



Published in final edited form as:

Science. 2014 December 12; 346(6215): 1254449. doi:10.1126/science.1254449.

Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs

A full list of authors and affiliations appears at the end of the article.

Abstract

To provide context for the diversifications of archosaurs, the group that includes crocodylians, dinosaurs and birds, we generated draft genomes of three crocodylians, *Alligator mississippiensis* (the American alligator), *Crocodylus porosus* (the saltwater crocodile), and *Gavialis gangeticus* (the Indian gharial). We observed an exceptionally slow rate of genome evolution within crocodylians at all levels, including nucleotide substitutions, indels, transposable element content and movement, gene family evolution, and chromosomal synteny. When placed within the context of related taxa including birds and turtles, this suggests that the common ancestor of all of these taxa also exhibited slow genome evolution and that the relatively rapid evolution of bird genomes represents an autapomorphy within that clade. The data also provided the opportunity to analyze heterozygosity in crocodylians, which indicates a likely reduction in population size for all three taxa through the Pleistocene. Finally, these new data combined with newly published bird genomes allowed us to reconstruct the partial genome of the common ancestor of archosaurs providing a tool to investigate the genetic starting material of crocodylians, birds, and dinosaurs.

Introduction

Crocodylians, birds, dinosaurs, and pterosaurs are a monophyletic group known as the archosaurs. Crocodylians and birds are the only extant members and thus crocodylians (alligators, caimans, crocodiles, and gharials) are the closest living relatives of all birds (1, 2). While crocodylians diverged from birds more than 240 million years ago (MYA), animals with morphology unambiguously similar to the extant crocodylian families (Alligatoridae, Crocodylidae, and Gavialidae) first appear in the fossil record between 80 and 90 MYA (3). Unlike other vertebrates such as mammals, squamates and birds, which underwent substantial diversification, extant crocodylian species have maintained morphological and ecological similarities (4). Slow divergence among living crocodylians is also observed at the level of karyotype evolution (5).

Crocodylians are important model organisms in fields as diverse as developmental biology, osmoregulation, cardiophysiology, paleoclimatology, sex determination, population genetics, paleobiogeography, and functional morphology (4). For example, the males and

*Correspondence to: Richard E Green - ed@soe.ucsc.edu or David A Ray – david.a.ray@ttu.edu.

¹⁴Current Address: Cancer Prevention and Control, Fox Chase Cancer Center, Philadelphia, PA 19111, USA.

¹⁵Current Address: Department of Biology, Temple University, Philadelphia, PA 19122, USA.

³⁹Current Address: Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, 70803, USA.

⁴¹Current address: Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA.

females of all crocodylians (like some, but not all, reptiles) are genetically identical. Sexual fate is determined during development by a temperature sensing mechanism whose molecular basis remains poorly understood (6). More broadly, reptilian genomes exhibit substantial variation in isochore content, chromosome sizes and compositions (e.g. some but not all species have GC-rich and gene-rich micro-chromosomes), and sex-determination mechanisms. Remarkably, this plasticity in large-scale genome features is often coincident with a slower rate of karyotype and sequence evolution (7).

We sequenced the genomes of the American alligator, the saltwater crocodile, and the Indian gharial, spanning the three major extant crocodylian lineages (3, 8–10). These crocodylian genomes augment the list of assembled genomes from avian and non-avian reptiles (11–16) allowing us to probe the lineage-specific novelties in avian and crocodylian evolution. They also provide the substrate for computational inference of the common ancestor archosaur genome.

Genome assembly and annotation

We generated high-coverage Illumina sequence data (Tables S1–S3) from paired-end and mate-pair libraries from each species: alligator, crocodile, and gharial. The assembly strategy for each taxon differed due to varying legacy data and developments in library preparation methods during the course of the project (17). Importantly, genome scaffolding of alligator and to a lesser extent, saltwater crocodile, was aided by the availability of bacterial artificial chromosome (BAC) sequences and BAC end-sequence data. RNASeq data were collected from the alligator and, to a lesser extent, the crocodile and gharial (17). Stringently filtered consensus gene sequences were used for quality assessment of drafts of the genome assemblies and finally to aid in scaffolding the assemblies. Details of the libraries and assembly statistics for each genome are summarized in Tables S1–S4.

