

## Research Article

**Cite this article:** Iwama R, Ocegüera-Figueroa A, Giribet G, Kvist S (2019). The salivary transcriptome of *Limnobdella mexicana* (Annelida: Clitellata: Praobdellidae) and orthology determination of major leech anticoagulants. *Parasitology* **146**, 1338–1346. <https://doi.org/10.1017/S0031182019000593>

Received: 1 March 2019

Revised: 20 April 2019

Accepted: 22 April 2019

First published online: 27 June 2019

**Key words:**

Anticoagulants; antistasin; bloodfeeding; hirudin; Praobdellidae; transcriptomics

**Author for correspondence:**

Rafael Iwama,

E-mail: [rafael.iwama@mail.utoronto.ca](mailto:rafael.iwama@mail.utoronto.ca)

# The salivary transcriptome of *Limnobdella mexicana* (Annelida: Clitellata: Praobdellidae) and orthology determination of major leech anticoagulants

Rafael Iwama<sup>1,2</sup>, Alejandro Ocegüera-Figueroa<sup>3</sup>, Gonzalo Giribet<sup>4</sup> and Sebastian Kvist<sup>1,2</sup>

<sup>1</sup>Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto, ON M5S 2C6, Canada; <sup>2</sup>Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 2B4, Canada; <sup>3</sup>Laboratorio de Helmintología, Instituto de Biología, Universidad Nacional Autónoma de México, Circuito Zona Deportiva 53, Ciudad Universitaria, 04510 Coyoacan Ciudad de México, México and <sup>4</sup>Museum of Comparative Zoology & Department of Organismic and Evolutionary Biology, Harvard University, Cambridge MA, 02138, USA

**Abstract**

Bloodfeeding requires several adaptations that allow the parasite to feed efficiently. Leeches and other hematophagous animals have developed different mechanisms to inhibit hemostasis, one of the main barriers imposed by their hosts. *Limnobdella mexicana* is a member of the leech family Praobdellidae, a family of host generalists known for their preference to attach on mucosal membranes of mammals, such as those in nasopharyngeal cavities, bladders and ocular orbits. Previous studies have hypothesized a positive relationship between diversity of anticoagulants and diversity of hosts in bloodfeeding leeches. However, orthology determination of putative anticoagulants and the lack of standardization of sequencing effort and method hinder comparisons between publicly available transcriptomes generated in different laboratories. In the present study, we examine the first transcriptome of a praobdellid leech and identify 15 putative anticoagulants using a phylogeny-based inference approach, amino-acid conservation, Pfam domains and BLAST searches. Our phylogenetic analyses suggest that the ancestral leech was able to inhibit factor Xa and that some hirudins that have been reported in previous studies on leech anticoagulants may not be orthologous with the archetypal hirudin.

**Introduction**

Bloodfeeding requires several adaptations to overcome biotic barriers imposed by hosts and to enable efficient feeding. Such adaptations include morphological, behavioral and molecular features that allow the parasites to create an incision and inhibit hemostasis, inflammation and other responses from the host immune system (Ribeiro and Francischetti, 2003; Ribeiro and Arcà, 2009; Min *et al.*, 2010). In this regard, leeches have attracted attention, especially due to their intense use in medicine in the 18<sup>th</sup> and 19<sup>th</sup> centuries and the potent anticoagulants isolated from salivary secretions of different species, especially from the European medicinal leech, *Hirudo medicinalis* Linnaeus, 1758 (Elliott and Kutschera, 2011).

Traditionally, leeches were subdivided into two orders based on the morphology of their feeding apparatus. Members of Rhynchobdellida are proboscis-bearing and insert their proboscis subcutaneously in order to feed on blood. By contrast, arhynchobdellid leeches are non-proboscis-bearing and the bloodfeeding taxa within this group use denticle-bearing jaws to create wounds from which they feed (Sawyer, 1986). Studies based on molecular data have suggested that Rhynchobdellida is paraphyletic (e.g. Apakupakul *et al.*, 1999) and, as a result, Tessler *et al.* (2018a) erected five orders to accommodate known leech diversity (Oceanobdelliformes, Glossiphoniformes, Erpobdelliformes, Americobdelliformes and Hirudiniformes). Additionally, phylogenetic studies have contributed to clarifying the evolution of bloodfeeding in leeches. Siddall and Burren (1995) were first to propose bloodfeeding as the behavior of the last common ancestor of all leeches, with subsequent independent transformations to macrophagy within Hirudinida. This notion is also supported by subsequent phylogenetic analyses (Apakupakul *et al.*, 1999; Borda and Siddall, 2004; Tessler *et al.*, 2018a).

Salivary transcriptomic data and the presence of putative anticoagulants in both non-proboscis and proboscis-bearing leeches have been used as evidence for an ancestral bloodfeeding state in Hirudinida (see Siddall *et al.*, 2016). However, some of these studies have focused on BLAST determination for the majority of the putative anticoagulants found, instead of tree-based orthology determination, as in other studies on non-proboscis-bearing leeches (Kvist *et al.*, 2013, 2014, 2017; Müller *et al.*, 2017). The lack of a proper phylogenetic hypothesis of the major leech anticoagulants questions previous conclusions that the presence of such molecules across the leech tree is evidence of bloodfeeding behavior in the last

common ancestor. Therefore, an analysis including putative anticoagulants of a wide phylogenetic diversity of leeches is needed to determine which molecules are orthologous to proteins that were present in the last common ancestor of leeches.

*Limnobdella mexicana* Blanchard, 1893 is a member of Praobdellidae, a family of jawed leeches. Although little is known about the host diversity and bloodfeeding behavior of this species, members of this family are known to cause internal hirudiniasis – leech infestations in orifices such as the buccal mucosa, nasal cavity and vesicles – and most have been found to feed indiscriminately on mammals (including humans) and birds (Oka, 1934; Phillips *et al.*, 2010; Siddall *et al.*, 2013); and recently a member of this family has been recorded as feeding on crustaceans (Nakano *et al.*, 2017). Previous leech salivary transcriptomic analyses have reported a positive relationship between host-generality and diversity of anticoagulants (Kvist *et al.*, 2014). Thus, it is predicted that leeches with higher levels of host-generality, such as that recorded for members of the family Praobdellidae, possess a proportionately higher diversity of anticoagulants. Despite the medical and veterinary importance of praobdellid leeches, there are no studies that describe the anticoagulant repertoire in members of this family. To remedy this, herein we explore the diversity of anticoagulants and other bioactive proteins secreted by the salivary glands of *L. mexicana*.

## Methods

### Specimen collection and identification

Leech specimens were collected in Río Pilón, Rayones, Nuevo León, Mexico (25°01'31.39"N, 99°59'38.7"W) on 11–12 July 2015. Salivary gland tissue was dissected from a single specimen using sterile tools while the leech was submerged in RNAlater (Thermo Fisher Scientific). Tissue masses were stored in RNAlater at –20 °C until RNA extraction. The specimen was identified based on morphology using specialized literature (Sawyer, 1986).

### RNA extraction, cDNA library construction and sequencing

Total RNA was extracted using a standard TRIzol-based (Life sciences) protocol, as described below. Salivary tissue mass was frozen in liquid nitrogen, ground to powder with a drill and homogenized in 500 µL of TRIzol and the sample was incubated for 5 min at room temperature (RT). Then, 100 µL of bromochloropropane was added and the sample was mixed by vortexing. After incubation for 10 min at RT, the sample was centrifuged at 16 000 rpm for 15 min at 4 °C. The aqueous layer was recovered, mixed with 500 µL of isopropanol and incubated overnight at –20 °C. The sample was then centrifuged for 15 min at 16 000 rpm at 4 °C to allow the formation of an RNA pellet, which was then washed twice with 1 mL of 75% isopropanol and centrifuged at 16 000 rpm at 4 °C for 15 and 5 min, respectively, in the first and second washing step, air dried and eluted in 100 µL of RNA Storage solution (Ambion, Austin, TX, USA). Dynabeads mRNA purification kit (Invitrogen, Carlsbad, CA, USA) was used following the manufacturer's protocol for total RNA purification. After incubation of total RNA at 65 °C for 5 min, the sample was incubated on a rocker for 30 min with 200 µL of magnetic beads and then washed twice with washing buffer. This step was repeated after an incubation for 5 min. Thereafter, mRNA was eluted in 15 µL of Tris-HCL buffer. Quality and quantity of mRNA as measured with a pico RNA assay in an Agilent 21 000 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and quantity was also measured using a RNA assay on a Qubit fluorometer (Life Technologies, Gaithersburg, MD, USA).

