

Which Came First: The Lizard or the Egg?
Robustness in Phylogenetic Reconstruction of Ancestral States

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

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

2: School of Biological Sciences, University of Sydney, NSW 2006,
Australia; and New York University – Sydney, The Rocks, NSW 2000,
Australia

Phone: 512-471-5792

E-mail:

AMW: wright.aprilm@gmail.com

KML: klyons19@gmail.com

MCB: mbrandley@gmail.com

DMH: dhillis@austin.utexas.edu

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 has evolved many times independently within amniotes (especially in lizards and snakes), with possibly a few reversions to oviparity. In amniotes, the shelled egg is considered a complex structure that is unlikely to re-evolve if lost (i.e., it is an example of Dollo's Principle). However, a recent ancestral state reconstruction analysis concluded that viviparity was the ancestral state of squamate reptiles (lizards and snakes), and that oviparity re-evolved from viviparity many times throughout the evolutionary history of squamates. Here, we re-evaluate support for this provocative conclusion by testing the sensitivity of the analysis to model assumptions and estimates of squamate phylogeny. We found that the models and methods used for parity mode reconstruction are highly sensitive to the specific estimate of phylogeny used, and that the point estimate of phylogeny used to suggest that viviparity is the root state of the squamate tree is far from an optimal solution. The ancestral state reconstructions are also highly sensitive to model choice and specific values of model parameters. A method that is designed to account for biases in taxon sampling actually accentuates, rather than lessens, those biases with respect to ancestral state reconstructions. In contrast to recent conclusions from the same data set, we find that ancestral state reconstruction analyses provide highly equivocal support for the number and direction of transitions between oviparity and viviparity in squamates. Moreover, the reconstructions of ancestral parity state are highly dependent on the assumptions of each model. We conclude that the common ancestor of squamates was oviparous, and subsequent evolutionary transitions to viviparity were common, but reversals to oviparity were rare. The three putative reversals to oviparity with the strongest phylogenetic support occurred in the snakes *Eryx jayakari* and *Lachesis*, and the lizard, *Liolaemus calchaqui*. Our results emphasize that because the conclusions of ancestral state reconstruction studies are often highly sensitive to the methods and assumptions of analysis, researchers should carefully consider this sensitivity when evaluating alternative hypotheses of character-state evolution.

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, defined by its water-retaining shell, four extra-embryonic membranes, and embryo-nourishing yolk, existed by the time the earliest amniotes (mammals and reptiles) diverged from one another about 325 mya (Hedges, 2009; Benton, 2014)—long before the first chicken walked the Earth. Today, the majority of living amniotes are oviparous, including all birds, crocodylians, tuataras, turtles, and monotreme mammals (Fig. 1). However, the distribution of parity modes in lizards and snakes (squamate reptiles) is far more diverse. Approximately 20% of squamate species are viviparous, and this complex of traits has been estimated to evolve independently over 100 times across the squamate phylogeny (Blackburn, 2006; 2014). In these viviparous species, developing embryos are retained in the mother’s uterus for the entire duration of embryonic development.

The traditional view of parity mode evolution in amniotes is that the most recent common ancestor of squamates, which lived ~200 mya (Jones et al., 2013), was oviparous, as it inherited the same ancestral parity mode that characterizes all other reptiles (Fig. 1). The transition from oviparity to viviparity requires extensive modification of uterine physiology and morphology (e.g., Blackburn, 2006, 2015; Blackburn and Flemming, 2008). For example, uterine shell glands in oviparous species secrete calcium during the discrete period of eggshell construction (Stinnett et al., 2010; Stewart, 2013). In viviparous species, shell gland function has been modified to provide calcium to the embryo throughout gestation (Herbert et al., 2006; Stewart, 2013). True “ovoviviparity” does not exist in squamates, as all examined viviparous squamates have some form of placenta composed of both maternal and embryonic tissue. The uterine structure of oviparous species is therefore modified into the maternal half of the placenta in viviparous species. The embryonic portion of the placenta is composed from the same extra-embryonic membranes that are present in all amniote eggs. Most squamate placentae are relatively simple structures used primarily for gas exchange and water transport (Thompson and Speake, 2006), but a more elaborate placenta that facilitates significant nutrient exchange has evolved at least six times in squamates (Blackburn, 2014; Stewart and Blackburn, 2014). Underlying the evolutionary transition to viviparity and a placenta are significant changes in gene expression of hundreds of genes (Brandley et al., 2012; Van Dyke et al., 2014).

Given the complexity of morphological, physiological, and genetic changes associated with the evolutionary transition to viviparity, the prevailing view is that when viviparity evolves in a lineage, reversals to oviparity are exceedingly unlikely to occur. Such a

reversal would be 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 reversal to oviparity would not simply be a matter of shutting off viviparity-specific functions, thereby returning to a default ancestral oviparous state. Rather, it would require the re-evolution of another relatively complex character suite—a shelled egg that supports external development, as well as uterine functions specific to oviparity that may have been lost while the lineage was viviparous (see Griffith et al., 2015).

Nonetheless, multiple studies using phylogenetic comparative methods and parity mode character-state data have challenged the traditional view that reversals to oviparity are highly unlikely. de Fraipont et al. ('96) estimated that the rate of transition from oviparity to viviparity has been similar to the rate of transition from viviparity to oviparity in squamates. In contrast, Shine and Lee ('99) strongly criticized de Fraipont et al. ('96), especially in relation to how the study incorporated phylogenetic information, and concluded that there is no support for reversals to oviparity (but see de Fraipont, '99).

