J. Parasitol., 102(1), 2016, pp. 143–150

COMPARATIVE TRANSCRIPTOMIC ANALYSES OF THREE SPECIES OF *PLACOBDELLA* (RHYNCHOBDELLIDA: GLOSSIPHONIIDAE) CONFIRMS A SINGLE ORIGIN OF BLOOD FEEDING IN LEECHES

Mark E. Siddall,* Mercer R. Brugler*†, and Sebastian Kvist‡

* Sackler Institute for Comparative Genomics, and Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024. Correspondence should be sent to: *siddall@amnh.org*

ABSTRACT: One of the recalcitrant questions regarding the evolutionary history of clitellate annelids involves the feeding preference of the common ancestor of extant rhynchobdellid (proboscis bearing) and arhynchobdellid (jaw bearing) leeches. Whereas early evidence, based on morphological data, pointed towards independent acquisitions of blood feeding in the 2 orders, molecular-based phylogenetic data suggest that the ancestor of modern leeches was a sanguivore. Here, we use a comparative transcriptomic approach in order to increase our understanding of the diversity of anticoagulation factors for 3 species of the genus *Placobdella*, for which comparative data have been lacking, and inspect these in light of archetypal anticoagulant data for both arhynchobdellid and other rhynchobdellid species. Notwithstanding the varying levels of host specificity displayed by the 3 different species of *Placobdella*, transcriptomic profiles with respect to anticoagulation factors were largely similar —this despite the fact that *Placobdella kwetlumye* only retains a single pair of salivary glands, as opposed to the 2 pairs more common in the genus. Results show that 9 different anticoagulant proteins and an additional 5 putative antihemostasis proteins are expressed in salivary secretions of the 3 species. In particular, an ortholog of the archetypal, single-copy, anticoagulant hirudin (not previously available as comparative data for rhynchobdellids) is present in at least 2 of 3 species examined, corroborating the notion of a single origin of blood feeding in the ancestral leech.

Both Arhynchobdellida and Rhynchobdellida, the 2 ordinal subdivisions of Hirudinida, harbor species that have evolved a behavioral preference for feeding exclusively on vertebrate blood, and hosts typically span a wide taxonomic range—mammals, amphibians, reptiles, and birds (Klemm, 1982; Sawyer, 1986; Oceguera-Figueroa et al., 2010; Kvist et al., 2011). It remains to be determined whether blood feeding evolved independently in the evolutionary history of leeches (Siddall and Burreson, 1995, 1996) or only once in the common ancestor of leeches (Siddall et al., 2011); members of the 2 orders comprise nonsanguivorous (e.g., Siddall and Budinoff, 2005) and sanguivorous taxa, with the latter solving the problem of blood feeding in contrasting ways. Whereas arhynchobdellid leeches use denticulate jaws in a serrating motion to create an incisional wound on the surface of their host, rhynchobdellid leeches possess a muscular tubular proboscis that is inserted subcutaneously into their prey (Sawyer, 1986; Siddall and Burreson, 1996). In addition, both arhynchobdellid and rhynchobdellid leeches possess repertoires of salivary gland peptides that are expressed during feeding to prevent the coagulation of the host blood. These bioactive peptides underlie the extensive use of leeches as an alternative to phlebotomy in humoralistic medical practices that prevailed prior to the 20th century (Fields, 1991; Whitaker et al., 2004). Leeches have found use in modern medicine, specifically in postsurgical applications following flap and replantation surgery, during a critical interval when the venous blood flow is far inferior to arterial blood flow (e.g., Markwardt, 2002; Mumcuoglu, 2014). Diverse leech anticoagulants antagonize different pathways of the coagulation cascade, in a manner for which the cascade may not have an opportunity to compensate regarding inefficiency of any 1 procoagulant (see Kvist et al., 2013).

The rhynchobdellid genus *Placobdella* (Clitellata: Glossiphoniidae) encompasses 22 currently recognized species (López-Jiménez and Oceguera-Figueroa, 2009; Moser et al., 2014). All members of the genus are sanguivorous and most are found feeding on freshwater turtles (Klemm, 1972; Brooks et al., 1990; Siddall and Gaffney, 2009; Richardson et al., 2010), with fewer being specific to amphibian hosts (McCallum et al., 2011; Oceguera-Figueroa, 2012). *Placobdella parasitica* (Say, 1824) is widely distributed in North America as a semipermanent ectoparasite on a wide range of turtle hosts, including snapping turtles (*Chelydra serpentina* Linnaeus, 1758), flattened musk turtles (*Sternotherus depressus* Tinkle and Webb, 1955), painted turtles (*Chrysemys picta* [Schneider, 1783]), and wood turtles (*Glyptemys insculpta* [LeConte, 1830]) (Brooks et al., 1990). In contrast, *Placobdella ali* Hughes and Siddall, 2007 is known only from snapping turtles in New England states (Hughes and Siddall, 2007; Richardson et al., 2010). Whereas the presence of compound eye spots, esophageal cecate mycetomes, and bilobed ovaries are consistent morphological characters for the genus (Light and Siddall, 1999; Siddall et al., 2005), salivary-gland structure is less uniform across species of *Placobdella*. Like the related sanguivorous genus *Haementeria*, species of *Placobdella* that feed on turtles all are equipped with 2 pairs of compact salivary glands with the exception of *Placobdella kwetlumye* Oceguera-Figueroa, Kvist, Watson, Sankar, Overstreet and Siddall, 2010 from the Pacific Northwest, which retains only the anterior pair (Oceguera-Figueroa et al., 2010).

Considerable data regarding the transcriptomic profiles of blood-feeding leeches have become available for "medicinal" leeches and their relatives among the jawed Arynchobdellida (e.g., Macagno et al., 2010; Min et al., 2010; Oliveira et al., 2012; Kvist et al., 2013, 2014; Alaama et al., 2014; Xiao et al., 2015; see also Hibsh et al., 2015). However, among proboscis-bearing rhynchobdellid leeches, fewer than 900 (unassembled) expressed sequence tags exist (Faria et al., 2005) and only for a single species, *Haementeria depressa* Ringuelet, 1972. Various bioactive compounds have been purified and characterized from the Giant

Received 10 June 2015; revised 14 October 2015; accepted 26 October 2015.

† College of Technology, City University of New York, 300 Jay Street, Brooklyn, New York 11201.

‡ Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada, and Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 2B4, Canada.

DOI: 10.1645/15-802

TABLE I. Sequencing and assembly metrics for each of the transcriptomic data sets used.

