Bacterial symbiont and salivary peptide evolution in the context of leech phylogeny

MARK E. SIDDALL[1]*, GI-SIK MIN[2], FRANK M. FONTANELLA[3], ANNA J. PHILLIPS[1,4] *and* SARA C. WATSON[1]

[1] *Sackler Institute of Comparative Genomics & Division of Invertebrate Zoology, American Museum of Natural History, New York, NY, USA*
[2] *Department of Biological Sciences, Inha University, Yonghyun-dong, Nam-Gu, Incheon 402-751, Korea*
[3] *Department of Biology, Brigham Young University, Provo, UT 84602, USA*
[4] *Department of Biology, The Graduate Center, The City University of New York, New York, NY, USA*

(*Received February 2011; revised 7 March 2011; accepted 8 March 2011; first published online 10 May 2011*)

SUMMARY

The evolutionary history of leeches is employed as a general framework for understanding more than merely the systematics of this charismatic group of annelid worms, and serves as a basis for understanding blood-feeding related correlates ranging from the specifics of gut-associated bacterial symbionts to salivary anticoagulant peptides. A variety of medicinal leech families were examined for intraluminal crop bacterial symbionts. Species of *Aeromonas* and Bacteroidetes were characterized with DNA gyrase B and 16S rDNA. Bacteroidetes isolates were found to be much more phylogenetically diverse and suggested stronger evidence of phylogenetic correlation than the gammaproteobacteria. Patterns that look like co-speciation with limited taxon sampling do not in the full context of phylogeny. Bioactive compounds that are expressed as gene products, like those in leech salivary glands, have 'passed the test' of evolutionary selection. We produced and bioinformatically mined salivary gland EST libraries across medicinal leech lineages to experimentally and statistically evaluate whether evolutionary selection on peptides can identify structure-function activities of known therapeutically relevant bioactive compounds like antithrombin, hirudin and antistasin. The combined information content of a well corroborated leech phylogeny and broad taxonomic coverage of expressed proteins leads to a rich understanding of evolution and function in leech history.

Key words: Leeches, phylogeny, microbiology, protein evolution.

INTRODUCTION

From the advent of single-gene amplification reactions in the late 1980s, and until not very long ago, many systematists have been content in fashioning phylogenetic trees for their group of interest from DNA sequence data on the basis of one or a few loci. Our motivations were, and largely still remain, centered on uncovering the specifics of the evolutionary relationships of the constituent higher taxonomic groups down to species so as to better describe their history of diversification. These self-styled 'tree of life' research programmes have borne considerable fruit, particularly in the last decade (e.g. Giribet, 2008; Hackett *et al*. 2008; McLaughlin *et al*. 2009), during which time at-the-bench research efforts and costs associated with generating data have dwindled even as the scope of loci and numbers of taxa has accelerated. Similarly, the difficulties associated with such problems as sequence alignment and optimal tree discovery have been ameliorated considerably through computational advances driven by an intersection of evolutionary biology and the computer sciences (e.g. Bader *et al*. 2006; Warren *et al*. 2007; Hittinger *et al*. 2010; Kristensen *et al*. 2010). Presently, the ease with which even model-based phylogenetic trees can be acquired for hundreds of terminals is such that the operationalism associated with the endeavour has progressively migrated from the rarefied realm of advanced research laboratories to being readily available to undergraduate and even secondary school instruction for basic biology course work (Kvist *et al*. 2011).

Our collective success in taking what has been a research goal and transforming that into a prerequisite for research is reminiscent of similar transformations associated with earlier technological advances such as electron microscopy. That is, while the 1960s and 1970s marked an 'age of discovery' for the sub-cellular organization of Apicomplexa and their complex development, or the surface architecture of cestode worms, electron microscopy is now but a tool in comparative biology, not a research nexus in and of itself. The same must now be admitted as it pertains to molecular and morphological phylogenetic systematics. We believe that it is appropriate for some introspection regarding how the powerful tool of a phylogeny, at one time the consequential end-point, is being re-imagined in a broader context of a progressive research programme that continues to

* Corresponding author: Mark E Siddall. E-mail: siddall@amnh.org

Parasitology (2011), **138**, 1815–1827.
doi:10.1017/S0031182011000539

deepen our understanding of the natural world. In short, knowing the *name* of the song, what the song *is called*, and even what the *name-of-the-song is called*, each belie what the song actually *is* (Carroll, 1871).

Imre Lakatos aptly contrasted progressive and degenerative research programmes; the latter being marked by entrenchment, the former by its pursuit of novelties both of method and prediction (Lakatos, 1971). Exemplary of the distinction between these two paradigms is embodied in the (socio-scientific) history of the pursuit of historical biogeography in a phylogenetic context. Early considerations of the spatiotemporal diversification of some clades consisted of little more than superimposition of phylogeny and cartography (Croizat, 1958). The idea was sufficiently novel to spur ever increasingly thoughtful methods of inference, some operationalist, others statistical, with a trajectory that has seen and continues to see advances like Component Analysis (Page, 1990), Biogeographic Parsimony Analysis (Brooks, 1990), Dispersal Vicariance Analysis (Ronquist, 1997), and most recently LaGrange (Ree and Smith, 2008) and BayesDIVA (Nylander *et al*. 2008). This field of inquiry clearly continues to satisfy the progressivism imagined by Lakatos (1971), even as the tendency to entrench oneself in one method or another might not be. Our intent is not to cast aspersions, having been as guilty of Croizatian generalized-arm-waving (Borda *et al*. 2008) as we have equally availed ourselves of more progressive considerations regarding the historical biogeography of leeches (Borda and Siddall, 2011). Rather, we hope to stimulate a deeper consideration of how the results of phylogenetic analyses (i.e. trees) might be brought to bear on the field of comparative biology in manner that already is technologically and conceptually well within reach; though perhaps ways underexploited by our field, systematics.

Using the Hirudinida, leeches, as a framework - historically notorious and yet an inexplicably understudied group of (mostly) ectoparasitic annelid worms – our aim here is to go 'beyond the tree'. This is not to minimize the efforts several have made in the last 15 years to generate hypotheses of phylogenetic relationship for leeches; indeed, those efforts are necessary and central prerequisites to the discovery operations of historical correlates we might now explore. Instead, and drawing on each of microbiology, co-speciation, evolutionary selection and genomic evolution, we hope to characterize and describe a progressive research programme for a clade of charismatic microfauna in a way that inspires our hirudinological colleagues as much as it might have others give greater consideration to what can be accomplished with a tree-in-hand.

THE TREE AS A PREMISE

Underpinning any contemporary approach to comparative biology is a phylogenetic tree, or more typically a constellation of trees pertaining to the group of interest. With respect to leeches, work towards understanding their place among Annelida more generally, and of the various family, genus and species level relationships more specifically, began first with morphological (Siddall and Burreson, 1995) and, quickly on the heels of that, molecular phylogenetic analyses (Siddall and Burreson, 1998). Among the earliest discoveries stemming from this work were that leeches do not deserve their own Class-level status as Hirudinea, on a par with Oligochaeta and Polychaeta, but in contrast are simply a highly specialized group of oligochaete worms closely related to the Lumbriculida (Siddall *et al*. 2001). While perhaps an unwelcome diminishment of the taxonomic stature of the group, the findings are fortuitous for any long-term attempts at understanding their development, biochemistry or other evolutionary-associated phenomena. Suppose, for example, that dinosaurs were sister to crocodiles and birds, no extant data from crocodiles and birds could convincingly shed light on the unknowable characteristics of dinosaurian physiology or soft anatomy. The discovery that dinosaurs are arranged as a paraphyletic grade between crocodilians and birds (relegating Aves to a mere subset of theropods), allows for a more convincing understanding of ancestral states. So too with the clitellate annelids. Determination of ancestral states, even as it might concern gene families associated with blood-feeding, can now proceed by examination of the utility and complexity of homolgous loci in the lumbriculids and other groups of oligochaetes.

Leeches remain monophyletic in all analyses of the group. The phylogenetic relationships of various leech groups remains a work in progress, but one that has already revealed considerable information about within and among group relationships. Apakupakul *et al*. (1999) remains the touchstone from which all other leech phylogenetic work derives. In that analysis we demonstrated the basic organization of leech evolutionary history with the early divergence of Glossiphoniidae, Ozobranchidae and Piscicolidae and confirming the sister group relationship of the hirudiniform and erpobdelliform leeches. Within those basic outlines, most of the suborders and families of Hirudinida have since been subject to phylogenetic scrutiny on the basis of molecular and morphological data including Glossiphoniidae (Siddall *et al*. 2005), Piscicolidae (Utevsky and Trontelj, 2004; Williams and Burreson, 2006), predaceous Erpobdelliformes (Siddall, 2002; Oceguera-Figueroa *et al*. 2011), and the Hirudiniformes of blood-feeding infamy (Borda and Siddall, 2004; Phillips and Siddall, 2005; Borda *et al*. 2008; Phillips and Siddall, 2009; Phillips *et al*. 2010). It is with respect to the latter that most of the rest of this contribution is concerned. That is, our recent research into phylogenetic correlates, both microbial

https://doi.org/10.1017/S0031182011000539 Published online by Cambridge University Press