Gene annotation was accomplished using a combination of RNASeq data and homology-based analyses (17). We identified 23,323 protein-coding genes in the alligator compared to 13,321 and 14,043 in crocodile and gharial, respectively (Table S5). The unevenness likely reflects the larger N50 of the alligator genome assembly (Table S4) and importantly that the bulk of the transcriptome data used to guide gene identification derives from alligator (Table S6). This unevenness of annotation complicates direct comparisons of gene content. Therefore, for protein-coding sequence analyses, we compared orthologous sequence of the crocodile and gharial to the more thoroughly annotated alligator genome. We assigned names to 55% of crocodylian genes, on the basis of orthology to vertebrates with existing standardized nomenclature (human, mouse, anole, chicken and zebrafish). Between 60 and 70% of crocodylian proteins had conserved functional motifs on the basis of comparison to other vertebrates and we provided 377,441 GO annotations for 43,436 crocodylian proteins.

Transposable elements (TEs) were identified *de novo* in all three crocodylians and analyses resulted in a library of 1269 different TEs (Table S7); a large number for a vertebrate. This high TE count in crocodylian genomes is due, at least in part, to the apparently low rate of base substitution in crocodylians as discussed below. We find that ~37.5% of each

crocodilian genome can be annotated as TE (Table S7), a value intermediate between mammals (40–60%) and birds (12–15%) (18–23).

Ultraconserved element (UCE) phylogeny and molecular evolution

UCEs were originally defined as orthologous segments that exhibit very high levels of sequence conservation (24). Subsequent work established that UCEs often occur in single copy regions of the genome. Regions immediately flanking the core of a UCE typically exhibit progressively greater evolutionary rates (25–27). The relative ease of assessing orthology for UCEs and their flanking regions (hereafter called UCE-anchored loci), combined with their ease of alignment and the fact that they exhibit little or no substitution saturation, makes them useful for estimating *relative* evolutionary rates across all tetrapods. We identified and extracted 965 UCE-anchored sequences from the three crocodilian genomes and compared them to their orthologs from representatives of all major tetrapod lineages [in addition to the archosaurs, we included mammals, lepidosaurs (lizards and snakes), turtles, and an amphibian along with the coelacanth outgroup; (17) (Table S8)]. Using these data, we inferred tetrapod phylogeny and examined rates of evolution along the branches (Figure 1a and Figures S1–S7).

The phylogeny estimated using UCE-anchored data largely agrees with other studies (8, 10, 28, 29). For example, we recovered Longirostres (crocodile + gharials) within Crocodylia, found crocodilians to be the sister-group of birds (supporting the clade Archosauria), and confirm turtles as the sister group to living archosaurs. Branch lengths across this phylogeny suggest that crocodilians exhibit a low rate of molecular evolution for UCE-anchored loci relative to all tetrapod groups (Figure 1a), including the slowly evolving turtles. To explore the evolutionary tempo of crocodilians, we estimated absolute substitution rates across the tree using divergence time estimates for critical nodes (17). These estimates suggest that the molecular evolution of crocodilians is slower than all other lineages (Figures S2, S4, S7, S15, and S16). Indeed, the crocodilian rate is approximately an order of magnitude slower than that of lepidosaurs and mammals. Perhaps more importantly, the availability of multiple bird, crocodilian, and turtle genomes allow us to estimate the ancestral rates for these groups (Figure 1b). Using a variety of calibration times for the TMRCA of birds, crocodilians, and archosaurs (Figure S8 and Table S9), we find that the rate of UCE evolution for the avian stem lineage was similar to extant avian lineages (Figure 1b and Figure S7). In contrast, the crocodilians stem lineage evolved more rapidly than its extant lineages. Given the low rates observed in both turtles and crocodilians and the reduced rate in the avian stem lineage, we propose that the ancestor of all archosaurs was likely characterized by an extremely slow rate of molecular evolution that subsequently increased on the avian stem lineage.