Assembly	*Placobdella ali*	*Placobdella kwetlumye*	*Placobdella parasitica*
No. of seqs after demultiplexing	30,848	39,391	45,496
No. of seqs after trimming and quality filtering	29,992	38,169	44,185
No. of contigs (MIRA)	1,434	2,766	3,763
No. of seqs in debris list (MIRA)	3,533	7,388	9,165
No. of contigs (CAP3)	103	153	188
No. of singlets (CAP3)	926	2,115	2,985

Amazonian leech (*Haementeria ghilianii* de Filippi, 1849) including antistasins, ghilantens, and hementin (Sawyer, 1991), but only 1 anticoagulant gene product has been characterized from a species of *Placobdella* (Mazur et al., 1991, 1993; Salzet, 2001). Here we examine 3 species of *Placobdella* (*P. parasitica*, *P. ali*, and *P. kwetlumye*) in order to evaluate more fully the diversity of anticoagulation-related information in salivary transcriptomes of this genus and contrast that with the related South American genus *Haementeria*.

MATERIALS AND METHODS

Specimen collection, dissection, RNA extraction, and cDNA synthesis

Leech specimens were collected from the underside of rocks, wood and other debris in Silvermine Lake, Harriman State Park, New York (*Placobdella ali*), Lake Skannatati, Harriman State Park, New York (*Placobdella parasitica*), and Squires Lake, Whatcom County, Washington (*P. kwetlumye*), and immediately preserved in RNA*later* (Life Technologies/Thermo Fisher Scientific; New York, New York). Specimens were identified with the use of a stereomicroscope while they were submerged in RNA*later*; identifications were later verified molecularly with the use mitochondrial cytochrome *c* oxidase subunit I (COI) and nuclear NADH dehydrogenase 1 (ND1). Prior to RNA extraction, leeches were washed in 0.5% bleach for 1 min and subsequently rinsed in deionized water for 1 min to minimize contamination with surface bacteria. With the use of sterilized dissecting tools, salivary tissue masses (glandular tissue) from single individuals of each species were removed aseptically and subsequently rinsed in 0.5% bleach for 1 min followed by rinsing in deionized water for 1 min. Total RNA then was isolated with the use of a hybrid protocol as described in Kvist et al. (2014). Total RNA was visualized on an Agilent 2100 BioAnalyzer (Agilent Technologies, Wilmington, Delaware) using an Agilent RNA 6000 Nano Kit (Agilent Technologies). First- and second-strand cDNA synthesis followed the protocol described in Kvist et al. (2014), which is a modified protocol originally described by Hale et al. (2009). Second strand cDNA was quantified on a NanoDrop1000 Spectrophotometer (Thermo Fisher Scientific) and visualized using a 1.1% agarose gel in 1XTBE buffer and an Agilent High Sensitivity DNA Kit.

Library construction, multiplexing, emPCR, and pyrosequencing

Approximately 500 ng (in 100 µl) of each cDNA library ($n = 3$) was processed individually by following the manufacturer's protocols as outlined in the Roche 454 *RL Preparation Method Manual* (Roche Applied Sciences; version June 2010). Following nebulization and fragment-end repair, RL MID (bar-coded) adaptors were ligated to each library. Prior to pooling, the individually bar-coded libraries underwent small fragment removal and library assessment/quantitation with the use of an Agilent 2100 Bioanalyzer and QuantiFluor ST Fluorometer (Promega, Madison, Wisconsin); the latter was used to prepare a working stock of each bar-coded library at 1×10^7 molecules/µl (Table I). Emulsion-based clonal amplification, bead washes and recovery, DNA library bead enrichment, and sequence primer annealing were carried out with the use of the GS Junior Titanium emPCR (Lib-L) Kit (454 Life Sciences, Branford, Connecticut) in accordance with the manufacturer's protocols as outlined in the *emPCR Amplification Method Manual* (*Lib-L*) (v. April 2011). Enriched beads were prepared for sequencing on a GS Junior PicoTiterPlate Device with the use of the GS Junior Titanium Sequencing Kit (454 Life Sciences) and in accordance with the manufacturer's protocols as outlined in the *Sequencing Method Manual* (v. November 2011). Single-end massively parallel pyrosequencing was carried out in multiplex on a 454 GS Junior at the Sackler Institute for Comparative Genomics (American Museum of Natural History).

Filtering, trimming, and assembly

Data were demultiplexed and passed through 5 standard quality filters (Dot, Mixed, Signal Intensity, Primer and TrimBack Valley) with the use of native GS Junior software. Thereafter, sff_extract (http://bioinf.comav.upv.es/sff_extract/index.html) was used to create .fasta, .fasta.qual, .fastq, and .xml files. In addition, sff_extract clipped key/adaptor sequences and removed low-quality reads (i.e., any base listed in lower case). Overall sequence quality was visualized with the use of FastQC v 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). The FASTX Toolkit v 0.0.13 (FASTX Trimmer [f = 38 {*P. ali*, *P. parasitica*} or 41 {*P. kwetlumye*}; l = 525 {all}] and FASTQ Quality Trimmer [t = 20 {all}; l = 20 {all}]; http://hannonlab.cshl.edu/fastx_toolkit/) was used to further remove low-quality regions. FastQC v 0.10.1 was again used to visualize and verify the overall quality of the reads. MIRA v3.4.0.1 (Chevreux et al., 1999) was used to assemble the filtered reads *de novo*. MIRA was called as follows: mira –project=MyFile –job=denovo,est,accurate,454 454_SETTINGS -CL:qc=no:cpat=no -AS:mrpc=1 -OUT:sssip=yes >andlog_assembly_MyFile.txt (where qc=quality clip; cpat=poly-A clipping; mrpc=minimum reads per contig; sssip=save simple singlets in project). Subsequently, CAP3 (Huang and Madan, 1999) was implemented (with the use of default parameters) to assemble longer contigs with the use of the contigs and the "debris list" (i.e., real singlets that have no overlap with any other reads) created by MIRA v3.4.0.1.

Open reading frames, BLAST queries, and Blast2Go annotation

The program TRANSEQ v 6.6.0, which is part of the EMBOSS package (Rice et al., 2000), was used to translate contigs in all 6 reading frames. Only open reading frames (ORFs) of size >50 amino acids were examined for signal peptide sequences with SignalP v.4.0 (Petersen et al., 2011). Note that all undisturbed reading frames were saved for downstream informatics, such that each contig could potentially result in several ORFs.

Initially, contigs were BLASTed (with the use of BLASTx) against a locally compiled data set of known anticoagulants (modified from table 1 in Kvist et al., 2014), in which the majority were leech-derived peptides. Single contigs that best matched the anticoagulant-related sequences in the local data set were then reciprocally BLASTed against GenBank NR and SwissProt to confirm that they did not find better matches against nonanticoagulants. Positive reciprocal hits involved significant matches of the contigs to sequences annotated as the same bioactive protein as its best match in the locally compiled data set. By contrast, negative reciprocal hits occurred when the contigs matched a different bioactive protein than the best hit against the local data set. High scoring matches found for species of *Placobdella* were subsequently compared (tBLASTx) to the 894 EST sequences available for *Haementeria depressa* (see Faria et al., 2005).

