



OPEN

# RNA-sequencing of the *Nyssomyia neivai* sialome: a sand fly-vector from a Brazilian endemic area for tegumentary leishmaniasis and pemphigus foliaceus

Sebastian Vernal<sup>1,4,5</sup>, Fabiano Oliveira<sup>2,5✉</sup>, Wanderson H. C. Oliveira<sup>3</sup>, Thais M. Goulart<sup>3</sup>, James Oristian<sup>2</sup>, Eric Calvo<sup>2</sup>, Mara C. Pinto<sup>3</sup>, Ana Maria Roselino<sup>1</sup> & José M. C. Ribeiro<sup>2</sup>

Leishmaniasis encompasses a spectrum of diseases caused by a protozoan belonging to the genus *Leishmania*. The parasite is transmitted by the bite of sand flies, which inoculate the promastigote forms into the host's skin while acquiring a blood meal. *Nyssomyia neivai* is one of the main vectors of tegumentary leishmaniasis (TL) in Brazil. Southeastern Brazil is an endemic region for TL but also overlaps with an endemic focus for pemphigus foliaceus (PF), also known as *Fogo Selvagem*. Salivary proteins of sand flies, specifically maxadilan and LJM11, have been related to pemphigus etiopathogenesis in the New World, being proposed as an environmental trigger for autoimmunity. We present a comprehensive description of the salivary transcriptome of the *N. neivai*, using deep sequencing achieved by the Illumina protocol. In addition, we highlight the abundances of several *N. neivai* salivary proteins and use phylogenetic analysis to compare with Old- and New-World sand fly salivary proteins. The collection of protein sequences associated with the salivary glands of *N. neivai* can be useful for monitoring vector control strategies as biomarkers of *N. neivai*, as well as driving vector-vaccine design for leishmaniasis. Additionally, this catalog will serve as reference to screen for possible antigenic peptide candidates triggering anti-Desmoglein-1 autoantibodies.

Leishmaniasis encompasses a broad spectrum of diseases caused by an obligate intramacrophage protozoan belonging to the genus *Leishmania* (Kinetoplastida: Trypanosomatidae), occurring predominantly in the tropical and subtropical world regions. The bite of sand fly vectors (Diptera: Psychodidae: Phlebotominae) transmits promastigote forms of *Leishmania* into the host's skin while acquiring a blood meal<sup>1,2</sup>. Additionally, the phlebotomine deliver the parasite in conjunction with salivary proteins<sup>3</sup>, whose pharmacological activities assist blood feeding by preventing host hemostasis and modulating the host's immune system<sup>4-6</sup>.

*Leishmania (Viannia) braziliensis* (Viana 1911) and *Leishmania (Leishmania) amazonensis* (Lainson & Shaw 1972) are the main species involved in both the cutaneous and mucocutaneous forms of Tegumentary Leishmaniasis (TL) in Brazil<sup>7</sup>. Autochthonous cases of TL have been reported in the northeastern São Paulo state (NSPS), Southeastern Brazil. Both sand flies: *Nyssomyia intermedia* (Lutz & Neiva 1912) (syn = *Lutzomyia (Nyssomyia) intermedia*) and *Nyssomyia neivai* (Pinto 1926) are the main vectors of *L. (V.) braziliensis* in São Paulo state; however, only *N. neivai* has been recognized in systematic research collections in Northeastern São Paulo State (NSPS) cities<sup>8</sup>.

Pemphigus foliaceus (PF) is an autoimmune bullous disease caused by autoantibodies against desmoglein-1 (Dsg-1). PF is subdivided into classic sporadic worldwide Cazenave's pemphigus and endemic pemphigus (known in Brazil as *Fogo Selvagem*). Although the pathogenesis of PF remains unclear, genetic and environmental factors have been implicated in the susceptibility to this disease<sup>9,10</sup>. Interestingly, some salivary proteins of sand flies

<sup>1</sup>Division of Dermatology, Department of Clinical Medicine, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil. <sup>2</sup>Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Rockville, Maryland, USA. <sup>3</sup>Department of Biological Sciences, School of Pharmaceutical Sciences, São Paulo State University (UNESP), Araraquara, São Paulo, Brazil. <sup>4</sup>Present address: Department of Infectious Diseases, University of São Paulo, São Paulo, Brazil. <sup>5</sup>These authors contributed equally: Sebastian Vernal and Fabiano Oliveira. ✉email: loliveira@niaid.nih.gov

have been associated with pemphigus etiopathogenesis. Maxadilan is a highly immunogenic salivary protein described in *Lutzomyia longipalpis* (Lutz and Neiva 1912), the vector of visceral leishmaniasis (VL) in South America. Higher levels of serum IgG against maxadilan were observed in PF patients compared to healthy controls living in the same endemic region<sup>11,12</sup>. There is also evidence that antibodies raised against LJM11 and LJM17, immunogenic proteins from *Lu. longipalpis*, cross-reacts with antibodies against Dsg-1, and it has been proposed as the antigen that triggers PF diagnosed in Amerindians living in Mato Grosso state, Brazil<sup>13,14</sup>. Of note, *Lu. longipalpis* is not widely distributed in NSPS<sup>10,15</sup>.

The profile of salivary components has been defined in ten Old World Phlebotomus species<sup>16–24</sup>, and one Sergentomyia (*Sergentomyia schwetzi*) species<sup>25</sup>; notwithstanding, the sialotranscriptome of the New World sand flies have been documented only in four species: *Bichromomyia olmeca* (Vargas and Diaz-Najera 1959) (syn = *Nyssomyia olmeca*), *Lutzomyia ayacuchensis* (Caceres and Galati 1988), *Lu. longipalpis*, and *N. intermedia*<sup>26–29</sup>. Considering the developments on sand fly saliva-based vaccines for *Leishmania sp.* infection, and discovery of possible candidate proteins that might be the trigger of anti-Dsg1 autoantibodies in Brazilian endemic PF in NSPS, we here report the identity and abundances of the putative secreted proteins on the sialome of *N. neivai* by RNA-sequencing. We take this opportunity to compare the most abundant proteins in the *N. neivai* sialome to *N. intermedia*<sup>27</sup> (the main vector of *Leishmania (V.) braziliensis* in the coastal SP state), and *Lu. longipalpis*<sup>28</sup> (the vector of VL in São Paulo state), besides other published salivary proteins in Old World sand flies.

## Methods

**Collection and maintenance of sand flies.** A colony of *N. neivai* was established at São Paulo State University<sup>30</sup> from sand flies collected in Santa Eudóxia, SP state, Brazil (along the edges of the *Mogi Guaçu* River) on the wall of a house, using a manual aspirator between 06:00 PM and 11:00 PM. In the laboratory, the sand flies were maintained in cages covered with voile (30 cm<sup>3</sup>) at 26 ± 1 °C, 80–90% humidity, and a 12:12 (L:D) photoperiod. Salivary glands were dissected as follows: Sand flies were transferred into a tube with mild soap solution. The contents were poured over a fine-mesh screen stretched over a beaker and rinsed with water. We transferred the flies from the mesh screen to a small Petri dish containing 1xPBS. We relocated a freshly rinsed sand fly to a drop of PBS on the microscope slide. The sand flies' legs and wings were removed. We pierced the thorax and held it against the glass and removed the sand fly's head. When the head was pulled from the body the salivary glands became visible at the back of the head. The glands were collected with the dissecting pins and transferred to a small, labeled Eppendorf tube for storage. Sand fly identification was based on morphological characteristics of the spermathecae in females and apical genital filaments on males as described by Andrade Filho et al.<sup>31</sup>.

**Salivary gland transportation.** Two hundred salivary gland pairs were dissected from starved and non-gravid, 2 to 3 days old *N. neivai* female sand flies. Samples were submitted in 1 mL RNAlater each to the North Carolina State Genomic Sciences Laboratory (Raleigh, NC, USA) for RNA extraction, Illumina library preparation, and sequencing. The salivary glands tissue samples export was approved by the United States Department of Agriculture—Veterinary Permit under the ID number #130339.

**Salivary gland RNA extraction.** Prior to extraction, salivary glands were pelleted by the addition of 1 mL PBS in a benchtop centrifuge at 5000×g for 10 m to remove the RNAlater. Total RNA was extracted using the RNeasy Mini Kit (Qiagen, MD, USA) following the manufacturer's protocol for purification of total RNA from animal tissue. Briefly, Qiagen RLT buffer with β-Mercaptoethanol (β-ME) was added to the tissue samples, and samples were homogenized using a Qiagen TissueLyser with 5 mm stainless steel beads (Qiagen). Samples were then purified with provided RNeasy spin columns. Total RNA was then assessed for purity and size integrity using an Agilent 2100 Bioanalyzer with an RNA 6000 Nano Chip (Agilent Technologies, CA, USA). Purification of messenger RNA (mRNA) was performed using oligo-dT beads provided in the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, MA, USA). Complementary DNA (cDNA) libraries for Illumina sequencing were constructed using the NEBNext Ultra Directional RNA Library Prep Kit (NEB) and NEBNext Multiplex Oligos for Illumina (NEB) using the manufacturer-specified protocol.

**Sialome library.** Briefly, the mRNA was chemically fragmented and primed with random Oligos for first strand cDNA synthesis. Second strand cDNA synthesis was then carried out with dUTPs to preserve strand orientation information. The double-stranded cDNA was then purified, end repaired, and “a-tailed” for adaptor ligation. Following ligation, the samples were selected for a final library size (adapters included) of 400–550 bp using sequential AMPure XP bead isolation (Beckman Coulter, USA). Library enrichment was performed, and specific indexes for each sample were added during the protocol-specified PCR amplification. The amplified RNA-seq library fragments were purified and checked for quality and final concentration using an Agilent 2200 TapeStation (D1000 chip, Agilent Technologies, CA, USA) combined with a Qubit fluorometer (Thermo-Fisher, MA, USA). The final quantified libraries were pooled in equimolar amounts for clustering and sequencing on an Illumina HiSeq 2500 DNA sequencer, utilizing a 125 bp single-end cycle sequencing kit (Illumina, CA, USA). The software package Real Time Analysis (RTA) was used to generate raw bcl, or base call files, which were then de-multiplexed by sample into fastq files using bcl2fastq Conversion Software v2.17 (Illumina, CA, USA).

**Bioinformatics.** Custom bioinformatic analysis were described elsewhere<sup>32</sup>. Succinctly, low quality reads were trimmed from Fastq files (<20) and contaminating adapter primer sequences removed. De novo assembly from reads was a result of Abyss<sup>33,34</sup> (using k parameters from 21 to 91 in fivefold increments) and SOAP de novo-trans<sup>35</sup> assemblers. The fasta files were combined and further assembled using an iterative blast and CAP3<sup>36</sup> pipeline as previously described<sup>37</sup>. Coding sequences (CDs) were extracted based on the existence of a

Class	Number of contigs	% of contigs	RPKM	% RPKM
Secreted	494	41.2	1,203,951	92.4
Housekeeping	399	33.3	42,119	3.2
Unknown	295	24.6	56,243	4.3
Viral products	3	0.3	31	0.0
Transposable elements	9	0.8	52	0.0
Total	1,200	100	1,302,396	100

**Table 1.** Classification of transcripts originating from the sialotranscriptome of *Nyssomyia neivai*.

Putative secreted proteins		Number of contigs	% of contigs	RPKM	% RPKM
Enzymes	Apyrase	3	0.6	18,214	1.5
	5' Nucleotidase	4	0.8	7537	0.6
	Hyaluronidase	7	1.4	6887	0.6
	Endonuclease	4	0.8	4927	0.4
	Adenosine deaminase	2	0.4	2923	0.2
	Proteases	53	10.7	1186	0.1
	Lipases*	8	1.6	53	0.0
	Alpha amylase*	1	0.2	2	0.0
Other ubiquitous protein families	Antigen 5 protein	6	1.2	47,405	3.9
	Antimicrobial peptides	8	1.6	781	0.1
	Protease inhibitors	7	1.4	48	0.0
Small molecule binding proteins	C type lectins	20	4.0	190,860	15.9
	ML domain salivary protein	13	2.6	69,404	5.8
	Yellow protein family	13	2.6	61,755	5.1
	Mucins	22	4.5	13,174	1.1
	Diptera conserved salivary secreted peptide	59	11.9	3523	0.3
	JH/PB/OBP	36	7.3	1393	0.1
	Lipocalins	3	0.6	147	0.0
	Hormones	3	0.6	29	0.0
Galectin	1	0.2	3	0.0	
Hematophagous nematocera specific families	D7 family	4	0.8	26,832	2.2
Sand fly specific families	SP13-15 protein family	22	4.4	423,542	35.09
	15 kDa family	15	3.0	278,824	23.2
	14 kDa family	2	0.4	8989	0.7
	13 kDa family	3	0.6	135,729	11.3
	Maxadilan	16	3.2	187,866	15.6
	10 kDa family	3	0.6	40,594	3.4
	Hypothetical secreted protein	149	30.2	31,893	2.6
	Toxin-like peptide	12	2.4	24,330	2.0
	30 kDa family	2	0.4	16,470	1.4
	32.4 salivary protein: Lufaxin	2	0.4	7913	0.7
	34 kDa family	1	0.2	6850	0.6
	Nyssomyia specific	1	0.2	3602	0.3
	38.8 kDa family	1	0.2	1958	0.2
	56.6 kDa salivary protein	3	0.6	1029	0.1
	Gly rich salivary protein	6	1.2	808	0.1
	15.5 kDa family	1	0.2	13	0.0