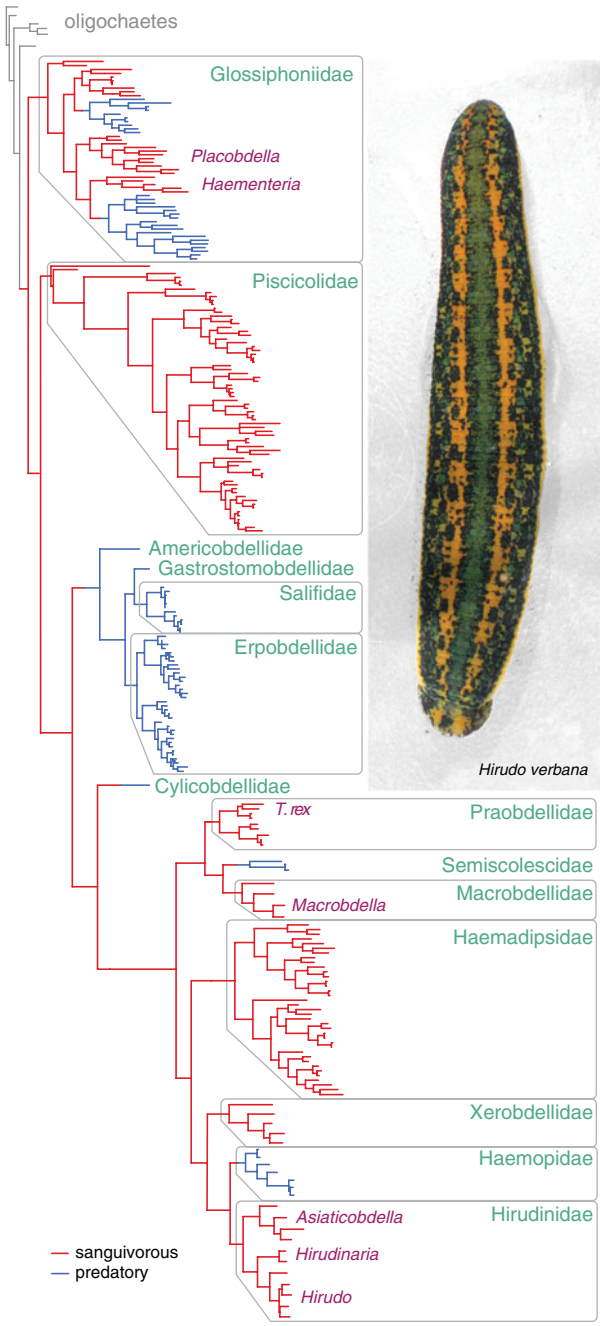

Fig. 1. Composite metaphylogeny of the order Hirudinida based on a collection of prior work illustrating current knowledge of the relationships of most leech families. The relationships of the Hirudindiformes (below *Hirudo verbana*) indicate the complex evolutionary history of the 'medicinal' leech families Hirudinidae, Praobdellidae and Macrobdellidae. Each terminal represents a species in a molecular phylogenetic analysis. Branches are proportional to change within families. Backbone phylogeny based on Apakupakul *et al*. (1999), Siddall *et al*. (2001) and Phillips and Siddall (2009). Blood-feeding lineages in red; non-sanguivorous lineages in blue.

and salivary, has focused on leeches once thought to comprise the Hirudinidae. These are vermiform, freshwater, swimming leeches with muscular jaws armed with denticles for cutting into flesh so as to allow the acquisition of a blood-meal. Fig. 1 amalgamates the current state of knowledge regarding leech phylogeny. As it pertains to the traditional composition of Hirudinidae, those taxa now are variously spread across the hirudiniforms in the families Hirudinidae, Macrobdellidae and Praobdellidae. Borda *et al*. (2008) detailed the intermediate phylogenetic position of terrestrial leech families Haemadipsidae and Xerobdellidae between the New World medicinal leeches, Macrobdellidae, and Old World Medicinal leeches, Hirudinidae *sensu stricto*. Furthermore, various non-blood-feeding groups have been found to place among the 'medicinal' leech taxa (Phillips and Siddall, 2009), and an entirely mammalophilic family, Praobdellidae, has been recently recognized as phylogenetically distinct from the other two 'medicinal' leech clades (Phillips *et al*. 2010).

MICROBIAL CORRELATES

Like many other blood-feeding animals, leeches harbour select prokaryotic flora in association with their digestive tracts. Proboscis-bearing sanguivorous species possess mycetomal organs specific to this task with intracellular alphaproteobacteria or gammaproteobacteria (Kikuchi and Fukatsu, 2002; Siddall *et al*. 2004; Perkins *et al*. 2005). In contrast, jaw-bearing medicinal leeches in the Hirudinidae and Macrobdellidae host a limited flora in the intraluminal fluid of the crop (Graf, 1999; Siddall *et al*. 2007*a*; Laufer *et al*. 2008). To date, only a single culturable bacterial species has been detected in any individual medicinal leech: *Aeromonas veronii* in the European *Hirudo verbana* (Graf, 1999, 2002), which was often mistakenly reported as *Hirudo medicinalis* (Siddall *et al*. 2007*b*), *Aeromonas jandaei* in the North American *Macrobdella decora* (Siddall *et al*. 2007*a*), and either of these two *Aeromonas* species (but never both) in the crop of the European *Hirudo orientalis* (Laufer *et al*. 2008). In addition to these individual culturable gammaproteobacteria, Worthen *et al*. (2006) demonstrated the co-presence of an unculturable Bacteroidetes microbe closely related to *Rikenella* species in *H. verbana*.

The crop, or gastric caeca, occupies approximately one-third of a leech's body somites allowing the annelid to expand more than six times its unfed body weight during feeding, permitting extended periods between feeding events (Munro *et al*. 1992). The role of the resident microbial flora is not yet well elucidated. Functions could range from the provision of essential nutrients not readily available in a diet that is limited exclusively to blood (e.g. Nogge, 1981) to antimicrobial activities inhibiting putrefaction of the blood meal (Rio *et al*. 2007).

Species of *Aeromonas*, including *A. veronii* and *A. jandaei*, are ubiquitous in circumglobal freshwater habitats raising questions regarding the historical

https://doi.org/10.1017/S0031182011000539 Published online by Cambridge University Press

maintenance of a single species of the genus in any given leech. Graf (2000) was first to suggest that oral vertical transmission is responsible insofar as all leeches must withdraw their oral anterior through the egg-bearing coccon after it is secreted by the clitellum. Corroborating this, the medicinal use of leeches has repeatedly demonstrated their propensity for introducing *Aeromonas* infections at a bite wound (Whitaker *et al*. 2009). Recent work confirms that *Aeromonas veronii* is present as soon as *H. verbana* coccons are deposited (Rio, 2008). The *Rikenella*-like symbiont is detectable later (Rio *et al*. 2008). While not a prerequisite, such vertical transmission of associated microbes hints at emergent co-evolutionary histories (Moran, 2001).

The revision of medicinal leeches into several families (Phillips *et al*. 2010) demonstrates that the two genera examined thus far for their intraluminal microbial crop symbionts are distantly related representatives of Macrobdellidae and the revised Hirudinidae (Fig. 1). Here we investigate the crop flora of a broader range of leech genera and families, and evaluate historical patterns of this tripartite symbiotic system.

Methodology

Intraluminal blood-meal was removed following transverse bisection of leeches at the region of the gastric tissue and well-anterior of the intestinal tract. DNeasy Tissue Kit (Qiagen Valencia, CA) was used for tissue lysis and DNA purification. *Aeromonas*-specific primers for DNA gyrase B (*gyr*B) were AerogyrBf TGTTGCTGACCATTCGTCGTAAC and AerogyrBr TTGGCATCGCTCGGGTTTTC with a predicted optimal annealing temperature of 59·4 °C. Amplification reactions employed Taq Gold (Applied Biosystems) and 50 cycles of 94 °C (45 sec), 55 °C (45 sec) and 72 °C (60 sec) following a 10 min pre-melt at 94 °C. Bacteroidetes-specific primers employed for amplification of 16S rDNA from the co-symbiont and to avoid co-amplification of the gammaproteobacterium were SSUrik416F GCAGGAAGACGGCTCTATGAGTTG and SSUrik781 RATCGTTTACGGCGTGGACTACC with a predicted optimal annealing temperature of 56·7 °C. Amplification reactions employed Ready-To-Go PCR Beads (GE Healthcare) and 35 cycles of 94 °C (15 sec), 50 °C (15 sec) and 72 °C (40 sec) following a 4 min pre-melt at 94 °C. PCR amplification products were purified with AMPure™ (Agencourt Bioscience Corporation). Cycle sequencing reactions were performed with an Eppendorf Mastercycler® using 1 μl Big Dye™ Extender Buffer v3.1, 1 μl of 1 μM primer and 3 μl of cleaned PCR template (13 μl total volume) and analyzed with an ABI PRISM® 3730 sequencer (Applied Biosystems). CodonCode Aligner (CodonCode Corporation) was used to edit and reconcile sequences. Sequences employed for comparative purposes were downloaded from NCBI. Alignments were accomplished using the European Bioinformatics Institute server for MUSCLE v. 3.7. Parsimony analyses were conducted in TNT v 1.1 (Goloboff *et al*. 2008) using ten replicates of random taxon addition, sectorial searching, the Ratchet (Nixon, 1999), and tree-fusing algorithms, with a requirement that the minimum length be found at least three times. Trees resulting from these new technology searches were submitted to tree-bisection-reconnection branch swapping retaining up to 10 000 trees. Resampling in TNT employed the parsimony jack-knife (Farris *et al*. 1996), with five replicates of random taxon addition, sectorial searching, the Ratchet (Nixon, 1999), and tree fusing, with no requirement that the minimum length be found multiple times.

