Molecular Phylogenetics and Evolution 81 (2014) 221–231

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Effectiveness of phylogenomic data and coalescent species-tree methods for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes: Caenophidia)

R. Alexander Pyron [a,*], Catriona R. Hendry [a], Vincent M. Chou [a], Emily M. Lemmon [b], Alan R. Lemmon [c], Frank T. Burbrink [d,e]

[a] Dept. of Biological Sciences, The George Washington University, 2023 G St. NW, Washington, DC 20052, USA
[b] Dept. of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, USA
[c] Dept. of Biological Science, Florida State University, Tallahassee, FL 32306-4295, USA
[d] Dept. of Biology, The Graduate School and University Center, The City University of New York, 365 5th Ave., New York, NY 10016, USA
[e] Dept. of Biology, The College of Staten Island, The City University of New York, 2800 Victory Blvd., Staten Island, NY 10314, USA

ARTICLE INFO

Article history:
Received 7 January 2014
Revised 29 July 2014
Accepted 22 August 2014
Available online 3 September 2014

Keywords:
Phylogenomics
Species trees
Gene trees
Coalescence
Colubroidea
Snakes

ABSTRACT

Next-generation genomic sequencing promises to quickly and cheaply resolve remaining contentious nodes in the Tree of Life, and facilitates species-tree estimation while taking into account stochastic genealogical discordance among loci. Recent methods for estimating species trees bypass full likelihood-based estimates of the multi-species coalescent, and approximate the true species-tree using simpler summary metrics. These methods converge on the true species-tree with sufficient genomic sampling, even in the anomaly zone. However, no studies have yet evaluated their efficacy on a large-scale phylogenomic dataset, and compared them to previous concatenation strategies. Here, we generate such a dataset for Caenophidian snakes, a group with >2500 species that contains several rapid radiations that were poorly resolved with fewer loci. We generate sequence data for 333 single-copy nuclear loci with ~100% coverage (~0% missing data) for 31 major lineages. We estimate phylogenies using neighbor joining, maximum parsimony, maximum likelihood, and three summary species-tree approaches (NJst, STAR, and MP-EST). All methods yield similar resolution and support for most nodes. However, not all methods support monophyly of Caenophidia, with Acrochordidae placed as the sister taxon to Pythonidae in some analyses. Thus, phylogenomic species-tree estimation may occasionally disagree with well-supported relationships from concatenated analyses of small numbers of nuclear or mitochondrial genes, a consideration for future studies. In contrast for at least two diverse, rapid radiations (Lamprophiidae and Colubridae), phylogenomic data and species-tree inference do little to improve resolution and support. Thus, certain nodes may lack strong signal, and larger datasets and more sophisticated analyses may still fail to resolve them.

1. Introduction

The ability to generate genomic datasets containing hundreds or thousands of loci promises to rapidly transform phylogenetic inference (Faircloth et al., 2012; Lemmon et al., 2012). While few examples currently exist for large-scale phylogenies, these approaches should allow researchers to estimate credible species-trees, especially in cases where there is significant genealogical discordance among loci, by providing hundreds of independent observations of tree structure (Edwards, 2009; Lemmon and Lemmon, 2013). Phylogenomic estimation of species trees might therefore be expected to help resolve difficult nodes in the Tree of Life that were poorly supported in smaller datasets (Dunn et al., 2008; Rokas et al., 2003), particularly those that were analyzed using concatenated gene-tree inference (Edwards, 2009).

Alternatively, certain nodes in the Tree of Life may remain beyond resolution, even when confronted with genome-scale datasets (Delsuc et al., 2005; Rokas and Carroll, 2006). In particular, rapid evolutionary divergence may produce short branch-lengths between speciation events. In this case, there is little time for informative substitutions to fix in populations, yielding gene trees that are unresolved with respect to the species tree. In some situations,

* Corresponding author.
E-mail addresses: rpyron@colubroid.org (R.A. Pyron), frank.burbrink@mail.csi.cuny.edu (F.T. Burbrink).

http://dx.doi.org/10.1016/j.ympev.2014.08.023
1055-7903/

these short branches may even be positively misleading, as in the "anomaly" zone, where the most probable gene trees fail to match the underlying species tree due to short branch lengths and large effective-population sizes (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007). Trees in the anomaly zone may not be readily reconcilable using normal methods of phylogenetic reconstruction (e.g., maximum likelihood or Bayesian inference) based on DNA sequence data, if phylogenetically informative sites are lacking in the genome.

Two recently-derived analytical strategies may alleviate these problems. First is the independent analysis of loci under the multi-species coalescent model, resulting in joint inference of the individual gene-trees and the underlying species-tree (Heled and Drummond, 2010). This method estimates gene trees directly conditioned on the model of nucleotide substitution, and species trees jointly conditioned on the gene trees. Such approaches can resolve discordance among loci when sufficient phylogenetic signal is present, but is computationally intensive (Edwards et al., 2007). Recently, it has been reported that the current implementation of this method in the program *BEAST (Heled and Drummond, 2010) performs poorly on datasets with more than 50 loci, exhibiting a general lack of convergence (O'Neill et al., 2013).

Second are approaches based on summarizing the relative rank order of divergences of DNA sequences under the coalescent, which has been mathematically shown to estimate the true species-tree with high accuracy when the number of loci is large (Liu et al., 2009b). These methods are conditioned only on the expected distribution of gene trees, from which a species tree can be estimated directly. This makes them extremely fast compared to full coalescent methods, apparently with a high degree of accuracy (Liu et al., 2009a). One in particular, the maximum pseudo-likelihood of estimating species trees (MP-EST), approximates the species tree from the multi-species coalescent with a probability of 1 as the number of genes increases, even in the anomaly zone for gene-tree discordance (Liu et al., 2010).

These strategies have not been compared to traditional concatenation methods using large empirical phylogenomic datasets. Concatenating data, particularly when partitions contain incongruent signal, can introduce significant biases and distortions into phylogenetic inference, leading to strong support for erroneous topologies (see review in Lemmon and Lemmon, 2013). We note, however, that no species tree methods yet adequately account for other sources of gene discordance (such as horizontal gene transfer and gene duplication), and if these problems are persistent across a dataset, then the methods discussed here are still likely to fail even at phylogenomic scales (Liu et al., 2009b).

Here we use the Anchored Phylogenomics protocol (Lemmon et al., 2012) to generate a large-scale phylogenomic dataset for colubroid snakes containing hundreds of loci, thousands of informative sites, and hundreds of thousands of base pairs in total. The infraorder Caenophidia, which represents one of the largest radiations of tetrapods, contains over 3000 species (~87% of snakes) including all medically important venomous species (Vitt and Caldwell, 2009). Species-level relationships within the group have been well studied, and the monophyly and content of the various genera, tribes, and subfamilies is not generally in question, although these inferences still rely on only a handful of markers (Grazziotin et al., 2012; Kelly et al., 2003, 2009; Lawson et al., 2005; Zaher et al., 2009). In contrast, relationships among these groups and monophyly of Colubroidea with respect to Acrochordoidea (the basal caenophidian divergence) are not well supported (Pyron et al., 2011, 2013a). None of these studies sampled many loci (<10), and all were based on incomplete datasets, with missing data often >50%.