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. They demonstrated this problem by comparing the fit and performance of state-dependent models to state-independent models in a simulation framework. Lynch and Wagner (2009) subsequently examined the oviparity reversal hypothesis in snakes by testing whether the oviparous parity mode in the sand boa, *Eryx jayakari*, re-evolved from viviparous ancestors. They reconstructed ancestral states of parity mode using the state-dependent BiSSE model (Maddison et al., 2007) while estimating the rate of transition between parity modes and parity-dependent diversification rates. They estimated statistically significant support for the evolutionary reversal from viviparity to oviparity in the sand boa (see also Griffith et al., 2015). Using similar methods, Fenwick et al. (2012) also inferred a possible reversion to oviparity from viviparity in pitvipers. 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, although the oviparity to viviparity transition has remained the dominant paradigm of squamate parity evolution, there is nonetheless provocative ancestral state reconstruction evidence that reversals to viviparity may occur, although rarely.

Recently, the results of Pyron and Burbrink's (2014) phylogenetic ancestral state reconstruction analyses once again challenged the traditional view of squamate parity mode evolution. That study

compiled binary oviparity/viviparity character state data for a massive data set including 8006 species, or about 85% of all squamates. For their phylogenetic framework, Pyron and Burbrink (2014) used the point estimate of a phylogeny from Pyron et al.'s (2013) maximum likelihood analysis of a squamate DNA supermatrix (4162 species). The phylogeny was then pruned to the 3952 species for which both phylogenetic and parity mode information had been collected. This single tree was the phylogenetic framework used in Pyron and Burbrink's ancestral state reconstruction analyses.

Pyron and Burbrink's (2014) ancestral state reconstruction analyses using the BiSSE model supported the hypothesis that the most-recent common ancestor of all extant squamates was viviparous rather than oviparous. Moreover, they concluded not only that reversals from viviparity to oviparity are common, they actually occur more frequently than transitions from oviparity to viviparity. For example, their 6-parameter BiSSE analysis estimated 34 origins of viviparity (including the root node), and 59 reversals to oviparity. In contrast, their Dollo (irreversible) parsimony model estimated 121 origins of viviparity—similar to the estimate of at least 115 independent origins of squamate viviparity by Blackburn (2006; 2014), who also assumed no reversals. Pyron and Burbrink's (2014) conclusion of frequent reversals from viviparity to oviparity is profoundly different than the prevailing concept of the evolution of squamate viviparity, and indeed the evolution of complex structures in general. This conclusion should therefore be subject to enhanced scrutiny before accepted as a new paradigm for future study (see also Pyron and Burbrink, 2015).

Here, we re-evaluate Pyron and Burbrink's (2014) ancestral state reconstruction analyses with respect to estimates of squamate phylogeny, parameters used in the various models of character evolution, and other aspects of model choice and taxon sampling. We also assess the effects of BiSSE model parameters on estimates of ancestral states, as well as the model's performance using massive phylogenies.

Although we discuss the implications of our re-analysis of Pyron and Burbrink (2014) to the study of viviparity, our primary goal is not to find a single analytical solution to the question of whether reversals to oviparity evolved. Rather, our motivation is to assess how differing model assumptions and phylogenetic uncertainty affect ancestral state reconstruction estimation. Because these issues may affect ancestral state estimation in general, and therefore the biological interpretation of these reconstructions, our results are applicable to any study that uses these methods, regardless of taxonomic focus.

METHODS

Phylogenetic Estimation

The DNA supermatrix used by Pyron et al. (2013) consisted of up to five mtDNA and seven nuclear genes from all available squamate sequences on Genbank, for 4162 taxa. One potential problem of most supermatrix studies is that not all phylogenetic studies use the same loci, and therefore the supermatrix contains a large amount of missing data (as not all loci are available for all species). For example, 81% of the cells in the DNA matrix of Pyron et al. (2013) are empty. Because these supermatrix studies also typically sample thousands of taxa, many of which may be represented by only fragmentary sequences, we expect a large degree of uncertainty in the resulting phylogenetic estimates (Sanderson et al. 2010).

The representation of squamate taxa for each locus in the Pyron et al. (2013) supermatrix is highly unequal. Squamate phylogenetic analyses have used nuclear loci mostly to infer “deep” relationships. These studies have typically sequenced multiple nuclear loci, but from only a few exemplars per family. Indeed, the deep, interfamilial phylogenetic relationships in the Pyron et al. (2013) tree generally agree with other squamate multi-locus phylogenies that sampled fewer taxa (e.g., Vidal and Hedges 2009; Mulcahy et al., 2012; Wiens et al., 2012; Reeder et al., 2015). On the other hand, there are many mtDNA sequences available for a diverse assemblage of squamate species, especially at the intergeneric and interspecific levels where nuclear sequences have not yet been sequenced. This distribution of data results in a situation where the “deep” relationships are reconstructed using multiple loci, but the tip relationships are driven by mtDNA sequences. We therefore also expect the tip relationships to be subject to more phylogenetic error that may artificially increase the estimated number of transitions in the ancestral state reconstruction analyses.

Given the large number of taxa sampled in the Pyron et al. (2013) data set and large amount of missing data, finding a near-optimal phylogenetic solution is a computationally difficult problem. However, because the ancestral state reconstruction analyses of Pyron and Burbrink (2014) rely on a single point estimate of squamate phylogeny, alternative phylogenetic estimates (many of which are likely to reflect improved fits to the data) may support conflicting results. We therefore used three sets of phylogenetic analyses to attempt to assess the degree to which the conclusions of Pyron and Burbrink (2014) are affected by phylogenetic uncertainty, and to assess the degree that their point estimate of phylogeny reflects the optimal solution space.

Our analyses re-estimated the 4162-taxon squamate phylogeny used in Pyron and Burbrink’s (2014) analysis under the same likelihood model and parameters used in the original phylogenetic study (Pyron et al., 2013). We used the same partitioning scheme used by Pyron et

al. (2013), including separate partitions for each of the three coding positions for the 10 protein-coding genes and a single partition each for the 12S and 16S rRNA, for a total of 32 partitions. Separate parameters of the GTR+ Γ model were estimated for each partition.

