

# The Origin and Evolution of Antistasin-like Proteins in Leeches (Hirudinida, Clitellata)

Rafael Eiji Iwama <sup>1,2,\*</sup>, Michael Tessler <sup>3,4,5,\*</sup>, Mark E. Siddall<sup>6</sup>, and Sebastian Kvist<sup>1,2</sup>

<sup>1</sup>Department of Natural History, Royal Ontario Museum, Toronto, Ontario, Canada

<sup>2</sup>Department of Ecology and Evolutionary Biology, University of Toronto, Ontario, Canada

<sup>3</sup>Department of Biology, St. Francis College, Brooklyn, New York, USA

<sup>4</sup>Division of Invertebrate Zoology, American Museum of Natural History, New York, New York, USA

<sup>5</sup>Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, New York, USA

<sup>6</sup>ORCID: 0000-0002-3977-7517

\*Corresponding authors: E-mails: rafael.iwama@mail.utoronto.ca; mtessler@sfc.edu.

Accepted: 15 November 2020

## Abstract

Bloodfeeding is employed by many parasitic animals and requires specific innovations for efficient feeding. Some of these innovations are molecular features that are related to the inhibition of hemostasis. For example, bloodfeeding insects, bats, and leeches release proteins with anticoagulatory activity through their salivary secretions. The antistasin-like protein family, composed of serine protease inhibitors with one or more antistasin-like domains, is tightly linked to inhibition of hemostasis in leeches. However, this protein family has been recorded also in non-bloodfeeding invertebrates, such as cnidarians, mollusks, polychaetes, and oligochaetes. The present study aims to 1) root the antistasin-like gene tree and delimit the major orthologous groups, 2) identify potential independent origins of salivary proteins secreted by leeches, and 3) identify major changes in domain and/or motif structure within each orthologous group. Five clades containing leech antistasin-like proteins are distinguishable through rigorous phylogenetic analyses based on nine new transcriptomes and a diverse set of comparative data: the trypsin + leukocyte elastase inhibitors clade, the antistasin clade, the therostasin clade, and two additional, unnamed clades. The antistasin-like gene tree supports multiple origins of leech antistasin-like proteins due to the presence of both leech and non-leech sequences in one of the unnamed clades, but a single origin of factor Xa and trypsin + leukocyte elastase inhibitors. This is further supported by three sequence motifs that are exclusive to antistastins, the trypsin + leukocyte elastase inhibitor clade, and the therostasin clade, respectively. We discuss the implications of our findings for the evolution of this diverse family of leech anticoagulants.

**Key words:** antistasin, anticoagulants, bloodfeeding, protein evolution, leeches.

## Significance

Antistasin-like proteins are common in leech saliva and have also been recorded in non-bloodfeeding invertebrates, such as cnidarians, polychaetes, and oligochaetes. Their function in non-bloodfeeding invertebrates is still unknown, and so is the origin and protein-level changes related to bloodfeeding in leeches. The present study establishes a root for the leech antistasin-like gene tree, delimits orthologous groups of leech antistasin-like proteins, and identifies conserved regions specific to each protein lineage. Our findings support multiple origins of leech antistasin-like proteins and identification of protein-level changes that might be related to bloodfeeding.

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

## Introduction

Bloodfeeding occurs in mammals, fishes, birds, and numerous invertebrate taxa, such as insects, arachnids, and leeches (Mans and Neitz 2004; Oviedo et al. 2009; Phillips and Siddall 2009; Gerdol et al. 2018). Several evolutionary innovations are required for bloodfeeding, including both morphological and molecular adaptations (Ribeiro and Arcà 2009). Although morphologically diverse, feeding structures and mouthparts of bloodfeeding animals usually include an organ designed for incision making, such as a proboscis in mosquitoes and glossiphoniid leeches, or specific adaptations to the teeth and mandibular systems such as in bats and lampreys (Potter and Gill 2003; Ribeiro and Arcà 2009; Tessler, de Carle, et al. 2018). Other examples of innovations include organs harboring bacteria that aid in the production of B vitamins that are lacking from the blood diet, specialized organs for fixation to the host's skin, metabolic adaptations against heme toxicity, and water excretion (Sawyer 1986; Graça-Souza et al. 2006; Kersch and Pietrantonio 2011; Siddall et al. 2011; Manzano-Marín et al. 2015; Smith et al. 2015). Some of the most important innovations required for bloodfeeding are related to the inhibition of hemostasis (blood clotting) through the secretion of salivary compounds that act as anticoagulants (Ribeiro and Arcà 2009; Min et al. 2010). Anticoagulants are proteins that inhibit the process of hemostasis by interaction with various components of the coagulation cascade. Such proteins have already been recorded for a subset of bloodfeeding species, including bats, insects, and leeches (Assumpção et al. 2008; Min et al. 2010; Mizurini et al. 2013; Ribeiro et al. 2018; Tessler, Marancik, et al. 2018), but they are not restricted to bloodfeeders. In some non-bloodfeeding organisms, anticoagulants prevent the formation of clots in normal conditions and allow blood flow in the circulatory system of the organism (Tsakiris et al. 1999; Hanumanthaiah et al. 2002; Lee et al. 2010; Mehr et al. 2015).

Leeches (*Hirudinida* sensu Tessler, de Carle, et al. 2018) are notorious for their bloodfeeding behavior, which was extensively leveraged in medicinal practices during the 18th and 19th centuries (Elliott and Kutschera 2011), and they continue to be used in modern, science-based medicine (Utley et al. 1998; Herlin et al. 2017). Despite the prominence of bloodfeeding as a feeding strategy in leeches, about one-third of the known leech diversity are non-bloodfeeding species that feed on other macroinvertebrates by two main strategies. Liquidosomatophagous leeches ingest body liquids and tissue of invertebrates with the use of a proboscis; this feeding strategy is manifested in leeches of the suborders *Glossiphoniiformes* and *Oceanobdelliformes* (Sawyer 1986). By contrast, some members of the *Americobdelliformes* + *Erpobdelliformes* + *Hirudiniformes* clade (formerly

*Arhynchobdellida*) are macrophagous, meaning that their prey are ingested whole. *Hirudinea* (leeches + *Branchiobdellida* [crayfish worms] + *Acanthobdellida* [salmonid-fish parasites]) is consistently recovered as the sister to the non-bloodfeeding oligochaete family *Lumbriculidae*, and two hypotheses about the origin of bloodfeeding have been proposed (Siddall et al. 2001; Erséus and Källersjö 2004; Tessler, de Carle, et al. 2018). The first hypothesis was proposed by Sawyer (1986) and it states that the bloodfeeding behavior in leeches evolved from a non-bloodfeeding ancestor. However, the detailed placement of non-bloodfeeding taxa on the leech tree has led to the second hypothesis, suggesting bloodfeeding as the plesiomorphic state for leeches with subsequent independent losses of the behavior (Apakupakul et al. 1999; Borda and Siddall 2004; Tessler, de Carle, et al. 2018).

Bloodfeeding leeches use three main strategies to inhibit hemostasis: 1) inhibition of crosslinked platelets by preventing platelet adhesion to collagen, 2) inhibition of platelet adhesion to fibrinogen, and 3) thrombin inhibition (Koh and Kini 2009; Francischetti 2010). Among leech thrombin inhibitors, antistatin targets factor Xa, an agonist of prothrombin (Dunwiddie et al. 1989). Antistatin is an inhibitor of serine proteases and is formed by two tandemly repeated domains that each includes ten cysteine residues (Nutt et al. 1988). These domains are also found in other leech anticoagulants that belong to the antistatin-like protein family, such as ghilanten, therostatin, theromin, bdellastatin, guamerin, piguamerin, hirustatin, poecistatin, and gigastatin. Both ghilanten and therostatin also target factor Xa, whereas theromin targets thrombin (Blankenship et al. 1990; Chopin et al. 2000; Salzet et al. 2000); piguamerin, guamerin, bdellastatin, hirustatin, and poecistatin are thought to target plasmin, trypsin, and leukocyte elastase, although their role in hemostasis inhibition is not well understood (Söllner et al. 1994; Jung et al. 1995; Mittl et al. 1997; Rester et al. 1999).

Proteins with antistatin-like domains have been recorded from several non-bloodfeeding invertebrate taxa, such as cnidarians, mollusks, polychaetes, and oligochaetes (Holstein et al. 1992; Lee et al. 2010; Nikapitiya et al. 2010; Mehr et al. 2015). Although it has been proposed that these proteins might be involved in immune response in non-bloodfeeding taxa, their specific function remains unclear (Mehr et al. 2015). The presence of antistatin-like proteins in non-bloodfeeding animals, coupled with the different targets for leech antistatin-like proteins, opens up questions regarding the origin and evolution of these proteins. The present study aims to clarify some of these questions by 1) rooting the antistatin-like tree and delimiting the major orthologous groups, 2) determining if independent origins should be inferred for leech antistatin-like proteins, and 3) identifying the major changes within each orthologous group.

