

2012

# WHEN MOLECULES AND MORPHOLOGY CLASH: REVISITING SPECIES TREE RECONSTRUCTION OF AMBYSTOMATID SALAMANDERS USING MULTIPLE NUCLEAR LOCI

Joshua Steven Williams
*University of Kentucky*, jswi223@g.uky.edu

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained and attached hereto needed written permission statements(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine).

I hereby grant to The University of Kentucky and its agents the non-exclusive license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless a preapproved embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's dissertation including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Joshua Steven Williams, Student

Dr. David. W. Weisrock, Major Professor

Dr. David W. Weisrock, Director of Graduate Studies

# WHEN MOLECULES AND MORPHOLOGY CLASH: REVISITING SPECIES TREE RECONSTRUCTION OF AMBYSTOMATID SALAMANDERS USING MULTIPLE NUCLEAR LOCI

---

THESIS

---

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the College of Arts and Sciences at the University of Kentucky


By

Joshua Steven Williams

Lexington, Kentucky

Director: Dr. David. W. Weisrock, Professor of Biology

Lexington, Kentucky


2012

ABSTRACT OF THESIS

WHEN MOLECULES AND MORPHOLOGY CLASH: REVISITING SPECIES TREE RECONSTRUCTION OF AMBYSTOMATID SALAMANDERS USING MULTIPLE NUCLEAR LOCI

The analysis of diverse data sets can yield different phylogenetic estimates that challenge systematists to explain the source of discordance. The Ambystomatidae are a classic example of this phylogenetic conflict. Previous attempts to resolve the ambystomatid species tree using allozymic, morphological, and mitochondrial sequence data have yielded different estimates, making it unclear which data source best approximates ambystomatid phylogeny. We present the first multi-locus DNA sequence-based phylogenetic study of the Ambystomatidae. Because independent loci can contain discordant gene tree histories, concatenating unlinked loci into a single data matrix can lead to strongly supported and erroneous results. Therefore, we utilized a range of analyses, including coalescent-based methods of phylogenetic estimation that account for incomplete lineage sorting and concordance-based methods that estimate the proportion of sampled loci that support a particular clade. We repeated these analyses with the removal of individual loci to determine if any locus has a disproportionate effect on our phylogenetic results. Many deep and relatively shallow clades within *Ambystoma* were robustly resolved. Analyses that excluded loci produced overlapping posterior distributions, suggesting no disproportionate influence of any particular locus. Our estimates differ from previous hypotheses, although there was greater similarity with previous molecular estimates, relative to morphological estimates.

KEYWORDS: Coalescent analysis, concordance analysis, species tree, gene tree, *Ambystoma*

_____Joshua S. Williams_____

_____6/22/12_____

WHEN MOLECULES AND MORPHOLOGY CLASH: REVISITING SPECIES TREE RECONSTRUCTION OF AMBYSTOMATID SALAMANDERS USING MULTIPLE NUCLEAR LOCI

By

Joshua Steven Williams

_____Dr. Brian C. Rymond_____
Director of Thesis

_____Dr. David W. Weisrock_____
Director of Graduate Studies

_____6/22/12_____
Date

ACKNOWLEDGEMENTS

We would like to thank Jim Demastes, Paul Moler, Greg Pauly, Brad Shaffer, the Louisiana State University Museum of Zoology, and the Museum of Vertebrate Zoology at the University of California, Berkley for their contribution of tissue samples. We also thank Eric O'Neill, Stephanie Mitchell, Alex Noble, and Ana Mendia for invaluable laboratory assistance. Randal Voss and Paul Hime provided comments that improved this manuscript. We would also like to thank the University of Kentucky Information Technology department and Center for Computational Sciences for computing time on the Lipscomb High Performance Computing Cluster and for access to other supercomputing resources. Funding and support for this work was provided from the University of Kentucky, a Commonwealth of Kentucky NSF EPSCoR grant (# 0814194) in support of Ecological Genomics training and research, NSF grant DEB0949532 (to DWW), and a Society of Systematic Biologists Graduate Student Award (to JSW).

TABLE OF CONTENTS

Acknowledgements ..... iii

List of Tables ..... v

List of Figures ..... vi

**Chapter 1, Revisiting species tree reconstruction of Ambystomatidae using multiple nuclear loci** ..... 1

Introduction ..... 1

Materials and Methods ..... 4

*Taxonomic and genomic sampling* ..... 4

*Phylogenetic reconstruction of individual gene trees* ..... 7

*Coalescent-based species tree estimation* ..... 8

*Bayesian concordance of gene trees* ..... 11

Results ..... 12

*Individual gene tree reconstruction* ..... 13

**BEAST species tree analyses* ..... 13

*STEM analyses* ..... 15

*Bayesian concordance analyses* ..... 16

Discussion ..... 18

*Interpreting phylogenetic resolution within the Ambystomatidae* ..... 18

*Phylogenetic ambiguity in the Ambystomatidae* ..... 21

*Effects of individual loci on phylogenetic reconstruction* ..... 23

*Comparison to previous hypotheses* ..... 25

**References** ..... 29

**Vita** ..... 55

LIST OF TABLES

Table 1. *Ambystoma* and *Dicamptodon* individuals used in this study.....35
Table 2. Forward and reverse primers for all amplified loci .....37
Table 3. Information for the loci sequenced in this study.....38
Table 4. Averaged Robinsons-Foulds distances for *BEAST posterior distributions.....39

LIST OF FIGURES

Figure 1. Previous hypotheses for ambystomatid phylogeny .....40
Figure 2. 50% majority-rule consensus trees generated by MrBayes for each locus .....42
Figure 3. Maximum clade credibility trees generated from Bayesian coalescent species trees reconstructions .....47
Figure 4. Ordination plots based on the posterior distributions (PDs) generated from Bayesian species tree analyses using all loci and excluding a single locus .....49
Figure 5. Species tree generated via a maximum likelihood analysis .....51
Figure 6. Primary concordance trees generated with Bayesian concordance analysis .....52
Figure 7. Ordination plots based on the posterior distributions (PDs) generated from Bayesian species tree analyses using all nuclear loci and excluding both mtDNA and a single nuclear locus .....54

# Chapter 1

## Revisiting species tree reconstruction of Ambystomatidae using multiple nuclear loci

### Introduction

Systematists have often been challenged to explain phylogenetic conflict arising from the analysis of diverse data sets (e.g., morphological and molecular data) (Shaffer et al., 1991; Wiens and Hollingsworth, 2000). Individual data sets can be phylogenetically misleading if convergent evolution has produced homoplastic characters, a problem that can be inherent in both morphological and molecular characters (Hillis, 1987). Furthermore, properties of the underlying phylogeny itself can facilitate inaccurate estimation when branches are long and the data are highly variable [e.g., long branch attraction (Felsenstein, 1978)]. While further exploration of individual data sets can sometimes identify the source of discordance (e.g., Wiens and Hollingsworth, 2000), in other studies individual data sets can each yield convincingly strong support so as to preclude resolution of the conflict. In these situations, collection of additional data from an independently evolving source will be necessary to elucidate phylogenetic history and shed light on the source of the initial phylogenetic conflict.

Phylogenetic reconstruction of multiple independent loci can also yield discordance among gene trees, a product of a number of processes, including incomplete lineage sorting (deep coalescence) and lateral gene transfer (Maddison, 1997). Analysis of these data as a concatenated supermatrix may be prone to yield inaccurate species trees due to the mixture of phylogenetic signal from different gene histories (Kubatko and Degnan, 2007; Weisrock et al., 2012). As an alternative, methods that estimate a species

tree from independently estimated gene trees, including those that account for the stochastic nature of genetic drift in the lineage sorting process, can be used to reconstruct the species phylogeny, despite strongly supported discordance among gene trees.

The phylogeny of salamander species of the family Ambystomatidae represents a classic example of phylogenetic conflict arising from the analysis of very different data sources (Shaffer et al., 1991). The Ambystomatidae are a broadly distributed group of species covering much of the United States and Mexico, and feature a diverse array of life history phenotypes. This includes a radiation of U.S. and Mexican species (the tiger salamanders) that vary in their propensity to metamorphose, and a group of unisexual populations in eastern North America that are putatively of hybrid origin. Phylogenetic analyses of allozymic and morphological data sets collected from sexual *Ambystoma* species (and representative tiger salamander species) yield a number of discordant topological patterns. The strongest of these involve a morphologically supported clade comprised of *Ambystoma annulatum, A. barbouri, A. cingulatum, A. mabeei,* and *A. texanum* (Fig. 1A) (Kraus, 1988). This clade, named the subgenus Linguaelapsus, is supported by four synapomorphic characters associated with hyoid musculature. An additional nine morphological synapomorphies support the clade of *A. annulatum, A. barbouri, A. cingulatum,* and *A. texanum*. In contrast, parsimony and maximum likelihood analysis of allozymic data strongly support very different placements of these taxa within the *Ambystoma* tree (Fig. 1B). A number of other relationships differ between the two data sources, including the morphological placement of *A. gracile* as the sister lineage to all other *Ambystoma* species. However, the Linguaelapsus clade has been highlighted as a particularly striking example of discordance between morphological and

molecular data sets (Shaffer et al., 1991). Combined analysis of the data only weakly supports the Linguaelapsus clade (Jones et al., 1993; Shaffer et al., 1991), but individual data sets each strongly support different placements for its component taxa. Whether or not this discordance results from homoplasy in one, or both, data sets, and extreme non-independence among convergently-evolved characters is not clear.

Phylogenetic reconstruction using independent sources of data represent one step towards elucidating the potential factors contributing to this strong discordance. There have been previous attempts to reconstruct ambystomatid phylogeny using mtDNA sequence data. Bogart (2003) presented a phylogenetic tree based on *cytb* and 16S sequence data that included all species of *Ambystoma*, except for *A. annulatum*. This tree placed *A. texanum* and *A. barbouri* together in a clade, with *A. mabeei* maintaining a close relationship with these two species; however, the Linguaelapsus clade was not recovered. In a subsequent mtDNA study of the origin of unisexual ambystomatids using *cytb* sequence data, Robertson et al. (2006) resolved a clade that contained Linguaelapsus species (*A. annulatum* was not sampled) along with representative tiger salamander taxa. However, this clade – and the majority of interspecific relationships – received very low parsimony and Bayesian branch support. This study may also have been misled by PCR amplification of a nuclear paralog of *cytb* (Bi and Bogart, 2010).

In this study, we have presented the first multi-locus nuclear DNA sequence data set ever collected to resolve phylogenetic relationships among sexual species of *Ambystoma* and shed light on the incongruence of previous morphological and molecular trees. We included representatives of nearly all sexual species and only limited our sampling to representatives within the diverse tiger salamander lineage. Sequence data

were generated from 14 nuclear loci, the majority of which were located on separate linkage groups, and a mitochondrial locus. We explored the potential for gene tree discordance among loci using a Bayesian estimate of clade concordance (Ane et al., 2007). When gene trees exhibit high levels of discordance, concatenated phylogenetic approaches to species tree reconstruction may be prone to produce inaccurate results (Kubatko and Degnan, 2007; Liu and Edwards, 2009). Therefore, we focused our species tree reconstruction on methods that use a coalescent model to account for gene tree discordance through mechanisms of incomplete lineage sorting. Simulation studies indicate that these methods outperform concatenation in species tree reconstruction when phylogenetic history features short branches and/or large effective population sizes (Leache and Rannala, 2011); conditions that should increase the probability of an incomplete lineage sorting event and discordance among gene trees. By utilizing best practices of phylogenetic reconstruction, we aimed to robustly resolve relationships within the Ambystomatidae and to compare these results with previous hypotheses to shed light on phylogenetic ambiguity within this group.

## Materials and Methods

### *Taxonomic and genetic sampling*

A total of 33 *Ambystoma* individuals were sampled from 18 extant *Ambystoma* species, with 1-4 representative individuals per species (Table 1). This sampling included five representative lineages of the taxonomically diverse *A. tigrinum* species complex clade based on divergent lineages in the mtDNA gene tree (Shaffer and McKnight, 1996), and all diploid sexual species outside of this clade. The Ambystomatidae also contains a

complex of unisexual populations with a complicated evolutionary history (Bi and Bogart, 2010) and representatives from this group were not included in this study. To root the *Ambystoma* tree, 1-2 two samples were included from all four extant *Dicamptodon* species (seven total individuals). *Dicamptodon* is the most appropriate outgroup for phylogenetic reconstruction within *Ambystoma* as numerous molecular studies consistently establish them as sister lineages (Frost et al., 2006; Larson, 1991; Roelants et al., 2007; Weisrock et al., 2005; Wiens et al., 2005). DNA was extracted from tissues using a Qiagen DNeasy Blood and Tissue kit following the standard protocol for DNA extraction. Genomic DNA quantity and quality were assessed using a NanoDrop 2000 Spectrophotometer (Thermoscientific) and through electrophoresis on a 1.3% agarose gel.

Mitochondrial DNA sequence data were collected from a region encompassing the *nad2* gene region and the $tRNA^{Trp}$ and $tRNA^{Ala}$ genes using primers previously published in (Weisrock et al., 2001). Nuclear sequence data were collected from 14 loci identified from EST-based genome resources developed for the Mexican Axolotl, *A. mexicanum*, and eastern tiger salamander, *A. tigrinum* (Putta et al., 2004). A list of primers for each locus can be found in Table 2. All PCR reactions were performed in a total volume of 20μL, and were comprised of 14.1μL of water, 2μL of *Taq* buffer (with $MgCl_2$), 0.4μL of dNTPs, 0.7μL of each primer, 0.1μL of *Taq* DNA polymerase, and 2μL of template DNA. In each reaction we aimed to use approximately 50ng of genomic DNA. Most loci were PCR amplified with 35 cycles of denaturing at 95°C for 45 sec., annealing at 55°C for 45 sec., and extension at 72°C for 30 sec. All PCR runs opened with 95°C for 3 min and concluded with a 5 min extension stage at 72°C. For loci and

individuals that were troublesome to amplify, a gradient PCR was used with the same PCR protocol as outlined above, except that the annealing phase consisted of a 45-65°C gradient across the 12 columns of the thermalcycler block. To confirm that there had not been contamination, negative controls were run for each set of PCRs using 2μL of water instead of DNA. All reactions were run on a 1.3% agarose gel, using 0.8 μL of EZ-Vision One 6X loading dye with 4 μL of PCR product for each well.