In our first set of analyses, we performed ten separate ExaML (Stamatakis, 2014; Jarvis et al., 2014) tree searches, each seeded from a starting tree obtained from ten separate parsimony searches in RAxML 7.7.6 (Stamatakis, 2014). ExaML is a version of the RaxML software that uses a new parallelization approach to speed up estimation of large trees. ExaML and RaxML produce rounded approximations of likelihood scores for trees, rather than precise likelihood scores. The scores calculated in different versions of the software are not comparable with one another (Stamatakis, 2014). To compare the likelihood scores of all trees found by different programs and the Pyron et al. (2013) tree, we re-calculated precise (rather than approximate) likelihood scores of all trees using the software Garli (Zwickl, 2006). These calculations revealed large discrepancies between reported scores from ExaML and Garli for the same trees. In all cases we examined, the approximate scores reported by RaxML and ExaML over-estimated the fit of the data to a given tree.

Our second set of analyses used Garli to optimize the branch lengths without changing the tree topology of the ten best ExaML trees. This branch-length optimization step accounted for substantial improvements in likelihood scores for all ten trees. For the final set of analyses, we used Garli to search for improvements in both topology and branch lengths of the ten ExaML trees.

We used the trees with the highest likelihood score from each of four sets of analyses for our subsequent ancestral state reconstruction analyses (Table 1): (1) The tree reported by Pyron et al. (2013) and used by Pyron and Burbrink (2014); (2) the best tree from the ten ExaML searches; (3) the best tree from the ExaML searches, but with branch lengths optimized under Garli; and (4) the best tree from the ExaML searches with both branch lengths and topology improved in Garli (this was the solution with the overall highest likelihood, or the “best tree” found across all analyses).

The model of character evolution used by Pyron and Burbrink (2014) to reconstruct ancestral states requires a time-calibrated phylogeny. We used identical methods as reported by Pyron and Burbrink (2014) by scaling all trees to absolute time using the software treePL (Smith and O’Meara, 2012), which is a C++ implementation of r8s (Sanderson, 2002). This method uses a penalized likelihood algorithm to smooth rates among branches. We also used the same seven calibration points used by Pyron and Burbrink (2014); these calibration nodes and dates are shown in Figure 2 (see the Supplemental Information for details of treePL implementation).

To compare topological distances between phylogenetic estimates, we computed Robinson–Foulds symmetric difference scores (Robinson and Foulds, '81) using the Dendropy Python library (Sukumaran and Holder, 2010). Because of the potentially large number of differences between trees, the unweighted version of this metric is preferable to the weighted version (Kuhner and Yamato 2014).

BiSSE Models and Taxon Sampling Parameters

Pyron and Burbrink (2014) estimated the ancestral state of squamate reproductive parity mode using the BiSSE model. BiSSE was developed by Maddison et al. (2007) 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 a separate speciation rate for taxa possessing state 1, an extinction rate for taxa possessing character state 0 and another extinction rate for taxa possessing state 1, a rate of change from state 0 to state 1, and a rate of change from state 1 to state 0. As Pyron and Burbrink's model-fitting work selected the 6-parameter BiSSE model as the best-fit to the data, and our goal was to examine sensitivity of the model to assumption violations, we did not explore alternative models. We used the default setting for the root state parameter. This setting allows the relative probabilities of the two states at the root to be estimated from the data (FitzJohn et al. 2009).

As originally conceived, the BiSSE model was only appropriate with trees for which all species were included. However, this model was extended by FitzJohn et al. (2009, 2012) to include sampling parameters that permit the use of trees with incomplete taxon sampling (using the software Diversitree). These sampling parameters require users to specify the proportion of the extant species of the clade that have been sampled. Users can specify sampling corrections using either one or two terms. A one-term setting informs the analysis of the proportion of the species sampled from within the clade, irrespective of character state. The two-term setting informs the analysis of the proportion of all species with state 0 and state 1 that are sampled. Pyron and Burbrink (2014) used a two-term sampling correction, specifying that 47% of oviparous squamate species and 63% of viviparous squamate species were sampled. Diversitree outputs ancestral states calculated in the process of estimating values for the six other parameters in the model. These states undergo no further optimization, as they are calculated as a step in estimating the model parameters.

An important note for later discussion is that the BiSSE model assumes that lineages were sampled at random from across the phylogeny. Pyron et al. (2013) sampled most lineages known to contain

transitions between parity mode states. Sampling of taxa was clearly not random, as it included all known lineages deep in the tree, with reduced sampling near the tips. If virtually all transitions between parity states are captured by the taxon sampling, it is unclear that either the one-term or the two-term sampling parameter should be included, as the goal is to estimate the number and direction of changes between parity mode states (as opposed to the rates of speciation and extinction with respect to character-state, which is the primary goal of the software). Because the exact values used in the one- and two-term sampling corrections will be highly dependent on the exact scale of the taxonomic question, we performed two tests to assess the impact of including either of these parameters on the analyses.

The first test was to use three different sampling parameter settings to estimate a BiSSE model and ancestral states on each tree: (1) the default setting (which involves no correction for taxon sampling); (2) a one-term model assuming the proportion of all squamate species sampled = 0.42; and (3) a two-term model assuming the proportion of known oviparous species sampled = 0.47, and proportion of known viviparous species sampled = 0.63. Pyron and Burbrink (2014) used the two-term model.

To assess the effect of the two-term sampling correction on the estimated parity state at the root of the squamate phylogeny, we estimated BiSSE models with incrementally-changing sampling parameters along four different trees: (1) the original Pyron et al. (2013) tree; (2) the best tree resulting from the ExaML searches; (3) the best ExaML tree with branch lengths re-optimized in Garli; and (4) the best ExaML tree with branch lengths and topology re-optimized in Garli (Table 1).