**Table 2.** Classification of secreted proteins originating from the sialotranscriptome of *Nyssomyia neivai*.

Protein ID	Abundance index (%)	Accession number	Best match to NCBI NR or TSA-NR databases	E-value	Identity (%)	Best match accession number	Seq size	MW	pI
JAV08233.1	26.14	GFDF01005851.1	SP15 family member <i>N. intermedia</i>	1.00E-87	90.98	AFP99232.1	133	15.61	9.2
JAV08232.1	23.80	GFDF01005852.1	SP15 family member <i>N. intermedia</i>	2.00E-87	90.23	AFP99232.1	133	15.58	9.1
JAV08238.1	23.28	GFDF01005846.1	SP15 family member <i>N. intermedia</i>	9.00E-101	98.56	AFP99232.1	139	16.48	9.16
JAV08231.1	18.15	GFDF01005853.1	SP15 family member <i>N. intermedia</i>	9.00E-84	90.00	AFP99232.1	158	18.63	9.44
JAV08239.1	5.00	GFDF01005845.1	SP15 family member <i>N. intermedia</i>	6.00E-98	97.12	AFP99239.1	139	16.09	9.16
JAV08235.1	1.29	GFDF01005849.1	SP15 family member <i>N. intermedia</i>	6.00E-97	97.12	AFP99266.1	139	16.33	9.57
JAV08236.1	1.25	GFDF01005848.1	SP15 family member <i>N. intermedia</i>	4.00E-97	96.40	AFP99266.1	139	16.33	9.45
JAV08237.1	1.08	GFDF01005847.1	SP15 family member <i>N. intermedia</i>	2.00E-96	97.12	AFP99266.1	139	16.38	9.5

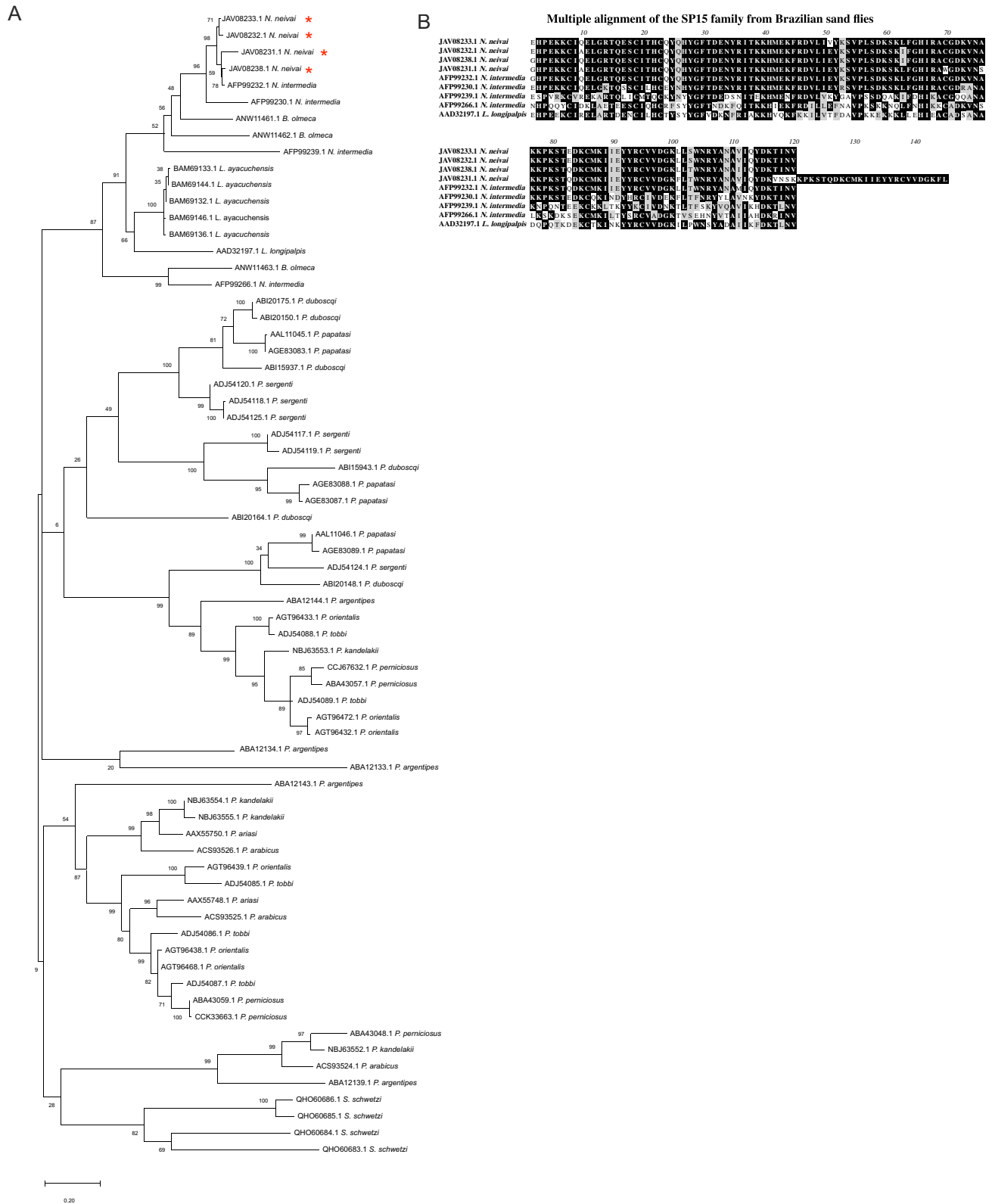
**Table 3.** SP15 secreted proteins originating from the sialotranscriptome of *Nyssomyia neivai*.

signal peptide in the longer open reading frame (ORF) and by similarities to other proteins found in the Refseq invertebrate database from the National Center for Biotechnology Information (NCBI), proteins from “Diptera [organism]” deposited at NCBI’s GenBank, and from Swiss-Prot. Contigs containing an open reading frame and any similarity to sequences in the chosen databases were selected for further analysis. Reads for each library were mapped on the deduced CDs using blastn<sup>38</sup> with a word size of 25, 1 gap allowed, and 95% identity or better required. Up to five matches were allowed if and only if the scores were the same as the largest score. Mapping of the reads was also included in the Excel spreadsheet. Values of the Reads Per Kilobase of transcript, per Million mapped reads (RPKM)<sup>39</sup> for each coding sequence were also mapped to a spreadsheet. Automated annotation of proteins was based on a vocabulary of nearly 350 words found in matches to various databases: Swiss-Prot, Gene Ontology, KOG, Pfam, SMART, Refseq-invertebrates, and the GenBank Diptera subset. Raw reads were deposited on the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under BioProject ID PRJNA359206 and read accession SRR5134059. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GFDF00000000.

**Phylogenetic analysis.** For multiple sequence alignment and phylogenetic analysis, abundant salivary proteins from *N. neivai*, had their predicted signal peptide signal (SignalP-5.0 server<sup>40</sup>) removed, and resulting protein sequence entered into a Basic Local Alignment Search tool (BLAST<sup>38</sup>) against NR and TSA databases. We selected the five most similar homolog sequences (based on the e-value) for each sand fly species. The cut-off to exclude a homologue was an e-value above  $1^{-10}$ , except for the *N. neivai* yellow family of proteins where a homologue from *Drosophila* was used to root the tree. Multiple sequence alignment and identity/similarity matrix were constructed on MacVector v15.5.3 with MUSCLE<sup>41</sup> using PAM 200 profile. We determined the best method for amino acid substitution using the “Find best protein Models” feature of MEGA7<sup>42</sup>. A score was given to each of 56 amino acid substitution models including the mixing Gamma and invariant sites likelihood. The option with the lowest Bayesian information criterion score was selected to build the tree. Through this feature, it was determined that the best amino acid substitution model for phylogeny as follows: WAG for the ML domain and Maxadlan trees. The model WAG with discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+ G, parameter = 1.6114)) for the Yellow proteins tree. For the SP15 family of proteins, the model WAG with a discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+ G, parameter = 4.0108)). The rate variation model allowed for some sites to be evolutionarily invariable ([+ I], 2.61% sites). For the C-type lectin, the best model was LG + F. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+ G, parameter = 3.6037)). For Gaps/Missing data treatment, a partial deletion option was utilized. Finally, the reliability of the trees was tested, by bootstrap method (N = 1000).

## Results and discussion

**cDNA library of *Nyssomyia neivai* salivary gland.** cDNA library was constructed from salivary glands of *N. neivai* females dissected up to 3 days after emergence. From this cDNA library 1,302,396 high quality reads were assembled in 1200 contigs (Table 1). Contigs were classified in five categories namely: secreted, housekeeping, transposable elements, viral, and unknown. Remarkably, the secreted proteins category comprised 92.4% of the number of reads, dispersed in 41.2% of the contigs. Most salivary transcriptomes from *Phlebotomus* and *Lutzomyia* genera were based in low output cDNA library sequencing; nevertheless, a high abundance of transcripts encoding secreted proteins were also reported<sup>17–20,23,27,43,44</sup>. Our data further validates the specialization of the salivary gland machinery and the specificity of the material obtained with the sand fly dissection. The housekeeping category had 33.3% of the clusters and only 3.2% of the total sequences. The category of “unknowns” comprised 24.6% of the clusters and 4.3% of the sequences. Finally, Viral products and transposable elements included less than 1% of the families (0.3% and 0.8% respectively) and less than 0.1% of total sequences. Recently,



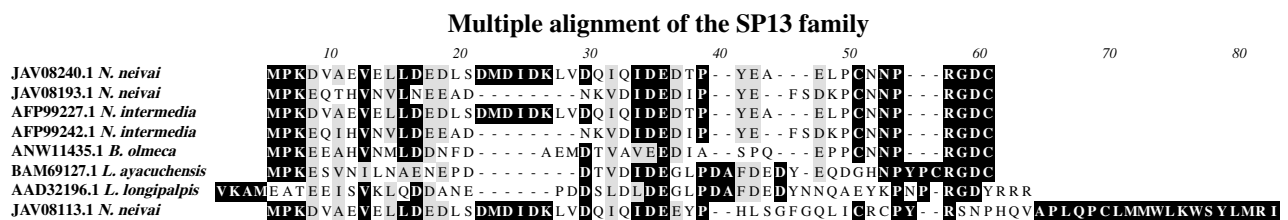
**Figure 1.** Molecular phylogenetic analysis and sequence alignment of *Nyssomyia neivai* 15 protein family. (A) The evolutionary history was inferred by using the Maximum Likelihood method. The tree with the highest log likelihood is shown. The tree is drawn to scale, with branch lengths measured in number of substitutions per site. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter = 4.0108)]. The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 2.61% sites). All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7. (B) Multiple alignments of SP15 from *Nyssomyia neivai* with *Nyssomyia intermedia* and *Lutzomyia longipalpis* SP15 proteins using Muscle. Black shading represents identical amino acids, light gray shading represents similar amino acids.

	JAV08233.1.N. <i>neivai</i>	JAV08232.1.N. <i>neivai</i>	JAV08238.1.N. <i>neivai</i>	JAV08231.1.N. <i>neivai</i>	AAD32197.1 L. <i>longipalpis</i>	AFP99239.1.N. <i>intermedia</i>	AFP99232.1.N. <i>intermedia</i>	AFP99230.1.N. <i>intermedia</i>	AFP99266.1.N. <i>intermedia</i>
JAV08233.1 N. <i>neivai</i>		97.5	95	74.3	57.1	55.5	95.8	81.5	52.9
JAV08232.1 N. <i>neivai</i>	98.3		95.8	75	56.3	56.3	95	80.7	52.9
JAV08238.1 N. <i>neivai</i>	98.3	98.3		75.7	56.3	54.6	98.3	79.8	52.9
JAV08231.1 N. <i>neivai</i>	76.4	77.8	77.8		43.1	41.7	76.4	63.2	41.7
AAD32197.1 L. <i>longipalpis</i>	74.8	73.9	73.1	55.6		48.7	56.3	53.8	50.4
AFP99239.1 N. <i>intermedia</i>	76.5	75.6	75.6	58.3	67.2		53.8	57.1	40.3
AFP99232.1 N. <i>intermedia</i>	98.3	98.3	100	77.8	73.1	75.6		80.7	52.9
AFP99230.1 N. <i>intermedia</i>	89.1	89.1	89.1	70.1	71.4	74.8	89.1		47.1
AFP99266.1 N. <i>intermedia</i>	72.3	73.1	73.9	58.3	64.7	63.9	73.9	66.4	

**Table 4.** Pairwise comparison matrix of identity and similarity percentages.

Protein ID	Abundance index (%)	Accession number	Best match to NCBI NR or TSA-NR databases	E-value	Identity (%)	Best match accession number	Seq size	MW	pI
JAV08240.1	72.86	GFDF01005844.1	SP13 family member <i>N. intermedia</i>	6.00E-42	98.53	AFP99227.1	71	8.03	3.86
JAV08113.1	23.73	GFDF01005971.1	SP13 family member <i>N. intermedia</i>	2.00E-28	82.09	AFP99227.1	94	10.98	4.49
JAV08193.1	3.41	GFDF01005891.1	SP13 family member <i>N. intermedia</i>	1.00E-36	96.72	AFP99242.1	69	7.96	4.57

**Table 5.** SP13 secreted proteins originating from the sialotranscriptome of *Nyssomyia neivai*.



**Figure 2.** Molecular phylogenetic analysis and sequence alignment of *Nyssomyia neivai* 15 protein family. (A) Multiple alignments of SP13 from *Nyssomyia neivai* with other SP13 sand fly salivary proteins using Muscle. Black shading represents identical amino acids, light gray shading represents similar amino acids.

