GA

GENOMIC PERSPECTIVES ON THE AMPHIBIAN TREE OF LIFE

Paul Hime, University of Kentucky
2016–2017 Graduate Fellow

EXECUTIVE SUMMARY

The relationships among extant amphibians (frogs, salamanders, and caecilians) has been a longstanding debate in phylogenetics. Previous studies have supported any of the three possible topologies relating these three extant orders of amphibians. This project developed a novel amphibian-specific gene-capture system to target and sequence 220 nuclear genes in a diverse set of 296 amphibian species representing all major amphibian lineages. Using an information-theoretic approach to compare the relative model support for interordinal amphibian relationships and to calculate the magnitude of support for these relationships across genes, this project has demonstrated substantial variation in how strongly different genes support any interordinal topology. A few genes supporting each of the three interordinal models resoundingly reject the alternative models (high ΔAIC values in Fig. 1), while most other genes only weakly support any model for relationships among the three amphibian models, suggesting an erosion of phylogenetic signal over these deep evolutionary timescales.

Phylogenetic information content in many genes in the genome appears to have eroded in the nearly 300 million years since the three amphibian orders diverged, yet our results support a model of frogs and salamanders sharing a common ancestor despite substantial variation in phylogenetic signal across different genes. Our results also suggest a revised hypothesis for the relationships among the extant families of amphibians and support a more recent origin for many of the hyperdiverse lineages of frogs. Overall, this project highlights the power of phylogenomics and
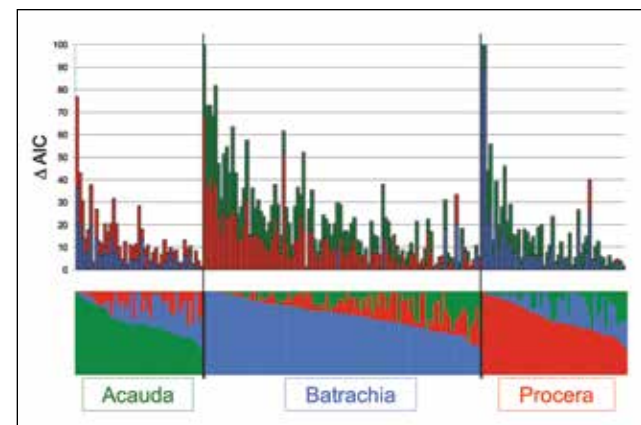

Figure 1: Support for competing models of interordinal amphibian relationships—194 genes are binned on the horizontal axis by support for frogs+caecilians (Acauda hypothesis in green), frogs+salamanders (Batrachia hypothesis in blue), or caecilians+salamanders (Procera hypothesis in red). The upper panel depicts the magnitude of support *against* competing hypotheses measured by the AIC.

a model-based theoretical framework for testing phylogenetic hypotheses in the era of genome-scale evolutionary biology.

RESEARCH CHALLENGE

All organisms trace their ancestry back to a single common ancestor nearly four billion years ago [1]. Yet today, life has diversified into tens of millions of species. Reconstructing these evolutionary relationships is a key aim of the field of phylogenetics, and such insights may inform nearly all aspects of modern biology. The recent advent of genome sequencing technologies has ushered in a new era of phylogenomics in which genetic data generation is no longer a rate-limiting step toward testing phylogenetic hypotheses. Yet, it is becoming increasingly clear that different regions of the genome can support conflicting phylogenetic hypotheses, and reconciling these discordant gene genealogies is a key problem facing evolutionary biologists today. Information-theoretic approaches have great promise for parsing signal from noise in large phylogenomic data sets.

METHODS & CODES

Three possible topologies exist for relationships among frogs, caecilians, and salamanders. The Procera hypothesis supports caecilians+salamanders, the Acauda hypothesis supports frogs+caecilians, and the (canonical) Batrachia hypothesis supports frogs+salamanders. To evaluate support for each of these three hypotheses across the genome, we conducted gene-by-gene tests of constrained topology, comparing the maximum likelihood estimates of gene trees for 194 nuclear genes between constraints for the three possible interordinal topologies. Gene trees were estimated in RAxML [2]. The Akaike information criterion (AIC) [3] was used to quantify the direction and magnitude of support across genes. Gene trees were then reconciled into an estimate of the species tree using Astral [4]. This species tree topology was then used in concert with a set of 25 fossil calibrations to estimate divergence times across Amphibia in MCMCTree [5].