The cDNA library was constructed on the Apollo 324 automated system using the PrepX mRNA kit (Wafergen, Fremont, CA, USA) following the manufacturer's protocol. Full-length cDNAs were fragmented to a mean insert size of 500 bp. The sample was inline barcoded with Illumina indices to allow for pooling with other libraries prior to sequencing. Library quality and size selection were checked on the Agilent 2100 Bioanalyzer (Agilent Technologies) with the HS DNA assay. Sequencing was performed in 1/6th of a lane on the Illumina HiSeq 2500 platform (paired end, 150 bp) at the FAS Center for Systems Biology at Harvard University, Cambridge, MA, USA.

### Data sanitation, assembly and gene prediction

Before assembly, raw data were filtered and trimmed with Trim Galore! ver. 0.5 ([http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore)) applying a PHRED score of 30. Assembly was conducted with Trinity on Galaxy ver. 2.2.0 with pair distance set to 800. Open Reading Frames (ORF) among the contigs were predicted with Transdecoder ver. 5.2.0.

### Annotation

Nucleotide and amino acid ORFs were BLASTed against a local anticoagulant database adapted from Kvist *et al.* (2014, 2017) and Tessler *et al.* (2018b) and the Swiss-Prot database, using BLASTp and BLASTx with a cut-off *e*-value of  $1 \times 10^{-10}$ . In order to identify protein domains within the predicted ORFs, we searched against the Pfam database (Finn *et al.*, 2016) using HMMER ver. 3.2.1 (Eddy, 2011). In addition, SignalP ver. 4.1 (Petersen *et al.*, 2011), with the sensitive D-cutoff value, was used to determine putative signal peptides and to select ORFs with higher *e*-values when compared to hits for the same ORF against the local anticoagulant database. This was done in order to ensure that hits against anticoagulants in our local database were better than any potential hit against a non-anticoagulant in any other database. The remaining ORFs were manually examined and significant hits with a higher *e*-value against Swiss-Prot were still reported when the hits against both databases belonged to the same protein family. ORFs were then aligned with putative anticoagulant orthologues for further phylogenetic analyses. All alignments were performed using the web version of MAFFT ver. 7 (Katoh and Standley, 2013) with automatic assignment of alignment strategy, the BLOSUM62 scoring matrix and gap opening penalty of 5.00. Gene trees were obtained employing the parsimony criterion in TNT ver. 1.5 (Goloboff and Catalano, 2016) with 1000 replicates, each with 5 rounds of ratcheting, drifting and tree fusing. Trees were left unrooted due to the lack of obvious outgroups for leech anticoagulants. Putative anticoagulants were then aligned with their respective archetypal anticoagulant sequence and visualized with JalView ver. 2 (Waterhouse *et al.*, 2009).

## Results

### Sequencing, assembly and BLAST searches

In total, 11 029 677 raw sequence reads were produced by the Illumina run and 65 833 contigs were assembled with Trinity – the data are deposited in the Short Read Archive at NCBI under accession number PRJNA 532 931. Transdecoder predicted a total of 1,019 ORFs with significant matches (*e*-value of  $1 \times 10^{-10}$  or below) to known anticoagulants and other salivary proteins from leeches, as well as other bloodfeeding organisms. A total of 137 hits were recovered against 15 widely known anticoagulation factors from different inhibitor families, suggesting the presence of homologues to several anticoagulants. Putative

orthologues to thrombin inhibitors included hirudin and hirustasin from *Hirudo medicinalis* Linnaeus, 1758, bufrudin from *Hirudinaria manillensis* (Lesson, 1842), and an antithrombin from the king cobra *Ophiophagus hannah* (Cantor, 1836). Further, 27 hits against the factor Xa inhibitor antistasin from *Haementeria officinalis* de Phillipi, 1849 were recovered, 6 hits against therostasin from *Theromyzon tessulatum* (Müller, 1774) and 4 hits against ghilanten from *Haementeria ghilianii* de Phillipi, 1849. Fully 35 hits against inhibitors of platelet aggregation were recovered, including destabilase, saratin, bdellin-B, and disintegrins and metalloproteases with thrombospondin motifs (ADAM[TS]), as well as peptides inhibiting other pathways, such as leech derived tryptase inhibitor (LDTI). Several ORFs matched non-leech putative anticoagulants, including Kazal-type serpins and 5'-apyrase. Many putative anticoagulants possessed signal peptides at the 5'-end, suggesting their secretion out of the salivary cells – these included manillase, destabilase, antistasin, bdellin, LDTI, 5'-apyrase and therostasin. Table 1 lists all of the recovered anticoagulants and other bioactive salivary proteins related to bloodfeeding, together with their respective Pfam domains and indication of presence or absence of signal peptides.

### Gene trees

We describe the unrooted tree topologies using the terminology proposed by Wilkinson *et al.* (2007), in which a 'clan' in an unrooted tree is equivalent to a 'monophyletic group' in a rooted tree and 'adjacent group' is equivalent to 'sister group'. Gene trees for multiple sequence alignments of putative antistasin-like factor Xa inhibitors, putative hirudin-like thrombin inhibitors and hirudin-like factors (HLFs), ADAM[TS], bdellin, destabilase, eglin c, ficolin, manillase and saratin are shown in Fig. 1 and Supplementary Fig. S1. Gene trees for the remaining putative anticoagulants were not inferred due to lack of comparative sequences. Hirudin-like and antistasin-like proteins showed a high degree of molecular diversity (Fig. 1A and B).

In the antistasin family tree (Fig. 1A), sequences from *Limnodynastes mexicanus* were nested in each of the three main clans that were recovered in our analysis. These represent the main lineages of putative antistasin-like factor Xa inhibitors: antistasin, therostasin and piguamerin/guamerin/bdellastasin. Within the antistasin clade, putative antistasin-like inhibitors from Piscicolidae form a clan adjacent to sequences from Glossiphoniiformes and Hirudiniformes. The archetypal antistasin and the archetypal ghilanten form a well-supported clan (99% parsimony bootstrap support [PBS]), adjacent to a sequence from *Ozobranchius margo* (Apathy, 1890). Putative antistasins (DN22758 and DN19809) recovered in the *L. mexicana* salivary transcriptome nest as the adjacent group of sequences from arhynchobdellid species. The therostasin clan exhibits a high degree of molecular diversity and is formed by sequences recovered in previous studies from members of Oceanobdelliformes, Glossiphoniiformes, as well as one sequence from *L. mexicana* (DN457). The third main clan is formed by guamerin, piguamerin, bdellastasin and other sequences that presented BLAST hits against antistasin and ghilanten in previous studies. A clan, adjacent to therostasins, formed exclusively by antistasins and ghilantens from *Placobdella* species is also present in our analysis.

Trees for hirudin-like thrombin inhibitors (Fig. 1B) showed that the archetypal hirudin, hirulin and bufrudin form a clan that is separated from the remaining sequences by a proportionally long branch. Several hirudin-like factors form a clade, in which HLF1 cluster as adjacent to the remaining HLFs. The remaining clan is composed of hirudin-like antithrombins from a wide range of leech taxa, including both rhynchobdellids and

arhynchobdellids, adjacent to *L. mexicana* hits against hirudin and bufrudin.

