EVOLUTION & DEVELOPMENT 8:6, 491–501 (2006)

Lessons from leeches: a call for DNA barcoding in the lab

Alexandra E. Bely[a,*] and David A. Weisblat[b]

[a]Department of Biology, University of Maryland, College Park, MD 20742, USA
[b]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720-3200, USA
*Author for correspondence (email: abely@umd.edu)

SUMMARY Many evolution of development labs study organisms that must be periodically collected from the wild. Whenever this is the case, there is the risk that different field collections will recover genetically different strains or cryptic species. Ignoring this potential for genetic variation may introduce an uncontrolled source of experimental variability, leading to confusion or misinterpretation of the results. Leeches in the genus *Helobdella* have been a workhorse of annelid developmental biology for 30 years. Nearly all early *Helobdella* research was based on a single isolate, but in recent years isolates from multiple field collections and multiple sites across the country have been used. To assess the genetic distinctness of different isolates, we obtained specimens from most *Helobdella* laboratory cultures currently or recently in use and from some of their source field sites. From these samples, we sequenced part of the mitochondrial gene cytochrome oxidase I (COI). Sequence divergences and phylogenetic analyses reveal that, collectively, the *Helobdella* development community has worked on five distinct species from two major clades. Morphologically similar isolates that were thought to represent the same species (*H. robusta*) actually represent three species, two of which coexist at the same locality. Another isolate represents part of a species complex (the "*H. triserialis*" complex), and yet another is an invasive species (*H. europaea*). We caution researchers similarly working on multiple wild-collected isolates to preserve voucher specimens and to obtain from these a molecular "barcode," such as a COI gene sequence, to reveal genetic variation in animals used for research.

INTRODUCTION

Since the birth of evolutionary developmental biology in the late 1970s, there has been a growing emphasis on organisms other than major model systems for developmental studies. Many researchers have ventured from the handful of traditional developmental model organisms (e.g., *Drosophila melanogaster*, *Caenorhabditis elegans*, *Danio rerio*, *Xenopus laevis*) to study taxa that hold key phylogenetic positions for understanding body plan evolution (e.g., Shankland and Seaver 2000; Tessmar-Raible and Arendt 2003; Gerhart et al. 2005; Martindale 2005), to investigate the evolution of features well characterized in a model organism (e.g., body regionalization by Hox genes, segmentation, vertebrate limb development: Grenier and Carroll 2000; Davis and Patel 2002; Hughes and Kaufman 2002; Metscher et al. 2005; Peel et al. 2005; Seaver and Kaneshige 2006), to investigate the microevolution of development through the study of closely related taxa (e.g., Jeffery 2001; Simpson 2002), or to study developmental phenomena not present or not easily studied in major model systems (e.g., regeneration, phenotypic plasticity, agametic reproduction, polyembryony: Emlen and Nijhout 1999; Bely and Wray 2001; Abouheif and Wray 2002; Saló and Baguñà 2002; Sánchez Alvarado 2004; Zhurov et al. 2004). Expanding the list of developmental subjects to include these less-well-studied organisms is essential for investigating novel evolution of development questions. Although the wealth of resources, techniques, and background information available for major model systems offers tremendous advantages that are unlikely to disappear, organisms other than these major models are nevertheless sure to be used increasingly in the field of evolutionary developmental biology, especially as new questions are investigated and broadly applicable techniques become available (e.g., RNAi, micro-arrays).

There are often significant challenges to working on organisms other than the few major model systems (for brevity, we refer to these as "nonmodel systems," recognizing of course that many of these organisms have large research communities working on them). Above and beyond the obvious challenge of needing to develop or adapt techniques to these less studied organisms, one of the most fundamental challenges is the lack of stocks and stock centers. Nonmodel organisms may be maintained by one or just a few labs. Instead of obtaining a standard strain from a stock center or colleague, those working on nonmodel systems often collect their organisms directly from the field. If the organisms cannot be maintained in the lab over multiple generations (as is the case for many marine invertebrates), then repeated field collections must be made. Even for those species that can be cultured in the lab, rearing conditions are typically less well

optimized than for major model systems, resulting in a higher risk of colony loss or the need to replenish colonies occasionally with wild-collected individuals. Field-collecting research organisms means that different researchers, or even the same researcher over time, may work on different strains, or even different species, either knowingly or unknowingly.

Freshwater leeches in the genus *Helobdella* (Annelida: Clitellata: Rhynchobdellida: Glossiphoniidae) have been a workhorse of annelid developmental biology for over 20 years (Weisblat and Huang 2001). They are arguably the best-characterized lophotrochozoan group with respect to embryonic development and the genome of an inbred strain is currently being sequenced as part of a lophotrochozoan genome initiative by the Joint Genome Institute. *Helobdella* leeches have featured prominently in discussions of body plan evolution, especially the evolution of segmentation (Seaver 2003). These leeches are common inhabitants of freshwaters, especially in North and South America. They are relatively small in size ($\sim$ 1–3 cm in length), hermaphroditic, and feed on other aquatic invertebrates such as snails, oligochaetes, and insects (Sawyer 1986). *Helobdella* leeches present several important advantages for experimental developmental work (Weisblat and Huang 2001): they produce easily accessible, roughly synchronized broods of up to $\sim$ 200 embryos that are relatively large, develop via stereotyped cleavages, can develop to hatching stage in a simple salt solution, and are amenable to a range of experimental manipulations.

Helobdella cultures require considerable care, however, and maintaining these leeches can be troublesome. Although they can be kept simply in small aquaria or bowls of freshwater, cultured animals are prone to infections that can compromise or even wipe out entire colonies. Also, because these leeches feed only on live material, primarily snails, large quantities of feeder snails must be collected or reared to sustain them. Feeder snail shortages, even temporary ones, can cause rapid crashes in leech colonies. For such reasons, *Helobdella* cultures occasionally need to be replenished or restarted with wild-collected individuals. In recent years, *Helobdella* cultures have been maintained by up to half a dozen development labs across the United States, but there are no standard strains or stock centers.

A number of different *Helobdella* isolates have been used in development studies. In the 1970s and 1980s, studies were based almost entirely on a species identified as *H. triserialis* originating from Golden Gate Park, San Francisco, CA. In 1992, a new species of *Helobdella*, *H. robusta*, was described from a small tributary of the American River running through Sacramento, CA (Shankland et al. 1992), and this species was used increasingly for development studies. Laboratory colonies of both *H. triserialis* and *H. robusta* were periodically supplemented (or even restarted) with wild-collected individuals from original source locations, leading to potential changes in the genetic makeup of laboratory colonies over time. (Although unlikely, there is also the possibility that leeches from the field sites of wild-collected feeder snails could have been unintentionally introduced into laboratory cultures, potentially introducing yet another source of genetic variation.) As laboratories were established in different regions of the U.S., investigators encountered and began culturing and working on *Helobdella* populations from at least three additional locations: a sturgeon farm near Galt, CA, a creek and fish ponds in Austin, TX, and a reservoir in Tarrytown, NY. Based on limited morphological inspection relying largely on body pigmentation, these new isolates have generally been thought to be *H. robusta* or a very closely related species. However, although pigmentation patterns are reliable species-specific characters in some leech genera (e.g., *Hirudo*: Trontelj and Utevsky 2005), they can be unreliable in *Helobdella* (Siddall and Borda 2003).

