Which Came First: The Lizard or the Egg?

April Wright1, Kathleen Lyons1, Matthew Brandley2, and David M. Hillis1

1: Department of Integrative Biology, University of Texas, Austin, TX 78712 USA

2:

Phone: 512-471-5792

E-mail:

Abstract

Changes in parity mode between egg-laying (oviparity) and live-bearing (viviparity) have occurred repeatedly throughout vertebrate evolution. Oviparity is the ancestral amniote state, and viviparity is thought to have evolved many times independently within amniotes (especially in lizards and snakes), with relatively rare reversions to oviparity. In amniotes, the shelled amniote egg is considered a complex structure that is unlikely to re-evolve if lost (i.e., it is an example of Dollo's Principle). A recent analysis, however, suggested that viviparity was the ancestral state of squamate reptiles (lizards and snakes), and that oviparity evolved from viviparity many times throughout the evolutionary history of squamates. Here, we evaluate whether this conclusion is robust to the assumptions of the analysis. In contrast to the earlier analysis, we find strong support for the traditional view of parity evolution in squamates (ancestral oviparity, many independent origins of viviparity, and rare reversions to oviparity) under reasonable assumptions. Our analysis emphasizes that the conclusions of ancestral state reconstruction studies are likely to be highly sensitive to the methods and assumptions of analysis.

Introduction

Non-biologists may still ponder the age-old riddle of “Which came first, the chicken or the egg?,” but the answer is clear from an evolutionary standpoint (Fig. 1). The amniote egg, with its water-retaining shell, four extraembryonic membranes, and embryo-nourishing yolk, had evolved by the time the earliest amniotes diverged, 306.1 ± 8.5 mya in the Carboniferous (Shedlock and Edwards, 2009). That date is well over 300 million years before the first chicken walked the Earth. Today, most living amniotes (including chickens) still produce shelled, externally developing eggs (i.e., they are oviparous). Nonetheless, there have been multiple independent origins of viviparity (in which the eggs are retained in the female, receiving nourishment and/or parental protection, until birth) among the amniotes. Oviparity is retained in basal lineages of mammals (Prototheria), as well as in the vast majority of reptiles (including all turtles, crocodilians, birds, tuataras, and many but not all lizards and snakes; Fig. 1).

The traditional view of parity evolution in amniotes (Fig. 1) holds that what is true for the chicken-and-egg question above also is true for squamates (lizards and snakes). The last common ancestor of squamates lived over 200 mya in the Triassic (Hedges and Vidal, 2009), and most biologists have argued that this first squamate was almost certainly oviparous, even though viviparity evolved many times independently in various lizard and snake lineages (Lee and Shine, 1998; Sites et al., 2011; Blackburn 2014; Stewart and Blackburn 2014). In addition, there is reasonably strong evidence that oviparity has re-evolved within a few otherwise viviparous groups of squamates (Lynch and Wagner, 2009; Fenwick et al., 2012). The re-evolution of a relative complex character (a shelled egg that supports external development) is of great interest from a biological standpoint, and has been viewed as a counter-example to Dollo’s Principle (sometimes misleadingly called “Dollo’s Law”): the proposition that complex characters, once lost from a lineage, are unlikely to be regained (Dollo, 1893).

The traditional view of parity evolution in squamates was challenged by de Fraipont et al. (1996), who estimated that the rate of transition from oviparity to viviparity has been similar to rate of transition from viviparity to oviparity in squamates. Shine and Lee (1999) strongly criticized de Fraipont et al. (1996), especially in relation to how the study incorporated phylogenetic information (but see de Fraipont 1999). Goldberg and Igić (2008) argued that existing ancestral state reconstructions methods were statistically inappropriate and would lead to erroneous support for re-evolution of complex traits. Goldberg and Igić (2008) solved this problem by developing a model that incorporated the possibility of state-dependent diversification. Taking heed of the criticisms of Goldberg and Igić (2008), later studies used phylogenetic comparative methods that modeled transition rates between parity modes, as well as parity-dependent diversification rates. Using both parsimony and likelihood-based ancestral state reconstruction methods, Lynch and Wagner (2009) made a strong case that oviparity re-evolved from a viviparous ancestor in the sand boa, *Eryx jayakari* (see Griffith et al., this issue), and that viviparous taxa experience a higher speciation rate. Using similar methods, Fenwick et al. (2012) also found support for a reversion to oviparity from viviparity in vipers. Finally, a study of the relationship between climate and parity mode in *Liolaemus* lizards found that transitions from oviparity to viviparity were far more likely than reversals, but the authors could not rule out the latter (Pincheira-Donoso et al. 2013). Thus, by 2013, the prevailing view of parity mode evolution in squamates largely supported the traditional view: once viviparity evolves in a lineage, reversals to oviparity occur only rarely.