There are at least four contentious nodes in the caenophidian tree that we attempt to resolve through broad phylogenomic sampling. Assessing the conflict among these nodes is important because each potential resolution suggests a completely different pattern of origin for each major group. First, we test for monophyly of Colubroidea, as some studies have suggested that Xenodermatidae is actually the sister group of Acrochordoidea (traditionally regarded as the sister group of Colubroidea), rather than the basal colubroid lineage (Kelly et al., 2003; Pyron et al., 2013a). Second, we test the placement of Homalopsidae, which has been inferred as the sister group to Elapidae + Lamprophiidae in some studies (Pyron et al., 2011), and Colubridae + (Elapidae + Lamprophiidae) in others (Pyron and Burbrink, 2012; Wiens et al., 2008). Third, we test the monophyly of Lamprophiidae and support for the relationships of its various subfamilies, which appear to represent a rapid radiation that has historically proven difficult to resolve (Kelly et al., 2009; Pyron et al., 2011). Fourth, we similarly test support for the relationships of the various subfamilies of Colubridae, which vary widely among recent studies (Chen et al., 2013; Zaher et al., 2012). Ultimately, we attempt to determine if new methods for species-tree estimation, in combination with broadly sampled phylogenomic datasets with little missing data, can resolve these nodes with high support – historically the most basic motive for phylogenetic studies, regardless of data or methods.

Using the Anchored Phylogenomic dataset, we compare several methods of phylogenetic inference, including maximum parsimony (MP) and maximum likelihood (ML) analysis of the concatenated dataset to estimate gene trees (GT), and summary coalescent-based methods including MP-EST, species tree estimation using average ranks of coalescence (STAR), and neighbor joining species tree (NJst) for estimating species trees (ST). We compare the species trees to each other, extended majority-rule trees and to the individual gene-trees quantitatively using Robinson-Foulds (RF) distances, and qualitatively to previous analyses with respect to the resolution and support for the four problematic areas of the tree described above. Most, but not all, methods support monophyly of Caenophidia, though some place Acrochordidae as the sister group to Pythonidae. All methods estimate monophyly of Colubroidea as traditionally defined, as Xenodermatidae is the sister group to the remaining colubroids, but not all with strong support. Homalopsidae is strongly supported as the sister group to Colubridae + (Elapidae + Lamprophiidae) by all methods. This suggests that phylogenomic species-tree analysis can reveal novel relationships and support for some nodes. In contrast, little progress is obtained regarding Lamprophiidae and Colubridae, with weak support for relationships among their constituent subfamilies. This appears to confirm previous suspicions that certain nodes in the Tree of Life may be permanently intransigent for ordinary phylogenetic inference using DNA sequence data, at least when deep coalescence is the only mechanism evaluated for gene-tree discordance.

2. Materials and methods

2.1. Molecular data

We obtained tissue samples (via museum loans) for representatives from the 7 colubroid families and 17 subfamilies (see Table S1 for vouchers). Some subfamilies (Crotalinae, Dipsadinae, Lamprophiinae, and Natricinae) were represented by more than one terminal species. We did not include several taxa considered *incertae sedis* in Lamprophiidae (e.g., *Micrelaps, Oxyrhabdium*). We used *Anolis carolinensis* as the squamate outgroup, and included *Acrochordus granulatus* to represent the basal caenophidian lineage. We also included *Python* as an additional alethinophidian outgroup, from a draft genome assembly (see below). It may be beneficial to include more outgroups (Zwickl and Hillis, 2002), but the

sampling here is constrained by time and finances. Hopefully, the adoption of this protocol over time will result in enlarged datasets to test these preliminary results. We thus carefully compare these results to previous analyses with more taxa but fewer loci (Pyron et al., 2013a). The final sampling consisted of 32 total species (1 outgroup, and 31 caenophidians). Tissues were typically visceral organs (e.g., heart, liver), extracted using Qiagen DNEasy kits to obtain ~1–2 μg of high molecular-weight genomic DNA, eluted in the standard buffers prior to library prep and enrichment.

We prepared the samples for next-generation genomic sequencing on the Illumina HiSeq platform using the Anchored Phylogenomics protocol. This is described in more detail elsewhere (Lemmon et al., 2012; Lemmon and Lemmon, 2013), but we outline the protocol here. We used the Vertebrate 1.0 probe kit (Lemmon et al., 2012), which contains 512 loci generated from the genomes of 5 model organisms (*Homo, Gallus, Anolis, Xenopus,* and *Danio*), each designed to capture ~1500 bp (240 bp probe, with 25 120 bp tiles overlapping every 5 bp, covering ~700 bp in the flanking regions on either side of the probe). Thus, 56,664 probes target 122,800 bp of the genome, designed to capture ~800,000 bp of total sequence data per individual. Of these, not all can be expected to capture, enrich, and sequence properly for all individuals, so the final dataset will almost always be somewhat smaller than this best-case scenario.

Using the genomic DNA extractions, we prepared 775 bp inserts by fragmenting the genome using a Covaris sonicator, to which sample-specific barcodes and Illumina sequencing adapters were ligated (Lemmon et al., 2012). The inserts were pooled and enriched using an Agilent Custom SureSelect kit containing the probes as described above, which hybridize to the genomic DNA *in vitro*, and allow the regions of interest to be isolated. These enriched libraries were then sequenced using the 100 bp paired-end protocol in a single lane on the HiSeq 2000. For contig assembly, we also included the raw draft of the *Python* genome (Castoe et al., 2013), to represent a non-caenophidian outgroup, which was included in the pipeline described below.

The raw sequencing reads were filtered for quality then de-multiplexed (separated into the different species by index with no mismatches tolerated) and assembled using the quasi-de novo approach described in Lemmon et al. (2012) as follows. First, read pairs in which the two reads overlapped were merged following Rokyta et al. (2012). The high-sensitivity approach described in Lemmon et al. (2012) was then used to match reads to probe region sequences of *Anolis carolinensis*. The matching reads were then aligned for each locus using the following iterative procedure: (i) reads were sorted by the % match to the *Anolis*, reference sequence (in descending order), (ii) the position of each read was determined by maximizing the % match to the previous reads in the sorted list (reads failing to achieve at least a 90% match for a consecutive segment of at least 20 bp were not included in the assembly during that iteration), (iii) steps 1–2 were repeated until no additional reads could be aligned (source code available from ARL).

Following read assembly, consensus bases were called using the following rules: (i) "N" was called for all sites with less than three-fold coverage and variant sites with less than 10-fold coverage, (ii) the observed nucleotide was called for invariant sites with coverage between 3 and 10, and (iii) the most common base was called for sites with greater than 9-fold coverage, unless the distribution of observed bases was unlikely to have arisen under a two-allele model with equal allele frequencies (probability approximated as p = 1-pbinom[nMax,n,0.5], with n being equal to the number of unambiguous characters and nMax equaling the abundance of the most common base). The two-allele model was determined to be unlikely (and thus the most common base was called) when $p < 0.05$. Consensus sequences were then aligned for each locus using MUSCLE v. 3.8.31 with the default parameters (Edgar, 2004), and ambiguous regions or those containing large amounts of missing data were trimmed manually from the final alignments. A total of 160 loci were removed due to the presence of missing data after trimming.

This generated ~100× coverage per locus, per species, with a total of 333 loci, totaling 225,140 bp per species with 79,520 segregating sites, 26,043 singletons, and 119,577 identical sites. Missing data, 39,877 gaps (length variation or end loss) and 1398 'N,' accounted for only 0.6% of the total matrix. The 333 loci range in size from 253 bp to 1799 bp, with an average length of 676 bp. All loci are variable, and range in % pairwise identity from 85.7% to 99.3%, with an average of 93.8% (6.2% divergence). Heterozygosity was low (0.09%), with a GC content of 44.9%, and thus a slight AT bias, which may actually reduce incomplete lineage-sorting and gene conflict (Romiguier et al., 2013).

2.2. Phylogenetic inference

Essentially all previous studies have been based on concatenated analysis of a few nuclear or mitochondrial genes (<10), based on traditional ML or BI methods (e.g., Pyron et al., 2011; 5 genes, up to 5814 total bp per species, 67% average missing data per species). Here, we compare those strategies to explicit species-tree methods (using phylogenomic data). We thus used a variety of methods to infer the colubroid phylogeny from the dataset described above. We did not use the methods *BEAST (Heled and Drummond, 2010) or STEM (Kubatko et al., 2009), because the former has shown poor performance with phylogenomic-scale data (O'Neill et al., 2013), which was confirmed in this case by preliminary exploratory analyses. Both approaches also require parameterization of values such as ancestral effective-population size that cannot be easily assessed at the deep phylogenetic scale used here.