Collectively, then, *Helobdella* developmental studies have been carried out on isolates stemming from a large number of collections made from at least five different localities across the country. Through the years, a variety of experiments have been performed on multiple isolates and most of these have yielded identical results, indicating that many aspects of development are similar between isolates. However, some differences in developmental timing, body pigmentation, nuclear DNA sequence, and even cell fates (Huang et al. 2002; Kuo and Shankland 2004) have been noted between isolates, suggesting the possibility that isolates represent more taxa than initially recognized. Unrecognized genetic variation within or between laboratory populations introduces an uncontrolled source of experimental variability, which can lead to ambiguous or misinterpreted results. Therefore, to assess the variation among *Helobdella* isolates in current or recent use, we obtained *Helobdella* specimens from most active laboratory cultures as well as from some of the field sites from which these isolates were originally collected. We sequenced from these specimens part of the mitochondrial gene cytochrome oxidase I (COI), a relatively fast-evolving gene that can be an excellent marker for generating a molecular "barcode" for taxa (Hebert et al. 2003; Smith et al. 2006). This gene, in a range of animals including annelids, has also proven useful in resolving relationships within species and among closely related species (e.g. Bely and Wray 2004; Oceguera-Figueroa et al. 2005; Ross and Shoemaker 2005; Trontelj and Utevsky 2005; Weigt et al. 2005). It can therefore be expected to elucidate the relationships among *Helobdella* laboratory isolates to provide a phylogenetic framework for interpreting developmental differences among them.

METHODS

Specimen collections and DNA extractions

We obtained *Helobdella* specimens from laboratory cultures being used for developmental work and from field sites from which

Table 1. *Helobdella* specimens from laboratory cultures and source populations

Identifier[1,2]	Name used by researchers	Collection site and year[3]	GenBank Accession # (COI)
CAGA-1	*H.* sp. (Galt)	D. Weisblat lab culture, 2001; original field collection from sturgeon fish ponds at The Fishery, Galt, CA	DQ995297
CAGA-2	*H.* sp. (Galt)	Ditch near sturgeon fish ponds at The Fishery, Galt, CA 2004 (collected by D.-H. Kuo)	DQ995298
CASA-1	*H. robusta* (Sacramento)	cDNA library stage 7–10, D. Weisblat lab culture, 1996; original field collection from an American River tributary near California Exposition and State Fair grounds, Sacramento, CA	DQ995299
CASA-2	*H. robusta* (Sacramento)	D. Weisblat lab culture, ca. 1998; original field collection from an American River tributary near California Exposition and State Fair grounds, Sacramento, CA	AF178680 (Boore and Brown 2000)
CASA-3	*H. robusta* (Sacramento)	American River tributary near California Exposition and State Fair grounds, Sacramento, CA, 2001	DQ995300
CASA-4	*H. robusta* (Sacramento)	D. Weisblat lab culture, inbred line BS7-1X (line used for complete genome sequence), 2004; original field collection from an American River tributary near California Exposition and State Fair grounds, Sacramento, CA, ca. 2002.	DQ995301
CASA-5	*H. robusta* (Sacramento)	American River tributary near California Exposition and State Fair grounds, Sacramento, CA, 2004	DQ995302
CASF	*H. triserialis*	S. Bissen lab culture, 2001; original field collection from pond in Golden Gate Park, San Francisco, CA	DQ995303
CAUCB	*H.* sp. (UCB-VLSB)	Artificial pond on roof of Valley Life Sciences Building, University of California, Berkeley, CA, 2004	DQ995304
NYTA	*H. robusta* (Valhalla)	C. Wedeen lab culture, ca. 1996; original field collection from Tarrytown Reservoir in Tarrytown, NY (near Valhalla College)	DQ995305
TXAU-1	*H. robusta* (Austin)	M. Shankland lab culture, 2001; original field collection from Shoal Creek, Austin, TX	DQ995306
TXAU-2	*H. robusta* (Austin)	Fish pond at Brackenridge Field Laboratory, Austin, TX, 2004 (collected by D.-H. Kuo)	DQ995307
TXAU-3	*H. robusta* (Austin)	Fish pond at Brackenridge Field Laboratory, Austin, TX, 2004 (collected by D.-H. Kuo)	DQ995308
TXAU-4	*H. robusta* (Austin)	Shoal Creek, Austin, TX, 2004 (collected by D.-H. Kuo)	DQ995309
TXAU-5	*H. robusta* (Austin)	Shoal Creek, Austin, TX, 2004 (collected by D.-H. Kuo)	DQ995310

[1]Specimen names include the two letters referring to the U.S. state and two letters referring to the (closest) city of the original collection site.
[2]One individual was sequenced for CASA-1, CASA-2, CASF, CAUCB, NYTA, and TXAU-4. Two individuals were sequenced and found to be identical in COI sequence for CAGA-1, CAGA-2, CASA-3, CASA-5, TXAU-1, TXAU-2, TXAU-3, TXAU-5. Three individuals were sequenced and found to be identical in COI sequence for CASA-4.
[3]For collections made from laboratory cultures, the presumed field collection site is also listed.

laboratory cultures have been established, as well as from a few other sources (Tables 1 and 2). In most cases, multiple individuals per culture or field collection were obtained and sequenced (see Table 1). Specimens were placed in 95–100% ethanol and ultimately frozen at $-20^{\circ}C$ for long-term storage. For DNA extractions, a small longitudinal strip of tissue was removed from one side of the leech, avoiding any gut tissue (and thus avoiding any potentially contaminating gut contents). Removing this tissue sample allowed us to extract DNA while retaining essentially uncompromised voucher specimens (i.e., no unique structures were removed). DNA was extracted using the DNeasy Tissue Kit (Qiagen, Valencia, CA, USA).