More recently, Pyron and Burbrink (2014) again challenged the traditional view of parity evolution among squamates, and argued that the ancestral squamate was viviparous. They argued that oviparity has re-evolved from viviparous ancestors many times among the squamates, seemingly in strong opposition to Dollo’s Principle. We undertook the present analysis to reexamine this conclusion, and to ask if this conclusion is robust to the assumptions of the analysis.

**Methods**

The Squamate Tree

1) The tree as published in 2013, and as compared to previous estimates and the morphological tree. Matt, this seems like it's up your alley?

2) Re-estimating the tree. What did we do? Matt: can you briefly discuss what you did in ExaML? I'll take over after that and discuss further optimizations performed.

There are known comparability issues between likelihood calculations for different versions of RaxML softwares (Stamatakis 2008). In order to compare the likelihood scores our trees with previously estimated trees, we calculated the exact likelihood scores of our data given the Pyron and Burbink tree and given our sample of trees using the software Garli (Zwickl 2006). These calculations revealed large discrepancies between reported scores from the ExaML package, the Pyron, Wiens and Burbrink (2013) tree and the Garli score. All trees scored worse with an exact calculation, indicating the trees discovered were worse than reported.

Using the ten best trees discovered by ExaML and the tree reported by Pyron, Wiens and Burbrink (2013), we performed two additional optimizations. One was an optimization of branch lengths only. The other was an co-optimization of topology and branch length. Both were performed in Garli (Zwickl 2006). The best topology estimated had a log likelihood score of -2592702.947982. The tree reported by Pyron, Wiens and Burbink (2013), which was used in Pyron and Burbink's 2014 analysis had a score of −2609551.07. This tree represents an 83,796 log likelihood unit improvement from the previously-reported tree. Fig. 2 shows a schematic of this tree with major clades colored. The distribution of scores can be seen in Figure [histogram].

The branch length and topology optimized trees had a smaller average Robinson-Foulds difference among themselves than did the branch-length optimized trees (865 RF units vs. 935 units). They also exhibited an overall shorter tree length, averaging 214 substitutions per site across the whole tree, compared to 344 substitutions/site across the tree.

The dramatic differences among analyses suggests several important points for practitioners of large-tree phylogenetic inference. On Fig. [histogram], trees 1, 2, 11, 13-16, 18 and 19 are all topology and branch-length optimized trees. The remainder are only branch-length optimized. The intercalation of the trees resulting from these two optimizations coupled with the branch length differences between them suggests that there are two regions of solutionspace that are similar in likelihood score, but differing predominantly in branch length. It should be noted, however, that 'similarity' in likelihood score in this context is a very coarse-grained metric. The trees recovered are thousands of log likelihood units from the optimal tree, and many trees are hundreds of likelihood units from their next best solution. Despite the additional optimizations we performed, these score differences and a lack of topological convergence, suggests treespace for this problem is still relatively poorly explored, and that there are more optimal trees that have not been discovered in our searches.

As these differences in scores and tree lengths are not trivial, we have several suggestions. Firstly, completing multiple independent tree searches will yield a sample of trees that can be used for comparison purposes. Secondly, for large-tree problems, additional optimizations can yield large improvements in likelihood scores over the first pass of estimation. To tease apart the relative contributions of branch lengths and topology to macroevolutionary questions, we suggest doing two rounds of optimizations: one in which topology is fixed and only branch lengths are optimized and one in which topology and branch lengths are jointly optimized. Lastly, all these comparisons are impossible without programmatic comparisons of trees. Full scripts used in analysis are provided in the supplemental materials.