RNA-seq of salivary glands of Old World *P. kandelakii*<sup>24</sup> and *Sergentomyia schwetzi* have been published<sup>25</sup> and the presence of representative salivary proteins has been confirmed.

**Housekeeping and unknown proteins sequences.** The 399 clusters (comprising 42,119 sequences) attributed to Housekeeping genes expressed in the salivary glands of *N. neivai* were further divided into 22 subgroups according to their function. Two sets were associated with (a) protein synthesis machinery (22 contigs), including translation, ribosomal structure and biogenesis, and (b) metabolism (94 contigs), a pattern also observed in other sialotranscriptomes. Proteins with unknown function (295 contigs comprising 4.3% of reads) were classified as “unknown”.

**Secreted proteins sequences.** The putative secreted salivary proteins of *N. neivai* were classified into 35 main protein families (Table 2). The most abundant transcripts were within the SP13-15 protein family (35.09%), followed by C-type lectins (15.9%), Maxadilan-like (15.6%), ML domain salivary proteins (5.8%), and the Yellow protein family (5.1%). Previously, the novel families 8-kDa, 6-kDa and 5-kDa that were only described in the *N. intermedia* sialome<sup>27</sup>, are now also present in *N. neivai* sialotranscriptome and grouped as Toxin-like peptides.

Protein ID	Abundance index (%)	Accession number	Best match to NCBI NR or TSA-NR databases	E-value	Identity (%)	Best match accession number	Seq size	MW	pI
JAV08563.1	19.52	GFDF01005521.1	C-type lectin <i>N. intermedia</i>	1.00E-105	92.26	AFP99244.1	156	18.33	8.47
JAV08583.1	15.90	GFDF01005501.1	C-type lectin <i>N. intermedia</i>	8.00E-118	100.00	AFP99236.1	157	18.6	8.64
JAV08561.1	15.69	GFDF01005523.1	C-type lectin <i>N. intermedia</i>	1.00E-99	90.67	AFP99244.1	152	17.77	8.69
JAV08565.1	14.82	GFDF01005519.1	C-type lectin <i>N. intermedia</i>	2.00E-105	94.08	AFP99243.1	158	18.92	8.81
JAV08584.1	13.44	GFDF01005500.1	C-type lectin <i>N. intermedia</i>	9.00E-111	96.15	AFP99243.1	156	18.58	9.21
JAV08562.1	10.08	GFDF01005522.1	C-type lectin <i>N. intermedia</i>	1.00E-96	95.07	AFP99243.1	149	17.21	8.82
JAV08549.1	3.31	GFDF01005535.1	tfid subunit <i>C. appendiculata</i>	2.00E-85	64.92	JAB55018.1	248	27.68	9.32
JAV08548.1	2.72	GFDF01005536.1	tfid subunit <i>C. appendiculata</i>	2.00E-84	65.29	JAB55018.1	229	25.62	9.1
JAV08581.1	2.41	GFDF01005503.1	C-type lectin <i>N. intermedia</i>	3.00E-132	98.37	AFP99256.1	184	21.93	6.18
JAV08559.1	2.10	GFDF01005525.1	C-type lectin <i>N. intermedia</i>	2.00E-96	98.54	AFP99271.1	141	16.31	9.37
JAV08596.1	0.01	GFDF01005488.1	C-type lectin <i>C. tarsalis</i>	3.00E-156	93.61	JAV29959.1	219	25	8.45
JAV08591.1	0.01	GFDF01005493.1	C-type lectin <i>P. kandelakii</i>	6.00E-53	51.37	NBJ58870.1	151	17.41	8.85

**Table 6.** C-type lectin secreted proteins originating from the sialotranscriptome of *Nyssomyia neivai*.

The following paragraphs describe the most abundant families in detail, focusing on protein family characteristics, possible function, biochemical, immune-modulatory, and antigenic properties; also, phylogenetic analysis in context with related proteins from other Brazilian sand flies and desmoglein proteins.

**SP15 family.** The SP15 is the most abundant secreted protein family in *N. neivai* sialome with 23.13% of the total RPKM in the transcriptome (Table 2). This salivary family is present among all species of sand flies studied so far. In *N. neivai*, we have categorized eight full-length members of this family (Table 3). For further analysis we considered the four most abundant members JAV08233.1, JAV08232.1, JAV08238.1 and JAV08231.1 with 91.37% of the SP13-15 family abundance as highlighted in Table 3.

The SP15 family was described for the first time in the sialome of *Lu. longipalpis* as the SL1 family<sup>45</sup>. This family was then also reported in the Old World sand fly *Phlebotomus perniciosus* (Newstead 1911), and later named as SP15 family due to 15-kDa salivary protein from *Phlebotomus papatasi* (Scopoli 1786) (PpSP15: AF335487)<sup>19</sup>. Thus far, SP15-like proteins have only been reported in sand flies and not in any other Dipteran; It has been suggested that SP15-like proteins were derived from an ancestral odorant-binding protein and were closely related to mosquitoes short D7 proteins<sup>16,19</sup>. Alvarenga et al.<sup>46</sup>, demonstrated that SP15 from *Phlebotomus duboscqi* (Neveu-Lemaire 1906) inhibit anionic surface-mediated reactions suggesting a role in anticoagulation, inhibiting the activation of FXII and FXI, and anti-inflammatory processes.

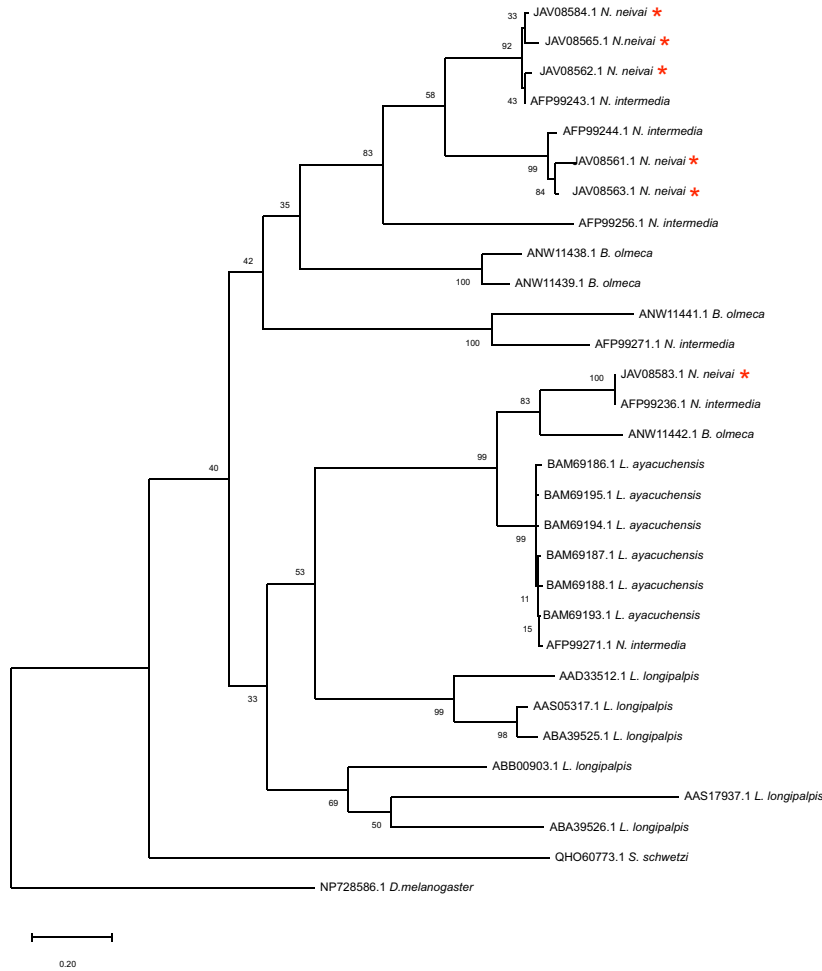
PpSP15-like proteins were reported as promising anti-*Leishmania* vaccine candidates. Immunization of mice with *P. papatasi* SP15 protein conferred partial protection against *Leishmania (Leishmania) major* (Yakimoff and Schokhor 1914) infection<sup>43</sup>; furthermore, a DNA vaccine containing the PpSP15 cDNA provided the same protection<sup>43</sup>. ParSP03 (AAX56359), a PpSP15-like protein from *Phlebotomus ariasi* (Tonnoir 1921), elicited similar delayed type hypersensitivity and humoral immune responses upon DNA vaccination<sup>20</sup>. Recently, BALB/c mice immunized to PsSP19 (HM56964), a protein member of the SP15 family from *Phlebotomus sergenti* (Parrot 1917), acted as an adjuvant to accelerate the cell-mediated immune response to co-administered *Leishmania* antigens, providing protection against *Leishmania (Leishmania) tropica* (Wright 1903) infection<sup>47</sup>.

Phylogenetic analysis comparing selected sequences from sand fly salivary transcriptomes to the *N. neivai* SP15 family clustered these 4 abundant members closely (Fig. 1A, red asterisks) in a New World sand fly clade next to members of *N. intermedia*, *B. olmeca*, *Lu. ayacuchensis* and to the only SP-15 family protein described in *Lu. longipalpis* so far, SL1<sup>28</sup> (AAD32197.1) (Fig. 1A). The remaining clades represent Old world VL and TL vectors and the three members of the *S. schwetzi* SP-15 family. We then aligned *N. neivai* SP15 proteins to SP15 salivary proteins from other sand fly vectors present in Brazil, namely *N. intermedia* and *Lu. longipalpis* (Fig. 1B). *N. neivai* SP15 family proteins shared a relatively high percentage of identity (41.7 to 98.3%) to *N. intermedia* and at a lesser extent to *Lu. longipalpis* SL1 (43.1 to 57.1%) (Table 4).

*N. neivai* SP15 members shared 76.4 to 98.3% identity (Table 4) with *N. intermedia* Linb-8 (AFP99232.1). Interestingly, BALB/c immunization with DNA plasmids encoding Linb-8 induced the highest humoral immune response against *N. intermedia* salivary gland homogenate (SGH), even greater than Linb-7 (AFP99230.1), another SP-15 protein family member also tested<sup>27</sup>. Linb-7-immunized mice induced a strong humoral response leading to a sustained local inflammatory process, which could exacerbate *Leishmania sp.* infection by *L. (V.) braziliensis*<sup>27</sup>. Regarding *N. neivai* and *N. intermedia*, which are the two vectors of *L. (V.) braziliensis* in South-eastern Brazil, the observed high similarities between *N. neivai* members and Linb-8 make them possible targets as biomarkers of vector exposure and as a vector-based vaccine for TL in Brazil.

