1,145 ultraconserved elements provide evidence that turtles are the sister group of archosaurs

Nicholas G. Crawford*[1], Brant C. Faircloth[2], John E. McCormack,[3] Robb T. Brumfield[3,4] Kevin Winker[5] & Travis C. Glenn[6]

[1]*Department of Biology, Boston University, Boston, MA 02215*, USA.

[2]*Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095*, USA.

[3]*Museum of Natural Science and* [4] *Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803*, USA.

[5]*University of Alaska Museum, 907 Yukon Drive, Fairbanks, AK 99775, USA.*

[6]*Department of Environmental Health Science and Georgia Genomics Facility, University of Georgia, Athens, GA 30602*, USA.

*Corresponding author: ngcrawford@gmail.com

Summary

We present the first genomic-scale analysis addressing the phylogenetic position of turtles, using over 1,000 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 1,145 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] (see Fig 1). Historically, turtles were thought to be early-diverging reptiles, called anapsids, based on their skull morphology and traits such as dermal armor [2]. Recent morphological studies that included soft tissue and developmental characters [3] allied turtles with lepidosaurs, a group including squamates (lizards and snakes) and tuataras. 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 [4,6-9] and nuclear DNA [5,10-16] typically place turtles sister to archosaurs (crocodilians and birds) (Fig 1). This molecular hypothesis was recently contradicted by a phylogeny reconstructed from microRNAs [17] that allied turtles with lepidosaurs. Lyson et al. [17] suggested that prior molecular evidence for a turtle-archosaur relationship may be the result of analytical artifacts. If true, the hypothetical relationship between turtles and lepidosaurs (Ankylpoda) 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 [18] and their flanking sequence (hereafter UCEs) 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., two of the most divergent crocodilians, lepidosaurs including tuatara, and turtles).

2. Material and methods

We enriched DNA libraries prepared with Nextera kits (Epicentre, Inc.) using a synthesis (Mycroarray, Inc. or Agilent, Inc.) of RNA probes [20] targeting 2,386 ultraconserved elements 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 (4 chains per run; burn-in: 50%; thinning: 100). We also used each alignment to estimate gene trees incorporating 1,000 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. [20,25].

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 2,648 (± 314 SD) 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 1,145 individual alignments with a mean length of 406 bp (± 100 bp SD) per alignment, totaling 465 Kbp of sequence. Tracer showed that both Bayesian analyses converged quickly, having 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 1,145 independent gene histories showed turtles to be the sister lineage of extant archosaurs with complete support (Fig. 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 1,145 nuclear UCE loci agreed with most other molecular studies [4,5,7-9,11-16], supporting a sister relationship between turtles and

archosaurs. We found no support for the turtles/lizard relationship predicted by the Ankylopoda hypothesis [17] (Fig 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 artifacts.

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 (Fig. 2b). Of the studies including tuatara, two [7,13] 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. [17] used the presence of four miRNA gene families, detected among turtles and lepidosaurs and undetected in the other taxa analyzed, 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 tissue and developmental-stage specific expression profiles [26,27], 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 [28]. 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 [29]. This approach to high-throughput phylogenomics – based on thousands of loci – is likely to fundamentally change the way that systematists gather and analyze data.

Additional Information

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

Acknowledgments

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, 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.

Author Contributions

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 data sets; J.E.M. performed laboratory work; all authors helped write the manuscript.

References

1 Lee, M. S. Y., Reeder, T. W., Slowinski, J. B. & Lawson, R. 2004 Resolving Reptile Relationships. In *Assembling the Tree of Life* (eds J. Cracraft & M. J. Donoghue), pp. 451–467. Oxford: Oxford University Press.

2 Lee, M. 1997 Reptile relationships turn turtle ... *Nature* **389**, 245–246.

3 Rieppel, O. 1999 Turtle Origins. *Science* **283**, 945–946. (doi:10.1126/science.283.5404.945)

4 A Janke, D. E. M. N. U. A. 2001 The mitochondrial genomes of the iguana (Iguana iguana) and the caiman (Caiman crocodylus): implications for amniote phylogeny. *Proc Biol Sci* **268**, 623. (doi:10.1098/rspb.2000.1402)

5 Hedges, S. & Poling, L. 1999 A molecular phylogeny of reptiles. *Science* **283**, 998–1001.

6 Hedges, S. B., Moberg, K. D. & Maxson, L. R. 1990 Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. *Mol Biol Evol* **7**, 607–633.

7 Rest, J. S., Ast, J. C., Austin, C. C., Waddell, P. J., Tibbetts, E. A., Hay, J. M. & Mindell, D. P. 2003 Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol Phylogenet Evol* **29**, 289–297.

