2005

COMPARATIVE ANALYSIS OF FIVE COMPLETE AMBYSTOMATID SALAMANDER MITOCHONDRIAL GENOMES

Amy K. Samuels
University of Kentucky, aksamu2@uky.edu

ABSTRACT OF THESIS

COMPARATIVE ANALYSIS OF FIVE COMPLETE AMBYSTOMATID SALAMANDER MITOCHONDRIAL GENOMES

In this study, mitochondrial transcript information from a recent EST project was extended to obtain complete mitochondrial genome sequence for 5 tiger salamander complex species (*Ambystoma mexicanum*, *A. t. tigrinum*, *A. andersoni*, *A. californiense*, and *A. dumerilii*). For the first time, aspects of mitochondrial transcription in a representative amphibian are described, and then complete mitochondrial sequence data are used to examine salamander phylogeny at both deep and shallow levels of evolutionary divergence. The available mitochondrial ESTs for *A. mexicanum* (*N*=2481) and *A. t. tigrinum* (*N*=1205) provided 92% and 87% coverage of the mitochondrial genome, respectively. Complete mitochondrial sequences for all species were rapidly obtained by using long distance PCR and DNA sequencing. A number of genome structural characteristics (base pair length, base composition, gene number, gene boundaries, codon usage) were highly similar among all species and to other distantly related salamanders. Overall, mitochondrial transcription in *Ambystoma* approximated the pattern observed in other vertebrates. From the mapping of ESTs onto mtDNA it was inferred that transcription occurs from both heavy and light strand promoters and continues around the entire length of the mtDNA, followed by post-transcriptional processing. However, the observation of many short transcripts corresponding to rRNA genes indicates that transcription may often terminate prematurely to bias transcription of rRNA genes; indeed an rRNA transcription termination signal sequence was observed immediately following the 16S rRNA gene. Phylogenetic analyses of salamander family relationships consistently grouped Ambystomatidae in a clade containing Cryptobranchidae and Hynobiidae, to the exclusion of Salamandridae. This robust result suggests a novel alternative hypothesis because previous studies have consistently identified Ambystomatidae and Salamandridae as closely related taxa. Phylogenetic analyses of tiger salamander complex species also produced robustly supported trees. The D-loop, used in previous molecular phylogenetic studies of the complex, was found to contain a relatively low level of variation and we identified mitochondrial regions with higher rates of molecular evolution that are more useful in resolving relationships among species. Our results show the benefit of using complete mitochondrial genome information in studies of recently and rapidly diverged taxa.

KEYWORDS: *Ambystoma*; EST; axolotl; Bayesian; phylogeny

Amy K. Samuels
12-01-05

COMPARATIVE ANALYSIS OF FIVE COMPLETE AMBYSTOMATID SALAMANDER MITOCHONDRIAL GENOMES

By

Amy Kathleen Samuels

Dr. S. Randal Voss
Director of Thesis

Dr. Peter Mirabito
Director of Graduate Studies

12-01-05

RULES FOR THE USE OF THESES

Unpublished theses submitted for the Master's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the thesis in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this thesis for use by its patrons is expected to secure the signature of each user.

Name	Date

THESIS

Amy Kathleen Samuels

The Graduate School

University of Kentucky

2005

COMPARATIVE ANALYSIS OF FIVE COMPLETE AMBYSTOMATID SALAMANDER MITOCHONDRIAL GENOMES

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

Amy Kathleen Samuels

Lexington, Kentucky

Director: Dr. S. Randal Voss, Professor of Biology

Lexington, Kentucky

2005

MASTER'S THESIS RELEASE

I authorize the University of Kentucky Libraries to reproduce this thesis in whole or part for purposes of research.

Signed: Amy K. Samuels

Date: 12-01-05

ACKNOWLEDGMENTS

I thank all of the people who helped me through this project, especially David Weisrock for his patience in teaching me about salamander phylogenetics and how to write about it. His contributions to this project were enormous. I also thank Jeramiah Smith for teaching me how to manipulate large amounts of data and for his consistent encouragement and support throughout my time in Kentucky. This project also benefited greatly from the hard work of Katie France, Sri Putta, and John Walker. I thank Randal Voss for turning my simple enthusiasm for salamanders into a deep appreciation for how important and powerful this animal can be in Biology, and for giving me the chance to earn a graduate degree at the same time. Thanks to the other members of the Voss lab: James Monaghan, Kevin Kump, and Robert Page for their time both inside and outside of the lab. I thank the faculty and staff of the Biology Department along with the Biology graduate students for help and friendship along the way. Thank you to my committee members Pete Mirabito and Shelly Steiner. I acknowledge NSF and NIH for grants to Randal Voss that helped fund this project. Finally, I thank both my Mom and Sara Bicker, who believed in me and gave me an incredible amount of encouragement from hundreds of miles away.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

Expressed sequence tag (EST) projects provide an efficient means to generate genome-wide information and resources. Although EST projects are used primarily to gather structural and functional information for nuclear transcripts, a relatively large proportion of sampled cDNA clones correspond to mitochondrial transcripts. As such, EST projects simultaneously provide information about the nuclear and mitochondrial transcriptomes (Gissi and Pesole, 2003). Such information is essential for correctly annotating gene boundaries and understanding transcriptional processes unique to the mitochondrial genome (Edqvist et al., 2000; Barth et al., 2001). In addition, because mitochondrial genomes are small, EST projects can rapidly yield scaffolds to obtain complete mitochondrial genome sequence (CMGS). Thus, ESTs projects are expected to benefit phylogenetic research in particular, because CMGSs will yield thousands of additional molecular characters of the type commonly used in evolutionary biology.

Reliable annotation of gene boundaries within genomes requires an assembly of both genomic DNA and mRNA sequences. Traditionally, mitochondrial genome annotation has been accomplished by comparing newly acquired mitochondrial sequences with previously published and annotated sequences obtained from genomic mtDNA. Unfortunately, genomic mtDNA does not reveal post-transcriptional modifications that determine the boundaries of protein-coding regions. For example, several mitochondrial protein-coding genes do not contain a full stop codon, but instead use polyadenylation of the terminating T nucleotide to complete the stop codon sequence (Gissi and Pesole, 2003). Identification of the correct terminating T nucleotide may not always be straightforward when comparing mtDNA sequences, but it is when mRNA sequences are available in the form of ESTs (Gissi and Pesole, 2003). For example, in *C. intestinalis* the *Cytb* reading-frame could potentially end at a TAG stop codon located within the adjacent $tRNA^{PRO}$ gene. However, the true stop codon revealed by transcript mapping of ESTs is a TA nine bp upstream that is polyadenylated to form a complete stop codon. EST projects are valuable in this regard because they provide transcriptional information necessary for genome annotation.

In some cases, complete mitochondrial sequences may provide greater phylogenetic resolution and branch support for deep evolutionary relationships that are difficult to resolve with smaller molecular datasets (Zardoya and Meyer, 2003). Presumably, in these cases, CMGSs provide a greater number of informative characters that collectively dampen misleading signals from fast evolving nucleotide sites. The use

of phylogenetic methods that can partition complete mitochondrial genomes into gene-specific subsets (that can each be fit to a different evolutionary model) may also increase the accuracy and support of deep phylogenetic relationships (Nylander et al., 2004).

The large number of characters provided by CMGS may also be useful in inter and intraspecific studies. In some cases, the use of a single gene may not provide sufficient variation to address questions of phylogeny or phylogeography, whereas CMGS from species or populations can provide ample variation (Nardi et al., 2003). This may apply especially in systems where divergence among lineages has occurred over a short evolutionary time scale, producing few substitutions marking internal branches. A comparative mitogenomic analysis among several closely related species would be informative to determine if CMGS data provides sufficient variation to resolve phylogeny in situations where individual genes or gene regions provide poor resolution. This comparative approach may also permit the identification of genes or gene regions that are more informative for low-level phylogenetic studies.

The tiger salamander complex (*Ambystoma tigrinum*) includes a diverse array of lineages distributed throughout North America and Mexico that exhibit considerable life history and morphological variation. Previous phylogenetic analysis of mtDNA D-loop sequence identified a number of clades containing geographically proximal sets of populations and species (Shaffer and McKnight, 1996). However, many phylogenetic relationships were poorly resolved because relatively few molecular characters were found to differentiate lineages of this recently derived complex (~5 million years old). To determine if CMGSs can better resolve relationships within the complex, we assembled mitochondrial transcripts from a recent EST project involving *A. mexicanum* and *A. t. tigrinum* (Putta et al., 2004), and used the resulting scaffold to obtain complete mitochondrial sequences for these and three other tiger salamander complex species (*A. andersoni, A. californiense,* and *A. dumerilii*). We used these data to explore the use of CMGS at a much deeper evolutionary level in salamanders.

At the start of this sequencing project, CMGS have been generated for single species within three salamander families: Cryptobranchidae (*Andrias davidianus*), Hynobiidae (*Ranodon sibiricus*), and Salamandridae (*Lyciasalamandra atifi*, formerly known as *Mertensiella luschani*). The availability of CMGS data for these families and the Ambystomatidae allowed us to generate new topologies and compare these to published topologies constructed using fewer mitochondrial and nuclear characters.

Since the completion of this sequencing project, an additional twenty-six salamander mitochondrial genomes have been sequenced and deposited in Genbank. Of these CMGS, 22 are from salamanders of the family Plethodontidae (Mueller et al., 2004), and the other 4 come from the families Rhyacotritonidae (*Rhyacotriton variegates,* Mueller et al., 2004), Salamandridae (*Paramesotriton hongkongensis*), Cryptobranchidae (*Andrias japonicus*), and Ambystomatidae (*Ambystoma laterale*, Mueller et al., 2004). This increase in the available number of salamander CMGS (i.e. 3 to 34 in a single year) is further support for the importance and utility of complete mitochondrial genome sequences in understanding salamander evolution.

CHAPTER TWO: MATERIALS AND METHODS

Mitochondrial EST Assembly and PCR Primer Design

ESTs were obtained from several *A. mexicanum* and *A. t. tigrinum* cDNA libraries (Putta et al., 2004). Out of 39,982 high quality transcripts, 3,686 (9.3%) were found to correspond to the salamandrid *L. atifi* mitochondrial genome sequence (GenBank accession no. **NC002756**) using BLAST searches against the NCBI database. ESTs were assembled using the DNASTAR Lasergene program to form contigs that were subsequently mapped onto the CMGS of *L. atifi* using BLASTN. Using sequence information from both *Ambystoma* contigs and *L. atifi*, PCR primers were designed using the web-based program Primer3 (http://frodo.wi.mit.edu/). Approximately 75 primer pairs were used for PCR amplification of the complete mitochondrial sequence of *A. mexicanum* and *A. t. tigrinum.* Additional primers were designed from these two sequences and used in conjunction with existing primers to amplify and sequence *A. andersoni, A. dumerilii,* and *A. californiense.* Primer sequences are available in Appendix 1.