## Materials and Methods

### Specimen Collection and Data Set Generation

All specimens used to generate the new transcriptomes were preserved in RNAlater at the time of collection. They were later identified using traditional microscopy, and the identities were secondarily confirmed using the cytochrome c oxidase subunit I and 18S rDNA loci from the transcriptomes described below. Salivary glands were dissected for larger specimens, medium sized specimens had their anterior used (where salivary tissue is found), and for the very small branchiobdellidans the entire specimen was used. The following species were newly sequenced: *Lumbriculus variegatus* (USA, NY), *Xironogiton victoriensis* (Germany), *Branchiobdella* cf. *kozarovi* (the Netherlands), *Acanthobdella peledina* (Sweden), *Mesobdella gemmata* (Chile), *Cylicobdella costaricae* (Costa Rica), *Patagoniobdella* sp. (Chile), *Erbobdella* sp. (USA, VT), and *Americobdella valdiviana* (Chile). Additionally, transcriptome data from 13 annelid species available on the NCBI Sequence Read Archive (SRA) were included in the analysis. Table 1 lists all transcriptomes used in the present study. The data set was supplemented by oligochaete expressed sequence tag (EST) sequence data available from GenBank and transcriptome sequences identified as antistasin, therostasin, guamerin, piguamerin, and bdellastasin from previous leech studies (Kvist et al. 2017; Tessler, Marancik, et al. 2018; Iwama et al. 2019). To further explore the root of the antistasin-like gene tree, a second data set was generated including all sequences in the original data set, as well as putative antistasin-like proteins from non-annelid taxa. The complete list of sequences included in the final data set, along with literature references, is available as [supplementary table S1, Supplementary Material](#) online. The final data sets are available as [supplementary data S2 and S3, Supplementary Material](#) online.

### RNA Extraction, cDNA Library Preparation, and Sequencing

RNA extractions of leech tissue followed Kvist et al. (2014) and Tessler, Marancik, et al. (2018). Briefly, a TRIzol-based (Life Sciences) protocol was used for extraction and, thereafter, the isolated RNA was further extracted using a Qiagen RNeasy Mini Kit. Finally, an Illumina HiSeq 2500 with a TruSeq 2 × 125-bp paired end library prep was used for transcriptome sequencing at the New York Genome Center.

### Data Sanitation and Annotation

Raw sequences were trimmed and filtered with Trimmomatic ver. 0.39 (Bolger et al. 2014) and TrimGalore ver. 0.6.4 ([http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore), last accessed November 24, 2020) applying a PHRED score of

30 and, thereafter, assembled with Trinity ver. 2.8.1 (Haas et al. 2013) with pair distance set to 800. Open reading frames (ORFs) among transcriptomic contigs and oligochaete ESTs with a minimum of 30 amino acids were predicted with Transdecoder ver. 5.2.0. Predicted ORFs were BLASTed both against a local database adapted from Kvist et al. (2014, 2017) and Tessler, Marancik, et al. (2018), and also against the Swiss-Prot database, using BlastP and BlastX with a cutoff e-value of  $1E^{-10}$ .

Protein domains were identified through searches against the Pfam database (Finn et al. 2016), Prosite (Sigrist et al. 2002), PRINTS (Attwood et al. 1994), and SUPERFAMILY (Wilson et al. 2007) with InterProScan ver. 5.4 (Jones et al. 2014). Signal peptides were predicted using SignalP ver. 4.1 (Petersen et al. 2011), with the sensitive *D*-cutoff value. The final antistasin-like data set was created from previously annotated antistasin-like sequences from other studies and BLAST hits against antistasin-like archetypal proteins that did not have any better hits against other proteins in Swiss-Prot and in which only antistasin-like domains were predicted.

In order to identify conserved regions among the antistasin-like proteins in our data set and to establish homology between domains, we used the MEME suite ver. 5.1.1 online tool (Bailey et al. 2015) to find motifs across putative orthologs of anticoagulants found in both oligochaetes and leeches. MEME uses a deterministic optimization algorithm to optimize a position weight matrix that describes the putative motif. We restricted the number of repeats that any motif was allowed within the same sequence and also restricted their length to 25 amino acids (corresponding to the length of antistasin-like domains) in order to differentiate the tandemly repeated antistasin-like domains.

### Alignments and Phylogenetic Analysis

All alignments were performed using the web version of MAFFT ver. 7 (Kato and Standley 2013) with automatic assignment of alignment strategy, the BLOSUM62 scoring matrix, and gap opening penalty of 3.00. IQTREE ver. 1.6.12 (Nguyen et al. 2015) was used to infer the best-fitting model of amino acid evolution and, using the same software, gene trees were reconstructed under maximum likelihood (ML). The best-fitting model, WAG + I + G4 (General Empirical Model of Protein Evolution), was used for all gene tree estimations, as suggested by IQTREE. The default heuristic search in IQTREE was performed with 1,000 replications. Likelihood bootstrap support values were calculated using 1,000 replicates with default settings. The trees were rooted at *Loimia bermudensis* (Terebellidae) following previous phylogenetic hypotheses (e.g., Rousset et al. 2007). Orthologous groups were defined according to the distribution of archetypal anticoagulants and the distribution of motifs predicted.

**Table 1**

List of Transcriptomes Used in the Present Study and Their Respective Statistics

Species	SRR	TSA	TSA Used in This Study	Reference	Raw Reads	Contigs	ORFs	N50
<i>Lumbriculus variegatus</i>	SRR12921559	GIWA00000000	GIWA01000000	Present study	32,973,677	250,739	135,066	1,975
<i>Xironogiton victoriensis</i>	SRR12921556	GIWI00000000	GIWI01000000	Present study	37,418,234	318,490	148,552	714
<i>Branchiobdella cf. kozarovi</i>	SRR12921562	GIVZ00000000	GIVZ01000000	Present study	36,321,044	157,890	76,134	1,202
<i>Acanthobdella peledina</i>	SRR12921564	GIWE00000000	GIWE01000000	Present study	37,413,030	145,245	65,145	1,179
<i>Mesobdella gemmata</i>	SRR12921558	GIWH00000000	GIWH01000000	Present study	38,530,097	116,307	79,332	2,312
<i>Cylicobdella costaricae</i>	SRR12921561	GIWF00000000	GIWF01000000	Present study	36,208,544	90,518	69,695	3,088
<i>Patagoniobdella sp.</i>	SRR12921557	GIWB00000000	GIWB01000000	Present study	36,495,218	107,805	75,331	2,573
<i>Erpobdella sp.</i>	SRR12921560	GIWG00000000	GIWG01000000	Present study	38,688,118	99,710	68,909	2,747
<i>Americobdella valdiviana</i>	SRR12921563	GIVY00000000	GIVY01000000	Present study	30,510,929	73,730	47,931	2,578
<i>Aulodrilus japonicus</i>	SRR10997429	—	—	Erséus et al. (2020)	30,043,367	95,949	74,793	1,543
<i>Branchiura sowerbyi</i>	SRR9668457	—	—	Zhao et al. (2020)	22,053,563	150,477	108,372	2,250
<i>Chaetogaster diaphanus</i>	SRR10997419	—	—	Erséus et al. (2020)	18,614,401	83,045	49,988	1,260
<i>Criodrilus lacuum</i>	SRR5353276	—	—	Anderson et al. (2017)	22,692,862	113,512	58,490	725
<i>Dendrobaena hortensis</i>	SRR5353263	—	—	Anderson et al. (2017)	29,866,771	142,713	89,432	1,448
<i>Dichogaster saliens</i>	SRR5353279	—	—	Anderson et al. (2017)	30,019,522	21,555	11,349	626
<i>Drawida sp.</i>	SRR5353252	—	—	Anderson et al. (2017)	25,982,583	152,918	77,737	1,042
<i>Eudrilus eugeniae</i>	SRR5353275	—	—	Anderson et al. (2017)	28,843,703	66,148	37,199	1,001
<i>Glossoscolex sp.</i>	SRR5353272	—	—	Anderson et al. (2017)	8,536,064	139,997	79,610	939
<i>Kynotus pittarelli</i>	SRR5353264	—	—	Anderson et al. (2017)	21,051,306	98,804	65,050	1,075
<i>Loimia bermudensis</i>	SRR11434463	—	—	Stiller et al. (2020)	54,545,318	116,168	67,198	2,168
<i>Lutodrilus multivesiculatus</i>	SRR5353257	—	—	Anderson et al. (2017)	6,930,865	58,513	41,218	1,098
<i>Propappus volki</i>	SRR5353250	—	—	Anderson et al. (2017)	25,630,160	120,312	64,644	1,260
<i>Rhinodrilus priollii</i>	SRR5353249	—	—	Anderson et al. (2017)	24,922,104	53,921	40,186	1,494

NOTE.—New transcriptomes are denoted in bold.