First, we used NJ methods on the entire dataset, in the Geneious Tree Builder (Biomatters Ltd.), assessing support with 100 BS replicates summarized in a consensus tree, to approximate the simplest possible phylogenetic inference. Second, we performed a standard Maximum Parsimony (MP) analysis using TNT, under the default settings with 100 BS replicates summarized in a consensus tree (Goloboff et al., 2008), for an analysis of all data that is putatively not influenced by model assumptions. Third, we analyzed the concatenated dataset (unpartitioned), using FastTree2.1 (Price et al., 2010), which has been shown to yield topological accuracy equivalent to RAxML in much less time (Liu et al., 2011). We assessed support using the SHL (Shimodaira-Hasegawa like) values estimated in FastTree, which have been shown to be equivalent to BS values for all but the shortest branches, and are more likely to be accurate for shorter branches (Anisimova et al., 2011; Pyron et al., 2011).

Fourth, for each of the 333 genes, we generated 100 bootstrap replicates in RAxML using the GTRGAMMA model (Stamatakis, 2006) for species-tree estimation. With each of these gene trees and estimates of error, we generated species trees with support using summary methods (MP-EST, STAR, and NJst). These methods are based on the relative rank order of divergences in gene trees to estimate the species tree (Liu and Yu, 2011; Liu et al., 2009b, 2010), and are available jointly in STRAW (Shaw et al., 2013), a web-server for estimating species trees (http://bioinformatics.publichealth.uga.edu/SpeciesTreeAnalysis/index.php). These methods also use the coalescent, but simplify assumptions to increase computational efficiency.

The MP-EST method takes rooted gene-trees and maximizes a pseudo-likelihood function of triplets in the species tree, with branch lengths in coalescent units. The STAR method uses the average ranks of coalescence times to infer the species tree based on a distance matrix of ranks. The NJst method is based on a distance

matrix of internodes across gene trees, which approximates the species tree under the coalescent. The latter two methods do not estimate meaningful branch-lengths. This pipeline, available from FTB, uses a Perl script to generate 100 bootstraps per locus using the RAxML GTRGAMMA model, packages the trees, and sends them to the STRAW server or a compiled version of those programs on a desktop or high performance computing cluster to generate species trees. With the aligned data, the completion of both bootstrapped gene-trees and the final species-trees took less than one day on a desktop computer.

Finally, we estimated the Extended Majority-Rule (EMR) consensus-tree from the ML gene trees using Rphylip (Revell, 2013), a wrapper for PHYLIP (Felsenstein, 2013). This approach simply shows all compatible clades found in the gene trees, even if they occur at a very low frequency, and have often been found to approximate the species tree and concatenated trees in recent genomic studies (Salichos and Rokas, 2013; Song et al., 2012). Consensus and concatenation have recently been assumed to be generally inferior to species-tree estimation (Degnan et al., 2009; Degnan and Rosenberg, 2009), but this has rarely been tested directly on the scale represented in these analyses.

2.3. Treespace

We compared concatenated gene-tree estimates and ML gene-trees to all three species trees and the EMR using unweighted Robinson–Foulds distances (Robinson and Foulds, 1981; Steel and Penny, 1993). This determines pairwise differences in topologies, and here assesses how close the topology of each gene tree is to the species tree, and yields a basic estimate of the information inherent in each gene with respect to the overall species tree. Additionally, we visualized the estimated treespace using multidimensional scaling (MDS), based on the un-weighted Robinson-Foulds distances of the estimated trees (Hillis et al., 2005). We combined the 100 per-gene BS replicates from the ML gene-tree searches used to estimate the coalescent-based species trees in STRAW (333 total). This allowed us to visualize the total gene-tree space. We used the TreeSetViz module in Mesquite v2.7.5 (Maddison and Maddison, 2010) to calculate the tree differences (55,444 total), as well as the majority-rule consensus of treespace. This approach only considers topological differences, not differences in estimated branch lengths or other parameters.

Finally, we plotted the net phylogenetic informativeness (Townsend, 2007), a rough measure of the per-locus power of a dataset to resolve a given node, over the time-scale of colubroid evolution. To do this, we scaled to the MP-EST topology to relative time, ultrametricizing it using the 'chronopl' command in APE (Paradis et al., 2004) with a scaling parameter $\lambda = 1$ (default). We then used the PhyDesign web-server with the DNARates algorithm (http://www.phydesign.townsend.yale.edu/) to calculate the power of each locus to resolve nodes over time, averaged across sites. This method assesses the power of genes in a prospective manner, to identify advantageous candidate loci for future analyses. Profiling in this way will allow us to determine the informativeness of the Anchored Phylogenomics probes over time, such as clusters of high- or low-power genes in different time-periods.

3. Results

3.1. Colubroid phylogeny

In general, all seven methods (NJ, MP, ML, EMR, MP-EST, NJst, and STAR) are fairly concordant in terms of both topology and support (Fig. 1, Figs. S1–S7), with a few notable differences. All families with multiple subfamilies are supported at ~100%, as are relationships between most families. The placement of Pareatidae, Viperidae, Homalopsidae, Elapidae + Lamprophiidae, and Colubridae is supported by all analyses. These results are generally similar to previous studies (Burbrink and Pyron, 2008; Grazziotin et al., 2012; Kelly et al., 2003; Lawson et al., 2005; Pyron and Burbrink, 2012; Pyron et al., 2011, 2013a; Vidal and Hedges, 2002; Wiens et al., 2008; Zaher et al., 2009). Importantly, the matrix generated in our study offers a test of these previous hypotheses with hundreds of loci and ~0% missing data, at least partially alleviating previous concerns about the reliability of matrices with a large proportion of ambiguous cells (Lemmon et al., 2009; Wiens and Morrill, 2011).

3.2. Monophyly of Caenophidia

In general, most methods support the monophyly of Caenophidia, as in previous studies (Pyron et al., 2013b; Vidal and Hedges, 2002; Wiens et al., 2008). However, both concatenated (NJ) and species-tree (NJst and STAR) place *Python* as the sister taxon to *Acrochordus* at 100% support. Most sparse concatenated analyses have not found this result, and the relationship only appears in 23% of the gene trees (Fig. 3a).

3.3. Monophyly of Colubroidea

A major difference between the concatenated and species-tree analyses is the monophyly of Colubroidea. Most analyses both support monophyly of Colubroidea as traditionally defined (with Xenodermatidae as the basal lineage) at 100%, in concordance with many previous studies (Pyron et al., 2011; Vidal and Hedges, 2002; Wiens et al., 2008). In contrast, this is only weakly supported (77%) in the NJst and STAR trees, and is not supported in a substantial fraction of the ML gene-trees (Fig. 3b). A monophyletic Colubroidea is found in 58% of the ML gene trees. Paraphyly of Colubroidea has also been found in previous analyses (Kelly et al., 2003; Pyron et al., 2013a) and was hypothesized by early researchers based on morphology (Boulenger, 1894), but was typically only weakly supported. Paraphyly of Colubroidea was also recovered in our preliminary exploratory analyses using *BEAST, which failed to converge fully with only 50 loci after >400 million generations across multiple runs (results not shown). We find that the non-monophyly of Colubroidea, with Acrochordidae as a sister to Xenodermatidae, is only inferred in ~10% of the ML gene trees. We also note that analysis of the dataset without *Python* yielded support for non-monophyly of Colubroidea (Acrochordidae + Xenodermatidae) from most methods (results not shown).