DNA isolation, PCR, Cloning and Sequencing

Genomic DNA was isolated from liver tissue using a SDS-Proteinase K/Phenol-Chloroform procedure. Multiple individuals were used for both *A. t. tigrinum* and *A. mexicanum*; *A. mexicanum* were obtained from a laboratory stock and *A. t. tigrinum* from Charles Sullivan Corp. *Ambystoma andersoni* tissue also came from several individuals, while *A. dumerilii* and *A. californiense* DNA each came from a single individual; all of these individuals were obtained from natural populations or laboratory descendants of natural populations. PCR reactions were performed in a Biometra T-gradient thermocycler with the following conditions: 94°C for 3 min; 32 cycles of 94°C for 45s, 58.5°C for 1 min, 72°C for 2 min; 72°C for 10 min. Each reaction included between 90 ng and 300 ng total DNA, 2.5 μl dNTPs, 2.5 μl Takara ExTaq PCR Buffer, 0.25 μl Takara ExTaq, and 25ng of each primer for a total reaction volume of 25 μl. Products larger than 1 Kb were cloned into the pGEM-T vector (Promega) and sequenced (Davis Sequencing, Davis, CA). Large PCR products (2-14 Kb) were cut with restriction enzymes (*Alu*I and *Rsa*I), cloned and sequenced. Sequences were assembled using DNAStar Lasergene and manually edited.

Prior to the completion of this work, the mitochondrial genome of the axolotl, *A. mexicanum* was published (Arnason et al., 2004; GenBank accession no. **NC005797**). We compared this sequence to our independently completed *A. mexicanum*

mitochondrial genome sequence using BLASTN and found 22 single nucleotide differences and an overall nucleotide identity of 99%.

Transcript Mapping

Vertebrate mitochondrial gene transcription typically occurs from either one or two promoters on each strand to create polycistronic heavy and light-strand transcripts (Boore, 1999). The transcripts are cleaved and polyadenylated to produce mature mRNAs, tRNAs, and rRNAs. In some cases, genes form mature bicistronic transcripts in which polyadenylation is present only on the 3' gene (Gissi and Pesole, 2003). Immature polycistronic mRNAs as well as mature (polyadenylated) mRNAs were present in the salamander ESTs. Mitochondrial ESTs were assembled separately for both *A. mexicanum* and *A. t. tigrinum* using DNASTAR Lasergene. The resulting contigs for each species were mapped to the full-length mtDNA sequences using BLAST. In cases of bicistronic mRNAs, ORFs were examined to identify stop codons and gene boundaries, while polyadenylation identified gene boundaries in mature single-gene mRNAs. Polyadenylation within contigs containing transcripts for more than one gene, as well as transcripts spanning annotated gene boundaries, suggested the presence of immature mRNAs in the libraries.

Some nucleotide variation was identified between mitochondrial transcripts and genomic mtDNA, presumably because transcripts and mitochondrial DNAs were sampled from different individuals. In these cases, we reported the nucleotide from the mitochondrial genomic sequence. In cases where variable nucleotides were found within a single species' set of PCR amplified mitochondrial DNA fragments and were likely the result of sequencing error, we made majority-rule consensus base calls.

Sequence alignments

Salamander family phylogeny was investigated by aligning the five complete *Ambystoma* mitochondrial genomes with CMGS from representatives of the families Cryptobranchidae (*Andrias davidianus*; GenBank accession no. **NC004926**), Hynobiidae (*Ranodon sibiricus*; **NC004021**), and Salamandridae (*Lyciasalamandra atifi*; GenBank accession no. **NC002756**). A single frog representative (*Bufo melanostictus*; **NC005794**) and caecilian representative (*Typhlonectes natans*; **NC002471**) were used as outgroups. Sequence alignment was performed using ClustalX (Jeanmougin et al., 1998) with default parameter settings. Manual adjustments of protein-coding nucleotide sequence alignments were facilitated through translation to amino-acid sequence. A number of protein-coding genes contained regions that could not be unambiguously aligned due to

length variation and were excluded from all analyses. Manual adjustment of tRNA nucleotide sequence was facilitated using secondary-structural models to align stem regions (Kumazawa and Nishida, 1993). Any unpaired loop region containing length variation was excluded in all analyses. Alignment of rRNA sequence followed a conservative approach whereby any region containing length variation was excluded. Also excluded from phylogenetic analysis were the mitochondrial control region and a large and variable intergenic spacer between the $tRNA^{THR}$ and $tRNA^{PRO}$ genes found in all salamander sequences. The *Bufo melanostictus* genome does not contain identifiable $tRNA^{LEU(CUN)}$, $tRNA^{THR}$, and $tRNA^{PRO}$ genes, and contains an incomplete and unalignable $tRNA^{HIS}$ gene. These genes were included in phylogenetic analysis, but scored as missing data in *Bufo*.

A more exclusive alignment containing only *Ambystoma* genomes was also constructed following the same criteria as described above. This alignment included the mitochondrial control region and the $tRNA^{THR}$-$tRNA^{PRO}$ spacer and overall, fewer exclusions were necessary.

Phylogenetic Analysis

We employed partitioned Bayesian phylogenetic analysis to account for different underlying evolutionary models and parameter estimates for different gene regions (Nylander et al., 2004). Individual protein-coding genes and rRNA genes were treated as separate data partitions. All tRNAs were included into a single data partition. For the *Ambystoma*-only analyses, the control region and the $tRNA^{THR}$-$tRNA^{PRO}$ intergenic spacer were each treated as separate data partitions. Best-fit evolutionary models for each partition were chosen using the Akaike Information Criterion as implemented in the program Modeltest version 3.06 (Posada and Crandall, 2000). This was done separately for the full amphibian alignment and for the *Ambystoma*-only alignment. Bayesian analysis was performed using the program MrBayes version 3.04 (Huelsenbeck and Ronquist, 2001). Four Markov chains were used with one million generations and a tree sampling taken every 1000 generations. Flat priors were used for parameter estimates and random trees were used to start each chain. Model parameter estimates were unlinked to allow for independent estimation of parameters for each partition. All trees sampled prior to the stationary distribution were discarded as burn-in. Two additional independent runs were performed for each data set using identical conditions to verify the stationary posterior distribution.

Parsimony analysis was performed using PAUP* v4.0 (Swofford, 2002). An exhaustive tree searching option was used with equal weighting of all characters and TBR branch swapping. To assess support for branches in parsimony trees, bootstrap analysis was applied using 1000 bootstrap replicates with branch and bound tree searching. Decay indices were calculated for each internal branch using PAUP.

Alternative phylogenetic topologies were tested using the Shimodaira and Hasegawa (SH) test using 1000 RELL bootstrap replicates (Shimodaira and Hasegawa, 1999) as implemented in PAUP* v4.0. To perform these tests, a maximum-likelihood tree was found in an unconstrained analysis treating the entire data set as a single partition and using the best-fit model of evolution determined using the methods described above and estimating all parameters. The unconstrained ML tree was compared to a constrained ML tree favoring a particular topological constraint.

CHAPTER THREE: RESULTS

Mitochondrial DNA content, gene organization, and molecular evolution

The complete mitochondrial genomes of five ambystomatid salamanders were sequenced with the following lengths: 16,370 (*A. mexicanum*, *A. andersoni*, and *A. dumerilii*), 16,374 (*A. californiense*), and 16,375 bp (*A. t. tigrinium*). These five mitochondrial genomes are listed in GenBank under the accession numbers **AY659991**, **NC006888**, **NC006889**, **NC006890**, and **NC006887**, respectively. The base compositions for the light strand vary by species (Table 1); however, all *Ambystoma* mtDNA genomes are AT rich, similar to patterns found in most vertebrates. The gene order of *Ambystoma* mtDNA genomes is identical to other salamander mitochondrial genomes and to most other vertebrates (Fig. 1; Table 2). All genomes contained 13 protein-coding genes, 22 tRNAs and two rRNAs. Eight tRNAs and one mRNA are encoded on the light strand, and 14 tRNAs, 12 mRNAs, and two rRNAs are encoded on the heavy strand. Additional features present in each mtDNA sequence were an identifiable D-loop (or control region), an origin of light strand replication (O_L) located in the WANCY tRNA region, and an intergenic spacer region located between the $tRNA^{THR}$ and $tRNA^{PRO}$ genes. This spacer region has been previously identified within the genus *Ambystoma* (Shaffer and McKnight, 1996; Arnason et al., 2004), as well as in the CMGSs of *L. atifi*, *A. davidianus,* and *R. sibiricus*, but its function remains unknown.

Nucleotide substitution patterns for most *Ambystoma* mitochondrial genes (partitions) are best fit to a General Time-reversible (GTR) model with six different nucleotide substitution rates and a proportion of sites estimated to be invariant (I) (Appendix 2). One gene, *atp6*, is best fit to the less complex Hasegawa-Kishino-Yano (HKY) model. Most genes exhibit no heterogeneity in substitution rates across sites (G); however, two genes, *atp6* and *cox1*, are best fit to a model that accounts for rate heterogeneity. For the *atp6* gene, the 95% credible set of trees from the Bayesian analysis contained a wide range of estimates for the gamma shape parameter (α), up to a value of 18.61, indicating that we cannot rule out a model in which most sites within this gene evolve at a similar rate, with only a small number of sites containing relatively high or low substitution rates (Yang, 1996). In contrast, the *cox1* gene contains considerable rate heterogeneity across sites with a 95% credibility interval of α much less than one (0.05-0.34). Average Bayesian estimates of substitution rates are generally much higher for transitions than for transversions across gene partitions. However, the Bayesian posterior distribution displayed a high variance in nucleotide

substitution rate estimates and in many partitions the 95% credibility intervals for transition and transversion rates were overlapping.

A 100 bp sliding-window plot of the frequency of polymorphism across the *Ambystoma* alignment reveals a relatively low level of nucleotide variation in the D-loop region (Fig. 2). Variation within the D-loop is comparable to levels of variation within the 12S and 16S rRNA genes. Overall, variation is lowest in regions coding for tRNAs and highest in protein-coding regions. The $tRNA^{THR}$-$tRNA^{PRO}$ spacer region also contains regions of relatively high variation. The relatively low variation detected in the D-loop region is surprising because it is typically considered the most variable mitochondrial region and is widely used as a polymorphic marker in phylogeographic studies of shallowly diverged populations and species (Avise, 2004). Based on this observation, we suggest that other regions of the *Ambystoma* mitochondrial genome may provide a stronger phylogenetic signal than the D-loop region at shallow levels of divergence (see below).