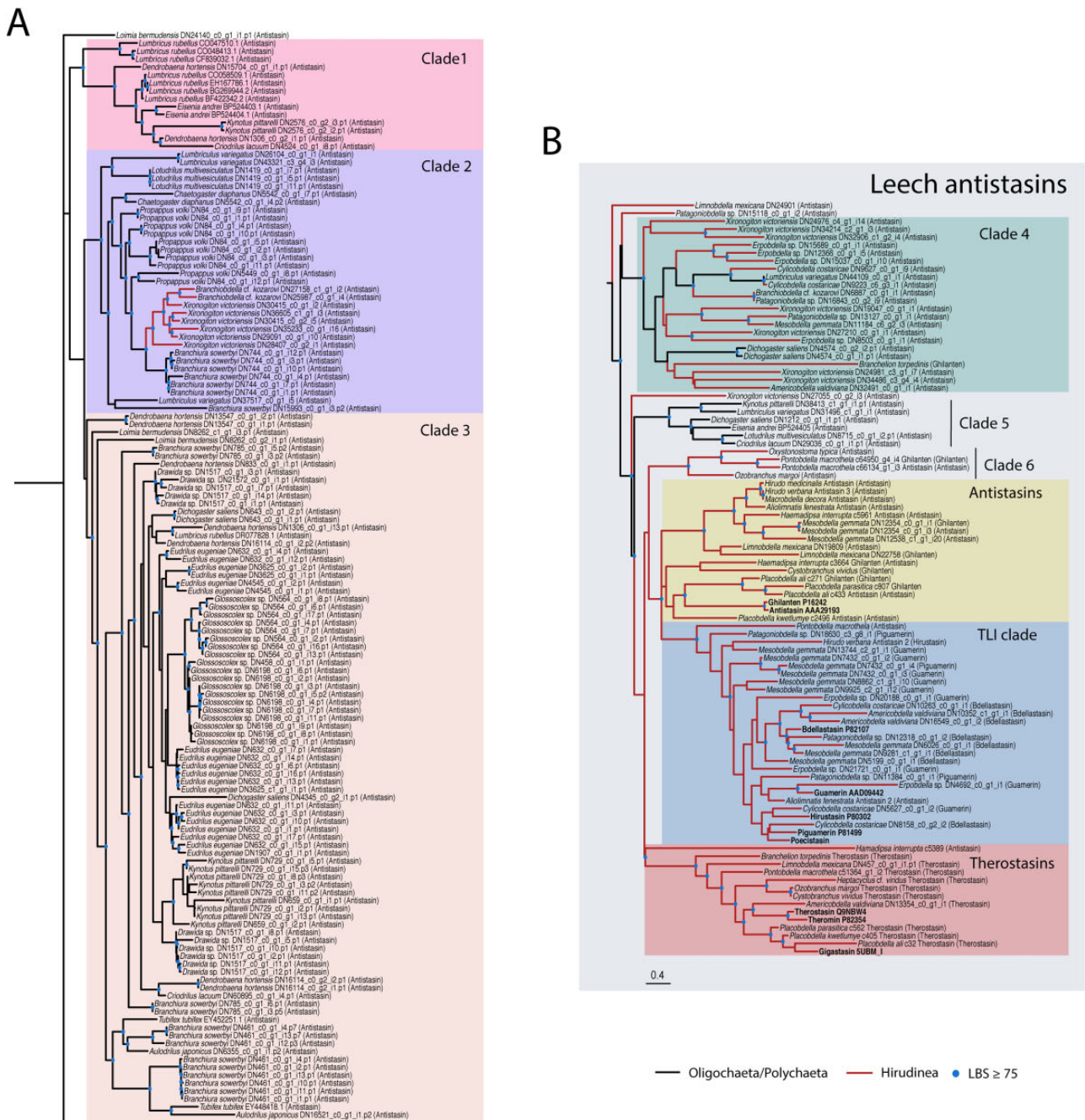
## Results

Raw sequences generated for this study and their respective assemblies are deposited in the SRA and the Transcriptome Shotgun Assembly (TSA) Sequences Database (BioProject accession number: PRJNA670722); SRA and TSA accession numbers are available in [table 1](#). Over 30,000,000 raw sequence reads were generated and 151,159 contigs were assembled on average for each of the nine new transcriptomes. Additionally, we included in our analyses both oligochaete ESTs deposited in GenBank, annotated transcriptomic data from previous studies of leech anticoagulants and additional annelid transcriptomes available on SRA (see [table 1](#) for statistics). Transdecoder predicted 85,121 ORFs on average for each of the new transcriptomes (a total of 57 of these found matches against one of antistasin, therostasin, guamerin, piguamerin, or bdellastasin and, at the same time, possessed a predicted antistasin-like domain). No hits against antistasin-like proteins were found for the *Acanthobdella peledina* transcriptome. Transdecoder predicted a total of 14,188 ORFs for the ESTs and 51,743 ORFs on average for each of the SRA transcriptomes and a total of 141 EST sequences showed significant matches against one of the aforementioned proteins. The final antistasin-like data set was composed of 232 sequences from the new and SRA transcriptomes, sequences from previous leech studies, and oligochaete ESTs.

## Gene Tree

The final alignment for the data set exclusively formed by annelid sequences included 2,289 aligned sites. The best scoring ML tree had a log likelihood ( $\ln L$ ) of  $-62,971.666$  ([fig. 1A and B](#)). The tree was rooted at *L. bermudensis*—note that this is the first time that an antistasin gene tree has been rooted. In our analysis, all leech sequences fall within a larger clade denoted “leech antistastins” in [figure 1B](#), though this clade is not formed exclusively by leech sequences. Clades 1–3 are formed only by non-Hirudinea sequences, except for a few sequences from Branchiobdellida that fall within clade 2. The larger leech antistasin clade can be further subdivided into six clades, three of which contain archetypal leech sequences that were used to define them as antistastins, trypsin and leukocyte elastase inhibitors (TLIs), and therostastins. The remaining three clades (4–6) do not contain any archetypal sequences and, therefore, these remain undefined. Clade 4 is formed by sequences from leeches, branchiobdellidans, and oligochaetes. Clade 5 is formed only by oligochaete sequences and is sister to larger clade formed by clade 6 and the three defined clades that contain the archetypal leech sequences.

The archetypal antistastins are sister to the archetypal ghilanten and these two anticoagulants are separated by very short branches. These fall within a clade formed by sequences



**Fig. 1.**—Best scoring ML tree resulting from the analysis of the putative antistatin-like proteins ( $ln L = -62,971.666$ ): (A) best scoring ML tree with Leech antistatins clade collapsed; (B) Leech antistatins clade of the best scoring ML tree. Nodes with likelihood bootstrap support values above 75 are denoted by blue circles. Branches are drawn proportionately to change. Archetypal leech bioactive proteins are denoted in bold. Red and black branches represent leech and non-leech sequences, respectively.

recorded from both glossiphoniids and non-proboscis bearing leeches. The TLI clade is sister to the antistasin clade and is formed by five archetypal sequences: bdellastasin, guamerin, hirustasin, piguamerin, and poecistasin. The archetypal therostasin, theromin, and gigastasin place within a monophyletic group (Therostasins) as sister to clade 6, antistasins and the TLI clade. Therostasin and theromin are recovered as sister to each other in our analysis, despite their functional divergence (see Introduction).

The tree resulting from the analysis that included putative non-annelid antistasin-like sequences is shown in [supplementary figure S4, Supplementary Material](#) online. The inclusion of these sequences and the rooting at *Hydra vulgaris* resulted in the paraphyly/polyphyly of clades that were previously reported monophyletic, including the leech antistasins clade. Therefore, we largely disregarded this hypothesis because of obvious artefactual issues (see Discussion).

### Motif and Domain Prediction

In total, 50 motifs (M1–M50) were predicted at an *e*-value superior to  $1E^{-10}$ . Figure 2 shows the distribution of 22 predicted motifs that are present among sequences from Hirudinea. M1–M12, M20, M40, and M50 are located within antistasin-like domains predicted by InterProScan, as illustrated in figure 3. M1 is present in almost all sequences included in the analysis except for six leech sequences that place in three separate clades in the gene tree. M2 and M3 are present in the majority of the oligochaete and branchiobdellid sequences (fig. 2A and B) but only three leech sequences in the antistasin clade possess M2 and only one sequence possesses M3 (fig. 2C). Three motifs (M26, M33, and M39) are exclusive or almost exclusive to certain clades. M26 and M33 are exclusive to the TLI and therostasin clades, respectively (fig. 2D and E). M39 occurs in the majority of the antistasin clade (fig. 2C), including in archetypal antistasin and ghilanten, but it is also present in sequences from the therostasin clade, including gigastasin. The only shared motifs between the archetypal theromin and therostasin is the M6, which corresponds to an antistasin-like domain.