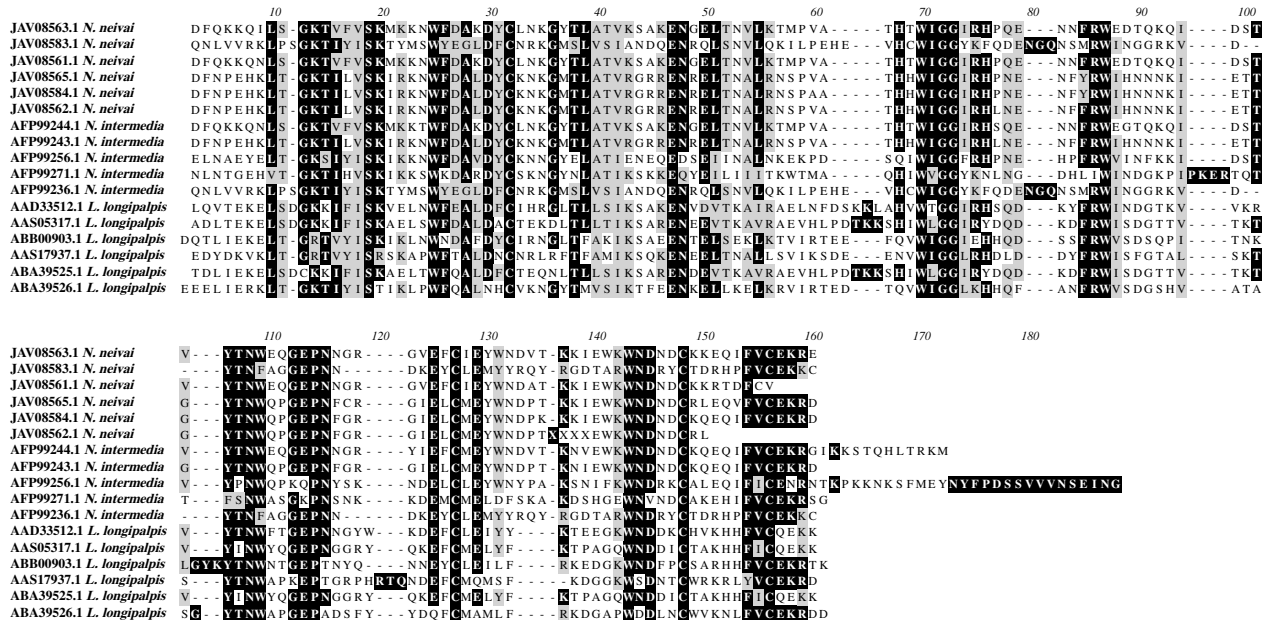
**SP13 family.** The SP13 *N. neivai* family (JAV08240.1, JAV08113.1 and JAV08193.1) represents 11.26% of the total RPKM in the transcriptome and has the most abundant salivary transcript (JAV08240.1) representing 72.86% of the reads belonging to this family (Table 5). Despite its abundance in *N. neivai*, searches for SP13 sub-family (JAV08240.1, JAV08113.1 and JAV08193.1) related members available in the NCBI NR and TSA-NR data-

A



B

Multiple alignment of the C-type lectin family from Brazilian sand flies





◀**Figure 3.** Molecular phylogenetic analysis and sequence alignment of *Nyssomyia neivai* C-type lectin protein family. (A) The evolutionary history was inferred by using the Maximum Likelihood method. The tree with the highest log likelihood is shown. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+ G, parameter = 3.6037)]. The tree is drawn to scale, with branch lengths measured in number of substitutions per site. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7. (B) Multiple alignments of C-type lectins from Brazilian sand flies using Muscle. Black shading represents identical amino acids, light gray shading represents similar amino acids.

bases (blast-P hits with e-value lower than  $1e-10$ ), yielded few sand fly salivary proteins. We identified only two members from *N. intermedia* (AFP99227.1 and AFP99242.1), one member for *L. ayacuchensis* (BAM69127.1) and one from *B. olmeca* (ANW11435.1) (Fig. 2). The *N. neivai* JAV08240.1 member is identical to *N. intermedia* Linb-1 (AFP99227.1) that was also the most abundant contig in the *N. intermedia* salivary transcriptome<sup>27</sup>. Interestingly, all the members identified, but JAV08113.1 (Fig. 2) share the RGD domain in the carboxy region that are common in members of the disintegrin family<sup>48,49</sup>. These RGD containing proteins had been previously observed in other New World sand flies such as *Lu. ayacuchensis* (LuayaRGD; BAM69127.1) and *Lu. longipalpis* (LuloRGD; AAD32196). The function and relevance of this family during blood feeding remains to be tested.

**C-type lectin family.** We have categorized twelve novel full-length transcripts as C-type lectins in the *N. neivai* sialome, as shown in Table 6. The C-type lectins are the third most abundant salivary family in *N. neivai*. Of those twelve, six members (JAV08563.1, JAV08583.1, JAV08561.1, JAV08565.1, JAV08584.1, JAV08562.1) corresponded to 89.5% of this family abundance and were further considered for in depth analysis (Table 6). In vertebrates, protein-carbohydrate interactions serve multiple functions in the immune system. C-type lectin family members are components of the innate immune response and work via pathogen neutralization through the activation of the complement pathway and adaptive immune response<sup>50</sup>. The C-type lectin putative domain may function as a  $Ca^{2+}$ -dependent carbohydrate-binding pocket involved in extracellular matrix organization, pathogen recognition, and cell-to-cell interactions<sup>50</sup>. Homologous salivary proteins with molecular weight of 16.2–16.5 kDa have been identified in New World sand flies. Recently, homologues were also found in Old World sand flies by next generation sequencing of salivary glands from *P. kandelakii*<sup>24</sup>, as a partial protein, and *Sergentomyia schwetzi*<sup>25</sup>. The most abundant members of the *N. neivai* C-type lectins family seem to have a close relationship to *N. intermedia* homologues (Fig. 3A). Interestingly, JAV08583.1N. *neivai* protein segregated from the other C-type lectin *N. neivai* members in a subtree that also encompasses members from *Lu. ayacuchensis* and *Lu. longipalpis*. The protein sequence alignment comparing Brazilian sand flies depicts a scenario of fast evolution of this family (Fig. 3B), indicated by the large ranges of amino acid identity scores (from 96.4 to 26.7%) across species in pairwise comparisons (Table 7). This may be associated with multiple events of gene duplication and high immune pressure from hosts. The exact role of these proteins in sand flies remains elusive.

**Maxadilan-like family.** Maxadilan (AAA29288.1) is a 7-kDa peptide present in the salivary gland of the sand fly *Lu. longipalpis*. Maxadilan was the first molecule to be identified in sand fly saliva<sup>51</sup>, and it is recognized for its powerful vasodilator effect. *N. neivai* maxadilan-like family corresponds of 11 full length abundantly expressed proteins (Table 8) representing 15.6% of the transcriptome. We will be further discussing 6 members of this family (JAV08475.1, JAV08473.1, JAV08474.1, JAV08472.1, JAV08471.1, JAV08462.1) accounting for 90.14% of this family abundance.

Comparative analyses of abundant transcripts from *N. neivai* maxadilan-like family were able to identify three homologues in *N. intermedia* (Fig. 4A). Phylogenetic topology shows that a main clade clustered the most abundant *N. neivai* members (JAV08475.1, JAV08473.1, JAV08474.1) with Linb-9 (AFP99245.1), while the other *N. neivai* members clustered with Linb-25 from *N. intermedia*. Maxadilan has its own branch and, JAV08462.1 from *N. neivai* was the closest relative to Maxadilan with 37.5% identity and 56.2% similarity suggesting that *N. neivai* JAV08462.1 (Fig. 4B) may have preserved its pharmacological properties<sup>52</sup>. *N. neivai* JAV08462.1 represents the sixth most abundant member with 6.44% of the maxadilan-like family abundance. Linb-147 (JK846521), a partial sequence from *N. intermedia*, showed a similar match to maxadilan, provided for only 34% identity and 70% similarity over a stretch of 50 amino acids. This sequence seems to be scarcely present in *N. intermedia* sialome, with only one transcript identified in its cDNA library, as compared to 30 transcripts of maxadilan present in *Lu. longipalpis* sialome<sup>27,28</sup>.

Maxadilan-like proteins have never been identified in Old World *Phlebotomus* species<sup>17–20,23,27,43</sup>. *Phlebotomus* sand flies, except for *P. duboscqi*, contain large amounts of vasodilatory adenosine and AMP in their saliva<sup>18–20,23,25</sup>. Interestingly, they lack adenosine deaminase (ADA), an enzyme that hydrolyzes adenosine and adenosine monophosphate (AMP)<sup>53,54</sup>. In contrast, *Lu. longipalpis* has ADA and lacks adenosine and AMP in the saliva<sup>28,45</sup>. Unexpectedly, neither ADA nor maxadilan were identified in *Lu. ayacuchensis*. Thus, with those exceptions in general, salivary vasodilators of sand flies are adenosine and AMP in the *Phlebotomus* complex, and maxadilan in the *Lu. longipalpis* complex.

Inoculation of maxadilan in experimental animals exacerbates *Leishmania* infection to the same degree as the whole salivary gland homogenate<sup>55</sup>. This peptide can drive a Th1 response to Th2, up-regulate IL-10 and TGF- $\beta$  production, and suppress IL-12p40, TNF- $\alpha$ , and NO production<sup>56,57</sup>.

In animal models, mice vaccinated with maxadilan become markedly protected against *Leishmania* infection, producing not only anti-maxadilan antibodies, but also immune CD4 + T cells specific to maxadilan generating IFN- $\gamma$  and inducing NO production<sup>55</sup>. Notoriously, immunization against maxadilan also inhibits blood meal acquisition by sand flies, a promising target to block the vector reproductive process<sup>58</sup>. Nonetheless, maxadilan

	JAV08563.1 <i>N. neivai</i>	JAV08583.1 <i>N. neivai</i>	JAV08561.1 <i>N. neivai</i>	JAV08565.1 <i>N. neivai</i>	JAV08584.1 <i>N. neivai</i>	JAV08562.1 <i>N. neivai</i>	AFP99244.1 <i>N. intermedia</i>	AFP99243.1 <i>N. intermedia</i>	AFP99256.1 <i>N. intermedia</i>	AFP99271.1 <i>N. intermedia</i>	AFP99236.1 <i>N. intermedia</i>	AAD33512.1 <i>L. longipalpis</i>	AAS05317.1 <i>L. longipalpis</i>	ABB00903.1 <i>L. longipalpis</i>	AAS17937.1 <i>L. longipalpis</i>	ABA39525.1 <i>L. longipalpis</i>	ABA39526.1 <i>L. longipalpis</i>
JAV08563.1 <i>N. neivai</i>		33.6	92	65.7	66.4	58	84.6	67.2	40	39.4	33.6	41	39.6	42.4	37.8	36.8	35
JAV08583.1 <i>N. neivai</i>	51.7		31.5	36.4	36.4	32.6	31	36.4	26.9	28.4		36.7	32	35.1	30.6	31.3	34.7
JAV08561.1 <i>N. neivai</i>	92	48.3		60.6	60.6	59.7	79.2	61.3	36.4	34.5	31.5	39.6	38.9	39.6	34.3	36.1	31.5
JAV08565.1 <i>N. neivai</i>	82.5	53.1	75.9		95.6	87.7	58.4	94.9	44.2	38.7	36.4	37.5	37.5	37.5	37.8	34.7	36.4
JAV08584.1 <i>N. neivai</i>	82.5	53.1	75.9	97.1		85.5	59.7	96.4	44.2	39.4	36.4	38.2	38.2	37.5	37.8	35.4	36.4
JAV08562.1 <i>N. neivai</i>	72.5	47.2	74.6	88.4	87		53.3	88.4	39.8	33.6	32.6	35.9	36.6	33.1	33.3	33.8	31.2
AFP99244.1 <i>N. intermedia</i>	87.9	49.7	82.6	73.2	73.2	66		62.4	39.4	36.6	31	37.8	37.2	38.1	34.8	35.3	33.1
AFP99243.1 <i>N. intermedia</i>	82.5	53.8	75.9	97.1	97.1	89.1	75.2		44.2	40.1	36.4	38.9	38.9	37.5	37.8	36.1	36.4
AFP99256.1 <i>N. intermedia</i>	54.5	42.7	50.3	59.4	57.6	52.4	55.8	57.6		34.3	26.9	29.1	28.5	31.6	27.5	26.7	28.2
AFP99271.1 <i>N. intermedia</i>	57.7	47.3	52.8	55.6	55.6	49.7	55.6	57.7	46.2		28.4	30.9	32.9	32.4	28.4	29.5	29.9
AFP99236.1 <i>N. intermedia</i>	51.7	100	48.3	53.1	53.1	47.2	49.7	53.8	42.7	47.3		36.7	32	35.1	30.6	31.3	34.7
AAD33512.1 <i>L. longipalpis</i>	58.3	53.7	55.6	59	59	54.5	55.8	59.7	45.3	49	53.7		63.8	42.5	36.8	63.8	35.4
AAS05317.1 <i>L. longipalpis</i>	53.5	55.1	50.7	52.1	52.8	47.6	50.6	53.5	42.4	48.3	55.1	77.3		38.4	36.8	92.9	34.7
ABB00903.1 <i>L. longipalpis</i>	58.3	54.7	54.9	56.2	56.2	49.7	52.9	55.6	43.9	52	54.7	63	60.3		42.8	37.7	49.3
AAS17937.1 <i>L. longipalpis</i>	55.2	49	50.3	55.2	55.9	47.9	49	55.2	40.4	50.7	49	54.2	58.3	57.2		35.4	47.6
ABA39525.1 <i>L. longipalpis</i>	54.2	56.5	51.4	52.8	53.5	48.3	50.6	54.2	43	49	56.5	78	96.5	62.3	57.6		36.1
ABA39526.1 <i>L. longipalpis</i>	53.8	53.7	49.7	55.2	53.8	46.5	49.4	53.1	42.9	50.3	53.7	59.7	55.6	64.8	62.9	57.6	

