Phylogenomic analyses of more than 4000 nuclear loci resolve the origin of snakes among lizard families

Jeffrey W. Streicher1,2 and John J. Wiens1

1Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721-0088, USA
2Department of Life Sciences, The Natural History Museum, London SW7 5BD, UK

Received: 17 June 2017
Accepted: 18 August 2017

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3865459.

1. Introduction

Squamates, lizards and snakes, are a particularly important group. They include more than 10 000 extant species [1], surpassing birds as the most diverse major tetrapod clade. They are used as study systems in many areas of research, including (among many) the evolution of sex, viviparity, herbivory, body form and toxins [2]. Venomous squamates also pose a serious hazard to humans in some regions [3], whereas their venoms offer many resources for medical research [4].

There has been considerable recent research on higher-level squamate phylogeny, but some branches have remained contentious. Molecular analyses have supported a very different set of relationships relative to those suggested by morphology. Morphological analyses have placed iguanian lizards at the base of the squamate tree [5,6]. By contrast, molecular analyses have placed gekkotans and dibamids at the base, and iguanians in a more recent clade with snakes and anguimorphs [7–11]. Combined analyses of molecular and morphological data have also supported the molecular trees [12,13], and have shown that the support for the morphological tree is problematic. Nevertheless, some authors have continued to question the placement of iguanians by molecular data [6,14,15].

Further, some key aspects of higher-level squamate phylogeny have remained uncertain, even in molecular analyses. One is the placement of snakes. Many molecular analyses have placed snakes as the sister to a clade
unifying iguanians and anguimorphs, but the support for the latter clade has been consistently weak [8–11]. Furthermore, previous studies generally used concatenated analyses, rather than species-tree analyses. Concatenated analyses may yield weak or even positively misleading results when there is extensive conflict among gene trees, whereas species-tree methods can yield well-supported and accurate results [16]. Such conflicts among gene trees are well documented for many major branches of squamate phylogeny [9]. Another key area of uncertainty is whether the sister group to all other squamates is dibamids [7,10,11] or a clade consisting of dibamids and gekkotans [9,12,13].

Here, we analyse higher-level squamate relationships using phylogenomic data and species-tree methods. We obtain nuclear data from 4178 loci, approximately 100 times more than previous analyses of higher-level squamate phylogeny [9,11,13]. We strongly support snakes as the sister group of a well-supported clade uniting Iguania + Anguimorpha. Thus, our analyses simultaneously resolve these two persistent controversies in squamate phylogeny. Overall, we provide a generally well-supported estimate of higher-level squamate phylogeny using phylogenomic data and species-tree methods.

2. Material and methods

We sampled 32 squamate species and five outgroup taxa, which included representatives of all major amniote clades (i.e., rhynchocephalians, crocodylians, birds, turtles and mammals). Sampling within squamates included all major groups and most families (outside of snakes, iguanians and amphisbaenians). Details of sampling are given in the electronic supplementary material, table S1.

We used anchored enrichment to sequence ultraconserved elements (UCEs, [17]) and obtained an average of 2738.7 UCEs per species. Laboratory methods (following [18,19]) are listed in the electronic supplementary material, appendix S1. Data for most outgroups (all but Sphenodon) and nine ingroup taxa were from previous studies (electronic supplementary material, table S1). We obtained new UCE data for 24 taxa. In theory, we could have included additional published data for iguanian lizards [18] and snakes [19]. However, to make the analyses more computationally tractable, we included only three species from each clade, representing major lineages (i.e., scolecocephalian, pythonid, and colubroid snakes; pleurodont and acrodont iguanians). We included the most complete taxon available within each major lineage.

We cleaned, assembled, identified and aligned UCE data using the software packages velvet 1.2.10 [20] and PhyInCE 2.0.0 [17,21]. We included UCEs with up to 50% missing taxa, which appears to give optimal results [18]. The final dataset had 4178 UCEs (1 530 015 total aligned base pairs), with 52.4% missing data overall (including gaps). The average number of variable sites per UCE was 108.3 (range 17–455) and mean UCE length was 380.0 base pairs (range 108–878). Raw FASTQ files for new samples were uploaded to the NCBI sequence read archive (electronic supplementary material, table S1). Alignments and gene trees for all analyses are available via the Dryad Digital Repository [22].

Phylogenetic analyses were conducted using two approaches: (i) concatenated maximum likelihood, and (ii) species-tree analysis using gene trees as input (i.e. one best likelihood tree per UCE). To perform the concatenated analysis and infer gene trees, we used RAxML 8.2.9 [23] with default settings and the GTRCAT model applied to the entire alignment (partitions are not generally appropriate for UCE data, given that most loci are not protein coding [17]). The concatenated analyses used the CIPRES server [24] and branch support was assessed with bootstrapping. Species-trees analyses used the program ASTRAL 4.10.11 [25,26], with branch support estimated using local posterior probabilities [27].