In the antiplatelet tree (Fig. 1C), there is no clear dichotomy between archetypal saratin and LAPP, suggesting that these nomenclaturally distinct anticoagulation factors may indeed be orthologous. The *L. mexicana* sequence is positioned on a long branch adjacent to saratin from *Aliolimnatis fenestrata* (Moore, 1939) and two different species of the genus *Haementeria*.

ADAM[TS] sequences recovered in the *L. mexicana* transcriptome (DN24510 and DN13743) form a clan adjacent to ADAM [TS] from *Oxytonostoma typica* Malm, 1863 (Supplementary Fig. S1A). In the manillase gene tree (Supplementary Fig. S1B), predicted ORFs from *L. mexicana* form a clan with archetypal manillase and an additional sequence from *Haemadipsa interrupta* (Moore, 1935). In the bdellin gene tree (Supplementary Fig. S1C), the *L. mexicana* (DN2497) sequence falls within a clade formed by *H. medicinalis*, *Hirudo nipponia* Whitman 1886 and the archetypal sequence of bdellin. The *L. mexicana*-derived eglin c putative orthologue also sits on a long branch when compared to other branches of the tree (Supplementary Fig. S1D). The hit against destabilase I falls within a clan formed by the archetypal sequence for destabilase and sequences belonging to *Hirudo medicinalis* and *Macrobdella decora* (Say, 1824) (Supplementary Fig. S1E). For ficolin (Supplementary Fig. S1F), the *L. mexicana* sequence and a *H. medicinalis* sequence form an adjacent group to another ficolin-like molecule from *H. medicinalis*.

### Sequence similarity and pairwise alignments

Sequence similarity varied considerably between the newly-sequenced, putative anticoagulants and their respective archetypal variants. Destabilase I showed amino acid sequence similarity of 76.47%, manillase showed 59%, eglin c showed 51.43%, ficolin showed 51.43%, Kazal type serpin showed 49.12%, DN24002\_c0\_g1\_i1 showed 39.58% against hirudin, DN24002\_c0\_g1\_i2 showed 41.18% against hirudin, cystatin showed 43.14%, therostasin showed 34.44%, hirustasin showed 32.73%, 5'-apyrase showed 29.25%, bdellin showed 28.22%, DN19809 showed 28.16% against antistasin, DN22758 showed 25.74% against antistasin, the snake antithrombin showed 23.33%, saratin (orthologous to LAPP; see Kvist *et al.*, 2011) showed 19.89%, DN24901 showed 12.5% against bdellastasin, ADAM showed 10.76%, and ADAM[TS] showed 8.88%.

Pairwise alignments of selected putative anticoagulants are shown in Fig. 2 and Supplementary Fig. S2. Analysis of sequences with hits against antistasin showed important differences when compared to the archetypal anticoagulant. For example, the number of cysteines, involved in disulfide bonds, and the number of antistasin-like domains differ between the comparisons. The alignment of the best hit against antistasin-like proteins is shown in Fig. 2A. DN22758 from *L. mexicana* is shorter than the archetypal sequence and 12 out of the 20 cysteines present in the archetypal sequence are conserved in the *L. mexicana*-derived putative orthologue. One of the other hits against antistasin, ORF DN19809, shows full conservation of all 20 cysteine residues when aligned with the archetypal sequence (Fig. 2B). However, the newly generated sequence also shows the presence of a C-terminal region that is rich in cysteines and that is not shared by the archetypal antistasin. ORFs DN457 (best hit against therostasin) and DN24901 (best hit against bdellastasin) are formed by a single antistasin-like domain (Fig. 2C and D), suggesting orthology with therostasin (DN457), and piguamerin, guamerin or bdellastasin (DN24901).

Pairwise alignment of the putative hirudin and the archetypal sequence (Supplementary Fig. S2A and S2B) evinced full conservation of the six cysteine residues present in both the newly

**Table 1.** Putative anticoagulant proteins with high scoring matches in salivary transcripts from *Limnobdella mexicana*

	BLASTx		BLASTp		Pfam	Signal peptide	GO term
	Local database	Swiss-Prot	Local database	Swiss-Prot			
DN24002_c0_g1_i2	Partial hirudin ( $3.49 \times 10^{-13}$ )	Hirudin variant 2 ( $1.41 \times 10^{-10}$ )	Partial hirudin ( $6.38 \times 10^{-14}$ )	Hirudin variant 2 ( $2.76 \times 10^{-11}$ )	Hirudin ( $4.70 \times 10^{-10}$ )	Yes	MF: serine-type endopeptidase inhibitor activity
DN24272_c0_g1_i1	Manillase ( $1.26 \times 10^{-176}$ )	Hyaluronoglucuronidase (0)	Manillase ( $3.13 \times 10^{-177}$ )	Hyaluronoglucuronidase (0)	Glyco hydro_79n ( $1.60 \times 10^{-25}$ )	Yes	CC: membrane; MF: hydrolase activity, acting on glycosyl bonds
DN16987_c0_g1_i1	Ficolin ( $7.13 \times 10^{-80}$ )	Tenascin ( $1.36 \times 10^{-32}$ )	Ficolin ( $6.41 \times 10^{-80}$ )	Tenascin ( $1.22 \times 10^{-32}$ )	Fibrinogen C ( $1.30 \times 10^{-41}$ )	No	NA
DN5526_c0_g1_i1	Destabilase I ( $5.63 \times 10^{-58}$ )	Lysozyme 1 ( $4.49 \times 10^{-34}$ )	Destabilase ( $4.41 \times 10^{-65}$ )	Lysozyme 1 ( $1.58 \times 10^{-37}$ )	Destabilase ( $9.60 \times 10^{-32}$ )	Yes	MF: lysozyme activity
DN24510_c0_g4_i2	ADAM 19 precursor ( $2.10 \times 10^{-52}$ )	ADAM 12 ( $1.66 \times 10^{-37}$ )	ADAM 19 precursor ( $4.50 \times 10^{-54}$ )	ADAM 12 ( $1.75 \times 10^{-38}$ )	ADAM CR ( $1.20 \times 10^{-25}$ )	No	NA
DN19249_c0_g1_i1	Cystatin ( $8.12 \times 10^{-38}$ )	Cystatin-B ( $2.19 \times 10^{-27}$ )	Cystatin ( $1.30 \times 10^{-38}$ )	Cystatin-B ( $1.61 \times 10^{-27}$ )	Cystatin ( $5.30 \times 10^{-11}$ )	No	MF: cysteine-type endopeptidase inhibitor activity
DN24901_c0_g1_i1	Antistasin ( $7.34 \times 10^{-25}$ )	Antistasin ( $2.96 \times 10^{-22}$ )	Antistasin precursor ( $7.21 \times 10^{-26}$ )	Antistasin ( $2.91 \times 10^{-23}$ )	Antistasin ( $2.00 \times 10^{-45}$ )	Yes	MF: endopeptidase inhibitor activity; CC: intracellular
DN24997_c0_g6_i2	Bdellin ( $5.49 \times 10^{-16}$ )	Bdellin B-3 ( $2.22 \times 10^{-13}$ )	Bdellin ( $1.45 \times 10^{-21}$ )	Bdellin B-3 ( $5.84 \times 10^{-19}$ )	Kazal 1 ( $2.00 \times 10^{-10}$ )	Yes	MF: protein binding
DN19378_c0_g1_i2	Hirustasin ( $1.30 \times 10^{-14}$ )	Hirustasin ( $5.24 \times 10^{-12}$ )	Hirustasin ( $1.17 \times 10^{-14}$ )	Hirustasin ( $4.72 \times 10^{-12}$ )	Antistasin ( $1.60 \times 10^{-5}$ )	No	MF: serine-type endopeptidase inhibitor activity; MF: enzyme inhibitor activity
DN24002_c0_g1_i1	Bufrudin ( $2.50 \times 10^{-11}$ )	Hirudin ( $1.01 \times 10^{-8}$ )	Hirudin ( $2.21 \times 10^{-12}$ )	Hirudin variant-2 ( $8.93 \times 10^{-10}$ )	Hirudin ( $7.60 \times 10^{-9}$ )	Yes	MF: serine-type endopeptidase inhibitor activity; MF: enzyme inhibitor activity
DN20090_c0_g2_i1	LDTI ( $1.91 \times 10^{-11}$ )	LDTI ( $7.71 \times 10^{-9}$ )	Bdellin ( $2.96 \times 10^{-11}$ )	Bdellin B-3 ( $1.19 \times 10^{-8}$ )	Kazal 1 ( $3.80 \times 10^{-9}$ )	Yes	MF: protein binding
DN25439_c1_g5_i1	Kazal type serpin ( $2.85 \times 10^{-15}$ )	NA	Kazal type serpin ( $2.57 \times 10^{-15}$ )	NA	PLAC (0.0081)	No	NA
DN22335_c0_g1_i1	Partial Antithrombin-III from <i>Ophiophagus hannah</i> ( $1.38 \times 10^{-30}$ )	Serpin B11 ( $8.56 \times 10^{-31}$ )	Antithrombin-III partial <i>Ophiophagus hannah</i> ( $4.40 \times 10^{-34}$ )	Serpin B11 ( $3.58 \times 10^{-34}$ )	Serpin ( $7.50 \times 10^{-44}$ )	No	CC: extracellular space
DN5213_c0_g2_i1	Eglin c ( $6.59 \times 10^{-11}$ )	Protease inhibitor ( $2.75 \times 10^{-11}$ )	Eglin C ( $2.63 \times 10^{-11}$ )	Protease inhibitor ( $1.07 \times 10^{-11}$ )	potato inhibit ( $7.00 \times 10^{-15}$ )	No	MF: serine-type endopeptidase inhibitor activity; BP: response to wounding