Data

Parsimony analysis of *gyr*B sequences (Fig. 2) for species of *Aeromonas* resulted in 100 equally parsimonious trees with 2232 steps for 504 informative characters and a retention index of 0·75. Each species of *Aeromonas* for which multiple sequences were available was resolved as monophyletic with jack-knife frequency values ranging from 67% for *Aeromonas bestiarum* towards 100% for most species. Isolates from European *H. verbana* and *H. orientalis*, Mexican *Limnobdella mexicana* and Southeast Asian *Hirudinaria manillensis* grouped in the *A. veronii* clade. Isolates from European *H. medicinalis* and African *Asiaticobdella fenestrata* grouped in the *Aeromonas hydrophila* clade. Isolates from North American *M. decora* and European *H. orientalis* grouped in the *A. jandaei* clade. Amplification reactions of *gyr*B were not successful with Asian *Limnatis paluda* and *Hirudo nipponia*, African *Asiaticobdella buntonensis*, or Australian *Goddardobdella elegans*.

Parsimony analyses of the 16S rDNA sequences obtained from Bacteroidetes symbionts only of hirudiniform leeches (Fig. 3) resulted in 1 tree of length 321 for 115 informative characters and a retention index of 0·92. The resulting consensus grouped isolates in a manner that was highly congruent with the phylogenetic history of 'medicinal' leeches. That is, six of eight vertices of the 16S rDNA tree map to nodes on the leech tree without conflict (Fig. 3). The two conflicting vertices resulted from non-monophyly of the isolates from *Asiaticobdella* species and from the pre-exisiting isolate from *H. verbana* not grouping with other European *Hirudo* species.

Parsimony analysis of 16S rDNA sequences for a broader sampling of Bacteroidetes (Fig. 4) resulted in 36 equally parsimonious trees with 10 560 steps for 870 informative characters and a retention index of 0·67. Isolates from European and Asian species of

https://doi.org/10.1017/S0031182011000539 Published online by Cambridge University Press

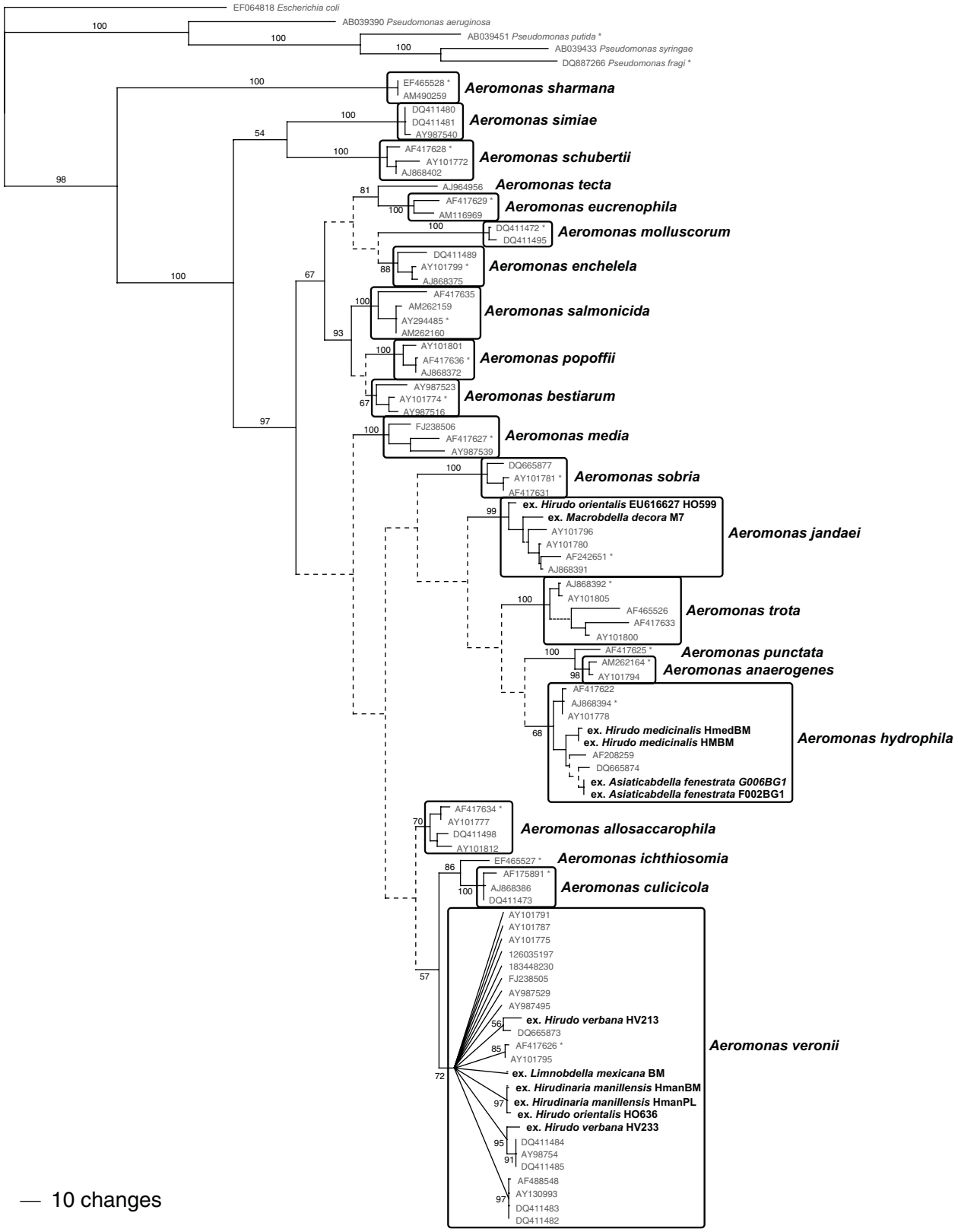

Fig. 2. Consensus of 100 equally parsimonious trees resulting from analysis of *gyr*B sequences of species of *Aeromonas* as well as those isolated from the crop of hirudiniform leeches (bold). Asterisks denote sequences obtained from type-strains for species. Numbers at nodes are jack-knife frequencies (not shown within species). Relationships supported in fewer than 50 jack-knife replicates are represented by interrupted lines.

Hirudo formed a clade sister to a fish gut symbiont, which together were more closely related to species of *Alistipes* than to *Rikenella microfusus*. The isolate from North American *M. decora* clustered nearby among a variety of uncultured and unidentified isolates, but closest to a termite gut symbiont.

https://doi.org/10.1017/S0031182011000539 Published online by Cambridge University Press

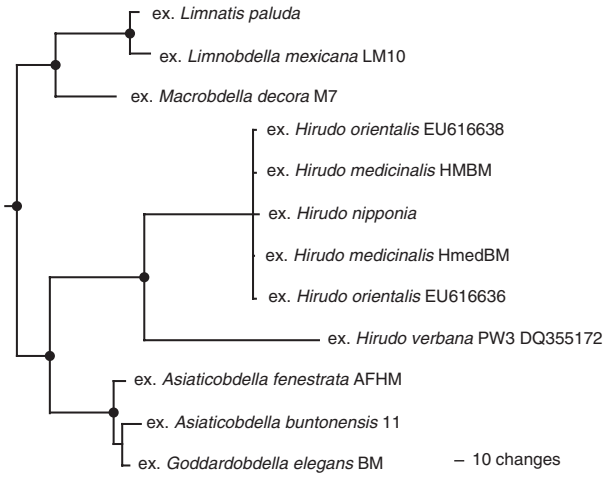

Fig. 3. Consensus of 105 equally parsimonious trees resulting from analysis of Bacteroidetes-specific 16S rDNA amplicons from hirudiniform leeches. Closed circles at nodes correspond to divergences that are consistent with leech phylogeny in Fig. 1.

Isolates from African species of *Asiaticobdella* and the Australian *G. elegans* formed a clade deriving from among *Pedobacter* species. Isolates from the Asian *L. paluda* and the Mexican *L. mexicana* formed a clade deriving from among a paraphyletic assemblage of *Flexibacter* species.

The two distinct microbial groups resident in the intraluminal crop fluid of hirudiniform leeches exhibit markedly different patterns of historical conservation with respect to their leech hosts. In terms of the genus *Aeromonas*, and even though there appears to be marked host specificity and vertical transmission (Rio *et al*. 2009), leech phylogenetic relationships are entirely non-predicitive of the associations; nor is geography. That is, the North American *M. decora* (Macrobdellidae) and European *H. orientalis* (Hirudinidae) each harbour *A. jandaei*. Similarly, European *H. medicinalis* harbours the same symbiont, *A. hydrophila*, as the African *A. fenestrata* and is thus distinct from its European congeners *H. verbana* and *H. orientalis*. *Aeromonas veronii* proved to be the most ubiquitous culturable symbiont, inhabiting the crop of leeches in three distinct genera from two families across three continents. Whether or not species of *Aeromonas* are involved in the origins or maintenance of species-level cohesion where recently derived leech congeners are sympatric is an intriguing possibility.

The culturable gut flora of the European medicinal leech, originally named *Pseudomonas hirudinis* Busing, 1953, was widely considered to be a strain of *A. hydrophila* and has been a matter of some concern in the post-operative use of commercially available leeches (Whitaker *et al*. 2009). Graf (1999) demonstrated that all isolates from commercially available leeches were actually *A. veronii* and that previous identifications as *A. hydrophila* were misled by the vagaries of chemotaxonomy. Since then, Siddall *et al*. (2007*a*) have demonstrated that commercially available European medicinal leeches actually are *H. verbana*, not *H. medicinalis*. It is as ironic to discover *H. medicinalis* harbouring *A. hydrophila*, as it is accidentally fortuitous that *H. verbana* has been the leech commercially available for clinical use; it harbours a considerably less-pathogenic *A. veronii* (Silver *et al*. 2007).