Blast2GO (Götz et al., 2008) was used to infer functional annotations of the DNA sequences with the gene ontology. The entire pool of assembled transcripts was submitted to Blast2GO to clarify the proportional allocation of the partial transcriptome to various classification categories. The resulting GO terms were subsequently interpreted with the use of WEGO (Ye et al., 2006), and Bonferroni corrections of all possible pairwise comparisons were performed in order to assess putative differences in transcriptomic profiles between the 3 species. Histograms illustrating the proportions were created in Microsoft Excel.

RESULTS

Pyrosequencing of the multiplexed *Placobdella* spp. cDNA libraries resulted in 223,280 wells, of which 218,418 yielded signal (library + controls). Of the 212,988 library wells that passed tests for key signal, 115,907 (54.4%) passed all 5 quality filters, with an

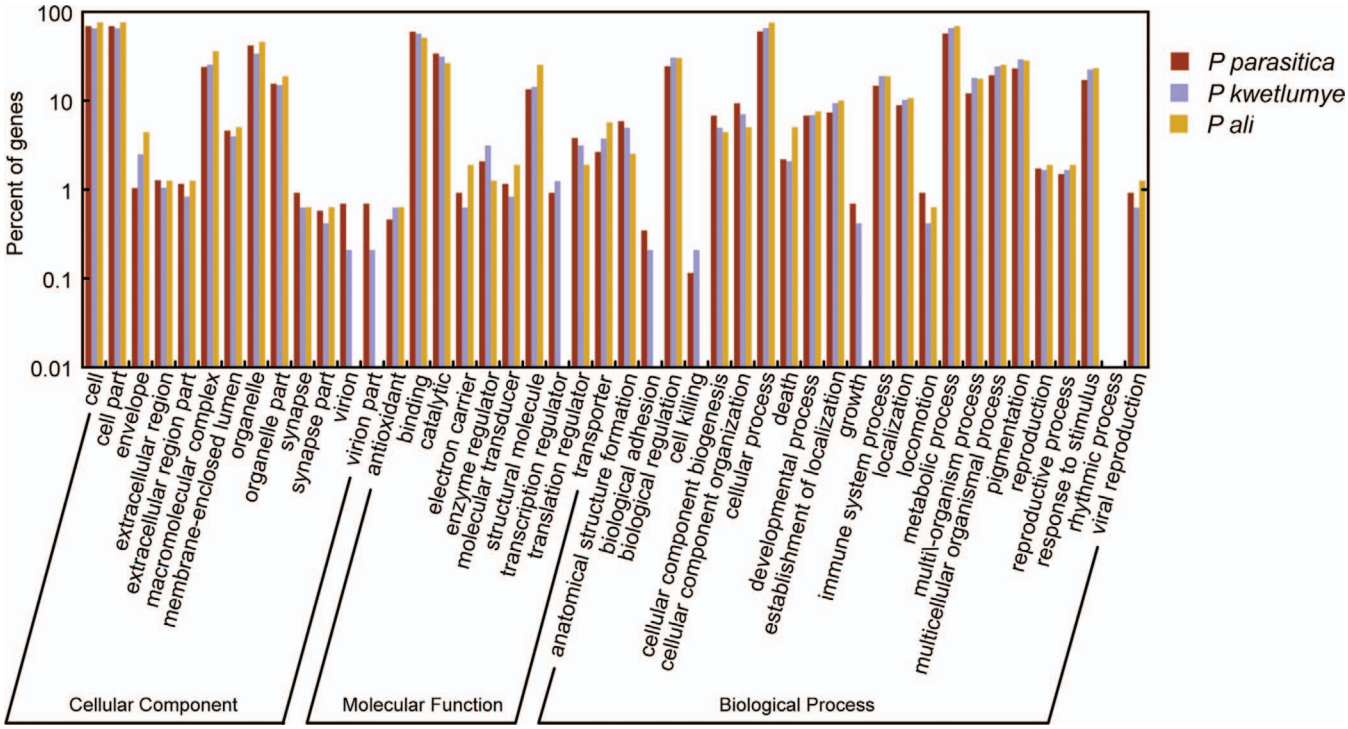

FIGURE 1. Results from gene ontology analysis, examined using WEGO, illustrating the (log scale) proportional devotion of the salivary transcriptomes to different processes regarding cellular component, biological process and molecular function. There were no significant differences between the 3 species of *Placobdella* after Bonferroni correction of all possible 138 pairwise comparisons. Missing values for *Placobdella ali* are expected relative to next-generation sequence coverage. Color version available online.

average read length of 408 +/− 134.4 nucleotides and thus a total of 47,291,904 bases sequenced. Assembly metrics from MIRA and CAP3 differed considerably (Table I); only contigs resulting from MIRA assembly were used for subsequent investigations of transcriptomic content. After demultiplexing 115,907 reads based on in-line bar codes and assembly with MIRA, there were 30,848 reads for *P. ali* (1,434 contigs), 39,391 for *P. kwetlumye* (2,766 contigs), and 45,496 for *P. parasitica* (3763 contigs). Of the 1,434 maximum possible, 728 contigs were found to be loci common to all 3 species of *Placobdella*. In addition to disintegrins, plasminogen, apyrases, and a tick-secreted peptide homologue, these included cytoskeletal proteins (22%, including the intermediate filament, filarin), ribosomal proteins (14%), and a large proportion (25%) corresponding to the metazoan-wide AC1147-like peptide of unknown function. Comparison of cellular component, molecular function, and biological process GO terms among the salivary transcriptomes of the 3 species of *Placobdella*, relative to whole-body EST data available for *Hirudo medicinalis* Linnaeus, 1758 were unremarkable (Fig. 1). All 115,907 reads passing quality filter assemblies have been deposited in the GenBank Sequence Read Archive (accession number SRX1335044).