To phase alleles from heterozygous individuals with indels or multiple polymorphic sites, PCR products were cloned using an Invitrogen TOPO-TA Cloning Kit. Culture plates were made with LB agar, 40mg of X-gal per mL of dimethylformamide, and mixed with 50 μg/mL of kanamycin. Colonies were grown overnight at 37°C, subsequently picked from plates, and lysed in 25μL TE Buffer for 5 minutes at 95°C. 2μL of lysed cells were used in the PCR protocol outlined above using standard M13 forward and reverse primers. Four separate colonies were sequenced for each cloning reaction. If only one allele was recovered, then four more clones were sequenced. Cloning sometimes resulted in sequences exhibiting patterns consistent with PCR recombination in the cloned PCR products (i.e. recovering three to four alleles in the clone products). For cloning products exhibiting PCR recombinant patterns, we performed a subsequent round of PCR on the genomic DNA aimed at minimizing the potential for PCR recombination by reducing the number of amplification cycles from 35 to 30 (Cronn et al., 2002; Zylstra et al., 1998). Overall, this had the intended effect of reducing the recovery of extra alleles from heterozygous individuals.

All PCR reactions were cleaned up with a 1:5 dilution of ExoSAP-IT following the standard manufacturer's protocol. Sequencing reactions were performed using

BigDye Terminator v3.1 and the individual PCR primers originally used for PCR. Samples were sequenced in both the forward and reverse directions on an ABI 3730 sequencer located in the University of Kentucky's Advanced Genetic Technologies Center. Sequences were analyzed, edited, and aligned using Geneious Pro version 5.3.3 (Drummond, 2010). All data alignments included two haploid gene copies from each individual (i.e., alleles; these were randomly labeled A and B) to accommodate heterozygous individuals.

*Phylogenetic reconstruction of individual gene trees*

To estimate the model of nucleotide substitution for each gene we analyzed individual gene alignments (including both of the intraindividual A and B alleles) in JModelTest 0.1.1 (Guindon and Gascuel, 2003; Posada, 2008). We performed these analyses on alignments that included all *Ambystoma* and *Dicamptodon* sequence data. In addition, for the purpose of using these gene trees in some downstream analyses, we also estimated evolutionary models for alignments that were exclusive to *Ambystoma* sequence data. For all data sets, the Akaike Information Criterion (AIC) was used to determine the best-fit substitution model. Bayesian posterior distributions of gene trees for each locus were estimated using MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003). For each locus, analyses were performed with four runs containing four MCMC chains each. Each analysis was run for 25 million generations with trees and parameters sampled every 5,000 generations. We performed four replicate analyses for each locus using different starting conditions determined by random number seeds. Tree scores (lnL values) and ESS estimates from the four independent MCMC runs were analyzed with

Tracer v1.5 (Rambaut and Drummond, 2007) to detect whether or not the posterior distribution of all runs for a locus had converged and had been run long enough to provide independent samples of the posterior distribution, where an ESS of 200 or greater for combined replicate runs was considered representative of adequate posterior sampling. In all analyses, replicate runs reached the same stable posterior distribution before 2.5 million generations. Using MrBayes, we generated a 50% majority-rule consensus tree based on the four replicate runs (using a 2.5 million generation burnin for each replicate).

*Coalescent-based species tree estimation*

We used two different analytical approaches to estimate a species tree within a coalescent framework. First, we used a Bayesian MCMC analysis implemented in the program *BEAST version 1.6.1 (Drummond and Rambaut, 2007; Heled and Drummond, 2010) to estimate a posterior distribution of the species tree based on gene trees estimated from the individual loci. *BEAST analyses were performed using the mitochondrial and nuclear loci, as well as only the nuclear loci. Gene trees were estimated for the individual loci using the best-fit substitution models identified for each locus (as described above) and using a relaxed uncorrelated lognormal clock (Drummond et al., 2006). Differences in ploidy between the mitochondrial and nuclear genome were set to account for the smaller effective population size of the mtDNA locus. Species tree estimation was modeled with a Yule process. Four replicate analyses were run, each for 500 million generations with sampling events every 50,000 generations. Replicate analyses were each started using a different random number seed. Tracer was used to assess –lnL and ESS

values for convergence across replicate analyses, where an ESS of 200 or greater for combined independent runs for each scenario was regarded as representative of sufficient posterior sampling. Analyses both using all loci and excluding mtDNA appeared to converge on the posterior distribution before 200 million generations. The program LogCombiner was used to combine posterior distributions across replicates using a burnin of 200 million generations, and we used the program TreeAnnotater to generate a Maximum clade credibility tree.

To estimate the effect of each gene on the species tree posterior distribution, we ran a number of additional analyses, including: 1) a series of analyses that excluded a single nuclear locus, and 2) a series of analyses that excluded both the mtDNA locus and a single nuclear locus. All of these subsequent analyses were performed as described above for the total set of loci. All these analyses appeared to converge on the posterior distribution before 250 million generations, with the exception of one replicate for an analysis that excluded both mtDNA and *TRMT5*. The program LogCombiner was used to combine posterior distributions of all *BEAST analyses across replicates using a burnin of 250 million generations. To calculate a measure of dissimilarity among trees from these analyses, we calculated Robinson-Foulds distances between species trees posterior distributions using the program Treedist in PHYLIP version 3.69 (Felsenstein, 2004). In addition, to visualize the relative degree of similarity among each posterior distribution, 100 random samples from each posterior distribution were plotted in ordination space using multidimensional scaling (MDS) in the Mesquite module Tree Set Viz v2.1 (Hillis et al., 2005; Maddison and Maddison, 2010). Unweighted Robinson-Foulds (RF) distances, which measure the dissimilarity between the topology of two

trees, were calculated for all pairwise tree comparisons and used in the MDS analyses. The default step size in Tree Set Viz was used in all analyses and MDS was allowed to proceed until the first six decimal positions of the stress function value ceased to change. To avoid being trapped in local optima, this procedure was repeated multiple times to insure that similar results were being achieved. The results of MDS analyses were plotted as two-dimensional representations of multidimensional space.

Finally, as a second method of coalescent-based species tree reconstruction, we estimated the maximum likelihood species trees using the program STEM version 2.0 (Kubatko et al., 2009). STEM requires the input of an ultrametric gene tree generated for each individual locus. Therefore, we estimated Bayesian posterior distributions of gene trees for each locus using BEAST version 1.6.1 (Drummond and Rambaut, 2007). Best fitting substitution models were used for each locus and analyses were performed using a relaxed uncorrelated lognormal molecular clock. Four replicate analyses were performed for each locus. All of these BEAST runs were run for 100 million generations, sampling every 10,000 generations. We assessed convergence by assessing the distribution of lnL and parameter values (with an ESS of 200 or greater as indicative of adequate sampling of the posterior) over the course of each run using Tracer and by comparing these values across replicate runs. All replicate analyses converged on the same stable distribution prior to 10 million generations and we excluded samples from this portion of the run prior to summarizing the posterior distribution. The Maximum clade credibility tree for each locus was summarized from the combined posterior distribution of each replicate analysis. Single locus-gene trees were input into STEM and maximum likelihood species trees were estimated using all loci and estimated using only nuclear loci. We used a

range of prior values for Θ (0.0000001, 0.000001, 0.00001, 0.0001, 0.0006, 0.001, 0.006, 0.01, 0.1, 1, 10, 100, 1000) to account for the potential effect that ancestral population size has on our results. Analyses were run using a simulated annealing search for 10 million generations, while discarding the first 1 million generations as burnin. A total of four replicate analyses were executed for each scenario.

*Bayesian concordance of gene trees*

To determine the proportion of gene trees that supported a particular clade, we calculated Bayesian concordance factors using BUCKy version 1.4.0 (Ane et al., 2007; Larget et al., 2010). All BUCKy analyses were performed on posterior distributions of individual gene trees generated from *Ambystoma*-specific data sets in MrBayes. We focused these analyses on *Ambystoma* based on the lack of complete nuclear gene sampling for *Dicamptodon* and as an effort to reduce the number of tips in the analyzed trees. BUCKy analyses were conducted using both mitochondrial and nuclear gene trees, and using nuclear gene trees exclusively. For all nuclear loci, both the A and B alleles from an individual were present in the input gene trees, and we used a function within BUCKy to choose the allele designated with the A label (which was randomly assigned to the two gene copies) from each individual to use for analysis. All BUCKy analyses were run for 10 million generations after an initial burnin of 1 million generations. For each analysis, four independent replicate runs were performed, each with four MCMC chains. We ran multiple analyses using a range of Dirichlet process priors ($\alpha = 0.001$, 0.01, 0.1, 1.0, 10.0, 100.0) where $\alpha$ is an *a priori* parameter indicating the degree of discordance between different genes. To assess the effect of the random sampling of

alleles from an individual on the posterior distribution of concordance factors, 100 replicate analyses using all nuclear loci were performed as described above, using an $\alpha$ prior of 1.0. The primary concordance tree was generated for each run and input into the PHYLIP (v3.69) program Treedist (Felsenstein, 2004) to calculate pairwise Robinson-Foulds distances between all the resulting primary concordance trees.

## Results

A total of 14 nuclear loci and one mitochondrial locus (*nad2* and the adjacent $tRNA^{Trp}$ and $tRNA^{Ala}$ genes) were sequenced for the majority of the 33 *Ambystoma* individuals. The exceptions were the *PSME3* locus (31 *Ambystoma* individuals) and the *CD81* locus (32 individuals). For *Dicamptodon*, we were able to sequence three nuclear loci for all seven individuals; however, we were unable to generate successful PCR or sequence data for the remaining nuclear loci. New mtDNA data were generated for only one *Dicamptodon copei* individual (MVZ223515). Two additional *Dicamptodon* sequences (GenBank Accessions AY916017 and AY916018) that were sequenced in a previous study (Weisrock et al., 2005) were used in the mtDNA alignment. Overall, this totaled to 4276 bp of aligned nuclear sequence data and 1183 bp of aligned mtDNA data. The nuclear DNA contained a total of 1688 variable sites and 988 parsimony informative sites across all *Ambystoma* and *Dicamptodon* individuals (Table 3). Within *Ambystoma*, the nuclear data contained a total of 1581 variable and 926 parsimony informative sites. The mtDNA contained a total of 1035 variable sites and 524 parsimony informative sites across all *Ambystoma* and *Dicamptodon* individuals (Table 3). Within *Ambystoma*, the mtDNA data contained a total of 827 variable and 441 parsimony informative sites.

*Individual Gene Tree Reconstruction*

For all loci analyzed with MrBayes, the independent replicate runs resulted in sampling patterns after 2.5 million generations that indicated convergence on the posterior distribution, including plots of stabilized lnL values and similar majority-rule consensus topologies generated from each replicate analysis. Consensus trees generated from the combined posterior distributions for each locus are presented in Figure 2. Likewise, replicate Bayesian posterior distributions of trees generated for each locus in BEAST (generated for use in STEM analyses) also exhibited similar signs of convergence. Furthermore, the topologies of majority-rule consensus trees generated from MrBayes were generally consistent with those of the Maximum clade credibility trees estimated in BEAST. Differences between these two analyses were generally the result of how each program resolved ambiguous regions of the tree; unresolved regions of the MrBayes consensus trees were left as polytomies, while corresponding relationships in BEAST trees were resolved as bifurcating branches with low posterior probabilities. The BEAST-generated Maximum clade credibility trees are not presented here.

**BEAST species tree analyses*

The *BEAST species tree generated using all mitochondrial and nuclear loci produced a monophyletic *Ambystoma* clade with a high posterior probability (PP) of 1.0 (Fig. 3A). Within the *Ambystoma* clade, eight lineages were resolved that exhibited moderate to strong PP support and either contained multiple species or were monotypic with no strong placement with other species. *Ambystoma gracile* and *A. talpoideum* were placed in a clade (clade A) with a PP = 0.91 and this clade was supported as the sister

lineage to a clade containing all remaining *Ambystoma* species (PP = 1.0). Within this larger clade, *A. maculatum* (clade B) was placed as sister to the remaining *Ambystoma* clades (C-H) with strong support (PP = 1.0). There was weak branch support for the relationships among *A. macrodactylum* (clade C), *A. opacum* (clade D), and a clade of remaining *Ambystoma* species (clades E-H), with the latter two clades placed as sister lineages with a PP = 0.48. Clades E-H were each individually supported by strong PPs, although relationships among these clades received lower measures of branch support. *Ambystoma jeffersonianum* and *A. laterale* (clade E) were resolved as sister taxa with a PP = 0.96. *Ambystoma mabeei*, *A. barbouri*, and *A. texanum* were placed in a clade (clade F) with a PP = 1.0. Clades E and F were resolved as sister lineages, although with weaker levels of branch support (PP = 0.73). *Ambystoma annulatum*, *A. bishopi*, and *A. cingulatum* were placed in a clade (clade G) with a PP = 1.0. Finally, all sampled tiger salamander taxa (*A. californiense*, *A. mexicanum*, *A. ordinarium*, and *A. tigrinum*) were placed in a clade with a PP = 1.0. Clades G and H were resolved as sister lineages, again, with weaker branch support (PP = 0.78) than that seen for the individually identified clades. Collectively, clades E-H were resolved as a monophyletic group with strong branch support (PP = 0.98).

Analyses that excluded individual loci produced species tree posterior distributions that largely overlapped in ordination space with that of the full data analysis (Fig. 4); however, there are two notable deviations. Exclusion of the mtDNA data resulted in an overlapping, but slightly different posterior distribution, compared to the full data analysis (Fig. 4A). These differences were manifested in two different ways. First, measures of branch support for some of the terminal clades described above

changed (Fig. 3B). When the mtDNA data were excluded, branch support increased slightly for clade A (PP = 0.93), which corresponds with the individual mtDNA gene tree not resolving *A. gracile* and *A. talpoideum* as sister species (Fig. 2). In addition, the exclusion of mtDNA data resulted in decreased branch support for clade E (PP = 0.77). Second, exclusion of mtDNA data resulted in some alternative phylogenetic relationships among clades C-H, although these involved branches that received low levels of branch support from both sets of analyses. Clade D was placed as the sister lineage to a clade containing clades C and clades E-H (PP = 0.54). Clades E-G were grouped together to the exclusion of clade H with a PP = 0.50. In addition, support for the placement of clades E-H in a larger clade was reduced to a PP = 0.71.

Average Robinson-Foulds distances between posterior distributions from these exclusion analyses are presented in Table 4.