Gene-based phylogeny and molecular evolution

We used the PhylomeDB pipeline (19) to identify 337 single copy orthologous gene sequences (17) from 22 tetrapod genomes (Table S10). Phylogenetic analysis of a concatenated alignment of these genes (Figure S10) produced a tree congruent with the UCE-based phylogeny shown in Figure 1a and other amniote phylogenies (28, 30). The

concatenated alignment of orthologs was then further filtered to extract four-fold degenerate (4D) sites (17), which evolve at a rate similar to the neutral rate. Although some 4D sites may be subject to purifying selection (31), studies in birds suggest that substitutions at 4D sites accumulate ~75% as rapidly as those at other sites thought to be neutral (32). Thus, their rate is expected to be much closer to the neutral rate than the rate estimated using UCE-anchored data. As expected, substitution rates at 4D sites were higher than the rate estimated using UCE-anchored regions (Fig. 1b). However, the pattern of relative rates for different taxa was qualitatively similar to that reconstructed using the UCE-anchored regions (Figure 1b and Figures S13–S16).

A larger survey of aligned genes (without the single copy orthology filters) found 9,574 trees that suggested monophyly of birds, turtles and crocodylians relative to squamates and the vast majority of those (6,880; 72%) placed crocodylians and birds together in a clade. Only 28% of trees supported alternate topologies (birds + turtles or crocodylians + turtles (17)). Although the placement of gharial within the crocodylian phylogeny has been contentious over the last several decades (33) a clear majority (78.4%) of protein coding gene trees supported Longirostres (8).

Rates of genome evolution in crocodylians, birds, and other reptiles

To explore the patterns of molecular evolution across the genomes of crocodylians, we created a whole genome alignment (WGA) (17) that included 23 reptile genomes, including the three crocodylians, 15 birds, four turtles, and the Carolina anole lizard as the outgroup (Table S12). Consistent with our other results, the WGA analysis revealed low genome-wide pairwise divergences among crocodylians (Table S13); for example, the alligator and crocodile (which shared a common ancestor ~80–100 MYA; node O in Table S9) have ~93% genome-wide identity. This is similar to the level of identity between human and rhesus macaque, whose common ancestor lived only ~23 MYA (34), indicating exceptionally low rates of evolution compared to mammals.

This WGA for birds and reptiles also provides an opportunity to assess the relative rates of different substitution types using a single alignment framework. We compared rates at 4D sites (Figure 2 and Table S14) with those occurring within orthologous TE insertions that are shared among the three crocodylians (Table S15). Substitutions in TEs, which presumably accumulate at close to the neutral rate (35), accumulated slightly more rapidly than those at 4D sites extracted from the WGA (Fig. 2A). The WGA also allowed us to estimate the rate of micro-indels (10 bp per event, filtered to avoid alignment errors) relative to substitutions. This ratio for crocodylians (0.064 microindels per substitution) is similar to that in birds and turtles (Fig. 2b), and is within the range of previous estimates for mammals (36, 37). The ratio of micro-deletions to micro-insertions was similar across the tree (avg. ~1.94; Table S16) and concordant with previous estimates from other taxa (36, 37), with no apparent bias towards either category in crocodylians, birds or turtles (Table S20).

Finally, we also used the multiple species WGA to examine the conservation of synteny between adjacent gene pairs in chicken and alligator, examining only those pairs where both

genes were unambiguously located (17). We find similarly high levels of gene order conservation between crocodylian genomes as between comparably separated bird genomes – a group marked by its extreme syntenic conservation relative to mammals (Figure 2C). Thus, the low evolutionary rates observed in crocodylians are not specific to substitutions, but also include micro-indels and gene-level rearrangements.

Transposable elements evolve slowly in crocodylians

Of the annotated TEs, 95% belong to families that appear in all three genomes with near-equal frequency. Thus, only ~5% of TE copies (representing <2% of the genome) arose after the split of Longirostres (crocodyle + gharial) from alligators, approximately 80–100 MYA (Table S9). Considering that there is an ascertainment bias against older repeats, these data suggest that the rate of new TE family invasion/evolution has generally been decreasing in crocodylians, with an exception of a minor burst of novel activity in the common ancestor of Longirostres (Figure S20). Indeed, in the ~235 MY between the mammal-crocodylian divergence and the origin of crown crocodylians, at least 823 TE families were active, a rate of around 3.5 TE families/MY. The rate has fallen below 1.0 in both crocodyle and gharial since their divergence.

