruSeq Stranded mRNA LT Sample Preparation Kit (Illumina, San Diego, CA, USA). Library quality control was performed using the 2100 Bioanalyzer System with the Agilent High Sensitivity DNA Kit (Agilent, Santa Clara, CA, USA). The library was quantified *via* qPCR using a KAPA Library Quantification Kits for Illumina platforms (KAPA Biosystems, Wilmington, MA, USA) and sequenced in an MiSeq sequencing system (Illumina) with paired-end reads (2×75 bp) obtained using MiSeq Reagent Kits v3 (150-cycles). Samples are available at SRA repository under the accession numbers SRR5934425 and SRR5934426.

All the experimental procedures using goats were in accordance with the principles and guidelines adopted by the Brazilian College of Animal Experimentation (COBEA) and approved by the Animal Care Ethics Committee/UNESP (CEEA).

2.2. Trypanosoma vivax bloodstream transcriptomes

Transcriptome data from bloodstream forms of three *T. vivax* strains were analyzed. Transcriptome for the African *T. vivax* IL1392 strain was acquired from bloodstream forms maintained *in vivo* by continuous passages in mice [11]. Trimmomatic v0.32 [41] was used for trimming the reads. All replicates in fastq format (SRA accession numbers: ERR236859, ERR236860, ERR236861, and ERR236862) were combined and aligned to the reference genome of *T. vivax* Y486, version 2013-01-16, downloaded from TriTrypDB [42] using TopHat2 [43]. Samtools v0.1.18 [44] was used to filter reads with mapping quality ≥ 10. To assemble these reads into contigs, the PASA [45] pipeline [http://pasa.sourceforge.net/#A_ComprehensiveTranscriptome] combining *de novo* and genome-guided assemblies were applied. For each assembly, the Trinity [46] package was used. In order to filter contaminants from the *Mus musculus* host genome present at *de novo* assembly, all contigs were blasted [47] against both the *T. vivax* Y486 and *Mus musculus* genomes, removing all sequences with best hit associated with mouse sequences. Finally, to determine highly probable protein-coding regions, Transdecoder [http://transdecoder.sourceforge.net/] was run with final assembled transcripts, requiring a minimum protein

length of 30 codons.

Replicates from bloodstream transcriptome for the South American *T. vivax* Lins strain were also combined and assembled using Trinity [46] and the *Capra hircus* (GCF_000317765) genome as reference for filtering contaminants.

Reads from bloodstream transcriptome of *T. vivax* Venezuelan isolate (LIEM-176) were downloaded from the SRA database (SRR527163 and SRR527235), and used only for FPKM quantitative analysis. Assembled transcripts were downloaded from http://bioinformatica.fcien.edu.uy/Tvivax [25]. Protein-coding regions were equally predicted with Transdecoder as above. To calculate FPKM values, reads were aligned to assembled transcripts with bowtie 2 [48] and HTSeq v0.5.4p5 (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html) was used to count reads uniquely aligned to predicted protein-coding regions. Transcriptome assembly completeness was assessed with BUSCO [49] v1.22, using 429 benchmarked universal single-copy genes from the Eukaryotes dataset and e-value cutoff of 1e-05.

2.3. Prediction of cell-surface and secreted proteins

Annotated proteins from a fourth strain, *T. vivax* Y486, were downloaded from TriTrypDB [42] (Build 28) and analyzed together with predicted proteins from transcriptomic data, through the screening of both predicted cell-surface/secreted and non-cell-surface/secreted proteins, based on a procedure described by Staats et al. [50]. Briefly, an automatic PERL pipeline was used with the MySQL database. Initially, all coding sequences smaller than 30 codons were removed. To detect a signal peptide, we combined predictions by both SignalP v4.1 [51] (with default D-score cutoffs 0.45 for *SignalP-noTM networks and 0.50 for SignalP-TM networks*; http://www.cbs.dtu.dk/services/SignalP/) and TargetP v1.1 [52] (LOC = S; http://www.cbs.dtu.dk/services/TargetP/). To predict GPI-anchors, PredGPI [53] (FRate ≤ 0.005) was used (http://gpcr2.biocomp.unibo.it/predgpi/pred.htm). WoLF PSort v0.2 [54] (Extr or plas ≥ 17) and ProtComp v9.1 (LocDB and PotLocDB, using NNets an Integral prediction; http://www.softberry.com) tools were combined to infer protein localization. Finally, to assign sequences associated with the Endoplasmic Reticulum, PROSITE scan [55] was used to search for pattern PS00014 (Endoplasmic reticulum targeting sequence). Any sequence not predicted as secreted or exposed at plasma membrane by both WoLF PSort and ProtComp was classified as non-cell-surface/secreted.

2.4. In silico prediction of linear B-cell epitopes

Complete protein sequences were screened for linear B-cell epitopes using BepiPred 1.0 [56] and LBtope [57]. Usually about five amino acids in each epitope and paratope sequences are the key contributors to the binding energy, but about 15 amino acids are present in this interaction, therefore, a PERL script was developed to screen for 15mer peptide windows with BepiPred mean score ≥ 1.3 and overlapping windows were combined into one. This combination may decrease the final mean somewhat below 1.3 due to the summation of local minimums. Selected peptides were then screened with LBtope (Lbtop_Confirm as database). Complete protein sequences from predicted linear B-cell epitopes were further screened with Immune Epitopes Database Analysis Resource (IEDB) tools for the prediction and analysis of linear B-cell epitopes: Chou & Fasman Beta-Turn Prediction, Emini Surface Accessibility Prediction, Karplus & Schulz Flexibility Prediction, Kolaskar & Tongaonkar Antigenicity, and Parker Hydrophilicity Prediction [58] (http://tools.immuneepitope.org/main/). LBtope and IEDB tools were used for characterization purposes only. The median scores obtained in each method for each protein were taken as thresholds and the percentages of residues at regions corresponding to the epitopes above the thresholds were computed. Additionally, intrinsically unstructured/disordered regions were predicted with IUPred [59] for characterization using complete protein sequences, calculating the percentage of residues in predicted epitopes which individual scores were above developers indicated cutoff of ≥ 0.5.

2.5. BLAST screening

A filtering step was performed to select only *T. vivax*-specific peptides, comparing sequence similarities between genomes from other Trypanosomatid TriTrypDB genomes, using tBLASTn search. This step aimed to reduce the possibility of cross-reactivity with other species capable of infecting *T. vivax* hosts. This filtering strategy does not exclude the possibility that selected B-cell epitopes will be recognized by *T. vivax* non-infected animals sera, which can only be tested by experimental validation. The selected genomes were: *T. brucei* TREU927, *T. congolense* IL3000, *T. evansi* STIB805, *T. cruzi* CL Brener, *T. cruzi* Dm28c, *T. cruzi* Esmeraldo, *T. cruzi* JRcl4, *T. cruzi* marinkellei B7, *T. cruzi* Sylvio X10–1, *T. cruzi* Tula cl2, *L. braziliensis* MHOM/BR/75/M2904, *L. donovani* BPK282A1, *L. infantum* JPCM5, and L. *major* SD75.1. BLAST parameters were adjusted for short peptide match (Word size = 2, SEG filter off, Composition-based Statistics, e-value = 20,000, and PAM30 score matrix). Linear B-cell epitopes containing alignments with ≥15mer and ≥70% identity were removed. Finally, BLASTp and tBLASTn searches using the complete protein sequences from each selected linear B-cell epitope against NCBI nr/nt databases (e-value ≤ $1e^{-03}$) were performed removing those with no significant hits found in the *Trypanosoma* genus, to eliminate the chance of including uncharacterized sequences from any sort of contamination. Clustering of peptide sequences was performed with local OrthoMCL [60] v2.0.9 with BLAST adjusted parameters for short sequences and OrthoMCL similarity cutoff (percentMatchCutoff) was set to 70%.

3. Results and discussion

3.1. T. vivax transcriptomes and epitope predictions

Parasites in the host's bloodstream trigger immune responses with activation of B-cells by specific regions present in antigens, known as B-cell epitopes. B-cell epitopes are recognized by both their linear amino acid sequences and three-dimensional structures, the former being easier to be predicted by bioinformatics tools as no prior knowledge of protein conformational structure is required. Hundreds or even thousands of linear B-cell epitopes can be predicted from complete genomes, making laborious the screening of antigenic and immunogenic epitopes [61] using recombinant antigens, as compared to methodologies using high-density peptide arrays [62]. Recently, the usage of synthetic peptides has shown high sensitivity and specificity for B-cell epitope diagnostic procedures [63], representing an important advance over recombinant technologies, but at elevated costs, when several peptides are to be tested [38]. In the present study, we have combined several bioinformatics tools in a complete pipeline (Fig. 1A and B), from RNA-seq assembly to epitope prediction, in order to characterize and allow the selection of the best linear B-cell epitope candidates for future experimental tests in the pursuit of efficient and species-specific serological tests for the diagnosis of *T. vivax* infection.