3.4. Placement of Homalopsidae

The family Homalopsidae is the sister group to (Elapidae + Lamprophiidae) + Colubridae, with strong support in all 7 analyses, similar to some previous studies (Lawson et al., 2005; Pyron and Burbrink, 2012; Vidal et al., 2007; Wiens et al., 2008). Previous studies that recovered a sister-group relationship between Homalopsidae and (Elapidae + Lamprophiidae) were generally poorly supported and based on fewer loci (Burbrink and Pyron, 2008; Pyron et al., 2011, 2013a). Different analytical methods (e.g., parsimony vs. likelihood) estimated variable placement of Homalopsidae in one previous study of a smaller dataset (Kelly et al., 2003). The strong support and concordance suggests that increased character sampling converges on what is presumably a more accurate placement of Homalopsidae as the sister group of (Elapidae + Lamprophiidae) + Colubridae, regardless of method.

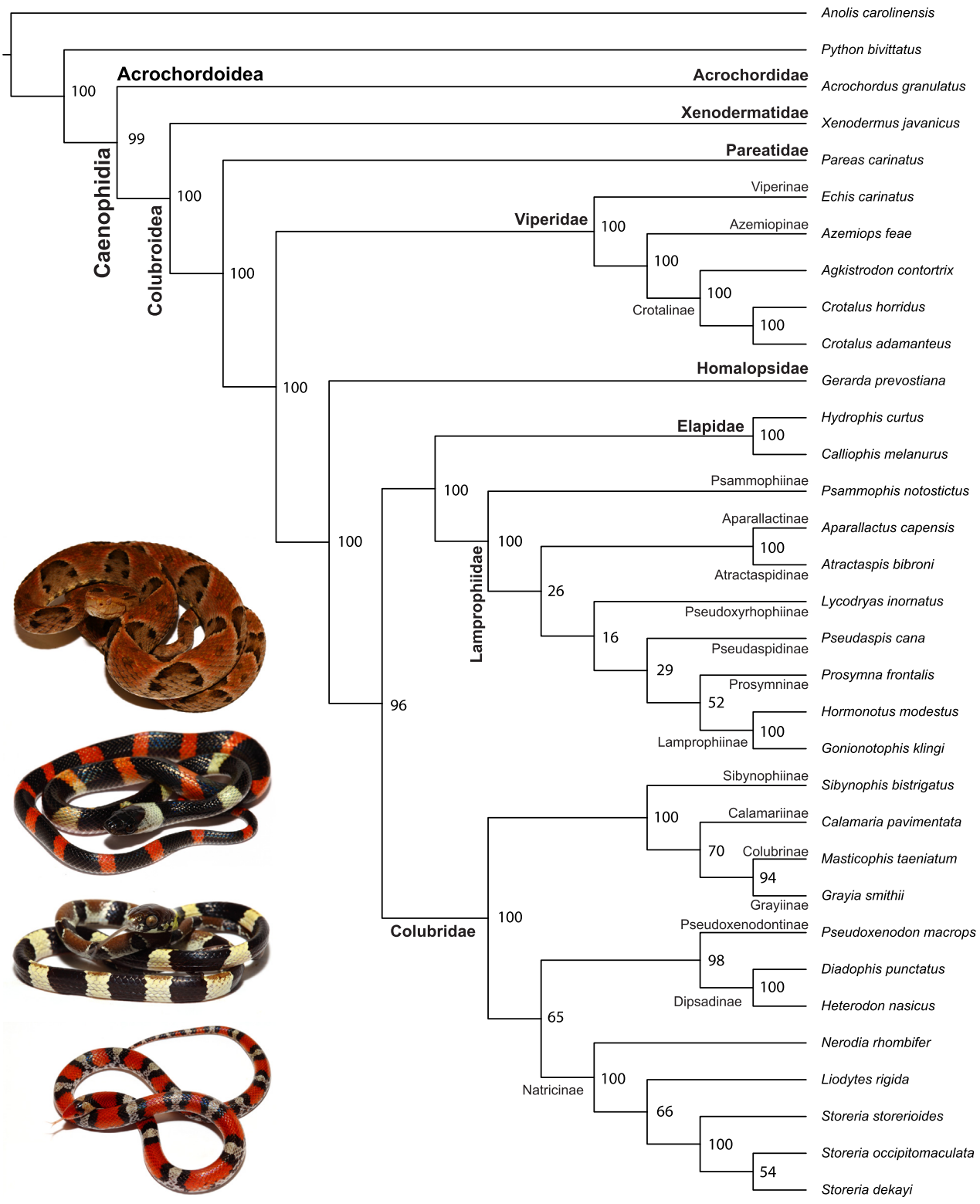

Fig. 1. Coalescent-based Maximum Pseudo-likelihood Estimation of the species tree (MP-EST; support from 100 BS replicates per locus = 33,300 total gene-trees). Branch lengths are measured in coalescent units (see Fig. S6), but are not shown here for topological clarity. Snakes shown from top to bottom are representatives from Viperidae (*Bothrops moojeni*), Dipsadinae (*Oxyrhopus petola* and *Dipsas sp.*), and Colubrinae (*Cemophora coccinea*). Photos by RAP.

3.5. Relationships in Lamprophiidae

All analyses support the monophyly of Lamprophiidae, as do most previous studies, outside of a few apparent rogue taxa such as *Micrelaps* and *Oxyrhabdium* (Pyron et al., 2013a). Within Lamprophiidae, however, relationships among the subfamilies differ between the methods, with little support for an overall resolution. The only consistent grouping is Atractaspidinae + Aparallactinae and monophyly of Lamprophiinae (*Hormonotus* + *Gonionotophis*), which are supported by all methods, as well as in

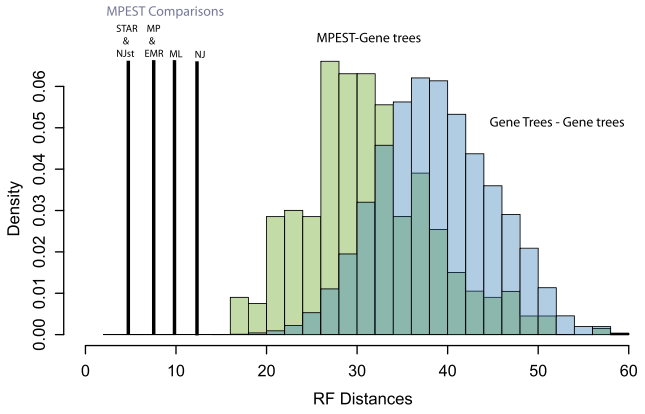

Fig. 2. Robinson-Foulds (RF) distances for among-gene tree comparisons (blue) and gene tree to species tree (MP-EST) comparisons (green). Differences between the MP-EST and other trees, including other ST methods and concatenation, are illustrated with lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

previous studies (e.g., Pyron et al., 2011). Relationships between this and other groups are essentially unresolved among methods (Figs. S1–7), as this group represents a rapid evolutionary radiation, with many short, difficult-to-resolve internodes between subfamilies (Kelly et al., 2009).