**BiSSE Models and the sampling of lineages**

Pyron and Burbrink (2014) estimated the ancestral state of squamate parity mode using a framework called BiSSE. BiSSE was first proposed in 2007 by Maddison, Midford and Otto as a way of estimating macroevolutionary rates (such as speciation and extinction) in a trait-dependent framework. This model involves six total parameters: a speciation rate for taxa possessing character state 0 and one for those possessing state 1, an extinction rate for taxa possessing character state 0 and one for those possessing state 1, a rate of change from state 0 to state 1, and one from state 1 to state 0. Originally proposed for fully-resolved trees, this model was extended in 2009 by FitzJohn et al. to incompletely-resolve trees. The implementation of the model used by Pyron and Burbrink (2014), Diversitree (FitzJohn 2012), outputs ancestral states calculated in the process of estimating values for the six parameters. These states undergo no further optimization, as they are calculated as a model convenience. That the ancestral states are neither optimized nor an intended output of the model casts doubt on the propriety of this particular use of the Diversitree software.

Diversitree incorporates a sampling parameter. This parameter allows users to inform the analysis of what proportion of the descendants of the root of the tree have been sampled. If not manipulated by the user, the analysis assumes that all possible lineages have been sampled. Users can alter this parameter using either one or two terms. A one-term setting tells the analysis what proportion of the lineages in the tree are sampled, irrespective of character state. The two-term setting tells the analysis what proportion of taxa with state 0 are sampled, and what proportion of taxa with state 1 are sampled. In their 2014 analysis, Pyron and Burbrink used a two-term sampling parameter, specifying that 43% of oviparous species and 67% of viviparous species had been sampled.

In the case of the squamate tree, all lineages known to contain changes between states are sampled, casting doubt of specifying a one- or two-term sampling parameter. In order to assess the impact this parameter has on the results of an analysis, we performed two tests.

The first test was to use three parameter settings to estimate a BiSSE model on each tree from our samples. We estimated the model using the default setting (complete sampling), a one-term model (sampling frequency = .44, the proportion of squamates sampled) and Pyron and Burbrink's two-term model. We then estimated the number of reversals to oviparity seen on each tree. To calculate the reversals, we used a very liberal cutoff (i.e. the number of estimated reversals on calculated are a maximum, with the true number likely being less). We counted a reversal to oviparity as any time a child node had more than 50% support as oviparous, but the parent had more than 50% support as viviparous. As can be seen on Fig. [histogram], the number of reversals estimated on the tree varies across trees. But the number also varies with the model estimated. For trees 1,2,11, 13-16, 18 and 19 (the short-tree solutions), using a two-term model strongly increases the number of reversals observed. For trees 3-10, 12, 17 (the longer-tree solutions), this parameter has little effect on the number of reversals estimated. This is also true on the Pyron and Burbrink original tree (tree 20).

To get a better handle on the effect of the two-term sampling parameter on the estimated state of parity at the root, we estimated BiSSE models with incrementally-changing sampling parameters along three different trees: the original Pyron, Wiens and Burbrink (2013) tree, the best tree resulting from the ExaML estimation and the previous tree, after it had been optimized for branch lengths and topology in Garli. As shown in Fig. [heatmap], there is a strong effect of both topology and parameter on estimated root state. On the Pyron, Wiens and Burbrink tree, there is a large region of parameter space in which, for any combination of proportion of state 0 and state 1 taxa sampled, the result will strongly support viviparity at the root. This region of space is smaller on the ExaML tree, and very small on the optimized tree. The sampling parameter chosen by Pyron and Burbink is indicated by a yellow star, and falls inside the region of space in which viviparity is favored on two of the three plots, indicating that use of this parameter will result in viviparity being recovered as the ancestral state of squamates for those trees.

The sensitivity of BiSSE analyses to this parameter is troubling. The parameter is doing something slightly unintuitive: one might expect that a parameter that codes what proportion of taxa sampled were from a given character state might increase the likelihood of a given ancestor of being in the undersampled state. The sampling parameter effectively does the opposite, weighting more heavily the more completely-sampled state.