Transcript Mapping

The relatively large number of mitochondrial transcripts generated from the *Ambystoma* EST project allowed us to accurately annotate mitochondrial gene boundaries and reconstruct transcriptional patterns and processes. There are several reasons why so many transcripts were sampled: (1) most eukaryotic cells contain hundreds to thousands of mitochondria that can make up 25% of the cytoplasmic volume (Lodish et al., 2000), (2) salamanders have some of the largest cell volumes of all vertebrates (Sessions and Larson, 1987), and (3) each mitochondrion can contain from 1-10 copies of mtDNA (Larsson and Clayton, 1995). Few studies have taken advantage of transcript mapping to understand mitochondrial transcription (Edqvist et al., 2000; Barth et al., 2001; Gissi and Pesole, 2003), and no studies have examined transcriptional patterns in salamander. Salamander mitochondrial gene transcription, as in other vertebrates likely occurs from both the heavy and light strand promoters and continues around the entire length of the mtDNA, followed by post-transcriptional processing. Evidence to support this transcriptional mechanism in salamanders includes the presence of bicistronic and polycistronic transcripts and the presence of intervening tRNA sequences between protein-coding and ribosomal RNA genes, indicating that multiple contiguous genes are transcribed in an uninterrupted fashion.

A total of 1,205 ESTs from *A. t. tigrinium* were assembled into 12 contigs covering 14,231 bp or 86.9% of the complete eastern tiger salamander mitochondrial

genome. These contigs mapped to 22 genes overall, including 10 complete genes (16S rRNA, *nd1*, *cox1*, *cox2*, *atp8*, *atp6*, *cox3*, *nd3*, *nd4L*, *cytb*), six partial genes (12S rRNA, *nd2*, *nd4*, *nd5*, *nd6*), and seven tRNAs (-Leu$^{(UUR)}$, -Trp, -Ala, -Asn, -Cys, -Tyr, -Ser$^{(UCN)}$) (Fig. 1). Four eastern tiger salamander ESTs from the light strand were also mapped to the non-coding D-loop region and intervening spacer region. This result suggests that both regions are post-transcriptionally processed in salamander because cDNA libraries were constructed by sampling the polyadenylated fraction of the cellular RNA pool, and no polyadenylated regions were observed flanking the intergenic spacer and D-loop regions.

The 2,481 mitochondrial ESTs from *A. mexicanum* were assembled into nine contigs covering 15,215 bp or 92.9% of the complete Mexican axolotl mitochondrial genome. These contigs mapped to 23 genes overall, including 11 complete genes (12S rRNA, 16S rRNA, *nd1*, *nd2*, *cox2*, *atp6*, *nd4L*, *nd4*, *nd5*, *nd6*, *cytb*), four partial genes (*cox1*, *atp8*, *cox3*, *nd3*), and eight tRNAs (-Val, -Leu$^{(UUR)}$, -Ser$^{(UCN)}$, -Asp, -Arg, -Leu$^{(CUN)}$; partial –Ser$^{(AGY)}$ and -Glu) (Fig. 1). As in *A. t. tigrinum*, three axolotl light strand ESTs were mapped to the non-coding D-loop region.

Assembled EST contigs from *A. mexicanum* and *A. t. tigrinum* confirmed previously published and annotated salamander mitochondrial gene boundaries where possible. The first few bases of mitochondrial genes could not always be confirmed by ESTs because either poor quality sequence ends were trimmed before assembly, or the 3' sequencing reads stopped short of the 5' gene start. In these cases, ORFs were identified from both mitochondrial genomic and EST sequence that matched annotated boundaries (*A. mexicanum: nd2*, *cox1*, *atp8*, *cox3*, *nd3*, *cytb*; *A. t. tigrinum: nd2*, *nd3*, *nd4L*, *nd5*). All protein-coding genes begin with the start codon ATG, with the exception of *cox1*, which uses GTG. The stop codons used in *Ambystoma* mtDNA genes were TAA, TAG, and AGA. Notably, the AGG stop codon was not observed in any *Ambystoma* mitochondrial genes, but is present in the mtDNAs of other salamanders (e.g. *L. atifi*). Full stop codons were present in the genes *nd1, nd2, cox1, atp8*, *atp6, nd3, nd4L*, *nd5*, and *nd6*. Polyadenylation of a T nucleotide completes the stop codon for the genes *cox2, cox3*, *nd4*, and *cytb*. The genes polyadenylated in both species were 12S rRNA, 16S rRNA, *nd1*, *nd2*, *nd3*, *nd4*, *atp6*, *cox2*, *cox3*, and *cytb*. We observed polyadenylation of *nd4L* in *A. t. tigrinum* but not *A. mexicanum*, most likely because there were relatively few transcripts representing this gene in the latter species. Though polyadenylation was only observed in these genes, the presence of D-loop and spacer

region transcripts in the *A. mexicanum* library suggests that all mitochondrial transcripts are polyadenylated in salamander.

A large percentage of mitochondrial ESTs mapped to the 12S and 16S rRNA genes in both *A. mexicanum* (64.3%) and *A. t. tigrinum* (72.7%) (Fig. 1). Mitochondrial ESTs assembled into a contig that included the 12S–$tRNA^{VAL}$–16S genes. As in human, rodents, and ascidians (Gissi and Pesole, 2003), polyadenylation was inferred in both rRNA genes in *Ambystoma*. An unusual pattern was observed for the 16S rRNA gene in both *A. mexicanum* and *A. t. tigrinum*. The majority of 16S rRNA ESTs terminated with a polyA stretch approximately 825 nucleotides upstream of the presumptive terminus. Approximately 18 out of 25 nucleotides in this region are adenine, suggesting that mispriming of oligo dT occurred here during cDNA library construction. The smaller transcript product was presumably synthesized more efficiently, resulting in a higher relative abundance in cDNA libraries.

In vertebrates, mitochondrial transcription begins with the rRNA genes and sometimes ends immediately following these genes by the conserved rRNA transcription termination signal ($TGGCAGAn_5G$). Early termination of transcription allows a greater number of rRNA products to be produced relative to the remainder of the mitochondrial genome. This signal was located in both salamander species' $tRNA^{LEU}$ gene immediately following the 16S rRNA gene. This suggests that, as in other vertebrates, rRNA genes are more heavily transcribed in the salamander mitochondrial genome than the rest of the mtDNA molecule, and correlates with the high proportion of mitochondrial rRNA ESTs detected in our libraries.

Salamander Family Phylogeny

Alignment of the amphibian CMGSs resulted in 12,474 included characters, of which 6235 are variable and 3941 are parsimony informative. Using the amphibian alignment, the HKY+G model of evolution was determined to be the best-fit model for the *atp6*, *atp8*, *cox1*, *cox2*, *cox3*, *cytb*, *nd2*, and *nd3* partitions. The more parameter rich GTR+I+G model was determined to be the best-fit model for the tRNA and rRNA partitions and for the *nd1*, *nd4*, *nd4L*, *nd5*, and *nd6* partitions. Parsimony and Bayesian analyses yielded identical tree topologies (Fig. 3). Parsimony analysis yielded a single tree of 12,656 steps in length. Bayesian analysis reached a stationary distribution by approximately 25,000 generations. The posterior distribution yielded an average lnL of -65,458.1 with a variance of 74.24. Bayesian analysis found no uncertainty regarding phylogenetic relationships among salamander families and all interfamilial branches had

posterior probabilities of 1.0. Parsimony analysis supported all interfamilial relationships with bootstrap values of 100% and decay indices of ≥63. These analyses placed the families Cryptobranchidae and Hynobiidae as sister lineages (Fig. 3). The Ambystomatidae is placed as the sister lineage to the Cryptobranchidae + Hynobiidae clade. The Salamandridae is placed as the sister lineage to the clade containing the Ambystomatidae, Cryptobranchidae, and Hynobiidae. An SH test comparing the unconstrained ML tree (lnL=-65466.38), which had the same topology as the Bayesian and parsimony tree, to a constrained ML tree (lnL=-65485.29), which placed the Salamandridae in a clade with the Ambystomatidae and made this clade the sister lineage to the clade containing the Cryptobranchidae and Hynobiidae, was nearly significant in the rejection of the alternative hypothesis (p=0.052).

These results are partially concordant with previous phylogenetic hypotheses, but also contain substantial conflict. The sister relationship between Cryptobranchidae and Hynobiidae is supported in previous analyses of nuclear and mitochondrial sequence data (Larson et al., 2003). The strong branch support from analysis of CMGS data and the lack of conflict across multiple data sets provide a high degree of confidence in the sister relationship of these two families.

However, our results conflict with previous studies in the placement of the family Salamandridae. Bayesian and ML analyses of separate and combined nuclear rRNA and mtDNA data place the families Ambystomatidae and Salamandridae in a clade with the families Dicamptodontidae, Proteidae, and Sirenidae, which is the sister clade to a clade containing the Cryptobranchidae and Hynobiidae. The families Amphiumidae, Plethodontidae, and Rhyacotritonidae are placed in more basal positions in salamander phylogeny (Larson et al., 2003). These results received strong support from Bayesian posterior probabilities (>0.95). In contrast, our analyses place the Salamandridae outside of a clade containing the Cryptobranchidae, Hynobiidae, and Ambystomatidae. This relationship is strongly supported by both Bayesian posterior probabilities and parsimony bootstrap values and decay indices (Fig. 3) and nearly rejects the previous phylogenetic hypothesis using an SH test.

Tiger Salamander Phylogeny

Alignment of the five *Ambystoma* genomes resulted in 16,309 included characters, of which 1,592 are variable and 501 are parsimony informative. Parsimony and Bayesian analysis yield identical tree topologies (Fig. 4). Parsimony yields a single tree of 1,824 steps. Bayesian analysis reached a posterior distribution by approximately

25,000 generations and the posterior distribution yielded an average lnL of -30,785.88 with a variance of 74.72. Bayesian analysis places posterior probabilities of 1.0 on every branch. Parsimony analysis supports every branch with a bootstrap value of 100% and decay indices of ≥127. *Ambystoma andersoni* and *A. mexicanum* are placed as sister taxa with *A. dumerilii* placed as the sister taxon to this clade. *Ambystoma californiense* and *A. t. tigrinum* are placed outside of the above-described clade.

