More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs

Nicholas G. Crawford1,*, Brant C. Faircloth2, John E. McCormack3, Robb T. Brumfield3,4, Kevin Winker5 and Travis C. Glenn6

1Department of Biology, Boston University, Boston, MA 02215, USA
2Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA
3Museum of Natural Science, and 4Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA
5University of Alaska Museum, 907 Yukon Drive, Fairbanks, AK 99775, USA
6Department of Environmental Health Science and Georgia Genomics Facility, University of Georgia, Athens, GA 30602, USA

*Author for correspondence (ngcrawford@gmail.com).

We present the first genomic-scale analysis addressing the phylogenetic position of turtles, using over 1000 loci from representatives of all major reptile lineages including tuatara. Previously, studies of morphological traits positioned turtles either at the base of the reptile tree or with lizards, snakes and tuatara (lepidosaurs), whereas molecular analyses typically allied turtles with crocodiles and birds (archosaurs). A recent analysis of shared microRNA families found that turtles are more closely related to lepidosaurs. To test this hypothesis with data from many single-copy nuclear loci dispersed throughout the genome, we used sequence capture, high-throughput sequencing and published genomes to obtain sequences from 1145 ultraconserved elements (UCEs) and their variable flanking DNA. The resulting phylogeny provides overwhelming support for the hypothesis that turtles evolved from a common ancestor of birds and crocodilians, rejecting the hypothesized relationship between turtles and lepidosaurs.

Keywords: turtles; ultraconserved elements; phylogenomics; evolution; archosaurs

1. INTRODUCTION

The evolutionary origin of turtles has confounded the understanding of vertebrate evolution [1] (figure 1). Historically, turtles were thought to be early-diverging reptiles, called anapsids, based on their skull morphology and traits such as dermal armour [2]. Recent morphological studies that included soft tissue and developmental characters [3] allied turtles with lepidosaurs, a group including squamates (lizards and snakes) and tuatars. However, homoplasy stemming from the derived skeletal specializations of turtles limits the utility of phylogenetic inference based on morphological data to resolve turtle placement [4,5].

Molecular studies using mitochondrial [4,6–8,16] and nuclear DNA [5,9–14,17] typically place turtles sister to archosaurs (crocodilians and birds; figure 1). This molecular hypothesis was recently contradicted by a phylogeny reconstructed from microRNAs [15] that allied turtles with lepidosaurs. Lyson et al. [15] suggested that prior molecular evidence for a turtle–archosaur relationship may be the result of analytical artefacts. If true, the hypothetical relationship between turtles and lepidosaurs (Ankylopoda) should appear throughout the genomes of these organisms.

Here, we test the Ankylopoda hypothesis and address the evolutionary origin of turtles. We reconstruct a reptile phylogeny using ultraconserved elements (UCEs) [18] and their flanking sequence that we obtained using sequence capture of DNA from a tuatara and two species each of crocodilians, squamates and turtles (table 1). We used UCEs because they are easily aligned portions of extremely divergent genomes [19], allowing many loci to be interrogated across evolutionary timescales, and because sequence variability within UCEs increases with distance from the core of the targeted UCE [20], suggesting that phylogenetically informative content in flanking regions can inform hypotheses spanning different evolutionary timescales. To break up long branches and mitigate potential problems with long-branch attraction, we selected species representing the span of diversity within major reptilian lineages (i.e. the most divergent crocodilians, lepidosaurs and turtles).

2. MATERIAL AND METHODS

We enriched DNA libraries prepared with Nextera kits (Epicentre, Inc., Madison, WI, USA) using a synthesis (Microarray, Inc., Ann Arbor, MI, USA or Agilent, Inc., Santa Clara, CA, USA) of RNA probes [20] targeting 2386 UCEs and their flanking sequence. We generated sequences for each enriched library using single-end, 100-base sequencing on an Illumina GAIIx. After quality filtering, we assembled reads into contigs using Velvet [21], and we matched contigs to the UCE loci, removing duplicate hits. We generated alignments using MUSCLE [22], and we excluded loci having missing data in any taxon. Following alignment, we estimated the appropriate finite-sites substitution model for each locus using MrAIC.