3.6. Relationships in Colubridae

In contrast to Lamprophiidae, relationships in Colubridae seem to be more consistent across methods and with previous studies, but are still not generally well-supported. All methods are concordant in recovering clades comprising (i) Sibynophiinae, Calamariinae, Colubrinae, and Grayiinae; (ii) Pseud oxenodontinae + Dipsadinae, and (iii) Natricinae. In all but the NJ analyses (Figs. S1–7), these form a group comprising i + (ii + iii), albeit with weak support. This basic topology has been recovered (at least partially, depending on the groups sampled) in several previous analyses (Kelly et al., 2003; Pyron et al., 2011; Wiens et al., 2008). The congruence here among methods, and between our phylogenomic dataset and previous mitochondrial analyses (Kelly et al., 2003) suggests that this topology is likely correct. As

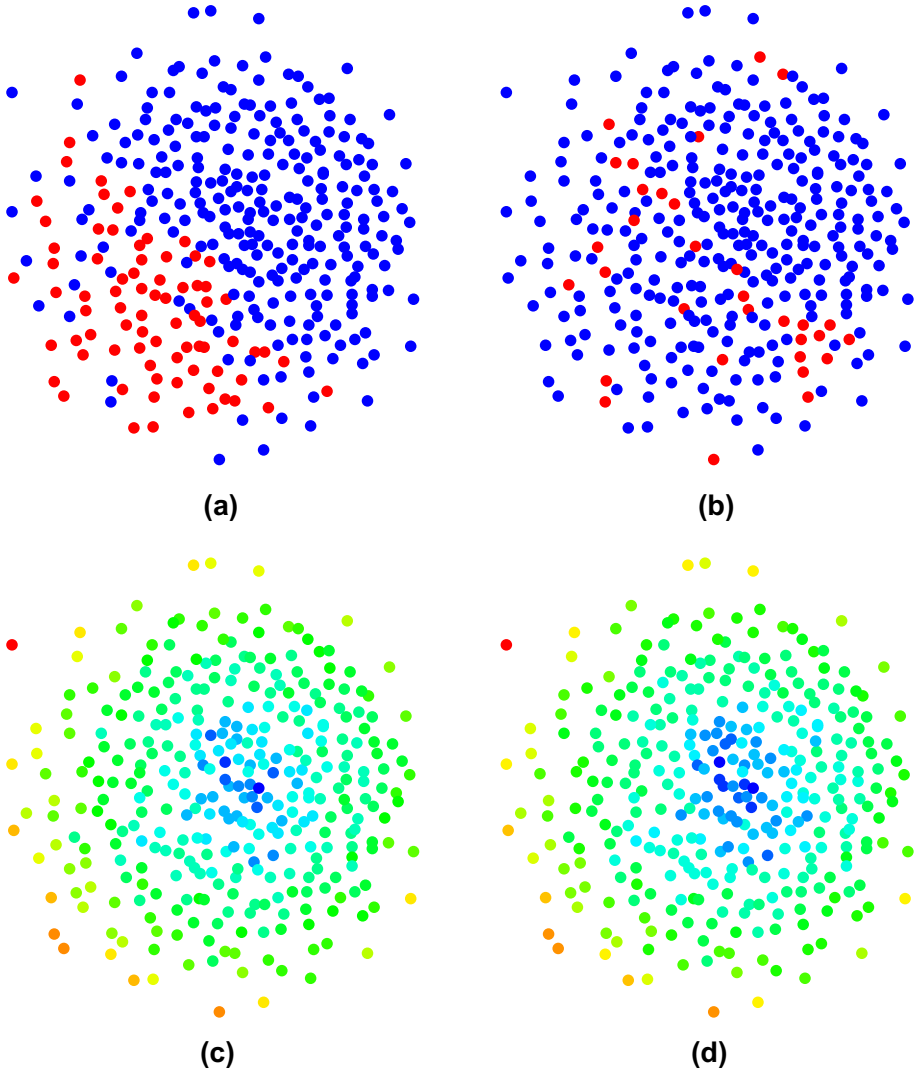

Fig. 3. Visualization of ML treespace using MDS for the 333 ML gene-trees generated from the individual loci. Red points show trees containing Acrochordidae + Pythonidae (a) or Acrochordidae + Xenodermatidae (b), whereas blue points indicate those taxa are not sister, though this does not necessarily mean the tree contains a monophyletic "Caenophidia" or "Colubroidea". Distance of each gene by RF is shown from the EMR (c) and MP-EST (d) estimates, with cooler colors showing closer topologies and warmer colors showing greater differences (see Fig. 2). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with Lamprophiidae, though, the number of short internodes due to an apparent rapid radiation makes this difficult to resolve conclusively. Other studies have identified a Dipsadinae + Colubrinae grouping (Vidal et al., 2007), which was also recovered in our NJ analysis. However, this topology is not well supported in any analysis. Relationships between Calamariinae, Colubrinae, Grayiinae, and Sibynophiinae continue to be poorly supported (Chen et al., 2013; Pyron et al., 2011).

3.7. Comparisons among species and gene trees

Each individual gene tree provided little information in comparison with the overall species-tree topology generated by any of the three methods (mean RF distance GT–ST = 32.3, SD = 7.2; Fig. 2) and were also extremely different from one another (mean among-gene-tree RF = 39.7, SD = 6.4). In comparison, gene trees were slightly more similar to the concatenated tree (mean RF = 32.1, SD = 7.1). The other methods revealed only small RF difference from MP-EST: STAR and NJst = 6, MP and EMR = 8, ML = 10, and NJ = 13, likely owing to the similarity of the rank-ordered methodologies. There is clearly little of the overall phylogenetic signal of the loci present in the individual gene-trees (Figs. 2, 3c–d, 4).

Considering the distribution of bootstrapped gene-trees to represent the most complete representation of treespace, we visualized the 333 ML gene-tree analyses in two-dimensional space using MDS. Contrary to some previous studies (Hillis et al., 2005), this did not yield clearly identifiable "islands" of treespace corresponding to the alternate topologies considered here (e.g., monophyly of Colubroidea; Fig. 3a). The treespace instead forms a single large spheroid, with alternate topologies such as Acrochordidae + Xenodermatidae and Acrochordidae + Pythonidae scattered throughout. The majority-rule consensus of these trees is similar to the EMR tree, suggesting that the individual gene-trees contain much of the signal for the alternate topologies (i.e., including the concatenates and species-tree estimates), but that none segregate clearly in the BS distribution (i.e., non-monophyly of Colubroidea).

3.8. Phylogenetic informativeness

The phylogenetic informativeness plot (Fig. 4) shows a strong early peak and leveling off of nearly all loci examined in this dataset. The Anchored Phylogenomics probes exhibit uniformly high power for resolving nodes across the time-scale of colubroid evolution, particularly near the root, where the contentious caenophidian and colubroid nodes occur. The ordinate in this plot is the "normalized asymptotic density for a true synapomorphy occurring in an asymptotically short, deep internode at historical time T of a quartet of taxa under an infinite-states Poisson model of character evolution," (http://www.phydesign.townsend.yale.edu/), roughly equivalent to the likelihood of informative characters (e.g., power to resolve the node) being present in that gene at that time-scale (Townsend, 2007). A small cluster of loci can be seen to peak during the colubroid radiation, and decrease in informativeness towards the root (Fig. 4), but most exhibit a consistently high net phylogenetic informativeness throughout the history of Colubroidea.

4. Discussion

4.1. Colubroid phylogeny and taxonomy

Coalescent-based species-tree estimates from phylogenomic data (Fig. 1) support the monophyly of Colubroidea as previously defined (Lawson et al., 2005; Pyron et al., 2011). In contrast, some previous concatenated analyses of nuclear and mitochondrial data (Kelly et al., 2003; Pyron et al., 2013a) supported paraphyly of Colubroidea (i.e., Acrochordidae + Xenodermatidae), as did preliminary species-tree analysis of our dataset without *Python* (not shown). This suggests that taxon sampling is still an important consideration (Heath et al., 2008; Zwickl and Hillis, 2002) for phylogenomic species-tree inference, in addition to the increased character-sampling (Graybeal, 1998).

Overall, the species trees and concatenated estimates are very similar to a recent analysis that included 4161 squamate species (Pyron et al., 2013a), suggesting that outgroup choice alone is not a determining factor in the species-tree relationships presented here. The mechanisms responsible for the discrepancy between the concatenated and species-tree analyses are unclear, but the Acrochordidae + Xenodermatidae and Acrochordidae + Pythonidae topologies are consistently recovered throughout the gene-tree space (Fig. 3). Thus, the inclusion of more outgroup taxa may alter some of these relationships in future analyses, but there is clearly signal among gene trees for non-monophyly of Caenophidia and Colubroidea. Therefore, monophyly of neither Caenophidia nor Colubroidea are certain, and we consider this issue still open for debate.