PCR and sequencing

A $\sim$ 700 bp region of the mitochondrial gene COI was amplified by PCR using LCO1490 (Folmer et al. 1994) or COI-F+(5′-GGAGTATGAGCAGGAATAGTAGGIAC-3′, corresponding to amino acids GVWAGMVGT) as a forward primer and HCO2198 (Folmer et al. 1994) as a reverse primer. Our PCR amplification profile was 94°C for 5–10 min, 35–40 cycles of 94°C for 30 sec, 45–49°C for 30 sec, 72°C for 30–60 sec, and a final extension at 72°C for 5–10 min. Amplified products were sequenced directly on an automated sequencer. Sequences were edited and assembled into alignments in Sequencher (version 4.1.2, Gene Codes Corporation, Ann Arbor, MI, USA).

Table 2. *Helobdella* taxa (other than those listed in Table 1) and close out-groups included in this study

Species	Collection site	GenBank Accession # (COI)	Reference
Helobdella			
H. sp. (MEX)	Mexico[1]	DQ995311	This study
H. bolivianita	Bolivia	AF329053	Siddall and Borda (2003)
H. elongata[2]	Michigan, USA	AF329045	Siddall and Borda (2003)
H. europaea (AU-1)	Australia	AF329052	Siddall and Borda (2003)
H. europaea[3] (AU-2)	Australia	AY856047	Siddall and Budinoff (2005)
H. europaea (GE)	Germany	AY576008	Pfeiffer et al. (2004)
H. europaea (NZ)	New Zealand	AY856049	Siddall and Budinoff (2005)
H. europaea (SA)	South Africa	AY856048	Siddall and Budinoff (2005)
H. fusca	Michigan, USA	AF329038	Siddall and Borda (2003)
H. lineata	Michigan, USA	AF329039	Siddall and Borda (2003)
H. michaelseni[2]	Chile	AF536824	Siddall and Borda (2003)
H. nununununojensis (BO-1)	Bolivia	AF329048	Siddall and Borda (2003)
H. nununununojensis (BO-2)	Bolivia	AF329047	Siddall and Borda (2003)
H. papillata (MI-1)	Michigan, USA	AF329042	Siddall and Borda (2003)
H. papillata (MI-2)	Michigan, USA	AF329043	Siddall and Borda (2003)
H. papillata (VA)	Virginia, USA	AF329046	Siddall and Borda (2003)
H. paranensis[4]	Ururguay	AF116019	Apakupakul et al. (1999)
H. ringueleti	Bolivia	AF329051	Siddall and Borda (2003)
H. sorojchi (BO-1)	Bolivia	AF329050	Siddall and Borda (2003)
H. sorojchi (BO-2)	Bolivia	AF329049	Siddall and Borda (2003)
H. stagnalis (FR)	France	AF116018	Apakupakul et al. (1999)
H. stagnalis (OH)	Ohio, USA	AF329040	Siddall and Borda (2003)
H. stagnalis (UK)	England, UK	AF329041	Siddall and Borda (2003)
H. transversa	Michigan, USA	AF329044	Siddall and Borda (2003)
H. triserialis (BO)	Bolivia	AF329054	Siddall and Borda (2003)
Out-groups			
Haementeria ghilianii	French Guiana	AF329035	Siddall and Borda (2003)
Ha. gracilis	Uruguay	AF329034	Siddall and Borda (2003)
Ha. lutzi	Ecuador	AF329033	Siddall and Borda (2003)
Hemiclepsis marginata	France	AF003259	Siddall and Burreson (1998)
Theromyzon rude	Ontario, Canada	AF003262	Siddall and Burreson (1998)

[1]Collected from Parque Ecológico de Xochimilco, Mexico in 2001 by D. A. Weisblat.
[2]Formerly in the genus *Gloiobdella* (Siddall and Borda 2003).
[3]Submitted to GenBank under the name *H. papillornata*. See Siddall and Budinoff (2005) for details of taxonomic revision.
[4]Formerly in the genus *Desmobdella* (Siddall and Borda 2003).

Phylogenetic analysis

We aligned our *Helobdella* COI sequences with published sequences for all *Helobdella* species for which COI data were available in GenBank as well as several close out-groups (Tables 1 and 2), as determined by recent phylogenies of the family Glossiphoniidae (Light and Siddall 1999; Siddall and Borda 2003). Our final data matrix consisted of 630 bp, 40 ingroup (*Helobdella*) taxa, and five out-group taxa. Uncorrected *p* distances between sequences were calculated using PAUP* (version 4.0b10, Sinauer Associates, Sunderland, MA, USA). Maximum parsimony (MP) analyses were performed using PAUP*, using the heuristic search option, 10 random additions of the sequences, and TBR branch swapping, with the remaining parameters set to PAUP* defaults. To evaluate node support, we generated 100 bootstrap pseudoreplicates of the data and analyzed these as described above. Maximum likelihood (ML) analyses were performed using PhyML (version 2.4.4; Guindon and Gascuel 2003). Modeltest (version 3.7, Posada and Crandall 1998) was used to select the best-fit model (TVM+I+G) and relative rate and gamma shape parameter values, using AIC criteria. Relative rates and the gamma shape parameter (using four rate categories) were then optimized by PhyML and the proporition of invariant sites was estimated. A ML bootstrap analysis was performed with 100 pseudoreplicates.

RESULTS

New COI sequences have been deposited in GenBank under accession numbers DQ995297–DQ995311 (Tables 1 and 2). Where multiple specimens from a particular laboratory culture or field collection were sequenced, these were usually found to be identical and in such cases only one sequence was

Table 3. COI sequence divergences (uncorrected ["*p*"] distances) among *Helobdella* isolates used for development studies

Isolate name		CAGA-1	CASA-1	CASA-2	NYTA	CASA-4	CASF
CAGA-1	(*H.* sp. [Galt])	—					
CASA-1	(*H.* sp. "*robusta*" [Sac.])	0.164	—				
CASA-2	(*H.* sp. "*robusta*" [Sac.])	0.172	0.004	—			
NYTA	(*H.* sp. "*robusta*" [Valhalla])	0.169	0.014	0.016	—		
CASA-4	(*H.* sp. "*robusta*" [Sac.])	0.157	0.068	0.073	0.065	—	
CASF	(*H. triserialis*)	0.043	0.170	0.180	0.176	0.165	—
TXAU-1	(*H.* sp. "*robusta*" [Austin])	0.167	0.104	0.110	0.100	0.081	0.175

submitted to GenBank (Table 1) and included in our analyses. Sequence alignments were straightforward, as there was no length variation among the amplified COI sequences.