### Discussion

Anticoagulants are important evolutionary innovations insofar as they allow bloodfeeders to feed efficiently. The results presented here provide much needed data for the understanding of the origin and evolution of antistasin-like anticoagulants in leeches. The inclusion of oligochaete and polychaete sequence data for antistasin-like proteins in a phylogenetic context provides an objective root to the gene tree and allowed the recognition of the major protein-level changes in each group of orthologous proteins. Additionally, the detailed analyses of conserved motifs allowed for the comparison of composition and motif number

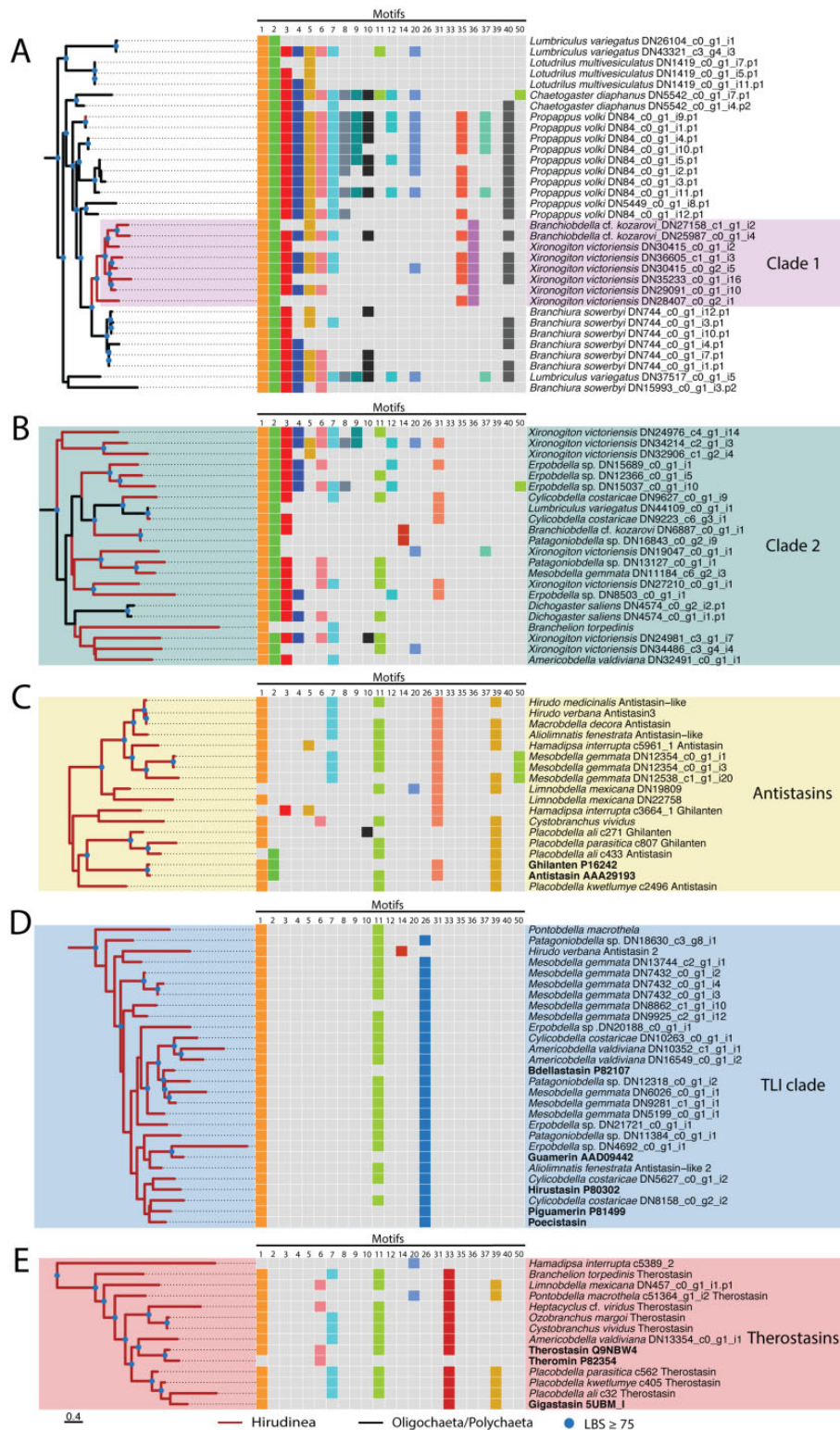
in antistasin-like domains across orthologous groups and granted the identification of major changes within antistasin-like sequences across evolutionary time. Our phylogenetic analysis suggests that the archetypal antistasin-like proteins isolated from leech saliva belong to three orthologous groups: 1) TLIs, 2) antistasins, and 3) therostasins. Beyond this, two additional clades (clades 4 and 6 in fig. 1B) likely represent new anticoagulants that have yet to be named. Without formal functional analyses, however, we refrain from naming these putative anticoagulants here.

Although antistasin-like proteins have been recorded in non-annelid taxa, previous studies on the evolution and classification of anticoagulants have left gene trees unrooted due to the lack of comparative data from outgroup taxa. The issue of long branches and random rooting is discussed in greater detail elsewhere (e.g., Wheeler 1990; Graham et al. 2002; Rosenfeld et al. 2012). However, the position of the root within an unrooted tree can render clans (equivalent to monophyletic group in an unrooted tree; see Wilkinson et al. 2007) paraphyletic if the root falls within a specific clade. Therefore, orthology determination and a nomenclatural system based on phylogenetic relationships are dependent on the root of a tree. Moreover, the lack of a root has so far prevented any conclusion about direction of evolution within leech anticoagulants. We rooted our tree at a sequence from *L. bermudensis*, a polychaete species that belongs to the family Terebellidae and included oligochaete sequence data to provide a meaningful outgroup. We also explored the possibility of a non-annelid root, but the resulting tree topology requires at least two extra origins of leech antistasin-like proteins and is discordant with previous studies (Amorim et al. 2015; Iwama et al. 2019; Kvist et al. 2020).

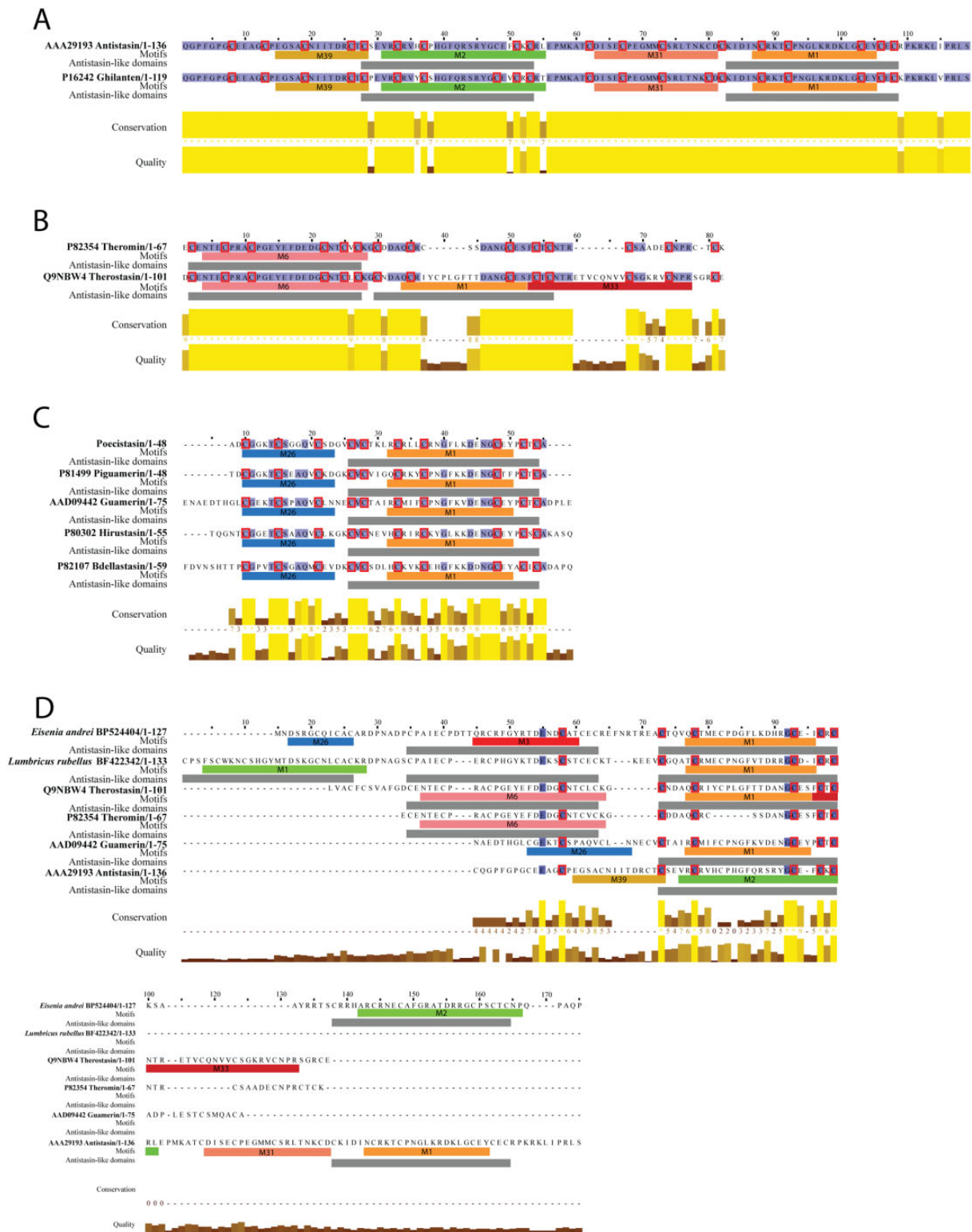
### Classification of Leech Antistasin-like Proteins