We prepared a concatenated dataset by partitioning loci by substitution model prior to analysis using two runs of MrBayes [23] for 5 000 000 iterations (four chains per run; burn-in: 50%; thinning: 100). We also used each alignment to estimate gene trees incorporating 1000 multi-locus bootstrap replicates, which we integrated into STEAC and STAR [24] species trees. Additional details concerning UCE sequence capture methods and phylogenetic methods are available in Faircloth et al. [20].

3. RESULTS

We enriched genomic DNA for UCEs in corn snake (Pantherophis guttata), African helmeted turtle (Pelomedusa subrufa), painted turtle (Chrysemys picta), American alligator (Alligator mississippiensis), saltwater crocodile (Crocodylus porosus) and tuatara (Sphenodon tuatara) (table 1). We sequenced a mean of 4.9 million reads from each library, and from these reads, we assembled an average of 2648 (± 314 s.d.) contigs.

We supplemented these taxa with UCEs extracted from the chicken (Gallus gallus), zebra finch (Taeniopygia guttata), Carolina anole lizard (Anolis carolinensis) and human (Homo sapiens) genome sequences. We combined the in silico and in vitro data and generated alignments across all taxa and excluded all loci having missing data from any taxon. This
resulted in 1145 individual alignments with a mean length of 406 bp (+100 bp s.d.) per alignment, totaling 465 Kbp of sequence. Tracer showed that both Bayesian analyses converged quickly, having effective sample size (ESS) scores for log likelihood of 170 and 220. Because posterior probabilities for all nodes were 1.0, AWTY (http://ceb.csit.fsu.edu/awty) showed zero variance in the tree topology throughout either run. Bayesian analysis of concatenated alignments and species-tree analysis of 1145 independent gene histories showed turtles to be the sister lineage of extant archosaurs with complete support (figure 2). Removing the snake, which had a very long branch, and re-running all analyses did not change the results.

4. DISCUSSION
Genomic-scale phylogenetic analysis of 1145 nuclear UCE loci agreed with most other molecular studies [4–14], supporting a sister relationship between turtles and archosaurs. We found no support for the turtles–lepidosaur relationship predicted by the Ankylopoda hypothesis [15] (figure 2). The combination of taxonomic sampling, the genome-wide scale of the sampling and the robust results obtained, regardless of analytical method, indicates that the turtle–archosaur relationship is unlikely to be caused by long-branch attraction or other analytical artefacts.

Although our results corroborate earlier studies, many of these studies did not include tuatara. Because tuatara is an early-diverging lepidosaur, it is important to include this taxon in studies of turtle evolution as it breaks up the long-branch leading to squamates (figure 2b). Of the studies including tuatara, two [6,11] found results similar to this study, but both were based on a single locus. The third study [5] was unable to produce a well-resolved tree from four nuclear genes when the authors included tuatara in the dataset. Our study is the first to produce a well-resolved reptile tree that includes the tuatara and multiple loci.

The discrepancy between our results showing a strong turtle–archosaur relationship and microRNA (miRNA) results, which showed a strong turtle–lepidosaur relationship, may be due to several factors. Lyson et al. [15] used the presence of four miRNA gene families, detected among turtles and lepidosaurs and undetected in the other taxa analysed, to support the turtle–lepidosaur relationship. Because complete genomes are unavailable for turtles, tuatara and crocodilians, and because expressed miRNA data are lacking for most reptiles, the authors collected miRNA sequences from small RNA expression libraries. miRNAs have