### *STEM Analyses*

Species tree estimation using STEM produced results that varied according to the θ value used in an analysis. Overall, phylogenetic relationships estimated in STEM analyses were minimally consistent with those obtained with the results from Bayesian species tree analyses and Bayesian concordance analyses. Here, we present the maximum likelihood tree favored when using a θ = 0.001 (-lnL = -241915.13843), which had the greatest amount of concordance with the *BEAST species trees (Fig. 3). This STEM tree placed *A. gracile* and *A. talpoideum* as sister lineages (Fig. 5) and resolved all tiger salamander lineages as a monophyletic clade. Other well-supported clades present in the *BEAST and BUCKy results (see below) were not resolved in the ML STEM trees

(using a range of θ values: 0.0000001, 0.000001, 0.00001, 0.0001, 0.0006, 0.001, 0.006, 0.01, 0.1, 1, 10, 100, 1000). *Ambystoma bishopi* and *A. cingulatum* were resolved as sister species; however, *A. annulatum* was not placed with these species in a clade.

*Bayesian Concordance Analyses*

In analyses using all 15 loci (mtDNA and nuclear data), and analyses using only nuclear loci, there was no difference in results across a range of α priors (0.001, 0.01, 0.1, 1, 10, and 100). There was also little difference in results across replicate analyses that randomly sampled either the A or B allele from each locus. These replicate analyses yielded similar Primary Concordance (PC) trees and exhibited limited variation in concordance factors for individual branches (results not shown).

Bayesian concordance analysis using all 15 loci (mtDNA and nuclear data) produced a primary concordance tree with many similarities to the total-data species tree estimated in *BEAST (Fig 3A, 6A). Clades A-H each received a minimum concordance factor (CF) of 3.7 with a 95% credible interval of 3 and 5 (a CF could not be estimated for clade D due to the sampling of a single *A. opacum* individual). The branch separating clade A from clades B-H received a CF of 11.7 (95% credibility interval: 10, 13). In addition, the monophyly of clades C-H was supported with a CF = 5.8 (95% CI: 4, 8), and the placement of clades E-H in a larger clade received a CF = 2.4 (with a lower bound 95% CI indicating support from at least two loci). Concordance factors for the remaining inter-clade relationships in the PC tree were low, with 95% CIs that included zero or one. Alternative phylogenetic relationships resolved in the *BEAST species tree that were not present in the PC tree also received low CFs (Fig. 3A). For example, the

*BEAST tree placed clades E and F as sister lineages with a PP = 0.73, while the concordance analysis gave it a very low CF of 0.1.

Bayesian concordance analysis of only the nuclear loci produced similar results to those that included mtDNA (Fig. 6B). Differences in the PC tree between the two sets of analyses were restricted to branches that had 95% credibility intervals that included a CF < 2. For example, the nuclear PC tree included the Linguaelapsus clade (clade F + clade G) with a CF = 1.4 and a 95% CI that included 1 and 3 (Fig. 6B), while the mtDNA + nuclear PC tree placed clades G and H as sister lineages with a CF = 2.1 and a 95% CI that included 1 and 3 (Fig. 6A). Overall, branches with lower bounds on their 95% CIs that included CFs ≥ 2 were consistent across the PC trees generated from all loci (mtDNA and nuclear), and only the nuclear loci. Removal of mtDNA did decrease CFs for most branches. Concordance factors for clades A-H each decreased by approximately one, as did the branch separating clade A from all other *Ambystoma*, and the branch ancestral to clades C-H. This effect was slightly less pronounced for the ancestral branch leading to clades E-H (CF = 2.4 for all loci vs. CF = 2.0 for all nuclear loci); however, the upper bound of the 95% CI did drop by two in the nuclear analyses. While many absolute CFs were decreased by the exclusion of mitochondrial data, concordance values presented in this way are relative to the number of loci used in BUCKy analyses. For analyses using 14 loci, a CF of 14 would be the highest value that could be given.

## Discussion

### *Interpreting phylogenetic resolution within the Ambystomatidae*

In contrast to the lack of resolution between conflicting morphological and allozymic data sets in previous phylogenetic studies of *Ambystoma*, we found considerable phylogenetic resolution among data sets generated from independent nuclear and mitochondrial loci. We reached this conclusion based primarily on patterns of posterior probability support resulting from Bayesian coalescent species tree reconstruction in *BEAST and from concordance factors resulting from Bayesian concordance analysis. In general, we considered clades that received high posterior probabilities (in the range of 0.95 or greater) and concordance factors with a minimum lower confidence estimate of 2.0 (as reported in the 95% credibility interval) to represent confidently supported relationships.

Whether or not the *BEAST posterior probabilities should be interpreted as the probability that a clade is present in the true species tree is not entirely clear (Alfaro and Holder, 2006), and can depend on a number of aspects of the analysis, including priors and model assumptions (e.g., no gene tree discordance due to gene flow). Recent simulation studies suggest that Bayesian implementations of the multispecies coalescent can produce very accurate estimates of the species tree (Leache and Rannala, 2011). Here, we interpret the species tree posterior probabilities as a measure of the certainty that our data support a particular clade, and given our broad sampling of loci across the genome, we infer such clades to be strongly supported estimates of the phylogeny for ambystomatids.

The interpretation of concordance factors as measures of branch support is less clear, and cannot be viewed in the same light as a posterior probability or bootstrap value. Instead, they represent the proportion of sampled loci with reconstructed gene trees that reflect a particular relationship. Because discordance for a relationship can exist across gene trees due to a number of factors (Maddison, 1997), CFs for a true branch in the species tree can be far less than 1.0 (measured as a proportion). This can present a challenge to the systematist when CFs do not overwhelmingly indicate concordance across the majority of loci. On their own, we chose to view CFs with a lower 95% confidence boundary of 2.0 (measured as the number of our sampled loci) as a substantial measure of support for a branch, with the caveat that an alternatively reconstructed branch did not share an equal or greater CF. While alternative relationships may lack concordant patterns (and low CFs) simply through poor resolution of the gene tree, we view this as a useful ad hoc interpretation of our results.

By combining these two sets of results, we robustly resolved a number of interspecific relationships within *Ambystoma*. First, *A. gracile* and *A. talpoideum* were resolved as sister species, and this clade was placed as the sister lineage to all remaining ambystomatids. The *BEAST posterior probability for the sister relationship of these two species increased slightly from 0.91 to 0.93 when the mtDNA were excluded from the analysis (Fig. 3), a result that was not surprising given that the mtDNA gene tree did not place them as sister species (Fig. 2). We also point out that the very high CF for these species listed in Figure 3 does not necessarily support *A. talpoideum* and *A. gracile* as sister species, but instead strongly supports the bipartition between these two species and all other ambystomatids in an unrooted framework (this is because *Dicamptodon* was not

included in the Bayesian concordance analyses). As a result, our inference of support for their placement in a clade comes primarily from the posterior probability support in the nuclear *BEAST tree. Nonetheless, the branch separating these two species from all other ambystomatids is one of the most strongly supported of all relationships within the Ambystomatidae.

Second, we resolved *A. maculatum* as the sister lineage to a clade of all remaining *Ambystoma* species (Fig. 3). Interestingly, these first two sets of relationships are somewhat similar to the previous allozyme-based results, which placed these three species outside of a clade containing all remaining ambystomatids (Fig. 1B). However, the strongly supported placement of *A. gracile* in our study is notably different from the allozyme-based results.

Third, we were able to robustly resolve a number of more terminal multi-species clades. This includes the placement of *A. laterale* and *A. jeffersonianum* in a clade (Clade E), the placement of *A. mabeei* in a clade with *A. barbouri* and *A. texanum* (Clade F), the placement of *A. annulatum* in a clade with *A. bishopi* and *A. cingulatum* (Clade G), and the resolution of a tiger salamander clade (Clade H). Many of these findings represent additional support for already well-accepted phylogenetic relationships. However, the sister relationship between *A. jeffersonianum* and *A. laterale* represents a substantial deviation from previous morphological and allozymic phylogenetic estimates (Kraus, 1988; Shaffer et al., 1991). Interestingly, our results are consistent with previous mtDNA (*cytb*) results, which placed these two species as sister lineages, albeit with weak measures of branch support (Robertson et al., 2006).

Finally, our results support the placement of Clades E, F, G, and H together in a larger clade. The posterior support for this combined relationship is strong when mitochondrial and nuclear loci are analyzed together (PP = 0.98; Fig. 3A), but drops when the mitochondrial data are excluded (PP = 0.71; Fig. 3B). Interestingly, however, the lower bound on the 95% credibility interval of CFs is 2.0 in both sets of analyses (Fig. 6), indicating that even when the mtDNA data are removed, this relationship is still supported by concordant patterns in at least two nuclear loci.

*Phylogenetic ambiguity in the Ambystomatidae*

While analysis of our multi-locus data resolves much of the species tree history for the Ambystomatidae, a fair degree of ambiguity still exists in some portions of the tree. In particular, the placement of *A. macrodactylum*, *A. opacum* and the branch leading to the combined clade of Clades E, F, G, and H remains unclear. There was disagreement among data sets (i.e., mtDNA and nuclear loci vs. nuclear loci) and analyses in the resolution of these relationships; however, this disagreement among trees coincided with very low branch support. A similar pattern of low branch support and poor phylogenetic resolution is seen among Clades E-H.

The poor resolution of these branches are roughly clustered in the same region of the phylogeny, occurring after the deeper and well supported branching events involving *A. gracile*, *A. talpoideum*, *A. maculatum*, and the clade of remaining ambystomatids, but before the well supported branches leading to Clades E through H. The lack of resolution in this region of the species tree could be due to a number of factors. This region of the species tree features short branch lengths (a pattern most evident in the *BEAST trees;

Fig. 3), potentially suggesting a period of rapid diversification in the history of *Ambystoma* (Shaffer, 1993). Such an event could affect phylogenetic resolution in two ways. First, short branching events coupled with large effective population sizes would be expected to increase the prevalence of deep coalescent events in gene trees (Maddison, 1997) and provide a challenge to their accurate reconstruction, even with a relatively large number of genes (Edwards et al., 2007). Second, as branches in the species tree become shorter, the probability of mutations in genes marking those events decreases. As a result, increasing the number of genes used in species tree reconstruction will not necessarily translate into an increase in phylogenetic information (Huang et al., 2010). It is not completely clear in this study whether the regions of poor resolution found here are tied to either of these factors. However, a number (but not all) of our individual gene trees contain strongly supported branches involving lineages that are poorly resolved in our species tree reconstructions (Fig. 2), suggesting that many loci contain adequate phylogenetic information at the gene tree level. A more complete resolution of the ambystomatid species tree may be possible by sampling a larger number of loci from the same EST-based pool of genomic resources, and by including greater numbers of individuals for each species (McCormack et al., 2009).

We also point out that the STEM-based maximum likelihood estimate of the species tree produced results that were largely inconsistent with the *BEAST estimates of the species tree and the primary concordance trees estimated in BUCKy. One factor that may contribute to this starkly contrasting estimate is that STEM uses a single reconstructed tree as a representative. In contrast, Bayesian species tree analyses reconstruct a joint posterior distribution for each gene tree and Bayesian concordance

analyses utilize a posterior distribution of gene trees for each locus. In both of these cases, the variance in gene tree reconstruction is accounted for in the reconstruction of the encompassing phylogenetic history. Several of our loci exhibited limited variation, and produced consensus gene trees (used as input for our STEM analyses) that contained many poorly resolved branches. The reduced information in these point estimates of the gene trees may have constrained the STEM analyses, and it may be necessary for either a higher number of gene trees or for more well-resolved gene tree estimates to be input into STEM for more robust species tree estimation.

*Effects of individual loci on phylogenetic reconstruction*

An important consideration in multi-locus species tree reconstruction studies involves an assessment of the influence of different components of the data on the overall phylogenetic signal. Recent studies have examined the effect of the sampling ratio of individuals to loci, demonstrating that greater numbers of loci lead to an increase in accuracy for more deeply diverged branches, while more recent rapid radiations can benefit from greater sampling of individuals per species (Maddison and Knowles, 2006; McCormack et al., 2009). Furthermore, while increasing the number of loci is generally expected to increase accuracy, not all loci are equal in their information content. As a result, the increase in phylogenetic resolution and accuracy is only expected to be proportional to the information content of the added loci (Camargo et al., 2012).

Here, we examined an equally important aspect of our molecular sampling: the effect that a particular locus has on our species tree estimates. By plotting samples from the posterior distributions of analyses that excluded a single locus in multidimensional

space (Hillis et al., 2005), we were able to assess the effect of each locus on the species tree estimate. We found that the exclusion of any one locus did not greatly change the posterior estimate of the species tree (Figs. 4, 7). The largest shift in the posterior distribution was seen with the removal of the mtDNA locus, which resulted in an overlapping, but slightly shifted distribution of sampled trees in ordination space (Fig. 4A). This result is not particularly surprising given that the mtDNA alignment was almost three times larger than any individual nuclear locus, and had far more informative sites than any individual nuclear locus. However, while excluding the mitochondrial locus could have had substantial effects on species tree estimates, given the disproportionate amount of information it contained, its removal still resulted in largely overlapping posterior distributions with nuclear-based analyses, and only resulted in changes in topology that were weakly supported with or without its inclusion (Fig. 3). Analyses that excluded a single nuclear locus, or that excluded the mtDNA and a single nuclear locus, also produced posterior distributions that overlapped with each other and with the posterior distribution of the total data analysis (Figs. 4B, 7). The MCC trees constructed from each of these permutations of excluding loci produced topologies that were very similar, with the only differences involving branches with short lengths and low posterior probability support (results not shown).

Overall, the most important conclusion derived from these exclusion analyses is that our species tree reconstruction estimates using *BEAST have not been biased by the substitution patterns of any individual locus. While this does not mean that all loci are contributing equally to the phylogenetic resolution of ambystomatid phylogeny, it does help to clarify that no single locus is driving the overall resulting species tree. This is

probably most useful in clarifying that the relatively large and variable mtDNA data set is not overwhelmingly influencing the analysis, an important result given the demonstration of a disproportionate influence of highly variable mitochondrial mtDNA data when combined with less-variable nuclear data in other salamander species tree studies (Fisher-Reid and Wiens, 2011).