These results provide robust resolution among representatives from four of the eight primary lineages of the tiger salamander complex identified by Shaffer and McKnight (1996). *Ambystoma andersoni* and *A. mexicanum* both represent a clade from the Sierra Madre Oriental and central Mexican plateau. Both parsimony and Bayesian analysis strongly support these as sister taxa (Fig. 4). The CMGS of *A. dumerilii* represents a central and western Mexican plateau clade and is placed as the sister taxon to the *A. andersoni-A. mexicanum* clade with strong support (Fig. 4). Finally, the CMGSs from *A. californiense* and *A. t. tigrinum* represent primary geographic lineages from the western and eastern United States, respectively, and are robustly placed outside of the clade of Mexican taxa (Fig. 4). The tree represented in Figure 4 is rooted with *A. californiense* because the results of Shaffer and McKnight (1996) and these phylogenetic analyses of all available salamander CMGSs (Fig. 3) both support a basal split between *A. californiense* and all remaining tiger salamander complex species.

The lack of strong branch support for branches leading to primary lineages in Shaffer and McKnight (1996) may be a product of the limited number of variable sites present in the D-loop region (Fig. 2). Their analyses used approximately 840 bases of mitochondrial sequence spanning from the 3' end of the $tRNA^{THR}$ gene through 500 bases of the 5' portion of the D-loop region. Across these five new tiger salamander mitochondrial genomes, this particular region contains 69 variable sites, of which 17 are parsimony informative. Parsimony analysis of this region produces a topology concordant with the results of Shaffer and McKnight (1996) with poor bootstrap support for the clade of Mexican species (64%, results not shown). In contrast, many individual protein-coding regions contain a considerably higher amount of variation (Fig. 2) and provide greater phylogenetic support. For example, the genes *nd1* and *nd2* are only slightly larger in size at 971 bases and 1041 bases, respectively. However, they contain over twice as many variable sites as the $tRNA^{THR}$-D-loop region used by Shaffer and McKnight (1996) (*nd1* variable bases=140, parsimony informative=44; nd2 variable bases=144, parsimony informative=35). Both of these genes individually provide robust

support (parsimony bootstrap=100%) for relationships among tiger salamander complex species.

TABLES AND FIGURES

Table 1. Base frequencies of complete *Ambystoma* mtDNA sequences.

Species	%A	%C	%G	%T	#sites
A. mexicanum	34.8	20.0	13.2	31.9	16370
A. t. tigrinum	34.7	19.7	13.3	32.2	16375
A. andersoni	34.7	20.0	13.2	31.9	16370
A. dumerilii	34.7	19.9	13.3	31.9	16370
A. californiense	34.2	20.3	13.7	31.7	16374

Table 2. Annotation and gene organization of *Ambystoma* mitochondrial genomes.

	Position[a]							
Gene	*A. mexicanum*	*A. t. tigrinum*	*A. andersoni*	*A. dumerilii*	*A. californiense*	Start	Stop[b]	Strand
$tRNA^{PHE}$	1-69	1-69	1-69	1-69	1-69			+
12S rRNA	70-998	70-997	70-998	70-998	70-996			+
$tRNA^{VAL}$	999-1068	998-1067	999-1068	999-1068	997-1066			+
16S rRNA	1069-2623	1068-2622	1069-2623	1069-2622	1067-2622			+
$tRNA^{LEU(UUR)}$	2624-2698	2623-2697	2624-2698	2623-2697	2623-2697			+
nd1	2699-3669	2698-3666	2699-3667	2698-3666	2698-3666	ATG	TAG/TAA	+
$tRNA^{ILE}$	3670-3740	3667-3739	3678-3740	3667-3739	3667-3739			+
$tRNA^{GLN}$	3741-3810	3740-3813	3741-3810	3740-3809	3740-3810			-
$tRNA^{MET}$	3811-3880	3814-3883	3811-3880	3810-3879	3811-3880			+
nd2	3881-4922	3884-4927	3881-4924	3880-4923	3881-4924	ATG	TAA	+
$tRNA^{TRP}$	4923-4993	4928-4996	4925-4993	4924-4992	4925-4993			+
$tRNA^{ALA}$	4994-5063	4997-5066	4994-5063	4993-5062	4994-5063			-
$tRNA^{ASN}$	5064-5136	5067-5139	5064-5136	5063-5135	5064-5136			-
O_L	5137-5176	5140-5179	5137-5176	5136-5176	5137-5178			-
$tRNA^{CYS}$	5177-5242	5180-5245	5177-5242	5177-5242	5179-5244			-
$tRNA^{TYR}$	5243-5310	5246-5314	5243-5310	5243-5310	5245-5313			-
cox1	5311-6858	5315-6862	5311-6858	5311-6858	5314-6861	GTG	TAA	+
$tRNA^{SER(UCN)}$	6859-6928	6863-6932	6859-6928	6859-6928	6862-6931			-
$tRNA^{ASP}$	6929-6998	6933-7002	6929-6998	6929-6998	6932-7001			+
cox2	6999-7683	7003-7688	6999-7683	6999-7683	7002-7686	ATG	T	+
$tRNA^{LYS}$	7686-7759	7689-7763	7686-7759	7686-7759	7687-7762			+
atp8	7760-7917	7764-7931	7760-7927	7760-7927	7763-7930	ATG	TAA	+
atp6	7918-8600	7922-8605	7918-8601	7918-8601	7921-8604	ATG	TAA	+
cox3	8601-9384	8605-9389	8601-9384	8601-9384	8604-9387	ATG	T	+
$tRNA^{GLY}$	9385-9453	9390-9457	9385-9453	9385-9453	9388-9456			+
nd3	9454-9799	9458-9805	9454-9801	9454-9801	9457-9804	ATG	TAG	+
$tRNA^{ARG}$	9800-9868	9806-9872	9802-9868	9802-9868	9805-9871			+
nd4L	9869-10165	9873-10169	9869-10164	9869-10165	9872-10168	ATG	TAA	+
nd4	10159-11533	10163-11537	10159-11533	10159-11533	10162-11536	ATG	T/TAG	+
$tRNA^{HIS}$	11534-11602	11521-11608	11534-11602	11534-11602	11537-11605			+
$tRNA^{SER(AGY)}$	11603-11671	11609-11677	11603-11671	11603-11671	11606-11674			+
$tRNA^{LEU(CUN)}$	11672-11743	11678-11749	11672-11743	11672-11743	11675-11746			+
nd5	11744-13540	11750-13550	11744-13540	11744-13540	11747-13546	ATG	TAA/AGA	+

Table 2. continued.

nd6	13520-14038	13526-14044	13526-14038	13520-14038	13529-14041	ATG	TAA/TAG	-
$tRNA^{GLU}$	14036-14108	14044-14114	14036-14108	14036-14108	14042-14111			-
cytb	14109-15249	14115-15253	14109-15249	14109-15249	14112-15252	ATG	T	+
$tRNA^{THR}$	15250-15318	15256-15324	15250-15318	15250-15318	15253-15321			+
Spacer	15319-15560	15325-15566	15319-15560	15319-15560	15322-15563			+
$tRNA^{PRO}$	15561-15632	15567-15638	15561-15632	15561-15632	15564-15635			-
D-loop	15633-16370	15639-16375	15633-16370	15633-16370	15636-16374			+

[a] Corresponds to nucleotide positions of each gene.
[b] Corresponds to last coding nucleotide in the gene. Incomplete stop codons are completed by polyadenylation.

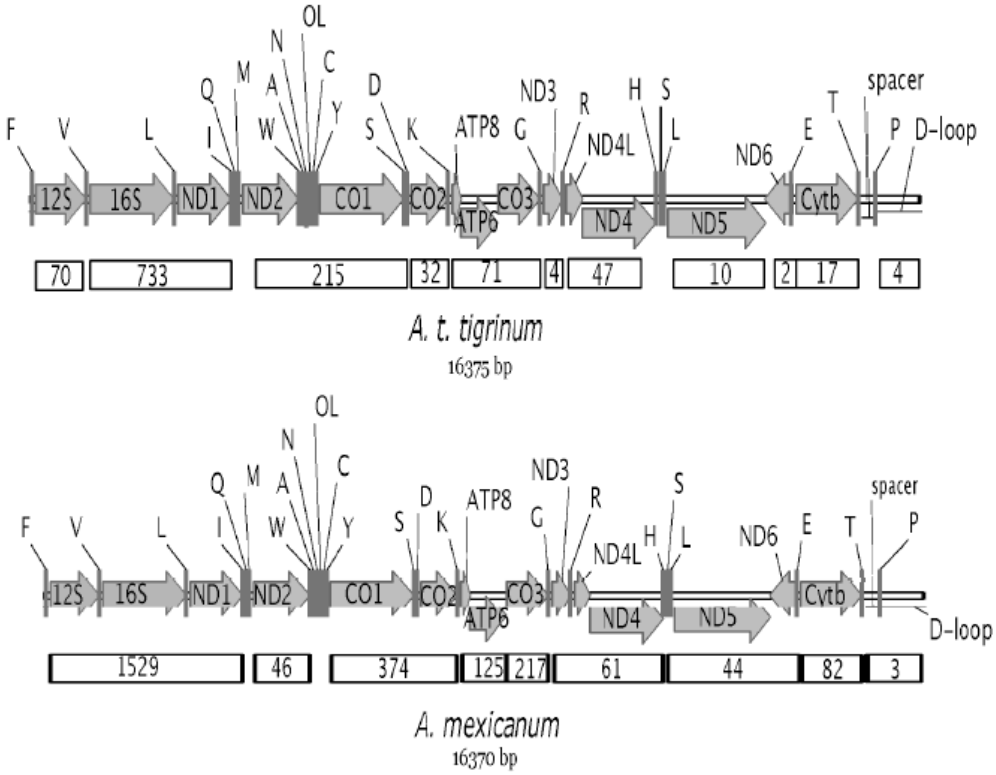

Figure 1. Transcript maps of mitochondrial EST contigs on the complete mitochondrial genome of *A. mexicanum* and *A. t. tigrinum*. Arrows represent rRNA and protein coding genes, and indicate the direction of transcription of each gene in both species. Vertical bars represent tRNAs. Boxes below the mtDNA map represent EST contigs, and include the number of ESTs that make up each contig. Note the large proportion of rRNA transcripts relative to all other transcribed genes. tRNA genes are labeled using the single letter amino acid code. OL (Origin of light strand replication); spacer ($tRNA^{THR}$-$tRNA^{PRO}$ intergenic spacer)