BLAST to a local database of bioactive compounds and NR

Resulting contigs for each of the 3 species of *Placobdella* yielded high-scoring (low E-value) matches when compared (BLASTx) to a variety of local databases, and yielded equivalently high scoring reciprocal matches when compared to SwissProt. Transcripts corresponding to 9 distinct proteins that have previously been found in leech salivary secretions were found in at least 1 species of *Placobdella* examined here (Table II), of which nearly half exhibited significant evidence of a signal peptide and cleavage site. The most frequently encountered proteins were c-type lectins. As well, 2 antistasins (ghilanten and therostasin), leech antiplatelet protein (LAPP), and an endoglucoronidase (matching mannilase) were recovered from the transcriptome assemblies of all 3 species. A leucocyte elastase inhibitor and an antithrombin hirudin homolog were found in each of *P. kewtlumye* and *P. parasitica*, but not in *P. ali*. Eglin-c was recovered from *P. ali* and *P. parasitica*, but not from *P. kwetlumye*. Ornatin was only evident in the assembled contigs from *P. ali*. Transcripts corresponding to 5 other proteins with putative blood-feeding related functionality were found in at least 2 species of *Placobdella* (Table III). These included disintegrins with matelloprotease activity, apyrases (nucleotidases), plasminogen analogs, galactose binding lectins, and a kazal-type serine protease inhibitors. Of the foregoing, publicly available EST data for *H. depressa* evidence transcripts only of therostasin ($4.0E^{-60}$), LAPP ($4.0E^{-22}$), ghilanten ($9.0E^{-06}$), eglin-c ($4.0E^{-23}$), and the disintegins with matelloprotease activity ($5.0E^{-20}$).

Examination of the 106 inferred amino acid–aligned sites for hirudins revealed high levels of conservation of residues between the 2 species of *Placobdella* for which this antithrombin protease inhibitor was found, as compared to previously known sequences from arynchobdellid leeches in the families Hirudinidae, Haemadipsidae, and Macrobdellidae (Fig. 2). Each of 7 cysteines present in hirudin from *P. kwetlumye* and *P. parasitica* were conserved in position between the 2 species of *Placobdella* and with respect to the arhynchobdellids included in the alignment. The 20 amino

TABLE II. Known antihemostatis related loci with high scoring matches to *Placobdella* salivary transcripts.

Protein	Species	Contigs (with signal)	E (min)	E (max)
C-type lectin*	*Placobdella ali*	7 (2)	$2.00E^{-14}$	$8.00E^{-06}$
	Placobdella. kwetlumye	16 (5)	$7.00E^{-14}$	$2.00E^{-06}$
	Placobdella. parasitica	6 (3)	$4.00E^{-12}$	$3.00E^{-06}$
Therostasin	*P. ali*	6 (4)	$7.00E^{-29}$	$5.00E^{-08}$
	P. kwetlumye	4 (2)	$7.00E^{-29}$	$7.00E^{-19}$
	P. parasitica	9 (4)	$2.00E^{-28}$	$2.00E^{-15}$
Mannilase	*P. ali*	8 (4)	$6.00E^{-84}$	$1.00E^{-08}$
	P. kwetlumye	1 (1)	$1.00E^{-28}$	
	P. parasitica	5 (3)	$1.00E^{-81}$	$1.00E^{-14}$
LAPP†	*P. ali*	8 (6)	$1.00E^{-22}$	$2.00E^{-13}$
	P. kwetlumye	2 (1)	$1.00E^{-15}$	$4.00E^{-06}$
	P. parasitica	4 (1)	$1.00E^{-21}$	$5.00E^{-07}$
Ghilanten	*P. ali*	3 (1)	$2.00E^{-12}$	$2.00E^{-07}$
	P. kwetlumye	4 (3)	$6.00E^{-11}$	$7.00E^{-07}$
	P. parasitica	2 (0)	$1.00E^{-09}$	$3.00E^{-08}$
LEI‡	*P. ali*	0		
	P. kwetlumye	3 (1)	$5.00E^{-29}$	$3.00E^{-14}$
	P. parasitica	1 (0)	$4.00E^{-21}$	
Hirudin	*P. ali*	0		
	P. kwetlumye	1 (1)	$5.00E^{-07}$	
	P. parasitica	1 (1)	$1.00E^{-07}$	
Eglin-c	*P. ali*	1 (0)	$4.00E^{-11}$	
	P. kwetlumye	0	($5.00E^{-04}$)	
	P. parasitica	1 (1)	$7.00E^{-11}$	
Ornatin	*P. ali*	2 (1)	$1.00E^{-06}$	$2.00E^{-06}$
	P. kwetlumye	0		
	P. parasitica	0	($1.00E^{-04}$)	

* Poor matches show homology to nucleocan core.
† Leech antiplatelet protein, copy # variation may = alt/splice.
‡ Leucocyte elastase inhibitors, poor matches show homology to serpins.

TABLE III. Putative antihemostatis-related loci with high scoring matches common to salivary transcripts from at least 2 *Placobdella* species.

Protein	Species	Contigs (with signal)	E (min)	E (max)
Disintegrin /metalloprotease	*Placobdella. ali*	24 (10)	$2.00E^{-49}$	$7.00E^{-06}$
	Placobdella kwetlumye	5 (5)	$2.00E^{-39}$	$3.00E^{-06}$
	Placobdella. parasitica	10 (3)	$2.00E^{-30}$	$1.00E^{-07}$
Apyrase /nucleotidase	*P. ali**	6 (1)	$8.00E^{-52}$	$8.00E^{-13}$
	P. kwetlumye	4 (2)	$4.00E^{-74}$	$1.00E^{-13}$
	P. parasitica†	3 (2)	$2.00E^{-98}$	$7.00E^{-16}$
Plasminogen analog	*P. ali*	0		
	P. kwetlumye	7 (1)	$3.00E^{-22}$	$2.00E^{-13}$
	P. parasitica	4 (0)	$3.00E^{-21}$	$3.00E^{-12}$
Galactose binding lectin	*P. ali*	0		
	P. kwetlumye‡	3 (2)	$1.00E^{-34}$	$1.00E^{-13}$
	P. parasitica	4 (2)	$5.00E^{-32}$	$2.00E^{-06}$
Kazal protease inhibitor	*P. ali*	0		
	P. kwetlumye	1 (0)	$2.00E^{-16}$	
	P. parasitica	1 (0)	$1.00E^{-09}$	

* Includes high-scoring match to *Anopheles darlingi* salivary apyrase.
† Includes high-scoring match to *Haemonchus contortus* apyrase.
‡ Includes high-scoring match to *Ixodes scapularis* lectin.

acid predicted secretory signal cleavage site also was conserved across all taxa, with the leading signal peptide region showing more than 50% amino acid identity between *Placobdella* species and arhynchobdellids. Parsimony analysis of the aligned hirudins yielded a tree with 194 steps, and that is identical to the expected relationships of these leeches with the exception that the 2 hirudin sequences from species of *Haemadipsa* in the gene tree were not monophyletic; forcing monophyly of the 2 *Haemadipsa* species required an additional 2 steps.