Bacteroidetes symbionts, unlike species of *Aeromonas*, exhibit a considerably tighter historical association with their respective hosts, and one that is more obviously phylogenetically than geographically constrained. The phylogenetic results of Bacteroidetes leech symbionts alone (Fig. 3), while topologically remarkably similar to historical expectations from hirudiniform phylogeny (Fig. 1) prove illusory when reconsidered in the context of Bacteriodetes more fully (Fig. 4). Half of the apparent co-evolutionary pattern depicted in the leech-symbiont-only tree evaporates under broader phylogenetic consideration in a manner that should serve as a caution to other investigations of host-symbiont co-speciation. This problem of scale in co-speciation studies has a precendent in Nishiguchi *et al*.'s (1998) work concerning light-emitting symbiotic vibrionids in sepiolid squid. That is, while a preliminary analysis based on only seven species evidenced tight co-evolutionary patterns between *Vibrio fischeri* strains and squid hosts, an error of scale that is frequently taken for granted (Kimbell *et al*. 2002; Kimbell and McFall-Ngai, 2003; Nishiguchi, 2002; Nichiguchi *et al*. 2004; Soto, 2009), that co-speciation pattern was erased under fuller consideration of squid and fish associated strains of bioluminescent vibrionids (Dunlap *et al*. 2007; Keading *et al*. 2007).

In the broader evaluation of Bacteroidetes, however, the symbiont-leech associations retain more phylogenetic than geographic constraint (Fig. 4). All symbionts from species of *Hirudo*, whether European or Asian (*H. nipponia*) form a clade of apparently indistinguishable *Alistipes* species. Reflecting leech genus-level diversificiation (Fig. 1), and notwithstanding their inhabiting well-separated continents, Australian species of *Goddardobdella* are host to a symbiont that is closely related to (and barely distinguishable from) a *Pedobacter* species found in two African *Asiaticobdella* species. Similarly, symbionts of leeches in the Praobdellidae form a clade of *Flexibacter*-like species despite the obvious geographic separation of Mexico and Afghanistan. Taken together these results are suggestive of recent, but not ancient, tight historical association between leech hosts and their unculturable Bacteroidetes symbionts. A similar recent historical pattern has been noted in relation to glossiphoniid leeches and their mycetomal symbionts in which three clades of leeches, the genera *Placobdella, Placobdelloides* and *Haementeria*, are host to three distinct clades of

https://doi.org/10.1017/S0031182011000539 Published online by Cambridge University Press

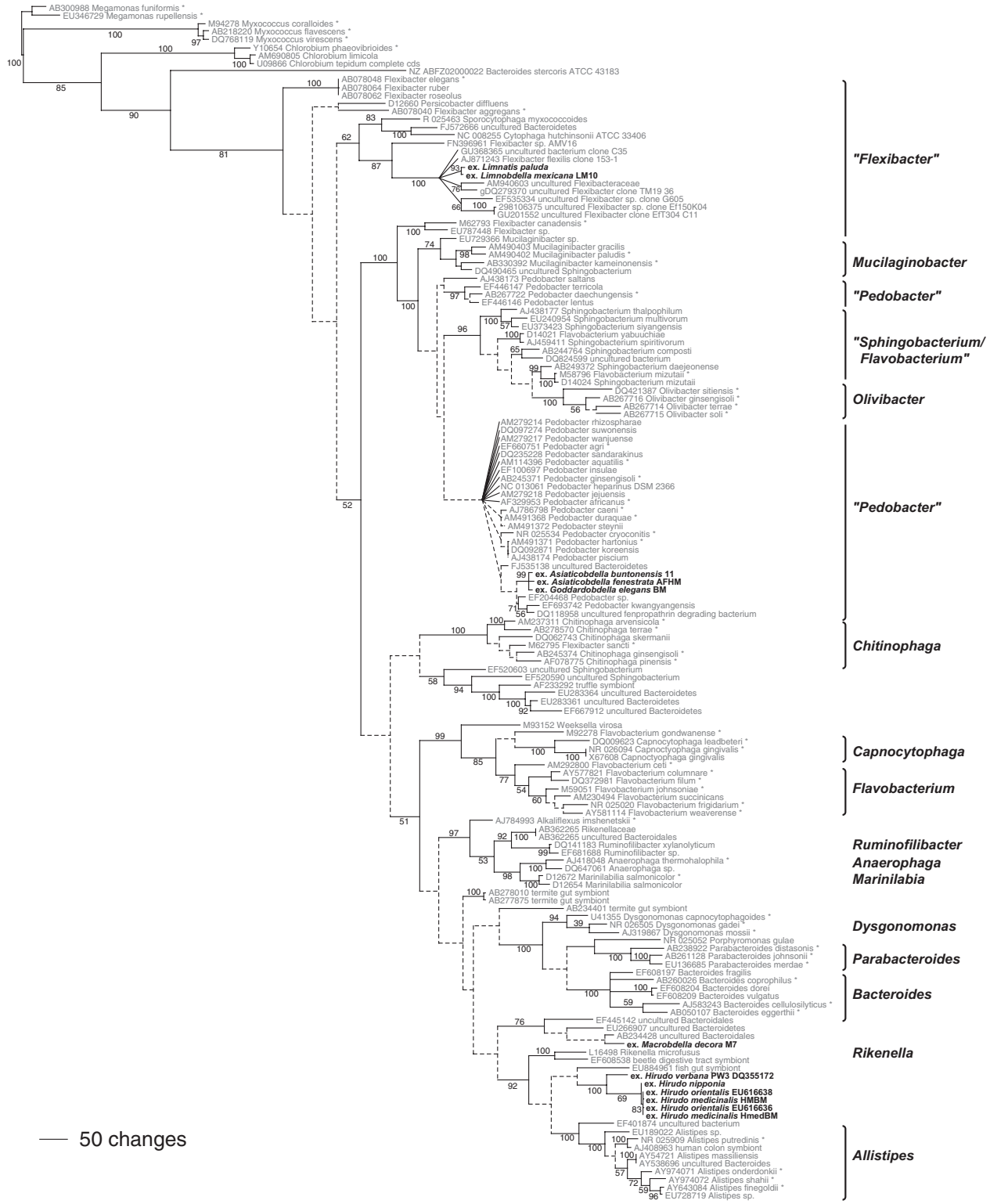

Fig. 4. Consensus of 36 equally parsimonious trees resulting from analysis of 16S rDNA sequences of a variety of species and strains of Bacteroidetes, as well as those isolated from the crop of hirudiniform leeches (bold). Asterisks denote sequences obtained from type-strains for species. Numbers at nodes are jack-knife frequencies. Relationships supported in fewer than 50 jack-knife replicates are represented by interrupted lines.

endosymbiotic bacteria each occupying three distinct mycetomal morphological types (Perkins *et al*. 2005).

Hints regarding a physiological role that the unculturable symbionts may play in the guts of their leech hosts comes from other Bacteroidetes symbionts of invertebrates. Flavobacteriaceae symbionts of termites, for example, are involved both in synthesizing essential amino acids and in recycling nitrogen from uric acid (Bourtzis and Miller, 2006). Whereas we were unable to amplify a Bacteroidetes

https://doi.org/10.1017/S0031182011000539 Published online by Cambridge University Press

16S rDNA from the Asian *H. manillensis*, given its belonging to the family Hirudinidae (Fig. 1), we would anticipate such an isolate to group with others in *Pedobacter*. Likewise, others in the mammalophilic clade of mucous membrane feeders (i.e. the Praobdellidae) should prove to be relatively closely related to the marine *Flexibacter flexis*. The inconsistency with which we were able to amplify *Aeromonas* species relative to Bacteroidetes reflects the crop dynamics reported for these two symbionts; both flourish in response to a blood-meal but the unculturable Bacteroidetes symbiont persists at higher levels for considerably longer periods after feeding (Kikuchi and Graf, 2007).

EVOLUTION OF SALIVARY PEPTIDES

Leeches have a long and storied history in medicine. Most of that seems to have been misguided optimism pertaining to the balancing of humors in the face of various ailments (Jackson, 2001). Prior even to Hippocrates describing the utility of leeches in phlebotomy, the practice of leech-mediated bloodletting already was already central in Ayurvedic and other oriental medical practices (Sawyer, 1999). Hardly the medieval European practice it is perceived to be, leeching reached a nexus under Napoleon's surgeon Broussais in the 19th century being considered safer than the raw and uncontrolled methods of venesection for phlebotomy that otherwise prevailed (Jackson, 2001). No one, it seems, has ever died from a leech bite. With the advent of clinical medicine, in perhaps the first controlled trials ever attempted, P.C.A. Lewis demonstrated the futility of leeching as applied to prognoses associated with pnuemonitis and pleurisy (Moraiba, 1996). Notwithstanding the dubious utility of leeches for the treatment of obesity, hysteria and other ailments in the 19th century, the European medicinal leech, *Hirudo medicinalis*, since then has come to play a valuable role in the postoperative treatment of venous congestion following flap and replantation surgery (Derganc and Zdravic, 1960; Batchelor *et al*. 1984). Leeches were recently approved as a medical device by the US-FDA (Rados, 2004) and the anticoagulant property of leech saliva remains the subject of some considerable scientific scrutiny. The first successful attempt at human clinical dialysis treatment, for example, was only made possible through Haas (1924) employing the anticoagulative properties of a newly purified hirudin from European medicinal leeches. It appears that this was the first use of an animal-derived compound for clinical purposes. The use of hirudin was quickly superceded by porcine-derived heparin, yet hirudin has remained a leech-derived protein of considerable interest; particularly in cases of heparin-induced thrombocytopenia (HIT) (Greinacher *et al*. 1999).