Sequence divergence

COI sequence divergences among *Helobdella* laboratory isolates range widely, from 0.4% to 18.0% (Table 3). As would be expected, specimens (either laboratory isolates or wild-collected individuals) originating from the same field site generally vary little, if at all, in sequence. Specimens originating from Galt, CA, are all identical and those from Austin, TX (either Shoal Creek or Brackenridge Field Laboratory), differ by at most 0.8%. Surprisingly, however, specimens originating from the Sacramento, CA locality have a sequence divergence of up to 7.3%. CASA-1, CASA-2, and CASA-5 are identical or nearly identical ($\leq$ 0.4% different) and CASA-3 and CASA-4 are very similar ($\leq$ 1.3%), but these two groups vary by 5.0–7.3%. Also noteworthy is the fact that the two specimens thought to represent *H. triserialis*, laboratory isolate CASF and *H. triserialis* from Bolivia, differ by a large amount, 10.6%. Yet another unexpected finding is that the nucleotide sequences of CAGA-1, CAGA-2, and NYTA are identical to *H. europaea* specimens from Australia (AU-1) and from South Africa and nearly identical (0.4–1.5%) to the other *H. europaea* specimens in the dataset, from New Zealand, Australia (AU-2), and Germany.

Phylogenetic analyses

The maximum parsimony and maximum likelihood analyses yielded largely congruent trees (Figs. 1 and 2) with no well-supported conflicts. The MP analysis recovered 12 best trees (tree length = 1264) that differed only trivially, with respect to the internal topology of clades comprised of nearly identical sequences.

In both the MP and ML analyses, *Helobdella* laboratory isolates are not united into a single clade but instead fall into five distinct lineages (Figs. 1 and 2, A–E). These are: (A) laboratory and wild-collected specimens originating from Galt, CA, and the specimen from an artificial pond on the roof of a UC Berkeley laboratory building (from which snails were being collected), along with specimens of *H. europaea* from several continents, (B) the laboratory isolate of *H. triserialis* from San Francisco, CA, which does not group with the other *H. triserialis* in the dataset (from Bolivia), (C) two of the three laboratory isolates originating from Sacramento, CA, specimens collected from the Sacramento field site in 2004 (but not those collected in 2001), and a laboratory isolate originating from Tarrytown (near Valhalla), NY, (D) the remaining laboratory isolate originating from Sacramento, CA, and specimens collected from the Sacramento field site in 2001, and (E) laboratory and wild-collected specimens originating from Austin, TX (from either Shoal Creek or Brackenridge Field Laboratory).

The five lineages listed above form parts of two major clades, recovered in both the MP and the ML analyses (Figs. 1 and 2). One clade includes lineages A and B along with *H.* sp. (from Mexico) and *H. triserialis* (from Bolivia). The internal branching structure of this clade is essentially identical between the ML and MP analyses and bootstrap support for the clade overall is fairly high (88%) in the ML analysis (although only 54% in the MP analysis). The second clade includes the remaining three lineages, C, D, and E, along with *H. lineata*, *H. papillata*, and *H. transversa*. This clade is very well supported by bootstrap analysis (100% support for both MP and ML), and the internal relationships recovered in both analyses are congruent.

These two major clades are closely allied to each other in both analyses, although this relationship receives low bootstrap support. *H. fusca* and *H. elongata* also appear to be closely related to these two clades, although their position is not well supported in either analysis and varies slightly between the MP and ML trees.

DISCUSSION

Important aspects of development can differ even between closely related taxa (Carroll et al. 2005). For this reason, if research on a group is conducted on several genetically

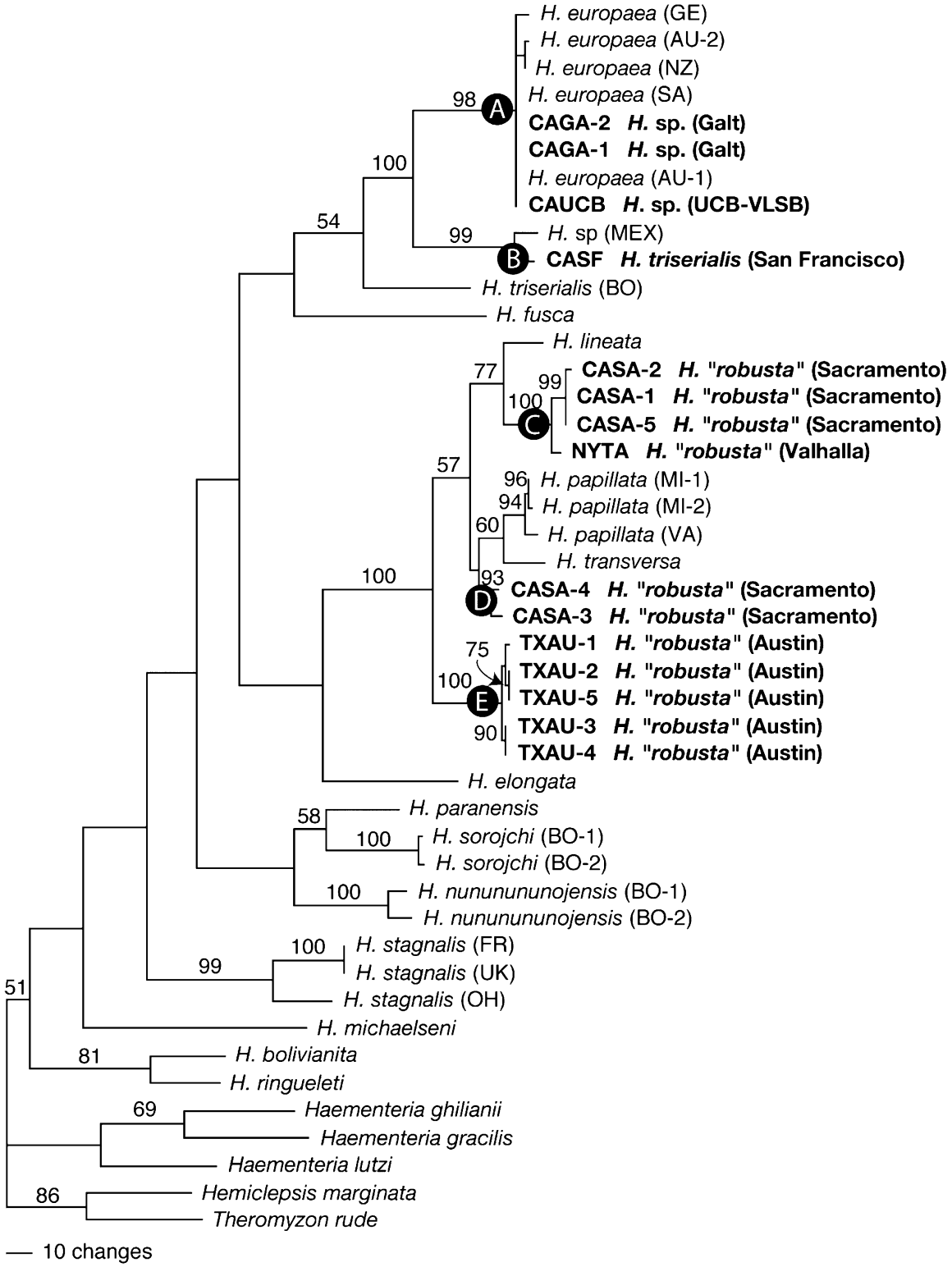

Fig. 1. Maximum parsimony tree of *Helobdella* based on COI sequences. This tree is one of 12 best trees and differs from the others in only trivial ways (see text). Bootstrap values above 50% are shown. Specimens from laboratory isolates and source populations (bold) fall into five distinct clades (A–E).