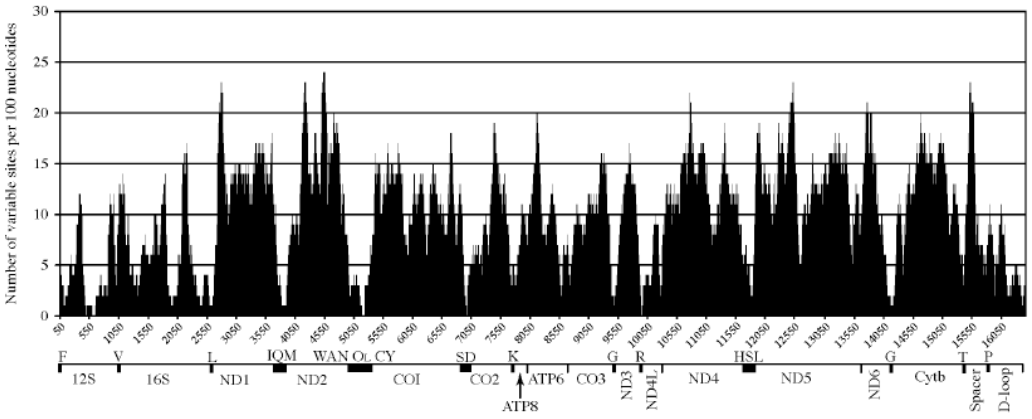

Figure 2. Plot from a sliding-window analysis of nucleotide polymorphism across an alignment of *Ambystoma* mitochondrial genomes. The sliding window calculated the number of variable sites in a block of 100 contiguous bases. Ambiguously aligned bases were excluded from the analysis. A generalized map of the mitochondrial genome is provided below the plot for reference.

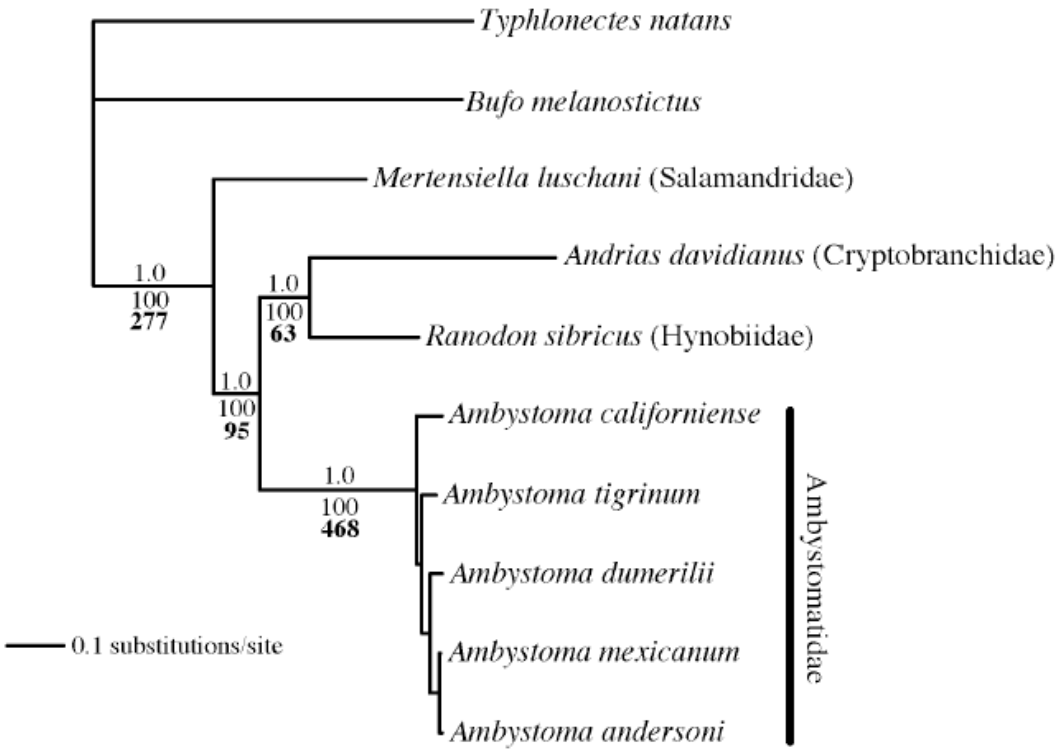

Figure 3. Consensus phylogram from a partitioned, family-level Bayesian analysis of complete salamander mitochondrial genomes. Parsimony analysis yielded a tree with an identical topology. Numbers above branches represent Bayesian posterior probabilities. Numbers directly below branches represent parsimony bootstrap values and bold numbers below branches represent decay indices. Names within parentheses indicate three of the four represented salamander families. The vertical black bar denotes representatives of the family Ambystomatidae. *Mertensiella luschani* = *Lyciasalamandra atifi.*

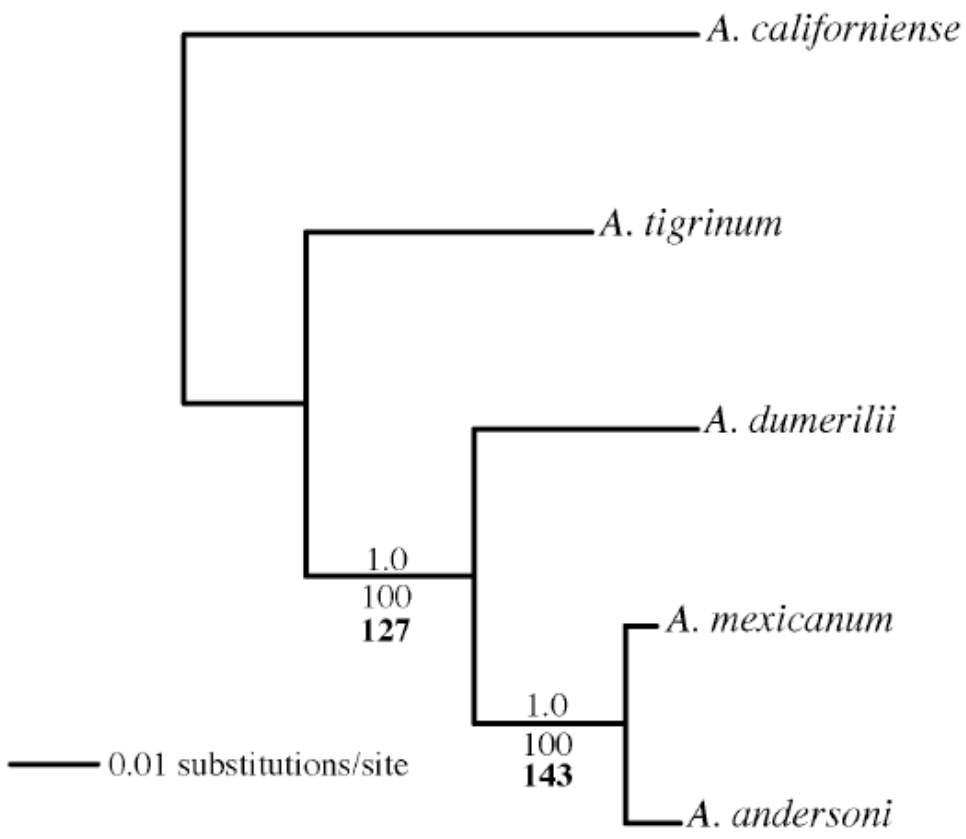

Figure 4. Consensus phylogram from a partitioned Bayesian analysis of *Ambystoma* mitochondrial genomes. Parsimony analysis yielded a tree with an identical topology. Numbers above branches represent Bayesian posterior probabilities. Numbers directly below branches represent parsimony bootstrap values and bold numbers below branches represent decay indicies.

CHAPTER FOUR: DISCUSSION

Transcriptional analyses

Nuclear integrations of mtDNA transcripts (Numts) have been identified for many species and may present a problem when trying to exclusively PCR amplify genuine mtDNA genes. It is unlikely that Numts are represented in these mtDNA sequences because of the methods used. Although mtDNA integrations into the nuclear genome are not uncommon (Bensasson et al., 2001), these genes are unlikely to be transcribed (Woischnik and Moraes, 2002). The PCR generated sequence from *A. t. tigrinum* and *A. mexicanum* matches the sequence derived from mitochondrial transcripts sampled from the cells RNA pool, indicating the amplified sequences are true mtDNA. The use of long PCR in amplifying mtDNA sequences also lowers the chance of amplifying Numts (Bensasson et al., 2001). Finally, these sequences contain a number of mtDNA characteristics, including high AT and low G nucleotide base frequencies, fully functional tRNAs, and no premature stop codons.

Phylogenetic Analyses

The basal phylogenetic position of the Salamandridae observed in this salamander family phylogeny may result from features of the genome data set. There has been a large amount of nucleotide substitution evolution since the divergence of family-level salamander lineages. Maximum-likelihood corrected sequence divergences between salamander families based on complete, but unpartitioned, CMGS data using a GTR+I+G model range from 54.4% to 83.4%, suggesting that multiple substitutions are common in our data. The inclusion of such a large number of nucleotides, of which many are saturated, coupled with extremely long branches since the divergence of salamander families [>150 Million years for the Cryptobranchidae (Gao and Shubin, 2003)] could be leading to inconsistency in the phylogeny, producing an inaccurate tree through long-branch attraction (Felsenstein, 1978). Saturated characters are expected to more strongly influence the parsimony results because all characters are equally weighted in the analysis; however, Bayesian analysis resolves the same tree.

The use of a model-based approaches and the gene partitioning of the data should act to minimize the effects of saturation in highly variable nucleotide positions or gene regions and maximize phylogenetic signal. Nonetheless, if these models do not adequately account for the underlying mode of molecular evolution, then the likelihood-based analyses may also be prone to providing an inaccurate topology (Buckley, 2002). Because a number of gene partitions in this data set were analyzed under an HKY

model of evolution (as determined through the Akaike Information Criterion), the effects of using the more parameter-rich GTR+I+G model were investigated for all partitions in a Bayesian analysis (with independent parameter estimates for each partition). The result was an identical topology and branch support measures as seen in Fig. 3 (results not shown), suggesting that the results are not a product of model underparameterization.