Posttranslation modification and predicted signal peptides

Salivary gland transcriptomic contigs common to at least 2 of the 3 species of *Placobdella* also comprised a variety of posttranslational protein modification enzymes including glycosyl- and oligosaccharyl-transferases, disulfide isomerases, carbonic anhydrases, and asparagines-linked glycosylation proteins (Table IV) of which about 1/3 exhibited significant evidence of a secretory signal peptide and cleavage site. The recovery of N-linked glycosylation posttranslational pathway enzymes (dolychildiphosphooligosaccharide glycosyl transferases and Asp-linked glycosylation proteins) reflects the importance of glycosylation in holding the stability, solubility, and activity of leech salivary gland components that must operate in the external environment of a host's blood (Table IV). Moreover, glycosylation itself strongly modifies the potency and specific activities of hemotoxic snake venoms (Soares and Oliveira, 2009). Cystein-rich hirudin, ornatin, and in particular the antistasins ghilanten and therostasin, which are equipped with 20 cysteine residues each, owe their protease-inhibiting functionality to a protein scaffold shape built by disulfide bridges formed between cysteines in their sequences (Duggan et al., 1999). Protein disulfide isomerases, found in all 3 species of *Placobdella*, would be critical to the proper formation of these bonds in the growing proteins.

In addition we recovered several transcripts with significant evidence of a secretory signal peptide and cleavage site that were unique to individual species of *Placobdella*. Nine distinct contigs from *P. ali* encoded both EGF (epidermal growth factor) and CUB (complement C1r/C1s, Uegf, Bmp1) domains ($3.0E^{-10}$ to $3.0E^{-07}$), 6 of which were equipped with secretory signal peptides. Two distinct contigs from *P. kwetlumye* showed strong homology to tick (*Ixodes scapularis* Say, 1821) secreted proteins ($2.0E^{-19}$ to $1.0E^{-06}$) of which 1 was equipped with a secretory signal peptide. *Placobdella parasitica* contigs included those with high-scoring matches to a viper (*Bitis gabonica* [Duméril, Bibron and Duméril, 1854]) venom gland protein ($2.0E^{-27}$); a King Cobra (*Ophiophagus hannah* Cantor, 1836) cysteine-rich ophanin venom protein ($5.0E^{-24}$) with a secretory signal peptide; a multiple coagulation factor deficiency protein ($3.0E^{-12}$); and 2 cathepsin L ($5.0E^{-90}$–$6.0E^{-63}$) sequences, of which 1 was equipped with a secretory signal peptide.

DISCUSSION

Salivary transcriptome sequence data from 3 species of *Placobdella* put to rest the question as to whether leeches have a single origin of blood feeding in a common ancestor or have

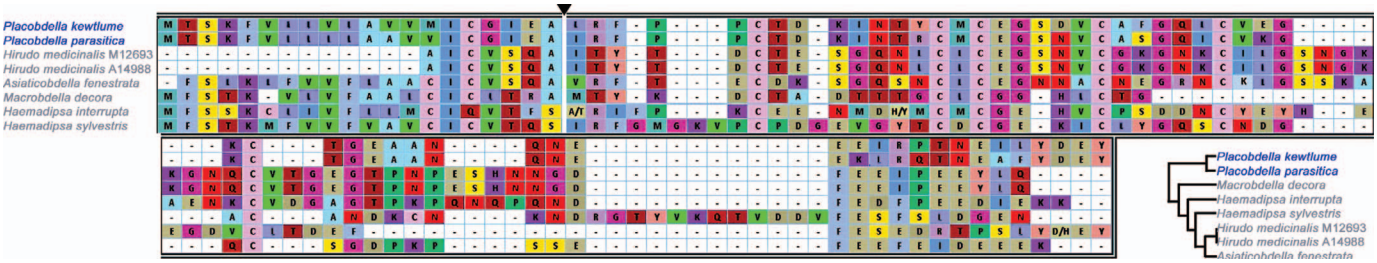

FIGURE 2. Predicted single-copy orthologs of hirudin from *Placobdella kwetlumye* and *Placobdella parasitica* aligned with those known from various arhynchobdellid taxa. The black triangle indicates predicted position of the mature secreted peptide cleavage site. The alignment donstrates high levels of conservation for this peptide across taxa, in particular regarding the 6 disulphide-bond-forming cysteines present in the mature peptide. Color version available online.

evolved sanguivory twice in their evolutionary history. The presence of single-copy hirudin genes in *P. kwetlumye* and in *P. parasitica*, each with unmistakable homology to those already known from *H. medicinalis* and other Hirudiniforms (Fig. 2) represents the first definitive genomic evidence that this potent antithrombin protease inhibitor was present in their mutual common ancestor at the origin of leeches. Sawyer (1986) may have been first to suggest that the proboscis bearing rhynchobdellid leeches and jawed arhynchobdellid leeches independently acquired sanguivory. Siddall and Burreson (1995) appeared to corroborate that supposition in the context of a morphological phylogeny of major leech groups in which these groups each were monophyletic. With the availability of DNA sequence data for phylogenetic inference, the proboscis bearing Rhynchobdellida proved paraphyletic (e.g., Borda and Siddall, 2004; Siddall et al., 2011), which would imply a blood-feeding ancestor of all leeches. However, the relative paraphyletic arrangement of Erpobdelliformes and Cylicobdelliformes left open whether these 2 groups independently lost blood feeding versus an equally parsimonious ancestral loss with a reacquisition by the Hirudiniformes, including medicinal and terrestrial leeches (Borda and Siddall, 2004). Transcriptome data previously available for *H. depressa* do not suggest anticoagulant salivary proteins common to that rhynchobdellid and the variety of bioactive proteins already known for hirudiniforms (Faria et al., 2005).

TABLE IV. Loci related to posttranslational modification of secreted proteins with high scoring matches to *Placobdella* salivary transcripts.

Protein	Species	Contigs (with signal)	E (min)	E (max)
DDOS* glycosyl-transferase	*Placobdella ali*	4 (0)	$4.00E^{-70}$	$1.00E^{-27}$
	Placobdella kwetlumye	2 (2)	$2.00E^{-25}$	$8.00E^{-21}$
	Placobdella parasitica	6 (1)	0	$5.00E^{-11}$
Protein disulfide isomerase†	*P. ali*	1 (0)	$1.00E^{-118}$	
	P. kwetlumye	2 (1)	$7.00E^{-154}$	$9.00E^{-75}$
	P. parasitica	6 (2)	$8.00E^{-168}$	$1.00E^{-15}$
GPI carbonic anhydrase‡	*P. ali*	1 (1)	$4.00E^{-27}$	
	P. kwetlumye	2 (1)	$2.00E^{-25}$	$8.00E^{-21}$
	P. parasitica	2 (0)	$2.00E^{-23}$	$3.00E^{-06}$
Asp-linked glycosylation protein	*P. ali*	1 (0)	$1.00E^{-32}$	
	P. kwetlumye	2 (0)	$2.00E^{-13}$	$1.00E^{-12}$
	P. parasitica	0		
Oligosaccharyl transferase	*P. ali*	1 (0)	$5.00E^{-25}$	
	P. kwetlumye	0		
	P. parasitica	1 (0)	$3.00E^{-06}$	

* Dolychil-diphosphooligoscahhacride protein glycosyl transferase.
† A3, A5 and A6 forms.
‡ Glycosyl-phosphatidylinositol-linked carbonic anhydrase.