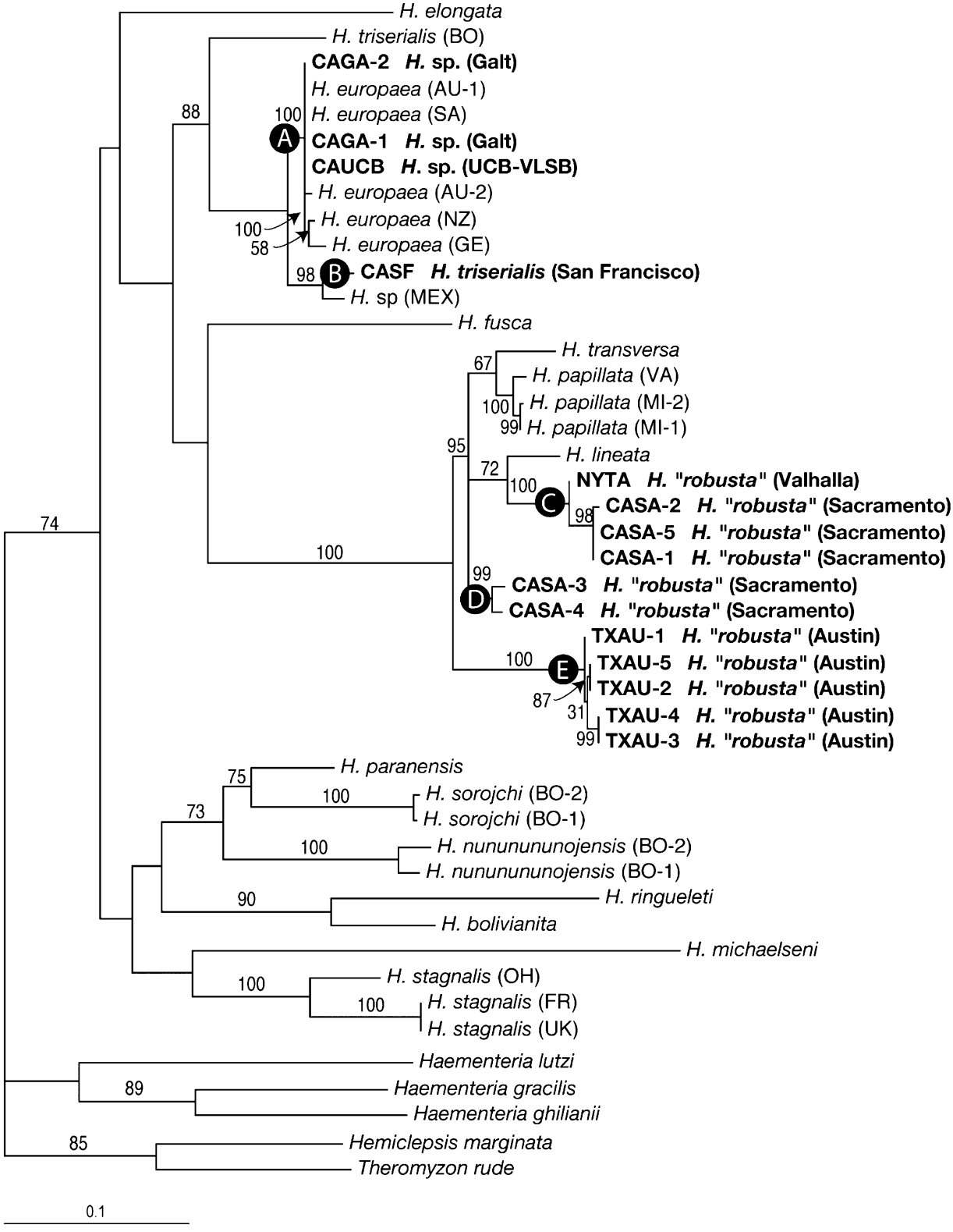

Fig. 2. Maximum likelihood tree of *Helobdella* based on COI sequences. Bootstrap values above 50% are shown. As in the maximum parsimony analysis, specimens from laboratory isolates and source populations (bold) fall into five distinct clades (A–E).

different isolates, one cannot assume that data collected on one isolate will necessarily apply to another. The research community working on *Helobdella* leech development has used multiple isolates stemming from repeated collections made at several localities across the United States. Differences between some of these isolates have been noted, and our results now clearly indicate that *Helobdella* laboratory isolates form a polyphyletic assemblage. Our analysis also provides a phylogenetic framework for interpreting developmental differences recognized within *Helobdella*.

Laboratory isolates of *Helobdella* probably represent five distinct species

All of the laboratory isolates and source populations that we sampled clearly fall within the genus *Helobdella*. However, they are not united into one clade but instead form five distinct lineages (Figs. 1 and 2, A–E), each of which is more closely related to other species or groups of species than they are to each other. These five lineages are also considerably diverged from each other at COI: inter-lineage divergences of laboratory isolates range from 4.3 to 18.0%. For comparison, closely related species of North American freshwater annelids (e.g., naidine oligochaetes, glossiphoniid leeches) can vary by as little as 4–5% at this gene (Siddall and Borda 2003; Bely and Wray 2004). Together, these data strongly suggest that *Helobdella* laboratory isolates represent five different species.

There is high congruence between our two analyses (maximum parsimony and maximum likelihood) in the placement of laboratory isolates and source populations, and nodes of interest are moderately to strongly supported by bootstrap analyses. Furthermore, relevant parts of our phylogenetic trees are highly concordant with those of another analysis of *Helobdella* relationships that was based on a combined morphological and molecular (COI and NDI genes) dataset (Siddall and Borda 2003), attesting further to the robustness of our findings.

We have refrained from making species-level assignments for most of the laboratory isolates for two reasons. First, based on COI gene sequences, the laboratory isolates we sequenced do not clearly match any previously sequenced species of *Helobdella*, with the exception of the CAGA specimens [*H.* sp. (Galt)], which unambiguously cluster with *H. europaea* (see below). Second, the morphological systematics of *Helobdella* is still being revised and we want to avoid adding confusion by prematurely assigning isolates to species. Indeed, even leading leech systematists have found initial identifications of *Helobdella* species based on morphology to be erroneous, needing to revise them once molecular data were obtained (Siddall and Borda 2003). For these reasons, we strongly believe that studies combining both molecular and morphological information will need to be performed in order to assign laboratory isolates to species.