(Continued)

Table 1. (Continued.)

Local database	BLASTx		BLASTp		Signal peptide	GO term
	Local database	Swiss-Prot	Local database	Swiss-Prot		
DN3752_c0_g1_i1	5'-apyrase (2.38 × 10 <sup>-60</sup> )	Soluble calcium-actiated nucleotidase (2.79 × 10 <sup>-100</sup> )	5'-apyrase (1.71 × 10 <sup>-63</sup> )	Soluble calcium-actiated nucleotidase (2.86 × 10 <sup>-104</sup> )	Yes	MF:calcium ion binding; MF: nucleoside-diphosphatase activity
DN22587_c0_g1_i1	Ca2+ dependant apyrase (1.99 × 10 <sup>-20</sup> )	Trifunctional nucleotide phosphoesterase protein (1.71 × 10 <sup>-26</sup> )	5'- apyrase	Trifunctional nucleotide phosphoesterase protein (1.71 × 10 <sup>-26</sup> )	No	BP: nucleotide catabolic process; MF: hydrolase activity
DN20856_c0_g1_i1	Saratin (1.25 × 10 <sup>-16</sup> )	NA	NA	Peptidase C54 (0.033)	No	NA
DN22758_c3_g7_i1	Ghilanten (7.62 × 10 <sup>-12</sup> )	Ghilanten (3.07 × 10 <sup>-9</sup> )	Ghilanten (6.86 × 10 <sup>-12</sup> )	Antistasin (2.77 × 10 <sup>-9</sup> )	No	MF: enzyme inhibitor activity
DN457_c0_g1_i1	Therostasin (3.11 × 10 <sup>-14</sup> )	Therostasin (1.25 × 10 <sup>-11</sup> )	Therostasin (7.95 × 10 <sup>-15</sup> )	Antistasin (2.77 × 10 <sup>-9</sup> )	Yes	MF: serine-type endopeptidase inhibitor activity; MF: enzyme inhibitor activity
DN19809_c0_g1_i1	Antistasin	Antistasin (1.24 × 10 <sup>-11</sup> )	Antistasin (8.37 × 10 <sup>-17</sup> )	Antistasin (4.33 × 10 <sup>-14</sup> )	Yes	MF: enzyme inhibitor activity

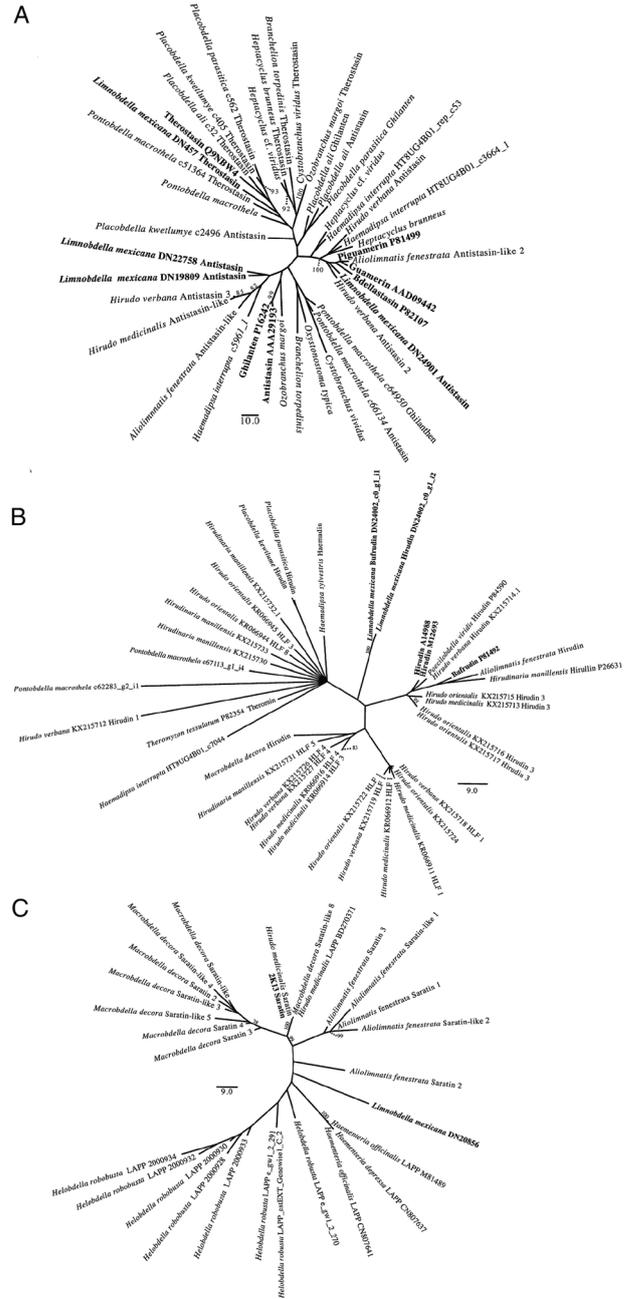
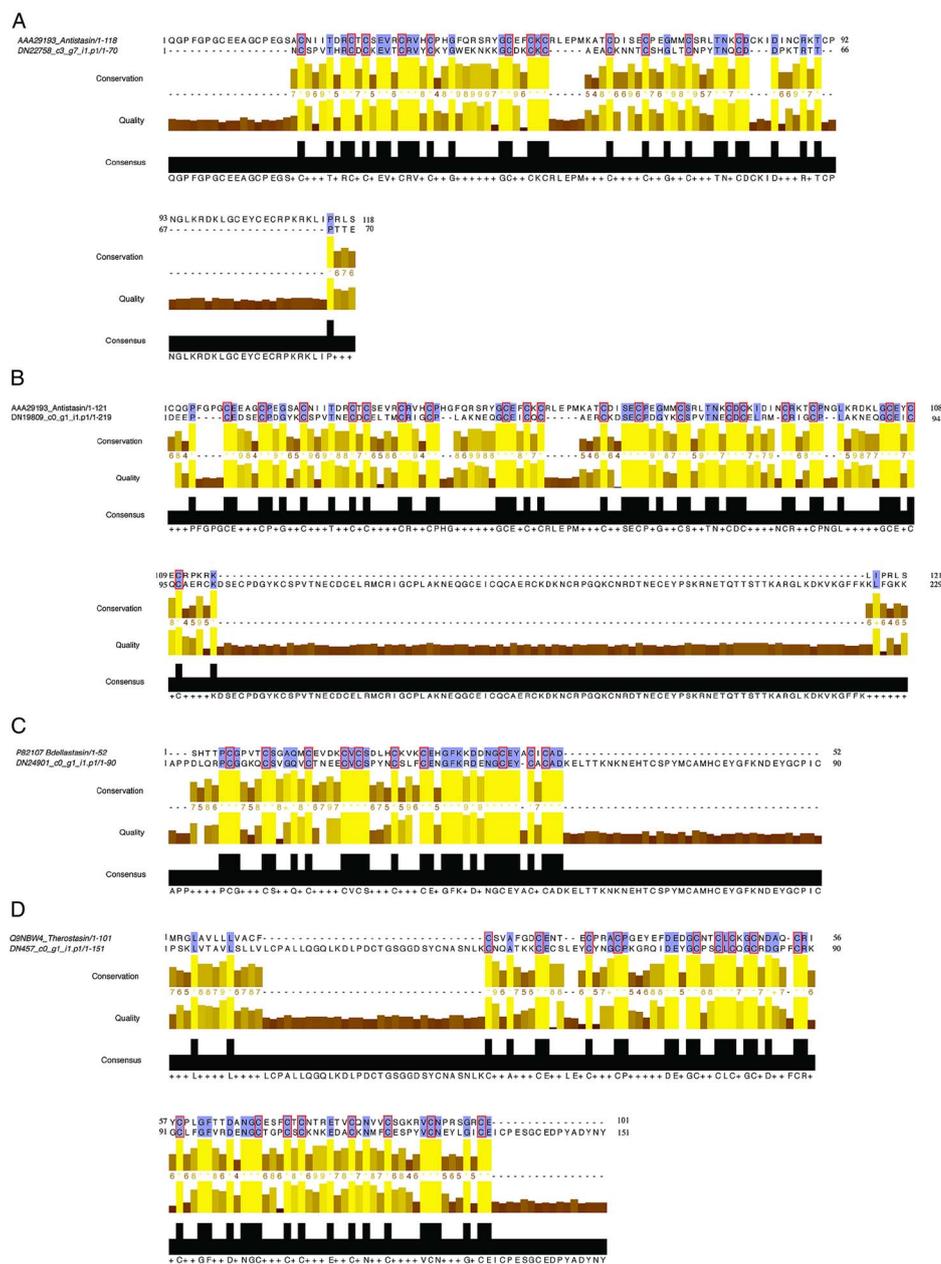