The ‘visibility’ of TE copies introduced before the divergence of mammals and reptiles, ~310–330 MYA (Table S9) (17), provides another line of evidence for the extraordinarily low rate of crocodylian genome evolution. Averaged over 74 unrelated families of such elements (17), crocodylian genomes contain five times more DNA that is recognizably derived from such elements as the typical mammalian genome, three times more than the reconstructed boreoeutherian (the mammalian clade comprising primates, rodents, carnivores, bats, and a number of additional orders, ‘boreo’ in Figure 3) genome (36), 3.8 times what is identifiable in the chicken genome and 15 times more than the anole genome. Surprisingly, the painted turtle genome contains on average 2.3 times more bases recognizably derived from each of these repeats compared to crocodylians (Figure 3 and Figures S21–S23), suggesting an even slower neutral decay rate. The consistency of the relative representation of these unrelated elements in each genome suggests that these ratios are not the result of differential lineage-specific accumulation but represent actual differences in mutation and deletion rates and that crocodylians exhibit a neutral mutation rate that is among the slowest found in vertebrates and may be the slowest within amniotes.

Gene family evolution suggests retention of ancestral orthologs in crocodylian lineages

We used gene trees from the phylome analysis to search for gene families that underwent duplications within the crocodylian lineage (17). One of several gene families on which we concentrated our efforts is olfactory receptors. These comprise one of largest vertebrate gene families and its members are small, single-exon genes, making them relatively easy to investigate. They have also played a central role in the development of our understanding of how gene families evolve (38). Similar to results found in other amniote lineages (39, 40), genes associated with olfactory perception were over-represented among duplicated genes in crocodylians. Crocodylians possess a diverse olfactory receptor (OR) repertoire, and each

species has approximately 1,000 ORs, half of which are likely functional (Table S21) (17) which is not unusual for a tetrapod genome. However, in other tetrapods the ORs derive from independent expansions of a small number of ancestral OR genes within those lineages (38), as we observed for the birds and turtles we examined (Figure 4 and Figure S24). In contrast, crocodylian OR repertoires almost exclusively reflect the retention of OR genes present in the common ancestor of crown crocodylian followed by few gains or losses (Figure 4 and Figure S24). This observation, many retained ancestral genes rather than novel expansion, suggests that crocodylians have achieved a diverse OR repertoire using a novel strategy: retention of ancient genes as opposed to the generation of novel variants.

Genetic diversity and natural history of Crocodylia

We used the genomic data generated here to investigate the population history of each crocodylian species. Mapping shotgun reads back to the assembly, we identified and quantified the rate of heterozygosity (17) within each species. All three genomes exhibited a low degree of heterozygosity compared to most mammalian and avian genomes (Figure 5). Among the three crocodylian taxa we examined, the crocodile had the highest observed genetic diversity with about three heterozygous sites per 10 kb. The lower heterozygosity of the other two crocodylians examined here is interesting considering their recent or current status as endangered species. The gharial is critically endangered due to habitat loss (41) and the American alligator recently survived an anthropogenic population bottleneck (42) and was removed from the endangered species list in 1987. We inferred the effective population sizes of alligator, crocodile, and gharial (Figure 5A) using the neutral mutation rate for crocodylians ($\mu = 7.9 \times 10^{-9}$ substitutions site⁻¹ generation⁻¹) calculated from the pairwise divergence (17) between the alligator and saltwater crocodile. We note that the alligator and crocodile were wild caught and thus likely to represent the genetic diversity of their species, whereas the gharial we sequenced was bred in captivity and of unknown recent ancestry.