The classification of leech anticoagulants is fraught with problems. A clear example of this is the classification and separation of hirudin and hirudin-like factors (Müller et al. 2016, 2017, 2019). Hirudin-like factors share several compositional characteristics with hirudin, making them a target for BLAST-based orthology determination, yet all but one lack the critical capacity for anticoagulation. As a result, Müller et al. (2020) suggested a classification system for hirudin and hirudin-like factors that hinges entirely on the function of these proteins. However, this system does not reflect historical events that have led to the diversification of the proteins. Herein, we suggest a classification scheme for antistasin-like proteins that is based on orthology as inferred by phylogenetic analysis (fig. 1).

Our results corroborate the findings of Iwama et al. (2019), and we suggest the following names for the three major groups of leech antistasin-like proteins: 1) the TLI clade, 2) the antistasin clade, and 3) the therostasin clade. The TLI clade is formed by guamerin, piguamerin, bdellastasin, hirustasin,



**Fig. 2.**—Motif distribution on the best scoring ML tree based on the MEME suite. (A) Clade 1; (B) Clade 2; (C) Antistasin clade; (D) TLI clade; (E) Therostasin clade. Each color denotes a different motif. Archetypal leech bioactive proteins are denoted in bold. Red and black branches represent leech and non-leech sequences, respectively



**Fig. 3.**—Alignment of antistasin-like proteins isolated from Oligochaeta and Hirudinea: (A) archetypal antistasin (AAA29193) and ghilanten (P16242); (B) archetypal theromin (P82354) and therostasin (Q9NBW4); (C) archetypal poecistasin, piguamerin (P81499), guamerin (AAD09442), hirustasin (P80302), and bdelastasin (P82107); and (D) contig BP524404 from *Eisenia andrei*, BF422342 from *Lumbricus rubellus*, therostasin (Q9NBW4), theromin (P82354), guamerin (AAD09442), and antistasin (AAA29193). See text for further discussion.



and poecistasin. Whereas guamerin is a leukocyte elastase inhibitor, bdellastasin, hirustasin, piguamerin, and poecistasin are strong inhibitors of trypsin and plasmin (Söllner et al. 1994; Kim and Kang 1998; Moser et al. 1998; Rester et al. 1999; Tang et al. 2018).

The antistasin clade is formed by archetypal antistasin, isolated from the leech *Haementeria officinalis* de Filippi 1849; ghilanten, isolated from *Haementeria ghilianii* de Filippi, 1849; and sequences from both proboscis- and non-proboscis-bearing leeches (fig. 1B). Previous studies have pointed to the sister relationship (with very short branch lengths), as well as the high sequence (95.8%) and functional similarity between these two antistasin-like archetypal anticoagulants, and have suggested their synonymization (Kvist et al. 2014; Amorim et al. 2015; Iwama et al. 2019). However, the lack of both a root for the antistasin-like gene tree and sufficient comparative data across the diversity of Clitellata has prevented the determination of orthologous groupings within this family of proteins and, therefore, the circumscription of a single name for these highly similar proteins. Our phylogenetic analysis confirms the sister relationship of antistasin and ghilanten and the motif prediction re-iterates the similarity in primary structure. Moreover, previous studies demonstrated that they both inhibit hemostasis by targeting factor Xa (Nutt et al. 1988; Condra et al. 1989; Dunwiddie et al. 1989; Blankenship et al. 1990; Lapatto et al. 1997). As a result of our findings, we suggest the formal synonymization of antistasin and ghilanten: Antistasin has seniority and should therefore be the formal name used for this group of anticoagulants.

In our rooted tree, as in Iwama et al. (2019), therostasins form a monophyletic group, separate from the other clades in the tree (fig. 1B). Archetypal therostasin and theromin are sister to each other, as reported by Amorim et al. (2015). However, unlike antistasin and ghilanten, these two anticoagulants are functionally divergent. Therostasin is a strong factor Xa inhibitor and theromin is a thrombin inhibitor (Chopin et al. 2000; Salzet et al. 2000). Moreover, the genetic distance between these two anticoagulants (amino acid similarity of 79.7%) is much higher than that between antistasin and ghilanten (95.8%). Also, the motif analysis revealed a lack of M1 and M33 (fig. 2E) in the theromin sequence. This motif is conserved in all previously identified members of the therostasin clade (except for the early diverging *Haemadipsa interrupta* (Moore, 1935) sequence). Therefore, the distribution of motifs within the therostasin clade supports the nomenclatural differentiation between therostasins and theromin. Our results also necessitate a broadening of the definition for antistasin-like proteins to include direct thrombin inhibitors, as well as trypsin, leukocyte elastase, plasmin, and factor Xa inhibitors.

Clades 1–6 do not possess any well-characterized proteins and do not all have strong nodal support. Because of the low nodal support for some of these clades, it remains impossible

to robustly infer whether or not these are antistasin-like proteins, therostasin-like proteins or represent a new lineage of anticoagulants; therefore, we choose not to name these clades.

### Functional Diversity of Antistasin-like Proteins in Leeches

The TLI clade is formed by archetypal anticoagulants that have been functionally characterized, although their role in anti-hemostasis in leeches is still not completely understood. Guamerin is the only archetypal protein within this clade that does not inhibit trypsin or plasmin. Instead, guamerin is a leukocyte elastase inhibitor, which inhibits several neutrophil serine proteases related to the inflammatory process and has a potential to influence immunological responses (Jung et al. 1995). Bdellastasin, poecistasin, hirustasin, and piguamerin are strong inhibitors of plasmin and trypsin. Plasmin is a key component of the fibrinolysis pathway, responsible for the degradation of several blood plasma proteins, including fibrin and fibrinogen that form the blood clot. Therefore, it seems counterintuitive to suggest that antiplasmin proteins are used by leeches to inhibit hemostasis. However, these four archetypal anticoagulants also inhibit trypsin. The effect of the serine protease trypsin in blood coagulation has long been described and it is analogous to factor Xa (Ferguson et al. 1960). In fishes, trypsin secretion is localized in nasopharyngeal mucosal and gill tissues, and its effect on hemostasis in blood of gills has been demonstrated (Kim et al. 2009). Nonetheless, the effect of trypsin inhibition in hemostasis has yet to be robustly demonstrated for other animal taxa, but an inhibitory impact on factor IX, factor VII, and/or factor V (such as in fishes) would suggest that they play important roles in antihemostasis (Furie and Furie 2005; Kim et al. 2009; Tavares-Dias and Oliveira 2009).

Our alignment of antistasin and ghilanten shows that differences in primary structure between these two anticoagulants are concentrated in their N-terminal domains (fig. 3A), including their reactive sites contained in M2. Previous studies have linked changes in the reactive sites and the exosite bindings of antistasin to subtle changes in anticoagulatory power (Dunwiddie et al. 1989; Blankenship et al. 1990).

The alignment of the sequences for (the functionally divergent) therostasin and theromin (fig. 3B) revealed relatively high sequence similarity (79.7%) except in the C-terminal region, which possesses a gap formed by 8 amino acid residues (between residues 53 and 54) in theromin. Interestingly, this gap region is aligned with M33, which is otherwise present in all other therostasins included in this study, but is not present in theromin (fig. 2E). This suggests a deletion event affecting this motif in theromin. The biological significance, if any, of M33 is still unknown, but our tree indicates that the origin of this motif occurred within the therostasin clade and that it was subsequently lost in theromin. This loss could either be a consequence of the accumulation of mutations following

gene duplication or by expression of different splice variants, a common source of toxin variation in other animal taxa (Siigur et al. 2001; Fry et al. 2010; Haney et al. 2019). Another gap is present in the C-terminal region of theromin, between residues 37 and 38. These residues are included in the putative active site of therostasin, suggesting that this region is degraded in theromin; this is further supported by an amino acid substitution (I/C) in this region of the theromin sequence.