Fig. 1. Consensus tree of the most parsimonious trees of leech-derived putative antistasin-like proteins (A), hirudin-like proteins (B) and saratin/LAPP (C).

generated sequence and the archetypal variant. However, the critical regions for anticoagulation activities in both N-terminal and C-terminal ends show high variation between the aligned sequences.

**Discussion**

The present study details, for the first time, the salivary transcriptome of a praobdellid leech. Species of this family are known for infestations in body orifices of a wide diversity of hosts. Beyond the 15 BLAST hits against well-known anticoagulants, we also found proteins that have previously been associated with anticoagulation processes, but that are not well characterized in leeches. Some of these hits include a thrombin inhibitor from the king cobra *O. hannah*, a Kazal type serpin and a 5'-apyrase. Similar hits have been recovered in previous studies of leech salivary transcriptomes, yet the phylogenetic comparison between the leech-



**Fig. 2.** MAFFT-based alignment of antistatin-like proteins from *Limnobdella mexicana* and their respective archetypal anticoagulant. The archetypal sequence is shown on top of each alignment. A: putative antistatin (DN22758) from *Limnobdella mexicana* aligned with the know sequence of antistatin (AAA29193). B: putative bdellastasin/piguamerin (DN24901) from *Limnobdella mexicana* aligned with the know sequence of bdellastasin (P82107). C: putative therostasin (DN457) from *Limnobdella mexicana* aligned with the know sequence of therostasin (Q9NBW4). D: putative antistatin (DN19809) from *Limnobdella mexicana* aligned with the know sequence of antistatin (AAA29193). Red boxes indicate conserved cysteines.

derived variants and those of other organisms, coupled with low sequence similarities, suggests that these are not orthologous molecules (Kvist *et al.*, 2017; Tessler *et al.*, 2018b). This is further exacerbated by the fact that functional characterization of these molecules is lacking for leeches. The significant BLAST-hits detailed here included inhibitors of platelet aggregation, factor Xa inhibitors and thrombin inhibitors. With regard to factor Xa inhibitors and thrombin inhibitors, whereas the initial BLAST searches of our newly generated sequences recovered crude identities, the phylogenetic analyses much refined their IDs. For example, one of the initial BLAST hits against antistatin was revealed to be a different factor Xa inhibitor (bdellastasin) by the phylogenetic analyses. This underscores the utility of phylogenetic analysis for orthology determination and the importance of including as much diversity as possible in the data matrices.

### Orthology determination

Given the low sequence similarity between some of the putative anticoagulants recovered in the *L. mexicana* transcriptome and their respective archetypal sequences, the detailed examination

of gene trees for each of the putative anticoagulants became of increasing importance. Indeed, the phylogenetic trees showed some evidence of paralogy within *L. mexicana*-derived putative factor Xa inhibitors (Fig. 1A), as well as within putative hirudin-like thrombin inhibitors (Fig. 1B), when compared to their highest scoring BLAST hits (Table 1).

Antistatin-like factor Xa inhibitors canonically inhibit the activation of prothrombin by factor Xa (Krowarsch *et al.*, 2003; Koh and Kini, 2009). They are formed by domains that are rich in cysteine and, whereas therostasin, guamerin and piguamerin are formed by a single domain, antistatin and ghilanten are formed by two antistatin-like domains repeated in tandem (Salzet, 2001; Koh and Kini, 2009). We found, in our analyses, BLAST hits against three antistatin-like factor Xa inhibitors: ghilanten, therostasin and antistatin. Consideration of the details of the pairwise alignments of factor Xa inhibitors disclosed low sequence similarity between these ORFs, with regards to both the number of cysteines and the number of antistatin-like domains. In the antistatin family tree (Fig. 1A), ORF DN457 forms a clan together with sequences annotated as both antistatin and therostasin, and this clan also includes the archetypal sequence of therostasin. DN457 is formed

by a high number of cysteines distributed in a single domain, 17 of which are shared with the archetypal sequence for therostasin (Fig. 2D). The placement of DN457 with a clan including the archetypal sequence of therostasin, coupled with the fact that our sequence seems to only include a single antistasin domain, strongly suggests orthology with therostasin.

ORF DN24901 presented a BLAST hit against antistasin, but the alignment against the archetypal sequence shows important sequence differences, including the number of cysteines involved in disulfide bonds, suggesting structural divergence. In the antistasin family tree, DN24901 falls within a clade formed by three archetypal anticoagulants: piguamerin, guamerin and bdellastasin (Fig. 1A). The phylogenetic position of these three archetypes and the presence of a single antistasin-like domain (Fig. 2B) suggest an orthologous relationship with one of these anticoagulants. Following the results of the phylogenetic tree, we hypothesize that this sequence is orthologous with bdellastasin, but detailed functional characterization of these poorly understood proteins is needed prior to any definitive statement.