Taxon sampling is also an important consideration when contrasting these results with previous studies because only four of the ten extant salamander families were included. Limited taxon sampling can have a substantial effect on phylogenetic error (Zwickl and Hillis, 2002). This may be especially so in data sets with large numbers of characters, such as from whole organellar genomes, from only a small number of representatives of the actual lineage diversity of a group (Soltis et al., 2004; Stefanovic, 2004). The old age and large genetic divergences among the four family lineages included in our analyses create a phylogeny that is exceptionally difficult to resolve, in which terminal branches are extremely long relative to internal branches. This scenario increases the probability of long-branch attraction when using highly homoplastic data. Inclusion of additional families or major lineages within families may act to subdivide long branches and diffuse homoplasy, thus minimizing the potential effects of long-branch attraction (Hillis, 1996). An example of the potential effect of taxon sampling in complete mtDNA genome studies is illustrated in birds where phylogenetic analysis of data for six avian orders placed Passeriformes in a basal position in bird phylogeny (Härlid and Arnason, 1999; Mindell et al., 1999), challenging their traditional and nuclear-based position as a recently derived lineage. Subsequent analyses using expanded sampling of CMGSs yielded trees consistent with traditional hypotheses (Braun and Kimball, 2002). Therefore, taxon sampling may be an especially important issue in phylogenetic studies of CMGS because limited resources typically lead to the generation of only one or a few genomes at a time; consequently, analyses of deep evolutionary relationships may be prone to producing inaccurate results as a byproduct of limited sampling (Soltis et al., 2004; Stefanovic, 2004). Recent studies of CMGS using much of the same data as used in these CMGS analyses along with diverse sampling of the family Plethodontidae yielded results consistent with previous studies of salamander family phylogeny (Mueller et al., 2004), further suggesting that taxon sampling is an important factor affecting these results. If substantial homoplasy and taxon sampling are affecting these results, they have produced strongly supported and concordant trees using different optimality criteria. These results offer a novel

phylogenetic hypothesis among salamander families. However, this hypothesis awaits further testing through the detailed analysis of additional salamander CMGSs and indicates that salamander family phylogeny may provide a good system to study the effects of taxon sampling in the recovery of old and divergent branching events.

This work demonstrates that CMGS information gathered from multiple closely related taxa can benefit population genetic and phylogeographic studies in at least two ways. First, it can provide a large number of characters for phylogenetic analysis of lineage representatives. This should be especially useful in studies involving species or populations that have recently diverged over a short period of time. However, the generation of complete mitochondrial genome sequences in studies involving hundreds of individuals is not practical. Therefore, complete mitochondrial genome sequence information from representative individuals can also be beneficial in the evaluation of molecular evolutionary patterns across the entire mitochondrial genome. Regions harboring the highest levels of variation can be identified and used in more extensive studies involving more detailed sampling.

APPENDIX I
PCR PRIMER LIST

Primer Name[a]	Primer Sequence	Position	Species[c]
Mito1_5.1	AAAAGGAACTCGGCAATCAAAGACT	209	*Ambystoma*
Mito1_3.1	ATCCCGGCTCAAGCACCAAATAC	2382	*Ambystoma*
Mito1_5.2	TGCTAAGCCACACCCACAA	1861	*Ambystoma*
Mito1_3.2	TAAACCTACTCGTCGATAAGAACTC	5383	*Ambystoma*
Mito1_5.3	GGTGCTTGAGCCGGGATAGTTGG	5367	*Ambystoma*
Mito1_3.3	CCCCCGGATTTAGGTTTTGTGTTGG	7386	*Ambystoma*
Mito1_5.4	ATTGCCCTCCCCCTTATCACAC	6790	*Ambystoma*
Mito1_3.4	GCAGCTGCTTCAAACCCAAAAT	9310	*Ambystoma*
Mito1_5.5	CCAGTTCAAAAAGGGTTACG	8814	*Ambystoma*
Mito1_3.5	AAAAATGCTAGTTGTGGTTGAT	12723	*Ambystoma*
Mito1_5.6	GGACTTAATCAACCACAAC	12695	*Ambystoma*
Mito1_3.6	GCATAAGCAAATAAGAAA	14964	*Ambystoma*
Mito1_5.7	AACATTTCAAGGCCATCATAC	8789	*Ambystoma*
Mito1_3.6	GCATAAGCAAATAAGAAA	14964	*Ambystoma*
Mito1_5.8	TATTATTCATVTTVTATTCCTTCAC	14707	*Ambystoma*
Mito1_3.8	ACTCTTTAGATTATTACTGGTTCAA	296	*Ambystoma*
Mito1_5.9	TTATACAAAATGATCGAAAAGAAGC	9046	*Ambystoma*
Mito1_3.9	TTAAAAATGGGAATAGGAAATGGA	14693	*Ambystoma*
Mito2_5.1	TTGGATCAGGACACCCAAAT	2250	*Ambystoma*
Mito2_3.1	AGGCTTAATGCAGTGCCAAC	5250	*Ambystoma*
Mito2_5.2	TTTGGGCACCCAGAGGTAT	6000	*Ambystoma*
Mito2_3.2	GGGTTCAATTCCCTCCTTTC	6750	*Ambystoma*
Mito2_5.3	CCAGGGCGACTAAATCAAAC	7500	*Ambystoma*
Mito2_3.3	TGTGATAAGCGTGAGCTTGG	8500	*Ambystoma*
Mito2_5.4	GCAGCTGCCTGATATTGACA	9500	*Ambystoma*
Mito2_3.4	TGAAAAGCCTCCTCAGATTCA	14600	*Ambystoma*
Mito2_5.4	GCAGCTGCCTGATATTGACA	9500	*Ambystoma*
Mito2_3.5	TAGAAGGGCCGATACTGGTG	12400	*Ambystoma*
Mito2_5.6	AATAATTCCTGCCACCACCA	12500	*Ambystoma*
Mito2_3.6	GAAGGCAAAGAATCGGGTTA	14750	*Ambystoma*
Mito2_5.7	AACCACCATTTTGGGTTTGA	9250	*Ambystoma*
Mito2_3.7	AAAGGGTCCGGCTATGAGTT	11150	*Ambystoma*
Mito2_5.8	ACTCATAGCCGGACCCTTTT	11150	*Ambystoma*
Mito2_3.8	TGGTGGGGTGATTAATGGAT	15000	*Ambystoma*

APPENDIX I, CONTINUED

Mito2_5.9	TTCCAGTTTTAGCAGCAGGA	5800	*Ambystoma*
Mito2_3.9	TCATTGATGTCCAATTGCTTT	7200	*Ambystoma*
Mito3_5.1	TTAACAACTTCGCCTAATATCTCAAAC	4900	*A. t. tigrinum*
Mito3_3.1	AAATTGCAAATTTTAAAGAACGACTTA	5180	*A. t. tigrinum*
Mito3_5.2	AACCCTATTATACGTTTAGCTTGAGGA	13500	*A. t. tigrinum*
Mito3_3.2	CGCCAATTCATGTTAGGATTATAGTA	14100	*A. t. tigrinum*
Mito3_5.3	TCTTCTATCCTTTATTTTTCACTTTTCA	15200	*A. mexicanum*
Mito3_3.3	TATTAGTAAATTTAGGGCATTTTCACC	70	*A. mexicanum*
Mito3_5.4	AGCGACAACATTATCAATCTTATTCTT	2700	*A. mexicanum*
Mito3_3.4	TTGCTAGCTAGTGTTAATATGGTTCCT	3880	*A. mexicanum*
Mito3_5.5	TCTCAATTTGACGACAAAATAATAAAA	4900	*A. mexicanum*
Mito3_3.5	GAAGTAGCTCCTCCATAATTGGTGA	6990	*A. mexicanum*
Mito3_5.6	GTTTGCCTATTTCGTCAAATTAACTAC	8590	*A. mexicanum*
Mito3_3.6	CAGCCATAAGTGTAAAAATAGTTGTGA	9860	*A. mexicanum*
Mito3_5.7	TAATTTCATCAGCCCTATTTTGTTTAG	11500	*A. mexicanum*
Mito3_3.7	TTTTGATACCTAAAACCAATGGATAAC	11600	*A. mexicanum*
Mito3_5.8	GCACTTCTAGTAACAATTATCGGACTT	13500	*A. mexicanum*
Mito3_3.8	CTACTTGATCGCTTTTTATTGAGCTA	13530	*A. mexicanum*
Mito3_5.9	TATAATCCTAACGTGAATTGGAGGA	16308	*A. mexicanum*
Mito3_3.9	AAAAATTTTCATTTAACCGCTCTTT	362	*A. mexicanum*
Mito3_5.10	GTCCTTTTGCCTTATTTTTCTTAGC	3370	*A. mexicanum*
Mito3_3.10	AAATATTTCGTAGCTGCTTCTGTTG	4040	*A. mexicanum*
Mito3_5.11	TCAATCTCTCTAACAACTTCCCCTA	4780	*A. mexicanum*
Mito3_3.11	AATATCATTGATGTCCAATTGCTTT	7264	*A. mexicanum*
Mito3_5.12	GCCTATTTCGTCAAATTAACTACCA	9247	*A. mexicanum*
Mito3_3.12	TGGAAGTGGAAAAATAGATGTTGAT	10028	*A. mexicanum*
Mito3_5.13	ATGATTTCACACGGACTAATTTCAT	11099	*A. mexicanum*
Mito3_3.13	GAGGAATACAAGGGAGTACTGGTCT	11995	*A. mexicanum*
Mito3_5.14	ATTATTAATTCCTCATCCCCCATAA	13135	*A. mexicanum*
Mito3_3.14	GATGAGAATGCTGATGATGTATCTG	14286	*A. mexicanum*
Mito3_5.41	GGAAGGAGGATTTAGCAGTAAAAAG	810	*A. mexicanum*
Mito3_3.41	GCTCTATTTTTAATTTCTTTCTCCAA	1153	*A. mexicanum*
Mito3_5.15	TTCGCCTAATATCTCAAACACTTTC	4821	*A. t. tigrinum*
Mito3_3.15	GGCTCAAGCACCAAATACTAAATAA	5423	*A. t. tigrinum*
Mito3_5.16	ATATTAAACCAACCCCACCATAATC	13639	*A. t. tigrinum*
Mito3_3.16	GGTTGGTAAATCAATAAACGAGTTG	14173	*A. t. tigrinum*