Transcriptomic data from bloodstream forms were analyzed for one Western African strain, TvIL1392: *T. vivax* IL1392 from Nigeria [9,11] (a clone derived from TvY486: *T. vivax* Y486) and two South American isolates: TvLins (*T. vivax* Lins, from São Paulo state, Brazil) and TvLIEM-176 (*T. vivax* LIEM-176, from Venezuela). The annotated genome of the Western African strain (TvY486, from Nigeria) was also included. Our study comprised the most representative dataset of *T. vivax* strains compared to date.

The transcriptome of TvIL1392 assembly yielded 9249 transcripts, which generated 21,697 putative protein sequences (13,829, or 63.7%, are complete Open Reading Frames – ORFs; and 11,021, or 50.8%, ≥100 residues). Similarly, the assembly for TvLins yielded 11,544

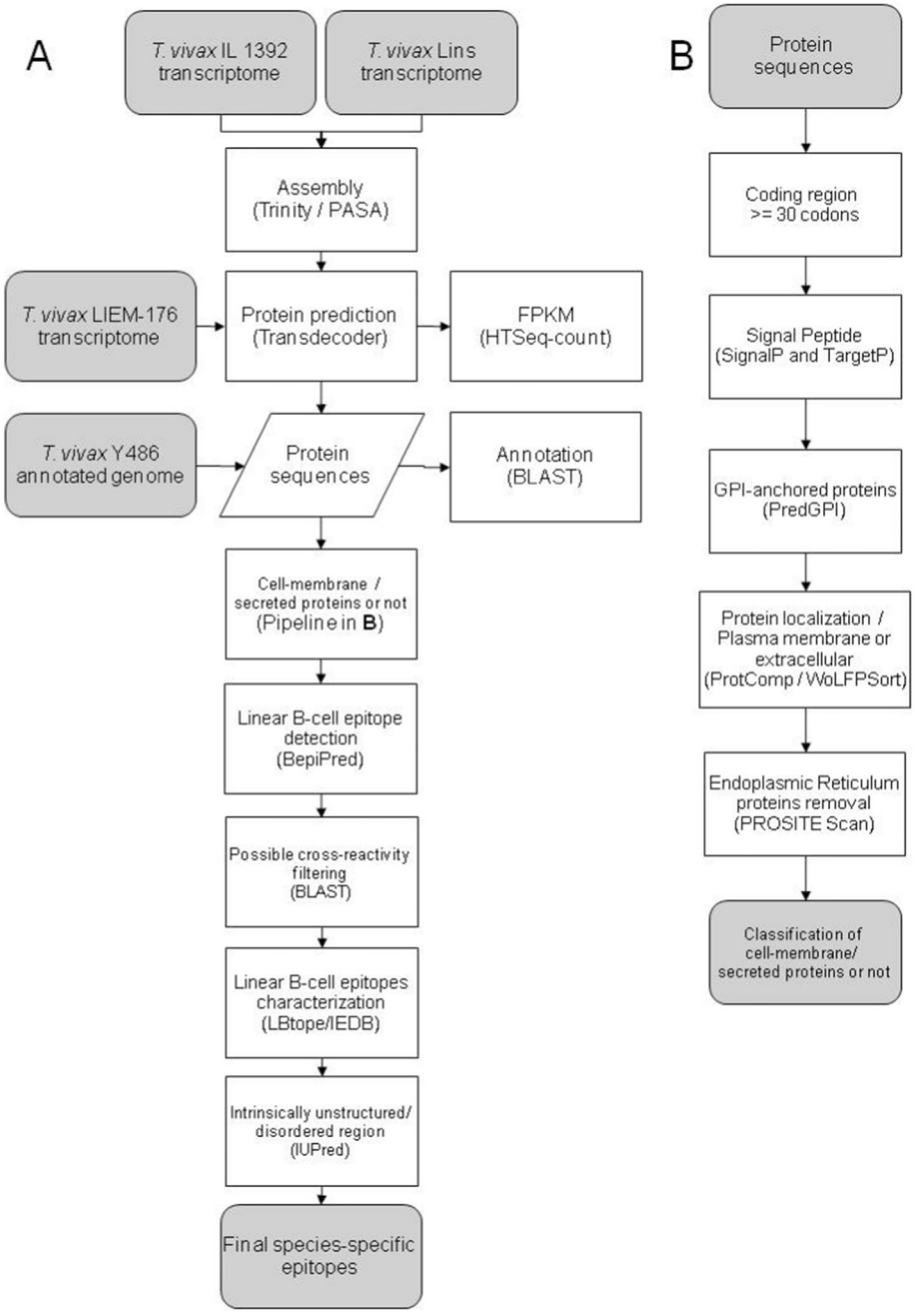

Fig. 1. Bioinformatics workflow. A) Complete pipeline from RNA-seq assembly to protein sequences and linear B-cell epitope predictions. B) Complete pipeline for classification of secreted proteins or potentially localized at cell-surface membrane.

Table 1
Main features from the three most expressed linear B-cell epitopes from the assembled transcriptomes of *Trypanosoma vivax* IL1392 (TvIL1392), *Trypanosoma vivax* Lins (TvLins), and *Trypanosoma vivax* LIEM-176 (TvLIEM-176).

Strain	Epitope ID	Annotation	Sequence	Size	FPKM
TvIL1392	asmbl_2096_m27533_1	Hypothetical protein, conserved in *T. vivax*	NKAQVFPGPSPEVTSAQDAARKA	23	3362.2
TvIL1392	asmbl_7052_m114735_1	Hypothetical protein, conserved in *T. vivax*	RRSHDEEKGKSKNSRKSK	18	2919.0
TvIL1392	asmbl_3484_m64070_1	Conserved hypothetical protein	EAKKLEKEKGDKPSKGEA	18	1933.9
TvLins	c6278_g1_i1_m131269_1	Surface glycoprotein	GSTKCGNETWTTSGPADYDTQAGKC	25	16,914.5
TvLins	c6278_g1_i1_m131269_2	Surface glycoprotein	ETETRRAAKEQTGPEAQKREQKEGNADMTTRAK	33	16,914.5
TvLins	c5909_g1_i1_m130763_1	Surface glycoprotein	GSTKCGNETWTTSGPADYDTQAGKC	25	13,888.1
TvLIEM-176	GCI4HUN02IGEU4_166066_1	Hypothetical protein	ASPQEGKGGRTKDSALG	17	2468.8
TvLIEM-176	GCI4HUN02FNGMF_164965_1	Hypothetical protein, conserved	AMERDAQTTASGDRR	15	1869.9
TvLIEM-176	TvMiraNov_c69_49206_1	Putative nucleoside diphosphate kinase	VLLGATNPADSQPGDDPWA	19	1580.8

Features for all linear B-cell epitopes are displayed in Additional File 2.

Table 2
Protein cellular localization and linear B-cell epitope parameters for the three epitopes with the highest BepiPred mean scores for all *Trypanosoma vivax* strains.

Strain	Epitope ID	Localization[a]	BepiPred mean score	LBtope probability scale %	Mean IUpred score	IEDB methods (%)[b]				
						Chou & Fasman	Emini	Karplus & Schultz	Kolaskar & Tongaonkar	Parker
TvIL1392	asmbl_4468_m84358_1	E/C/E	1.82	77.68	0.81	60.87	66.67	78.26	13.04	82.61
TvIL1392	asmbl_1223_m5649_4	N/N/N	1.76	66.81	0.80	77.08	66.67	87.50	4.17	95.83
TvIL1392	asmbl_2838_m46579_2	N/E/M	1.76	59.15	0.73	94.44	38.89	82.35	50.00	72.22
TvLins	c6312_g1_i4_m131330_1	C/C/E	1.83	64.85	0.65	94.23	84.62	86.54	11.54	86.54
TvLins	asmbl_8120_m118522_1	G/G/E	1.78	100.0	0.75	89.09	60.00	89.09	29.09	87.27
TvLins	c8205_g1_i3_m144643_1	C/C/C	1.78	100.0	0.75	83.64	60.00	78.18	32.73	81.82
TvLIEM-176	F35ERS101APT45_32246_1	E/E/E	1.84	67.81	0.65	82.69	67.31	84.62	17.31	82.69
TvLIEM-176	TvMiraNov_c11077_22717_1	N/N/N	1.81	78.97	0.95	53.85	48.15	52.00	42.31	53.85
TvLIEM-176	F35ERS101AT76C_15521_1	G/G/C	1.78	68.45	0.65	100.00	73.68	100.00	0.00	100.00
TvY486	TvY486_0028930_1	E/E/E	2.13	36.15	0.63	90.71	84.39	86.62	8.18	88.10
TvY486	TvY486_1010870_1	PM/PM/PM	2.12	80.58	0.77	85.00	68.12	77.99	32.50	81.88
TvY486	TvY486_0303314_1	PM/PM/PM	2.04	62.14	0.73	82.14	71.43	100.00	17.86	100.00

[a] Protein cellular localization predicted with: ProtComp Neural Nets/ProtComp Integral prediction/WoLF Psort. C: Cytoplasmic; E: extracellular; G: Golgi; M: mitochondrial; ME: Membrane bound or extracellular; N: nuclear; PM: Plasma Membrane.