8 Kumazawa, Y. & Nishida, M. 1999 Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. *Mol Biol Evol* **16**, 784–792.

9 Zardoya, R. & Meyer, A. 1998 Complete mitochondrial genome suggests diapsid affinities of turtles. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14226–14231.

10 Cao, Y., Sorenson, M. D., Kumazawa, Y., Mindell, D. P. & Hasegawa, M. 2000 Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes. *Gene* **259**, 139–148.

11 Katsu, Y., Braun, E. L., Guillette, L. J., Jr & Iguchi, T. 2009 From Reptilian

Phylogenomics to Reptilian Genomes: Analyses of c-*Jun* and *DJ-1* Proto-Oncogenes. *Cytogenet Genome Res* **127**, 79–93. (doi:10.1159/000297715)

12 Shedlock, A. M., Botka, C. W., Zhao, S., Shetty, J., Zhang, T., Liu, J. S., Deschavanne, P. J. & Edwards, S. V. 2007 Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2767–2772. (doi:10.1073/pnas.0606204104)

13 Hugall, A. F., Foster, R. & Lee, M. S. Y. 2007 Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Syst Biol* **56**, 543–563. (doi:10.1080/10635150701477825)

14 Shen, X.-X., Liang, D., Wen, J.-Z. & Zhang, P. 2011 Multiple Genome Alignments Facilitate Development of NPCL Markers: A Case Study of Tetrapod Phylogeny Focusing on the Position of Turtles. *Mol Biol Evol* (doi:10.1093/molbev/msr148)

15 Tzika, A. C., Helaers, R., Schramm, G. & Milinkovitch, M. C. 2011 Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent Sauropsida lineages and the phylogenetic position of turtles. *EvoDevo* **2**, 19. (doi:10.1186/2041-9139-2-19)

16 Iwabe, N., Hara, Y., Kumazawa, Y., Shibamoto, K., Saito, Y. & Miyata, T. 2005 Sister Group Relationship of Turtles to the Bird-Crocodilian Clade Revealed by Nuclear DNA-Coded Proteins. *Mol Biol Evol* **22**, 810–813. (doi:10.1093/molbev/msi075)

17 Lyson, T. R., Sperling, E. A., Heimberg, A. M., Gauthier, J. A., King, B. L. & Peterson, K. J. 2011 MicroRNAs support a turtle + lizard clade. *Biol Lett* (doi:10.1098/rsbl.2011.0477)

18 Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S. & Haussler, D. 2004 Ultraconserved elements in the human genome. *Science* **304**, 1321–1325. (doi:10.1126/science.1098119)

19 Siepel, A. et al. 2005 Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034–1050. (doi:10.1101/gr.3715005)

20 Faircloth, B. C. et al. In press. Ultraconserved Elements Anchor Thousands of Genetic Markers for Target Enrichment Spanning. **In press**. (doi:10.1093/sysbio/SYS004)

21 Zerbino, D. R. & Birney, E. 2008 Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829. (doi:10.1101/gr.074492.107)

22 Edgar, R. C. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797. (doi:10.1093/nar/gkh340)

23 Ronquist, F. et al. 2012 MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space. *Syst Biol* (doi:10.1093/sysbio/sys029)

24 Liu, L., Yu, L., Kubatko, L., Pearl, D. K. & Edwards, S. V. 2009 Coalescent methods for estimating phylogenetic trees. *Mol Phylogenet Evol* **53**, 320–328. (doi:10.1016/j.ympev.2009.05.033)

25 Lee, M. S. Y., Reeder, T. W., Slowinski, J. B. & Lawson, R. 2004 Resolving Reptile Relationships. In *Assembling the Tree of Life* (eds J. Cracraft & M. J. Donoghue), pp. 451–467. Oxford: Oxford University Press.

26 Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. & Tuschl, T. 2002 Identification of Tissue-Specific MicroRNAs from Mouse. *Current Biology* **12**, 735–739. (doi:10.1016/S0960-9822(02)00809-6)

27 Landgraf, P. et al. 2007 A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414. (doi:10.1016/j.cell.2007.04.040)

28 MacKenzie, D., Nichols, J., Lachman, G., Droege, S., Royle, J. & Langtimm, C. 2002 Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**, 2248–2255.

29 Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Robb T Brumfield & Glenn, T. C. In press. Ultraconserved Elements Anchor Thousands of Genetic Markers for Target Enrichment Spanning Multiple Evolutionary Timescales. *Syst Biol*

Short Title: "UCEs place turtles sister to archosaurs"

Legends

Figure 1. a) Depicts the primary morphological hypotheses: turtles most ancestral reptile [2][2] or turtles related to lepidosaurs [3][1]. b) Depicts the primary molecular hypothesis of a turtle-archosaur alliance [4,5,7-9,11-16]. c) Depicts the tree derived from miRNA loci [17].