Different species of even closely related leeches exhibit known variation in the anticoagulant cocktail (Min *et al*. 2010). Evolution's own site-directed mutagenesis has determined the components necessary for a compound to function successfully. Bioactive compounds that are expressed as gene products have 'passed the test' of evolutionary selection, unlike much of the results of *in vitro* structure-activity relationships. Phylogenetic examination of site-by-site values of the relative rates of non-synonymous to synonymous substitution over deep evolutionary time can identify negatively and positively selected sites in a peptide, thus permitting the identification of functional domains in otherwise too-large antigenic molecules.

The development, characterization and structure of novel therapeutic agents and molecular diagnostic tools from living organisms is an active area of biomedical research (hundreds of published papers in the last 5 years alone just for leeches). A determination of the molecular variation of salivary bioactive peptides is critical to the development of new therapies and tools for haematology. Besides hirudin, a variety of bioactive compounds already has been isolated (typically with HPLC and peptide sequencing) from the salivary secretions of *Hirudo* and *Haementeria*. These include (Baskova and Zavalova, 2001; Salzet, 2001): the original angiogenesis-inhibiting antistasin, as well as orthologues such as ghilanten and bdellastasin, each serine-protease (factor Xa) inhibiting antistasins with potent anti-metastatic abilities; the fibrinogenolytic hementin; bdellin, a non-classical Kazal-type plasmin-trypsin inhibitor; eglin, a leucocyte/mast cell elastase inhibitor; orgelase, a heparanase-like endoglucuronidase; destabilase, which promotes the dissolution of polymerized fibrin; and calin, which, unlike the preceding protease inhibitors, acts by blocking vWF-mediated binding of platelets to collagen glycoproteins. Already, some of these bioactive compounds are in pre-clinical development or in clinical trials for drug delivery and glaucoma (orgelase), emphysema and inflammation (eglin), or reduction of tumour metastases (bdellastasin).

Hirudo medicinalis and *Haementeria ghilianii* are only distantly related (Fig. 1), having diverged evolutionarily about 200 million years ago. *Hirudo* feeds by making a cutaneous incision whereas *Haementeria* inserts a muscular proboscis. *Hirudo* species are restricted to Europe where they feed on frogs, fish and only occasionally mammals. *Haementeria* species are confined to New World Tropics where the Giant Amazonian leech specializes on anacondas, crocodilians and the plentiful aquatic mammalian fauna like capybaras. Different species of medicinal leech (*Hirudo* sp. and their close allies) are already known to have evolved to produce distinct suites of bioactive compounds in their salivary secretions (Min *et al*. 2010); for example, the North

https://doi.org/10.1017/S0031182011000539 Published online by Cambridge University Press

American medicinal leech is unique in secreting a 39 amino-acid peptide, decorsin, inhibiting platelet aggregation by blocking membrane glycoprotein IIb-IIIa integrins. Thus, the far more distantly related Giant Amazonian leech is certain to have even more radically diverged and potentially valuable components in its salivary secretions. We anticipate that hementin, which appears unique in its ability to promote the dissolution of platelet-rich clots, is just the first example.

At present, little is known regarding the genomic organization of these peptides, their copy number or to what degree orthologous loci are distributed across the various kinds of blood-feeding leeches. The available leech genome from *Helobdella robusta* (from the Joint Genome Institute) sheds little light on these questions, because this species is a predator on aquatic invertebrates rather than feeding on blood. It is surprising then to discover, through screening the annotated *H. robusta* genome via JGI's portal, three loci orthologous to antistasin (i.e. scaffold_49:464509-465641, scaffold_49:517938-518879, scaffold_49:1286706-1287330), one of which is actually expressed in *Helobdella robusta* embryos (CAXA11664 corresponds to scaffold_49:1286706-1287330), and a tandem array of six copies of leech antiplatelet protein (LAPP) on Helro1 scaffold 2, five of which are expressed in the living organism (CAWZ13451, CAXA9903, CAWZ1735, CAWZ1685, CAWZ7874). As expected, however, none of the other previously characterized leech bioactive salivary peptides appear in the *Helobdella robusta* genome draft.

The evolutionary history of leech salivary peptides associated with blood-feeding is only just beginning to be revealed. Faria *et al.* (2005) generated the first salivary Expressed Sequence Tag (EST) library from a leech; specifically *Haementeria depressa*. In that library of comparatively few (898) clones they only found ESTs homologous to LAPP, tridegin, and therostasin. In contrast, with over 2,000 transcripts we found a much wider array of bioactive proteins from the salivary glands of *Macrobdella decora* (Min *et al.* 2010). These included the antiplatelet proteins saratin and decorsin, protease inhibitors like antistasins, eglin, bdellin and hirudin, the fibrinolytic destabilase, and an endoglucuronidase. All but decorsin had previously only been known from other leech species. In addition, Min *et al.* (2010) noted lectoxins, ficolins and histidine-rich proteins among the most frequent transcripts, raising the possibility that leeches have even more biomedically interesting secretions than previously thought.

With the same techniques as were detailed in Min *et al.* (2010), we have now successfully generated additional EST libraries from the European *Hirudo verbana* and from the African *Asiaticobdella fenestrata* (Fig. 1). Even the choice of these taxa was driven by the phylogenetic premise in Fig. 1 so as to include a broad array of medicinal leeches both in terms of known phylogenetic diversity as well as geographic diversity. From that work it appears that the platelet disintegrin decorsin is unique to the Macrobdellidae, but that each of hirudin, destabilase, antistasins, saratin, eglin and bdellin predate the origin of the various medicinal leech families.

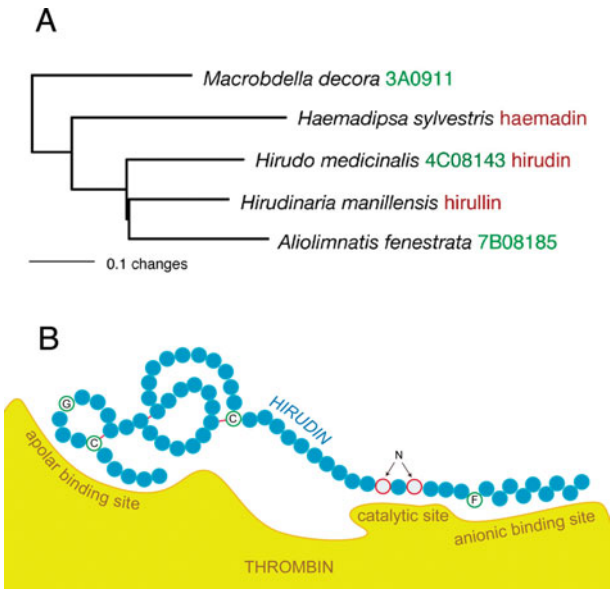

Fig. 5. Evolutionary relationships (A) for single-copy transcripts of thrombin-inhibiting hirudin orthologues from hirudiniform leeches as determined by model-based analyses in HyPhy suggest that the protein is under strong positive selection ($P<0{\cdot}05$). Hirudin, hirullin and haemadin had previously been characterized from European medicinal, Asian medicinal and Indian terrestrial leeches, respectively. Transcripts 3A0911, 4C08143 and 7B08185 were newly characterized from North American, European and African medicinal leeches, respectively. All divergences correspond to leech relationships depicted in Fig. 1. Fixed effects likelihood estimates indicate strong purifying selection (green) on two cysteines involved in disulfide bonds of the hirudin core, as well as on residues (G and F) associated with each of the two thrombin-binding domains of hirudin. Bayesian evolutionary networks revealed compensatory evolutionary changes in which an asparagine is required in exactly one of two amino acid positions that associate with thrombin's catalytic site (red).

Because this work comes at a time when we have well-corroborated evolutionary trees for medicinal leech phylogeny (Fig. 1), a variety of analytical approaches can enrich our understanding of functional and phenotypic constraints on these various protease inhibitors and other anticoagulants. It is now well established that simple pairwise comparisons of orthologous peptide sequences is insufficient (Rocha *et al.* 2005) for proper detection of rates of non-synonymous to synonymous substitutions (dN/dS or ω). The HyPhy statistical phylogenetic computing package (Kosakovsky Pond *et al.* 2005) provides three possibilities for evaluating positive

https://doi.org/10.1017/S0031182011000539 Published online by Cambridge University Press