*Comparison to previous hypotheses*

This multi-locus study of ambystomatid salamanders has yielded a phylogeny that has some similarity to previous hypotheses (Bogart, 2003; Kraus, 1988; Robertson et al., 2006; Shaffer et al., 1991), but also contrasts in numerous ways. Like both the morphological and the allozyme-based trees, estimates using our multi-locus sequence data maintain the sister species relationships between *A. barbouri* and *A. texanum*, between *A. annulatum* and *A. cingulatum*, and between *A. tigrinum* and *A. californiense*. The placement of *A. gracile* as one of the early diverging lineages was the only remaining similarity between our multi-locus tree and the morphological hypothesis; no other species relationship contained in the morphological estimate was found in the sequence-based phylogenies.

In contrast, our multi-locus phylogenetic results have many more similarities to previous phylogenetic hypotheses generated by allozyme data and *cytb* mtDNA data. This includes the placement of *A. gracile*, *A. talpoideum*, and *A. maculatum* as early diverging lineages in the tree (although our study provides robust support for a novel placement of the *A. gracile – A. talpoideum* clade as the sister lineage to all remaining

ambystomatids), and the placement of *A. mabeei* as sister to the clade of *A. barbouri* and *A. texanum*.

The greater overall similarity between our multi-locus sequence-based results and the allozyme-based hypotheses and the rather strong overall discordance with the morphological-based hypotheses may shed light on what was previously seen as an unresolved discordance (or "clash") in phylogenetic estimates between these two types of data (Shaffer et al., 1991). In particular, our results seem to indicate that the morphologically-hypothesized Linguaelapsus clade is a phylogenetic artifact, likely resulting from the presence of homoplastic characters in the morphological data set. Similar to the allozyme-based results, our multi-locus species tree results resolve two components of the Linguaelapsus as strongly supported clades (Fig. 3): (1) *A. barbouri*, *A. mabeei*, and *A. texanum* (clade F), and (2) *A. annulatum*, *A. bishopi*, and *A. cingulatum* (clade G) (*A. bishopi* and *A. cingulatum* were only recently designated as two species (Pauly et al., 2007)). However, none of our coalescent analyses supported the placement of these two lineages together in a clade. Concordance analysis of only nuclear loci did produce a primary concordance tree that contained the Linguaelapsus clade; yet, this received a very low concordance factor (CF = 1.4) with a 95% credibility interval indicating it may only be supported by a single locus. Indeed, only one out of our 15 loci produced a gene tree that resolved the Linguaelapsus clade (Fig. 2). We point out that none of our phylogenetic results contained a set of relationships that strongly conflicted with the Linguaelapsus clade. Instead, clades F and G were intermingled with two other clades (clades E and H). Each of these was individually strongly supported, but there was at best weak support for relationships among clades.

From a morphological perspective, the Linguaelapsus clade was supported by synapomorphies in four characters derived from skull morphology. Furthermore, a clade containing *A. barbouri*, *A. texanum*, *A. annulatum*, and *A. cingulatum* was supported by an additional nine synapomorphies primarily based on skull morphology. We believe our multi-locus phylogenetic results provide two lines of evidence that these morphological character patterns do not represent a robust set of independently derived synapomorphies for component branches of the Linguaelapsus clade. First, our results are strongly discordant with a clade comprised of *A. barbouri*, *A. texanum*, *A. annulatum*, and *A. cingulatum*. Instead, we found strong support for the placement of *A. mabeei* as the sister lineage to an *A. barbouri – A. texanum* clade. This is inconsistent with the evolution of nine morphological synapomorphies for the (*A. barbouri*, *A. texanum*, *A. annulatum*, *A. cingulatum*) clade, unless there were corresponding losses of these character states in *A. mabeei* after it diverged from other Linguaelapsus species. Second, while the weak support for relationships among clades E, F, G, and H in our multi-locus results does not preclude a true Linguaelapsus clade, the potentially rapid diversification of these lineages seems unlikely to provide the evolutionary time necessary for the independent evolution of multiple skull-based morphological synapomorphies in the ancestor of the Linguaelapsus clade. If Linguaelapsus is indeed a true clade within the Ambystomatidae, it seems more likely that these apparent synapomorphies are not independent of one another and instead represent a single linked character state. Alternatively, the Linguaelapsus clade may most likely be the result of considerable convergent evolution and the presence of homoplastic character states, a pattern that is not uncommon in morphological phylogenetic analyses of salamanders, particularly in groups that feature

paedomorphic morphologies (Wiens et al., 2005), a common life-history phenotype found in various ambystomatid lineages. Overall, our results are in agreement with previous allozyme and mtDNA phylogenetic studies in finding a lack of support for many of the morphologically-derived relationships, including the Linguaelapsus clade.

## References


Alfaro, M.E., Holder, M.T., 2006. The posterior and the prior in Bayesian phylogenetics. Annu. Rev. Ecol. Evol. Syst. 37, 19-42.

Ane, C., Larget, B., Baum, D.A., Smith, S.D., Rokas, A., 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24, 412-426.

Bi, K., Bogart, J.P., 2010. Time and time again: unisexual salamanders (genus *Ambystoma*) are the oldest unisexual vertebrates. BMC Evol. Biol. 10, 238.

Bogart, J.P., 2003. Genetics and systematics of hybrid species. In: Sever, D.M. (Ed.), Reproductive biology and phylogeny of Urodela (Amphibia). Science Publishers, Inc., Enfield, New Hampshire, pp. 109-134.

Camargo, A., Avila, L.J., Morando, M., Sites, J.W., 2012. Accuracy and precision of species trees: effects of locus, individual, and base pair sampling on inference of species trees in lizards of the *Liolaemus darwinii* Group (Squamata, Liolaemidae). Syst. Biol. 61, 272-288.

Cronn, R., Cedroni, M., Haselkorn, T., Grover, C., Wendel, J.F., 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. Theor. Appl. Genet. 104, 482-489.

Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J.A., B.; Buxton, S.; Cheung, M.; Cooper, A.; Duran, C.; Field, M.; Heled, J.; Kearse, M.; Markowitz, S.; Moir, R.; Stones-Havas, S.; Sturrock, S.; Thierer,

T.; Wilson, A., 2010. Geneious v5.3. <http://www.geneious.com>.

Edwards, S., Liu, L., Pearl, D., 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA 104, 5936-5941.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401-410.

Felsenstein, J., 2004. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Fisher-Reid, M.C., Wiens, J.J., 2011. What are the consequences of combining nuclear and mitochondrial data for phylogenetic analysis? Lessons from *Plethodon* salamanders and 13 other vertebrate clades. BMC Evol. Biol. 11, 300.

Frost, D.R., Grant, T., Faivovich, J., Bain, R.H., Haas, A., Haddad, C.F.B., De Sa, R.O., Channing, A., Wilkinson, M., Donnellan, S.C., Raxworthy, C.J., Campbell, J.A., Blotto, B.L., Moler, P., Drewes, R.C., Nussbaum, R.A., Lynch, J.D., Green, D.M., Wheeler, W.C., 2006. The amphibian tree of life. Bull. Am. Mus. Nat. Hist. 297, 1-370.

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

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570-580.

Hillis, D.M., 1987. Molecular versus morphological approaches to systematics. Annu. Rev. Ecol. Syst. 18, 23-42.

Hillis, D.M., Heath, T.A., St John, K., 2005. Analysis and visualization of tree space. Syst. Biol. 54, 471-482.

Huang, H.T., He, Q.I., Kubatko, L.S., Knowles, L.L., 2010. Sources of error inherent in species-tree estimation: impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. Syst. Biol. 59, 573-583.

Jones, T.R., Kluge, A.G., Wolf, A.J., 1993. When theories and methodologies clash - a phylogenetic reanalysis of the North-American ambystomatid salamanders (Caudata, Ambystomatidae). Syst. Biol. 42, 92-101.

Kraus, F., 1988. An empirical-evaluation of the use of the ontogeny polarization criterion in phylogenetic inference. Syst. Zool. 37, 106-141.

Kubatko, L.S., Carstens, B.C., Knowles, L.L., 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25, 971-973.

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17-24.

Larget, B.R., Kotha, S.K., Dewey, C.N., Ane, C., 2010. BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26, 2910-2911.

Larson, A., 1991. A molecular perspective on the evolutionary relationships of the salamander families. Evol. Biol. 25, 211-277.

Leache, A.D., Rannala, B., 2011. The accuracy of species tree estimation under simulation: a comparison of methods. Syst. Biol. 60, 126-137.

Liu, L., Edwards, S.V., 2009. Phylogenetic analysis in the anomaly zone. Syst. Biol. 58, 452-460.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523-536.

Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55, 21-30.

Maddison, W.P., Maddison, D.R., 2004. Mesquite: a modular system for evolutionary analysis. Version 1.05. <http://mesquiteproject.org>.

McCormack, J.E., Huang, H., Knowles, L.L., 2009. Maximum likelihood estimates of species trees: how accuracy of phylogenetic inference depends upon the divergence history and sampling design. Syst. Biol. 58, 501-508.

Pauly, G.B., Piskurek, O., Shaffer, H.B., 2007. Phylogeographic concordance in the southeastern United States: the flatwoods salamander, *Ambystoma cingulatum*, as a test case. Mol. Ecol. 16, 415-429.

Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253-1256.

Putta, S., Smith, J.J., Walker, J.A., Rondet, M., Weisrock, D.W., Monaghan, J., Samuels, A.K., Kump, K., King, D.C., Maness, N.J., Habermann, B., Tanaka, E., Bryant, S.V., Gardiner, D.M., Parichy, D.M., Voss, S.R., 2004. From biomedicine to natural history research: EST resources for ambystomatid salamanders. BMC Genomics 5, 54.

Rambaut, A., Drummond, A.J., 2009. Tracer v1.5. <http://tree.bio.ed.ac.uk/software/tracer>.

Robertson, A.V., Ramsden, C., Niedzwiecki, J., Fu, J.Z., Bogart, J.P., 2006. An unexpected recent ancestor of unisexual *Ambystoma*. Mol. Ecol. 15, 3339-3351.

Roelants, K., Gower, D.J., Wilkinson, M., Loader, S.P., Biju, S.D., Guillaume, K., Moriau, L., Bossuyt, F., 2007. Global patterns of diversification in the history of

modern amphibians. Proc. Natl. Acad. Sci. USA 104, 887-892.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Shaffer, H.B., 1993. Phylogenetics of model organisms - the laboratory axolotl, *Ambystoma mexicanum*. Syst. Biol. 42, 508-522.

Shaffer, H.B., Clark, J.M., Kraus, F., 1991. When molecules and morphology clash - a phylogenetic analysis of the North-American ambystomatid salamanders (Caudata, Ambystomatidae). Syst. Zool. 40, 284-303.

Shaffer, H.B., McKnight, M.L., 1996. The polytypic species revisited: genetic differentiation and molecular phylogenetics of the tiger salamander *Ambystoma tigrinum* (Amphibia: Caudata) complex. Evolution 50, 417-433.

Smith, J.J., Kump, D.K., Walker, J.A., Parichy, D.M., Voss, S.R., 2005. A comprehensive expressed sequence tag linkage map for tiger salamander and Mexican axolotl: enabling gene mapping and comparative genomics in *Ambystoma*. Genetics 171, 1161-1171.

Weisrock, D.W., Harmon, L.J., Larson, A., 2005. Resolving deep phylogenetic relationships in salamanders: analyses of mitochondrial and nuclear genomic data. Syst. Biol. 54, 758-777.

Weisrock, D.W., Macey, J.R., Ugurtas, I.H., Larson, A., Papenfuss, T.J., 2001. Molecular phylogenetics and historical biogeography among salamandrids of the "true" salamander clade: rapid branching of numerous highly divergent lineages in *Mertensiella luschani* associated with the rise of Anatolia. Mol. Phylogenet. Evol. 18, 434-448.

Weisrock, D.W., Smith, S.D., Chan, L.M., Biebouw, K., Kappeler, P.M., Yoder, A.D., 2012. Concatenation and concordance in the reconstruction of mouse lemur phylogeny: an empirical demonstration of the effect of allele sampling in phylogenetics. Mol. Biol. Evol. 29, 1615-1630.

Wiens, J.J., Bonett, R.M., Chippindale, P.T., 2005. Ontogeny discombobulates phylogeny: paedomorphosis and higher-level salamander relationships. Syst. Biol. 54, 91-110.

Wiens, J.J., Hollingsworth, B.D., 2000. War of the iguanas: conflicting molecular and morphological phylogenies and long-branch attraction in iguanid lizards. Syst. Biol. 49, 143-159.

Zylstra, P., Rothenfluh, H.S., Weiller, G.F., Blanden, R.V., Steele, E.J., 1998. PCR amplification of murine immunoglobulin germline V genes: strategies for minimization of recombination artefacts. Immunol. Cell Biol. 76, 395-405.