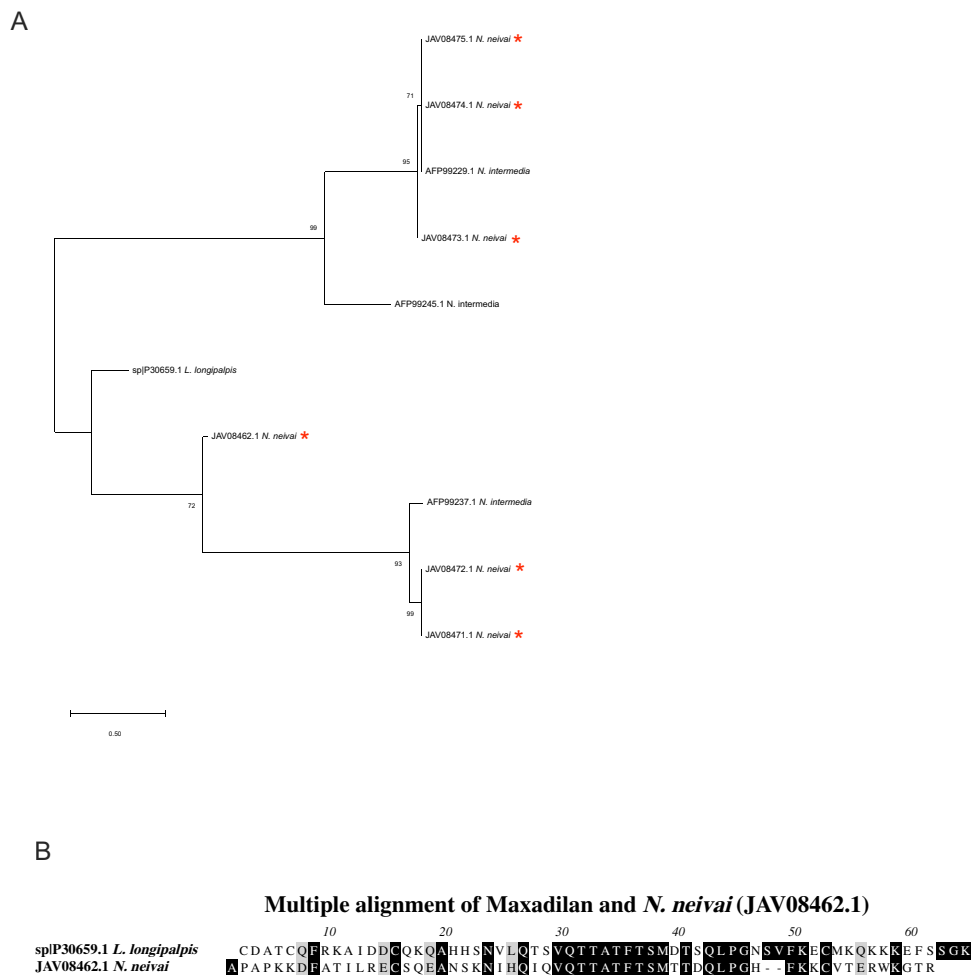
**Table 7.** Pairwise comparison matrix of identity and similarity percentages.

Protein ID	Abundance index (%)	Accession number	Best match to NCBI NR or TSA-NR databases	E-value	Identity (%)	Best match accession number	Seq size	MW	pI
JAV08475.1	22.83	GFDF01005609.1	Maxadilan related protein <i>N. intermedia</i>	1.00E-65	100.00	AFP99229.1	98	10.63	7.76
JAV08473.1	21.44	GFDF01005611.1	Maxadilan related protein <i>N. intermedia</i>	3.00E-65	98.98	AFP99229.1	98	10.66	7.76
JAV08474.1	13.14	GFDF01005610.1	Maxadilan related protein <i>N. intermedia</i>	4.00E-64	97.96	AFP99229.1	98	10.65	7.76
JAV08472.1	12.54	GFDF01005612.1	Maxadilan related protein <i>N. intermedia</i>	4.00E-40	87.32	AFP99237.1	71	8.043	7.71
JAV08471.1	7.84	GFDF01005613.1	Maxadilan related protein <i>N. intermedia</i>	3.00E-41	90.14	AFP99237.1	71	8.072	8.53
JAV08462.1	6.44	GFDF01005622.1	Maxadilan <i>L. longipalpis</i>	3.00E-07	38.75	P30659.1	79	8.901	9.45
JAV08466.1	5.91	GFDF01005618.1	Maxadilan related protein <i>N. intermedia</i>	3.00E-36	83.10	AFP99237.1	71	7.914	8.52
JAV08467.1	4.25	GFDF01005617.1	Maxadilan related protein <i>N. intermedia</i>	7.00E-40	88.73	AFP99237.1	71	7.944	7.72
JAV08469.1	2.63	GFDF01005615.1	Maxadilan related protein <i>N. intermedia</i>	8.00E-46	97.22	AFP99260.1	72	8.319	8.85
JAV08476.1	2.53	GFDF01005608.1	Maxadilan related protein <i>N. intermedia</i>	1.00E-44	97.26	AFP99260.1	73	8.411	8.85
JAV08468.1	0.44	GFDF01005616.1	Maxadilan related protein <i>N. intermedia</i>	1.00E-35	81.69	AFP99237.1	71	7.951	9.12

**Table 8.** Maxadilan secreted proteins originating from the sialotranscriptome of *Nyssomyia neivai*.

is not free of polymorphisms, its amino acid substitution rate is around 23%, with most amino acids positions not being conserved among homologues<sup>59,60</sup>. Considering *Lu. longipalpis* as the vector of VL and *N. neivai* as the vector of TL in the same endemic Brazilian regions<sup>8,61</sup>, maxadilan-like proteins could bring a new insight for a common vaccine targeting VL and TL.

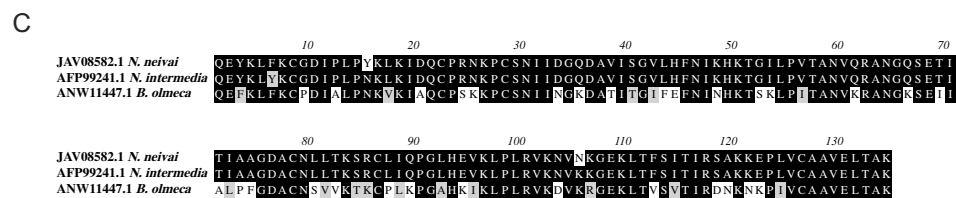
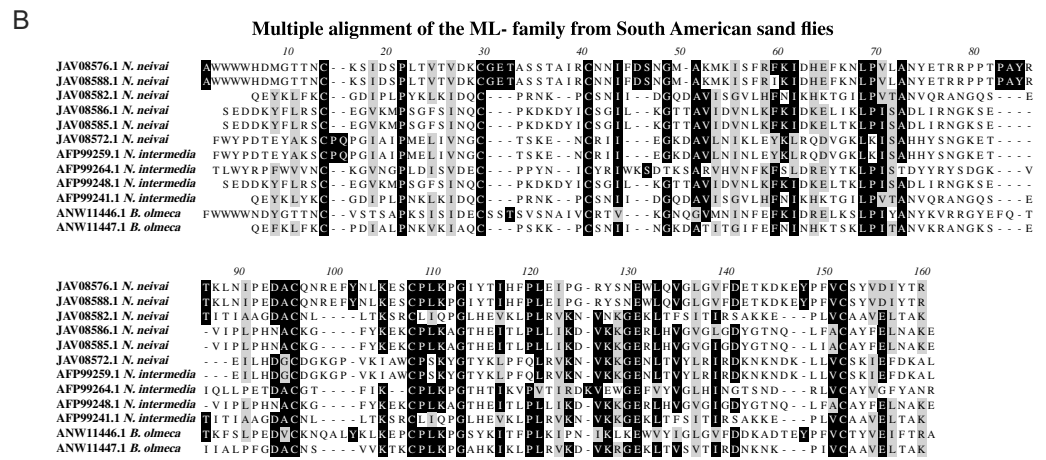
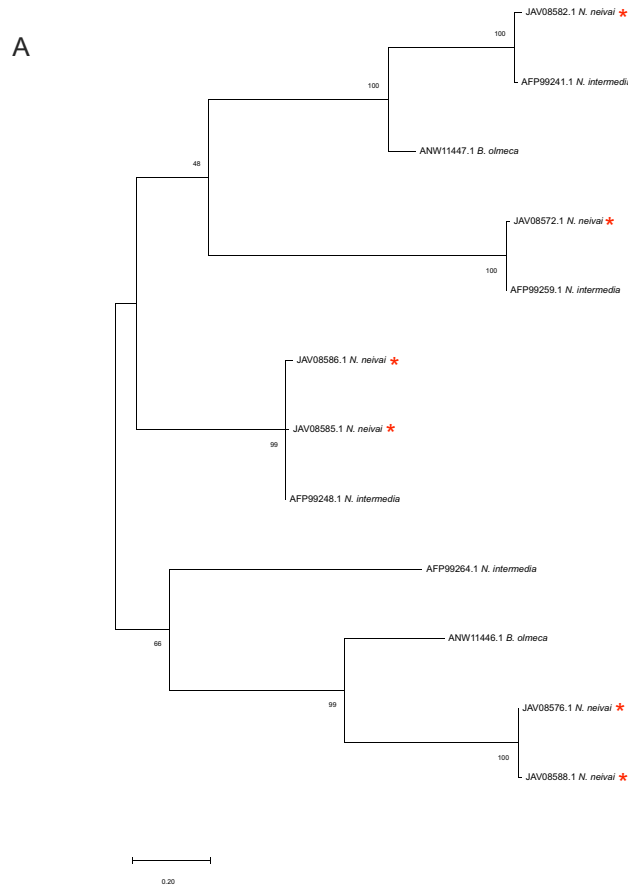
**ML domain peptide family.** We have categorized 10 full length abundant proteins as being part of the ML domain family that mapped to 5.8% of the *N. neivai* transcriptome (Table 9). We focused on the six most abundant members (JAV08576.1, JAV08588.1, JAV08582.1, JAV08586.1, JAV08585.1, JAV08572.1) for a comparative analysis representing 88.01% of this family representativity (Table 9). The MD-2-related lipid-recognition (ML)



**Figure 4.** Molecular phylogenetic analysis and sequence alignment of *Nyssomyia neivai* Maxadilan-simile protein family. **(A)** The evolutionary history was inferred by using the Maximum Likelihood method. The tree with the highest log likelihood is shown. The tree is drawn to scale, with branch lengths measured in number of substitutions per site. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7. **(B)** Multiple alignments of Maxadilan and *N. neivai* (JAV08462.1) proteins using Muscle. Black shading represents identical amino acids, light gray shading represents similar amino acids.

Protein ID	Abundance index (%)	Accession number	Best match to NCBI NR or TSA-NR databases	E-value	Identity (%)	Best match accession number	Seq size	MW	pI
JAV08576.1	19.01	GFDF01005508.1	Putative ML domain salivary peptide <i>C. appendiculata</i>	5.00E-112	87.06	JAB54888.1	170	19.90	7.71
JAV08588.1	18.94	GFDF01005496.1	Putative ML domain salivary peptide <i>C. appendiculata</i>	5.00E-112	86.55	JAB54888.1	171	20.00	7.55
JAV08582.1	16.91	GFDF01005502.1	ML domain salivary peptide <i>N. intermedia</i>	2.00E-105	98.01	AFP99241.1	151	16.60	9.51
JAV08586.1	13.53	GFDF01005498.1	ML domain salivary peptide <i>N. intermedia</i>	3.00E-109	98.71	AFP99248.1	155	17.25	8.47
JAV08585.1	13.49	GFDF01005499.1	ML domain salivary peptide <i>N. intermedia</i>	1.00E-109	99.35	AFP99248.1	155	17.20	8.47
JAV08572.1	6.22	GFDF01005512.1	ML domain salivary peptide <i>N. intermedia</i>	4.00E-112	98.72	AFP99259.1	165	18.99	9.39
JAV08554.1	3.36	GFDF01005530.1	ML domain salivary peptide <i>N. intermedia</i>	4.00E-88	100.00	AFP99264.1	124	14.41	8.99
JAV08568.1	3.25	GFDF01005516.1	LolMLc <i>B. olmea</i>	7.00E-65	57.74	ANW11448.1	165	19.16	9.45
JAV08580.1	2.80	GFDF01005504.1	ML domain salivary peptide <i>N. intermedia</i>	5.00E-114	98.73	AFP99264.1	157	18.12	8.87
JAV08567.1	2.49	GFDF01005517.1	LolMLc <i>B. olmea</i>	9.00E-65	57.74	ANW11448.1	165	19.12	9.45

**Table 9.** ML domain secreted proteins originating from the sialotranscriptome of *Nyssomyia neivai*.



◀ **Figure 5.** Molecular phylogenetic analysis and sequence alignment of ML-domain protein family of *Nyssomyia neivai*. (A) The evolutionary history was inferred by using the Maximum Likelihood method. The tree with the highest log likelihood is shown. The tree is drawn to scale, with branch lengths measured in number of substitutions per site. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7. (B) Multiple alignments of ML-domain from *Nyssomyia neivai* with South American sand flies ML-domain proteins using Muscle. Black shading represents identical amino acids, light gray shading represents similar amino acids. (C) Alignment of *Nyssomyia neivai* JAV08582.1, *N. intermedia* AFP99241.1 and *B. olmeca* ANW11447.1 from the ML-domain protein family using Muscle. Black shading represents identical amino acids, light gray shading represents similar amino acids.