The infraorder Caenophidia (Hoffstetter, 1939), comprising Acrochordidae + Colubroidea, seems to be monophyletic based on most previous analyses (Pyron et al., 2013a; Wiens et al., 2012). As in previous analyses, the superfamily Colubroidea (Oppel, 1811) refers to the branch subtending Xenodermatidae (Gray, 1849), the stem lineage subtending the MRCA of Xenodermatidae + Colubridae and all descendants thereof. The sister group of this taxon is Acrochordidae (Bonaparte, 1831), which receives an equivalent but redundant rank of superfamily (Acrochordoidea).

As in most previous analyses (Pyron and Burbrink, 2012; Wiens et al., 2012), Homalopsidae is the sister taxon of (Elapidae + Lamprophiidae) + Colubridae (Fig. 1). As this topology is strongly supported in both concatenated and species-tree analyses, it seems likely to be correct (or at least, not overturned by any traditional phylogenetic analyses). Similarly, Lamprophiidae is strongly supported as a monophyletic group, excluding a few apparently rogue taxa such as *Micrelaps* and *Oxyrhabdium* (Lawson et al., 2005; Pyron et al., 2011). However, relationships among the constituent subfamilies are highly variable and poorly supported, suggesting a potentially intractable rapid radiation (Kelly et al., 2009; Pyron et al., 2013b).

In Colubridae, most analyses seem to converge on a consistent but weakly supported topology also estimated in previous studies (Pyron et al., 2011). Thus, it is unclear whether sampling either more characters or more taxa will result in additional resolution for Lamprophiidae or Colubridae. This can be tested in future analyses that sample more species using the Anchored Phylogenomics protocol, which can be combined with this dataset. Preliminary results (A.R. Lemmon, unpubl. data) suggest that longer loci captured using the Anchored Phylogenomics approach yield better resolution and support for gene trees, which translates into better resolution and support for species trees.

We use a consistent taxonomy that has been employed by numerous recent authors (Lawson et al., 2005; Pyron et al., 2011; Wiens et al., 2008). We note that the ranks applied to these groups are somewhat arbitrary, and have been assigned differently by other authors (Grazziotin et al., 2012; Kelly et al., 2009; Vidal et al., 2007). However, the composition of the various groups discussed here at the family level and below has not typically been in dispute, and are given in previous papers (Pyron et al., 2013a). Hence, while alternative classification systems may elevate Colubrinae to Colubridae for instance (concurrently increasing the

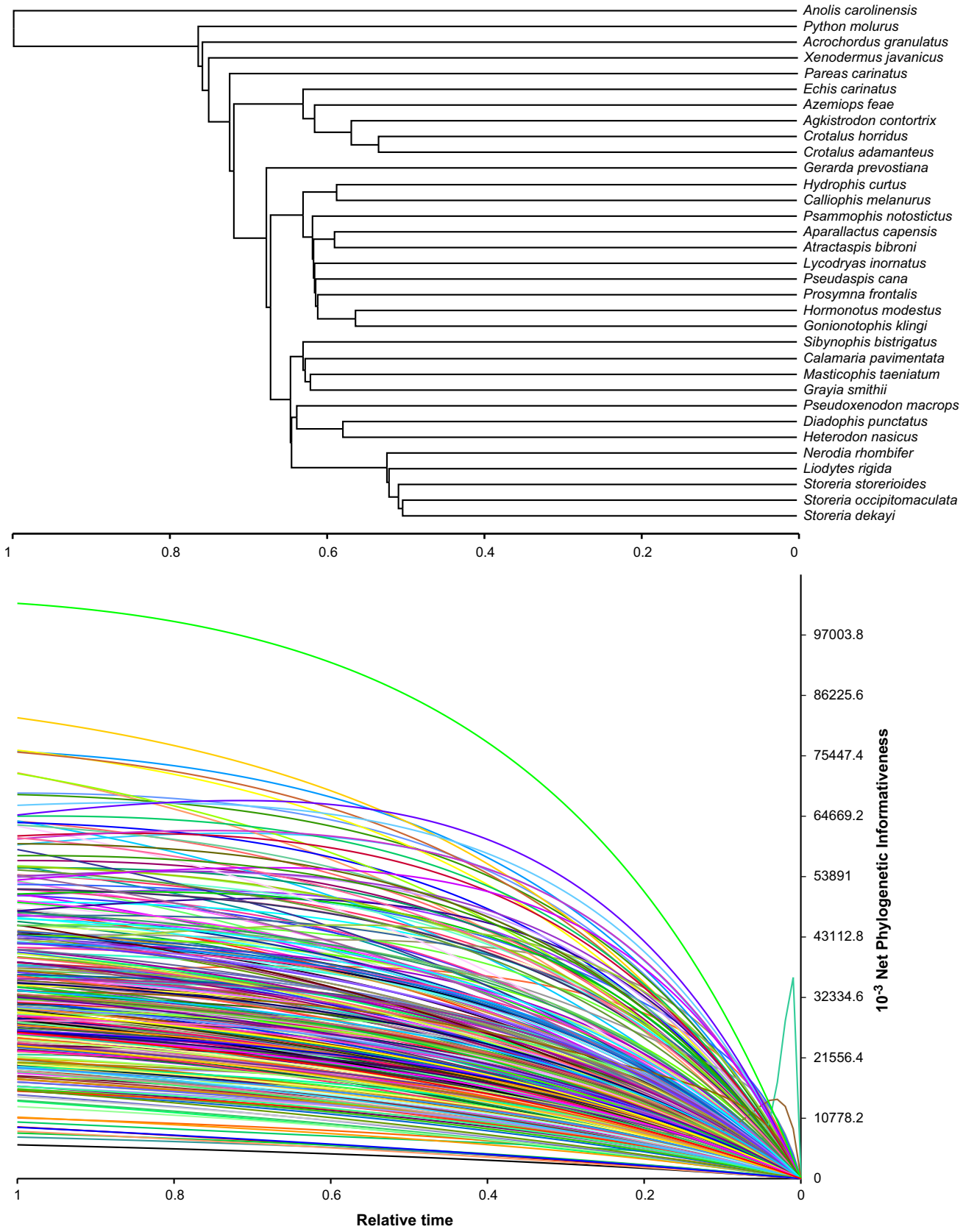

Fig. 4. Net phylogenetic informativeness plot for the 333 loci used here, plotted over the time-scale of the MP-EST tree scaled to absolute time. All loci show considerable power for resolving nodes over the time-scale of colubroid evolution, particularly near the root of Caenophidia.

ranks of the higher-level taxa), the species in those groups are unlikely to change.

The crucial points here are the phylogenetic relationships, which appear to be at least somewhat consistent for Homalopsidae and Colubridae, potentially intractable given available species-tree methods within Lamprophiidae, and poorly supported in the case of Caenophidia and Colubroidea (Fig. 1). Taxonomic histories and full contents of these groups are given in recent papers (Grazziotin et al., 2012; Pyron et al., 2011, 2013a; Vidal et al., 2007; Zaher et al., 2009). Future taxonomic revisions should be careful to consider these results when including advanced snakes in phylogenetic analyses, as concatenated and consensus-based gene-tree analyses seem likely to yield strongly supported topologies, apparently erroneously.