<table>
<thead>
<tr>
<th>common name</th>
<th>binomial</th>
<th>specimen ID/genome build</th>
<th>reads</th>
<th>assembled UCEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>African helmeted turtle</td>
<td>Pelomedusa subrufa</td>
<td>H20145&lt;sup&gt;a&lt;/sup&gt;</td>
<td>11 200 032</td>
<td>1972</td>
</tr>
<tr>
<td>American alligator</td>
<td>Alligator mississippiensis</td>
<td>HCD-2620&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3 528 983</td>
<td>2320</td>
</tr>
<tr>
<td>Carolina anole</td>
<td>Anolis carolinensis</td>
<td>H16061&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3 100 147</td>
<td>2111&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>corn snake</td>
<td>Pantherophis guttata</td>
<td>H15909&lt;sup&gt;e&lt;/sup&gt;</td>
<td>3 362 738</td>
<td>2168</td>
</tr>
<tr>
<td>human</td>
<td>Homo sapiens</td>
<td>UCSC hg19</td>
<td>NA</td>
<td>1748</td>
</tr>
<tr>
<td>painted turtle</td>
<td>Chrysemys picta</td>
<td>H2662&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4 467 644</td>
<td>2261</td>
</tr>
<tr>
<td>red junglefowl</td>
<td>Gallus gallus</td>
<td>UCSC galGal3</td>
<td>NA</td>
<td>2360&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>saltwater crocodile</td>
<td>Cocodylus porosus</td>
<td>LM-67&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3 261 088</td>
<td>2218</td>
</tr>
<tr>
<td>tuatara</td>
<td>Sphenodon tuatara</td>
<td>UMFS-10956&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5 651 932</td>
<td>2199</td>
</tr>
<tr>
<td>zebra finch</td>
<td>Taeniopygia guttata</td>
<td>UCSC taeGut1</td>
<td>NA</td>
<td>2345&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>From the LSU Museum of Natural Science.
<sup>b</sup>From the Darwin Crocodile Farm courtesy of L. Miles, S. Isberg and C. Moran.
<sup>c</sup>From the University of Michigan Museum of Zoology courtesy of R. Nussbaum and G. Schneider.
<sup>d</sup>Although we identified 2386 UCEs in these organisms, from which we designed capture probes, owing to slight adjustments to matching and filtering algorithms, we only recover ca 98% of these UCEs when re-screening these genomic sequences.
UCEs place turtles sister to archosaurs. N. G. Crawford et al. 785

Figure 2. (a) Reptilian phylogeny estimated from 1145 ultra-conserved loci using Bayesian analysis of concatenated data and species-tree methods, yielding identical topologies. Node labels indicate posterior probability/bootstrap support. (b) Phylogram of the UCE phylogeny generated with STEAC.

tissue and developmental-stage-specific expression profiles [25,26], which could make the detection of certain miRNAs challenging. Because preparing and sequencing libraries is a biased sampling process, the detection probability for specific targets is variable, and some miRNAs are likely to be more easily detected than others. Thus, failures to detect miRNA families are not equivalent to the absence of miRNA families [27]. We suggest that at least some of the four miRNA families currently thought to be unique to lizards and turtles may be present but as yet undiscovered in other reptiles.

This work is the first to investigate the placement of turtles within reptiles using a genomic-scale analysis of single-copy DNA sequences and a complete sampling of the major relevant evolutionary lineages. Because UCEs are conserved across most vertebrate groups [20] and found in groups including yeast and insects [19], our framework is generalizable beyond this study and relevant to resolving ancient phylogenetic enigmas throughout the tree of life [28]. This approach to high-throughput phylogenomics—based on thousands of loci—is likely to fundamentally change the way that systematists gather and analyse data.

(a) Additional information

We provide all data and links to software via Dryad repository (doi:10.5061/dryad.75nv22qj) and GenBank (JQ868813–JQ885411).

We thank R. Nilsen, K. Jones, M. Harvey, R. Nussbaum, G. Schneider, D. Ray, D. Peterson, C. Moran, L. Miles, S. Isberg, C. Mancuso, S. Herke, two anonymous reviewers and the LSU Genomic Facility. National Science Foundation grants DEB-1119734, DEB-0841729 and DEB-0956069, and an Amazon Web Services Education Grant supported this study. N.G.C., B.C.F., J.E.M. and T.C.G. designed the study; N.G.C. and B.C.F. performed phylogenetic analysis; B.C.F. created datasets; J.E.M. performed laboratory work; all authors helped write the manuscript.