Table 1. *Ambystoma* and *Dicamptodon* individuals used in this study

| Species | Tissue Source | Locality | # of Loci Sequenced |
|---|---|---|---|
| *Ambystoma annulatum* | DWW 0364/PD1 | Warren Co., MO, USA | 15 |
| *A. barbouri* | DWW 0363 | Jessamine Co., KY, USA | 15 |
| *A. bishopi* | HBS 18028 | Jackson Co., FL, USA | 15 |
| *A. bishopi* | HBS 18036 | Okaloosa Co., FL, USA | 15 |
| *A. cingulatum* | HBS 8197 | Liberty Co., FL, USA | 15 |
| *A. cingulatum* | HBS 18030 | Baker Co., FL, USA | 14 |
| *A. gracile* | MVZ 161801 | Mendocino Co., CA, USA | 15 |
| *A. gracile* | MVZ 173465 | Lane Co., OR USA | 15 |
| *A. jeffersonianum* | LSU H1207 | PA, USA | 15 |
| *A. laterale* | MVZ 188017 | Hants Co., Nova Scotia, Canada | 15 |
| *A. laterale* | DWW0376/JSE60a | Beherens Ponds, Linn Co., IA, USA | 15 |
| *A. laterale* | MVZ 173468 | Cook Co., IL, USA | 15 |
| *A. mabeei* | MVZ 144890 | Scotland Co., SC, USA | 15 |
| *A. macrodactylum* | MVZ 137198 | Missoula Co., MT, USA | 15 |
| *A. macrodactylum* | MVZ 144895 | Linn Co., OR, USA | 15 |
| *A. macrodactylum* | MVZ 161822 | Santa Cruz Co., CA, USA | 15 |
| *A. maculatum* | MVZ 144934 | Wake Co., NC, USA | 15 |
| *A. maculatum* | MVZ 187999 | Halifax Co., Nova Scotia, Canada | 15 |
| *A. maculatum* | LSU H15983 | LA, USA | 15 |
| *A. opacum* | LSU H513 | LA, USA | 15 |
| *A. talpoideum* | LSU H15996 | LA, USA | 15 |
| *A. talpoideum* | MVZ 144946 | Berkeley Co., SC, USA | 15 |
| *A. texanum* | LSU H18514 | LA, USA | 15 |
| *A. texanum* | MVZ 144954 | Douglas Co., KS, USA | 15 |
| *A. tigrinum* | DWW 1881/7247 | Goshen Co., WY, USA | 15 |
| *A. tigrinum* | DWW 1891/7877 | Washington Co., UT, USA | 15 |
| *A. mexicanum* | DWW 1774/Am3 | Area Laguna del Toro, Mexico | 15 |
| *A. californiense* | HBS 6687 | Jepson Praire Solano Co., CA, USA | 13 |

| | | | |
|---|---|---|---|
| *A. californiense* | HBS 26367 | Sonoma Co., CA, USA | 15 |
| *A. tigrinum* | DWW 2548 | Alachua Co., FL, USA | 15 |
| *A. tigrinum* | DWW 2554 | Alachua Co., FL, USA | 15 |
| *A. ordinarium* | HBS 25134 | San Jose Lagunillas, Mexico | 15 |
| *A. ordinarium* | HBS 24978 | El Pedregoso, Mexico | 15 |
| *Dicamptodon aterrimus* | MVZ 203271 | Idaho Co., ID, USA | 1 |
| *D. aterrimus* | MVZ 187983 | Valley Co., ID, USA | 3 |
| *D. aterrimus* | MVZ 187986 | Valley Co., ID, USA | 3 |
| *D. copei* | MVZ 197777 | Grays Harbor Co., WA, USA | 3 |
| *D. copei* | MVZ 223515 | Mason Co., WA, USA | 4 |
| *D. ensatus* | MVZ 230027 | San Mateo Co., CA, USA | 3 |
| *D. ensatus* | MVZ 249022 | Napa Co., CA, USA | 3 |
| *D. tenebrosus* | MVZ 246114 | Mendocino Co., CA, USA | 3 |
| *D. tenebrosus* | MVZ 187929 | Trinity Co., CA, USA | 1 |

Table 2: Forward and reverse primers for all amplified loci.

| Locus | Forward Primer | Reverse Primer |
|---|---|---|
| *AMOTL2* | 5'-AATTATATTCCCTTTCCATGTCTGTC-3' | 5'-TGCAGAAATATTTACGATTCTAGCAC-3' |
| *CD163L1* | 5'-TACTACTGTCCTCACAACACATGAAC-3' | 5'-AAACAGCTGCAGATATGTTAAACAAG-3' |
| *CD81* | 5'-CTACAGGACACATTTAGCAGATCACT-3' | 5'-ACATTCAGGTTACCAAGACAAGAAG-3' |
| E14E10 | 5'-TGAGGACTTCATCTTACACTCTGAAC-3' | 5'-TATATAGCTGCGAGACCACAAAATAC-3' |
| E16C7 | 5'-GACAGGAGAATGAGTGAGTTACAAAA-3' | 5'-AGAAGTGTTTCAACAGCATTATATCG-3' |
| *FMO3* | 5'-CAGTATCGTTTAACAGGGCCAG-3' | 5'-GTTACTAACCAATCAAACAGCAAGAA-3' |
| *IQGAP1* | 5'-AGTTATGCATTGGTTCTTATGTTCAC-3' | 5'-AAACAAAGGAATGTTTTGAATGACTT-3' |
| *KCTD3* | 5'-CTTCACCAACAAAGTTAAGCACATCT-3' | 5'-AAATTAACCCTGAATAGTGCCATC-3' |
| *LHX2* | 5'-TAACTGACTTGACTAACCCCACTATG-3' | 5'-GTCCATTGTACAAAGCCTCTATTAAA-3' |
| M13 | 5´-GTAAAACGACGGCCAG-3´ | 5´-CAGGAAACAGCTATGAC-3´ |
| mtDNA | 5'-AAGCTTTCGGGCCCATACC-3' | 5'-GCGTTTAGCTGTTAACTAAA-3' |
| *PDXDC1* | 5'-ACATAGGTTTAAAATGTGAACAGTGC-3' | 5'-GTCGTCAAATACAAAGCAAACAGTAT-3' |
| *PSME3* | 5'-GGAGAACACTGAAGTGAAAATAACAA-3' | 5'-GCATGTACCACTACTGATCTGAAACT-3' |
| *SEC22B* | 5'-ATCATGTTAATAGTGTATGTGCGGTT-3' | 5'-ATTTACACAGATTCTGCAGTACAAGG-3' |
| *TRMT5* | 5'-CCAGCTGTTAAAGTAAAGAAGGAAGT-3' | 5'-GTTTTAAAAATTTCATAAGGCAGCTC-3' |
| *ZFR* | 5'-TGATAGCTCTTAAAAGAAACCAGACA-3' | 5'-GTAGCTCAAAATCCATGACAGTAAGA-3' |

Table 3. Information for the loci sequenced in this study.

| Locus | Alignment Length (bp) | Number of *Ambystoma* Individuals Sequenced | Number of *Dicamptodon* Individuals Sequenced | Linkage Group[a] | Number of Variable Sites | Number of Parsimony Informative Sites | Substitution Model | Number of distinct topologies[b] |
|---|---|---|---|---|---|---|---|---|
| *AMOTL2* | 349 | 33 | 0 | 8 | 165 | 113 | HKY+I | 18004 |
| *CD163L1* | 379 | 33 | 0 | 4 | 116 | 92 | HKY+Γ | 18004 |
| *CD81* | 372 | 32 | 0 | 6 | 202 | 110 | HKY+Γ | 18004 |
| E14E10[c] | 184 | 33 | 7 | 5 | 34/41* | 18/25* | GTR+I/GTR+Γ* | 18004 |
| E16C7[c] | 373/387* | 33 | 7 | 5 | 133/223* | 94/139* | HKY+I+Γ | 18004 |
| *FMO3* | 384 | 33 | 0 | 10 | 143 | 107 | GTR+Γ | 18004 |
| *IQGAP1* | 271 | 33 | 0 | 6 | 136 | 63 | GTR+Γ | 18004 |
| *KCTD3* | 211 | 33 | 0 | 13 | 94 | 56 | HKY+Γ | 18004 |
| *LHX2* | 157 | 33 | 7 | 7 | 23/33* | 18/28* | HKY+Γ | 18004 |
| mtDNA | 1183 | 33 | 3** | mtDNA | 827/1035* | 441/524* | GTR+I+Γ | 2143/3957* |
| *PDXDC1* | 225 | 33 | 0 | 3 | 89 | 54 | GTR+Γ | 18004 |
| *PSME3* | 475 | 31 | 0 | 11 | 176 | 110 | GTR+Γ | 18004 |
| *SEC22B* | 397 | 33 | 0 | 10 | 130 | 64 | GTR+Γ | 18004 |
| *TRMT5* | 214 | 33 | 0 | 14 | 50 | 27 | GTR+I+Γ | 18004 |
| *ZFR* | 271 | 33 | 0 | 2 | 90 | 47 | GTR+I | 18004 |
| Total nuclear | 4262/4276* | - | - | - | 1581/1688* | 926/988* | - | - |
| Total nuclear + mtDNA | 5445/5459* | - | - | - | 2408/2723* | 1367/1512* | - | - |

[a] Based on the linkage map of Smith et al. (2005).
[b] Based on the combined posterior distributions from MrBayes analyses.
[c] Original EST locus name; human ortholog not determined.
* Value that differs when *Dicamptodon* is included.
** Two of these sequences were originally published in Weisrock et al. (2005).

Table 4. (A) Averaged Robinson-Foulds distances comparing *BEAST posterior distributions from runs using all loci with runs excluding one locus. (B) Averaged Robinson-Foulds distances comparing *BEAST posterior distributions from runs using all loci with runs excluding mtDNA and one nuclear locus. C) Averaged Robinson-Foulds distances comparing *BEAST posterior distributions from runs using all nuclear loci with runs excluding one nuclear locus.

| Excluded Locus | A | B | C |
|---|---|---|---|
| *AMOTL2* | 6.3852 | 7.56 | 7.2776 |
| *CD163L1* | 6.3664 | 7.4528 | 7.1412 |
| *CD81* | 5.9784 | 7.2696 | 6.9832 |
| E14E10 | 5.804 | 7.5672 | 7.4772 |
| E16C7 | 5.784 | 7.5672 | 7.2936 |
| *FMO3* | 6.072 | 7.2536 | 6.8744 |
| *IQGAP1* | 6.2188 | 8.5748 | 7.6104 |
| *KCTD3* | 6.3324 | 7.6352 | 7.1728 |
| *LHX2* | 6.7448 | 7.8468 | 7.7972 |
| mtDNA | 7.2016 | - | - |
| *PDXDC1* | 6.2384 | 9.3464 | 8.36 |
| *PSME3* | 5.9304 | 7.046 | 6.8124 |
| *SEC22B* | 5.6036 | 6.9816 | 6.6608 |
| *TRMT5* | 5.8196 | 7.0389 | 6.9461 |
| *ZFR* | 6.0916 | 8.7132 | 7.8472 |

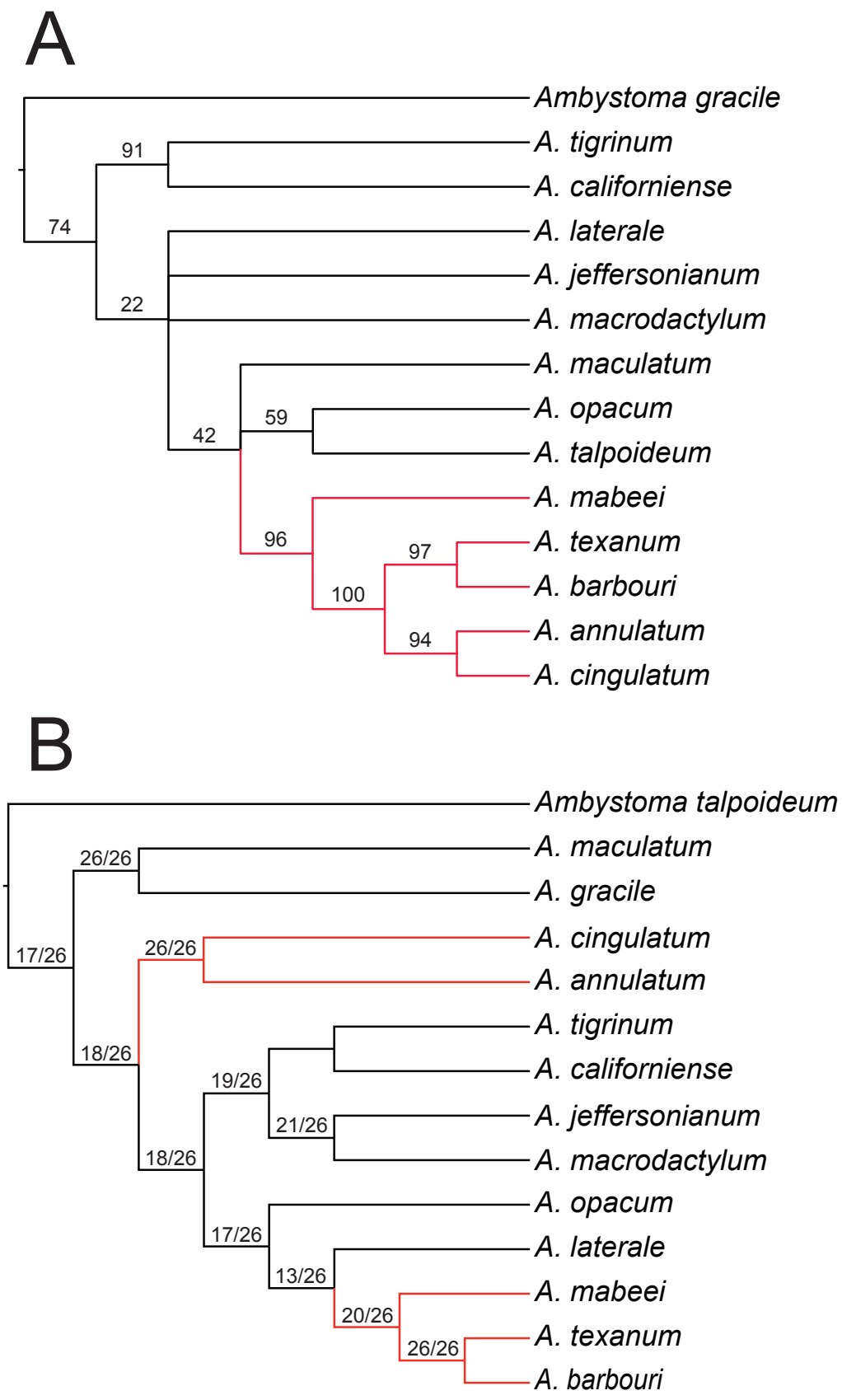
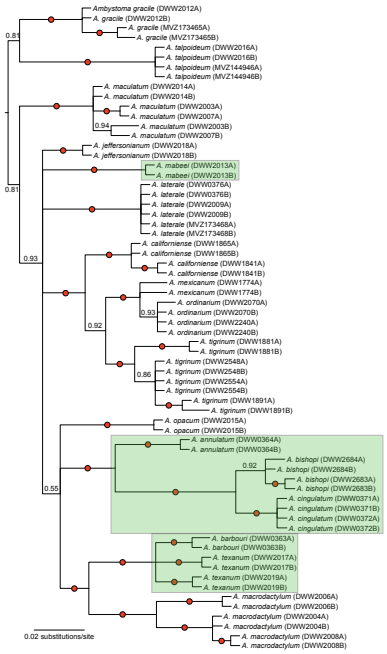
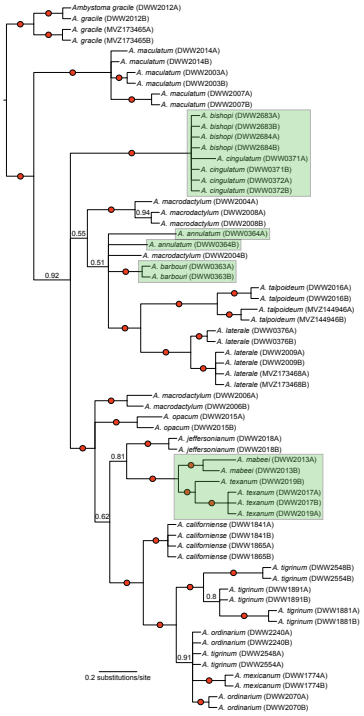

A
Ambystoma gracile
A. tigrinum
A. californiense
A. laterale
A. jeffersonianum
A. macrodactylum
A. maculatum
A. opacum
A. talpoideum
A. mabeei
A. texanum
A. barbouri
A. annulatum
A. cingulatum
91
74
22
59
42
96
97
100
94
B
Ambystoma talpoideum
A. maculatum
A. gracile
A. cingulatum
A. annulatum
A. tigrinum
A. californiense
A. jeffersonianum
A. macrodactylum
A. opacum
A. laterale
A. mabeei
A. texanum
A. barbouri
26/26
17/26
26/26
18/26
19/26
21/26
18/26
17/26
13/26
20/26
26/26

Figure 1. Morphological (A) and allozyme (B) based phylogenetic hypotheses for *Ambystoma* from previous published work (Kraus, 1988; Shaffer et al., 1991). (A) The morphological tree was based on 32 morphological characters. Values on branches represent bootstrap values. (B) This allozyme tree was based on 26 allozyme characters. Values on branches represent jackknife values. In both trees members of Linguaelapsus are highlighted in red.