The crocodylians comprise many of the largest extant ectothermic species. As such, their success through recent geologic time is of special interest. Given their long generation time and slow mutation rate the pairwise sequentially Markovian coalescent (PSMC) model (43) approach can probe population sizes further into the past than is possible for faster evolving lineages. All three lineages experienced distinct changes in their estimated N_e over the past seven million years (Figure 5B and Figure S26). We also included estimates of air temperature data (44) to identify any potential relationship of demographic histories to climate change. We identified that both crocodile and gharial maintained relatively stable population sizes through the Pleistocene and Pliocene but both experienced sharp declines during the last cooling cycle between ~100 and 10 thousand years before present (Figure 5B). In contrast, the population size of alligators declined continuously throughout the Pleistocene, perhaps because they inhabit more temperate latitudes and experienced greater effects from global cooling. A generally declining effective population size over the past million years was also shown for the Chinese alligator (15) using the PSMC approach.

A draft archosaur genome

One exciting use of genome sequence spanning archosaurs is the potential to infer the ancestral archosaur genome. As part of the WGA analysis, we computationally inferred the ancestral archosaur genome, along with ancestral genomes for all the internal nodes of the tree. Due to the constant turnover of sequence during the ~300 MY since the divergence of birds and crocodylians and the likelihood that some data are missing in the assemblies of extant taxa, the reconstructed genome assembly is limited to 584 Mb of sequence, less than the genome assemblies for extant taxa. Determining the nucleotide at each position in the ancestral archosaur genome with a standard continuous time substitution model, the average expected reconstruction accuracy of archosaur bases is 91% (Figure 6A and Figure S17).

The ancestral genome reconstruction exhibits a strong bias toward the recovery of functional elements. For example, we mapped alligator regions with various annotations, TEs, coding DNA sequences (CDS), 3' and 5' untranslated regions (UTR), exons, upstream sequences (defined by a 500bp window upstream of the putative transcription start site for each gene) and introns, to the archosaur genome using the WGA to map the annotations by projection through the alignment. Compared to putatively neutrally evolving elements like TEs, we found CDSs, 3' UTRs, and 5' UTRs (in decreasing order), to have substantially higher base level reconstruction accuracy (e.g., 97% of base calls mapped by CDS annotations are expected to be correct, Figure 6A). Concordantly, while on average only 26% of alligator bases have an aligned base in the archosaur reconstruction, the proportion of annotated bases mapping to archosaur is higher (Figure 6B) (17). The reconstruction bias toward functional elements is correlated with differences in purifying selection as measured with PhyloP on the WGA (17). Transcribed elements annotated in alligator or chicken are also more likely to have remained stably ordered and oriented mapping back to the archosaur, suggesting that intra-gene ordering constraints have helped preserve sequence structure (Figures S17–S19).

Discussion

The draft genome assemblies of these three crocodylian taxa add to the growing list of available reptilian genomes and allow a more comprehensive analysis of vertebrate genome evolution. Because crocodylians are the sister group of birds, these three genome also provide a critical resource for examining the ancestral state of various genomic features for birds, for which multiple genomes are now available (45). The most striking of our results is the remarkably low rate of genome-wide molecular evolution among all major crocodylian lineages. This low rate was observed for the accumulation of base substitutions at many different types of sites (those in UCE-anchored loci, 4D sites in protein coding regions, and the presumably neutral sites in TE insertions) and for other types of genomic changes, like micro-indels and TE movement. Recent genomic analyses of turtles suggest a low rate of evolution in that lineage as well (13), a finding we confirmed and extended. Taken as a whole, this provides strong evidence that that a slow rate of genomic change is the ancestral state for archosaurs.

Our evidence that the low rate of molecular evolution applies to multiple types of genomic changes makes it tempting to speculate that there is a single underlying cause. Within

mammals, the accelerated rate of molecular evolution for rodents relative to primates (also observed here; Figure 1) is often attributed to shorter generation times along the rodent lineage (46). However, there have also been suggestions that the high rate in rodents could reflect differences in DNA repair efficiency (47). More broadly, rates of molecular evolution may be correlated with a number of factors, including body size and metabolic rate (48, 49). However, these and other life history characters are themselves correlated (50, 51), making it very difficult to untangle the relevant causal factors.