As in previous studies (e.g. Kvist et al., 2014), the archetypal sequences for antistasin and ghilanten form a well-supported clan (PBS = 100%), suggesting that these are orthologous sequences that could be circumscribed by a single name (see Fig. 1A). Antistasin was first isolated from the glossiphoniid leech *Haementeria officinalis* and, later, ghilanten was isolated from *Haementeria ghilianii* and showed higher anticoagulant potency (Dunwiddie et al., 1989; Blankenship et al., 1990). The alignment of ghilanten and antistasin shows a high degree of sequence similarity between the C-terminal region of both sequences, with most variations lying on the N-terminal region (data not shown); the active site of antistasin is located in the N-terminal region, and the amino acid differences found between both anticoagulants were previously linked to differences in anticoagulatory power (Dunwiddie et al., 1989; Blankenship et al., 1990). Given the phylogenetic position of antistasin and ghilanten and their high sequence similarity, coupled with their identical function, we suggest that ghilanten should be recognized as an orthologue of antistasin. Two putative antistasin sequences (DN22758 and DN19809) were recovered in *L. mexicana* (Fig. 2A and D). These two sequences are adjacent to putative antistasin recovered from *H. medicinalis*, *Hirudo verbana* Carena, 1820, *M. decora*, *A. fenestrata* and *H. interrupta*, as well as the archetypal sequences for ghilanten and antistasin (Fig. 1A).

Two putative hirudin-like thrombin inhibitors were recovered in our BLAST search. ORFs DN24002\_c0\_g1\_i1 and DN24002\_c0\_g1\_i2 recovered BLAST hits against bufrudin and hirudin, respectively. Two regions are critical for the inhibition of thrombin by hirudin-like thrombin inhibitors. The first three amino acids of the N-terminal regions bind to thrombin, penetrating into its active site. In addition, the last 10 amino acids of the C-terminal region of hirudin are important residues involved in the specific binding with the exosite. As a result, changes to those amino acids cause critical alteration to its binding efficiency (Sharp, 1996; Salzet, 2002). Recently, the discovery of hirudin-like factors, proteins similar to hirudin but that lack anticoagulant capabilities, has raised questions about the capacity of hirudin-like antithrombin proteins for inhibiting coagulation (Müller et al., 2016, 2017). Our alignments show a low degree of amino acid residue conservation in the critical C-terminal region in our hits against hirudin and bufrudin, when compared to the archetypal hirudin. Previous studies have demonstrated that a high level of variation within this region is common among hirudin-like proteins, even among proteins that show anticoagulatory activities (Scacheri et al., 1993; Müller et al., 2016, 2017; Siddall et al., 2016). Our recovered hits are highly similar (84.54%) to each other, and they fall within a clan formed by a

polytomy of putative antithrombin inhibitors and hirudin-like factors. Further investigations are needed to clarify whether sequences within this polytomy are paralogous or orthologous and if they possess any anticoagulatory activity. Furthermore, hirudins (including the archetypal sequence) from *H. verbana*, and *H. medicinalis*, as well as the archetypal bufrudin from *H. manillensis* form a clan (Fig. 1B), indicating paralogy between hirudins and hirudin-like factors, as suggested by Müller et al. (2017).

BLAST searches also recovered hits against numerous putative inhibitors of platelet aggregation. In our bdellin, destabilase I and manillase phylogenetic reconstructions (Supplementary Figs S1B, S1C and S1E), the ORFs with significant hits against those anticoagulants nested within clades formed by species of Hirudiniformes, always including the archetypal sequences, suggesting orthology between our newly generated sequences and the archetypal anticoagulants. Saratin, ADAM[TS], eglin c and ficolin showed low sequence similarity in pairwise alignments with their respective archetypal anticoagulant (Supplementary Figs 2H, 2I). Saratin, is an inhibitor of the von Willebrand factor-mediated platelet aggregator that has been proposed to be orthologous to leech antiplatelet protein (LAPP) (Barnes et al., 2001; Cruz et al., 2001; Kvist et al., 2011). The placement in our analysis of saratin from *L. mexicana* as adjacent to LAPP sequences from glossiphoniid species and the presence of a clade formed by the archetypal saratin, saratin sequences from *H. medicinalis* and *M. decora*, and one LAPP sequence from *H. medicinalis*, suggest orthology between LAPP and saratin (Fig. 1C).

ADAM and ADAM[TS] are members of the metalloproteinase superfamily. Both of these proteins have been linked to biological processes, including the cleavage of the von Willebrand factor (ADAM[TS]13) that plays a major role in blood coagulation (Tang, 2001; Furie and Furie, 2005; Porter et al., 2005; Brocker et al., 2009). Putative metalloproteinase proteins have been recorded in some leech salivary transcriptomes (Kvist et al., 2017; Tessler et al., 2018b), but no functional characterizations have been performed and their role in leech salivary secretions is still unknown. Hits against both families of metalloproteinases were recovered with low sequence similarity against their respective archetypal anticoagulants (ADAM = 10.76% and ADAM[TS] = 8.88%). In our phylogenetic reconstruction, putative ADAM and ADAM[TS] from *L. mexicana* form a clan, adjacent to a sequence from *Oxytonostoma typica* (Supplementary Fig. S1A).

Ficolins are reported in numerous mammals, in which they play an important role as pattern-recognition receptors of pathogens in the immune system (Fujita et al., 2004). Non-immune functions of ficolin-like proteins have been reported in snake venoms, where they seem to act as a trigger for platelet aggregation (Ompraba et al., 2010). Regarding bloodfeeding animals, ficolin has been reported in several leech and arthropod salivary transcriptome studies, which suggests a regulatory role in coagulation during bloodfeeding (Ribeiro et al., 2007; Ribeiro and Arcá, 2009; Min et al., 2010; Kvist et al., 2014, 2017). As in those studies, we suggest that *L. mexicana* possesses proteins orthologous to leech ficolins, based on the fact that our best hit nests within a clade formed by two sequences from *Hirudo medicinalis* (Supplementary Fig. S1F).

Eglin c is an inhibitor of elastase that has been extensively considered as a candidate for drug development (Salzet, 2001). Previous studies have demonstrated low sequence similarity between putative eglin c, sometimes to the inclusion of various cysteine residues and sometimes in their complete absence (Min et al., 2010; Kvist et al., 2013). In *L. mexicana*, the best hit against eglin c possesses a cysteine in position 47 that is not shared with any other eglin c from other leech species. The phylogenetic reconstruction of leech eglin c showed that the *L. mexicana* hit forms a clan with a sequence from *H. interrupta* that also possesses a

single cysteine, but in a different position. The archetypal eglin c forms a separate clan with sequences from *A. fenestrata* and *H. medicinalis* (Supplementary Fig. S1D).

Although our analyses clarify the phylogenetic relationships of some major leech anticoagulants, especially antistasin-like and hirudin-like anticoagulants, we recognize that the position of the root in a tree may affect orthology determination (see Ballesteros and Hormiga, 2016); if the root of a tree is placed within a monophyletic assemblage this assemblage will be rendered paraphyletic after rooting. Since there is no objective way to determine the root for the phylogeny of these major putative leech anticoagulants, coupled with the fact that no study has investigated the presence of putative orthologues within non-hirudinean clitellates, we left the trees unrooted. Given the importance of the root for estimations of orthology using phylogenies, future efforts should focus on finding a logical root for the trees of major leech anticoagulants.