Counting Evolutionary Changes in Parity State

Counting changes between the ancestral reconstructions of character states in BiSSE is not straightforward, because BiSSE does not assign unambiguous states to ancestral nodes. Instead, BiSSE estimates the relative proportion of marginal likelihoods for the two states at each node of the tree. One approach to counting potential changes would be to count any change from >0.50 support for viviparity to >0.50 support for oviparity, or vice versa. However, many of the internal nodes on the squamate tree have very weak support for one parity-mode state or the other, so switches back and forth between ambiguous states are common. This method results in many counts of character changes that have virtually no statistical support. Another method of counting changes was used by Pyron and Burbrink (2014). They only counted changes when support (the relative proportion of marginal likelihoods for the two states) for the alternative state at a node on the tree exceeded 0.95. We implemented this

algorithm, but discovered that it also results in over-counting changes. For example, consider the case of two oviparous sister species, nested deeply within a clade of otherwise viviparous species. If the state of the common ancestor of these two sister species is estimated in BiSSE with less than 0.95 support for oviparity, then a transition from viviparity to oviparity will be counted twice (once in each of the two sister species), rather than once in their common ancestor.

We found the most straightforward method of conservatively counting character changes on the BiSSE trees was to minimize the number of reversals to oviparity under a parsimony model that equally weighted changes between the two parity states. We used PAUP* 4.0d142 (Swofford, 2003) for these analyses. These counts match the expectations of changes in the BiSSE ancestral states, without over-counting changes as in the methods described above.

We also calculated the minimum number of reversals from oviparity to viviparity on the best phylogenetic estimate using weighted parsimony (as implemented in PAUP* 4.0d142). To determine which putative reversals to oviparity had the most support across the squamate tree, we evaluated transitions in parity mode under models that penalized reversals from oviparity to viviparity twofold, threefold, and fourfold relative to changes from oviparity to viviparity.

RESULTS

Phylogenetic Estimation

The ln-likelihood score of our overall best phylogeny was -2,592,702.950. The tree reported by Pyron et al. (2013), which was also used in Pyron and Burbrink's (2014) analysis, had a ln-likelihood score of -2,676,499.387 as precisely scored by Garli. Thus, the best tree we found represented an improvement of >83,796 ln-likelihood units from the tree estimated by Pyron et al. (2013). Figure 2 shows a schematic of this tree with the major clades of squamates indicated.

Figure 3 shows the tree used by Pyron and Burbrink (2014), compared to the best trees from each of our phylogenetic analyses, with branches colored by the marginal likelihoods for the respective reconstructed ancestral states (oviparity or viviparity). Table 1 shows the likelihood score differences and symmetric topological differences among these trees. Symmetric difference scores (Robinson and Foulds, '81) indicate the total number of internal branches that need to be collapsed or added to convert between two trees. On these trees, the major groupings of squamates (colored on Fig. 2) are always recovered, with differences being found mostly among tip lineages.

BiSSE Models and Taxon Sampling Parameters

Sensitivity to the sampling parameters

As shown in Figure 4, there are strong effects of both tree topology and sampling parameters on the estimated parity state for the ancestor of squamates. On the Pyron and Burbrink (2014) tree, there is a large region of parameter space (the values of the two sampling parameters) in which viviparity is supported at the root of the squamate tree (Fig. 4A). However, as trees of higher likelihood are examined, viviparity at the root of the squamate tree is favored in a decreasing region of the parameter space (Fig. 4B-D). In all of the analyses, higher proportional sampling of viviparous taxa results in increasing support for viviparity at the root of the tree, and higher proportional sampling of oviparous taxa results in increasing support for oviparity at the root of the tree.

Ancestral State Reconstruction

As can be seen on Figure 3 and Table 2, the estimated number of origins of viviparity, as well as the estimated number of reversals from viviparity to oviparity, varies across trees and models. When changes in either direction between parity-mode states were weighted equally, we estimated between 98 and 106 origins of viviparity, with between 12 and 18 reversals to viviparity across the set of examined trees. However, if reversals to oviparity from viviparity are treated as less likely (penalized in the counting algorithm), then we found support for only 2–3 reversals under the 2X weighting, 0–1 under 3X weighting, and none under 4X weighting. The strongest supported reversal to oviparity (the only one supported under 3X weighting, on all but the best tree) was in the ancestor of *Lachesis*. The next most consistently supported reversal (in all trees at 2X weighting) was in *Eryx jayakari*. The only other reversal that received support in 2X-weighted analyses (but not for all trees) was in *Liolaemus calchaqui*.

DISCUSSION

As DNA data have become available for a greater diversity of species, phylogenies with thousands of species have been increasingly utilized to explore large-scale patterns of evolution using phylogenetic comparative methods (e.g., Zanne et al., 2013). However, relatively unexplored aspects of large-scale comparative methods analyses are the assumptions of the respective models (Thomas et al., 2014) and the effects of alternative tree topologies and branch length optimization on parameter estimates. This is particularly important because phylogenies derived from supermatrices of available GenBank sequences typically contain massive amounts of missing data that could result in high phylogenetic uncertainty.

Recently several studies have scrutinized the performance of trait-dependent speciation models such as BiSSE. Maddison and Fitzjohn (2015) and Robosky and Goldberg (2015) showed that

statistical tests of character evolution and speciation rate using these models are subject to very high rates of Type I error. Using the Pyron and Burbrink (2014) squamate data, King and Lee (2015) found that incorporating clade-specific rates dramatically reduced the estimated number of viviparous to oviparous character-state changes to the same three reversals that were most strongly supported in our analyses—in *Eryx jayakari*, *Lachesis*, and *Liolaemus calchaqui*. Our results further demonstrate that estimates of character-state changes using the BiSSE model are highly sensitive to variance in phylogenetic estimates and assumptions of taxon and character sampling.