[b] Percentage of epitope residues above threshold. Minimum, maximum, and threshold values obtained for each protein sequence are available in Additional File 2. TvIL1392: *T. vivax* IL1392, TvLins: *T. vivax* Lins, TvLIEM-176: *T. vivax* LIEM-176, TvY486: *T. vivax* Y486.

transcripts with 21,370 putative proteins (12,686, or 59.4%, complete ORFs; and 10,355, or 48.5%, ≥100 residues). The assembled transcriptome [25] for TvLIEM-176 had a much higher number of contigs (67,850) from which 42,893 putative proteins were predicted (only 8735, or 20.4%, of complete ORFs; and 9841, or 22.9%, ≥100 residues), indicating a fragmented assembly. This difference in TvLIEM-176 assembly is probably related to low coverage, distinct sequencing, and assembling procedures. For comparison, the TvY486 reference genome currently has 11,885 predicted proteins (9530, or 80.2%, complete ORFs; and 11,605, or 97.6%, ≥100 residues).

Completeness of transcriptome assemblies was assessed by measuring and comparing the number of detected genes from benchmarked universal single-copy orthologs, using the BUSCO [49] software. The assembled transcriptomes, with the exception of TvLIEM-176, presented results comparable or better than the completely annotated genomes. The genome of TvY486 presented 74.4% of the expected eukaryote orthologs while the transcriptomes of TvIL1392, TvLins, and TvLIEM-176 presented 73.9%, 77.2% and 23.8%, respectively (Additional File 1).

A pipeline (Fig. 1B) was applied to protein sequences in order to discriminate between those predicted to be localized at cell-membrane and those secreted to extracellular media or intracellular. Extracellular proteins have an enhanced probability of exposing B-cell epitopes, but it is known that intracellular-derived B-cell epitopes are also recognized, probably due to the parasite destruction with subsequent release of intracellular content [29]. Bloodstream forms of trypanosomes express a repertoire of such exposed proteins such as Variant Surface Glycoprotein. We detected the presence of the secretory pathway targeting signal peptide at N-terminal portion [64], the cell-membrane signaling C-terminal GPI-anchor (Glycosylphosphatidylinositol) [65], and used bioinformatics tools commonly applied to predict protein cellular localization, totaling 4322 (19.9%), 4411 (20.6%), 10,206 (23.8%), and 1278 (10.8%) protein sequences from TvIL1392, TvLins, TvLIEM-176, and TvY486, respectively. All proteins were submitted to linear B-cell epitope predictions, combining three highly sensitive tools. Epitopes were first predicted by BepiPred 1.0 Server, a method based on the hidden Markov model and propensity scale, whose predictions have already been experimentally validated for *T. cruzi* [29,38,63] and L. *braziliensis* [66]. Selected linear B-cell epitopes were further analyzed by LBtope, which is a method based on the Support Vector Machine (SVM). Next, five Immune Epitope Database (IEDB) methods were used for characterization: Chou & Fasman Beta-Turn Prediction, Emini Surface Accessibility Prediction, Karplus & Schulz Flexibility Prediction, Parker Hydrophilicity Prediction, and Kolaskar & Tongaonkar Antigenicity [58]. Finally, to ensure that peptides come from naturally disordered regions in the native protein, with higher accessibility, IUPred was used to screen for amino acid composition that does not allow favorable interactions to form well-defined tertiary structures.

To avoid possible cross-reactivity, we conducted a tBLASTn search against complete genomes of other trypanosomatids. For instance, since wild and domestic ungulates such as sheep, goats, cattle, and horses may be sporadically infected with *T. cruzi* and *Leishmania* [67,68], genomes of these species were also included in this filtering step. This similarity filtering accounted for the elimination of 72.4%, 72.1%, 71.8%, and 73.5% of all predicted epitopes for TvIL1392, TvLins, TvLIEM-176, and TvY486, respectively. Unfortunately, genomes of different genotypes of *T. congolense* and other species of trypanosome infective to ungulates are not currently available. However, the high stringency of our filtering strategy including *T. cruzi* and *Leishmania* that are more distantly related from *T. vivax* than all African trypanosomes will most likely reduced the probability of cross-reactive epitopes.

The final sets of *T. vivax*-specific epitopes are presented in Additional File 2 and examples are shown in Tables 1 and 2. A total of 3936, 4062, 2278, and 4684 linear B-cell epitopes were selected for TvIL1392, TvLins, TvLIEM-176, and TvY486, respectively (Additional File 2). Complete nucleotide and protein sequences are available in Additional File 3–6. As predictions for TvY486 were made from the completely annotated genome, a higher number of epitopes was observed, as opposed to TvLIEM-176, which presented a fragmented transcriptomic assembly.

Most of the selected sequences (81.2%) from the four *T. vivax* strains predicted to be present at cell-surface or secreted and containing putative linear B-cell epitopes were annotated as hypothetical proteins (Additional File 2). For TvLIEM-176, it has been shown that about 56% and 41% of cell-surface expressed proteins were VSGs and hypothetical proteins [25], respectively. As a comparison, for *T. brucei* the VSGs represent about 97% of protein membrane composition. Among the analyzed strains, several other linear B-cell epitopes were found to be present in proteins with annotated functions such as putative mannose-specific lectin, putative major facilitator superfamily (MFS) transporter, putative methyltransferase, glucose transporter, cysteine peptidase C, heat shock protein 70-related protein, and major surface protease gp63 (Additional File 2).

The expression levels were determined through fragments per kilobase of transcript per million mapped reads (FPKM) metric for TvIL1392, TvLins, and TvLIEM-176, revealing the presence of