The platelet glycoprotein IIb-IIIa inhibitor ornatin, from *Placobdella rugosa* (previously *Placobdella ornata*, see Moser et al., 2012) shows only weak amino acid homology to the North American medicinal leech (*Macrobdella decora* [Say, 1824]) platelet disintegrin decorsin; whereas the 6 cysteines are conserved, overall amino acid identitiy was only 34% (Min et al., 2010). Similarly, although arhynchobdellids and rhynchobdellids both prevent platelet aggregation by disrupting von Willebrand factor (vWf) –mediated collagen interactions with saratin and LAPP, respectively (Connolly et al., 1992; Barnes et al., 2001; Min et al., 2010), the longest stretch of conserved residues in the 2 proteins is H(V/I)FLHSS, which is not readily distinguishable from the 2 groups having converged on a common motif for a common problem (inhibiting platelet activation). With respect to hirudin, the first drug isolated from an animal for pharmacological use (Jacoby, 1904) with a powerful femtomolar inhibition constant, the most potent natural direct thrombin inhibitor (DTI) known (Greinacher and Warkentin, 2008), the profound conservation of cysteine residues likewise implies not only primary sequence conservation but conservation of the functionally important disulfide bonding secondary structure that affords bivalent binding to thrombin both in its active proteolytic site and its fibrinogen binding exosite (Min et al., 2010). Corroborating hirudin sequence homology, the gene tree derived from those data itself exhibits a topological arrangement (Fig. 2) that is highly congruent with expectations from recent molecular phylogenetic relationships of leeches (e.g., Borda and Siddall, 2004; Siddall et al., 2011) as one would expect from what in all taxa in which it has been found appears to be a short, nonrecombining, single-copy locus. The evidence from hirudin, while affirming a single origin of sanguivory, simultaneously implies independent losses in the Cyclicobdelliformes and Erpobdelliforms, while also corroborating ancient (if poorly conserved) homology for eglin-c (an elastase and cathepsin inhibitor), for extracellular matrix disrupting enodglucoronidases (e.g., orgelase and mannilase), for LAPP with saratin, and for ornatins with decorsin, all present in the ancestral blood-feeding leech.

Notwithstanding that inferences about expression levels from next-generation DNA sequencing results are difficult to interpret in light of size variation of target mRNA sequences, all 3 species of *Placobdella* yielded very large numbers of reads (and of

assembled contigs) for disintegrins and for various lectins. Min et al. (2010) noted the presence of high copy numbers of c-type lectins in the salivary transcriptome of *M. decora*, and so did Kvist et al. (2014) for the haemadipsid leech *Haemadipsa interrupta* (Moore, 1935). That appears to be equally true for all 3 species of *Placobdella*. These lectins also are important hemotoxic components of snake venoms like bothrojaracin and bothroalterin (Kini, 2006) and mosquito salivary extracts (Valenzuela et al., 2003). In addition, galactose-binding lectins have previously been isolated from a variety of annelids, including polychaetes, oligochaetes, and leeches (Hirabayashi et al., 1998). Although this family of lectins shares weak homology with galactose-binding, carbohydrate-relating proteins (e.g., ricin B-chain), a conserved and repeated amino acid motif is present among annelid galactose-binding lectins (GXXXQXW), such that homology can be determined with relative clarity among sequences possessing the motif. Interestingly, among the top 4 hits for this protein in *P. parasitica*, 75% possessed the common motif. The platelet glycoprotein IIb-IIIa inhibitor ornatin/decorsin are themselves disintegrins with RGD or KGD amino acid motifs. The additional disintegrins with matalloprotease activity found for species of *Placobdella* (Table III) appear to be distinct from those previously characterized. Thrombin-activating coagulation factor Xa is the primary target for the antistasin class of proteins common to a variety of blood-feeding invertebrates from hookworm nematodes to ticks and leeches (from which they were first isolated) and have been of great interest for their disrupting angiogenesis and preventing tumor metastases (Saito et al., 1980; Zacharski, 1981). Glossiphoniid leeches like *Placobdella* and *Haementeria* species appear to have a larger repertoire of distinct anti-Xa proteins, such as ghilanten and therostasin, than do the arhynchobdellid medicinal leeches. Whether the putative plasminogen activating contigs found for at least 2 species of *Placobdella* (Table III) correspond to the reported clot-dissolving and fibrinogenolytic protease hementin (Swadesh et al., 1990) described from the Giant Amazonian leech, *Haementeria ghilianii*, remains to be determined. Presently sequence data for the proprietary hementin do not exist in public databases. Apyrases from sanguivorous arthropods have also been shown to promote the inhibition of platelet aggregation in the host by hydrolyzing extracellular adenosine diphosphate (Smith et al., 2002). Interestingly, although numerous copies of apyrases seem to be present among the transcriptomic data for the 3 species studies here (see Table III), only a single copy for *P. ali*, and 2 copies for each of *P. kwetlumye* and *P. parasitica* possess a predicted signal peptide at the N-terminus, suggesting that only these copies are in fact secreted or, alternatively, that only these were activated at the time of preservation of the specimens.

We anticipated substantial differences in secretory composition among the 3 species of *Placobdella*. In particular, though *P. kwetlumye* lacks the posterior pair of salivary glands more typical of species in this genus of glossiphonid leeches (Moser and Desser, 1995), few differences were notable (Fig. 1; Tables II, III). In fact, *P. ali* lacked more of the overall representation of recovered loci, but this is readily explained by the relatively fewer reads acquired for *P. ali* from the multiplexed run. Much like our examination of putative blood-feeding related transcripts and posttranslational modification transcripts, *P. kwetlumye* (which has only 1 pair of salivary glands instead of the 2 pairs typical for the genus) does not obviously differ in its transcriptomic profile. Additional apparent lack of overlap in transcripts may be due merely to the precise physiological state of the leeches, or due to the relative stringency of search requirements ($1.0E^{-05}$). For example, the ornatin recovered for *P. ali* at $1.0E^{-06}$, while significant, is only just-so. A putative ornatin at $1.0E^{-04}$ was also found for *P. parasitica* (Table II), but not evaluated here because it did not cross the $10E^{-05}$ threshold—one that may be too stringent for such an exceedingly short protein of only 50 amino acids.

ACKNOWLEDGMENTS

We thank Jo Wolfe (Massachusetts Institute of Technology) for her assistance with the initial stages of library construction and Apurva Narechania (AMNH) for his assistance running the command-line version of Blast2GO.