Our analyses include all major amniote lineages and it is clear that crocodylians and turtles exhibit the lowest rates of molecular evolution; both of those clades are characterized by long generation times. Indeed, using a 20-year generation time along the crocodylian lineage (17), the inferred rate of molecular evolution per site per *generation* (7.9×10^{-9} substitutions per site per generation) is not substantially different from estimates in other lineages; it is the rate per *year* that is much lower for crocodylians. The higher rate for stem birds, which is actually similar to that observed for extant birds, could indicate that this species had already decreased its generation time. Indeed, recent analyses of paleontological data are highly consistent with decreased body size on the lineage ancestral to extant birds (52). Given the strong correlation between body size and generation time (51) this would be consistent with our observed changes in the average rate of molecular evolution. It will be of substantial interest to establish whether similar morphological correlates can be established for stem crocodylians and other lineages.

Materials and Methods

Sequencing and assembly

Genomic DNA was isolated using blood from four individuals including two *A. mississippiensis*, individuals of *C. porosus* and *G. gangeticus*. Sequencing depth and assembly strategies differed depending on legacy data available for each taxon (17). Briefly, alligator data consisted of Illumina sequences from five libraries ranging from 5.5 to 88.7x coverage. These reads were assembled using AllPaths-LG (53) with default parameters. Legacy data from 21 fully sequenced BACs, 1309 BAC-end read pairs (54), and RNASeq data, described below, were also used to aid the assembly. Crocodile data consisted of Illumina reads from three libraries ranging from 21.6 to 90.2x coverage. AllPaths-LG was used to assemble the raw data. As with the alligator genome draft, sequences from 360 MHC region BAC assemblies as well as RNASeq data were used to aid the assembly. The gharial genome was assembled using SOAPdenovo v2.04 (55) and data from four Illumina libraries ranging from 50 to 170x coverage. No legacy data was available to improve the gharial assembly.

Transcriptome sequencing and sequence annotation

Total RNA was extracted from multiple alligator and crocodile tissues as well as gharial whole blood (17). RNA was extracted and subjected to library preparation and Illumina RNASeq. While variable, most libraries had insert sizes between 300 and 350 bp and were sequenced both individually and as pools. In total, 11 Gb of high quality sequence data were generated.

Gene predictions were made using Augustus (version 2.5.5) (56). RNASeq data from *A. mississippiensis* was aligned to the genome draft of American alligator draft using Tophat version 2.0.6 (57) and Bowtie version 2.0.5 (58). Augustus used these alignments to improve its gene predictions. Protein coding genes predicted for alligator were then aligned to the other crocodylian assemblies using Genblastg version 1.38, and those alignments were used by Augustus to improve the gene predictions for those species. Functional annotation was accomplished by assigning gene nomenclature, Gene Ontology (GO) and pathway information. Gene names were assigned based upon orthology or homology to species with a gene nomenclature project by transferring names to the crocodylian genes. GO was assigned to predicted proteins based upon a combinatorial approach that is fully described in the Supplemental Materials (17). Pathway information was assigned based upon reciprocal BLAST. Annotated genes, gene products and genome assemblies are available at NCBI, CrocBase (<http://crocgenome.hpc.msstate.edu/crocbase/gene.php>) and via the Comparative Genomics (CoGe) browser (<http://genomevolution.org/CoGe/>).

Transposable elements (TEs) were identified and annotated semi-independently in the genomes by three laboratories semi-independently. Briefly, TEs were identified *de novo* in a given genome draft with either RepeatModeler (59) or a combination of PILER (60), RepeatScout (61), and LTRHarvest (62). Output from each method was curated using a combination of manual inspection and computational tools. Combining TE consensus sequences from all three crocodylians resulted in a library of 1269 different TEs. Full details of all sequence annotation protocols are available in the Supplemental Materials (17).