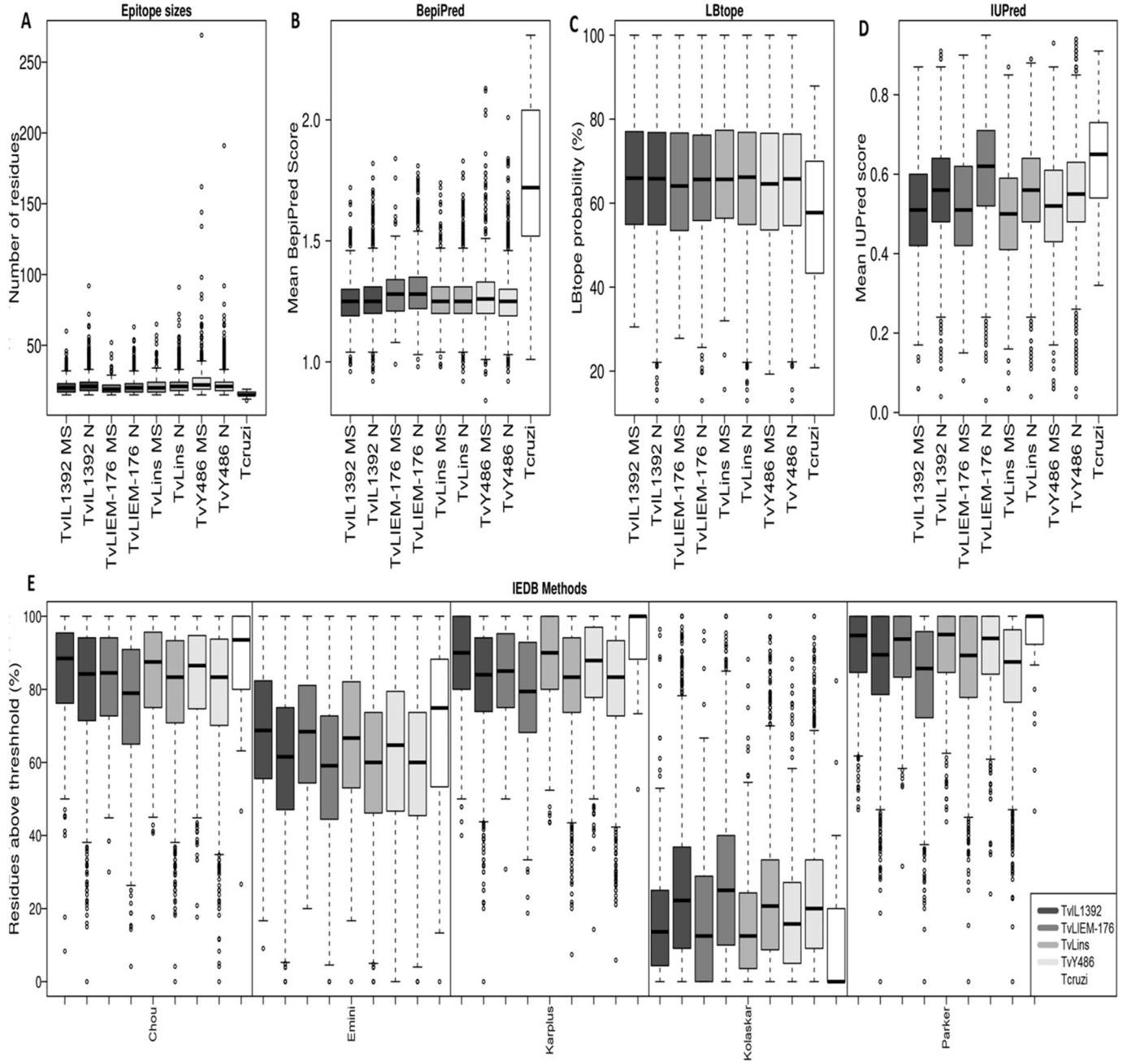

Fig. 2. Boxplots with distinct metrics for all selected epitopes and a control set from *Trypanosoma cruzi*. A) Epitope sizes. B) Mean BepiPred score. C) LBtope probability. D) Mean IUpred score. E) Percentage of epitope residues above thresholds for IEDB methods, Chou: Chou & Fasman Beta-Turn Prediction, Emini: Emini Surface Accessibility Prediction, Karplus: Karplus & Schulz Flexibility Prediction, Kolaskar: Kolaskar & Tongaonkar Antigenicity and Parker: Parker Hydrophilicity Prediction. MS: Complete protein sequences predicted as Membrane-bound or Secreted; N: Complete protein sequences not predicted as Membrane-bound or Secreted. TvIL1392: *T. vivax* IL1392, TvLins: *T. vivax* Lins, TvLIEM-176: *T. vivax* LIEM-176, TvY486: *T. vivax* Y486 and Tcruzi: *Trypanosoma cruzi.*

hypothetical proteins as the most abundant linear B-cell epitopes containing proteins (Table 1 and Additional File 2). The higher observed FPKM values from the selected TvLins surface glycoprotein (16,914.5), TvLIEM-176 (2468.8), and IL1392 (3362.2) hypothetical proteins were originated from 0.19%, 0.05%, and 0.01% of reads mapped to their final transcriptome assemblies, respectively. To put into perspective, the percentages of mapped reads corresponding to the highly expressed VSGs, and alpha and beta tubulins in TvLIEM-176, were estimated as 0.7%, 0.18% and 0.33%, respectively [25]. This leads to the assumption that these hypothetical linear B-cell epitopes containing proteins are less abundant but may play important biological roles.

3.2. Linear B-cell epitope characterization

TvY486 complete protein sequences have been clustered with *T. b. brucei* and *T. congolense* into ortholog families and organized in a comprehensive database of African trypanosome genes predicted to encode cell-surface proteins (Cell-surface Phylome – CSP) [69]. A total of 368 TvY486 epitopes (7.9%) were observed in proteins assigned to a CSP family, from which 279 (75.8%) were from *T. vivax*-specific CSPs (from Fam23 to Fam45, Additional File 2). Although care was taken to remove those with significant similarity to other trypanosome species, some proteins assigned to other CSP families grouping *T. vivax*, *T. brucei,* and *T. congolense* were found. The reason behind this is that even though the whole proteins were classified as orthologs, regions

Table 3
Selected *Trypanosoma vivax* linear B-cell epitope sequences based on observed parameters for *Trypanosoma cruzi* control set.

Organism	Epitope ID	Cluster	Epitope sequence
TvY486	TvY486_0000570_3	TvY486	DQTATSKGTEKAQAQRNSHSPASDNAGNARDPREGNSEDNTEEHKNSAITET
TvY486	TvY486_0006120_2	–	QSQRGNARQQGTTHKPETENTGASTQQHADTDPPSADTQVQDRKARNG
TvY486	TvY486_0008160_1	TvY486 TvIL1392	QRGTARKQGTTHKSETGNTGASTQQHADTDPPSADTQVQDRKARNG
TvY486	TvY486_0019430_2	–	PVFQKAEGGNESDQDAKGNESDKDAKGNESDQDAKGNESDQDAKGNESDQ DAKGNESDQDAKGNESDQDAKGNESDQDAKGNESDQDAKGNESDQDAKGNESDQDAKGNESDQDAKGNESEQDAKGNESDEDAKGNES- DQDTKGNESDQDAKGNESDQDAKGNESDEDAKGNESEQDAKGNE
TvY486	TvY486_0021710_1	–	MNQPRTDPTEPFETPPPLPSAPTSTFSPRDTKSNSSDKCSDG
TvY486	TvY486_0029190_4	–	LRQGGATTSRDSKGKDTATSHQASRNTGNTQSEEHTTGTETTGNQSSQTDSKRECDRTHPNWDE
TvY486	TvY486_0035170_3	TvY486	DQTATSKGTEKAQAQRNSHSPASDNAGNARDPREGNSEDNTEEHKNSAITET
TvY486	TvY486_0302580_1	TvY486 TvIL1392 TvLins	YGTPEVGEIGSTAEQADPSEEVPTNAGNGNGRQE
TvY486	TvY486_0303314_1	TvY486	EVVEENACPDTAGGDEAPAPAPAPTPPPQYR
TvY486	TvY486_0400320_2	TvY486 TvIL1392 TvLins TvLIEM-176	EDDNEKGIASTNREASPGEHQ
TvY486	TvY486_0500070_1	–	RRREPLRPANEAAPDAENPKENREPQEDEEKGKETEGPEGIRSL
TvY486	TvY486_0501850_1	TvY486 TvIL1392 TvLins	SLAPVSGNNKPRDANAGEPSTPADDHTPRTLLDS
TvY486	TvY486_0905360_1	–	MDTHESAEGPQAHRPKNHSKPSPAPNGCKTKMPKID
TvY486	TvY486_1002180_1	TvY486 TvIL1392	HGSVEDCDMGGDTPNDGSEDTGDGYAEDAEVF
TvY486	TvY486_1006800_2	TvY486 TvIL1392 TvLins	DGLEASDKEQVESEDNTDETDDKSNETKEKQ
TvY486	TvY486_1014540_3	–	KTNIIGEGPEPYPEEVDRNDDADGDGGVQQC
TvY486	TvY486_1100860_1	TvY486 TvIL1392 TvLins	QRVEDNQTSKPYPGHRSPSDSPGCKKTEEDDPCGDNFF
TvY486	TvY486_1102420_4	TvY486 TvIL1392 TvLins TvLIEM-176	YFSNRNNKSNERPPPQYDTGRTEPY
TvY486	TvY486_1108060_1	TvY486 TvIL1392	KSEEAGGGLEDTTSNGEPAMKKSRPDGCDDGTEKA
TvY486	TvY486_1116470_1	TvY486 TvIL1392 TvLins	KVSDKRESSEEDDGASTNDTESDRT
TvIL1392	asmbl_2491_m38052_1	TvIL1392 TvLins TvLIEM-176	MPPDKSETDTYGNGGERVK
TvIL1392	asmbl_2562_m39418_1	TvY486 TvIL1392	HGSVEDCDMGGDTPNDGSEDTGDGYAEDAEVF
TvIL1392	asmbl_2811_m45856_2	TvY486 TvIL1392 TvLins	DGLEASDKEQVESEDNTDETDDKSNETKEKQA
TvIL1392	asmbl_3284_m59070_1	TvY486 TvIL1392 TvLins	QRVEDNQTSKPYPGHRSPSDSPGCKKTEEDDPCGDNFF
TvIL1392	asmbl_3375_m61251_4	TvY486 TvIL1392 TvLins TvLIEM-176	YFSNRNNKSNERPPPQYDTGRTEPYA
TvIL1392	asmbl_4098_m78740_1	TvY486 TvIL1392 TvLins	KVSDKRESSEEDDGASTNDTESDRTA
TvIL1392	asmbl_439_m83426_1	TvY486 TvIL1392 TvLins	YGTPEVGEIGSTAEQADPSEEVPTNAGNGNGRQEA
TvIL1392	asmbl_4844_m88675_1	TvY486 TvIL1392	QRGTARKQGTTHKSETGNTGASTQQHADTDPPSADTQVQDRKARNG
TvIL1392	asmbl_6390_m106500_1	TvIL1392	CPNAEGAQNQGKKAEGAQNQGKKAEGAQNQGKKAEGAQNQGKKSEG
TvIL1392	asmbl_869_m123222_1	TvY486 TvIL1392 TvLins	SLAPVSGNNKPRDANAGEPSTPADDHTPRTLLDS
TvIL1392	c419_g1_i1_m127007_1	TvY486 TvIL1392	KSEEAGGGLEDTTSNGEPAMKKSRPDGCDDGTEKAA
TvLins	asmbl_1599_m12035_1	TvIL1392 TvLins TvLIEM-176	MPPDKSETDTYGNGGERVK
TvLins	asmbl_1990_m20143_1	TvY486 TvIL1392 TvLins	QRVEDNQTSKPYPGHRSPSDSPGCKKTEEDDPCGDNFF
TvLins	asmbl_2018_m20766_1	TvY486 TvIL1392 TvLins	SLAPVSGNNKPRDANAGEPSTPADDHTPRTLLDS
TvLins	asmbl_2352_m27174_2	TvY486 TvIL1392 TvLins	DGLEASDKEQVESEDNTDETDDKSNETKEKQA
TvLins	asmbl_4470_m66624_9	–	SLREDQQDDFDEDSNERDCSQSDSGEYVEESEDDYDSSGSSECPDDESDASYGRA
TvLins	asmbl_4851_m73276_1	–	KPENKLDEVSSDGGETDQTPKG
TvLins	asmbl_5366_m81191_1	TvY486 TvIL1392 TvLins	KVSDKRESSEEDDGASTNDTESDRTA
TvLins	asmbl_5659_m85745_1	TvY486 TvIL1392 TvLins	YGTPEVGEIGSTAEQADPSEEVPTNAGNGNGRQEA
TvLins	asmbl_7330_m109248_4	TvY486 TvIL1392 TvLins TvLIEM-176	YFSNRNNKSNERPPPQYDTGRTEPYA
TvLins	asmbl_7999_m117163_1	TvLins TvLIEM-176	TSTLSYKGANNESPNSKQAPNA
TvLIEM-176	F35ERS101ALXLD_60137_1	–	APSVIGTTTKGTPPQGEDANNNDEQKSA
TvLIEM-176	F35ERS101AT76C_15521_1	–	AQPRPPDPRNDVEKSSKSLDA
TvLIEM-176	F491XYR02GF556_12112_1	–	DDGGDSDNDKNAKGSSQIGNK
TvLIEM-176	TvMiraNov_c1340_176622_1	–	SNRRHAQRTKPGDSESK
TvLIEM-176	TvMiraNov_c15191_69060_1	TvY486 TvIL1392 TvLins TvLIEM-176	MPDSRTLSNSGNNNEVGDDEGDDC
TvLIEM-176	TvMiraNov_c3982_30916_1	TvY486 TvLIEM-176	SCSDMTPGANSSNDNDGEDVEDDDSDYYEKEVDDSDAVDT