### Origin of Leech Salivary Antistasin-like Proteins

Although the archetypal antistasin was originally described from the glossiphoniid leech *Haementeria officinalis* (Nutt et al. 1988; Condra et al. 1989; Dunwiddie et al. 1989), several studies have recorded antistasin-like serine proteases in non-leech taxa, such as oligochaetes, polychaetes, and cnidarians (Holstein et al. 1992; Lee et al. 2010; Mehr et al. 2015). Their function is still largely unknown but it has been suggested that they might be related to physiological processes other than antihemostasis (Kwak et al. 2019), such as immune response (Nikapitiya et al. 2010). Therefore, there seems to be an obvious functional change between non-leech antistasin-like proteins and leech salivary antistasin-like proteins that may be related to bloodfeeding. Such functional change opens up questions about the origin and evolution of anticoagulatory capabilities within the antistasin family of serine protease inhibitors, including structural and sequence changes that are involved in the inhibition of hemostasis.

The presence of non-leech antistasin-like sequences in clades 4 and 5 necessarily invokes at least two independent origins of leech antistasin-like proteins: one independent origin for clade 4 and one for the factor Xa and TLI inhibitors. Clades 4 and 5 lack functionally characterized anticoagulants, such that the proteins may not be involved in anticoagulation or bloodfeeding. In fact, BLAST searches of the final antistasin-like data set (formed exclusively by antistasin-like domains) recovered hits against the leech cocoon protein described by Mason et al. (2004) that is formed by six repeats of antistasin-like and NOTCH domains (Mason et al. 2006). Moreover, a recent study has suggested that proteins with antistasin-like domains might be involved in segment polarity signaling (Kwak et al. 2019). By contrast, archetypal anticoagulants in the TLI and the factor Xa inhibitor clades are produced and secreted by leech salivary gland cells and have been shown to possess anticoagulatory capabilities. The topology shown in figure 1 suggests a single origin of TLIs and factor Xa inhibitors, and that the ancestor of Hirudinida was capable of inhibiting factor Xa, trypsin, and/or leukocyte elastase. In the last decade, it has been proposed that the presence of antistasin, hirudins, mannase, piguamerin, bdellastasin, destabilase, and leech-derived tryptase inhibitors in modern leeches is derived from orthologs present in the ancestor of Hirudinida (Min et al. 2010; Kvist et al. 2011,

2013; Siddall et al. 2016; Tessler, Marancik, et al. 2018; Iwama et al. 2019; Babenko et al. 2020). Nevertheless, these analyses lacked robust orthology determination methodology and/or an objective root in the trees. These methodological difficulties impose limitations on any evolutionary conclusion regarding the origin of anticoagulants.

### Evolution of Leech Antistasin-like Anticoagulants

Holstein et al. (1992), in the description of the *Hydra* antistasin-like serine protease inhibitor, noticed that the sequence was composed of 6-fold internal repeats of 25 and 26 amino acids and that these internal repeats shared the same distribution of cysteines with the C-terminal region of the two antistasin-like domains in the archetypal leech antistasin. Therefore, they suggested that members of the antistasin-like family of serine proteases evolved by duplication of the repeated internal folds that form the domains in the archetypal antistasin. In our analyses, all oligochaete antistasin-like sequences possess multiple domains that are shared by all sequences in non-leech clades. These results suggest that the ancestral leech antistasin was a multidomain protein (varying between two and six domains). On the other hand, sequences in the TLI clade, as well as in the antistasin and therostasin clades, exhibit a reduction in number of domains predicted by InterProScan and the motif analysis suggests that the antistasin-like domain present in leech sequences predate bloodfeeding and the diversification of Hirudinida (figs. 2 and 3D).

Despite efforts to elucidate the evolution and origin of leech anticoagulants, few studies have focused on specific protein-level changes that resulted in anticoagulatory activity. One of the only recent attempts demonstrated the importance of specific residues in different regions of hirudin and hirudin-like factors in regards to anticoagulatory activity (Müller et al. 2020). However, the direction of changes within each orthologous group was not elucidated due to the lack of a phylogenetic hypothesis, much less a proper root for the tree. In hirudins and hirudin-like factors, this task is particularly difficult because these proteins seem to be restricted to Hirudiniformes, which are thought to have a bloodfeeding ancestor (Iwama et al. 2019). Unlike hirudins and hirudin-like factors, antistasin-like proteins are widely distributed within leeches, as well as their relatives. However, most proteins belonging to the antistasin family possess a variable number of antistasin-like domains repeated in tandem (fig. 3D), which complicates comparisons between domains in different antistasin-like proteins, because orthology between domains is difficult to establish. In the present study, this problem was partially alleviated by comparing the distribution and organization of motifs among antistasin-like domains.

Sequence motifs are a common resource for establishing homology between distantly related domains, and some important tools for protein and domain identification (e.g., PROSITE profiles and PRINTS) are based on motif identification (Attwood et al. 1994; Sigrist et al. 2002; Attwood et al. 2012; Karlin and Belshaw 2012; Gerdol et al. 2019). Moreover, several studies attempt to establish tentative functional characterization based on the presence and organization of certain motifs that are associated with a specific function within a family of proteins (e.g., Fang et al. 2016; Müller et al. 2016; Wang et al. 2019).

Our motif analyses found homology between domains belonging to different orthologous groups of leech and oligochaete antistatin-like proteins. In the archetypal antistatin and ghilanten, the two tandem domains seem to be homologous to two domains present in the oligochaete sequences (fig. 3D). The first domain is characterized by the presence of M2 and the second domain by the presence of M1. Furthermore, the N-terminal region of each domain seems to be modified in relation to the oligochaete sequences, with the acquisition of M31 and M39 in the leech sequences, respectively. The alignment shown in figure 3D failed to capture the putative homologous relationship between domains in the antistatin and the oligochaete sequences. This might be due to artifacts in the motif analysis, because we expect a high similarity between tandem domains. Another explanation is a major domain rearrangement during the evolution of antistatin-like proteins, a common process in protein evolution (Björklund et al. 2005; Bornberg-Bauer et al. 2005). The archetypal proteins in the TLI clade are all formed by a single domain characterized by the presence of M1 and M26 in a N-terminal position, suggesting a homologous relationship between domains in the oligochaete sequences that are also characterized by M1 and an acquisition of M26 in the TLI clade. The archetypal therostatin is formed by two antistatin-like domains, characterized by M1 and M6. Moreover, the presence of M4 in the C-terminal region of the second domain represents putative modifications in the therostatin clade. Unlike the other members of the therostatin clade, in theromin, M6 is the only domain predicted. The sequence alignment of therostatin and theromin demonstrates that the majority of changes in theromin are located in the C-terminal region of the alignment, within the second antistatin-like domain (fig. 3B). This may explain why M1 is not predicted to be present in the second domain of theromin.

Although the motif analysis and domain prediction successfully identified changes within each orthologous group of leech antistatin-like proteins, their role in hemostasis inhibition (if any) is still unknown. Therefore, it is still not possible to directly link these motifs and domain number to anticoagulatory activity and future efforts should focus on functionally characterizing antistatin-like proteins in order to better understand the transitions related to bloodfeeding and the

significance of each identified motif in the inhibition of hemostasis.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

Many thanks to José J. Nuñez, Bernald Pacheco-Chaves, Stanisław Cios, and others for help with collecting and providing specimens for this project; it is truly appreciated. This project was chiefly funded by a National Science Foundation Doctoral Dissertation Improvement grant (#1601562 awarded to M.E.S. [P.I.] and M.T. [Co-PI]). Expeditionary work in Chile was provided by Explorers Club Exploration Fund Grant (to M.T.). Partial funding came from a National Sciences and Engineering Research Council of Canada Discovery grant, Royal Ontario Museum Peer Review grant, and a stipend from the Olle Engkvist Byggnästars Foundation (to S.K.). We thank Danielle de Carle, Kevin Anderson, and two anonymous reviewers for valuable help with early drafts of this manuscript.

## Data Availability

The data underlying this article are available in its [supplementary material](#). Raw sequences generated and their respective assemblies are deposited in the Sequence Read Archive (SRA) and the Transcriptome Shotgun Assembly (TSA) Database at NCBI and can be accessed under the BioProject number PRJNA670722.