### Association of host specificity and anticoagulant diversity

Previous studies on leech salivary transcriptomes have suggested a positive association between the number of anticoagulants and host-generality (Kvist *et al.*, 2014). Therefore, it is hypothesized that host-generalist species have a wider anticoagulant repertoire compared to host-specialists. However, Tessler *et al.* (2018b) found no significant pattern between piscicolid leeches restricted to elasmobranch hosts and generalist leeches of the same family, though the number of anticoagulants found in elasmobranch-restricted parasitic leeches was slightly lower. *Limnodynastes mexicanus* is a member of Praobdellidae, a family that is known to be associated with nasopharyngeal infections; host records for members of this family include mammals, reptiles, amphibians and invertebrates (Phillips *et al.*, 2010; Siddall *et al.*, 2013; Nakano *et al.*, 2017). The number of anticoagulants found in *L. mexicanus* was slightly higher than most of the previous studies of leech salivary transcriptomes, except for that of *H. interrupta*. However, comparison between different available datasets is exacerbated by the fact that sequencing efforts and technologies vary among different studies. Moreover, previous studies have focused on the presence of homologues of archetypal anticoagulants, not accounting for the presence of multiple duplicates (Min *et al.*, 2010; Kvist *et al.*, 2014; 2017; Siddall *et al.*, 2016) that might possess minute sequence and structural differences such as those recorded for hirudin variants, specifically hirudin-like factors (Scacheri *et al.*, 1993; Müller *et al.*, 2017). Nonetheless, the sequencing effort and determinations of putative anticoagulants applied in the present study is exactly the same as in Kvist *et al.* (2017). Thus, the higher number of putative anticoagulants found might be an indication of a positive association between host diversity and diversity of anticoagulants in leeches, but further investigations, putatively involving full genome sequencing, are still needed to confirm such a relationship.

### Implication for evolution of bloodfeeding in leeches

The origin of bloodfeeding in leeches has been widely debated. Sawyer (1986) suggested a bloodfeeding ancestor for proboscis-bearing leeches, but a macrophagous ancestor for Arhyncobdellida. However, more recent phylogenetic studies have suggested bloodfeeding as the ancestral behaviour of Hirudinida, with independent losses within proboscis and non-proboscis-bearing lineages (Apakupakul *et al.*, 1999; Tessler *et al.*, 2018a). Transcriptomic data have been used to test the hypothesis of a bloodfeeding ancestor of leeches, under the perception that non-bloodfeeding leeches will maintain anticoagulation factors based on this ancestral donation,

as well as the fact that bloodfeeding leeches across the leech phylogeny should share orthologous proteins if these were derived from a common ancestor. The presence of orthologues in Glossiphoniidae and Piscicolidae, when compared to Hirudiniformes, for each of hirudin, mannase, piguamerin, bdellastasin, destabilase and LDIT, recorded in previous studies, suggest that these were derived from an ancestor of modern leeches that approaches the time frame of the ancestor of all leeches (Siddall *et al.*, 2016; Tessler *et al.*, 2018b). In other words, it has been suggested that if (e.g.) Piscicolidae shares orthologous anticoagulant proteins with Hirudinidae, the ancestor that provided these proteins existed early in leech evolution, approaching the base of the leech tree. However, previous studies that recorded the presence of these anticoagulants in many Piscicolidae and Glossiphoniidae species have approached the problem of orthology determination using a simple BLAST search or performed phylogenetic analysis that did not include the majority of the sequences available for the anticoagulant families studied (Siddall *et al.*, 2016; Tessler *et al.*, 2018b).

Comparison between BLAST hits and gene trees of their respective putative orthologous proteins showed incongruence of antistasin-like and hirudin-like putative anticoagulants. Therefore, previous records of these anticoagulants can be questioned. Our phylogenetic analysis of putative anticoagulants found in *L. mexicana* contributes to the orthology determination of several families of leech anticoagulants and, consequently, serves to clarify the origin of these major protein lineages. Our antistasin family tree shows that putative antistasin, bdellastasin and therostasin are found in both proboscis- and non-proboscis-bearing leeches, suggesting that the ancestral leech possessed the molecular armoury to inhibit factor Xa, even if this inhibition is not directly associated with bloodfeeding.

On the other hand, hirudin derived from hirudiniform species form a clan in the hirudin gene tree, and this clan is clearly separated from hirudin-like factor (HLFs) and other putative thrombin inhibitors from proboscis- and non-proboscis-bearing leeches. Since HLFs do not possess anticoagulatory capabilities and their function is still unknown (Müller *et al.*, 2016, 2017), not much can be inferred about the anticoagulatory activity of hirudin-like proteins in the ancestral leech, as suggested by Siddall *et al.* (2016) and Tessler *et al.* (2018b). Our analyses also present evidence of orthology between hits for manillase and LAPP/saratrin recovered from early and late branching lineages of leeches, pointing to a possible single origin of these proteins. Therefore, with evidence detailed above, the sole presence of a BLAST hit against a putative anticoagulant should not be interpreted as evidence for the origin of certain molecular function. BLAST homology determination should be followed by robust orthology tests.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0031182019000593>

**Author ORCIDs.**  Rafael Eiji Iwama, 0000-0002-2430-4178

**Acknowledgements.** The authors acknowledge Alejandro Manzano-Marín for assistance with software usage and Michael Tessler for the curation of the anticoagulant local data used on BLAST searches and for creating the script used to run TNT. We thank Rosa Fernández for technical assistance during RNA extraction and we thank the Willi Hennig Society for making TNT freely available. Two anonymous reviewers provided comments that help to make this article clearer.

**Financial support.** This project was funded by a NSERC Discovery Grant and an Olle Engkvist Byggnästare Foundation scholarship (SK), as well as internal funds from the Faculty of Arts and Sciences, Harvard University (GG).