Cutoff parameters were: BepiPred Mean ≥ 1.5; LBTope score ≥ 43; IUPred mean ≥ 54; Chou & Fasman Beta-Turn Prediction ≥ 80; Emini Surface Accessibility Prediction ≥ 55; Karplus & Schulz Flexibility Prediction ≥ 88; Kolaskar & Tongaonkar Antigenicity ≤ 19 and Parker Hydrophilicity Prediction ≥ 92. TvIL1392: *T. vivax* IL1392, TvLins: *T. vivax* Lins, TvLIEM-176: *T. vivax* LIEM-176, TvY486: *T. vivax* Y486.

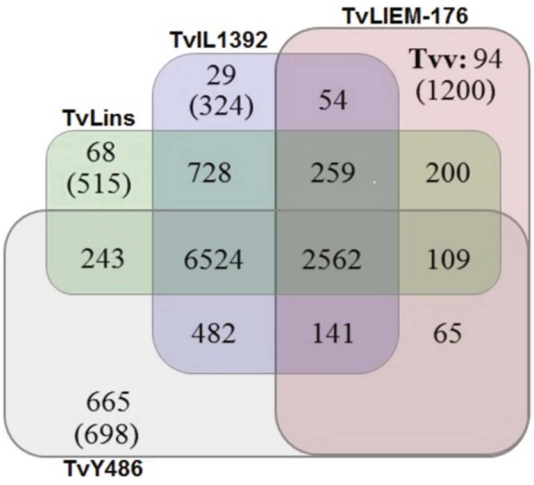

Fig. 3. Epitope clustering. Veen diagram with clustered epitopes with 70% BLAST identity for the four *T. vivax* strains. Numbers inside parentheses represent unclustered sequences. TvIL1392: *T. vivax* IL1392, TvLins: *T. vivax* Lins, TvLIEM-176: *T. vivax* LIEM-176, TvY486: *T. vivax* Y486.

corresponding to the linear B-cell epitope residues were highly variable (Additional File 7).

All the linear B-cell epitope parameters were evaluated and compared to a control set of 34 linear B-cell epitopes retrieved from experimental validation from *T. cruzi* with mice sera [29]. The article from Oliveira Mendes et al. originally describes the validation of 36 peptides, but we removed a duplicated and inconsistent epitope sequence (TcCLB.506401.320; epitope: PIELVLWMPTLCRAN) with negative mean BepiPred prediction score. Epitope sizes ranged from only 15 to 269 residues (both minimum and maximum were observed in TvY486, Fig. 2A and Additional File 2) with mean values of approximately 20 residues. The mean BepiPred scores were maintained near the selected threshold of 1.3 (Fig. 2B and Additional File 2), a value with specificity of 0.96 according to the software developers [56], although a significantly higher mean of 1.7 was observed for the control set. The LBtope mean probabilities were above 65% (Fig. 2C and Additional File 2), while the control set mean was close to the recommended threshold of 60% [57]. As indicated by the mean IUPred scores above the recommended threshold of ≥ 0.5, the linear B-cell epitopes seemed to be located at intrinsically disordered regions (Fig. 2D and Additional File 2), increasing the chances of exposition. Curiously, proteins not predicted as membrane-bound or secreted presented more disordered linear B-cell epitopes and the control set showed the higher mean (0.64).

For IEDB methods, the higher mean scores (Chou & Fasman, Emini, Karplus & Schulz and Parker) and the lower mean score (Kolaskar & Tongaonkar) were all observed for the *T. cruzi* control epitopes, with similar values among the *T. vivax* strains (Fig. 2E and Additional File 2). Similarly to the observed for the disordered values, differences are evident between intracellular and exposed proteins. Based on the control set mean values, a subset was filtered for each strain as the most promising linear B-cell epitopes for future experiments (Table 3). It is worth mentioning that although being valuable tools, these methods were mainly trained with complete absence of *Trypanosoma* sequences and more specific software is still required to increase prediction accuracy.

3.3. Clustering linear B-cell epitopes

Strain-specific linear B-cell epitopes may be useful for genotyping while *T. vivax* conserved epitopes (or a combination of genotype-

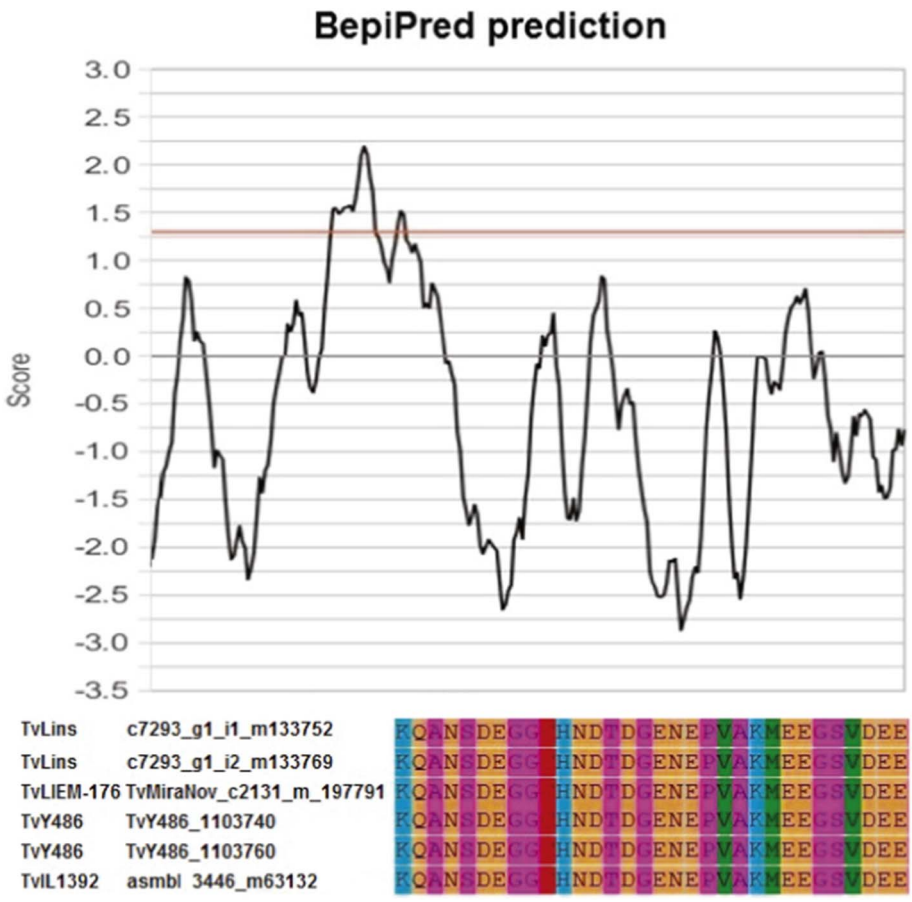

Fig. 4. Graphical view of BepiPred prediction for a portion of TvY486_1103740 protein sequence containing a selected epitope. A red line represents the mean score of 1.3. Protein sequence alignment is representing the linear B-cell epitope above the red line and shared between the *Trypanosoma vivax* strains. TvIL1392: *T. vivax* IL1392, TvLins: *T. vivax* Lins, TvLIEM-176: *T. vivax* LIEM-176, TvY486: *T. vivax* Y486. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