In our reanalysis of the squamate data set, we found dramatic differences in likelihood score, tree topology, and branch lengths among analyses. The previously estimated tree from this data set (Pyron et al., 2013) is far from optimal (>83,796 ln-likelihood units and 809 symmetric distance units from the best tree we estimated; Fig. 3). However, the purpose of our study was not simply to find an improved phylogenetic solution for this data set, but rather to test the robustness of parity mode evolution to deviations in tree estimation, as well as the sensitivity of the models and parameters used to reconstruct ancestral states on these trees.

One major difference between the new analyses shown in Figure 3 and the Pyron and Burbrink (2014) analysis is the estimated age for the origin of Squamata. In the new analyses, we estimate an age for this node that is very close to the 228 Mya calibration date for the more inclusive Lepidosauria clade (Squamata + *Sphenodon*). In contrast, the estimated age of Squamata on the Pyron and Burbrink (2014) tree is 174.1 Mya. This discrepancy is due to the different lengths of the distantly related outgroup branch (*Sphenodon*) relative to the diversification of Squamata, as estimated on the suboptimal versus improved trees. As expected for an imperfect model of sequence evolution, the length of the long branch leading to *Sphenodon* is underestimated in the near-optimal trees relative to its well-sampled sister lineage (Squamata), which forces the estimated squamate divergence time to be very close to the constrained calibration date for Lepidosauria. Additional outgroups and calibration dates are needed to more accurately estimate the age of Squamata from these data. Regardless, both the 228 and 174.1 Mya estimates fall within the range of other estimated ages for Squamata (Jones et al. 2013), and most importantly, assuming either age has no appreciable effect on our results (see below).

Ancestral State Reconstructions Are Sensitive to Model Parameters and Assumptions

We performed ancestral state reconstructions analyses using three different taxon-sampling models implemented in BiSSE: (1) sampling all species in a clade—the default model; (2) a one-parameter model that accounted for the percentage of all squamate species sampled; and (3) a two-parameter model that accounted for the separate percentages of sampled oviparous and viviparous species.

We examined the estimated parity state of the root when sequentially changing values of the two-parameter sampling model (and therefore different assumed proportions of sampled oviparous to viviparous species). The two-term analysis of Pyron and Burbrink (2014) reconstructed the common ancestor of squamate reptiles as viviparous. However, the estimated state of this common ancestor becomes more ambiguous as more optimal phylogenetic solutions are considered (Table 2; Fig. 4). Moreover, inspection of Figure 4 shows that as the percentage of oviparous taxa sampled is increased (or if the percentage of viviparous taxa is decreased), then the inferred ancestral state for squamates switches from viviparous to oviparous. This is exactly opposite of the expected behavior of a “bias correction,” if the purpose of the two-parameter correction is to account for possibly unsampled groups of oviparous or viviparous taxa. If we sampled mostly oviparous taxa, and relatively few viviparous taxa, then we might be concerned about biasing the ancestral state to oviparity, and wish to correct that problem. However, the two-parameter sampling model in BiSSE has the opposite effect: the more biased the sampling is toward oviparous taxa, the more the “corrected” analysis favors oviparity as the root state (Fig. 4). Pyron and Burbrink (2014) inferred the ancestral parity state of squamates as viviparous because they relatively oversampled viviparous taxa compared to oviparous taxa; if their sampling bias had been in the opposite direction, they would have inferred the ancestral state as oviparous, even on their tree and preferred model. Thus, the two-parameter sampling model in BiSSE is not accounting for sampling bias in the expected manner for ancestral state reconstruction. Even given the biases in the two-parameter sampling model of BiSSE shown in Figure 4, improvements in the tree topology (i.e., analyzing trees of higher likelihood) shift the inferred ancestral state towards a more ambiguous result (Table 2).

It is worthwhile considering why taxon sampling needs to be taken into account when estimating ancestral states on a phylogenetic tree. If taxa are sampled randomly from the complete squamate tree, and there are a few large oviparous clades of squamates, then many small clades of viviparous taxa would be expected to be missing from the sample. In this case, there would be a legitimate concern that the number of origins of viviparity (and the subsequent reversions to oviparity) would be underestimated, even if the rates of change

between states were estimated correctly. However, in the study of Pyron et al. (2013), squamate taxa were not selected randomly. Instead, taxa have been sequenced to include as many genera and families as possible, so that the tree is well represented with respect to phylogenetic diversity. In addition, most previously suggested origins of viviparity (as well as suggested reversions to oviparity) in squamates are represented in the tree. Under these conditions, taxon sampling of squamates is not likely to be a major concern with respect to biasing the ancestral state reconstructions.

Although the sampling of squamates in the analysis by Pyron and Burbrink (2014) was comprehensive, their sampling of amniotes outside of squamates is a much greater concern for this analysis, especially with regard to the estimate of the ancestral squamate condition. Their study included the oviparous tuatara, the closest living relative of Squamata, as the single outgroup. They used 228 MY as the constraint for the root between the tuatara and Squamata. Given the long branch from the root to the single outgroup (see Fig. 1), the parity state for tuatara (oviparity) is relatively uninformative about the root state of squamates. Although viviparity is common in squamates, it has never been recorded from any other extant reptile group, including birds, crocodylians, tuataras, and turtles, all of which are oviparous (Fig. 1). Thus, there is little doubt that the ancestor of birds, crocodylians, and turtles (which existed much closer to the ancestor of squamates in time, compared to the extant tuatara) was oviparous. By including other reptiles as outgroups (in addition to the tuatara), the ancestral parity state of squamates would be even more clearly reconstructed as oviparous.

Ancestral State Reconstructions Are Sensitive to Estimates of Tree Topology

We found that multiple independent runs of ExaML and Garli using different starting trees each infer a different suboptimal tree. This inconsistency of results is expected given the extremely complex solution space that results from the large number of taxa sampled and the substantial amount of missing data. Any of these approximate solutions might give a reasonable estimate of ancestral state reconstructions, *if* the process of ancestral state reconstruction is found to be robust across point-estimates of phylogeny. However, as illustrated in Figure 3 and detailed in Table 2, ancestral state reconstructions differ widely across different point estimates of phylogeny, different sampling models, and different methods of counting state changes, with the number of nominal reversals from viviparity to oviparity ranging from 0 to 18.