**Conflicts of interest.** None.

**Ethical standards.** Not applicable.

## References

- Apakupakul K, Siddall ME and Bureson EM (1999) Higher level relationships of leeches (Annelida: Clitellata: Euhirudinea) based on morphology and gene sequences. *Molecular Phylogenetics and Evolution* **12**, 350–359.
- Ballesteros JA and Hormiga G (2016) A New orthology assessment method for phylogenomic data: unrooted phylogenetic orthology. *Molecular Biology and Evolution* **33**, 2117–2134.
- Barnes CS, Krafft B, Frech M, Hofmann UR, Papendieck A, Dahlems U, Gellissen G and Hoylaerts MF (2001) Production and characterization of saratin, an inhibitor of von willebrand factor-dependent platelet adhesion to collagen. *Seminars in Thrombosis and Hemostasis* **27**, 337–347.
- Blankenship DT, Brankamp RG, Manley GD and Cardin AD (1990) Amino acid sequence of ghilanten: anticoagulant-antimetastatic principle of the south American leech, *Haementeria ghilianii*. *Biochemical and Biophysical Research Communication* **166**, 1384–1389.
- Borda E and Siddall ME (2004) Arhynchobdellida (Annelida: Oligochaeta: Hirudinida): Phylogenetic relationships and evolution. *Molecular Phylogenetics and Evolution* **30**, 213–225.
- Brocker CN, Vasilou V and Nebert DW (2009) Evolutionary divergence and functions of the ADAM and ADAMTS gene families. *Human Genomics* **4**, 43–55.
- Cruz CP, Eidt J, Drouilhet J, Brown AT, Wang YF, Barnes CS and Moursi MM (2001) Saratin, an inhibitor of von willebrand factor-dependent platelet adhesion, decreases platelet aggregation and intimal hyperplasia in a rat carotid endarterectomy model. *Journal of Vascular Surgery* **34**, 724–729.
- Dunwiddie C, Thornberry NA, Bull HG, Sardana M, Friedman PA, Jacobs JW and Simpson E (1989) Antistasin, a leech-derived inhibitor of factor Xa. Kinetic analysis of enzyme inhibition and identification of the reactive site. *Journal of Biological Chemistry* **264**, 16694–16699.
- Eddy SR (2011) Accelerated profile HMM searches. *PLoS Computational Biology* **7**, 1–17. doi: 10.1371/journal.pcbi.1002195.
- Elliott JM and Kutschera U (2011) Medicinal leeches: historical use, ecology, genetics and conservation. *Freshwater Reviews* **4**, 21–41.
- Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J and Bateman A (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **44**, D279–D285.
- Fujita T, Matsushita M and Endo Y (2004) The lectin-complement pathway – Its role in innate immunity and evolution. *Immunological Reviews* **198**, 185–202.
- Furie BBC and Furie BBC (2005) Review series Thrombus formation in vivo. *The Journal of Clinical Investigation* **115**, 3355–3362.
- Goloboff PA and Catalano SA (2016) TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* **32**, 221–238.
- Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780.
- Koh Y and Kini R (2009) Molecular diversity of anticoagulants from haematophagous animals. *Thrombosis and Haemostasis* **102**, 437–453.
- Krowarsch D, Cierpicki T, Jelen F and Otlewski J (2003) Canonical protein inhibitors of serine proteases. *Cellular and Molecular Life Sciences* **60**, 2427–2444.
- Kvist S, Sarkar IN and Siddall ME (2011) Genome-wide search for leech anti-platelet proteins in the non-blood-feeding leech *Helobdella robusta* (Rhynchobdellida: Glossiphoniidae) reveals evidence of secreted anticoagulants. *Invertebrate Biology* **130**, 344–350.
- Kvist S, Min GS and Siddall ME (2013) Diversity and selective pressures of anticoagulants in three medicinal leeches (Hirudinida: Hirudinidae, Macrobdellidae). *Ecology and Evolution* **3**, 918–933.
- Kvist S, Brugler MR, Goh TG, Giribet G and Siddall ME (2014) Pyrosequencing the salivary transcriptome of *Haemadipsa interrupta* (Annelida: Clitellata: Haemadipsidae): Anticoagulant diversity and insight into the evolution of anticoagulation capabilities in leeches. *Invertebrate Biology* **133**, 74–98.
- Kvist S, Ocegüera-Figueroa A, Tessler M, Jiménez-Armenta J, Freeman RM, Giribet G and Siddall ME (2017) When predator becomes prey: investigating the salivary transcriptome of the shark-feeding leech *Pontobdella macrothela* (Hirudinida: Piscicolidae). *Zoological Journal of the Linnean Society* **179**, 725–737.
- Min G-S, Sarkar IN and Siddall ME (2010) Salivary transcriptome of the north American medicinal leech, *Macrobdella decora*. *The Journal of Parasitology* **96**, 1211–1221.
- Müller C, Mescke K, Liebig S, Mahfoud H, Lemke S and Hildebrandt JP (2016) More than just one: multiplicity of hirudins and hirudin-like factors in the medicinal leech, *Hirudo medicinalis*. *Molecular Genetics and Genomics* **291**, 227–240.
- Müller C, Haase M, Lemke S and Hildebrandt JP (2017) Hirudins and hirudin-like factors in Hirudinidae: implications for function and phylogenetic relationships. *Parasitology Research* **116**, 313–325.
- Nakano T, Tomikawa K, Sakono T and Yoshikawa N (2017) Praobdellidae (Hirudinida: Arhynchobdellida) is not specific only to the mucous-membrane after all: discovery of a praobdellid leech feeding on the Japanese freshwater crab *Geothelphusa dehaani*. *Parasitology International* **66**, 210–213.
- Oka A (1934) Note sur les moeurs dela *Myxobdella sinanensis*. *Proceeding of the Imperial Academy (Tokyo)* **10**, 519–520.
- Ompraba G, Chapeaurouge A, Doley R, Devi KR, Padmanaban P, Venkatraman C, Velmurugan D, Lin Q and Kini RM (2010) Identification of a novel family of snake venom proteins veficolins from cerberus rynchops using a venom gland transcriptomics and proteomics approach. *Journal of Proteome Research* **9**, 1882–1893.
- Petersen TN, Brunak S, Von Heijne G and Nielsen H (2011) Signalp 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785–786.
- Phillips AJ, Arauco-Brown R, Ocegüera-Figueroa A, Gomez GP, Beltrán M, Lai YT and Siddall ME (2010) *Tyrannobdella rex* n. gen. n. sp. and the evolutionary origins of mucosal leech infestations. *PLoS ONE* **5**, 1–8. doi: 10.1371/journal.pone.0010057.
- Porter S, Clark IM, Kevorkian L and Edwards D (2005) The ADAMTS metalloproteinases Sarah. *Biochemical Journal* **386**, 15–27.
- Ribeiro JMC and Arcà B (2009) From Sialomes to the sialoverse. An insight into salivary potion of blood-feeding insects. *Advances in Insect Physiology* **37**, 59–118. doi: 10.1016/S0065-2806(09)37002-2.
- Ribeiro JMC and Francischetti IMB (2003) Role of arthropods saliva in blood feeding: Sialome and post-sialome perspectives. *Annual Review of Entomology* **48**, 73–88.
- Ribeiro JMC, Arcà B, Lombardo F, Calvo E, Phan VM, Chandra PK and Wikel SK (2007) An annotated catalogue of salivary gland transcripts in the adult female mosquito, *Aedes aegypti*. *BMC Genomics* **8**, 1–27.
- Salzet M (2001) Anticoagulants and inhibitors of platelet aggregation derived from leeches. *FEBS Letters* **492**, 187–192.
- Salzet M (2002) Leech thrombin inhibitors. *Current Pharmaceutical Design* **8**, 493–503.
- Sawyer RT (1986) *Leech Biology and Behavior*. Oxford: Clarendon Press.
- Scacheri E, Nitti G, Valsasina B, Orsini G, Visco C, Ferrera M, Sawyer RT and Sarmientos P (1993) Novel hirudin variants from the leech *Hirudinaria manillensis*: amino acid sequence, cDNA cloning and genomic organization. *European Journal of Biochemistry* **214**, 295–304.
- Sharp KA (1996) Electrostatic interactions in hirudin-thrombin binding. *Biophysical Chemistry* **61**, 37–49.
- Siddall ME and Bureson EM (1995). Phylogeny of the Euhirudinea: Independent evolution of blood feeding by leeches? *Canadian Journal of Zoology* **73**, 1048–1064.
- Siddall ME, Rood-goldman R, Barriot A and Barboutist C (2013) The eyes have it: long-distance dispersal by an intraorbital leech parasite of birds. *The Journal of Parasitology* **99**, 1137–1139.
- Siddall ME, Brugler MR and Kvist S (2016) Comparative transcriptomic analyses of three species of *Placobdella* (rhynchobdellida: glossiphoniidae) confirms a single origin of blood feeding in leeches. *Journal of Parasitology* **102**, 143–150.
- Tang BL (2001) ADAMTS: a novel family of extracellular matrix proteases. *International Journal of Biochemistry and Cell Biology* **33**, 33–44.
- Tessler M, de Carle D, Voiklis ML, Gresham OA, Neumann JS, Cios S and Siddall ME (2018a) Worms that suck: Phylogenetic analysis of Hirudinea solidifies the position of Acanthobdellida and necessitates the dissolution of Rhynchobdellida. *Molecular Phylogenetics and Evolution* **127**, 129–134.
- Tessler M, Marancik D, Champagne D, Dove A, Camus A, Siddall ME and Kvist S (2018b) Marine leech anticoagulant diversity and evolution. *Journal of Parasitology* **104**, 210–220.
- Waterhouse A. M., Procter J. B., Martin D. M. A., Clamp M. and Barton G. J. (2009). Jalview version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)* **25**, 1189–1191.
- Wilkinson M, McInerney JO, Hirt RP, Foster PG and Ebley TM (2007) Of clades and clans: terms for phylogenetic relationships in unrooted trees. *Trends in Ecology and Evolution* **22**, 114–115.