LITERATURE CITED

Alaama, M., A. H. Uddin, A. M. Ghawi, A. Merzouk, A. M. Abdualkader, and M. B. Awang. 2014. Isolation and characterization of antithrombin peptides from the saliva of Malaysian leeches. Tropical Journal of Pharmaceutical Research **13:** 553–557.

Barnes, C. S., B. Krafft, M. Frech, U. R. Hofmann, A. Papendieck, U. Dahlems, G. Gellissen, and M. F. Hoylaerts. 2001. Production and characterization of saratin, an inhibitor of von Willebrand factor-dependent platelet adhesion to collagen. Seminars in Thrombosis and Hemostasis **27:** 337–348.

Borda, E., and M. E. Siddall. 2004. Arhynchobdellida (Annelida: Oligochaeta: Hirudinida): phylogenetic relationships and evolution. Molecular Phylogenetics and Evolution **30:** 213–225.

Brooks, R. J., D. A. Galbraith, and J. A. Layfield. 1990. Occurrence of *Placobdella parasitica* (Hirudinea) on snapping turtles, *Chelydra serpentina*, in southeastern Ontario. Journal of Parasitology **76:** 190–195.

Chevreux B., T. Wetter, and S. Suhai. 1999. Genome sequence assembly using trace signals and additional sequence information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) **99:** 45–56.

Connolly, T. M., J. W. Jacobs, and C. Condra. 1992. An inhibitor of collagen-stimulated platelet activation from the salivary glands of the *Haementeria officinalis* leech. I. Identification, isolation, and characterization. Journal of Biological Chemistry **267:** 6893–6898.

Duggan, B. M., H. J. Dyson, and P. E. Wright. 1999. Inherent flexibility in a potent inhibitor of blood coagulation, recombinant nematode anticoagulant protein c2. European Journal of Biochemistry **265:** 539–548.

Faria, F., I. de LM Junqueira-de-Azevedo, P. L. Ho, M. U. Sampaio, and A. M. Chudzinski-Tavassi. 2005. Gene expression in the salivary complexes from *Haementeria depressa* leech through the generation of expressed sequence tags. Gene **349:** 173–185.

Fields, W. S. 1991. The history of leeching and hirudin. Pathophysiology of Haemostasis and Thrombosis **21**(Suppl. 1)**:** 3–10.

Götz, S., J. M. García-GÓmez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talón, J. Dopazo, and A. Conesa. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research **36:** 3420–3435.

Greinacher, A., and T. E. Warkentin. 2008. The direct thrombin inhibitor hirudin. Thrombosis and Haemostasis **99:** 819.

Hale, M. C., C. R. McCormick, J. R. Jackson, and J. A. DeWoody. 2009. Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (*Acipenser fulvescens*): The relative merits of normalization and rarefaction in gene discovery. BMC Genomics **10:** 203.

Hibsh, D., H. Schori, S. Efroni, and O. Shefi. 2015. *De novo* transcriptome assembly databases for the central nervous system of the medicinal leech. Scientific Data **2:** 150015.

Hirabayashi, J., S. K. Dutta, and K. I. Kasai. 1998. Novel galactose-binding proteins in Annelida characterization of 29-kDa tandem

repeat-type lectins from the earthworm *Lumbricus terrestris*. Journal of Biological Chemistry **273:** 14450–14460.

Huang, X., and A. Madan. 1999. CAP3: A DNA sequence assembly program. Genome Research **9:** 868–877.

Hughes, J. L., and M. E. Siddall. 2007. A new species of leech from the New York metropolitan area. American Museum Novitates **3578:** 1–6.

Jacoby, C. 1904. Über Hirudin. Deutsche Medizinische Wochenschrift **30:** 1786.

Kini, R. 2006. Anticoagulant proteins from snake venoms: Structure, function and mechanism. Biochemical Journal **397:** 377–387.

Klemm, D. J. 1972. Freshwater leeches (Annelida: Hirudinea) of North America. Biota of freshwater ecosystems: Identification manual no. 8, U.S. Government Printing Office, Washington, D.C., **14:** 12–894.

Klemm, D. J. 1982. Leeches (Annelida: Hirudinea) of North America. U.S. Environmental Protection Agency Bulletin, EPA-600/3-82-025, Cincinnati, Ohio, 177 p.

Kvist, S., M. R. Brugler, T. G. Goh, G. Giribet, and M. E. Siddall. 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., G. S. Min, and M. E. Siddall. 2013. Diversity and selective pressures of anticoagulants in three medicinal leeches (Hirudinida: Hirudinidae, Macrobdellidae). Ecology and Evolution **3:** 918–933.

Kvist, S., A. Narechania, A. Oceguera-Figueroa, B. Fuks, and M. E. Siddall. 2011. Phylogenomics of *Reichenowia parasitica*, an alphaproteobacterial endosymbiont of the freshwater leech *Placobdella parasitica*. PLoS ONE **6:** e28192.

Light, J. E., and M. E. Siddall. 1999. Phylogeny of the leech family Glossiphoniidae based on mitochondrial gene sequences and morphological data. Journal of Parasitology **85:** 815–823.

López-Jiménez, S., and A. Oceguera-Figueroa. 2009. New species of rhynchobdellid leech (Hirudinea: Glossiphoniidae): A parasite of turtles from Chiapas, Mexico. Journal of Parasitology **95:** 1356–1359.

Macagno, E. R., T. Gaasterland, L. Edsall, V. Bafna, M. B. Soares, T. Scheetz, T. Casavant, C. Da Silva, P. Wincker, A. Tasiemski, et al. 2010. Construction of a medicinal leech transcriptome database and its application to the identification of leech homologs of neural and innate immune genes. BMC Genomics **11:** 407.

Markwardt, F. 2002. Hirudin as alternative anticoagulant—A historical review. Seminars in Thrombosis and Hemostasis **28:** 405–414.

Mazur, P., M. S. Dennis, J. L. Seymour, and R. A. Lazarus. 1993. Expression, purification, and characterization of recombinant ornatin E, a potent glycoprotein IIb-IIIa antagonist. Protein Expression and Purification **4:** 282–289.

Mazur, P., W. J. Henzel, J. L. Seymour, and R. A. Lazarus. 1991. Ornatins: Potent glycoprotein IIb–IIIa antagonists and platelet aggregation inhibitors from the leech *Placobdella ornata*. European Journal of Biochemistry **202:** 1073–1082.

McCallum, M. L., W. E. Moser, B. A. Wheeler, and S. E. Trauth. 2011. Amphibian infestation and host size preference by the leech *Placobdella picta* (Verrill, 1872) (Hirudinida: Rhynchobdellida: Glossiphoniidae) from the Eastern Ozarks, USA. Herpetology Notes **4:** 147–151.