APPENDIX I, CONTINUED

Mito3_5.42	ATTATTATTCCAACAATCGGAATTT	15300	*A. t. tigrinum*
Mito3_3.42	GGATGGAAAAATTAGTTCATGTTCA	15800	*A. t. tigrinum*
Mito3_5.17	TGATCCCTCAGCCTCCTCTA	16027	*A. andersoni*
Mito3_3.17	AGGCTCCTCTAGGTGGGTGT	637	*A. andersoni*
Mito3_5.18	AACCCGTCTATGTGGCAAAA	1357	*A. andersoni*
Mito3_3.18	TCTTTGATTGCCGAGTTCCT	1938	*A. andersoni*
Mito3_5.19	CCGGAGTAATCCAGGTCAGT	2509	*A. andersoni*
Mito3_3.19	AGACCAATTGGGCCTACAAT	2857	*A. andersoni*
Mito3_5.20	AACCAATTCGCCCCTCTACT	2731	*A. andersoni*
Mito3_3.20	ATGGAACGGTGCAATTCCTA	4223	*A. andersoni*
Mito3_5.21	TAAAGATATTGGCGCCCTTT	5338	*A. andersoni*
Mito3_3.21	CGAATCCCCCGATTATAACA	5537	*A. andersoni*
Mito3_5.22	GCTTTTGGCTTCTTCCTCCT	5603	*A. andersoni*
Mito3_3.22	ACCAGTTGGGATGGCAATAA	6350	*A. andersoni*
Mito3_5.23	TGCCCTCCTCCTTATCACAC	6805	*A. andersoni*
Mito3_3.23	TCATTGATGTCCAATTGCTTT	7309	*A. andersoni*
Mito3_5.24	GATGCACAAGAAATTGGGATAG	7159	*A. andersoni*
Mito3_3.24	ACACCTGGTCGAGAAGCAAT	7561	*A. andersoni*
Mito3_5.25	TGACCAATTCATAAGCCCAAC	7932	*A. andersoni*
Mito3_3.25	TGCTGCAGTTGGTTGATTTC	8281	*A. andersoni*
Mito3_5.26	AGCCCATGACCACTTACAGG	8632	*A. andersoni*
Mito3_3.26	GTTGGAAGAAGGAGGGCAAT	9669	*A. andersoni*
Mito3_5.27	AGCGCAGGCCTAGCATTA	10003	*A. andersoni*
Mito3_3.27	GATGGTGGAAGCGCTATGTT	11265	*A. andersoni*
Mito3_5.28	GGTGCAACCCCAAGTAAAAGTA	11723	*A. andersoni*
Mito3_3.28	TTTGGTTGAGGATTTTTAGTAGCA	13626	*A. andersoni*
Mito3_5.29	TCATACGAAAAACACACCCTTTAAT	14218	*A. andersoni*
Mito3_3.29	AAGAGAAATATGGGTGGAATGAAAT	14914	*A. andersoni*
Mito3_5.30	ATACGGAATATTCACGCAAATG	14360	*A. dumerilii*
Mito3_3.30	CCTTCCGGTAAACTTACCATGT	983	*A. dumerilii*
Mito3_5.31	GAGCCATAGAGAGAGTACTGCAAA	1200	*A. dumerilii*
Mito3_3.31	TCAGTGAGTTTTTCGTTCTGATG	1468	*A. dumerilii*
Mito3_5.32	AACCCGCAATTAAACAAATCAC	3606	*A. dumerilii*
Mito3_3.32	AAAGGGTGCCAATATCTTTATGA	5436	*A. dumerilii*

APPENDIX I, CONTINUED

Mito3_5.33	CCTCCTTCGTTCCTCCTTCTAT	5632	*A. dumerilii*
Mito3_3.33	CAGCAGGATCAAAGAATGTTGT	5818	*A. dumerilii*
Mito3_5.34	GATGAAATTAATGACCCGCACT	7253	*A. dumerilii*
Mito3_3.34	TTGGAAGTGTCCTAATGGTGTG	7651	*A. dumerilii*
Mito3_5.35	CTTCACCAACAGATCGATGATT	8005	*A. dumerilii*
Mito3_3.35	AGACGAACGCCTAAAGCTAATG	8356	*A. dumerilii*
Mito3_5.36	CCAAGCTCACGCTTATCACATA	8601	*A. dumerilii*
Mito3_3.36	TGTCGGTGCTAGGCTAGAATTA	8918	*A. dumerilii*
Mito3_5.37	ACCGACACCAGAATTAGGAGAA	8914	*A. dumerilii*
Mito3_3.37	TCCTCATCAGTAGATTGAGACGTAA	9371	*A. dumerilii*
Mito3_5.38	CCCAGACACAGAAAAACTTTCA	9534	*A. dumerilii*
Mito3_3.38	AGCCTCCTTGGATTCATTCATA	9776	*A. dumerilii*
Mito3_5.39	AATACTTCACCGAACCCATTTTT	9916)	*A. dumerilii*
Mito3_3.39	GTTTGAATAATTGCGGCTGAA	11058	*A. dumerilii*
Mito3_5.40	CTCTTACATTACTTGCAACCTCCAT	12824	*A. dumerilii*
Mito3_3.40	ATACGGTGAATTACTGTGGGATAGA	13455	*A. dumerilii*
Mito3_5.43	GAAAATACTTGATAAAACCAAGTTGA	1029	*A. dumerilii*
Mito3_3.43	TTGGATTTAAGTTCATTTCTTGAGC	1442	*A. dumerilii*
Mito3_5.44	GTCATGATTAAAGAACCCGGTAATA	4578	*A. dumerilii*
Mito3_3.44	AGGCTTTGAAGGTCTTTAGTCTGAT	4960	*A. dumerilii*
Mito3_5.45	CTCGATGACTATTTTCTACAAATCATA	5311	*A. dumerilii*
Mito3_3.45	ATTATTACAAATGCGTGTGCTGTTA	5497	*A. dumerilii*
Mito3_5.46	TGCTATAGAAGGACCAACTCCTGTA	12448	*A. andersoni*
Mito3_3.46	ATCTTTAGAGAAAAATCCGGCTAAA	12917	*A. andersoni*
Mito3_5.47	CATTAATTACCCCACCACACATT	14910	*A. dumerilii*
Mito3_3.47	GTTTTCGACTTACAAGGTCGATGT	15310	*A. dumerilii*
Mito3_5.48	AATCCTAGCAGCAGTTCTTCTTAAA	10838	*A. dumerilii*
Mito3_3.48	AAAATTACTGCTCCTGTAAGGCTTC	11095	*A. dumerilii*
Mito3_5.49	TTAAGTAGTGGTGAAAAGCCTAACG	1376	*A. t. tigrinum*
Mito3_3.49	TTAAAAGACAGGTGATTACGCTACC	2011	*A. t. tigrinum*
Mito3_5.50	ATTAATACTTCACCGAACCCATTTT	9915	*A. t. tigrinum*
Mito3_3.50	CACATCCCTATTAGTGGAAGAATTG	11224	*A. t. tigrinum*
Mito3_5.51	GGACCAAAATATATAACAGCCCAAT	13415	*A. t. tigrinum*
Mito3_3.51	TAGATTCCATCGGATATGATTATGG	13688	*A. t. tigrinum*
Mito3_5.52	CCGAGATGATATTTTTAGAAGGACA	2239	*A. californiense*
Mito3_3.52	CGTTCTACAAGGGTTAAAAATGCTA	2762	*A. californiense*

APPENDIX I, CONTINUED

Tgmito_5.1	AAAAACGAAAAACCAAATGAAAA	700	*A. t. tigrinum*
Tgmito_3.1	TCCTTTGCAGTACTCTCTCTATGG	1200	*A. t. tigrinum*
Tgmito_5.2	GCACAAACAATCTCATATGAAGTTACA	3330	*A. t. tigrinum*
Tgmito_3.2	TGTTTAATTGCGGGTTAATTTG	4400	*A. t. tigrinum*
Tgmito_5.3	TTGCTATCACAACAATTATTGCAC	4800	*A. t. tigrinum*
Tgmito_3.3	TCGAGTAATTATCACAGGTAAAATGG	5300	*A. t. tigrinum*
Tgmito_5.4	GCTGGCCATCTTCTTATTCAA	7770	*A. t. tigrinum*
Tgmito_3.4	TGGCCTTGAAATGTTCCTTC	8600	*A. t. tigrinum*
Tgmito_5.5	GAAATTGCCCTCCTTCTTCC	9675	*A. t. tigrinum*
Tgmito_3.5	CGTGTTGCTGCTACTATTAATGC	10040	*A. t. tigrinum*
Tgmito_5.6	TCATATAGGACTTGTCATTTCAGC	11130	*A. t. tigrinum*
Tgmito_3.6	TCAGTTAAAATTAGTTGTTGTTGATTC	12150	*A. t. tigrinum*
Tgmito_5.7	GCTGCAACAGGAAAATCTGC	12510	*A. t. tigrinum*
Tgmito_3.7	AAAGAATGAAGCGCCATTTG	14130	*A. t. tigrinum*

[a] Primer 5.N is forward, primer 3.N is reverse.

[b] Refers to approximate nucleotide position of primer on completed mitochondrial genome of corresponding species. Position numbers for primers designed from ESTs (*Ambystoma*) correspond to homologous base positions in *L. atifi.*

[c] Species from which primer was designed. "*Ambystoma*" refers to a combination of *A. mexicanum* and *A. t. tigrinum* EST contigs.