4.2. Phylogenomic species-tree estimation

The push for explicit analytical methods for estimating species trees by considering the independent genealogical histories of individual loci seems poised to offer a revolution in systematics (Edwards, 2009). Interestingly, the results here suggest that given a sufficient volume of phylogenetically informative data (Lemmon and Lemmon, 2013), most tree-building methods (gene trees, concatenation, consensus, species trees, etc.) will converge on what is presumably the "correct" answer for nodes that are not otherwise problematic (Dunn et al., 2008). For rapid radiations with multiple short internodes, species-tree methods based on the multi-species coalescent offer little advantage over neighbor-joining or maximum likelihood. These branches may be permanently intractable, at least using ordinary methods for phylogenetic reconstruction from DNA sequence data (Rokas et al., 2003). This is apparent in cases with (i) few taxa and genes (Vidal et al., 2007), (ii) many taxa and few genes (Pyron et al., 2011), and (iii) few taxa and many genes (results here). Hopefully in future studies, some intransigent nodes may be resolved by sampling many taxa and genes (and longer loci) and/or applying as yet undeveloped methods that account for more sources of topological error.

We underscore that most previous colubroid trees used at least one gene from the mitochondria. Here, our mtDNA-free species trees are similar to the basic structure of mtDNA trees (Kelly et al., 2003; Pyron et al., 2011). Importantly, we note that most nuclear genes are not topologically congruent relative to the species-tree topologies (Fig. 2). When genes are analyzed jointly in a species-tree framework that uses the rank ordered times, the resulting topology is remarkably similar to existing mtDNA trees and to concatenated gene-trees. This may also be strengthened in future studies if longer loci are sequenced. All the loci examined here exhibit a uniformly high phylogenetic informativeness across the time-scale of colubroid evolution (Fig. 4), the implication being that they will also do so in other groups.

Similarly, while NJ, ML, and ST approaches may yield congruent results >90% of the time, the occasional well-supported discordances underscore the need for careful data analysis and explicit species-tree reconstruction (Edwards, 2009). This point is illustrated here by the potential non-monophyly of Caenophidia and Colubroidea as traditionally defined, and the placement of Homalopsidae. Gene-tree methods can estimate the "wrong" topology (Fig. 3), regardless of the number of loci (Degnan et al., 2009).

We acknowledge that this point of view assumes that the species-tree results are correct, and most of the other concatenated and gene-tree analyses are wrong. The exact evolutionary mechanisms (e.g., incomplete lineage-sorting) for this discordance among gene trees (Fig. 3) are unclear at present. It is also unclear what effects this type of discordance might have on phylogenetic comparative analyses (Burbrink and Pyron, 2011) and other downstream applications of the phylogeny and branch lengths. Nonetheless, these results reinforce the observation that strong support from multiple genes, either alone or in concatenation, is not necessarily a guarantee that a given node is correct, without consideration of the coalescent.

Furthermore, fully satisfactory approaches for phylogenomic estimation of species trees are still elusive. Methods based on full-likelihood approaches to the multi-species coalescent (*BEAST) currently appear unable to deal with datasets containing large numbers of genes and even modest numbers of taxa (O'Neill et al., 2013). Of the three approaches employed here, only MP-EST returns meaningful branch-lengths in coalescent units (Shaw et al., 2013). Even still, these cannot be estimated for terminal branches when only a single species is sampled per lineage (as is the case for many lineages here). They can also only be converted to a more intuitive unit of "substitutions per site" if the population-genetic parameter θ is known. Similarly, STEM (Kubatko et al., 2009) can estimate a species tree with topology and meaningful branch-lengths, but only if θ is constant across the tree, conditions unlikely to be met in most datasets.

However, these approaches do apparently give a quick and robust assessment of the species tree topology, particularly for large-scale datasets such as the one presented here (DeGiorgio and Degnan, 2014), even when individual loci have comparatively little signal (Lanier et al., 2014). A major goal for future species-tree research should be the estimation of robust and accurate branch lengths that can easily be scaled to intuitive units such as substitutions per site or millions of years. Such estimates are currently available in *BEAST, which proved to not be useful for a dataset of this scale. These estimates are crucial for downstream phylogenetic comparative analyses, which are sensitive to errors in branch lengths (Burbrink and Pyron, 2011; Diaz-Uriarte and Garland, 1998; Wertheim and Sanderson, 2011).

4.3. Future directions

Phylogenomics may not have brought us closer to a stable resolution of the Tree of Life (at least, not yet), but molecular and computational techniques now allow for a fast, quick, and cheap evaluation of phylogenetic hypotheses on a previously unprecedented scale. Given the availability of genome-scale datasets and the computational simplicity of many new species-tree methods, analyses using an explicit coalescent framework should be prerequisite for future studies interested in establishing branching patterns in the Tree of Life. The potential for nodes, even those supported at 100% by most genes, to shift radically when analyzed using the coalescent, should necessitate the use of such approaches in a thorough analysis. Not all changes in topology are likely to be as drastic as the ones observed here, nor must one assume that the species tree is correct and the other analyses are wrong. In contrast, congruence will hopefully be the most common outcome, allowing for a more robust consensus on topology. Newly developed phylogenomic methods and the ability to gather extremely large datasets with ~0% missing data at a rapid pace and low cost is ushering in a new era of phylogenetic inference.

Acknowledgments

This research was supported in part by a U.S. National Science Foundation Grant to RAP (DBI-0905765) and to FTB (DEB-1257926). We would like to thank A. Resetar (FMNH), J. Vindum (CAS), R. Brown (KU), A. Bauer (VU), J. Losos (MCZ), J. McGuire (MVZ), C. Schmidt (FHSM), and R. Somaweera, T. LaDuc, and D. Shepard for access to tissue specimens. Data are available in DataDryad repository http://dx.doi.org/10.5061/dryad.rb5nc as a single concatenated alignment, with associated partition data.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2014.08.023.

References

Anisimova, 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.

Bonaparte, C.L.J.L., 1831. Saggio di una distribuzione metodia degli animali vertebrati. Giornale Arcadico di Scienze Lettere ed Arti 49, 3–77.

Boulenger, G.A., 1894. Catalogue of Snakes in the British Museum. British Museum of Natural History, London.

Burbrink, F.T., Pyron, R.A., 2008. The taming of the skew: estimating proper confidence intervals for divergence dates. Syst. Biol. 57, 317–328.

Burbrink, F.T., Pyron, R.A., 2011. The impact of gene-tree/species-tree discordance on diversification-rate estimation. Evolution 65, 1851–1861.

Castoe, T.A., de Koning, A.P., Hall, K.T., Card, D.C., Schield, D.R., Fujita, M.K., Ruggiero, R.P., Degner, J.F., Daza, J.M., Gu, W., Reyes-Velasco, J., Shaney, K.J., Castoe, J.M., Fox, S.E., Poole, A.W., Polanco, D., Dobry, J., Vandewege, M.W., Li, Q., Schott, R.K., Kapusta, A., Minx, P., Feschotte, C., Uetz, P., Ray, D.A., Hoffmann, F.G., Bogden, R., Smith, E.N., Chang, B.S., Vonk, F.J., Casewell, N.R., Henkel, C.V., Richardson, M.K., Mackessy, S.P., Bronikowski, A.M., Yandell, M., Warren, W.C., Secor, S.M., Pollock, D.D., 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proc. Natl. Acad. Sci. USA 110, 20645–20650.

Chen, X., Huang, S., Guo, P., Colli, G.R., de Oca, A.N.M., Vitt, L.J., Pyron, R.A., Burbrink, F.T., 2013. Understanding the formation of ancient intertropical disjunct distributions using Asian and Neotropical hinged-teeth snakes (Sibynophis and Scaphiodontophis: Serpentes: Colubridae). Mol. Phylogenet. Evol. 66, 254–261.

DeGiorgio, M., Degnan, J.H., 2014. Robustness to divergence time underestimation when inferring species trees from estimated gene trees. Syst. Biol. 63, 66–82.

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. Plos. Genet. 2, 762–768.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340.