Protein ID	Abundance index (%)	Accession number	Best match to NCBI NR or TSA-NR databases	E-value	Identity (%)	Best match accession number	Seq size	MW	pI
JAV07960.1	36.18	GFDF01006124.1	Hypothetical protein <i>L. ayacuchensis</i>	0	76.94	BAM69111.1	398	45.44	9.49
JAV07959.1	34.61	GFDF01006125.1	Hypothetical protein <i>L. ayacuchensis</i>	0	76.69	BAM69111.1	398	45.40	9.47
JAV07968.1	28.94	GFDF01006116.1	Yellow-related salivary protein <i>N. intermedia</i>	0	98.03	AFP99235.1	406	46.13	8.64
JAV07963.1	0.13	GFDF01006121.1	43 kDa salivary yellow-related protein SP04 <i>P. argentipes</i>	5.00E-25	24.72	ABA12136.1	408	47.19	6.39
JAV07967.1	0.05	GFDF01006117.1	Putative major royal jelly protein <i>P. kandelakii</i>	0	76.00	NBJ60408.1	447	49.99	6.89
JAV07966.1	0.05	GFDF01006118.1	Putative major royal jelly protein <i>P. kandelakii</i>	0	76.00	NBJ60408.1	460	51.49	6.83
JAV07964.1	0.05	GFDF01006120.1	Major royal jelly protein <i>P. kandelakii</i>	4.00E-33	27.17	NBJ63551.1	450	51.58	5.95
JAV07955.1	0.00	GFDF01006129.1	AAEL011699-PA <i>A. aegypti</i>	0	62.69	EAT36216.1	490	55.03	5.52

**Table 10.** Yellow-related secreted proteins originating from the sialotranscriptome of *Nyssomyia neivai*.

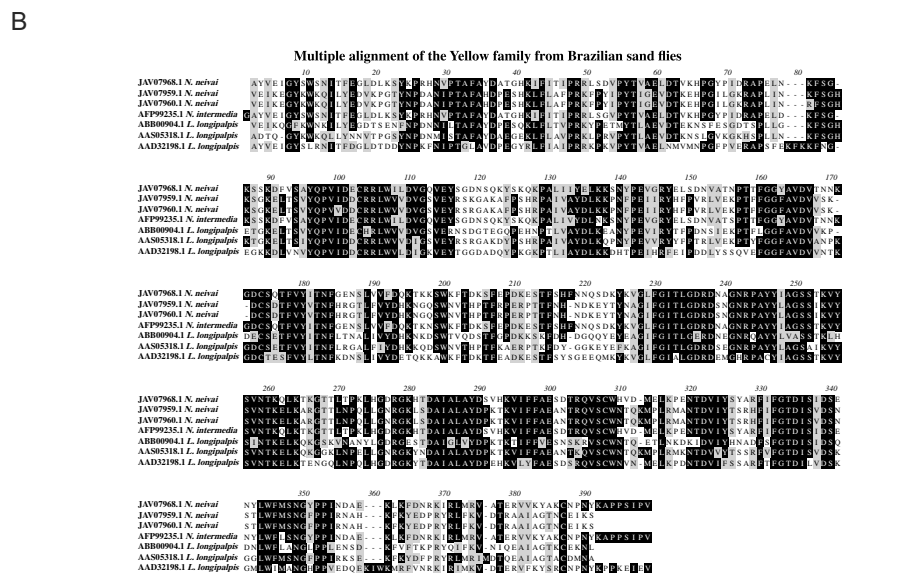
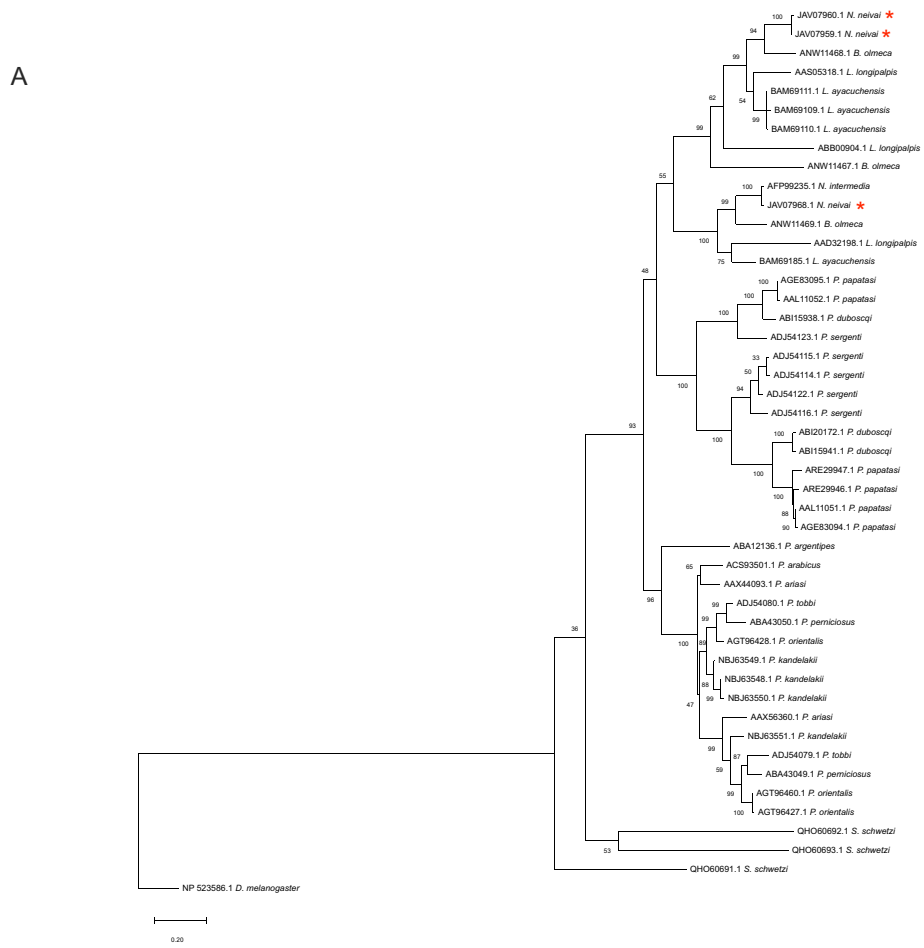
domain is implicated in lipid-mediated membrane binding mechanisms with a role in the execution and regulation of many cellular processes, including cell signaling and membrane trafficking<sup>62</sup>. The ML domain from the SMART database hints that this proteins may be involved in innate immunity or as an antagonist of lipid mediators of hemostasis and inflammation<sup>63</sup>.

This family is relatively common in tick sialomes<sup>63</sup> but it had only been described in *N. intermedia* and *B. olmeca* sialomes so far<sup>27,29</sup>. *N. neivai* phylogenetic analysis indicates the presence of 3 subfamilies with the ML domain, with several likely events of gene duplication occurring in *N. neivai* (Fig. 5A). ML domain salivary proteins present in sand flies seems to be a very divergent family with few conserved amino acids across species (Fig. 5B); however, when comparing the clustered molecules within each of the clades we start to appreciate a higher degree of conservation (Fig. 5C), for example comparing the *N. intermedia* ML family sequence AFP99241.1 and *B. olmeca* (ANW11447) that share the same subtree with *N. neivai* JAV08582.1 in Fig. 5A, we observe 97.8% and 64.9% identity (Fig. 5C), respectively, likely hinting at possibly three independent proteins families within the ML domain.

**Yellow protein family.** *N. Neivai* salivary transcriptome encompasses eight members of the yellow salivary proteins corresponding to 5.1% of the total sialome. From these eight members three of them corresponds to 99.72% of the yellow family abundance (Table 10).

Yellow-related proteins are abundantly expressed in salivary glands of phlebotomies, mainly in Old World sand flies<sup>17-20,23-25,27,43,44</sup>. The Yellow family was the most abundant salivary protein family detected, also by next generation sequencing, on *Phlebotomus kandelakii* salivary glands (accounting to 31.7% of the mRNA on salivary glands)<sup>24</sup> contrasting with the limited presence of this family in *N. neivai* (5.1%). Phlebotomine yellow-related proteins are characterized by having the major royal jelly protein domain (MRJP). Originally, MRJP proteins were described from honeybee larval jelly, making up to 90% of the protein content<sup>64</sup>. Sequences related to MRJP proteins were described in *Drosophila*, where it is related to cuticle pigmentation and, when mutated, it produced a yellow phenotype and thus named Yellow proteins<sup>65</sup>. It was later found that in Diptera they had a dopachrome oxidase function<sup>66,67</sup>.

In bloodsucking Diptera, salivary yellow-related proteins have only been described in sand flies (all sand fly species studied to date)<sup>17-20,23-25,27,43,44</sup>, and *Glossina morsitans morsitans* (Westwood 1851)<sup>68</sup>. The proteins of this family are immunogenic and host antibody responses to this protein can be a potential marker for sand fly exposure in experimentally bitten mice and dogs, as well as naturally exposed dogs, humans, and foxes<sup>3,69</sup>. *Lu. longipalpis* proteins, LJM11, LJM111 and LJM17, act as high affinity binders of pro-inflammatory biogenic amines such as serotonin, catecholamines, and histamine, suggesting that the proteins play a role for the reduction of inflammation during sand fly blood-feeding<sup>3,70</sup>, this activity has also been confirmed in salivary yellows from Old World *P. orientalis* and *P. perniciosus*<sup>71,72</sup>.



**Figure 6.** Molecular phylogenetic analysis and sequence alignment Yellow-related protein family of *Nyssomyia neivai*. (A) The evolutionary history was inferred by using the Maximum Likelihood method. The tree with the highest log likelihood is shown. The tree is drawn to scale, with branch lengths measured in number of substitutions per site. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter = 1.6114)]. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA7. (B) Multiple alignments of Yellow-related protein from *Nyssomyia neivai* with *Nyssomyia intermedia* and *Lutzomyia longipalpis* Yellow-related protein using Muscle. Black shading represents identical amino acids, light gray shading represents similar amino acids.

A combination of recombinant LJM17 and LJM11 successfully substituted *Lu. longipalpis* whole salivary gland homogenate in probing sera of individuals for vector exposure<sup>73</sup>. Yellow proteins are also under consideration for anti-*Leishmania* vector-based vaccines. LJM17 from *Lu. longipalpis* elicited leishmanicidal Th1 cytokines in immunized dogs<sup>74,75</sup>, and LJM11 protected laboratory animals against *L. (L.) infantum* (Nicolle 1908), *L. (L.) major*, and *L. (V.) braziliensis*<sup>70,76,77</sup>.

In contrast, mice immunized with *P. papatasi* yellow-related proteins PpSP42 or PpSP44 (AAL11052 and AAL11051, respectively) elicited Th2 cytokines and exacerbated *L. (L.) major* infection<sup>78</sup>. Other yellow-related proteins from *P. papatasi*, specifically PPTSP44 (AGE83095.1), induced a strong Th1 response constituting a potential vaccine candidates against leishmaniasis<sup>79</sup>. It remains to be elucidated whether the protection induced by yellow-related proteins is related to particular protein immunogenicity, to sand fly species, or to the vector-*Leishmania* host combination, as all of these factors can contribute to vaccine efficacy. New approaches using novel vaccine techniques, consisting of a single dose of plasmid, followed by two doses of recombinant Canary-poxvirus expressing *Lu. longipalpis* yellow-related salivary proteins, are a promising strategy to control *Leishmania* infection<sup>75</sup>.

Phylogenetic analysis segregated the New World sand fly yellow proteins in its own clade separated from VL and TL Old World sand fly yellow proteins (Fig. 6A). The New World sand fly yellow proteins clade branched out in two subclades, one with the presence of *Lu. longipalpis* LJM11 (AAS05318.1) and LJM111 (ABB00904.1) yellow proteins that clustered with *N. neivai* yellow-related protein JAV07960.1 and JAV07959.1 (Fig. 6A). Of note, we can observe that these two *N. neivai* yellows are closely related to the yellow from *B. olmeca* (ANW11468.1) (Fig. 6A). In the other subclade *N. neivai* yellow-related protein JAV07968.1 clustered closely with the *N. intermedia* (AFP99235.1) and with the *Lu. longipalpis* LJM17 (AAD32198.1) member (Fig. 6A). Multiple alignment of *N. neivai*, *Lu. longipalpis*, and *N. intermedia* (Fig. 6B) shows a high level of conservation among these proteins. For example, *N. neivai* JAV07960.1 and *Lu. longipalpis* LJM11 (AAS05318.1) share 73.3% identity and 84.6% similarity). In the other hand, *N. neivai* yellow-related protein JAV07968.1 share a closer relationship with *N. intermedia* Linb-21 (AFP99235.1) highlighted by its 97.7% identity compared to 62.4% identity with *Lu. longipalpis* LJM17 (AAD32198.1).