APPENDIX II

NUCLEOTIDE SUBSTITUTION MODELS AND PARAMETER ESTIMATES FOR *AMBYSTOMA* MITOCHONDRIAL GENE PARTITIONS

Gene	Model[a]	Ti:Tv[b]	C↔T[c]	C↔G	A↔T	A↔G	A↔C	I[d]	α[e]
tRNAs	GTR+I		19.16 (2.8-76.71)	1.91 (0.01-11.91)	2.82 (0.31-11.13)	16.1 (2.34-63.02)	2.44 (0.17-11.53)	0.72 (0.21-0.91)	Equal Rates
12S	GTR+I	-	45.51 (7.26-96.6)	0.61 (0.01-4.27)	21.5 (2.96-57.04)	24.08 (3.19-64.05)	0.89 (0.01-6.24)	0.72 (0.17-0.91)	Equal Rates
16S	GTR+I	-	22.86 (4.91- 72.62)	0.23 (0.01-1.35)	4.65 (0.95-15.27)	7.56 (1.67-26.66)	0.74 (0.01-3.36)	0.84 (0.72-0.9)	Equal Rates
atp6	HKY+G	9.47 (5.44-16.25)	-	-	-	-	-	-	1.39 (0.05-18.61)
atp8	GTR+I	-	43.64 (5.5-94.97)	6.11 (0.01-45.59)	0.84 (0.01-3.99)	6.64 (0.07-35.83)	2.32 (0.01-14.39)	0.84 (0.71-0.91	Equal Rates
cox1	GTR+G	-	53.35 (15.77-97.46)	10.41 (2.24-25.43)	5.63 (1.2-11.93)	48.32 (13.93-90.86)	3.21 (0.31-9.04)	-	0.15 (0.05-0.34)
cox2	GTR+I	-	6.43 (2.15-17.46)	0.53 (0.01-2.19)	0.32 (0.02-1.1)	2.98 (1.03-8.32)	0.42 (0.05-1.47)	0.68 (0.36-0.82)	Equal Rates
cox3	GTR+I	-	57.7 (17.03-97.52)	0.34 (0.01-2.71)	2.61 (0.01-9.25)	34.69 (9.53-77.91)	3.78 (0.02-12.36)	0.83 (0.69-0.089)	Equal Rates
cytb	GTR+I	-	52.6 (16.38-95.63)	0.58 (0.01-3.45)	0.63 (0.01-2.77)	44.31 (13.6-90.28)	5.85 (1.26-13.95)	0.80 (0.75-0.84)	Equal Rates
nd1	GTR+I	-	30.44 (6.05-81.71)	1.25 (0.01-7.48)	3.46 (0.52-9.68)	32.09 (6.68-80.05)	3.43 (0.21-11.54)	0.75 (0.67-0.81)	Equal Rates
nd2	GTR+I	-	22.05 (3.67-60.42)	0.71 (0.01-5.29)	1.83 (0.12-6.04)	50.9 (11.64-96.2)	2.57 (0.24-8.26)	0.76 (0.68-0.82)	Equal Rates
nd3	GTR+I	-	8.46 (1.54-31.12)	1.36 (0.02-6.7)	0.68 (0.01-3.21)	36.0 (6.79-91.57)	0.24 (0.01-1.41)	0.74 (0.52-0.84)	Equal Rates
nd4	GTR+I	-	31.01 (7.67-78.84)	2.69 (0.13-10.21)	2.65 (0.53-7.67)	36.94 (9.67-87.99)	3.57 (0.68-10.15)	0.64 (0.44-0.75)	Equal Rates
nd4L	GTR+I	-	24.0 (1.39-86.36)	1.01 (0.01-8.58)	0.41 (0.01-2.92)	33.16 (0.88-94.52)	0.24 (0.01-1.55)	0.91 (0.85-0.95)	Equal Rates

APPENDIX II, CONTINUED

nd5	GTR+I	-	33.07 (7.64-86.89)	2.74 (0.02-11.97)	2.04 (0.37-6.02)	28.0 (6.65-82.76)	3.2 (0.66-9.66)	0.75 (0.69-0.79)	Equal Rates
nd6	GTR+I	-	14.93 (1.76-67.99)	0.36 (0.01-2.95)	0.83 (0.01-5.24)	12.97 (2.15-61.21)	0.24 (0.01-1.37)	0.79 (0.72-0.85)	Equal Rates
D-loop	GTR+I	-	22.78 (2.29-77.12)	0.56 (0.01-4.02)	3.42 (0.03-15.66)	49.54 (9.27-97.40)	2.09 (0.02-10.26)	0.90 (0.81-0.94)	Equal Rates
Thr-Pro Spacer	GTR	-	22.28 (2.58-66.82)	21.84 (1.91-73.91)	7.72 (0.73-26.32)	37.6 (5.01-93.72)	2.88 (0.04-12.54)	-	Equal Rates

[a] Best-fit model determined from an alignment of five *Ambystoma* complete mitochondrial genomes using Akaike-Information criteria implemented in the program Modeltest (Posada and Crandall, 1998).

[b] Mean Bayesian estimate and 95% credibility interval (parentheses) for the Transition (Ti):Transversion (Tv) ratio for genes that fit a Hasegawa-Kishino-Yano (HKY) model of evolution (Hasegawa et al., 1985).

[c] Mean Bayesian estimate and 95% credibility interval (parentheses) for the six classes of substitutions of the General Time-reversible (GTR) model of evolution (Yang, 1994). All rates are standardized against a G↔T substitution rate of one.

[d] Mean Bayesian estimate and 95% credibility interval (parentheses) for the proportion of invariant sites (I).

[e] Mean Bayesian estimate and 95% credibility interval (parentheses) for the shape parameter (α) describing the gamma distribution of rate heterogeneity.

REFERENCES

The data in this thesis are reprinted from: Gene, Vol 349, Samuels, A. K, D. W. Weisrock, J. J. Smith, K. J. France, J. A. Walker, S. Putta, and S. R. Voss, Transcriptional and phylogenetic analysis of five complete ambystomatid salamander mitochondrial genomes, Pages 43-53, Copyright 2005, with permission from Elsevier.

Arnason U., A. Gullberg, A. Janke, J. Joss, and C. Elmerot, 2004. Mitogenomic analyses of deep gnathostome divergences: a fish is a fish. Gene. 333, 61-70.

Avise J. C, 2004. Molecular markers, natural history, and evolution. Sinauer Associates, Sunderland, Mass.

Barth C., U. Greferath, M. Kotsifas, Y. Tanaka, S. Alexander, H. Alexander, and P. R. Fisher, 2001. Transcript mapping and processing of mitochondrial RNA in *Dictyostelium discoideum*. Curr. Genet. 39, 355-364.

Bensasson D., D. X. Zhang, D. L. Hartl, and G. M. Hewitt, 2001a. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16, 314-321.

Boore J. L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767-1780.

Braun E. L., and R. T. Kimball, 2002. Examining basal avian divergences with mitochondrial sequences: model complexity, taxon sampling, and sequence length. Syst. Biol. 51, 614-625.

Buckley T. R., 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51, 509-523.

Edqvist J., G. Burger, and M. Gray, 2000. Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis*. J. Mol. Biol. 297, 381-393.

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

Gao K. Q., and N. H. Shubin, 2003. Earliest known crown-group salamanders. Nature 422, 424-428.

Gissi C. and G. Pesole, 2003. Transcript mapping and genome annotation of Ascidian mtDNA using EST data. Genome Res. 13, 2203-2212.

Härlid A. and U. Arnason, 1999. Analyses of mitochondrial DNA nest ratite birds within the Neognathae: supporting a neotenous origin of ratite morphological characters. Proc. R. Soc. Lond. B 266, 305-309.

Hillis D. M., 1996. Inferring complex phylogenies. Nature 383, 130-131.

Huelsenbeck J. P., and F. Ronquist, 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.

Jeanmougin F., J. D. Thompson, M. Gouy, D. G. Higgins. and T. J. Gibson, 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. 23, 403-405.

Kumazawa Y., and M. Nishida, 1993. Sequence Evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J. Mol. Evol. 37, 380-398.

Larson A., D. W. Weisrock, and K. H. Kozak, 2003. Phylogenetic systematics of salamanders (Amphibia: Urodela), a review. Pp. 31-108 *In* D. M. Server and B. G. M. Jamieson, eds. Reproductive biology and phylogeny of Urodela. Science Publishers, Enfield, New Hampshire.

Larsson N. G., D. A. Clayton, 1995. Molecular genetic aspects of human mitochondrial disorders. Annu. Rev. Genet. 29, 151-178.

Lodish H., A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, J. E. Darnell, 2000. Molecular cell biology. W. H. Freeman and Company, New York.

Mindell D. P., M. D. Sorenson, D. E. Dimcheff, M. Hasegawa, J. C. Ast, and T. Yuri, 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst. Biol. 48, 138-152.

Mueller R. L., J. R. Macey, M. Jaekel, D. B. Wake, and J. L. Boore, 2004. Morphological homoplasy, life history evolution, and historical biogeography of plethodontid salamanders inferred from complete mitochondrial genomes. Proc Natl. Acad. Sci. USA. 101, 13820-13825.

Nardi F., A. Carapelli, R. Dallai, F. and Frati, 2003. The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographical locations. Insect Mol. Biol. 12, 605-611.

Nylander J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey, 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47-67.

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

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

Sessions S. K. and A. Larson, 1987. Developmental correlates of genome size in plethodontid salamanders and their implications for genome evolution. Evolution. 41, 1239-1251.

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

Shimodaira H., and M. Hasegawa, 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114-1116.

Soltis D. E., V. A. Albert, V. Savolainen, K. Hilu, Y. L. Qiu, M. W. Chase, J. S. Farris, S. Stefanovic, D. W. Rice, J. D. Palmer, and P. S. Soltis, 2004. Genome-scale data, angiosperm relationships, and 'ending incongruence': a cautionary tale in phylogenetics. Trends Plant Sci. 9, 477-483.

Stefanovic S., D. W. Rice, and J. D. Palmer, 2004. Long branch attraction, taxon sampling, and the earliest angiosperms: Amborella or monocots? BMC Evol. Biol. 4, 35

Swofford D. L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4. Sinauer Associates, Sunderland, Mass.

Woischnik M. and C. T. Moraes, 2002. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res. 12, 885-893.

Yang Z., 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11, 367-372.

Zwickl D. J. and D. M. Hillis, 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51, 588-598.

VITA

Amy Kathleen Samuels

DATE OF BIRTH

January 26, 1980

PLACE OF BIRTH

Wilkes-Barre, Pennsylvania

EDUCATION

Bachelor of Arts, Biology, cum laude, 2002
Hartwick College, Oneonta, New York

PROFESSIONAL POSITIONS HELD

Research Assistant, Summer 2001, Summer 2002
Dept. of Biology, Hartwick College
Teaching Assistant, Fall 2002, Fall 2005
Dept. of Biology, University of Kentucky
Research Assistant, Spring 2003-Spring 2005
Dept. of Biology, University of Kentucky

SCHOLASTIC AND PROFESSIONAL HONORS

Dean's List, Fall 1998-Spring 2002
Hartwick College
Deborah M. Allen Brennan Scholarship Award, Spring 2001
Hartwick College
Highest GPA of Freshman Biology Major Award, Spring 1999
Hartwick College

PROFESSIONAL PUBLICATIONS

Smith J.J., Putta S., Walker J.A., Kump D.K., **Samuels A. K.**, Monaghan J.R., Weisrock D. W., Staben C., and Voss S.R. 2005. Sal-Site: Integrating new and existing Ambystomatid salamander research resources. BMC Genomics. In Press.

Samuels A.K., Weisrock D.W., Smith J.J., France K., Walker J.A., Putta S., and Voss S.R. 2005. Transcriptional and phylogenetic analysis of five complete ambystomatid salamander mitochondrial genomes. Gene 349:43-53

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., and Voss S. R. 2004 From biomedicine to natural history research: EST resources for ambystomatid salamanders. BMC Genomics 5(1):54