Figure 2. a) Reptilian phylogeny estimated from 1,145 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.

Table 1. UCSC genome build or specimen ID for each sample, the number of ~100 bp sequence reads, and the total number of UCEs assembled.

Common Name	Binomial	Specimen ID /Genome Build	Reads	Assembled UCEs
African helmeted turtle	*Pelomedusa subrufa*	H20145[a]	11,200,032	1972
American alligator	*Alligator mississippiensis*	HCD-2620[a]	3,528,983	2320
Carolina anole	*Anolis carolinensis*	H16061[a]	3,100,147	2111[d]
Corn snake	*Pantherophis guttata*	H15909[a]	3,362,738	2168
Human	*Homo sapiens*	UCSC hg19	NA	1748
Painted turtle	*Chrysemys picta*	H2662[a]	4,467,644	2261
Red junglefowl	*Gallas gallus*	UCSC galGal3	NA	2360[d]
Saltwater crocodile	*Crocodylus porosus*	LM-67[b]	3,261,088	2218
Tuatara	*Sphenodon tuatara*	UMFS-10956[c]	5,651,932	2199
Zebra finch	*Taeniopygia guttata*	UCSC taeGut1	NA	2345[d]

[a]from the LSU museum of Natural Science; [b]from the Darwin Crocodile Farm courtesy of L. Miles, S. Isberg, and C. Moran; [c]from the University of Michigan Museum of Zoology courtesy of R. Nussbaum and G. Schneider.

[d]Although we identified 2,386 UCEs in these organisms, from which we designed capture probes, due to slight adjustments to matching and filtering algorithms we only recover ca. 98% of these UCEs when re-screening these genomic sequences.

Figure 1.