**Sand flies' salivary proteins and pemphigus foliaceus.** Sand fly salivary proteins have been associated with endemic PF pathogenesis in Brazil and Tunisia<sup>10–12,52,80–82</sup>. Screening of sera from pemphigus foliaceus patients from NSPS antibodies against *Lu. longipalpis* maxadilan<sup>11</sup> while Mato Grosso do Sul State patients reacted to *Lu. longipalpis* LJM11<sup>13</sup> and LJM17<sup>14</sup>. Similarly, Tunisian PF patients also reacted to several salivary proteins from *P. papatasi*<sup>80</sup>.

The mechanisms through which exposure to sand fly bites may induce an autoimmunity triggering production of IgG autoantibodies against Dsg1 in genetically susceptible individuals remains to be teased out. In fact, IgG antibodies against salivary homogenates from *N. neivai* correlated positively with IgG anti-Dsg1 in PF patients<sup>10</sup>. An antigenic cross-reactivity between salivary proteins and Dsg1 is the most plausible hypothesis for anti-Dsg1 autoantibodies production following sand fly bites. Nevertheless, both BLAST and PSIBLAST do not detect any highly significant homology between *N. neivai*, *Lu. longipalpis* or *P. papatasi* salivary proteins and Dsg1<sup>14</sup>. Other pathogenic mechanisms that could explain the loss of Dsg1 self-antigen tolerance induced by exposure to sand fly salivary proteins remains to be tested.

The presence of an abundant maxadilan-simile transcripts in *N. neivai* (JAV08462.1) sialome may explain the antibody reactivity to maxadilan in NSPS patients, where *N. neivai* but no *Lu. longipalpis* is present<sup>9</sup>. Further testing with a *N. neivai* recombinant maxadilan can help confirm this assumption. Moreover, testing the PF sera from NSPS against recombinant yellow proteins identified in *N. neivai* would also be desirable in our PF casuistic.

Considering that non-homologous salivary proteins from different sand fly species were associated with PF pathogenesis<sup>10–14,81,82</sup>, we may expect that not a single peptide may act as an independent antigen to trigger PF. More than one protein may be involved in the PF pathogenesis considering the shared pharmacological properties and conformational mimotopes of these proteins in distinct biting sand fly species.

## Conclusion

Leishmaniasis is still a frequent and neglected disease in Brazil. Our results add valuable data related to New World Phlebotomine salivary proteins, expanding the findings reported in *Lu. longipalpis* and *N. intermedia* sialomes. The availability of the identity of the most abundant *N. neivai* salivary proteins of the three main species of sand flies widely distributed in Brazil will bring new insights into the host-vector-parasite relationship of *L. (L.) infantum* and *L. (V.) braziliensis* infections and may point to targets of interest for a vector-based vaccine. We hope the availability of this compilation of *N. neivai* salivary proteins by abundance can inform researchers on the selection *N. neivai* candidates for future experiments. Production of distinct abundant *N. neivai* recombinant proteins can be used to test individual candidates for the etiology of PF as the trigger of anti-Dsg1 autoantibodies, and also used as biomarkers of vector exposure translating into monitoring tools for vector intervention campaigns.

## Data availability

All relevant data are within the paper.

Received: 10 April 2020; Accepted: 17 September 2020

Published online: 19 October 2020

## References

- Burza, S., Croft, S. L. & Boelaert, M. Leishmaniasis. *Lancet* **392**, 951–970. [https://doi.org/10.1016/S0140-6736\(18\)31204-2](https://doi.org/10.1016/S0140-6736(18)31204-2) (2018).
- Serafim, T. D. *et al.* Sequential blood meals promote *Leishmania* replication and reverse metacyclogenesis augmenting vector infectivity. *Nat. Microbiol.* **3**, 548–555. <https://doi.org/10.1038/s41564-018-0125-7> (2018).
- Lestonova, T., Rohousova, I., Sima, M., de Oliveira, C. I. & Volf, P. Insights into the sand fly saliva: blood-feeding and immune interactions between sand flies, hosts, and *Leishmania*. *PLoS Negl. Trop. Dis.* **11**, e0005600. <https://doi.org/10.1371/journal.pntd.0005600> (2017).
- Abdeladhim, M., Kamhawi, S. & Valenzuela, J. G. What's behind a sand fly bite? The profound effect of sand fly saliva on host hemostasis, inflammation and immunity. *Infect. Genet. Evol.* **28**, 691–703. <https://doi.org/10.1016/j.meegid.2014.07.028> (2014).
- Luz, N. F. *et al.* Saliva induces heme oxygenase-1 expression at bite sites. *Front. Immunol.* **9**, 2779. <https://doi.org/10.3389/fimmu.2018.02779> (2018).
- Ribeiro, J. M. Role of saliva in blood-feeding by arthropods. *Annu. Rev. Entomol.* **32**, 463–478 (1987).
- Anversa, L., Tiburcio, M. G. S., Richini-Pereira, V. B. & Ramirez, L. E. Human leishmaniasis in Brazil: a general review. *Rev. Assoc. Med. Bras.* **1992**(64), 281–289. <https://doi.org/10.1590/1806-9282.64.03.281> (2018).
- AndradeFilho, J. D., Galati, E. A. & Falcão, A. L. Nyssomyia intermedia (Lutz & Neiva, 1912) and Nyssomyia neivai (Pinto, 1926) (Diptera: Psychodidae: Phlebotominae) geographical distribution and epidemiological importance. *Mem. Inst. Oswaldo Cruz* **102**, 481–487 (2007).
- Brochado, M. J. *et al.* Differential HLA class I and class II associations in pemphigus foliaceus and pemphigus vulgaris patients from a prevalent Southeastern Brazilian region. *J. Autoimmun.* <https://doi.org/10.1016/j.jaut.2016.04.007> (2016).
- Vernal, S. *et al.* Insights into the epidemiological link between biting flies and pemphigus foliaceus in southeastern Brazil. *Acta Trop.* **176**, 455–462. <https://doi.org/10.1016/j.actatropica.2017.09.015> (2017).
- Roselino, A. M., Figueiredo, J. F. C., Kouniga, K., Reddy, V. & Lerner, E. A. Serum IgG from pemphigus foliaceus patients reacts against maxadilan. *J. Invest. Dermatol.* **117**, 460 (2001).
- Vernal, S., De Paula, N. A., Bollela, V. R., Lerner, E. A. & Roselino, A. M. Pemphigus foliaceus and sand fly bites: assessing the humoral immune response to the salivary proteins maxadilan and LJM11. *Br. J. Dermatol.* <https://doi.org/10.1111/bjd.19221> (2020).
- Qian, Y. *et al.* Non-infectious environmental antigens as a trigger for the initiation of an autoimmune skin disease. *Autoimmun. Rev.* <https://doi.org/10.1016/j.autrev.2016.07.005> (2016).
- Diaz, L. A. *et al.* A lutzomyia longipalpis salivary protein induces cross-reactive antibodies to pemphigus autoantigen desmoglein 1. *J. Invest. Dermatol.* <https://doi.org/10.1016/j.jid.2020.02.041> (2020).
- Casanova, C., Colla-Jacques, F. E., Hamilton, J. G., Brazil, R. P. & Shaw, J. J. Distribution of *Lutzomyia longipalpis* chemotype populations in Sao Paulo state, Brazil. *PLoS Negl. Trop. Dis.* **9**, e0003620. <https://doi.org/10.1371/journal.pntd.0003620> (2015).
- Abdeladhim, M. *et al.* Updating the salivary gland transcriptome of *Phlebotomus papatasi* (Tunisian strain): the search for sand fly-secreted immunogenic proteins for humans. *PLoS ONE* **7**, e47347. <https://doi.org/10.1371/journal.pone.0047347> (2012).
- Rohousova, I. *et al.* Salivary gland transcriptomes and proteomes of *Phlebotomus tobbi* and *Phlebotomus sergenti*, vectors of leishmaniasis. *PLoS Negl. Trop. Dis.* **6**, e1660. <https://doi.org/10.1371/journal.pntd.0001660> (2012).
- Hostomska, J. *et al.* Analysis of salivary transcripts and antigens of the sand fly *Phlebotomus arabicus*. *BMC genomics* **10**, 282. <https://doi.org/10.1186/1471-2164-10-282> (2009).
- Anderson, J. M. *et al.* Comparative salivary gland transcriptomics of sandfly vectors of visceral leishmaniasis. *BMC Genom.* **7**, 52. <https://doi.org/10.1186/1471-2164-7-52> (2006).
- Oliveira, F. *et al.* From transcriptome to immunome: identification of DTH inducing proteins from a *Phlebotomus ariasi* salivary gland cDNA library. *Vaccine* **24**, 374–390. <https://doi.org/10.1016/j.vaccine.2005.07.085> (2006).
- Martin-Martin, I., Molina, R. & Jiménez, M. Identifying salivary antigens of *Phlebotomus argentipes* by a 2DE approach. *Acta Trop* **126**, 229–239. <https://doi.org/10.1016/j.actatropica.2013.02.008> (2013).
- Vlkova, M. *et al.* Comparative analysis of salivary gland transcriptomes of *Phlebotomus orientalis* sand flies from endemic and non-endemic foci of visceral leishmaniasis. *PLoS Negl. Trop. Dis.* **8**, e2709. <https://doi.org/10.1371/journal.pntd.0002709> (2014).
- Kato, H. *et al.* High degree of conservancy among secreted salivary gland proteins from two geographically distant *Phlebotomus duboscqi* sandflies populations (Mali and Kenya). *BMC Genom.* **7**, 226. <https://doi.org/10.1186/1471-2164-7-226> (2006).
- Oliveira, F. *et al.* Immunity to vector saliva is compromised by short sand fly seasons in endemic regions with temperate climates. *Scientific reports* **10**, 7990. <https://doi.org/10.1038/s41598-020-64820-9> (2020).
- Polanska, N. *et al.* *Sergentomyia schwetzi*: Salivary gland transcriptome, proteome and enzymatic activities in two lineages adapted to different blood sources. *PLoS ONE* **15**, e0230537. <https://doi.org/10.1371/journal.pone.0230537> (2020).
- Kato, H. *et al.* Analysis of salivary gland transcripts of the sand fly *Lutzomyia ayacuchensis*, a vector of Andean-type cutaneous leishmaniasis. *Infect. Genet. Evol.* **13**, 56–66. <https://doi.org/10.1016/j.meegid.2012.08.024> (2013).
- de Moura, T. R. *et al.* Functional transcriptomics of wild-caught *Lutzomyia intermedia* salivary glands: identification of a protective salivary protein against *Leishmania braziliensis* infection. *PLoS Negl. Trop. Dis.* **7**, 2242. <https://doi.org/10.1371/journal.pntd.0002242> (2013).
- Valenzuela, J. G., Garfield, M., Rowton, E. D. & Pham, V. M. Identification of the most abundant secreted proteins from the salivary glands of the sand fly *Lutzomyia longipalpis*, vector of *Leishmania chagasi*. *J. Exp. Biol.* **207**, 3717–3729. <https://doi.org/10.1242/jeb.01185> (2004).
- Abdeladhim, M. *et al.* Molecular diversity between salivary proteins from new world and old world sand flies with emphasis on bichromomyia olmeca, the sand fly vector of *Leishmania mexicana* in mesoamerica. *PLoS Negl. Trop. Dis.* **10**, e0004771. <https://doi.org/10.1371/journal.pntd.0004771> (2016).
- Goulart, T. M. *et al.* Breeding protocol for the sand fly *Nyssomyia neivai* (Diptera: Psychodidae) in laboratory conditions. *Acta Trop.* **174**, 102–105. <https://doi.org/10.1016/j.actatropica.2017.07.010> (2017).
- AndradeFilho, J. D., Galati, E. A. & Falcao, A. L. Redescription of *Nyssomyia intermedia* (Lutz & Neiva, 1912) and *Nyssomyia neivai* (Pinto, 1926) (Diptera: Psychodidae). *Mem. Inst. Oswaldo Cruz* **98**, 1059–1065. <https://doi.org/10.1590/s0074-02762003000800015> (2003).
- Araujo, R. N. *et al.* RNA-seq analysis of the salivary glands and midgut of the Argasid tick *Ornithodoros rostratus*. *Sci. Rep.* **9**, 6764. <https://doi.org/10.1038/s41598-019-42899-z> (2019).
- Simpson, J. T. *et al.* ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123. <https://doi.org/10.1101/gr.089532.108> (2009).
- Biol, I. *et al.* De novo transcriptome assembly with ABySS. *Bioinformatics* **25**, 2872–2877. <https://doi.org/10.1093/bioinformatics/btp367> (2009).
- Luo, R. *et al.* Erratum: SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **4**, 30. <https://doi.org/10.1186/s13742-015-0069-2> (2015).
- Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877. <https://doi.org/10.1101/gr.9.9.868> (1999).
- Karim, S., Singh, P. & Ribeiro, J. M. A deep insight into the salivary transcriptome of the gulf coast tick, *Amblyomma maculatum*. *PLoS ONE* **6**, e28525. <https://doi.org/10.1371/journal.pone.0028525> (2011).



38. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) (1990).
39. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628. <https://doi.org/10.1038/nmeth.1226> (2008).
40. Almagro-Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423. <https://doi.org/10.1038/s41587-019-0036-z> (2019).
41. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **5**, 113. <https://doi.org/10.1186/1471-2105-5-113> (2004).
42. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. <https://doi.org/10.1093/molbev/msw054> (2016).
43. Valenzuela, J. G. *et al.* Toward a defined anti-Leishmania vaccine targeting vector antigens: characterization of a protective salivary protein. *J. Exp. Med.* **194**, 331–342 (2001).
44. Coutinho-Abreu, I. V. & Valenzuela, J. G. Comparative evolution of sand fly salivary protein families and implications for biomarkers of vector exposure and salivary vaccine candidates. *Front. Cell. Infect. Microbiol.* **8**, 290. <https://doi.org/10.3389/fcimb.2018.00290> (2018).
45. Charlab, R., Valenzuela, J. G., Rowton, E. D. & Ribeiro, J. M. Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis*. *Proc. Natl. Acad. Sci. USA* **96**, 15155–15160 (1999).
46. Alvarenga, P. H. *et al.* Novel family of insect salivary inhibitors blocks contact pathway activation by binding to polyphosphate, heparin, and dextran sulfate. *Arterioscler. Thromb. Vasc. Biol.* **33**, 2759–2770. <https://doi.org/10.1161/atvbaha.113.302482> (2013).
47. Gholami, E. *et al.* DNA plasmid coding for Phlebotomus sergenti salivary protein PsSP9, a member of the SP15 family of proteins, protects against *Leishmania tropica*. *PLoS Negl. Trop. Dis.* **13**, e0007067. <https://doi.org/10.1371/journal.pntd.0007067> (2019).
48. Niewiarowski, S., McLane, M. A., Kloczewiak, M. & Stewart, G. J. Disintegrins and other naturally occurring antagonists of platelet fibrinogen receptors. *Semin. Hematol.* **31**, 289–300 (1994).
49. McLane, M. A., Kowalska, M. A., Silver, L., Shattil, S. J. & Niewiarowski, S. Interaction of disintegrins with the alpha IIb beta 3 receptor on resting and activated human platelets. *Biochem. J.* **301**(Pt 2), 429–436. <https://doi.org/10.1042/bj3010429> (1994).
50. Weis, W. I., Taylor, M. E. & Drickamer, K. The C-type lectin superfamily in the immune system. *Immunol. Rev.* **163**, 19–34 (1998).
51. Lerner, E. A., Ribeiro, J. M., Nelson, R. J. & Lerner, M. R. Isolation of maxadilan, a potent vasodilatory peptide from the salivary glands of the sand fly *Lutzomyia longipalpis*. *J. Biol. Chem.* **266**, 11234–11236 (1991).
52. Aires, J. *et al.* Maxadilan-simile expression in Nyssomyia neivai, a sandfly vector in an endemic region of Brazil, and its immunogenicity in patients with American tegumentary leishmaniasis. *Mem. Inst. Oswaldo Cruz* **112**, 116–122. <https://doi.org/10.1590/0074-02760160351> (2017).
53. Ribeiro, J. M., Katz, O., Pannell, L. K., Waitumbi, J. & Warburg, A. Salivary glands of the sand fly *Phlebotomus papatasi* contain pharmacologically active amounts of adenosine and 5'-AMP. *J. Exp. Biol.* **202**, 1551–1559 (1999).
54. Katz, O., Waitumbi, J. N., Zer, R. & Warburg, A. Adenosine, AMP, and protein phosphatase activity in sandfly saliva. *Am. J. Trop. Med. Hyg.* **62**, 145–150 (2000).
55. Morris, R. V., Shoemaker, C. B., David, J. R., Lanzaro, G. C. & Titus, R. G. Sandfly maxadilan exacerbates infection with *Leishmania major* and vaccinating against it protects against *L. major* infection. *J. Immunol.* **167**, 5226–5230 (2001).
56. Brodie, T. M., Smith, M. C., Morris, R. V. & Titus, R. G. Immunomodulatory effects of the *Lutzomyia longipalpis* salivary gland protein maxadilan on mouse macrophages. *Infect. Immun.* **75**, 2359–2365. <https://doi.org/10.1128/IAI.01812-06> (2007).
57. Titus, R. G., Bishop, J. V. & Mejia, J. S. The immunomodulatory factors of arthropod saliva and the potential for these factors to serve as vaccine targets to prevent pathogen transmission. *Parasite Immunol.* **28**, 131–141 (2006).
58. Milleron, R. S., Ribeiro, J. M., Elnaime, D., Soong, L. & Lanzaro, G. C. Negative effect of antibodies against maxadilan on the fitness of the sand fly vector of American visceral leishmaniasis. *Am. J. Trop. Med. Hyg.* **70**, 278–285 (2004).
59. Lanzaro, G. C. *et al.* Variation in the salivary peptide, maxadilan, from species in the *Lutzomyia longipalpis* complex. *Insect. Mol. Biol.* **8**, 267–275 (1999).
60. Warburg, A., Saraiva, E., Lanzaro, G. C., Titus, R. G. & Neva, F. Saliva of *Lutzomyia longipalpis* sibling species differs in its composition and capacity to enhance leishmaniasis. *Philos. Trans. R. Soc. Lond. B.* **345**, 223–230 (1994).
61. Casanova, C., Colla-Jacques, F. E., Hamilton, J. G., Brazil, R. P. & Shaw, J. J. Distribution of *Lutzomyia longipalpis* chemotype populations in São Paulo state, Brazil. *PLoS Negl. Trop. Dis.* **9**, e0003620. <https://doi.org/10.1371/journal.pntd.0003620> (2015).
62. Stahelin, R. V. Lipid binding domains: more than simple lipid effectors. *J. Lipid Res.* **50**(Suppl), S299–304. <https://doi.org/10.1194/jlr.R800078-JLR200> (2009).
63. Francischetti, I. M., Sa-Nunes, A., Mans, B. J., Santos, I. M. & Ribeiro, J. M. The role of saliva in tick feeding. *Front. Biosci.* **14**, 2051–2088 (2009).
64. Schmitzova, J. *et al.* A family of major royal jelly proteins of the honeybee *Apis mellifera* L. *Cell Mol. Life Sci.* **54**, 1020–1030 (1998).
65. Albert, S., Bhattacharya, D., Klaudiny, J., Schmitzova, J. & Simuth, J. The family of major royal jelly proteins and its evolution. *J. Mol. Evol.* **49**, 290–297 (1999).
66. Han, Q. *et al.* Identification of *Drosophila melanogaster* yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes. *Biochem. J.* **368**, 333–340 (2002).
67. Johnson, J. K., Li, J. & Christensen, B. M. Cloning and characterization of a dopachrome conversion enzyme from the yellow fever mosquito, *Aedes aegypti*. *Insect Biochem. Mol. Biol.* **31**, 1125–1135 (2001).
68. Alves-Silva, J. *et al.* An insight into the sialome of *Glossina morsitans morsitans*. *BMC Genom.* **11**, 213. <https://doi.org/10.1186/1471-2164-11-213> (2010).
69. Teixeira, C. *et al.* Discovery of markers of exposure specific to bites of *Lutzomyia longipalpis*, the vector of *Leishmania infantum* chagasi in Latin America. *PLoS Negl. Trop. Dis.* **4**, e638. <https://doi.org/10.1371/journal.pntd.0000638> (2010).
70. Xu, X. *et al.* Structure and function of a “yellow” protein from saliva of the sand fly *Lutzomyia longipalpis* that confers protective immunity against *Leishmania major* infection. *J. Biol. Chem.* **286**, 32383–32393. <https://doi.org/10.1074/jbc.M111.268904> (2011).
71. Spitzova, T. *et al.* Interactions between host biogenic amines and sand fly salivary yellow-related proteins. *Parasit. Vectors* **13**, 237. <https://doi.org/10.1186/s13071-020-04105-2> (2020).
72. Sumova, P. *et al.* Amine-binding properties of salivary yellow-related proteins in phlebotomine sand flies. *Insect Biochem. Mol. Biol.* **115**, 103245. <https://doi.org/10.1016/j.ibmb.2019.103245> (2019).
73. Souza, A. P. *et al.* Using recombinant proteins from *Lutzomyia longipalpis* saliva to estimate human vector exposure in visceral leishmaniasis endemic areas. *PLoS Negl. Trop. Dis.* **4**, e649. <https://doi.org/10.1371/journal.pntd.0000649> (2010).
74. Collin, N. *et al.* Sand fly salivary proteins induce strong cellular immunity in a natural reservoir of visceral leishmaniasis with adverse consequences for *Leishmania*. *PLoS Pathog.* **5**, e1000441 (2009).
75. Abbehussen, M. M. C. *et al.* Immunization of experimental dogs with salivary proteins from. *Front. Immunol.* **9**, 2558. <https://doi.org/10.3389/fimmu.2018.02558> (2018).
76. Gomes, R. *et al.* Immunity to a salivary protein of a sand fly vector protects against the fatal outcome of visceral leishmaniasis in a hamster model. *Proc. Natl. Acad. Sci. USA* **105**, 7845–7850. <https://doi.org/10.1073/pnas.0712153105> (2008).
77. Cunha, J. M. *et al.* Immunization with LJM11 salivary protein protects against infection with *Leishmania braziliensis* in the presence of *Lutzomyia longipalpis* saliva. *Acta Trop.* **177**, 164–170. <https://doi.org/10.1016/j.actatropica.2017.10.009> (2018).

78. Oliveira, F., Lawyer, P. G., Kamhawi, S. & Valenzuela, J. G. Immunity to distinct sand fly salivary proteins primes the anti-Leishmania immune response towards protection or exacerbation of disease. *PLoS Negl. Trop. Dis.* **2**, e226. <https://doi.org/10.1371/journal.pntd.0000226> (2008).
79. Tlili, A. *et al.* Phlebotomus papatasi yellow-related and apyrase salivary proteins are candidates for vaccination against human cutaneous leishmaniasis. *J. Invest. Dermatol.* **138**, 598–606. <https://doi.org/10.1016/j.jid.2017.09.043> (2018).
80. Zarea, I. *et al.* Exposure to Phlebotomus papatasi and/or Leishmania major: possible etiologic link to Tunisian pemphigus. *J. Invest. Dermatol.* **132**, 479–482. <https://doi.org/10.1038/jid.2011.291> (2012).
81. Qian, Y. *et al.* IgE anti-LJM11 sand fly salivary antigen may herald the onset of fogo selvagem in endemic Brazilian regions. *J. Invest. Dermatol.* **135**, 913–915. <https://doi.org/10.1038/jid.2014.430> (2015).
82. Qian, Y. *et al.* Cutting Edge: Brazilian pemphigus foliaceus anti-desmoglein 1 autoantibodies cross-react with sand fly salivary LJM11 antigen. *J. Immunol.* **189**, 1535–1539. <https://doi.org/10.4049/jimmunol.1200842> (2012).

## Acknowledgements

The authors are very grateful to the laboratorial staff of the Department of Biological Sciences, São Paulo State University Julio de Mesquita Filho, Araraquara, who assisted the maintenance of the sand flies samples, as well as the laboratorial staff of the North Carolina State Genomic Sciences Laboratory, Raleigh for RNA extraction, Illumina library preparation, and sequencing of the salivary samples. The authors thank Brian Brown, NIH Library Writing Center, for manuscript editing assistance. This work utilized the computational resources of the NIH HPC Biowulf cluster (<https://hpc.nih.gov>).

## Author contributions

S.V. and F.O. contributed with data collection, data analysis, data interpretation, literature search and writing, W.O.C.O. and T.M.G. contributed with samples collections, samples maintenance, salivary glands extraction and transportation, and writing, E.C. and J.O. contributed with data analysis, data interpretation and writing. M.C.P., A.M.R. and J.M.C.R. contributed with study design, data interpretation, and writing.

## Funding

This research was supported in part by the Intramural Research Program of the NIH, National Institute of Allergy and Infectious Diseases. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This project was also partially supported by FAPESP (Fundação de Apoio à Pesquisa do Estado de São Paulo #2010/51729-2). SV received a PhD scholarship from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to F.O.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2020