We generally included only taxa with more than 1000 UCEs. However, we included two more incomplete taxa that were potentially important for the tree’s base: the dibamid Anelytropsis (90 UCEs; electronic supplementary material, table S1) and the closest squamate outgroup (Sphenodon; 596 UCEs, electronic supplementary material, table S1). Although Anelytropsis is highly incomplete, its inclusion or exclusion had no impact on the estimated topologies (see the electronic supplementary material, figures S1 and S2).

3. Results

The concatenated and species-tree analyses provided strong and concordant support for most higher-level squamate relationships (figure 1; electronic supplementary material, figure S3). There was strong support for a clade uniting anguimorphs, iguanians, and snakes (Toxicofera) and strong support for a clade linking anguimorphs and iguanians. Thus, these results strongly resolve the controversial placement of snakes and iguanians with species-tree methods and a dataset of unprecedented size. There was strong, concordant support for most other relationships across the tree, such as the clade uniting scincids, xantusiids, gerrhosaurids and cordylids (Scincioidea), and the clade uniting gymnophthalmids, teids, lacertids, and amphisbaenians (Lacertioidea). The backbone relationships among these major clades were also strongly supported.

There were also a few remaining areas of uncertainty. Both approaches agreed that dibamids and gekkotans are at the base of the squamate tree, but differ as to whether dibamids are sister to all other squamates (ASTRAL), or whether dibamids and gekkotans are sister taxa (concatenated). The conflicting relationships were moderately supported by each approach. There were also conflicts (and weak support from ASTRAL) among some gekkotan families. Finally, there were conflicting results and weak support regarding monophyly of amphisbaenians relative to lacertids.

4. Discussion

In this study, we assembled a large phylogenomic dataset to address higher-level squamate phylogeny, using species-tree methods and sampling approximately 100 times more loci than previous studies. Our results strongly resolve the phylogenetic placement of snakes among lizard families, as the sister group to anguimorphs and iguanians. Concomitantly, they provide overwhelming evidence to conclusively resolve the contentious placement of iguanians. We acknowledge that some readers might consider the placement of snakes and iguanians as already known, but previous analyses had only weak support and generally used concatenated analyses [7–13]. Some authors have also argued for a snake + anguimorph clade instead [28]. More generally, we provide strong support for most higher-level relationships across squamates. We note that Squamata may be among the first major groups of vertebrates to have their higher-level relationships largely resolved by species-tree analyses of thousands of loci (along with birds [29]).
Nevertheless, we acknowledge three areas of remaining uncertainty in our tree. Two involve lower-level relationships, and can probably be resolved with increased taxon sampling (e.g. relationships among gekkotan families and among lacertids and amphisbaenians). Thus, adding more taxa within these groups should help subdivide long branches associated with these problematic lower-level relationships. The third area (gekkotans, dibamids, other squamates) is more difficult. It is unclear if adding more gekkotans or dibamids can significantly shorten the relevant branches. Resolving these relationships might require major methodological advances, or even more loci. Intriguingly, we note that the relationships for these taxa from ASTRAL are consistent with concatenated analyses of fewer taxa and loci [9,13].

In summary, our results here provide strong support for most higher-level squamate relationships. We resolve two contentious aspects of squamate phylogeny using approximately 100 times more loci than in previous studies, and with species-tree methods. Although some areas of uncertainty remain, squamates may now be among the most phylogenetically well-resolved of the major tetrapod clades.

Data accessibility. All data are accessible via Dryad (http://dx.doi.org/10.5061/dryad.66gt5) [22].

Authors’ contributions. J.W.S. and J.J.W. designed the study and wrote the paper, J.W.S. collected data and performed analyses, all authors provided tissues used here (listed in the electronic supplementary material, table S1), but especially T. Reeder. Funding was provided by the University of Arizona. The authors declare no competing interests. Funding. Funding was provided by the University of Arizona.

Acknowledgements. We thank the many individuals and institutions that provided tissues used here (listed in the electronic supplementary material, table S1), but especially T. Reeder.

References


Figure 1. Phylogeny of squamate reptiles inferred from a species-tree analysis (ASTRAL) of 4178 nuclear loci and 1 530 015 aligned base pairs. Tree is rooted with Homo. Light blue highlighting indicates clades congruent with the concatenated maximum-likelihood analysis (see the electronic supplementary material, figure S3). Grey highlighting indicates incongruent clades. Support values are: local posterior probabilities (species tree)/bootstrap support (concatenated). NS indicates ‘no support’ from the concatenated analysis (incongruent branch). When no numbers are shown, support values are equal to 1.0/100 (maximum support from both methods).