Table 2 presents the estimated root state and the number of estimated changes between oviparity and viviparity (in both directions) among four phylogenetic estimates and three BiSSE parameterizations.

With no correction for incomplete taxon sampling, the root state is consistently reconstructed with stronger support for oviparity (75.0% – 90.4%). Using the one-parameter sampling model (which accounts for the percentage of all squamate species sampled), all of the solutions except the tree used by Pyron and Burbrink (2014) also indicate greater support for ancestral oviparity in squamates (87.9% – 97.5%). With the two-parameter sampling model, there is much less apparent support for oviparity as the ancestral state for squamates across all four trees (7.4% – 24.4%). However, as shown in Figure 4 and discussed earlier, the two-parameter “correction” for taxon sampling in BiSSE actually accentuates taxon-sampling bias (at least with respect to ancestral-state reconstruction), rather than correcting it.

As we noted previously, our age estimates for the origin of Squamata from the re-analyses are older than the estimate from the sub-optimal tree reported by Pyron and Burbrink (2014)(Fig. 3). To determine if this age discrepancy is responsible for the differing support for root states estimated by Pyron and Burbrink (2014) and this study, we performed additional ancestral state reconstruction analyses using our newly estimated tree topologies, but with an added calibration date for the age of Squamata (174.1 Mya—the estimate from the Pyron and Burbrink analysis). Adding this additional calibration point did not substantially change support for the root state (see Supplemental Information for the details of the analyses that include this additional calibration point). In the analyses that used the one-term taxon sampling correction or no sampling correction, support for oviparity as the root state remained high (92% and 87.7%, respectively). For the two-term sampling correction, the support for oviparity at the root remained low (3.3%).

If we consider just the unweighted model for counting character changes shown in Table 2, the estimated 12–18 reversals are nonetheless surprisingly high for a trait that is widely thought to be irreversible, or exceedingly rare. However, closer inspection reveals that many of those putative reversions are nearly equivocal (Table 2). The weak support for multiple reversions from viviparity to oviparity is a result of using a model that assumes that changes between parity states are equally likely. Given the starting assumption that reversions to oviparity should be expected to be rare (following Dollo’s Principle), a model that assumes equal changes between states does not appear to be warranted. For example, four of the 18 putative reversals to oviparity in the BiSSE analysis of the best tree are concentrated within pitvipers, along with three origins of viviparity (Fig. 5A; see also Fenwick et al., 2012). But a model that disfavors reversibility requires just ten origins of viviparity in pitvipers (as opposed to the seven character changes suggested in the BiSSE analysis) to account for the same distribution of states on the tree (Fig. 5B). Given that there is support for 98 or more origins of viviparity in squamates across all

analyses and trees (Table 2), a few additional origins of viviparity in pitvipers (as shown in Fig. 5B) would not be surprising.

Of the 18 putative reversals to oviparity suggested by the BiSSE analysis on the best tree, only three are still favored if we penalize reversals to oviparity just twice as heavily as new origins of viviparity (Table 2). Given the large number of transitions to viviparity throughout the squamate tree, most of the weakly supported possible cases of reversion to oviparity should be viewed with some skepticism. Nonetheless, it does appear that the three strongest cases (*Eryx jayakari*, *Lachesis*, and *Liolaemus calchaqui*) represent species that could be good subjects for additional genetic, morphological, and physiological study to test the hypothesis that parity mode shifted from viviparity to oviparity.

CONCLUSIONS

Our analyses demonstrate that modeling the evolution of parity mode and performing ancestral state reconstruction in a phylogenetic context is highly sensitive to both the underlying phylogenetic estimate (including both tree topology and branch lengths) as well as to specific model parameters used in the analysis. Given the sensitivity of ancestral state reconstruction analyses to specifics of phylogenetic topology and branch lengths, it is critical to examine more than just a single, approximate solution from a complex solution space. One solution to this problem of ancestral-state ambiguity across phylogenetic estimates would be to incorporate phylogenetic uncertainty by calculating ancestral states over distributions of trees inferred by Bayesian or non-parametric bootstrap analyses. However, the computational complexity of the analysis, combined with the large amount of missing data, make approximate searches of the solution space in “fast bootstrapping” analyses unreliable (Ansimova et al., 2011). Furthermore, thorough searches of solution space in every bootstrap replicate would require years of computation time with current technology.

After examining improved phylogenetic solutions for the squamate tree, and considering the potential effects of taxon sampling biases, our best estimates support the traditional view that the ancestor of all extant squamates was oviparous. From an oviparous squamate ancestor, there were clearly numerous transitions to viviparity and (possibly) a small number of reversals back to oviparity. The best cases of possible reversals to oviparity (such as the ones in *Eryx jayakari*, *Lachesis*, and *Liolaemus calchaqui*) should be examined in greater detail, including a close examination of genetic and developmental evidence for reversal to oviparity.

Finally, we suspect that the general conclusions of our analysis are not unique to the squamate parity data set, and apply to other large-scale ancestral-state reconstruction analyses using the BiSSE

model. Ancestral state reconstruction analyses of data sets composed of thousands of taxa, but with massive amounts of missing data, should be interpreted with caution, and only after accounting for phylogenetic uncertainty in the reconstruction process and the underlying assumptions of the reconstruction methods.

ACKNOWLEDGMENTS

We thank R. A. Pyron and F. T. Burbrink for access to data used in their 2014 paper. We also thank R. G. FitzJohn for help with the Diversitree package, K. R. Kelley for facilitating reproducibility of the computational pipeline, and the University of Texas at Austin and Intersect Australia Ltd. for supercomputing resources. M.C.B was supported by the Discovery Early Career Researcher Award from the Australian Research Council (DE120101615).