RESULTS & IMPACT

This project has demonstrated that different genes in the amphibian genome support conflicting topologies for the relationships among the three amphibian orders. Roughly half of the genes examined support frogs and salamanders as each other's closest relatives (sister taxa). Roughly one-quarter of the genes support frogs and caecilians as sister taxa, while another quarter support caecilians and salamanders as sister taxa. Significant numbers of genes are found to support each of these three competing topologies, suggesting that either there has been an erosion of phylogenetic signal over deep time and/or
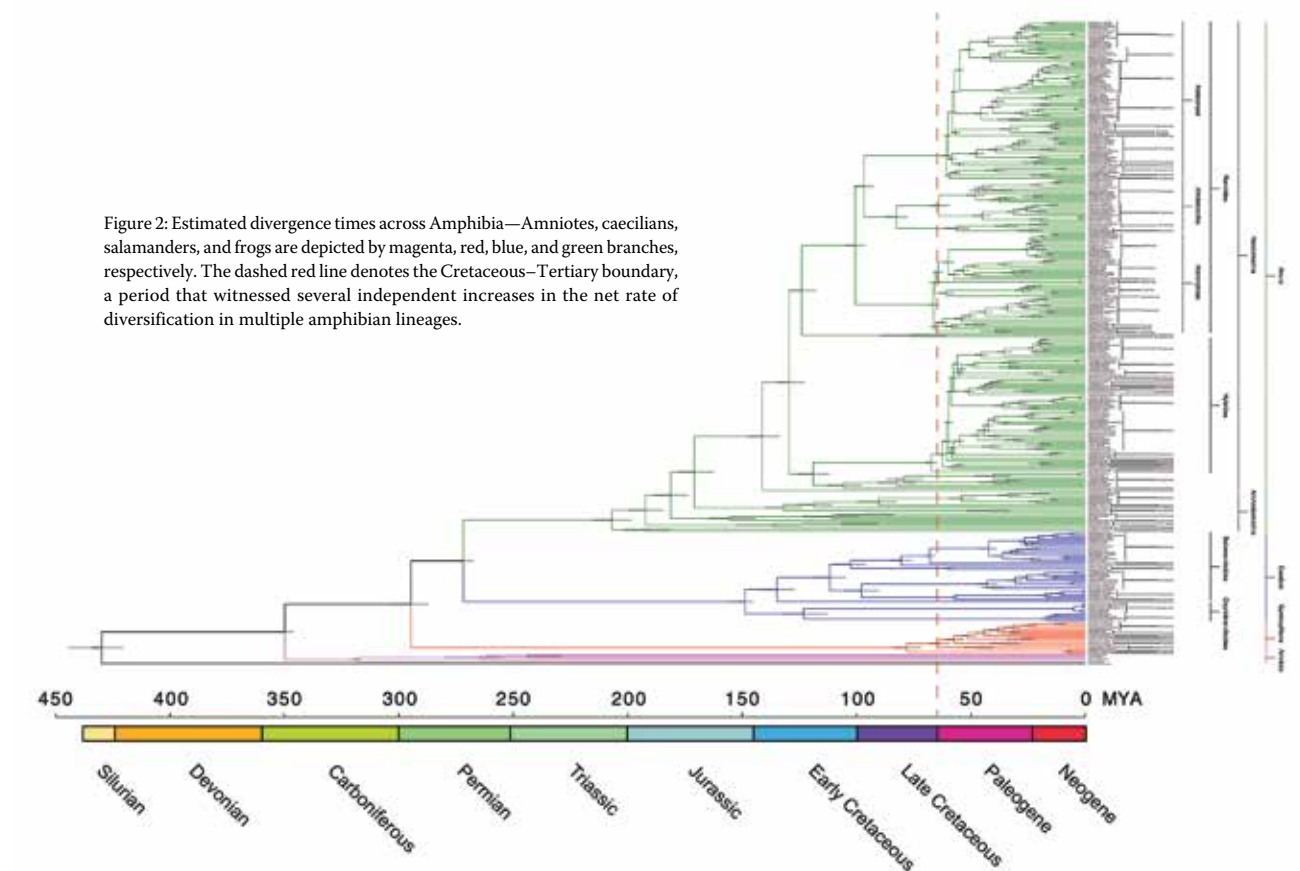

Figure 2: Estimated divergence times across Amphibia—Amniotes, caecilians, salamanders, and frogs are depicted by magenta, red, blue, and green branches, respectively. The dashed red line denotes the Cretaceous–Tertiary boundary, a period that witnessed several independent increases in the net rate of diversification in multiple amphibian lineages.

that population-level processes (large effective population sizes and rapid divergences) have led to incomplete lineage sorting and gene tree, or species tree discordance. Results from our AIC-based topology testing support the Batrachia hypothesis (frogs and salamanders are each other's closest relatives) at the level of the species tree and suggest that large numbers of genes may need to be analyzed in order to overcome stochastic phylogenetic noise at deep timescales. Our divergence time analyses also provide a revised timescale for amphibian diversification through time and indicate that a rapid increase in net rates of species diversification occurred at the Cretaceous–Tertiary boundary. Analyzing hundreds of genes for hundreds of species has traditionally been computationally intractable for empirical data sets, and this study is one of the first to use an information-theoretic framework to address not only the direction of support for phylogenetic hypotheses across the genome but also the magnitude of that support.

WHY BLUE WATERS

The Blue Waters fellowship has provided essential support to enable me to make progress on this project and to successfully write and defend my Ph.D. dissertation. Although many of the analyses I intended to execute on Blue Waters have been delayed due to unforeseen issues with our data set, I am hopeful that with a time extension to this allocation, I will be able to leverage the massively parallel resources on Blue Waters to perform topological testing in Neobatrachian frogs at a level of gene- and species-sampling that has not been attempted in any other empirical data set.

Paul Hime—a sixth-year Ph.D. student in evolutionary biology at the University of Kentucky—successfully defended his dissertation in April 2017. His work was directed by David W. Weisrock.