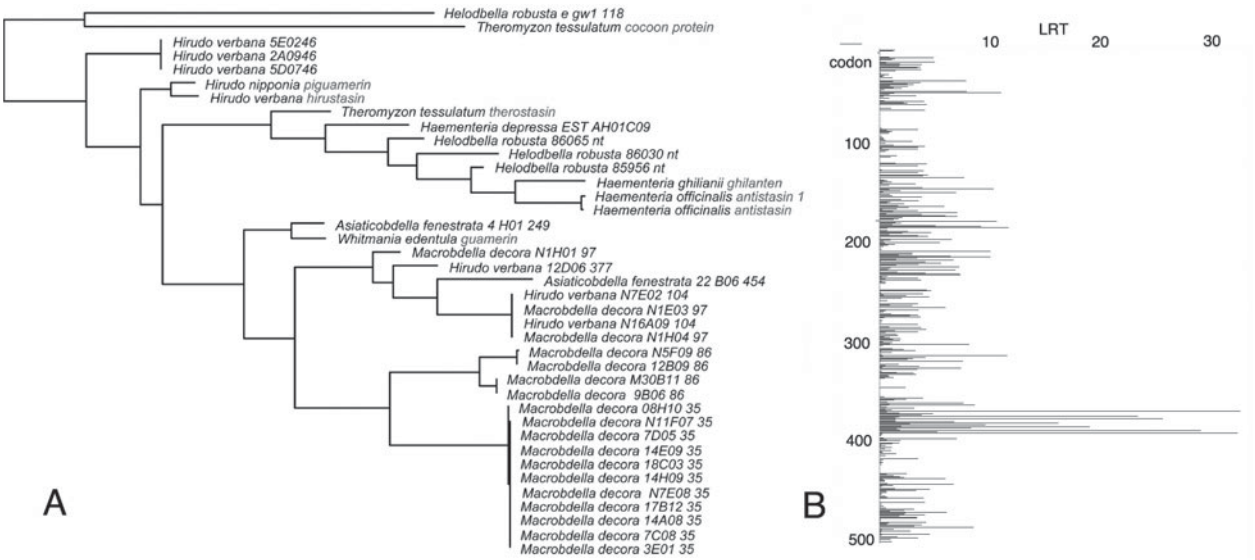

Fig. 6. Evolutionary relationships (A) for mutiple-copy transcripts of cysteine-rich antistasin family protease inhibitors from across leech diversity as determined by HyPhy and in which there is little overall concordance with leech phylogeny. A region exhibiting negative selection was detected (B) in the second antisasin domain in which likelihood ratios exceeded 10, and within which there was a single residue under Darwinian (positive) selection.

selection ($\omega >> 1{\cdot}0$) and negative selection ($\omega << 1{\cdot}0$) for whole molecules as well as for residue-level information. Amino acid residues that are under negative (or purifying) selection exhibit fewer amino acid changes than expected and, thus, may be critical to the historical functioning of an anticoagulant (useful for active site prediction). Whole peptides exhibiting overall more amino acid changes than expected are under positive selection as would be anticipated in evolutionary arms-race scenarios (Kosiol *et al*. 2008).

Implementation of the PARRIS method (Scheffler *et al*. 2006) in HyPhy entails codon-based likelihood ratio tests on whole sequences even where the history of a protein is confounded by domain-shuffling recombination; a potentially confounding phenomenon in the history of leech antistasins (Mason *et al*. 2004). While robust, the foregoing is best suited to examination of whole transcripts. In contrast, Kosakovsky Pond *et al*. (2005 - and implemented in HyPhy) have found that for relatively small data-sets, an implementation of fixed effects likelihood (FEL) models can accurately identify individual amino acid residues that deviate significantly from neutral evolution expectations. Moreover, the FEL method, and a Bayesian random effects (REL) approach overcome the lack of statistical power inherent in simple counting strategies (Pond and Frost, 2005). Both FEL and REL determinations take into account topological relationships and relative branch lengths for orthologues and their hypothesized ancestral sequences. A more computationally intensive (yet ultimately tractable) module in HyPhy employs Bayesian evolutionary network graphical modeling to identify pairs (and higher-order combinations) of residues that have changed in concert across time. This mapping of residue-residue interactions accurately predicts tertiary structural adjacency for amino acid sites operating in concert or compensatory changes in adjacent residues (Poon *et al*. 2007).

For the orthologous copies of hirudin already obtained from our EST libraries, and adding to this the orthologous hirullin and haeamdin (Fig. 5), these apparently single-copy transcripts sort out phylogenetically in a manner that exactly mirrors that expected from higher level relationships (Fig. 1). Hirudin binds irreversibly to the fibrinogen binding exosite of thrombin as well as to the catalytic active site pocket. With an inhibition constant in the picomolar range, it remains the most potent natural direct thrombin inhibitor (DTI) known (Greinacher and Warkentin, 2008). Moreover, hirudin, unlike heparin, requires no cofactor and is more effective in accessing clot-bound thrombin, promoting dissolution of mural thrombi and utility for acute coronary syndrome or deep vein thrombosis. However, hirudin has some undesirable properties. The irreversible 1:1 binding nature of hirudin to thrombin, carries with it the risk of severe bleeding in patients with reduced renal function necessary for clearing the peptide (Greinacher and Warkentin, 2008). Moreover, early reports of low antigenicity (in light of being only 65 amino acids in length) proved premature and risk of IgG-mediated anaphylaxis is 0·16% in re-exposed patients. Hirudin orthologues from other species, like *Hirudo verbana* or *Asiaticobdella fenestrata*, while retaining the requisite thrombin-binding abilities, may reveal substantially different binding affinities or antigenicities that can mitigate against unwanted side-effects of current hirudin treatment regimes. HyPhy reveals evidence of positive selection on the molecule as a whole ($P = 0{\cdot}0116$) and that (excluding

https://doi.org/10.1017/S0031182011000539 Published online by Cambridge University Press

the signal peptide region) cysteines at positions 6 and 39, a glycine at position 10, and a phenylalanine at position 56 are under strong pressure not to change. Notably, the two cysteines are the first and last of six involved in forming the three disulphide bonds in the hirudin core, a region that also contains the constrained glycine and which inhibits the fibrinogen-binding exosite of thrombin. The constrained phenylalanine corresponds to that portion of the N-terminal domain of hirudin blocking the catalytic pocket of thrombin (Markwardt, 1992). Bayesian network models reveal compensatory changes involving asparagine for non-adjacent residues known to associate with the Asp-His-Ser triad in the catalytic pocket of thrombin (Fig. 5).

This sort of comparative phylogenetic approach to understanding molecular function could prove more cost effective and more robust than alternatives like X-ray crystallography. The protease inhibiting antistasin family is revealing in this regard (Fig. 6). Antistasin has fully 20 cysteines involved in 10 disulphide bonds. Crystallography of antistasin complexed with factor Xa indicate that residues between the 17th and 18th cycteines in the C-terminal domain 2 of the protease inhibitor, are involved in the reactive site (Lapatto *et al*. 1997). Likelihood ratio plots from FEL analysis pinpoint this same region as one with codons having an overall evolutionary history of strong negative selection (Fig. 6), but within which there is a residue adjacent to the 18th cysteine that is under strong positive (i.e. Darwinian) selection having changed nine times in the course of leech evolution. Taken together these results point to this region as being sufficiently significant in the history of the molecule's functioning that most of the amino acid residues are under strong pressure not to change. However, there may also be a single residue that may be responsible for the various changes in binding affinity; changes associated with switches from inhibiting factor Xa, to factor XIIIa, to elastase or to kallikrein.

While 15 years of concerted effort has now provided a broad and compelling picture of the phylogenetic relationships of a wide array of parasitic groups from flatworms to roundworms and lice to leeches, it would be a shame if the utility of those analyses and the resulting trees were limited to mere systematic circumscription of natural groups with stable taxonomic names. While that was, perhaps, the initial *raison d'etre* of Phylogenetic Systematics, the toolbox available to phylogeneticists has become considerably more rich. With respect to the crop endosymbionts of medicinal leeches, it is the leech tree that adds depth and context to interpretation of the bacterial trees, and yet in neither case in a manner that corresponds to correlated co-speciation. In terms of potential biomedically relevant salivary secretions from leeches, it is again, in part, the leech phylogeny that adds a historically correlative framework for comparison; but also the emergent phylogeny of the proteins themselves. Together these trees within trees, and the power they provide for tracking and interpreting molecular change lead to enhanced understanding of evolutionary forces and ultimately function at the molecular level. We anticipate that these layers of intellectual pursuit driven by phylogenetic perspectives will only become richer, and the tools for their elucidation more intricate, as whole genomes of parasitic taxa (and their hosts) become more readily (and cheaply) available.

ACKNOWLEDGEMENTS

This research was supported by the US National Science Foundation, and by the Stavros Niarchos Foundation. We thank Sebastian Kvist and Alejandro Oceguera Figueroa for reviews of earlier drafts of portions of this contribution.

REFERENCES

Apakupakul, K., Siddall, M. E. and Burreson, E. M. (1999). Higher level relationships of leeches (Annelida: Clitellata: Euhirudinea) based on morphology and gene sequences. *Molecular Phylogenetics and Evolution* **12**, 350–359. doi:10.1006/mpev.1999.0639.

Bader, D. A., Roshan, U. and Stamatakis, A. (2006). Computational grand challenges in assembling the tree of life: problems and solutions. *Advances in Computers* **68**, 127–176. doi:10.1016/S0065-2458(06)68004-2.

Baskova, I. P. and Zavalova, L. L. (2001). Proteinase inhibitors from the medicinal leech *Hirudo medicinalis*. *Biochemistry Moscow* **66**, 703–717. doi:10.1023/A:1010223325313.

Batchelor, A. G. G., Davison, P. and Sully, L. (1984). The salvage of congested skin flaps by the application of leeches. *British Journal of Plastic Surgery* **37**, 358–360. doi:10.1016/0007-1226(84)90079-1.

Borda, E., Oceguera-Figueroa, A. and Siddall, M. E. (2008). On the evolution, classification and biogeography of some haemadipsoid leeches (Hirudinida: Arhynchobdellida: Hirudiniformes). *Molecular Phylogenetics and Evolution* **46**, 142–54. doi:10.1016/j.ympev.2007.09.006.

Borda, E. and Siddall, M. E. (2004). Arhynchobdellida (Annelida: Oligochaeta: Hirudinida): phylogenetic relationships and evolution. *Molecular Phylogenetics and Evolution* **30**, 213–225. doi:10.1016/j.ympev.2003.09.002.