UCE identification and analysis

To create a large set of ultraconserved element loci (UCEs) (17), we combined two sets of ultraconserved elements (25, 63) and kept unique and non-duplicate loci in the set ($n=8,047$ UCE loci). Using the positions of these loci in the chicken genome (galGal3), we designed capture probes ($n=12,237$) for each locus to use for *in silico* identification of orthologous UCEs in other tetrapods (Table S6) and aligned each capture probe to those genomes. Following identification of putative UCE loci in each genome, we sliced the match location of all probes ± 2000 bp from each genome assembly and recovered slices derived from multiple probes targeting the same locus, we re-assembled sequences back into full UCE loci. We then trimmed all slices to approximately the length of the UCE locus ± 1000 bp and identified the set of all loci found in all taxa (a complete matrix) from two different taxon samples (Table S8). We named these taxon-set-1 and taxon-set-2. Taxon-set-1 includes the Western clawed frog (*Xenopus tropicalis*), and as a result, contains fewer orthologous loci in a complete matrix.

Using the complete data matrices, we aligned FASTA data corresponding to each re-assembled UCE locus for each taxon. Following alignment and trimming, we removed any loci containing ambiguous base calls. The remaining alignment data for taxon-set-1 contained 604 loci totaling 495,744 characters and 93,374 alignment patterns (mean locus length=820 bp; 95 CI = 47 bp). The remaining alignment data for taxon-set-2 contained 965 loci totaling 878,786 characters and 172,112 alignment patterns (mean locus length=911 bp; 95 CI = 40 bp). We concatenated all loci in each set, and we analyzed the resulting,

concatenated alignments using RAxML 7.3.4 (64), conducting 20 maximum likelihood (ML) tree searches and 500 bootstrap replicates for each data set. Using RAxML, we checked for bootstrap replicate convergence using the “autoMRE” function. Both data sets converged after 50 replicates, and we used RAxML to reconcile each best, ML tree with each set of 500 bootstrap replicates. We also conducted partitioned, concatenated analyses of the UCE data, but these results did not differ from the unpartitioned results (17).

Phylome analysis

Complete collections of Maximum Likelihood (ML) gene trees for every gene encoded in each of the three crocodylian genomes (phylomes) were reconstructed using the phylomeDB pipeline (17, 65). In brief sequence searches were used to retrieve homologs (e-value $1e-5$, 50% overlap) in a set of vertebrates (17). These were aligned using three different programs in forward and reverse orientation. Consensus alignments were built with T-coffee (66) and trimmed with trimAl (67). The evolutionary model best fitting the data was used to build an ML tree with PhyML (68) using four rate categories and a fraction of invariable sites, estimated from the data. Branch support was computed using an aLRT (approximate likelihood ratio test) parametric test. Orthology and paralogy relationships among crocodylian genes and those encoded by the other genomes were inferred from the phylomes, using a species-overlap algorithm (69), as implemented in ETE (70). The resulting trees and orthology and paralogy predictions can be accessed through phylomeDB.org (19). The crocodylian phylomes were scanned to detect and date duplication events using a previously described algorithm (71). For species tree reconstruction two complementary approaches were used. First, a super-tree was inferred from all trees in the three phylomes by using a Gene Tree Parsimony approach as implemented in the dup-tree algorithm (72). Second, the alignments of 337 gene families with one-to-one orthology in all considered species were concatenated and used to build a ML phylogeny as described above.

Gene family analysis

We conducted bioinformatic searches to characterize the repertoires of Olfactory Receptors (ORs), Vomeronasal Receptors types 1 and 2 (V1R and V2R), Taste Receptors type 1 and 2 (T1R and T2R) and trace amine-associated receptors (TAAR) of the three crocodylians in our study, and we compared the repertoires with representative vertebrates (Table S21) (17). We focused the majority of our analyses on the ORs. Briefly, we performed TblastN searches of the three crocodylian genomes using known vertebrate ORs as queries and the best non-overlapping BLAST hits were extracted. Putative complete OR genes were added to the amino acid query and a new TblastN search was conducted to annotate pseudogenes and truncated genes. Putative ORs were annotated to their subfamily by comparing amino acid sequences against a BLASTP database of known OR amino acid sequences. Phylogenetic analyses were conducted using MEGA v5 (73). We inferred neighbor-joining phylogenies to assess patterns of divergence and diversity of intact crocodylian ORs compared to other vertebrates using a Poisson model of substitution and evaluated support for the nodes with 1,000 pseudoreplicates. We compared the evolution of ORs for the three crocodylians, chicken and zebra finch (74), and green sea turtle and Asian softshell turtle (16).