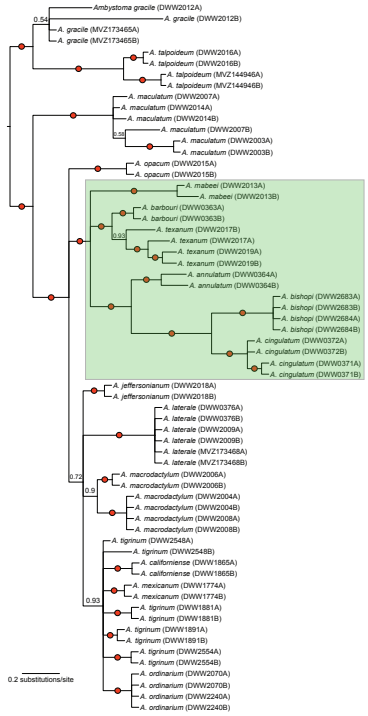
## *AMOTL2*



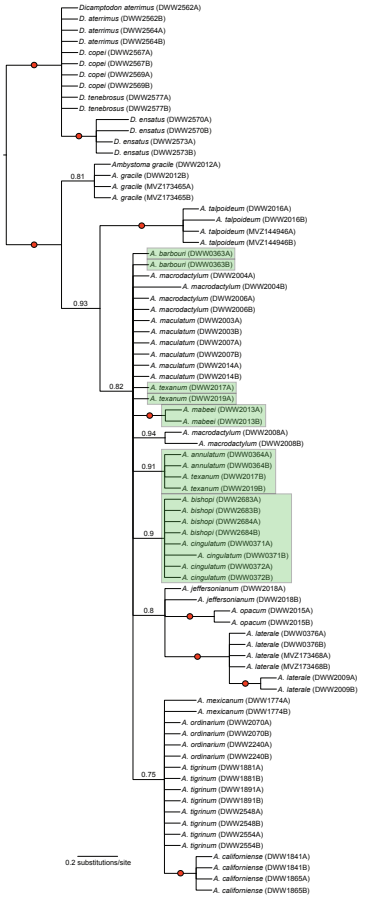

## *CD163L1*



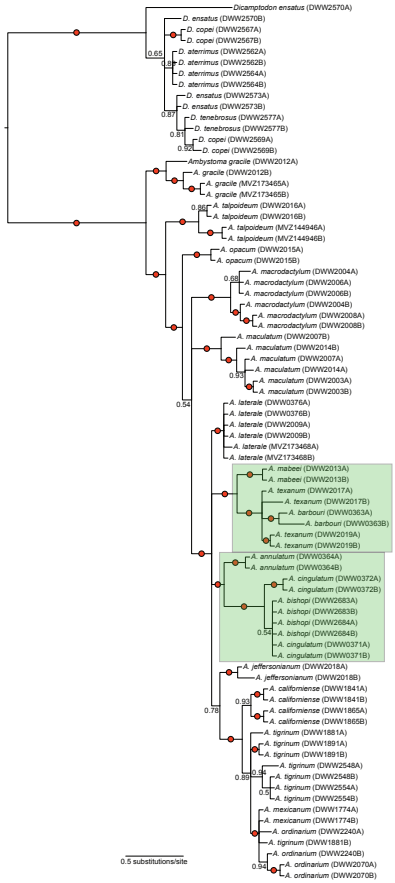

## *CD81*



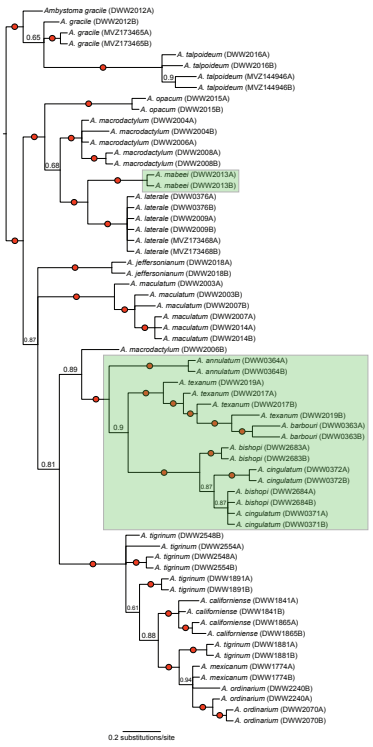

## E14E10



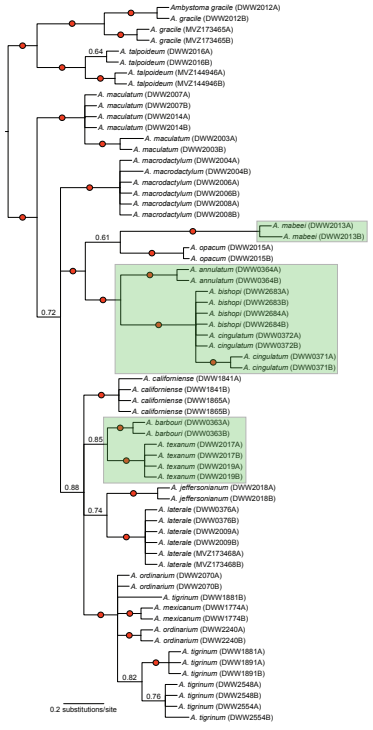
E16C7



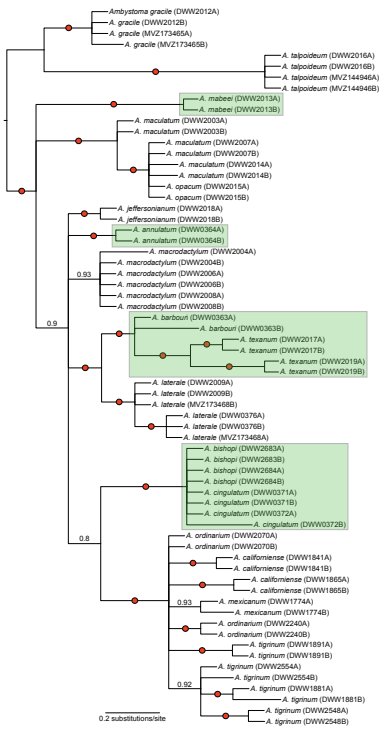

FMO3



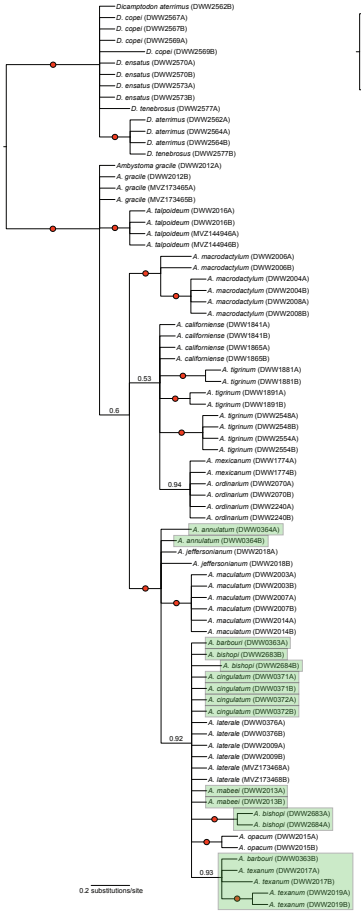

IQGAP1



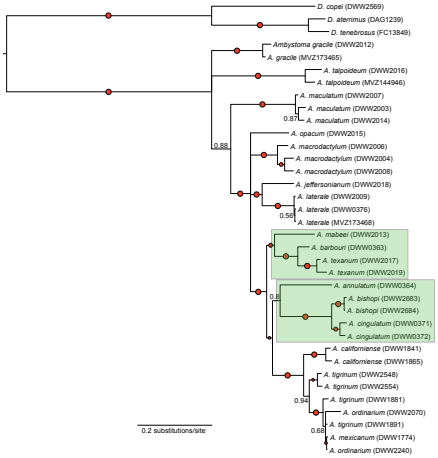

KCTD3



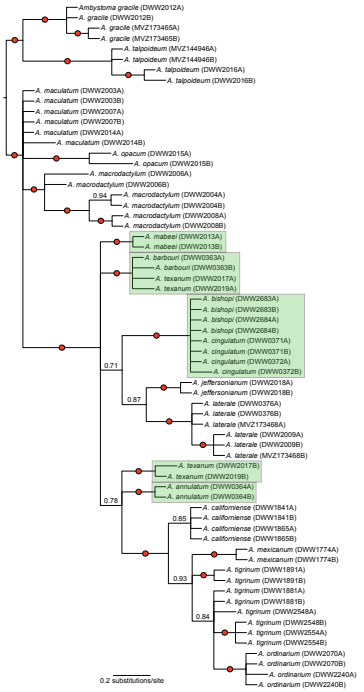
*LHX2*



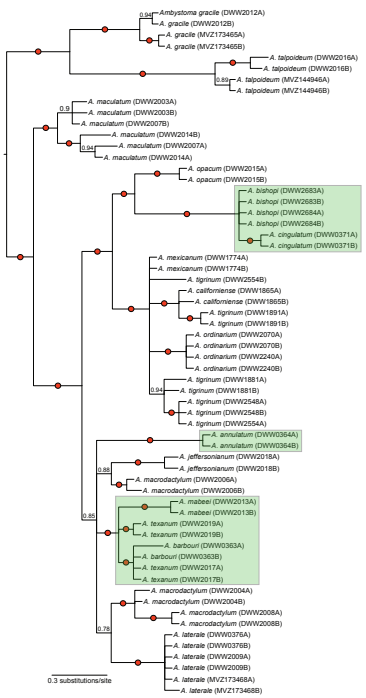

mtDNA



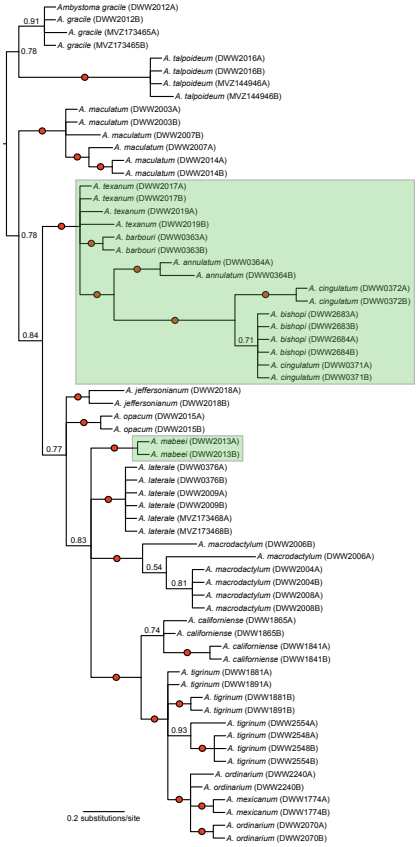

*PDXDC1*



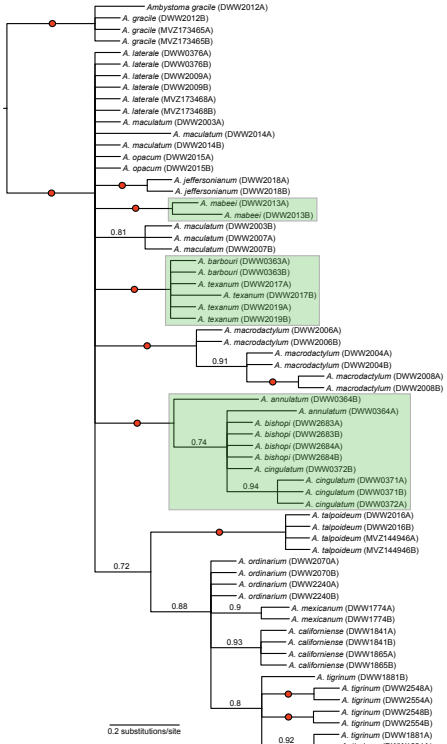

*PSME3*



## *SEC22B*




## *TRMT5*




## *ZFR*

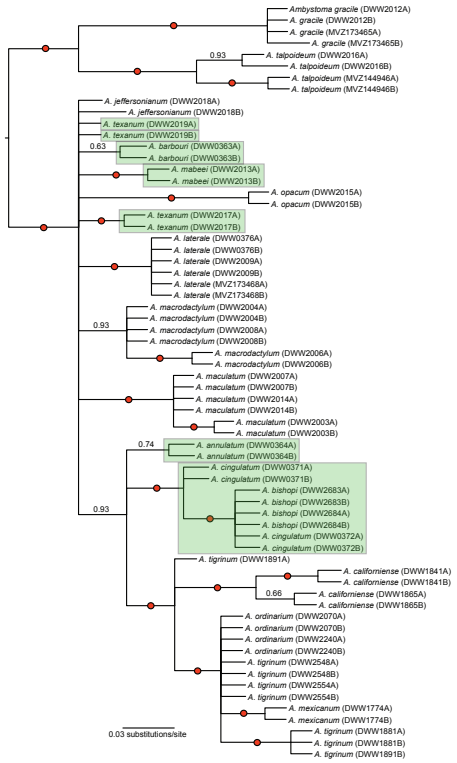
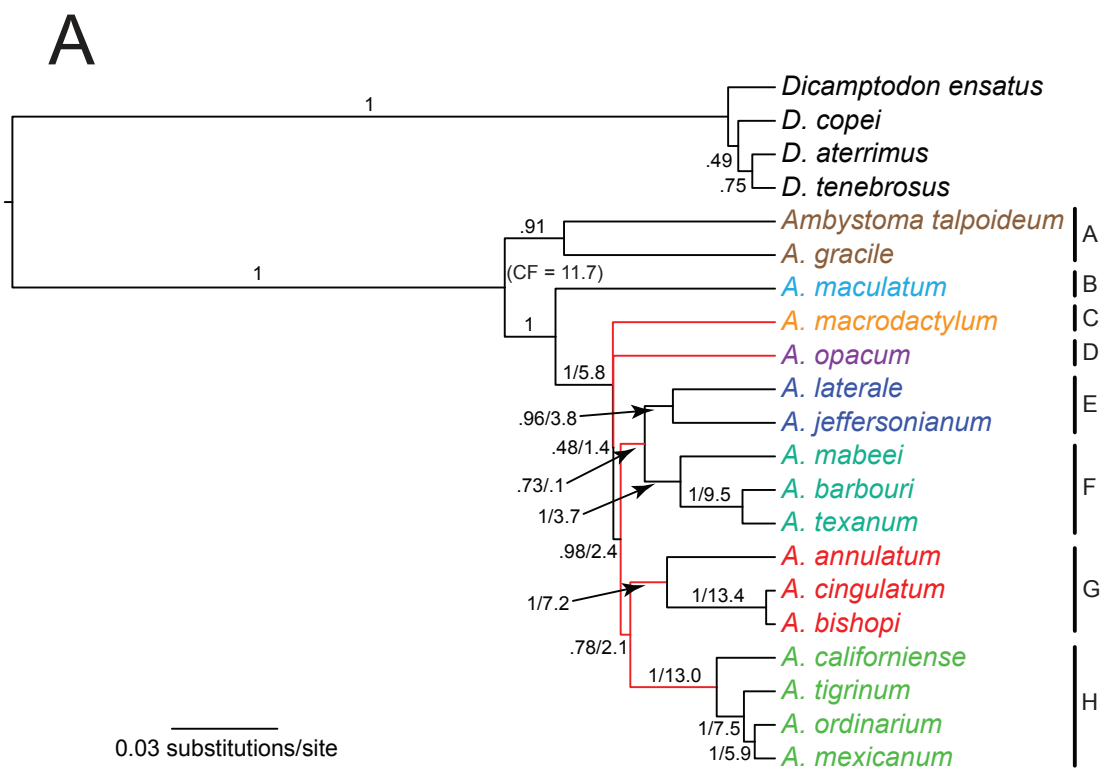


Figure 2. 50% majority-rule consensus trees generated by MrBayes for each locus. *AMOTL2*, *CD163L1*, *CD81*, E14E10, E16C7, *FMO3*, *IQGAP1*, *KCTD3*, *LHX2*, mtDNA, *PDXDC1*, *PSME3*, *SEC22B*, *TRMT5*, and *ZFR*.


A
Dicamptodon ensatus
D. copei
D. aterrimus
D. tenebrosus
Ambystoma talpoideum
A. gracile
A. maculatum
A. macrodactylum
A. opacum
A. laterale
A. jeffersonianum
A. mabeei
A. barbouri
A. texanum
A. annulatum
A. cingulatum
A. bishopi
A. californiense
A. tigrinum
A. ordinarium
A. mexicanum
A
B
C
D
E
F
G
H
1
.49
.75
.91
1
(CF = 11.7)
1
1/5.8
.96/3.8
.48/1.4
.73/.1
1/3.7
1/9.5
.98/2.4
1/7.2
1/13.4
.78/2.1
1/13.0
1/7.5
1/5.9
0.03 substitutions/site

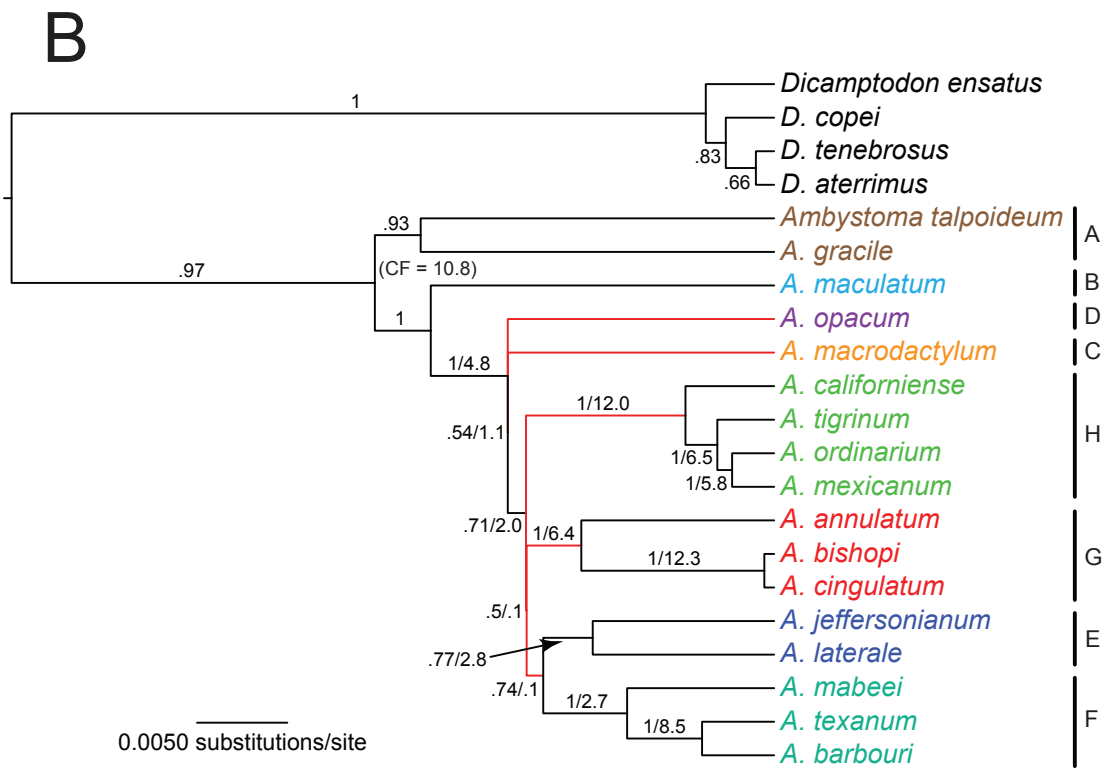
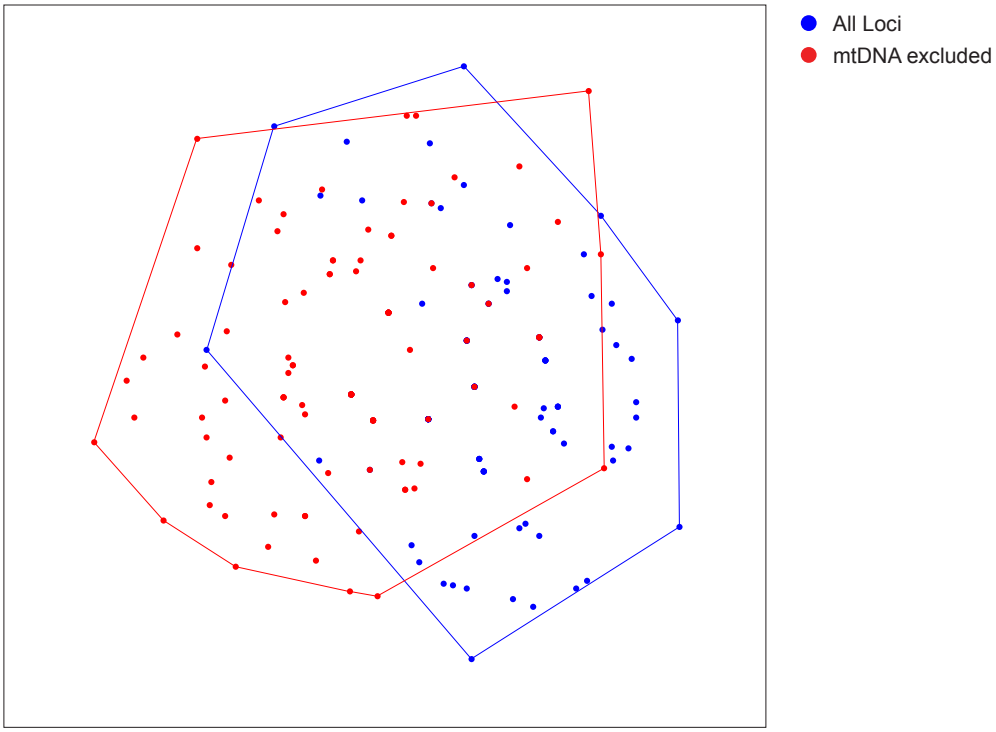

B
Dicamptodon ensatus
D. copei
D. tenebrosus
D. aterrimus
Ambystoma talpoideum
A. gracile
A. maculatum
A. opacum
A. macrodactylum
A. californiense
A. tigrinum
A. ordinarium
A. mexicanum
A. annulatum
A. bishopi
A. cingulatum
A. jeffersonianum
A. laterale
A. mabeei
A. texanum
A. barbouri
A
B
D
C
H
G
E
F
1
.83
.66
.93
.97
(CF = 10.8)
1
1/4.8
1/12.0
.54/1.1
1/6.5
1/5.8