Borda, E. and Siddall, M. E. (2011). Insights into the evolutionary history of IndoPacific bloodfeeding terrestrial leeches (Hirudinida: Arhynchobdellida: Haemadipisdae). *Invertebrate Systematics*. **24**, 456–472. doi:10.1071/IS10013.

Bourtzis, K. and Miller, T. A. (2006). *Insect Symbiosis*, Volume 2. CRC Press/Taylor and Frances, Boca Raton, FL, London.

Brooks, D. R. (1990). Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. *Systematic Biology* **39**, 14–30. doi:10.2307/2992205.

Carroll, L. (1871). *Through the Looking Glass and What Alice Found There*. Macmillan, London.

Croizat, L. (1958). *Panbiogeography*, Volumes 1, 2a, 2b. Published by author. Caracas, Venezuela.

Derganc, M. and Zdravic, F. (1960). Venous congestion of flaps treated by application of leeches. *British Journal of Plastic Surgery* **13**, 187–192.

Dunlap, P. V., Ast, J. C., Kimura, S., Fukui, A., Yoshino, T. and Endo, H. (2007). Phylogenetic analysis of host-symbiont specificity and codivergence in bioluminescent symbioses. *Cladistics* **23**, 507–532. doi:10.1111/j.1096-0031.2007.00157.x.

Faria, F., Junqueira-de-Azevedo Ide, L., Ho, P. L., Sampaio, M. U. and Chudzinski-Tavassi, A. M. (2005). Gene expression in the salivary complexes from *Haementeria depressa* leech through the generation of expressed sequence tags. *Gene* **349**, 173–185. doi:10.1016/j.gene.2004.12.022.

Farris, J. S., Albert, V. A., Källersjö, M., Lipscomb, D. and Kluge, A. G. (1996). Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12**, 99–124. doi:10.1111/j.1096-0031.1996.tb00196.x.

Giribet, G. (2008). Assembling the lophotrochozoan (=spiralian) tree of life. *Philosophical Transactions of the Royal Society B* **363**, 1513–1522. doi:10.1098/rstb.2007.2241.

Goloboff, P. A., Farris, J. S. and Nixon, K. C. (2008). TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774–786. doi:10.1111/j.1096-0031.2008.00217.x.

Graf, J. (1999). Symbiosis of *Aeromonas veronii* Biovar sobria and *Hirudo medicinalis*, the medicinal leech: a novel model for digestive tract associations. *Infection and Immunity* **67**, 1–7.

Graf, J. (2000). Symbiosis of *Aeromonas* and *Hirudo medicinalis*, the medicinal leech. *American Society for Microbiology News* **66**, 147–153.

Graf, J. (2002). The effect of symbionts on the physiology of *Hirudo medicinalis*, the medicinal leech. *Invertebrate Reproduction and Development* **41**, 269–275.

Greinacher, A., Völpel, H., Janssens, U., Hach-Wunderle, V., Kemkes-Matthes, B., Eichler, P., Mueller-Velten, H. G. and Pötzsch, B. (1999). Recombinant hirudin (Lepirudin) provides safe and effective anticoagulation in patients with heparin-induced thrombocytopenia. *Circulation* **99**, 73–80.

Greinacher, A. and Warkentin, T. E. (2008). The direct thrombin inhibitor hirudin. *Thrombosis and Haemostasis* **99**, 819–29.

Haas, G. (1924). Versuche der Blutauswaschung am Lebenden mit Hilfe der Dialyse. *Wiener Klinische Wochenschrift* **4**, 13–14. doi:10.1007/BF01745400.

Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C. K., Braun, E. L., Braun, M. J., Chojnowski, J. L., Cox, W. A., Han, K.-L., Harshman, J., Huddleston, C. J., Marks, B. D., Miglia, K. J., Moore, W. S., Sheldon, F. H., Steadman, D. W., Witt, C. C., and Yuri, T. (2008). A phylogenetic study of birds reveals their evolutionary history. *Science* **230**, 1763–1768. doi:10.1126/science.1157704

Hittinger, C. T., Johnston, M., Tossberg, J. T. and Rokas, A. (2010). Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *Proceedings of the National Academy of Sciences*, USA **107**, 1476–1481. doi: 10.1073/pnas.0910449107.

Jackson, W. A. (2001). A short guide to humoral medicine. *TRENDS in Pharmacological Sciences* **22**, 487–489.

Keading, A. J., Ast, J. C., Pearce, M. M., Urbanczyk, H., Kimura, S., Endo, H., Nakamura, M. and Dunlap, P. V. (2007). Phylogenetic diversity and cosymbiosis in the bioluminescent symbioses of "*Photobacterium mandapamensis*". *Applied and Environmental Microbiology* **73**, 3173–3182. doi:10.1128/AEM.02212-06.

Kiskuchi, Y. and Fukatsu, T. (2002). Endosymbiotic bacteria in the esophageal organ of glossiphoniid leeches. *Applied and Environmental Microbiology* **68**, 4637–4641. doi:10.1128/AEM.68.9.4637-4641.2002.

Kikuchi, Y. and Graf, J. (2007). Spatial and temporal population dynamics of a naturally occurring two-species microbial community inside the digestive tract of the medicinal leech. *Applied and Environmental Microbiology* **73**, 1984–1991. doi:10.1128/AEM.01833-06.

Kimbell, J. R. and McFall-Ngai, M. J. (2003). The squid-Vibrio symbioses: from demes to genes. *Integrative and Comparative Biology* **43**, 254–260. doi: 10.1093/icb/43.2.254.

Kimbell, J. R., McFall-Ngai, M. J. and Roderick, G. K. (2002). Two genetically distinct populations of bobtail squid, *Euprymna scolopes*, exist on the island of Oahu. *Pacific Science* **56**, 347–355. http://hdl.handle.net/10125/2570.

Kosakovsky Pond, S. L., Frost, S. D. W. and S. V. Muse. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679. doi:10.1093/bioinformatics/bti079.

Kosiol, C., Vinar, T., da Fonseca, R. R., Hubisz, M. J., Bustamante, C. D., Nielsen, R. and Siepel, A. (2008). Patterns of positive selection in six mammalian genomes. *PLoS Genetics* **4**, e1000144. doi:10.1371/journal.pgen.1000144.

Kristensen, D. M., Kannan, L., Coleman, M. K., Wolf, Y. I., Sorokin, A., Koonin, E. V. and Mushegian, A. (2010). A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* **26**, 1481–1487.

Kvist, S., Montanari, S. A., Yi, H., Fuks, B. and Siddall, M. E. (2011). Teaching Biodiversity & Evolutionary Biology in a North American Marine Coastal Environment. *The American Biology Teacher* **73**, 72–77. doi:10.1525/abt.2011.73.2.4.

Lakatos, I. (1971). History of science and its rational reconstruction. *Proceedings of the Biennial Meeting of the Philosophy of Science Association* **1970**, 91–136 http://www.jstor.org/stable/495757.

Lapatto, R., Krengel, U., Schreuder, H. A., Arkema, A., de Boer, B., Kalk, K. H., Hol, W. G., Grootenhuis, P. D., Mulders, J. W., Dijkema, R., Theunissen, H. J. and Dijkstra, B. W. (1997). X-ray structure of antistasin at 1·9 A resolution and its modelled complex with blood coagulation factor Xa. *EMBO Journal* **16**, 5151–5161. doi:10.1093/emboj/16.17.5151.

Laufer, A. S., Siddall, M. E. and Graf, J. (2008). Characterization of the digestive-tract microbiota of *Hirudo orientalis*, a European medicinal leech. *Applied and Environmental Microbiology* **74**, 6151–6154. doi:10.1128/AEM.00795-08.

Mason, T. A., McIlroy, P. J. and Shain, D. H. (2004). A cysteine-rich protein in the Theromyzon (Annelida: Hirudinea) cocoon membrane. *Federation of European Biochemical Societies Letters* **561**, 167–172. doi:10.1016/S0014-5793(04)00167-X.

Markwardt, F. (1992). Hirudin: The promising antithrombotic. *Cardiovascular Drug Reviews* **10**, 211–232. doi:10.1111/j.1527-3466.1992.tb00247.x.

McLaughlin, D. J., Hibbett, D. S., Lutzoni, F., Spatafora', J. W. and Vilgalys, R. (2009). The search for the fungal tree of life. *Trends in Microbiology* **17**, 488–497. doi:10.1016/j.tim.2009.08.001.

Min, G-S., Sarkar, I. N., and Siddall, M. E. (2010). Salivary transcriptome of the North American medicinal leech, *Macrobdella decora*. *Journal of Parasitology* **96**, 1211–1221. doi:10.1645/GE-2496.1.

Morabia, A. (1996). P. C. A. Louis and the birth of clinical epidemiology. *Journal of Clinical Epidemiology* **49**, 1327–1333. doi:10.1016/S0895-4356(96)00294-6.

Moran, N. A. (2001). The coevolution of bacterial endosymbionts and phloem-feeding insects. *Annals of the Missouri Botanical Garden* **88**, 35–44. http://www.jstor.org/stable/2666130.

Munro, R., Siddall, M., Desser, S. S., and Sawyer, R. T. (1992). The leech as a tool for studying comparative haematology. *Comparative Haematology International* **2**, 75–78. doi:10.1007/BF00186263.

Nishiguchi, M. K. (2002). Host-symbiont recognition in the environmentally transmitted sepiolid squid-*Vibrio* mutualism. *Microbial Ecology* **44**, 10–18. doi:10.1007/BF03036870.