specific sequences) can improve the efficiency of species-specific serodiagnosis. To detect exclusive and common linear B-cell epitopes from the four strains, we clustered their sequences considering 70% of similarity (Fig. 3 and Additional File 8). A multiple sequence alignment including a common linear B-cell epitope is represented in Fig. 4 (from K179 to E209 on TvY486_1103760). The peptide was observed only in *T. vivax* strains. A graphical view of BepiPred output is also shown, evidencing the amino acid scores above threshold. Additionally, the two Western African strains (TvIL1392 and TvY486) and the two South American strains (TvLins and TvLIEM-176) shared 482 and 200 exclusive linear B-cell epitopes, respectively (Fig. 3). Common and specific epitopes were detected for all strains, but the fragmentary nature of the TvLIEM-176 transcriptome assembly influenced the clustering and the excessive number of observed TvLIEM-176 specific sequences.

Recently, quantitative proteomics was employed to identify candidate diagnostic antigens in whole *T. vivax* lysate that bound selectively to the immobilized IgG from *T. vivax* infected calves [36]. *T*wo proteins selected in Fleming's study (TvY486_0045500 and TvY486_0019690) tested by ELISA exhibited high sensitivity and specificity for *T. vivax* when compared with *T. congolense.* These proteins presented linear B-cell epitopes identified in our analyses. In the TvY486_0045500 protein, the peptide "KEKKRQAIGGDSEGPKSSDAKSTDATPTSSASQKV" from K296 to V330 was selected, although it was eliminated in the BLAST filtering steps due to shared similarities with translated genomic regions from *T. brucei.* For the TvY486_0019690 protein, three peptides were detected: "AEEESKEWEKEQQDAESDL" from A185 to L203; "EIKEKKRQANNGDSEGPKSSDSKSADATPTSSASQKV" from E294 to V330; and "QTADKPSSANNSKLSP" from Q349 to P364. The peptide from E294 to V330 shares ≥ 89% identity with TvY486_0045500 selected residues, and was equally discarded in our analysis, while the other two were retained and analyzed (Additional File 2). The presence of peptides of experimentally tested proteins among the selected linear B-cell epitopes reinforces the potential of this *in silico* workflow.

3.4. Experimental validation

In order to provide a preliminary experimental evidence of the pipeline potential, we have randomly selected, among the complete list (Additional File 2), a peptide sequence containing a putative B-cell epitope (Additional File 9). The sequence was cloned and used to perform serological validation using sera from 11 mice infected or not with *T. vivax*. Mice sera obtained after *T. vivax* infection significantly recognized the peptide at 15 and 30 days post-infection. The results show highly significant reactions, as compared to sera from non-infected mice. Further experiments to test this, and preferentially a higher amount of B-cell epitopes, with sera from wild animals infected or not with *T. vivax* and other parasites are still required.

4. Conclusions

Here we present the first *in silico* linear B-cell epitope prediction and characterization for South American and African *T. vivax* strains, based on a complete bioinformatics pipeline designed to select the best putative species and strain-specific candidates for *T. vivax* diagnosis. Parameter comparisons with positively tested *T. cruzi* peptides allowed for the selection of a subset of *T. vivax* linear B-cell epitopes with high diagnostic potential. The applied control set was based on a limited amount of validated sequences due to the lack of equivalent data in the literature, therefore, biases could be incorporated. Stringent criteria were deliberately applied to provide a filtered list for more accessible short scale validations, representing time and cost savings, even at the cost of loosing true positives. The findings provided in this *in silico* analyses require further and broader experimental validation in order to specify its efficacy. Only preliminary experimental results could be evidenced for a single peptide encompassing a selected B-cell epitope in the current work, however, the usage of high scale experiments, such as high-density peptide arrays, would allow the validation of all the thousands of putative linear B-cell epitopes provided in this study in a single experiment. Proving the efficacy of the selected epitopes will certify the pipeline and serve as a reference for additional searchers. Finally, new transcriptomic data for the highly pathogenic *T. vivax* Lins strain were made available and will be useful for further comparative analyses.

Acknowledgments

This work was supported by the LNCC/USP/Pasteur Project Tvivax DIAG-DI-2015-17 and the Brazilian Agency CNPq (PROFRICA, process 490735/2010-0). The authors are indebted to the MAEDI, Chaire USP/Pasteur, RLMG and CMR received three months Calmette & Yersin fellowships from the Institut Pasteur International Division. HAG and CMR are postdoctoral fellows of FAPESP and CNPq, respectively. We are also grateful for the collaboration of Andrew Jackson (Liverpool University, UK).

Author contributions

Conceived and designed the experiments: RLMG, MMGT, PMCM, and ATRV. Performed experimental infections and prepared *T. vivax* Lins samples: CMR, RZM, and FAC. Performed experimental validation: NC and AC. Performed sequencing experiments: ALG.; Analyzed the data and conceived figures and tables: RLMG. Contributed with reagents/materials: MMGT, PMCM, RZM, FAC, and ATRV. Wrote the paper: RLMG. Contributed to the writing of the manuscript: CMR, HAG, MMGT, PMCM, and ATRV. All authors revised and approved the final version of the manuscript.

Competing interests

The authors have declared that no competing interests exist.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2018.02.017.

References

[1] P.R. Gardiner, A.J. Wilson, Trypanosoma (Duttonefla) vivax, Parasitol. Today 3 (2) (1987) 49–52, http://dx.doi.org/10.1016/0169-4758(87)90213-4.
[2] L.J. Morrison, L. Vezza, T. Rowan, J.C. Hope, Animal African trypanosomiasis: time to increase focus on clinically relevant parasite and host species, Trends Parasitol. 32 (8) (2016) 599–607, http://dx.doi.org/10.1016/j.pt.2016.04.012.
[3] H.A. Garcia, O.J. Ramírez, C.M.F. Rodrigues, et al., *Trypanosoma vivax* in water buffalo of the Venezuelan llanos: an unusual outbreak of wasting disease in an endemic area of typically asymptomatic infections, Vet. Parasitol. 230 (2016) 49–55, http://dx.doi.org/10.1016/j.vetpar.2016.10.013.
[4] H.A. Garcia, C.M.F. Rodrigues, A.C. Rodrigues, et al., Remarkable richness of trypanosomes in tsetse flies (Glossina morsitans morsitans and Glossina pallidipes) from the Gorongosa National Park and Niassa National Reserve of Mozambique revealed by fluorescent fragment length barcoding (FFLB), *Infect. Genet. Evol.* (July 2017), http://dx.doi.org/10.1016/j.meegid.2017.07.005.
[5] H. Nimpaye, F. Njiokou, T. Njine, et al., *Trypanosoma vivax*, T. congolense "forest type" and T. simiae : prevalence in domestic animals of sleeping sickness foci of Cameroon, Parasite 18 (2) (2011) 171–179, http://dx.doi.org/10.1051/parasite/2011182171.
[6] H. Birhanu, R. Fikru, M. Said, et al., Epidemiology of Trypanosoma evansi and *Trypanosoma vivax* in domestic animals from selected districts of Tigray and Afar regions, Northern Ethiopia, Parasit. Vectors 8 (2015) 212, , http://dx.doi.org/10.1186/s13071-015-0818-1.
[7] C.M. Rodrigues, J.S. Batista, J.M. Lima, et al., Field and experimental symptomless infections support wandering donkeys as healthy carriers of *Trypanosoma vivax* in the Brazilian Semiarid, a region of outbreaks of high mortality in cattle and sheep, Parasit. Vectors 8 (1) (2015) 564, http://dx.doi.org/10.1186/s13071-015-1169-7.
[8] E. Mossaad, B. Salim, K. Suganuma, et al., *Trypanosoma vivax* is the second leading cause of camel trypanosomosis in Sudan after Trypanosoma evansi, Parasit. Vectors 10 (1) (2017) 176, http://dx.doi.org/10.1186/s13071-017-2117-5.
[9] N. Chamond, A. Cosson, M.C. Blom-Potar, et al., *Trypanosoma vivax* infections: pushing ahead with mouse models for the study of Nagana. I. Parasitological,