Given this performance, it is clear that specifying a sampling parameter is unlikely to be appropriate for our particular case. We have sampled all of the transitions between the character states. Further, as seen in Fig 1., oviparity is more undersampled than implied by the two-term correction used by Pyron and Burbrink (2014). We would expect, given this further information, for oviparity to be more likely at the root state. Future developers of BiSSE models should consider adding sampling corrections to the sampling parameter.

**What came first: the lizard or the egg?**

Due to the issues with the sampling parameter outlined in the previous section, we chose to fit our BiSSE models using the default setting (i.e., all taxa assumed to be sampled). The results of this fitting can be seen in Fig [3-tree]. Larger versions of each tree can be found in the supplemental material. As can be seen in Fig [3 tree], the Pyron, Wiens and Burbrink tree generally has fairly long branches deep in the tree, with short branches separating many of the major clades. This tree supports a total of 88 reversions (see counting criterion outlined under the section 'BiSSE models'). This is a very liberal count, likely representing the maximum number of possible reversions. See below for a discussion on how many reversals show strong evidence.

The optimized trees support far fewer reversals. As seen in the middle panel of Fig [3-tree], the deepest branches on these trees are much shorter than those on the Pyron, Wiens and Burbrink tree. [ Matt and David, here would be a good place for the two of you to make some contributions, being generally more familiar with the tree and its historical relationships than I].

d) A specific discussion of the clades we find to generally have reversions. What are these cases, and why they may not be 'solid'.

Conclusions

Here, we demonstrate that modeling the evolution of parity mode in a phylogenetic context is highly sensitive to both the underlying phylogeny and to specific parameters used in the analysis. By looking at a sample of trees, we have demonstrated that Pyron and Burbrink's 2014 conclusion that the ancestor of the squamate tree was viviparous is strongly sensitive to the sampling parameter of the BiSSE model framework. We have also computed a range of possible solutions for the question of how many reversals to oviparity are seen in the squamates. In totality, we conclude that the ancestor of all squamates was most likely oviparous, and that a number of reversals to oviparity have occurred in the squamate tree. Oviparity, therefore, is unlikely to be a Dollo trait.

Phylogenetically-informed inferences using large trees provide a powerful tool for understanding evolutionary patterns, including those that extend far into the past. However, making use of large trees in comparative biology presents major challenges, most clearly the optimization of phylogenetic trees and appropriately fitting models of character evolution on said trees. For users conducting such analyses, we strongly suggest performing multiple rounds of topology searches. We also suggest performing sensitivity analyses (i.e. altering the value of key parameters of models used in the estimation of character evolution). Conducting multiple rounds of optimization, fitting models on many trees, and comparing trees in terms of score and structure is not a task that can be completed by hand. Researchers conducting large-tree inferences would be well-served by developing computational pipelines to conduct this work. All scripts to replicate analyses conducted in this paper, and a sample pipeline can be found in the supplemental material.

A. M. Shedlock and S. V. Edwards. Amniotes (Amniota). Pp. 375–379 in

Timetree of Life, S. B. Hedges and S. Kumar, Eds. (Oxford University Press, 2009).

S. B. Hedges and N. Vidal. Lizards, snakes, and amphisbaenians (Squamata). Pp. 383–389 in Timetree of Life, S. B. Hedges and S. Kumar, Eds. (Oxford University Press, 2009)

Dollo, L (1893). Les lois de l’évolution. *Bull. Soc. Belge Geol. Pal. Hydr*, VII:164-166.

A. Stamatakis. RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. In Bioinformatics 22(21):2688-2690, 2006.

[Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010) "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1 - 8.](http://www.phylo.org/sub_sections/portal/sc2010_paper.pdf)

A. Stamatakis, A.J. Aberer, C. Goll, S.A. Smith, S.A. Berger, F. Izquierdo-Carrasco: "RAxML-Light: A Tool for computing TeraByte Phylogenies", Bioinformatics 2012; doi: 10.1093/bioinformatics/bts309.

Zwickl, D. J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.