## Literature Cited

- Amorim AMXP, et al. 2015. Transcripts involved in hemostasis: exploring salivary complexes from *Haementeria vizottoi* leeches through transcriptomics, phylogenetic studies and structural features. *Toxicon* 106:20–29.
- Anderson FE, et al. 2017. Phylogenomic analyses of Crassidictylozoa support major Northern and Southern Hemisphere clades and a Pangaeian origin for earthworms. *BMC Evol Biol.* 17(1):1–123.
- Apakupakul K, Siddall ME, Burreson EM. 1999. Higher level relationships of leeches (Annelida: Clitellata: Euhirudinea) based on morphology and gene sequences. *Mol Phylogenet Evol.* 12(3):350–359.
- Assumpção TCF, et al. 2008. An insight into the sialome of the blood-sucking bug *Triatoma infestans*, a vector of Chagas' disease. *Insect Biochem Mol Biol.* 38(2):213–232.
- Attwood TK, Beck ME, Bleasby AJ, Parry-Smith DJ. 1994. PRINTS—a database of protein motif fingerprints. *Nucleic Acids Res.* 22:3590–3596.
- Attwood TK, et al. 2012. The PRINTS database: a fine-grained protein sequence annotation and analysis resource—its status in 2012. *Database* 2012:bas019.
- Babenko VV, et al. 2020. Draft genome sequences of *Hirudo medicinalis* and salivary transcriptome of three closely related medicinal leeches. *BMC Genomics.* 21(1):503.

- Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. *Nucleic Acids Res.* 43(W1):W39–W49.
- Björklund ÅK, Ekman D, Light S, Frey-Skött J, Elofsson A. 2005. Domain rearrangements in protein evolution. *J Mol Biol.* 353(4):911–923.
- Blankenship DT, Brankamp RG, Manley GD, Cardin AD. 1990. Amino acid sequence of ghilanten: anticoagulant-antimetastatic principle of the South American leech, *Haementeria ghilianii*. *Biochem Biophys Res Commun.* 166(3):1384–1389.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
- Borda E, Siddall ME. 2004. Arhynchobdellida (Annelida: Oligochaeta: Hirudinida): phylogenetic relationships and evolution. *Mol Phylogenet Evol.* 30(1):213–225.
- Bornberg-Bauer E, Beaussart F, Kummerfeld SK, Teichmann SA, Weiner J. 2005. The evolution of domain arrangements in proteins and interaction networks. *Cell Mol Life Sci.* 62(4):435–445.
- Chopin V, et al. 2000. Therostasin, a novel clotting factor Xa inhibitor from the rhynchobdellid leech, *Theromyzon tessulatum*. *J Biol Chem.* 275(42):32701–32707.
- Condra C, et al. 1989. Isolation and structural characterization of a potent inhibitor of coagulation factor Xa from the leech *Haementeria ghilianii*. *Thromb Haemost.* 61(03):437–441.
- Dunwiddie C, et al. 1989. Antistatin, a leech-derived inhibitor of factor Xa. Kinetic analysis of enzyme inhibition and identification of the reactive site. *J Biol Chem.* 264(28):16694–16699.
- Elliott JM, Kutschera U. 2011. Medicinal leeches: historical use, ecology, genetics and conservation. *Freshw Rev.* 4(1):21–41.
- Erséus C, Källersjö M. 2004. 18S rDNA phylogeny of Clitellata (Annelida). *Zool Scr.* 33:187–196.
- Erséus C, et al. 2020. Phylogenomic analyses reveal a Palaeozoic radiation and support a freshwater origin for clitellate annelids. *Zool Scr.* 49(5):614–640.
- Fang Y, et al. 2016. De novo transcriptome analysis reveals distinct defense mechanisms by young and mature leaves of *Hevea brasiliensis* (Para Rubber Tree). *Sci Rep.* 6:1–10.
- Ferguson BJH, Wilson EG, Iatridis SG, Rierson HA, Johnston BR. 1960. Enzymes and blood clotting. I. Trypsin as an accessory factor. *J Clin Invest.* 39(12):1942–1952.
- Finn RD, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44:D279–D285.
- Francischetti IMB. 2010. Platelet aggregation inhibitors from hematophagous animals. *Toxicon* 56(7):1130–1144.
- Fry BG, et al. 2010. Novel venom proteins produced by differential domain-expression strategies in beaded lizards and gila monsters (genus *Heloderma*). *Mol Biol Evol.* 27(2):395–407.
- Furie BBC, Furie BBC. 2005. Review series thrombus formation in vivo. *J Clin Investig.* 115(12):3355–3362.
- Gerdol M, Cervelli M, Oliverio M, Modica MV. 2018. Piercing fishes: porin expansion and adaptation to hematophagy in the vampire snail *Cumia reticulata*. *Mol Biol Evol.* 35(11):2654–2668.
- Gerdol M, et al. 2019. A recurrent motif: diversity and evolution of ShKT domain containing proteins in the vampire snail *Cumia reticulata*. *Toxins (Basel)* 11(2):106–118.
- Graça-Souza AV, et al. 2006. Adaptations against heme toxicity in blood-feeding arthropods. *Insect Biochem Mol Biol.* 36(4):322–335.
- Graham SW, Olmstead RG, Barrett SCH. 2002. Rooting phylogenetic trees with distant outgroups: a case study from the Commelinoid monocots. *Mol Biol Evol.* 19(10):1769–1781.
- Haas BJ, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8(8):1494–1512.
- Haney RA, Matte T, Forsyth FAS, Garb JE. 2019. Alternative transcription at venom genes and its role as a complementary mechanism for the generation of venom complexity in the common house spider. *Front Ecol Evol.* 7:1–13.
- Hanumanthaiah R, Day K, Jagadeeswaran P. 2002. Comprehensive analysis of blood coagulation pathways in Teleostei: evolution of coagulation factor genes and identification of Zebrafish factor VIII. *Blood Cells Mol Dis.* 29(1):57–68.
- Herlin C, et al. 2017. Leech therapy in flap salvage: systematic review and practical recommendations. *Ann Chir Plast Esthet.* 62(2):e1–e13.
- Holstein TV, et al. 1992. The primitive metazoan *Hydra* expresses antistatin, a serine protease inhibitor of vertebrate blood coagulation: cDNA cloning, cellular localisation and developmental regulation. *FEBS Lett.* 309(3):288–292.
- Iwama RE, Ocegüera-Figueroa A, Giribet G, Kvist S. 2019. The salivary transcriptome of *Limnodynastes maculosa* (Annelida: Clitellata: Praobdellidae) and orthology determination of major leech anticoagulants. *Parasitology* 146(10):1338–1346.
- Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.
- Jung HI, Kim SI, Ha KS, Joe CO, Kang KW. 1995. Isolation and characterization of guamerin, a new human leukocyte elastase inhibitor from *Hirudo nipponia*. *J Biol Chem.* 270(23):13879–13884.
- Karlin D, Belshaw R. 2012. Detecting remote sequence homology in disordered proteins: discovery of conserved motifs in the N-termini of mononegavirales phosphoproteins. *PLoS One* 7(3):e31719.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Kersch CN, Pietrantonio PV. 2011. Mosquito *Aedes aegypti* (L.) leucokinin receptor is critical for in vivo fluid excretion post blood feeding. *FEBS Lett.* 585(22):3507–3512.
- Kim DR, Kang KW. 1998. Amino acid sequence of piguamerin, an antistatin-type protease inhibitor from the blood sucking leech *Hirudo nipponia*. *Eur J Biochem.* 254(3):692–697.
- Kim S, Carrillo M, Kulkarni V, Jagadeeswaran P. 2009. Evolution of primary hemostasis in early vertebrates. *PLoS One* 4(12):e8403.
- Koh Y, Kini R. 2009. Molecular diversity of anticoagulants from hematophagous animals. *Thromb Haemost.* 102(09):437–453.
- Kvist S, Brugler MR, Goh TG, Giribet G, 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. *Invertebr Biol.* 133(1):74–98.
- Kvist S, Manzano-Marín A, de Carle D, Trontelj P, Siddall ME. 2020. Draft genome of the European medicinal leech *Hirudo medicinalis* (Annelida, Clitellata, Hirudiniformes) with emphasis on anticoagulants. *Sci Rep.* 10:9885.
- Kvist S, Min GS, Siddall ME. 2013. Diversity and selective pressures of anticoagulants in three medicinal leeches (Hirudinida: Hirudinidae, Macrobdellidae). *Ecol Evol.* 3(4):918–933.
- Kvist S, Sarkar IN, 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. *Invertebr Biol.* 130(4):344–350.
- Kvist S, et al. 2017. When predator becomes prey: investigating the salivary transcriptome of the shark-feeding leech *Pontobdella macrothela* (Hirudinea: Piscicolidae). *Zool J Linn Soc.* 179:725–737.
- Kwak HJ, Park JS, Jiménez BIM, Park SC, Cho SJ. 2019. Spatiotemporal expression of anticoagulation factor *Antistatin* in freshwater leeches. *Int. J. Mol. Sci.* 20(16):3994.
- Lapatto R, et al. 1997. X-ray structure of antistatin at 1.9 Å resolution and its modelled complex with blood coagulation factor Xa. *EMBO J.* 16(17):5151–5161.
- Lee MS, et al. 2010. Eisenstasin, new antistatin family inhibitor from the earthworm. *Biologia* 65(2):284–288.