hematological and pathological parameters, Raper J, ed, PLoS Negl. Trop. Dis. 4 (8) (2010) e792, http://dx.doi.org/10.1371/journal.pntd.0000792.
[10] M.C. Blom-Potar, N. Chamond, A. Cosson, et al., *Trypanosoma vivax* infections: pushing ahead with mouse models for the study of Nagana. II. Immunobiological dysfunctions, Raper J, ed, PLoS Negl. Trop. Dis/ 4 (8) (2010) e793, http://dx.doi.org/10.1371/journal.pntd.0000793.
[11] A.P. Jackson, S. Goyard, D. Xia, et al., Global gene expression profiling through the complete life cycle of *Trypanosoma vivax*, Büscher P, ed, PLoS Negl. Trop. Dis. 9 (8) (2015) e0003975, http://dx.doi.org/10.1371/journal.pntd.0003975.
[12] A.P. Cortez, R.M. Ventura, A.C. Rodrigues, et al., The taxonomic and phylogenetic relationships of *Trypanosoma vivax* from South America and Africa, Parasitology 133 (Pt 2) (2006) 159–169, http://dx.doi.org/10.1017/S0031182006000254.
[13] A.C. Rodrigues, L. Neves, H.A. Garcia, et al., Phylogenetic analysis of *Trypanosoma vivax* supports the separation of South American/West African from East African isolates and a new T. vivax-like genotype infecting a nyala antelope from Mozambique, Parasitology 135 (11) (2008), http://dx.doi.org/10.1017/S0031182008004848.
[14] H.A. Garcia, A.C. Rodrigues, C.M. Rodrigues, et al., Microsatellite analysis supports clonal propagation and reduced divergence of *Trypanosoma vivax* from asymptomatic to fatally infected livestock in South America compared to West Africa, Parasit. Vectors 7 (1) (2014) 210, http://dx.doi.org/10.1186/1756-3305-7-210.
[15] C.M. Rodrigues, H.A. Garcia, A.C. Rodrigues, et al., New insights from Gorongosa National Park and Niassa National Reserve of Mozambique increasing the genetic diversity of *Trypanosoma vivax* and *Trypanosoma vivax*-like in tsetse flies, wild ungulates and livestock from East Africa, Parasit. Vectors 10 (1) (2017) 337, http://dx.doi.org/10.1186/s13071-017-2241-2.
[16] J.S. Batista, F. Riet-Correa, M.M.G. Teixeira, C.R. Madruga, S.D.V. Simões, T.F. Maia, Trypanosomiasis by *Trypanosoma vivax* in cattle in the Brazilian semiarid: description of an outbreak and lesions in the nervous system, Vet. Parasitol. 143 (2) (2007) 174–181, http://dx.doi.org/10.1016/j.vetpar.2006.08.017.
[17] Cadioli FA, de Barnabé PA, Machado RZ, et al. First report of *Trypanosoma vivax* outbreak in dairy cattle in São Paulo state, Brazil. *Rev. Bras. Parasitol. Vet.*. 21(2):118–124. http://www.ncbi.nlm.nih.gov/pubmed/22832751.
[18] T.S.A. Bastos, A.M. Faria, C. Madrid DM de, et al., First outbreak and subsequent cases of *Trypanosoma vivax* in the state of Goiás, Brazil, Rev. Bras. Parasitol. Vet. (2017), http://dx.doi.org/10.1590/s1984-29612017019.
[19] O.L. Fidelis Junior, P.H. Sampaio, R.Z. Machado, M.R. André, L.C. Marques, F.A. Cadioli, Evaluation of clinical signs, parasitemia, hematologic and biochemical changes in cattle experimentally infected with *Trypanosoma vivax*, Rev. Bras. Parasitol. Vet. 25 (1) (2016) 69–81, http://dx.doi.org/10.1590/S1984-29612016013.
[20] J.S. Batista, C.M. Rodrigues, H.A. García, et al., Association of *Trypanosoma vivax* in extracellular sites with central nervous system lesions and changes in cerebrospinal fluid in experimentally infected goats, Vet. Res. 42 (1) (2011) 63, http://dx.doi.org/10.1186/1297-9716-42-63.
[21] G.J.N. Galiza, H.A. Garcia, A.C.O. Assis, et al., High mortality and lesions of the central nervous system in Trypanosomosis by *Trypanosoma vivax* in Brazilian hair sheep, Vet. Parasitol. 182 (2–4) (2011) 359–363, http://dx.doi.org/10.1016/j.vetpar.2011.05.016.
[22] S. D'Archivio, A. Cosson, M. Medina, T. Lang, P. Minoprio, S. Goyard, Non-invasive in vivo study of the *Trypanosoma vivax* infectious process consolidates the brain commitment in late infections, Pimenta PF, ed, PLoS Negl. Trop. Dis. 7 (1) (2013) e1976, , http://dx.doi.org/10.1371/journal.pntd.0001976.
[23] M.J. Turner, M.L. Cardoso de Almeida, A.M. Gurnett, J. Raper, J. Ward, Biosynthesis, attachment and release of variant surface glycoproteins of the African trypanosome, Curr. Top. Microbiol. Immunol. 117 (1985) 23–55 http://www.ncbi.nlm.nih.gov/pubmed/3896675.
[24] A.P. Jackson, A. Berry, M. Aslett, et al., Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species, Proc. Natl. Acad. Sci. 109 (9) (2012) 3416–3421, http://dx.doi.org/10.1073/pnas.1117313109.
[25] G. Greif, M. Ponce de Leon, G. Lamolle, et al., Transcriptome analysis of the bloodstream stage from the parasite *Trypanosoma vivax*, BMC Genomics 14 (2013) 149, http://dx.doi.org/10.1186/1471-2164-14-149.
[26] A.P. Cortez, A.C. Rodrigues, H.A. Garcia, et al., Cathepsin L-like genes of *Trypanosoma vivax* from Africa and South America – characterization, relationships and diagnostic implications, Mol. Cell. Probes 23 (1) (2009) 44–51, http://dx.doi.org/10.1016/j.mcp.2008.11.003.
[27] F.A. Cadioli, O.L. Fidelis Junior, P.H. Sampaio, et al., Detection of *Trypanosoma vivax* using PCR and LAMP during aparasitemic periods, Vet. Parasitol. 214 (1–2) (2015) 174–177, http://dx.doi.org/10.1016/j.vetpar.2015.09.001.
[28] Z.K. Njiru, A.S.J. Mikosza, E. Matovu, et al., African trypanosomiasis: sensitive and rapid detection of the sub-genus Trypanozoon by loop-mediated isothermal amplification (LAMP) of parasite DNA, Int. J. Parasitol. 38 (5) (2008) 589–599, http://dx.doi.org/10.1016/j.ijpara.2007.09.006.
[29] T.A. de Oliveira Mendes, J.L. Reis Cunha, R. de Almeida Lourdes, et al., Identification of strain-specific B-cell epitopes in *Trypanosoma cruzi* using genome-scale epitope prediction and high-throughput immunoscreening with peptide arrays, PLoS Negl. Trop. Dis. (2013), http://dx.doi.org/10.1371/journal.pntd.0002524.
[30] M.C. Eisler, P. Lessard, R.A. Masake, S.K. Moloo, A.S. Peregrine, Sensitivity and specificity of antigen-capture ELISAs for diagnosis of Trypanosoma congolense and *Trypanosoma vivax* infections in cattle, Vet. Parasitol. 79 (3) (1998) 187–201 http://www.ncbi.nlm.nih.gov/pubmed/9823059.
[31] D. Pillay, J. Izotte, R. Fikru, et al., *Trypanosoma vivax* GM6 antigen: a candidate antigen for diagnosis of African animal Trypanosomosis in cattle, Rodrigues MM, ed, PLoS One 8 (10) (2013) e78565, , http://dx.doi.org/10.1371/journal.pone.0078565.
[32] T.-T. NGUYEN, M.S. MOTSIRI, M.O. TAIOE, et al., Application of crude and recombinant ELISAs and immunochromatographic test for serodiagnosis of animal trypanosomosis in the Umkhanyakude district of KwaZulu-Natal province, South Africa, J. Vet. Med. Sci. 77 (2) (2015) 217–220, http://dx.doi.org/10.1292/jvms.14-0330.
[33] T.T. Nguyen, N. Ruttayaporn, Y. Goto, Kawazu S. ichiro, T. Sakurai, N. Inoue, A TeGM6-4r antigen-based immunochromatographic test (ICT) for animal trypanosomosis, Parasitol. Res. (2015), http://dx.doi.org/10.1007/s00436-015-4672-z.
[34] G.L. Uzcanga, Y. Pérez-Rojas, R. Camargo, et al., Serodiagnosis of bovine trypanosomosis caused by non-tsetse transmitted Trypanosoma (Duttonella) vivax parasites using the soluble form of a Trypanozoon variant surface glycoprotein antigen, Vet. Parasitol. 218 (2016) 31–42, http://dx.doi.org/10.1016/j.vetpar.2016.01.007.
[35] J.W. Magona, J.S.P. Mayende, J. Walubengo, Comparative evaluation of the antibody-detection ELISA technique using microplates precoated with denatured crude antigens from Trypanosoma congolense or *Trypanosoma vivax*, Trop. Anim. Health Prod. 34 (4) (2002) 295–308 http://www.ncbi.nlm.nih.gov/pubmed/12166331.
[36] J.R. Fleming, L. Sastry, S.J. Wall, L. Sullivan, M.A.J. Ferguson, Proteomic identification of immunodiagnostic antigens for *Trypanosoma vivax* infections in cattle and generation of a proof-of-concept lateral flow test diagnostic device, Raper J, ed, PLoS Negl. Trop. Dis. 10 (9) (2016) e0004977, , http://dx.doi.org/10.1371/journal.pntd.0004977.
[37] J.L. Reis-Cunha, T.A.D.O. Mendes, R. De Almeida Lourdes, et al., Genome-wide screening and identification of new *Trypanosoma cruzi* antigens with potential application for chronic chagas disease diagnosis, PLoS One (2014), http://dx.doi.org/10.1371/journal.pone.0106304.
[38] S.J. Carmona, M. Nielsen, C. Schafer-Nielsen, et al., Towards high-throughput immunomics for infectious diseases: use of next-generation peptide microarrays for rapid discovery and mapping of antigenic determinants, Mol. Cell. Proteomics 14 (7) (2015) 1871–1884, http://dx.doi.org/10.1074/mcp.M114.045906.
[39] R. Frank, The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports—principles and applications, J. Immunol. Methods 267 (1) (2002) 13–26 http://www.ncbi.nlm.nih.gov/pubmed/12135797.
[40] L.E. González, J.A. García, C. Núñez, et al., *Trypanosoma vivax*: a novel method for purification from experimentally infected sheep blood, Exp. Parasitol. 111 (2) (2005) 126–129, http://dx.doi.org/10.1016/j.exppara.2005.05.008.
[41] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (15) (2014) 2114–2120, http://dx.doi.org/10.1093/bioinformatics/btu170.
[42] M. Aslett, C. Aurrecoechea, M. Berriman, et al., TriTrypDB: a functional genomic resource for the Trypanosomatidae, *Nucleic Acids Res.* 38 (Database) (2010) D457–D462, http://dx.doi.org/10.1093/nar/gkp851.
[43] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, Genome Biol. 14 (4) (2013) R36, http://dx.doi.org/10.1186/gb-2013-14-4-r36.
[44] H. Li, B. Handsaker, A. Wysoker, et al., The sequence alignment/map format and SAMtools, Bioinformatics 25 (16) (2009) 2078–2079, http://dx.doi.org/10.1093/bioinformatics/btp352.
[45] B.J. Haas, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies, Nucleic Acids Res. 31 (19) (2003) 5654–5666, http://dx.doi.org/10.1093/nar/gkg770.
[46] M.G. Grabherr, B.J. Haas, M. Yassour, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (7) (2011) 644–652, http://dx.doi.org/10.1038/nbt.1883.
[47] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (3) (1990) 403–410, http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
[48] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (4) (2012) 357–359, http://dx.doi.org/10.1038/nmeth.1923.
[49] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics 31 (19) (2015) 3210–3212, http://dx.doi.org/10.1093/bioinformatics/btv351.
[50] C. Staats, Â. Junges, R.L. Guedes, et al., Comparative genome analysis of entomopathogenic fungi reveals a complex set of secreted proteins, BMC Genomics 15 (1) (2014) 822, http://dx.doi.org/10.1186/1471-2164-15-822.
[51] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat. Methods 8 (10) (2011) 785–786, http://dx.doi.org/10.1038/nmeth.1701.
[52] O. Emanuelsson, H. Nielsen, S. Brunak, G. von Heijne, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, J. Mol. Biol. 300 (4) (2000) 1005–1016, http://dx.doi.org/10.1006/jmbi.2000.3903.
[53] A. Pierleoni, P. Martelli, R. Casadio, PredGPI: a GPI-anchor predictor, BMC Bioinforma. 9 (1) (2008) 392, http://dx.doi.org/10.1186/1471-2105-9-392.
[54] P. Horton, K.-J. Park, T. Obayashi, et al., WoLF PSORT: protein localization predictor, Nucleic Acids Res. 35 (Web Server) (2007) W585–W587, http://dx.doi.org/10.1093/nar/gkm259.
[55] A. Gattiker, E. Gasteiger, A. Bairoch, ScanProsite: a reference implementation of a PROSITE scanning tool, Appl. Bioinforma. 1 (2) (2002) 107–108 http://www.ncbi.nlm.nih.gov/pubmed/15130850.
[56] J.E.P. Larsen, O. Lund, M. Nielsen, Improved method for predicting linear B-cell epitopes, Immunome Res. 2 (2006) 2, http://dx.doi.org/10.1186/1745-7580-2-2.
[57] H. Singh, H.R. Ansari, G.P.S. Raghava, Improved method for linear B-cell epitope prediction using Antigen's primary sequence, Schönbach C, ed, PLoS One 8 (5)