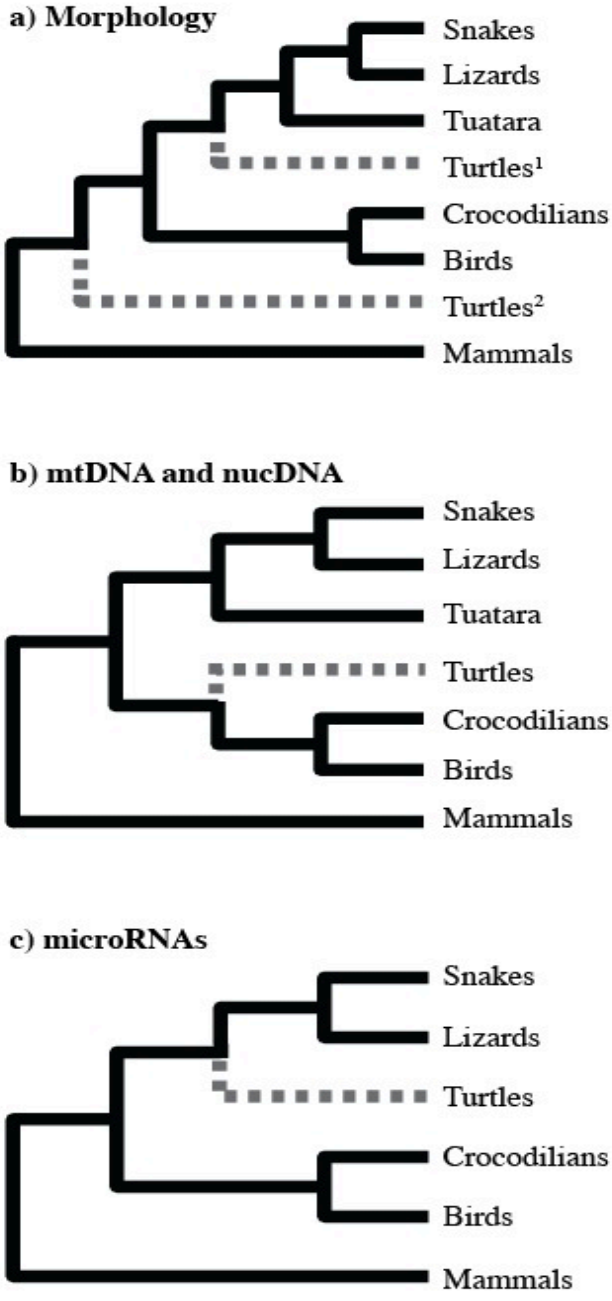
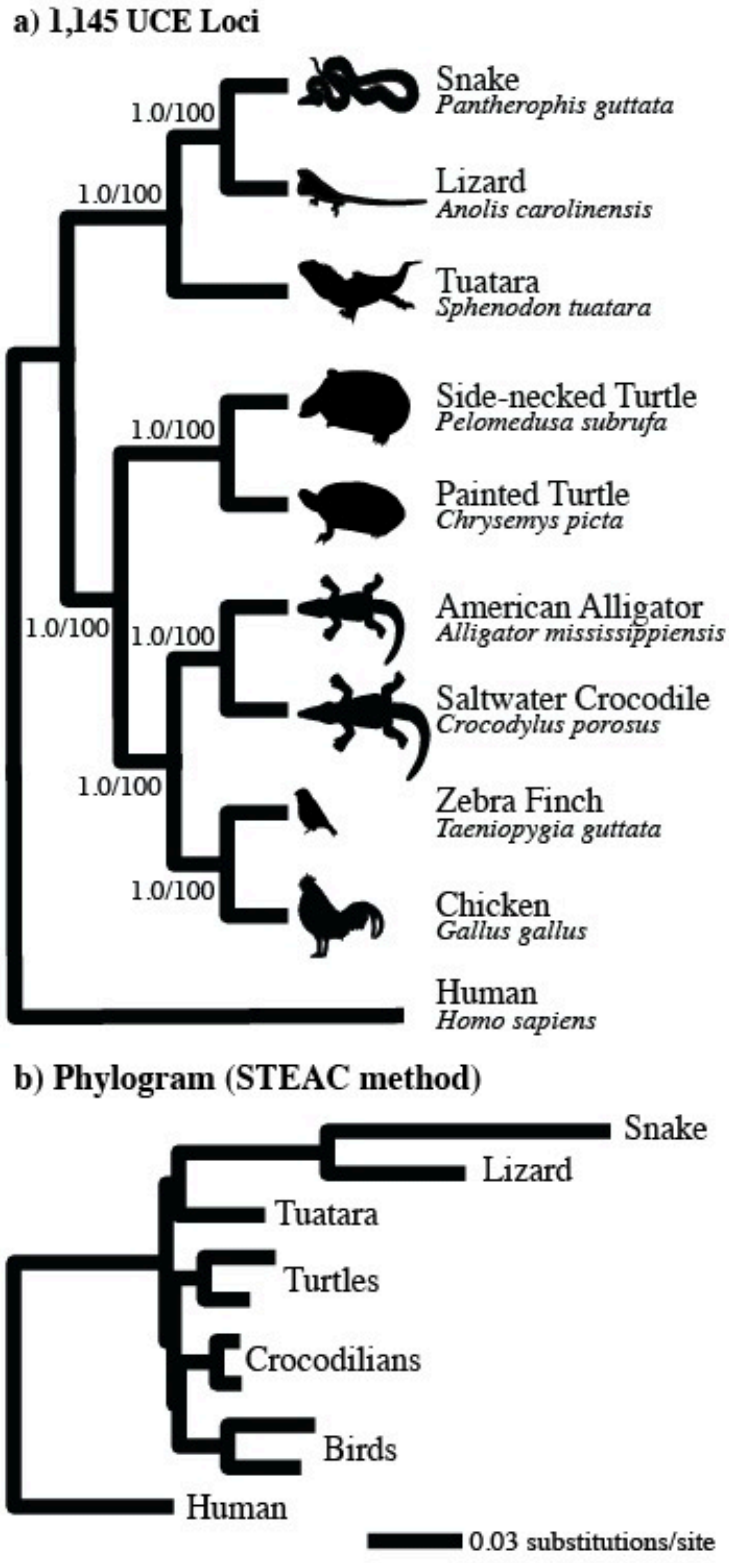

a) Morphology
Snakes
Lizards
Tuatara
Turtles[1]
Crocodilians
Birds
Turtles[2]
Mammals
b) mtDNA and nucDNA
Snakes
Lizards
Tuatara
Turtles
Crocodilians
Birds
Mammals
c) microRNAs
Snakes
Lizards
Turtles
Crocodilians
Birds
Mammals

Figure 2.


a) 1,145 UCE Loci
Snake
Pantherophis guttata
Lizard
Anolis carolinensis
Tuatara
Sphenodon tuatara
Side-necked Turtle
Pelomedusa subrufa
Painted Turtle
Chrysemys picta
American Alligator
Alligator mississippiensis
Saltwater Crocodile
Crocodylus porosus
Zebra Finch
Taeniopygia guttata
Chicken
Gallus gallus
Human
Homo sapiens
1.0/100
1.0/100
1.0/100
1.0/100
1.0/100
1.0/100
b) Phylogram (STEAC method)
Snake
Lizard
Tuatara
Turtles
Crocodilians
Birds
Human
0.03 substitutions/site