- Mans BJ, Neitz AWH. 2004. Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. *Insect Biochem Mol Biol.* 34(1):1–17.
- Manzano-Marín A, Ocegüera-Figueroa A, Latorre A, Jiménez-García LF, Moya A. 2015. Solving a bloody mess: B-vitamin independent metabolic convergence among gammaproteobacterial obligate endosymbionts from blood-feeding arthropods and the leech *Haementeria officinalis*. *Genome Biol Evol.* 7(10):2871–2884.
- Mason TA, McIlroy PJ, Shain DH. 2004. A cysteine-rich protein in the *Theromyzon* (Annelida: Hirudinea) cocoon membrane. *FEBS Lett.* 561(1–3):167–172.
- Mason TA, McIlroy PJ, Shain DH. 2006. Structural model of an antistatin/notch-like fusion protein from the cocoon wall of the aquatic leech, *Theromyzon tessulatum*. *J Mol Model.* 12(6):829–834.
- Mehr S, et al. 2015. Transcriptome sequencing and annotation of the polychaete *Hermodice carunculata* (Annelida, Amphinomidae). *BMC Genomics.* 16(1):1–13.
- Min G-S, Sarkar IN, Siddall ME. 2010. Salivary transcriptome of the North American medicinal leech, *Macrobdella decora*. *J Parasitol.* 96(6):1211–1221.
- Mittl PRE, et al. 1997. A new structural class of serine protease inhibitors revealed by the structure of the hirustasin-kallikrein complex. *Structure* 5(2):253–264.
- Mizurini DM, et al. 2013. Desmolaris, a novel factor Xla anticoagulant from the salivary gland of the vampire bat (*Desmodus rotundus*) inhibits inflammation and thrombosis in vivo. *Blood* 122(25):4094–4106.
- Moser M, et al. 1998. Bdelellastatin, a serine protease inhibitor of the antistatin family from the medical leech (*Hirudo medicinalis*): primary structure, expression in yeast, and characterisation of native and recombinant inhibitor. *Eur J Biochem.* 253(1):212–220.
- Müller AC, Lukas P, Lemke S, Hildebrandt J. 2019. Hirudin and decorsins of the North American medicinal leech *Macrobdella decora*: gene structure reveals homology to hirudins and hirudin-like factors of Eurasian medicinal leeches. *J Parasitol.* 105(3):423–431.
- Müller C, Haase M, Lemke S, Hildebrandt JP. 2017. Hirudins and hirudin-like factors in Hirudinidae: implications for function and phylogenetic relationships. *Parasitol Res.* 116(1):313–325.
- Müller C, Lukas P, Böhmert M, Hildebrandt JP. 2020. Hirudin or hirudin-like factor—that is the question: insights from the analyses of natural and synthetic HLF variants. *FEBS Lett.* 594(5):841–850.
- Müller C, et al. 2016. More than just one: multiplicity of hirudins and hirudin-like factors in the medicinal leech, *Hirudo medicinalis*. *Mol Genet Genomics.* 291(1):227–240.
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Nikapitiya C, et al. 2010. Disk abalone (*Haliotis discus discus*) expresses a novel antistatin-like serine protease inhibitor: molecular cloning and immune response against bacterial infection. *Fish Shellfish Immunol.* 28(4):661–671.
- Nutt E, et al. 1988. The amino acid sequence of antistatin. *J Biol Chem.* 263(21):10162–10167.
- Oviedo M, et al. 2009. The salivary transcriptome of *Anopheles gambiae* (Diptera: Culicidae) larvae: a microarray-based analysis. *Insect Biochem Mol Biol.* 39(5–6):382–394.
- Petersen TN, Brunak S, Von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 8(10):785–786.
- Phillips AJ, Siddall ME. 2009. Poly-paraphyly of Hirudinidae: many lineages of medicinal leeches. *BMC Evol Biol.* 9(1):246.
- Potter IC, Gill HS. 2003. Adaptive radiation of lampreys. *J Great Lakes Res.* 29:95–112.
- Rester U, et al. 1999. Structure of the complex of the antistatin-type inhibitor bdelellastatin with trypsin and modelling of the bdelellastatin-microplasmin system. *J Mol Biol.* 293(1):93–106.
- Ribeiro JMC, Arcà B. 2009. From sialomes to the sialoverse. An insight into salivary potion of blood-feeding insects. *Adv Insect Phys.* 37:59–118.
- Ribeiro JMC, Martin-Martin I, Moreira FR, Bernard KA, Calvo E. 2018. A deep insight into the male and female sialotranscriptome of adult *Culex tarsalis* mosquitoes. *Insect Biochem Mol Biol.* 95:1–9.
- Rosenfeld JA, Payne A, DeSalle R. 2012. Random roots and lineage sorting. *Mol Phylogenet Evol.* 64(1):12–20.
- Rousset V, Pleijel F, Rouse GW, Erséus C, Siddall ME. 2007. A molecular phylogeny of annelids. *Cladistics* 23(1):41–63.
- Salzet M, Chopin V, Baert JL, Matias I, Malecha J. 2000. Theromin, a novel leech thrombin inhibitor. *J Biol Chem.* 275(40):30774–30780.
- Sawyer RT. 1986. *Leech biology and behavior*. Oxford: Clarendon Press.
- Siddall ME, Brugler MR, Kvist S. 2016. Comparative transcriptomic analyses of three species of Placobdella (Rhynchobdellida: Glossiphoniidae) confirms a single origin of blood feeding in leeches. *J Parasitol.* 102(1):143–150.
- Siddall ME, Min GS, Fontanella FM, Phillips AJ, Watson SC. 2011. Bacterial symbiont and salivary peptide evolution in the context of leech phylogeny. *Parasitology* 138(13):1815–1827.
- Siddall ME, et al. 2001. Validating Livanow: molecular data agree that leeches, branchiobdellidans, and *Acanthobdella peledina* form a monophyletic group of oligochaetes. *Mol Phylogenet Evol.* 21(3):346–351.
- Sigrist CJA, et al. 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. *Briefings Bioinf.* 3(3):265–274.
- Siigur E, Aaspõllu A, Siigur J. 2001. Sequence diversity of *Vipera lebetina* snake venom gland serine proteinase homologs—result of alternative-splicing or genome alteration. *Gene* 263(1–2):199–203.
- Smith TA, Driscoll T, Gillespie JJ, Raghavan R. 2015. A *Coxiella*-like endosymbiont is a potential vitamin source for the lone star tick. *Genome Biol Evol.* 7(3):831–838.
- Söllner C, Mentele R, Eckerskorn C, Fritz H, Sommerhoff CP. 1994. Isolation and characterization of hirustasin, an antistatin-type serine-proteinase inhibitor from the medical leech *Hirudo medicinalis*. *Eur J Biochem.* 219(3):937–943.
- Stiller J, Tilic E, Rousset V, Pleijel F, Rouse GW. 2020. Spaghetti to a tree: a robust phylogeny for Terebelliformia (Annelida) based on transcriptomes, molecular and morphological data. *Biology (Basel)* 9(4):1–28.
- Tang X, et al. 2018. Isolation and characterization of poecistatin, an anti-thrombotic antistatin-type serine protease inhibitor from Leech *Poecilobdella manillensis*. *Toxins (Basel)* 10(11):429–413.
- Tavares-Dias M, Oliveira SR. 2009. A review of the blood coagulation system of fish. *Rev Bras Biociências* 4849:205–224.
- Tessler M, de Carle D, et al. 2018. Worms that suck: phylogenetic analysis of Hirudinea solidifies the position of Acanthobdellida and necessitates the dissolution of Rhynchobdellida. *Mol Phylogenet Evol.* 127:129–134.
- Tessler M, Marancik D, et al. 2018. Marine leech anticoagulant diversity and evolution. *J Parasitol.* 104(3):210–220.
- Tsakiris DA, Scudder L, Hodivala-Dilke K, Hynes RO, Coller BS. 1999. Hemostasis in the mouse (*Mus musculus*): a review. *Thromb Haemostas.* 81(02):177–188.
- Utley DS, Koch RJ, Goode RL. 1998. The failing flap in facial plastic and reconstructive surgery: role of the medicinal leech. *Laryngoscope* 108(8):1129–1135.

- Wang Q, et al. 2019. Integrative transcriptomic and genomic analysis of odorant binding proteins and chemosensory proteins in aphids. *Insect Mol Biol.* 28(1):1–22.
- Wheeler WC. 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* 6(4):363–367.
- Wilkinson M, McInerney JO, Hirt RP, Foster PG, Embley TM. 2007. Of clades and clans: terms for phylogenetic relationships in unrooted trees. *Trends Ecol Evol.* 22(3):114–115.
- Wilson D, Madera M, Vogel C, Chothia C, Gough J. 2007. The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Res.* 35:308–313.
- Zhao Y, et al. 2020. Transcriptomic analysis of oligochaete immune responses to myxosporeans infection: *Branchiura sowerbyi* infected with *Myxobolus cultus*. *J Invertebr Pathol.* 169:1–13.

**Associate editor:** Lavrov Dennis