LITERATURE CITED

- Ansimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol* 60:685–699.
- Blackburn DG. 2014. Evolution of vertebrate viviparity and specializations for fetal nutrition: a quantitative and qualitative analysis. *J Morph*, *in press*. doi: 10.1002/jmor.20272.
- Blackburn DG. 2006. Squamate reptiles as model organisms for the evolution of viviparity. *Herp Monographs* 20:131–146.
- Blackburn DG. 2015. Viviparous placentotrophy in reptiles and the parent – offspring conflict. *J Exp Zool B: Mol Dev Evol*, *in press*. doi: 10.1002/jez.b.22624
- Blackburn DG, Flemming, AF. 2009. Morphology, development, and evolution of fetal membranes and placentation in squamate reptiles. *J Exp Zool B: Mol Dev Evol* 312:579–589.
- Brandley MC, Young RL, Warren DL, Thompson MB, Wagner GP. 2012. Uterine gene expression in the live-bearing lizard, *Chalcides ocellatus*, reveals convergence of squamate reptile and mammalian pregnancy mechanisms. *Genome Biol Evol* 4:394–411.
- de Fraipont M, Clobert J, Barbault R. 1996. The evolution of oviparity with egg guarding and viviparity in lizards and snakes: a phylogenetic analysis. *Evolution* 50:391–400.
- de Fraipont M, Clobert J, Meylen S, Barbault R. 1999. On the evolution of viviparity and egg-guarding in squamate reptiles: a reply to R. Shine and M.S.Y. Lee. *Herpetologica* 55:550–555.
- Dollo, L. 1893. Les lois de l'évolution. *Bull. Soc. Belge Geol. Pal. Hydr*, VII:164–166.

- Fenwick AM, Greene HW, Parkinson CL. 2012. The serpent and the egg: unidirectional evolution of reproductive mode in vipers? *J Zool Syst Evol Res* 50:59–66.
- FitzJohn RG, 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Meth Ecol Evol* 3:1084–1092.
- FitzJohn RG, Maddison WP, Otto SP. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst Biol* 58:595–611.
- Goldberg EE, Igić B. 2008. On phylogenetic tests of irreversible evolution. *Evolution* 62:2727–2741.
- Griffith OW, Blackburn DG, Brandley MC, Van Dyke J, Whittington C, Thompson MB. 2015. Ancestral state reconstructions require biological evidence to test evolutionary hypotheses: A case study examining the evolution of reproductive mode in squamate reptiles. *J Exp Zool B: Mol Dev Evol.*, *in press*. doi: 10.1002/jez.b.22614
- Hedges SB, Vidal N. 2009. Lizards, snakes, and amphisbaenians In: Hedges SB and Kumar S, editors. *Timetree of Life*, New York: Oxford University Press, 2009. p 383–389.
- Herbert JF, Lindsay LA, Murphy CR, Thompson MB. 2006. Calcium transport across the uterine epithelium of pregnant lizards. *Herpetol Monogr.* 20:205–211.
- Jarvis ED, Mirarab S, Aberer AJ, and 102 additional authors. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331.
- Jones MEH, Anderson CL, Hipsley CA, Müller J, Evans SE, Schoch RR. 2013. Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). *BMC Evol Biol* 13: 208.
- King B and Lee MSY. 2015. Ancestral state reconstruction, rate heterogeneity, and the evolution of reptile viviparity. *Syst Biol* 64:532–544.
- Kuhner, M. and Yamato, J. 2014. Practical performance of tree comparison metrics. *Syst Biol* 64: 205–214.
- Lee MSY, Shine R. 1998. Reptilian viviparity and Dollo's law. *Evolution* 52:1441–1450.
- Lynch VJ, Wagner GP. 2010. Did egg-laying boas break Dollo's law? Phylogenetic evidence for reversal to oviparity in sand boas (*Eryx*: Boidae). *Evolution* 64:207–216.
- Maddison WP and RG FitzJohn. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. *Syst Biol* 64:127–136.
- McKinney W. 2013. Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*: 51–56.
- Mulcahy DG, Noonan BP, Moss T, Townsend TM, Reeder TW, Sites Jr JW, Wiens JJ. 2012. Estimating divergence dates and evaluating

- dating methods using phylogenomic and mitochondrial data in squamate reptiles. *Mol Phylogenet Evol.* 65:974–991.
- Pincheira-Donoso D, Tregenza T, Witt MJ, Hodgson DJ. 2013. The evolution of viviparity opens opportunities for lizard radiation but drives it into a climatic cul-de-sac. *Global Evol and Biogeo* 22:857–867.
- Pyron RA, Burbrink FT, Wiens JJ. 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol Biol* 13:93.
- Pyron RA, Burbrink FT. 2014. Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. *Ecol Lett* 17:13–21.
- Pyron RA and Burbrink FT. 2015. Contrasting models of parity-mode evolution in squamate reptiles. *J Exp Zool Part B: Mol Dev Evol*, *in press*. doi: 10.1002/jez.b.22593
- Rabosky DL and EA Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst Biol* 64:340–355.
- Reeder TW, Townsend TM, Mulcahy DG, Noonan BP, Wood Jr PL, Sites Jr JW, Wiens JJ. 2015. Integrated analyses resolve conflicts over squamate reptile phylogeny and reveal unexpected placements for fossil taxa. *PloS ONE* 10: e0118199.
- Robinson, DR and Foulds LR. 1981. Comparison of phylogenetic trees. *Math Biosci* 53: 131–147.
- Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol* 19:101–109.
- Sanderson MJ, McMahon MC and Steel, M. 2010. Phylogenomics with incomplete taxon coverage: the limits to inference. *BMC Evol Biol* 10: 155–168.
- Shedlock AM and Edwards SV. 2009. Amniotes (Amniota). In: Hedges SB and Kumar S, editors. *Timetree of Life*, New York: Oxford University Press, 2009. p 375–379.
- Smith SA, O'Meara BC. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28:2689–2690.
- Stamatakis A, Aberer AJ, Goll C, Smith SA, Berger SA, Izquierdo-Carrasco F. 2012. RAxML-Light: a tool for computing TeraByte phylogenies. *Bioinformatics*. 28:2064–2066.
- Stamatakis A. 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.
- Stewart JR. 2013. Fetal nutrition in lecithotrophic squamate reptiles: toward a comprehensive model for evolution of viviparity and placentation. *J Morph* 274:824–843.
- Stewart JR, Blackburn DG. 2014. Viviparity and placentation in lizards. *Reproductive Biology and Phylogeny of Lizards*. Boca Raton, FL: CRC Press.