.71/2.0
1/6.4
1/12.3
.5/.1
.77/2.8
.74/.1
1/2.7
1/8.5
0.0050 substitutions/site

Figure 3. Maximum clade credibility trees generated from the posterior distributions of Bayesian coalescent species trees reconstructions performed in *BEAST. (A) Maximum clade credibility tree generated from analyses that used all mitochondrial and nuclear loci. (B) Maximum clade credibility tree generated from analyses that used only nuclear loci. Numbers on branches represent *BEAST posterior probabilities (left of the slash) and Bayesian concordance factors (right of the slash). Bayesian concordance factors are not included on all branches due to the exclusion of *Dicamptodon* outgroup taxa in Bayesian concordance analysis. Lettered clade labels A-H correspond to ambystomatid clades that received high posterior support. Branches in red highlight relationships that differ between the two trees.

A


All Loci
mtDNA excluded

B

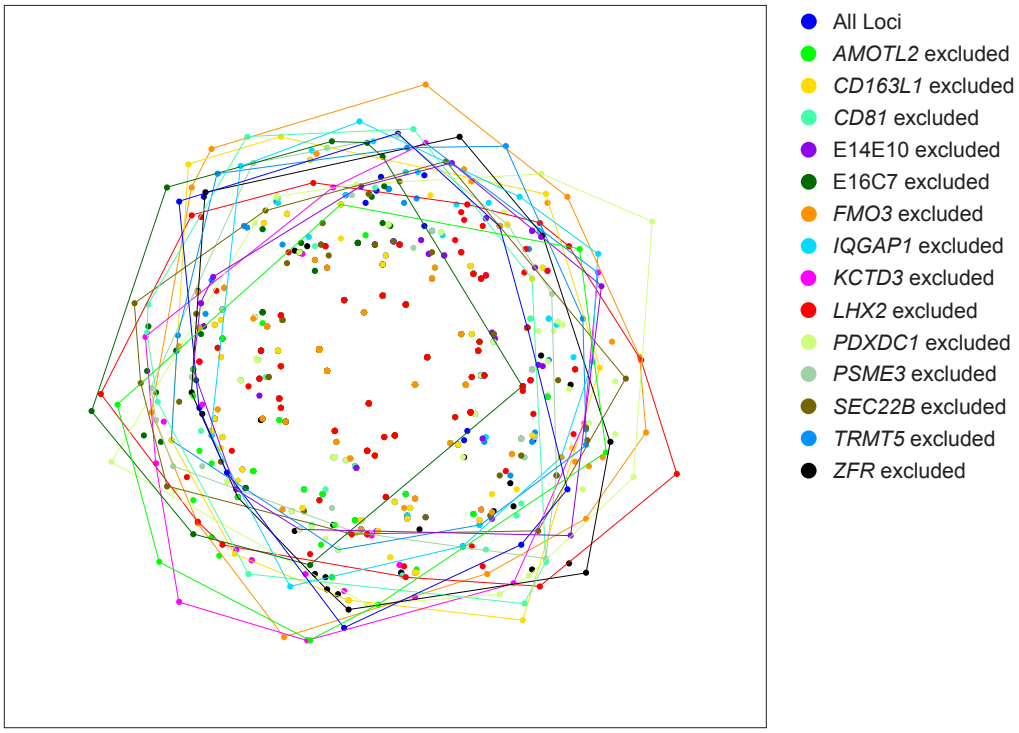

All Loci
AMOTL2 excluded
CD163L1 excluded
CD81 excluded
E14E10 excluded
E16C7 excluded
FMO3 excluded
IQGAP1 excluded
KCTD3 excluded
LHX2 excluded
PDXDC1 excluded
PSME3 excluded
SEC22B excluded
TRMT5 excluded
ZFR excluded

Figure 4. Ordination plots based on multidimensional scaling of trees sampled from the posterior distributions (PDs) generated from Bayesian species tree analyses. (A) The PDs resulting from analysis of all 15 (mitochondrial + nuclear) loci and analysis of all nuclear loci (excluding mtDNA). The final stress value of this analysis was 0.276388. Minimum convex polygons encompass the distribution of trees from each analysis. (B) The PDs resulting from analysis of all 15 (mitochondrial + nuclear) loci and analyses that excluded a single nuclear locus. The final stress value of this analysis was 0.301906. Minimum convex polygons encompass the distribution of trees from each analysis.