Nishiguchi, M. K., Lopez, J. E. and Boletzky, S. V. (2004). Enlightenment of old ideas from new investigations: more questions regarding the evolution of bacteriogenic light organs in squids. *Evolution and Development* **6**, 41–49. doi:10.1111/j.1525-142X.2004.04009.x.

Nishiguchi, M. K., Ruby, E. G. and McTall-Ngai, M. J. (1998). Competitive dominance among strains of luminous bacteria provides an unusual form of evidence for parallel evolution in sepiolid squid-vibrio symbioses. *Applied and Environmental Microbiology* **64**, 3209–3213.

Nixon, K. C. (1999). The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* **15**, 407–414. doi:10.1111/j.1096-0031.1999.tb00277.x.

Nogge, G. (1981). Significance of symbionts for the maintenance of an optional nutritional state for successful reproduction in haematophagous arthropods. *Parasitology* **82**, 101–104.

Nylander, J. A. A., Olsson, U., Alström, P. and Sanmartín, I. (2008). Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to Dispersal-vicariance Analysis of the thrushes (Aves: Turdus). *Systematic Biology* **57**, 257–268. doi: 10.1080/10635150802044003.

Oceguera-Figueroa, A., Phillips, A. J., Pacheco-Chaves, B., Reeves, W. K. and Siddall, M. E. (2011). Phylogeny of macrophagous leeches (Hirudinea, Clitellata) based on molecular data and evaluation of the barcoding locus. *Zoologica Scripta* **40**, 194–203. doi:10.1111/j.14636-409.2010.00465.x.

Page, R. D. M. (1990). Component analysis: a valiant failure? *Cladistics* **6**, 119–136. doi:10.1111/j.1096-0031.1990.tb00532.x.

Perkins, S. L., Budinoff, R. B. and Siddall, M. E. (2005). New gammaproteobacteria associated with blood-feeding leeches and a broad phylogenetic analysis of leech endosymbionts. *Applied and Environmental Microbiology* **71**, 5219–5224. doi:10.1128/AEM.71.9.5219-5224.2005.

Phillips, A. J., Arauco-Brown, R., Oceguera-Figueroa, A., Gomez, G. P., Beltran, M., Lai, Y.-T. and Siddall, M. E. (2010). *Tyrannobdella rex* n. gen. n. sp. and the evolutionary origins of mucosal leech infestations. *PLoS ONE* **5**, e10057. doi:10.1371/journal.pone.0010057.

Phillips, A. J. and Siddall, M. E. (2005). Phylogeny of the New World medicinal leech family Macrobdellidae (Oligochaeta: Hirudinida: Arhynchobdellida). *Zoologica Scripta* **34**, 559–564. doi:10.1111/j.1463-6409.2005.00210.x.

Phillips, A. J. and Siddall, M. E. (2009). Poly-paraphyly of Hirudinidae: Many lineages of medicinal leeches. *BMC Evolutionary Biology* **9**, 246–256. doi:10.1186/1471-2148-9-246.

Pond, S. L. K. and Frost, S. D. W. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531–2533. doi: 10.1093/bioinformatics/bti320.

Poon, A. F. Y., Lewis, F. I., Kosakovsky Pond, S. L. and Frost, S. D. W. (2007). An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. *PLoS Computational Biology* **3**, e231. doi:10.1371/journal.pcbi.0030231.

Rados, C. (2004). Beyond bloodletting: FDA gives leeches a medical makeover. *FDA Consumer* **38**, 9.

Ree, R. H. and Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* **57**, 4–14. doi: 10.1080/10635150701883881.

Rio, R. V. M. (2008). Tsetse fly and medicinal leech symbioses: providing insights into microbial species interactions within gastrointestinal systems. *Vie Milieu* **58**, 153–163.

Rio, R. V. M., Anderegg, M. and Graf, J. (2007). Characterization of a catalase gene from *Aeromonas veronii*, the digestive-tract symbiont of the medicinal leech. *Microbiology* **153**, 1897–1906. doi:10.1099/mic.0.2006/003020-0

Rio, R. V. M., Maltz, M., McCormick, B., Reiss, A. and Graf, J. (2009). Symbiont succession during embryonic development of the European medicinal leech, *Hirudo verbana*. *Applied and Environmental Microbiology* **75**, 6890–6895. doi:10.1128/AEM.01129-09.

Rocha, E. P., Smith, J. M., Hurst, L. D., Holden, M. T., Cooper, J. E., Smith, N. H. and Feil, E. J. (2005). Comparisons of dN/dS are time dependent for closely related bacterial genomes. *Journal of Theoretical Biology* **239**, 226–35.

Ronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* **46**, 195–203. doi: 10.1093/sysbio/46.1.195.

Salzet, M. (2001). Anticoagulants and inhibitors of platelet aggregation derived from leeches. *Federation of European Biochemical Societies Letters* **429**, 187–192. doi:10.1016/S0014-5793(01)02212-8.

Sawyer, R. T. (1999). The trade in medicinal leeches in the southern Indian Ocean in the nineteenth century. *Medical History* **43**, 241–245.

Scheffler, K., Martin, D. P. and Seoighe, C. (2006). Robust inference of positive selection from recombining coding sequences. *Bioinformatics* **22**, 2493–2499. doi: 10.1093/bioinformatics/btl427.

Siddall, M. E. (2002). Phylogeny and revision of the leech family Erpobdellidae (Hirudinida: Oligochaeta). *Invertebrate Systematics* **16**, 1–6. doi:10.1071/IT01011.

Siddall, M. E., Apakupakul, K., Burreson, E. M., Coates, K. A., Erséus, C., Gelder, S. R., Källersjö, M. and Trapido-Rosenthal, H. (2001). Validating Livanow: Molecular data agree that leeches, Branchiobdellidans and *Acanthobdella peledina* form a monophyletic group of Oligochaetes. *Molecular Phylogenetics and Evolution* **21**, 346–351. doi:10.1006/mpev.2001.1021.

Siddall, M. E., Budinoff, R. B., and Borda, E. (2005). Phylogenetic evaluation of systematics and biogeography of the leech family Glossphoniidae. *Invertebrate Systematics* **19**, 105–112. doi:10.1071/IS04034.

Siddall, M. E. and Burreson, E. M. (1998). Phylogeny of leeches (Hirudinea) based on mitochondrial cytochrome *c* Oxidase Subunit I. *Molecular Phylogenetics and Evolution* **9**, 156–162. doi:10.1006/mpev.1997.0455.

Siddall, M. E. and Burreson, E. M. (1995). Phylogeny of the Euhirudinea: Independent evolution of blood feeding by leeches? *Canadian Journal of Zoology* **73**, 1048–1064. doi:10.1139/z95-125.

Siddall, M. E., Perkins, S. L. and Desser, S. S. (2004). Leech mycetome endosymbionts are a new lineage of alphaproteobacteria related to the Rhizobiaceae. *Molecular Phylogenetics and Evolution* **30**, 178–186. doi:10.1016/S1055-7903(03)00184-2.

Siddall, M. E., Trontelj, P., Utevsky, S. Y., Nkamany, M. and Macdonald, K. S. (2007*a*). Diverse molecular data demonstrate that commercially available medicinal leeches are not *Hirudo medicinalis*. *Proceedings of the Royal Society B* **274**, 1481–1487. doi: 10.1098/rspb.2007.0248.

Siddall, M. E., Worthen, P. L., Johnson, M. and Graf, J. (2007*b*). Novel role for *Aeromonas jandaei* as a digestive tract symbiont of the North American medicinal leech. *Applied and Environmental Microbiology* **73**, 655–658. doi: 10.1128/AEM.01282-06.

Silver, A., Rabinowitz, N. M., Küffer, S. and Graf, J. (2007). Identification of *Aeromonas veronii* genes required for colonization of the medicinal leech, *Hirudo verbana*. *Journal of Bacteriology* **189**, 6763–6772. doi:10.1128/JB.00685-07.

Soto, W., Gutierrez, J., Remmenga, M. D. and Nishiguchi, M. K. (2009). Salinity and temperature effects of physiological responses of *Vibrio fischeri* from diverse ecological niches. *Microbial Ecology* **57**, 140–150. doi:10.1007/s00248-008-9412-9.

Utevsky, S. Y. and Trontelj, P. (2004). Phylogenetic relationships of fish leeches (Hirudinea, Piscicolidae) based on mitochondrial DNA sequences and morphological data. *Zoologica Scripta* **33**, 375–385. doi:10.1111/j.0300-3256.2004.00156.x.

Warren, R. L., Sutton, G. G., Jones, S. M. J. and Holt, R. A. (2007). Assembling millions of short DNA sequences using SSAKE. *Bioinformatics* **23**, 500–501. doi:10.1093/bioinformatics/btl629.

Whitaker, I. S., Kamya, C., Azzopardi, E. A., Graf, J., Kon, M. and Lineaweaver, W. C. (2009). Preventing infective complications following leech therapy: Is practice keeping pace with current research? *Microsurgery* **29**, 619–625. doi:10.1002/micr.20666.

Williams, J. I. and Burreson, E. M. (2006). Phylogeny of the fish leeches (Oligochaeta, Hirudinida, Piscicolidae) based on nuclear and mitochondrial genes and morphology. *Zoologica Scripta* **35**, 627–639. doi:10.1111/j.1463-6409.2006.00246.x.

Worthen, P. L., Gode, C. J. and Graf, J. (2006). Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis. *Applied and Environmental Microbiology* **72**, 4775–4781. doi:10.1128/AEM.00356-06.