Min, G. S., I. N. Sarkar, and M. E. Siddall. 2010. Salivary transcriptome of the North American medicinal leech, *Macrobdella decora*. Journal of Parasitology **96:** 1211–1221.

Moser, W. E., J. Bowerman, P. Hovingh, C. A. Pearl, and A. Oceguera-Figueroa. 2014. New host and distribution records of the leech *Placobdella sophieae* Oceguera-Figueroa et al., 2010 (Hirudinida: Glossiphoniidae). Comparative Parasitology **81:** 199–202.

Moser, W. E., and S. S. Desser. 1995. Morphological, histochemical, and ultrastructural characterization of the salivary glands and proboscises of three species of glossiphoniid leeches (Hirudinea: Rhynchobdellida). Journal of Morphology **225:** 1–18.

Moser, W. E., D. J. Richardson, C. I. Hammond, F. R. Govedich, and E. A. Lazo-Wasem. 2012. Resurrection and redescription of *Placobdella rugosa* (Verrill, 1874) (Hirudinida: Glossiphoniidae). Bulletin of the Peabody Museum of Natural History **53:** 375–381.

Mumcuoglu, K. Y. 2014. Recommendations for the use of leeches in reconstructive plastic surgery. Evidence-Based Complementary and Alternative Medicine **205929:** 1–7.

Oceguera-Figueroa, A. 2012. Molecular phylogeny of the New World blood feeding leeches of the genus *Haementeria* and reconsideration of the biannulate genus *Oligobdella*. Molecular Phylogenetics and Evolution **62:** 508–514.

Oceguera-Figueroa, A., S. Kvist, S. C. Watson, D. F. Sankar, R. M. Overstreet, and M. E. Siddall. 2010. Leech collections from Washington State, with the description of two new species of *Placobdella* (Annelida: Glossiphoniidae). American Museum Novitates **3701:** 1–14.

Oliveira, D. G. L., M. P. Alvarez-Flores, A. R. Lopes, and A. M. Chudzinski-Tavassi. 2012. Functional characterisation of vizottin, the first factor Xa inhibitor purified from the leech *Haementeria vizottoi*. Thrombosis and Haemostasis **108:** 570–578.

Petersen, T. N., S. Brunak, G. von Heijne, and H. Nielsen. 2011. SignalP 4.0. discriminating signal peptides from transmembrane regions. Nature Methods **8:** 785–786.

Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: The European molecular biology open software suite. Trends in Genetics **16:** 276–277.

Richardson, D. J., W. E. Moser, C. I. Hammond, A. C. Shevchenko, and E. Lazo-Wasem. 2010. New geographic distribution records and host specificity of *Placobdella ali* (Hirudinida: Glossiphoniidae). Comparative Parasitology **77:** 202–206.

Saito, D., M. Sawamura, K. Umezawa, Y. Kanai, C. Furihata, T. Matsushima, and T. Sugimura. 1980. Inhibition of experimental blood-borne lung metastasis by protease inhibitors. Cancer Research **40:** 2539–2542.

Salzet, M. 2001. Anticoagulants and inhibitors of platelet aggregation derived from leeches. FEBS Letters **492:** 187–192.

Sawyer, R. T. 1986. Leech biology and behaviour. Vol. 1. Clarendon Press, Oxford, U.K., 418 p.

Sawyer, R. T. 1991. Thrombolytics and anticoagulants from leeches. Nature Biotechnology **9:** 513–518.

Siddall, M. E., and R. B. Budinoff. 2005. DNA barcoding evidence for widespread introductions of a leech from the South American *Helobdella triserialis* complex. Conservation Genetics **6:** 467–472.

Siddall, M. E., R. B. Budinoff, and E. Borda. 2005. Phylogenetic evaluation of systematics and biogeography of the leech family Glossiphoniidae. Invertebrate Systematics **19:** 105–112.

Siddall, M. E., and E. M. Burreson. 1995. Phylogeny of the Euhirudinea: Independent evolution of blood feeding by leeches? Canadian Journal of Zoology **73:** 1048–1064.

Siddall, M. E., and E. M. Burreson. 1996. Leeches (Oligochaeta?: Euhirudinea), their phylogeny and the evolution of life-history strategies. Hydrobiologia **334:** 277–285.

Siddall, M. E., and E. S. Gaffney. 2009. Observations on the leech *Placobdella ornata* feeding from bony tissues of turtles. Journal of Parasitology **90:** 1186–1188.

Siddall, M. E., G. S. Min, F. M. Fontanella, A. J. Phillips, and S. C. Watson. 2011. Bacterial symbiont and salivary peptide evolution in the context of leech phylogeny. Parasitology **138:** 1815–1827.

Smith, T. M., C. A. Hicks-Berger, S. Kim, and T. L. Kirley. 2002. Cloning, expression, and characterization of a soluble calcium-activated nucleotidase, a human enzyme belonging to a new family of extracellular nucleotidases. Archives of Biochemistry and Biophysics **406:** 105–115.

Soares, S. G., and L. L. Oliveira. 2009. Venom-sweet-venom: N-linked glycosylation in snake venom toxins. Protein and Peptide Letters **16:** 913–919.

Swadesh, J. K., I. Y. Huang, and A. Z. Budzynski. 1990. Purification and characterization of hementin, a fibrinogenolytic protease from the leech *Haementeria ghilianii*. Journal of Chromatography A **502:** 359–369.

Valenzuela, J. G., I. M. Francischetti, V. M. Pham, M. K. Garfield, and J. M. Ribeiro. 2003. Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito. Insect Biochemistry and Molecular Biology **33:** 717–732.

Whitaker, I. S. J. Rao, D. Izadi, and P. E. Butler. 2004. *Hirudo medicinalis*: Ancient origins of, and trends in the use of medicinal

leeches throughout history. British Journal of Oral and Maxillofacial Surgery **42:** 133–137.

Xiao, L., J. Nie, D. Li, and K. Chen. 2015. Peptides from two sanguinovorous leeches analyzed by ultra-performance liquid chromatography coupled with electrospray ionization quadrupole time-of-flight mass spectrometric detector. Pharmacognosy Magazine **11:** 32–37.

Ye, J., L. Fang, H. Zheng, Y. Zhang, J. Chen, Z. Zhang, J. Wang, S. Li, R. Li, L. Bolund, and J. Wang. 2006. WEGO: A web tool for plotting GO annotations. Nucleic Acids Research **34**(Suppl. 2)**:** W293–W297.

Zacharski, L. R. 1981. Anticoagulation in the treatment of cancer in man. *In* Malignancy and the hemostatic system, M. B. Donati, J. E. Davidson, and E. Garattinii (eds.). Raven Press, New York, New York, p. 113–128.