The species *H. triserialis* may not be monophyletic

A population of *Helobdella* leeches from ponds in Golden Gate Park, San Francisco, CA, was the near-exclusive source of leeches used for *Helobdella* development research in North America from 1976 until 1989 and was sampled multiple times during this period. (The main exception is a study performed on *H. stagnalis* (Zackson 1984).) Much of the research on this San Francisco population used individuals derived from a breeding colony originally established in 1976, although it should be noted that colonies were supplemented several times with wild-collected individuals collected at later dates.

Our findings show that this San Francisco population, as represented in our study by *H. triserialis* isolate CASF, is quite genetically divergent (10.6% at COI) from an *H. triserialis* individual from Bolivia. Furthermore, these two *H. triserialis* individuals do not form a clade but instead form a paraphyletic assemblage (Figs. 1 and 2). Thus, the San Francisco population and the Bolivia population may well represent two distinct species.

H. triserialis isolate CASF and *H. triserialis* from Bolivia form part of a larger clade with *H. europaea* (including CAGA isolates of *H.* sp. (Galt)) and *H.* sp. from Mexico (the morphology of which closely resembles that of *H. triserialis* isolate CASF; D. A. Weisblat, personal observation). This clade represents the "*H. triserialis* species complex," a group of morphologically similar leeches, primarily from South America, whose taxonomy is still in flux (Siddall and Budinoff 2005). Much broader geographic sampling is needed to determine the species boundaries within this complex and to achieve a stable taxonomy of this group.

Publications on "*H. robusta*" have confused at least three different species, two of which coexist at the same locality

In 1989, a new species of *Helobdella*, which was named *H. robusta*, was collected from a minor, channelized tributary of the American River near the California Exposition and State Fair grounds in Sacramento, CA (Shankland et al. 1992). Since then, dozens of collections have been made from this site to provide material for development research. It has long been assumed that all *Helobdella* leeches that are collected from the type locality and that resemble *H. robusta* are indeed this species. However, we found that specimens originating from this site form two clades that differ considerably in COI sequence (5.0–7.3%). Each of these clades includes specimens collected directly from the field site and specimens obtained from laboratory cultures. Importantly, one of the two Sacramento clades is more closely related to a *Helobdella* isolate from the east coast (NYTA) than it is to the other Sacramento clade, consistent with the idea that the two Sacramento

clades are reproductively isolated. Under some circumstances, a species may harbor two or more highly divergent mitochondrial haplotypes, for example as a result of incomplete lineage sorting (i.e., the retention of ancestral polymorphisms) or inter-species hybridization. Barring such scenarios, however, the data at hand strongly indicate that the Sacramento field site, the type locality of *H. robusta*, harbors two closely related species of *Helobdella* that coexist simultaneously or sequentially through time.

Leech populations from two sources other than the Sacramento locality have also been referred to as *H. robusta* in publications: those from Tarrytown, NY, and those from two sites (Shoal Creek and Brackenridge Field Laboratory) in Austin, TX. The COI sequence divergences between the laboratory isolate from Tarrytown, NY (NYTA), and specimens in the Sacramento, CA clade that clusters with it are quite low (just 1.4–1.6%), despite the fact that the sources of these specimens are on opposite coasts of North America. Thus, the Tarrytown, NY population and one of the two Sacramento, CA populations probably represent the same species (although resolving whether or not these are *H. robusta* will need to await further work). The *Helobdella* populations from Austin, TX do not group with either of the two species from the Sacramento locality and differ from them by 7.7–11.1% at COI. *Helobdella* sp. (Austin) is therefore distinct from both Sacramento species and clearly not *H. robusta*.

A large proportion of *Helobdella* development work has been performed on "*H. robusta*" specimens from the Sacramento, CA site. It is now clear from our study, however, that development publications referencing "*H. robusta*" from Sacramento as their subjects could have used either of two species. Unfortunately, in many if not most cases it will be impossible to determine retrospectively which of these two species was used. Previous studies must therefore be interpreted with this uncertainty in mind. To determine which of the two mitochondrial haplotypes corresponds to *H. robusta*, mitochondrial sequence will have to be obtained from a type specimen. Until this is done, we recommend using the name "*H.* sp. 1" to refer to the species represented by clade C, which includes the specimens used for an early cDNA library (CA-SA-1) and for which a complete mitochondrial genome was obtained (CASA-2) (Boore and Brown 2000), and "*H.* sp. 2" to refer to the other species, represented by clade D, which includes the inbred line (CASA-4) for which a complete genomic sequence is being obtained. It will be critical to sequence the mitochondrial haplotype of any new isolates, at least until reliable morphological markers are found that distinguish these two species.

H. sp. (Galt) is an invasive species, *H. europaea*

Helobdella specimens collected from Galt, CA have a COI sequence identical or nearly identical to isolates of *H. europaea* and are clearly representatives of this species. This species has recently been recognized as invasive, based on the finding that individuals collected from distant geographic locations exhibit an identical or nearly identical COI and ND-I sequence (Pfeiffer et al. 2004; Siddall and Budinoff 2005). The population identified near Galt, CA is located on a sturgeon farm, consistent with the hypothesis that this invasive species is being spread by human activities. Populations of *H. europaea* have previously been recorded from Australia, New Zealand, South Africa, Germany, and Hawaii (Pfeiffer et al. 2004; Siddall and Budinoff 2005). The population that we have identified from California represents the first record of this species from the North American continent and demonstrates that it has spread even more widely than previously recognized. Consistent with its propensity to be dispersed by humans, *H. europaea* was discovered in an artificial pond used for holding various aquatic animals of diverse geographic origins on the roof of a UC Berkeley laboratory building.

Evolution of development among *Helobdella* species

Notwithstanding this confusion surrounding the taxonomy of laboratory isolates, the widespread sampling and parallel investigations undertaken by the *Helobdella* development community have made *Helobdella* an excellent group in which to study the microevolution of development, especially the evolution of cell lineages and cell fate specification. *Helobdella* is already a proven and powerful system for investigating cellular and molecular aspects of embryogenesis, owing in large part to *Helobdella* embryos being amenable to intracellular injection of lineage tracers and other reagents (Weisblat and Huang 2001). It is now clear that at least five closely related species can be easily obtained, reared, and manipulated in the lab; their phylogenetic relationships are now known; and some interesting developmental differences between species have already been noted.