- Stinnett HK, Stewart JR, Eday TW, Pyles RA, Herbert JF, Thompson MB. 2012. Placental development and expression of calcium transporting proteins in the extraembryonic membranes of a placentotrophic lizard. *J Morph* 273:347–359.
- Sukumaran J, Holder MT. 2010. DendroPy: A Python library for phylogenetic computing. *Bioinformatics* 26:1569–1571.
- Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Thomas GH, Cooper N, Venditti C, Meade A, Freckleton RP. 2014. Bias and measurement error in comparative analyses: a case study with the Ornstein Uhlenbeck model. doi: <http://dx.doi.org/10.1101/004036>.
- Thompson MB, Speake BK. 2006. A review of the evolution of viviparity in lizards: structure, function and physiology of the placenta. *J Comp Physiol B*. 176: 179–189.
- Van Dyke JU, Brandley MC, Thompson MB. 2014. The evolution of viviparity: molecular and genomic data from squamate reptiles advance understanding of live birth in amniotes. *Reproduction* 147:R15–R26.
- Vidal N, Hedges SB. 2009. The molecular evolutionary tree of lizards, snakes, and amphisbaenians. *Comptes Rendus Biologies* 332:129–139.
- Wiens JJ, Hutter CR, Mulcahy DG, Noonan BP, Townsend TM, Sites JW, Reeder TW. 2012. Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. *Biol Lett* 8:1043–1046.
- Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB, Royer DL, Soltis DE, Stevens PF, Westoby M, Wright IJ, Aarssen L, Bertin RI, Calaminus A, Govaerts R, Hemmings F, Leishman MR, Oleksyn J, Soltis PS, Swenson NG, Warman L, Beaulieu JM. 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506:89–92.
- Zwickl, DJ. 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.

Figure Legends

Figure 1: The evolution of parity mode in amniotes. Clades colored red are viviparous, clades colored blue are viviparous, and clades colored purple are represented by both parity modes. Branch lengths represent time (y-axis).

Figure 2: The best estimate of squamate phylogeny found in this study. This solution has a lnL score of -2,592,702.95 or approximately 83,796 lnL units better than the tree used by Pyron and Burbrink (2014). Calibration points for time scaling used by Pyron and Burbrink (2014) are indicated by asterisks and numbers.

Figure 3: Results of ancestral state reconstruction using the BiSSE model (with no taxon sampling correction parameter) for the four trees discussed in the text. Branches are colored by marginal likelihoods for reconstructed ancestral states on a blue (support for oviparity) to red (support for viviparity) scale. Calibration points for time-scale tree are shown with black squares.

Figure 4: Support for the estimated parity mode root state of squamates as a function of values for the two-term taxon sampling parameter in FitzJohn's (2009) revised BiSSE model. The yellow star indicates the parameter values used by Pyron and Burbink (2014). Note that increasing the sampling of taxa with a given parity state increases the support for that state at the root of the squamate tree.

Figure 5. Two estimates of parity mode evolution in pitvipers. Ancestral states are estimated using (left) the BiSSE model without a taxon sampling correction, and (right) a Dollo parsimony model that does not allow reversals from viviparity to oviparity. Diamonds represent transitions to viviparity, while stars represent reversals to oviparity. The tree at right (Dollo parsimony) is also favored under a model of evolution that considers changes from oviparity to viviparity twice as likely as changes from viviparity to oviparity.

Table 1. Comparisons in likelihood score and tree topology (as measured in symmetric distance) between the Pyron et al. (2013) tree used by Pyron and Burbrink (2014) to the best trees found in the phylogenetic analyses described in the text. The “best tree” is the best tree found in our searches with both ExaML and Garli, but is unlikely to represent the optimal solution.

Tree	Ln-Likelihood score	Difference in Ln-likelihood from best tree found	Symmetric distance from best tree found
Pyron and Burbrink tree	-2,676,499.387	83,796.437	809
ExaML tree	-2,598,386.107	5,683.157	667
Branch-length improved tree	-2,595,843.385	3,140.435	551
Best tree	-2,592,702.950	0	0

Table 2. Number of changes to viviparity and oviparity, and support for the root state, as estimated across four trees and three methods of taxon-sampling correction. The ranges for origins of viviparity indicate the number of origins under the equal weighted model to the number of origins assuming no reversals to oviparity. Reversals to oviparity are shown for four models of character change: 1X = equal weighting of changes between viviparity and oviparity; 2X = twofold greater penalty of reversals to oviparity; 3X = threefold greater penalty of reversals to oviparity; 4X = fourfold greater penalty of reversals to oviparity.

Tree	Origins of viviparity	Minimum number of reversals to oviparity				Root state (% support for oviparity)		
		1X	2X	3X	4X	No sampling correction	One term correction	Two term correction
Pyron and Burbrink tree	106-129	13	2	1a	0	80.8	7.2	7.4
ExaML tree	103-126	12	3	1	0	75.0	87.9	11.2
Branch-length improved tree	103-124	12	3	1	0	90.4	97.5	24.4
Best tree	98-129	18	3	0	0	88.6	94.5	14.3