(2013) e62216, , http://dx.doi.org/10.1371/journal.pone.0062216.
[58] Q. Zhang, P. Wang, Y. Kim, et al., Immune epitope database analysis resource (IEDB-AR), Nucleic Acids Res. 36 (Web Server) (2008) W513–W518, http://dx.doi.org/10.1093/nar/gkn254.
[59] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics 21 (16) (2005) 3433–3434, http://dx.doi.org/10.1093/bioinformatics/bti541.
[60] L. Li, C. Stoeckert, D. Roos, OrthoMCL: identification of ortholog groups for eukaryotic genomes, Genome Res. (2003) 2178–2189, http://dx.doi.org/10.1101/gr.1224503.candidates.
[61] M.H.V. Regenmortel, What is a B-cell epitope? in: M. Schutkowski, U. Reineke (Eds.), Epitope Mapping Protocols. Methods in Molecular Biology™ (Methods and Protocols), vol 524, Humana Press, 2009, pp. 3–20, , http://dx.doi.org/10.1007/978-1-59745-450-6_1.
[62] S. Buus, J. Rockberg, B. Forsström, P. Nilsson, M. Uhlen, C. Schafer-Nielsen, High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays, Mol. Cell. Proteomics 11 (12) (2012) 1790–1800, http://dx.doi.org/10.1074/mcp.M112.020800.
[63] S.J. Carmona, P.A. Sartor, M.S. Leguizamón, O.E. Campetella, F. Agüero, Diagnostic peptide discovery: prioritization of pathogen diagnostic markers using multiple features, Rodrigues MM, ed, PLoS One 7 (12) (2012) e50748, , http://dx.doi.org/10.1371/journal.pone.0050748.
[64] T.A. Rapoport, B. Jungnickel, U. Kutay, Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes, Annu. Rev. Biochem. 65 (1) (1996) 271–303, http://dx.doi.org/10.1146/annurev.bi.65.070196.001415.
[65] J.S. Silverman, J.D. Bangs, Form and function in the trypanosomal secretory pathway, Curr. Opin. Microbiol. 15 (4) (2012) 463–468, http://dx.doi.org/10.1016/j.mib.2012.03.002.
[66] D. Menezes-Souza, O. Mendes TA de, S. Gomes M de, D.C. Bartholomeu, R.T. Fujiwara, Improving serodiagnosis of human and canine leishmaniasis with recombinant Leishmania braziliensis cathepsin L-like protein and a synthetic peptide containing its linear B-cell epitope, PLoS Negl. Trop. Dis. (2015), http://dx.doi.org/10.1371/journal.pntd.0003426.
[67] M.S. Alam, D. Ghosh, M.G.M. Khan, et al., Survey of domestic cattle for anti-Leishmania antibodies and Leishmania DNA in a visceral leishmaniasis endemic area of Bangladesh, BMC Vet. Res. 7 (1) (2011) 27, http://dx.doi.org/10.1186/1746-6148-7-27.
[68] C. Gao, J. Wang, S. Zhang, Y. Yang, Y. Wang, Survey of wild and domestic mammals for infection with Leishmania infantum following an outbreak of desert zoonotic visceral Leishmaniasis in Jiashi, People's Republic of China, Munderloh UG, ed, PLoS One 10 (7) (2015) e0132493, , http://dx.doi.org/10.1371/journal.pone.0132493.
[69] A.P. Jackson, H.C. Allison, J.D. Barry, M.C. Field, C. Hertz-Fowler, M.A. Berriman, Cell-surface Phylome for African Trypanosomes, Tschudi C, ed, PLoS Negl. Trop. Dis. 7 (3) (2013) e2121, , http://dx.doi.org/10.1371/journal.pntd.0002121.