Variation in micromere cell lineage contributions has been documented between what we now recognize are different species of *Helobdella* (Huang et al. 2002). Specifically, Huang et al. (2002) reported two differences in micromere cell lineage contributions between *H. europaea* (Galt) and an isolate of *H.* sp. "*robusta*" (Sacramento). (Unfortunately, it cannot be determined which of the two Sacramento species was the subject of this work.) First, in *H. europaea* micromeres b″ and c″ each contribute to specific regions of the proboscis (with b″ contributing primarily to ventral regions and c″ contributing primarily to dorsal regions), whereas in *H.* sp. "*robusta*," the b″ clone normally dies and c″ alone contributes to the entire, combined region. In this latter species, b″ and c″ form an equivalence group, such that if c″ is killed, b″ fails to die and instead replaces c″ in contributing to those regions of the proboscis. Second, in *H.* sp. "*robusta*," micromere c‴ gives

rise to prominent circular muscle fibers of the proboscis, whereas in the *H. europaea* isolate, c‴ normally dies and instead these fibers arise entirely from a different micromere, dm′.

Variation between *Helobdella* species has also been documented with respect to the signals required for cell fate specification within one of the ectodermal lineages, O/P. The O fate is induced by the Q lineage in *H. europaea* (Galt) and in a Sacramento isolate of *H.* sp. "*robusta*," whereas it is induced redundantly by the Q lineage and the M lineage in *H.* sp. (Austin) (Huang and Weisblat 1996; Kuo and Shankland 2004). (Unfortunately, as with the Huang et al. (2002) study, it is not known which of the two Sacramento species was used for Huang and Weisblat (1996).) Our phylogenetic results suggest that redundant induction by the Q and M lineages is a derived condition within *Helobdella* and that induction solely by the Q lineage is ancestral. Another cell specification difference noted between species is that the provisional integument is required to specify the O fate in *H. triserialis* (San Francisco, Ho and Weisblat 1987) but apparently not in *H.* sp. (Austin) (Kuo and Shankland 2004).

Clearly, cell lineage and cell fate specification have evolved even among closely related and morphologically very similar *Helobdella* species. Indeed, the differences summarized above may reflect developmental system drift, by which underlying developmental processes may evolve without concomitant changes in resulting phenotypes (True and Haag 2001). Comparative developmental studies of *Helobdella* species provide an exquisite opportunity for investigating the evolutionary steps and developmental mechanisms by which such features evolve.

Importance of DNA barcoding in the lab

DNA barcoding is a molecular approach to taxonomic identification in which DNA sequences serve as taxon identifiers. It has been proposed as a needed, viable solution to the otherwise overwhelming problem of identifying the earth's biota (Hebert et al. 2003; Tautz et al. 2003), especially in the face of the regrettable recent erosion of specialized taxonomic expertise. Although it cannot replace the information held in the morphology and other characters of whole specimens, DNA barcoding offers some significant advantages over traditional morphology-based taxonomic identification using keys. It can distinguish morphologically cryptic taxa, is insensitive to phenotypically plastic characters, can be based on material from any stage of the life cycle, and does not require expert knowledge or subjective judgments regarding the morphology of a group. The mitochondrial gene COI has been proposed as a good choice for DNA barcoding (Hebert et al. 2003), although theoretically any sequence with appropriate variability could be used.

Although DNA barcoding has been championed primarily as a method to inventory biodiversity, our study of *Helobdella* highlights a specific need for DNA barcoding of organisms being used for research. By obtaining DNA barcodes (COI gene sequence) for laboratory isolates of *Helobdella*, we determined that isolates that have been referred to in the literature as one species ("*H. robusta*") appear to represent three different species, two of which coexist at the same field site. We were fortunate to be able to obtain and sequence specimens from most relevant cultures in order to sort out retroactively the identity of organisms used in some prior research. Yet, some important information is not recoverable. The name "*H. robusta* (Sacramento)" has been used to refer to two similar yet distinct species in a significant body of literature, representing a large fraction of development work on *Helobdella*. It will not be possible to determine which of these two species was the subject of many, if not most, prior studies. If a DNA barcode had been obtained immediately from every new *Helobdella* isolate, the genetic distinctness of different isolates would have been recognized from the beginning.

The perils of misidentification and taxonomic confusion in the lab are numerous and costly, and the risk of such problems is high when organisms are routinely recollected from the field, as is common in many evolution of development labs. If, unknowingly, genetically different isolates are used for research, previously obtained results may not be repeatable or valid and reagents such as nucleotide probes and antibodies may not work consistently (if they do not cross-react between isolates). Understandably, evolution of development researchers are typically not experts on the taxonomy of their study organisms, hindering their ability to make accurate morphology-based species identifications. Even if these can be done, there is always the risk that isolates whose adult morphology is indistinguishable still differ genetically and in key aspects of development. We caution researchers similarly working on multiple wild-collected isolates to collect voucher specimens and to obtain a molecular "barcode," such as a COI gene sequence, for each new isolate collected in order to maintain permanent records of the genetic stock of animals used.

Acknowledgments

We thank S. Bissen, D.-H. Kuo, M. Shankland, C. Wedeen, and members of the Weisblat lab for generously providing specimens and for helpful discussions, M. Siddall for help with the PhyML analysis, J. Sikes and R. Baker for technical assistance, and L. Shapiro, M. Siddall, and an anonymous reviewer for useful comments on the manuscript. F. Huang noticed some of the first developmental differences between isolates. This work was supported by funds from the University of Maryland to A. E. B. and NASA grant NAG-1359 to D. A. W.

REFERENCES

Abouheif, E., and Wray, G. A. 2002. Evolution of the gene network underlying wing polyphenism in ants. *Science* 297: 249–252.

Apakupakul, K., Siddall, M. E., and Burreson, E. M. 1999. Higher level relationships of leeches (Annelida: Clitellata: Euhirudinea) based on morphology and gene sequences. *Mol. Phylogenet. Evol.* 12: 350–359.

Bely, A. E., and Wray, G. A. 2001. Evolution of regeneration and fission in annelids: insights from *engrailed*- and *orthodenticle*-class gene expression. *Development* 128: 2781–2791.

Bely, A. E., and Wray, G. A. 2004. Molecular phylogeny of naidid worms (Annelida: Clitellata) based on cytochrome oxidase I. *Mol. Phylogenet. Evol.* 30: 50–63.

Boore, J. L., and Brown, W. M. 2000. Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Mol. Biol. Evol.* 17: 87–106.

Carroll, S. B., Grenier, J. K., and Weatherbee, S. D. 2005. *From DNA to Diversity*. Blackwell Publishing, Oxford.

Davis, G. K., and Patel, N. H. 2002. Short, long, and beyond: molecular and embryological approaches to insect segmentation. *Ann. Rev. Entomol.* 47: 669–699.

Emlen, D. J., and Nijhout, H. F. 1999. Hormonal control of male horn length dimorphism in the dung beetle *Onthophagus taurus* (Coleoptera: Scarabaeidae). *J. Insect. Physiol.* 45: 45–53.