Degnan, J.H., DeGiorgio, M., Bryant, D., Rosenberg, N.A., 2009. Properties of consensus methods for inferring species trees from gene trees. Syst. Biol. 58, 35–54.

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6, 361–375.

Diaz-Uriarte, R., Garland, T., 1998. Effects of branch length errors on the performance of phylogenetically independent contrasts. Syst. Biol. 47, 654–672.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sorensen, M.V., Haddock, S.H.D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745–U745.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32, 1792–1797.

Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19.

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

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726.

Felsenstein, J., 2013. PHYLIP (Phylogeny Inference Package).

Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. Cladistics 24, 774–786.

Gray, J.E., 1849. Catalogue of the Specimens of Snakes in the Collection of the British Museum. Newman, London.

Graybeal, A., 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47, 9–17.

Grazziotin, F.G., Zaher, H., Murphy, R.W., Scrocchi, G., Benavides, M.A., Zhang, Y.-P., Bonatto, S.L., 2012. Molecular phylogeny of the New World Dipsadidae (Serpentes: Colubroidea): a reappraisal. Cladistics 28, 437–459.

Heath, T.A., Hedtke, S.M., Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46, 239–257.

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

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

Hoffstetter, R., 1939. Contribution à l'étude des Elapidae actuels et fossiles et de l'ostéologie des ophidiens. Archives du Muséum d'Histoire Naturelle de Lyon 15, 1–78.

Kelly, C.M.R., Barker, N.P., Villet, M.H., 2003. Phylogenetics of advanced snakes (Caenophidia) based on four mitochondrial genes. Syst. Biol. 52, 439–459.

Kelly, C.M.R., Barker, N.P., Villet, M.H., Broadley, D.G., 2009. Phylogeny, biogeography and classification of the snake superfamily Elapoidea: a rapid radiation in the late Eocene. Cladistics 25, 38–63.

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

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

Lanier, H.C., Huang, H., Knowles, L.L., 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. Mol. Phylogenet. Evol. 70, 112–119.

Lawson, R., Slowinski, J.B., Crother, B.I., Burbrink, F.T., 2005. Phylogeny of the Colubroidea (Serpentes): new evidence from mitochondrial and nuclear genes. Mol. Phylogenet. Evol. 37, 581–601.

Lemmon, E.M., Lemmon, A.R., 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44.

Lemmon, A.R., Brown, J.M., Stanger-Hall, K., Lemmon, E.M., 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and bayesian inference. Syst. Biol. 58, 130–145.

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61, 727–744.

Liu, L., Yu, L.L., 2011. Estimating species trees from unrooted gene trees. Syst. Biol. 60, 661–667.

Liu, L., Yu, L.L., Kubatko, L., Pearl, D.K., Edwards, S.V., 2009a. Coalescent methods for estimating phylogenetic trees. Mol. Phylogenet. Evol. 53, 320–328.

Liu, L., Yu, L.L., Pearl, D.K., Edwards, S.V., 2009b. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58, 468–477.

Liu, L.A., Yu, L.L., Edwards, S.V., 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10.

Liu, K., Linder, C.R., Warnow, T., 2011. RAxML and FastTree comparing two methods for large-scale maximum likelihood phylogeny estimation. Plos One 6.

Maddison, W.P., Maddison, D.R., 2010. Mesquite: a modular system for evolutionary analysis. Version 2.72, http://mesquiteproject.org.

O'Neill, E.M., Schwartz, R., Bullock, C.T., Williams, J.S., Shaffer, H.B., Aguilar-Miguel, X., Parra-Olea, G., Weisrock, D.W., 2013. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (Ambystoma tigrinum) species complex. Mol. Ecol. 22, 111–129.

Oppel, M., 1811. Die Ordnungen, Familien und Gattungen der Reptilien, als Prodrom einer Naturgeschichte derselben. J. Lindauer, München.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. Plos One 5.

Pyron, R.A., Burbrink, F.T., 2012. Extinction, ecological opportunity, and the origins of global snake diversity. Evolution 66, 163–178.

Pyron, R.A., Burbrink, F.T., Colli, G.R., de Oca, A.N.M., Vitt, L.J., Kuczynski, C.A., Wiens, J.J., 2011. The phylogeny of advanced snakes (Colubroidea), with discovery of a new subfamily and comparison of support methods for likelihood trees. Mol. Phylogenet. Evol. 58, 329–342.

Pyron, R.A., Burbrink, F.T., Wiens, J.J., 2013a. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol. Biol. 13, 93.

Pyron, R.A., Kandambi, H.K.D., Hendry, C.R., Pushpamal, V., Burbrink, F.T., Somaweera, R., 2013b. Genus-level phylogeny of snakes reveals the origins of species richness in Sri Lanka. Mol. Phylogenet. Evol. 66, 969–978.

Revell, L.J., 2013. Rphylip: an R interface for PHYLIP R package.

Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. Math. Biosci. 53, 131–147.

Rokas, A., Carroll, S.B., 2006. Bushes in the tree of life. Plos Biol. 4, 1899–1904.

Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798–804.

Rokyta, D.R., Lemmon, A.R., Margres, M.J., Aronow, K., 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (Crotalus adamanteus). BMC Genom. 13.

Romiguier, J., Ranwez, V., Delsuc, F., Galtier, N., Douzery, E.J.P., 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. Mol. Biol. Evol. 30, 2134–2144.

Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 317–+.

Shaw, T.I., Ruan, Z., Glenn, T.C., Liu, L., 2013. STRAW: species tree analysis web server. Nucl. Acids Res. 41, W238–W241.

Song, S., Liu, L., Edwards, S.V., Wu, S.Y., 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc. Natl. Acad. Sci. USA 109, 14942–14947.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

Steel, M.A., Penny, D., 1993. Distributions of tree comparison metrics – some new results. Syst. Biol. 42, 126–141.

Townsend, J.P., 2007. Profiling phylogenetic informativeness. Syst. Biol. 56, 222–231.

Vidal, N., Hedges, S.B., 2002. Higher-level relationships of caenophidian snakes inferred from four nuclear and mitochondrial genes. Cr Biol. 325, 987–995.

Vidal, N., Delmas, A.S., David, P., Cruaud, C., Coujoux, A., Hedges, S.B., 2007. The phylogeny and classification of caenophidian snakes inferred from seven nuclear protein-coding genes. Cr Biol. 330, 182–187.

Vitt, L.J., Caldwell, J.P., 2009. Herpetology. Elsevier, Burlington, MA.

Wertheim, J.O., Sanderson, M.J., 2011. Estimating diversification rates: how useful are divergence times? Evolution 65, 309–320.

Wiens, J.J., Morrill, M.C., 2011. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. Syst. Biol. 60, 719–731.
Wiens, J.J., Kuczynski, C.A., Smith, S.A., Mulcahy, D.G., Sites, J.W., Townsend, T.M., Reeder, T.W., 2008. Branch lengths, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. Syst. Biol. 57, 420–431.
Wiens, J.J., Hutter, C.R., Mulcahy, D.G., Noonan, B.P., Townsend, T.M., Sites, J.W., Reeder, T.W., 2012. Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. Biol. Lett. 8, 1043–1046.
Zaher, H., Grazziotin, F.G., Cadle, J.E., Murphy, R.W., de Moura, J.C., Bonatto, S.L., 2009. Molecular phylogeny of advanced snakes (Serpentes, Caenophidia) with an emphasis on South American Xenodontines: a revised classification and descriptions of new taxa. Papeis Avulsos de Zoologia 49, 115–153.
Zaher, H., Grazziotin, F.G., Graboski, R., Fuentes, R.G., Sánchez-Martinez, P., Montingelli, G.G., Zhang, Y.-P., Murphy, R.W., 2012. Phylogenetic relationships of the genus Sibynophis (Serpentes: Colubroidea). Papéis Avulsos de Zoologia (São Paulo) 52, 141–149.
Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51, 588–598.