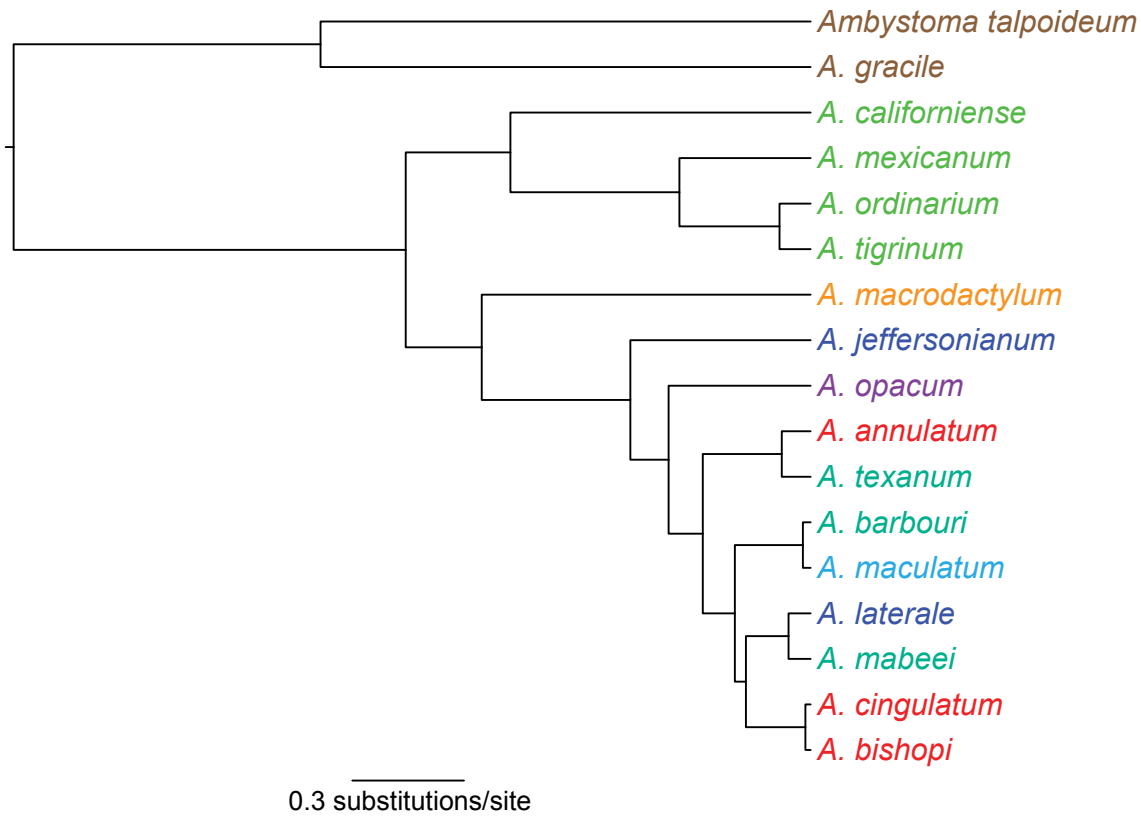



Figure 5. Coalescent-based species tree generated via a maximum likelihood analysis in STEM. The tree is based on input gene trees from all nuclear loci and the mtDNA locus. A range of θ values was used in STEM analyses. The tree presented here was generated using a $\theta = 0.001$.

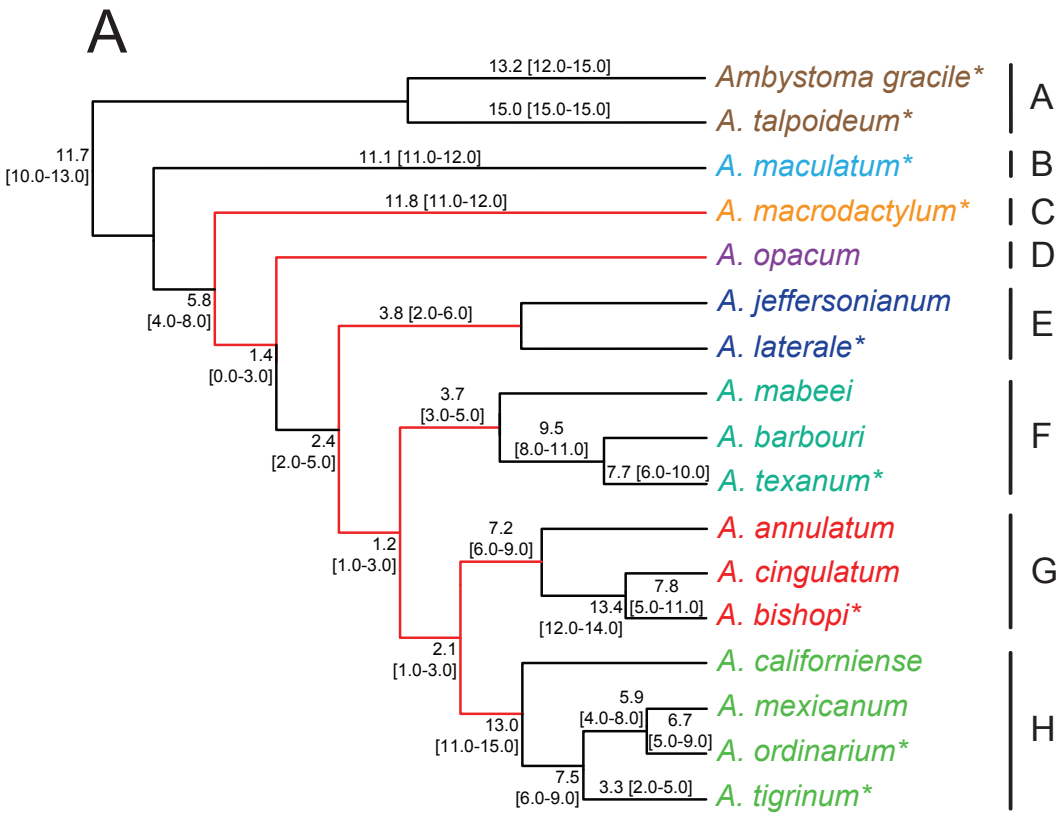

A
13.2 [12.0-15.0]
Ambystoma gracile*
15.0 [15.0-15.0]
A. talpoideum*
A
11.7
[10.0-13.0]
11.1 [11.0-12.0]
A. maculatum*
B
11.8 [11.0-12.0]
A. macrodactylum*
C
A. opacum
D
5.8
[4.0-8.0]
3.8 [2.0-6.0]
A. jeffersonianum
A. laterale*
E
1.4
[0.0-3.0]
3.7
[3.0-5.0]
A. mabeei
9.5
[8.0-11.0]
A. barbouri
2.4
[2.0-5.0]
7.7 [6.0-10.0]
A. texanum*
F
7.2
[6.0-9.0]
A. annulatum
1.2
[1.0-3.0]
A. cingulatum
7.8
13.4
[5.0-11.0]
[12.0-14.0]
A. bishopi*
G
2.1
[1.0-3.0]
A. californiense
5.9
[4.0-8.0]
A. mexicanum
13.0
[11.0-15.0]
6.7
[5.0-9.0]
A. ordinarium*
7.5
[6.0-9.0]
3.3 [2.0-5.0]
A. tigrinum*
H
B
12.2 [11.0-14.0]
Ambystoma gracile*
14.0 [14.0-14.0]
A. talpoideum*
A
10.8
[9.0-12.0]
10.1 [10.0-11.0]
A. maculatum*
B
A. opacum
D
10.8 [10.0-11.0]
A. macrodactylum*
C
4.8
[3.0-6.0]
6.4
[5.0-8.0]
A. annulatum
A. cingulatum
6.8
1.4
[1.0-3.0]
12.3
[4.0-10.0]
[11.0-13.0]
A. bishopi*
G
1.1
[0.0-3.0]
2.7
[2.0-4.0]
A. mabeei
8.5
[7.0-10.0]
A. barbouri
6.7
[5.0-9.0]
2.0
[2.0-3.0]
A. texanum*
F
2.8 [1.0-5.0]
A. jeffersonianum
12.0 [11.0-13.0]
A. laterale*
E
0.6
[0.0-2.0]
A. californiense
5.8
[4.0-8.0]
A. mexicanum
12.0
[10.0-14.0]
6.7
[5.0-9.0]
A. ordinarium*
6.5
[5.0-8.0]
3.3 [2.0-5.0]
A. tigrinum*
H

Figure 6. Primary concordance trees generated with Bayesian concordance analysis. (A) Primary concordance tree based on analyses using gene trees from all nuclear loci and the mtDNA locus. (B) Primary concordance tree based on analyses using gene trees from all nuclear loci. Numbers on branches are concordance factors with their 95% credibility interval presented inside brackets. Lettered clade labels correspond to those presented in Figure 2. Red branches highlight relationships that differ between primary concordance trees from the total data analysis and the nuclear analysis. Asterisks indicate species with alleles from multiple individuals.

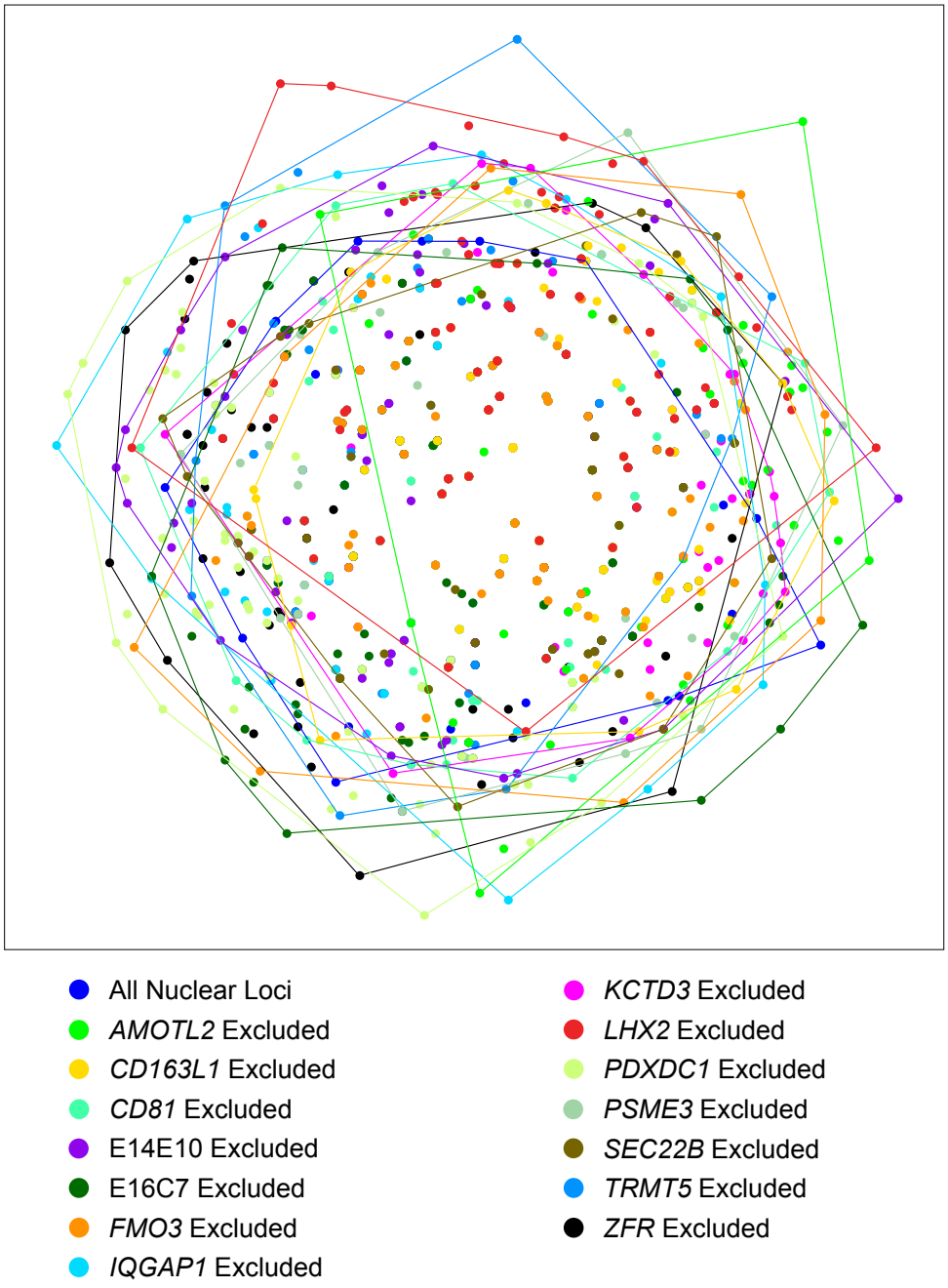



Figure 7. Ordination plot based on multidimensional scaling of trees sampled from the posterior distributions (PDs) generated from Bayesian species tree analyses. The PDs resulting from analysis of all nuclear loci (excluding mtDNA) and analyses that excluded both mtDNA and a single nuclear locus. The final stress value of this analysis was 0.306790.

# Vita

**Author's Name** – Joshua S. Williams

**Birthplace** – Ashland, KY

**Birthdate** – October 10th, 1985

## Education

2004 – 2009 B.S. Biology
Eastern Kentucky University, Richmond, KY 40475. GPA: 3.76

## Research Experience

Graduate Researcher at University of Kentucky, Lexington, KY
2011 – present Phylogenetic analysis of *Ambystoma tigrinum* complex
2009 – present Multi-locus species tree reconstruction of Ambystomatidae

Undergraduate Researcher at Eastern Kentucky University, Richmond, KY
2008 – 2009 Conservation Genetics of Gopher Tortoises (*Gopherus polyphemus*)
2007 – 2009 Evolutionary Pattern of Mitochondrial Genome Order in Amphibians

## Teaching Experience

Teaching Assistant at University of Kentucky
2009 Principles of Biology (BIO 153)
2010 – present Introductory Biology (BIO 155)

## Publications

**Williams, JS**, JH Niedzwiecki, DW Weisrock. When molecules clash: revisiting phylogenetic reconstruction of the Ambystomatid salamanders using multiple nuclear loci. *Molecular Phylogenetics and Evolution* (In Review).

O'Neill EM, R Schwartz, T Bullock, G Parra-Olea, **JS Williams**, DW Weisrock. Targeted next generation re-sequencing for population genetic and phylogenetic analysis: an implementation and informatic example using the North American tiger salamander radiation (*Ambystoma*). *Molecular Ecology* (In Press).

## Awards and Honors

2011 Society of Systematic Biologists Graduate Student Award, $1700.
2009 – present Ribble Fellowship, University of Kentucky, $4000.

| | |
|---|---|
| 2008 | LaFuze Scholarship, Eastern Kentucky University, $625. |
| 2005 – 2007 | Dean's Honor List, Eastern Kentucky University. |
| 2004, 2008 | President's Award, Eastern Kentucky University. |
| 2004 – 2008 | EKU Founder's Scholarship, Eastern Kentucky University. |
| 2004 – 2008 | KEES Scholarship. |

**Professional Organizations**

Society of Systematic Biologists
Biological Graduate Students Association, University of Kentucky