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Mol. Marine Biol. Biotech.* 3: 294–299.

Gerhart, J., Lowe, C., and Kirschner, M. 2005. Hemichordates and the origin of chordates. *Curr. Opin. Genet. Dev.* 15: 461–467.

Grenier, J., and Carroll, S. 2000. Functional evolution of the Ultrabithorax protein. *Proc. Natl. Acad. Sci. USA* 97: 704–709.

Guindon, S., and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Sys. Biol.* 52: 696–704.

Hebert, P. D. N., Cywinski, A., Ball, S. L., and deWaard, J. R. 2003. Biological identifications through DNA barcodes. *Proc. Royal. Soc. London B* 270: 313–321.

Ho, R. K., and Weisblat, D. A. 1987. A provisional epithelium in leech embryo: Cellular origins and influence on a developmental equivalence group. *Dev. Biol.* 120: 520–534.

Huang, F. Z., Kang, D. M., Ramirez-Weber, F. A., Bissen, S. T., and Weisblat, D. A. 2002. Micromere lineages in the glossiphoniid leech *Helobdella*. *Development* 129: 719–732.

Huang, F. Z., and Weisblat, D. A. 1996. Cell fate determination in an annelid equivalence group. *Development* 122: 1839–1847.

Hughes, C., and Kaufman, T. 2002. Hox genes and the evolution of the arthropod body plan. *Evol. Dev.* 4: 459–499.

Jeffery, W. 2001. Cavefish as a model system in evolutionary developmental biology. *Dev. Biol.* 231: 1–12.

Kuo, D.-H., and Shankland, M. 2004. Evolutionary diversification of specification mechanisms within the O/P equivalence group of the leech genus *Helobdella*. *Development* 131: 5859–5869.

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

Martindale, M. Q. 2005. The evolution of metazoan axial properties. *Nat. Rev. Genet.* 6: 917–927.

Metscher, B., Takahashi, K., and Crow, K. 2005. Expression of Hoxa-11 and Hoxa-13 in the pectoral fin of a basal ray-finned fish, *Polyodon spathula*: implications for the origin of tetrapod limbs. *Evol. Dev.* 7: 186–195.

Oceguera-Figueroa, A., León-Règagnon, V., and Siddall, M. E. 2005. Phylogeny and revision of Erpobdelliformes (Annelida, Arhynchobdellida) from Mexico based on nuclear and mitochondrial gene sequences. *Revista Mexicana Biodiversidad* 76: 191–198.

Peel, A., Chipman, A., and Akam, M. 2005. Arthropod segmentation: beyond the *Drosophila* paradigm. *Nat. Rev. Genet.* 6: 905–916.

Pfeiffer, I., Brenig, B., and Kutschera, U. 2004. The occurrence of an Australian leech species (genus *Helobdella*) in German freshwater habitats as revealed by mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 33: 214–219.

Posada, D., and Crandall, K. A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.

Ross, K. G., and Shoemaker, D. D. 2005. Species delimitation in native South American fire ants. *Mol. Ecol.* 14: 3419–3438.

Sánchez Alvarado, A. 2004. Regeneration and the need for simpler model organisms. *Phil. Trans. R. Soc. Lond. B* 359: 759–763.

Saló, E., and Baguñà, J. 2002. Regeneration in planarians and other worms: new findings, new tools, and new perspectives. *J. Exp. Zool.* 292: 528–539.

Sawyer, R. T. 1986. *Leech Biology and Behavior (Volume II): Feeding Biology, Ecology and Systematics*. Clarendon Press, Oxford.

Seaver, E. C. 2003. Segmentation: mono or polyphyletic? *Int. J. Dev. Biol.* 47: 583–595.

Seaver, E., and Kaneshige, L. 2006. Expression of 'segmentation' genes during larval and juvenile development in the polychaetes *Capitella* sp. I and *H. elegans*. *Dev. Biol.* 289: 179–194.

Shankland, M., Bissen, S. T., and Weisblat, D. A. 1992. Description of the Californian leech *Helobdella robusta* sp. nov., and comparison with *Helobdella triserialis* on the basis of morphology, embryology, and experimental breeding. *Can. J. Zool.* 70: 1258–1263.

Shankland, M., and Seaver, E. C. 2000. Evolution of the bilaterian body plan: what have we learned from annelids? *Proc. Natl. Acad. Sci. USA* 97: 4434–4437.

Siddall, M. E., and Borda, E. 2003. Phylogeny and revision of the leech genus *Helobdella* (Glossiphoniidae) based on mitochondrial gene sequences and morphological data and a special consideration of the *triserialis* complex. *Zool. Scripta*. 32: 23–33.

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

Siddall, M. E., and Burreson, E. M. 1998. Phylogeny of leeches (Hirudinea) based on mitochondrial cytochrome *c* oxidase subunit I. *Mol. Phylogenet. Evol.* 9: 156–162.

Simpson, P. 2002. Evolution of development in closely related species of flies and worms. *Nat. Rev. Gene.* 3: 907–917.

Smith, M. A., Woodley, N. E., Janzen, D. H., Hallwachs, W., and Hebert, P. D. N. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proc. Natl. Acad. Sci. USA* 103: 3657–3662.

Tautz, D., Arctander, P., Minelli, A., Thomas, R. H., and Vogler, A. P. 2003. A plea for DNA taxonomy. *Trend. Ecol. Evol.* 18: 70–74.

Tessmar-Raible, K., and Arendt, D. 2003. Emerging systems: between vertebrates and arthropods, the Lophotrochozoa. *Curr. Opin. Gene. Dev.* 13: 331–340.

Trontelj, P., and Utevsky, S. Y. 2005. Celebrity with a neglected taxonomy: molecular systematics of the medicinal leech (genus *Hirudo*). *Mol. Phylogenet. Evol.* 34: 616–624.

True, J. R., and Haag, E. S. 2001. Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* 3: 109–119.

Weigt, L. A., Crawford, A. J., Rand, A. S., and Ryan, M. J. 2005. Biogeography of the túngara frog, *Physalaemus pustulosus*: a molecular perspective. *Mol. Ecol.* 14: 3857–3876.

Weisblat, D. A., and Huang, F. Z. 2001. An overview of glossiphoniid leech development. *Can. J. Zool.* 79: 218–232.

Zackson, S. L. 1984. Cell lineage, cell-cell interaction and segment formation in the ectoderm of a glossiphoniid leech embryo. *Dev. Biol.* 104: 43–60.

Zhurov, V., Terzin, T., and Grbic', M. 2004. Early blastomere determines embryo proliferation and caste fate in a polyembryonic wasp. *Nature* 432: 764–769.