Genome alignments and ancestral genome reconstruction

The whole genome alignment of 23 taxa (Table S12) (17) was computed using progressive-cactus (github.com/glennhickey/progressiveCactus) with default parameters and the phylogeny shown in Fig. 2A (75). The topology of the phylogeny was derived by manually merging a subtree of the UCE tree (17) with results from the avian phylogeny sister paper (76) along with published phylogenies for passerine birds (77), parrots, (78), and turtles (79). Nucleotide level ancestral reconstruction of all internal nodes was performed as part of the process using a phylogenetically weighted form of the algorithm described in Nguyen et al. (80) and appropriate for partial genome assemblies. To improve the ancestral base calls we used the ancestorsML tool in the HAL tools library (github.com/glennhickey/hal) (81) to call bases by maximum likelihood, using the general reversible continuous time nucleotide substitution model. To parameterise the model and estimate branch lengths we used phyloFit (82) on conserved fourfold degenerate sites in alligator genes, as described in the Supplemental Materials (17). A complete technical exposition of the alignment computation and statistics calculated is available in the Supplemental Materials.

Mutation rate estimation

We used a phylogenetic approach to estimate the overall mutation rate, μ , along the crocodylian lineage. From both the whole genome alignment between alligator and crocodile and the multiple sequence alignment that includes alligator and crocodile, we estimate the overall divergence between alligator and crocodile to be 7.1%. Because of the remarkably small divergence between these two, we assumed an infinite sites model of evolution and ignored back mutations. We calculate a per generation mutation rate using 90 MY as the time of the most recent common ancestor of alligator and crocodile and an average generation time of 20 years (Table S23) as shown in the Supplemental Materials (17).

Heterozygosity and population history estimation

For each genome, we mapped paired-end genome reads from a single individual back to the final genome assembly using BWA (83) as described in the Supplemental Materials (17). We used tools in the GATK package (<http://www.broadinstitute.org/gatk/>) to perform indel realignment of each read around possible insertion-deletion positions, then analyzed all genomic positions where the read depth was exactly equal to the genome-wide mean. We derived cutoffs to distinguish *bona fide* heterozygous positions from sequence error by analysis of mutation spectra at these sites (Table S24). From this analysis, we calculated the observed rate of heterozygosity, H , at intergenic sequence in each species: alligator $H = 0.000136$; gharial $H = 0.000217$; crocodile $H = 0.000360$. Using these values as an estimate for theta and the substitution rate, m , calculated above, we estimated the effective population size for each species as shown in Figure 6 of the main manuscript.

To estimate historical population sizes, we called SNPs with SAMtools using reads with a >30 map score and base calls with a >20 quality score. We applied the pairwise sequential Markovian coalescent (43) model using 20 years for the generation time (Table S23). We used 90 MYA as the TMRCA of *C. porosus* and *A. mississippiensis* and our analyses indicate 7.1% divergence. Therefore, given a 20 year generation time, we calculated a mutation rate of 7.89×10^{-9} /year/site. We conducted bootstrap tests for each of the three

taxa by splitting the scaffolds into smaller segments and randomly sampling the segments with replacement (Figure S26). We used 100 replicates to test the robustness of the returned population demographic history. We also gathered ancestral Northern Hemisphere air temperature data from (44) and took averages for 200k year bins. Climate oscillations over the last one million years were calculated in 20k year bins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Richard E Green^{1,*}, Edward L Braun², Joel Armstrong^{1,3}, Dent Earl^{1,3}, Ngan Nguyen^{1,3}, Glenn Hickey^{1,3}, Michael W Vandewege⁴, John A St John¹, Salvador Capella-Gutiérrez^{5,6}, Todd A Castoe^{7,8}, Colin Kern⁹, Matthew K Fujita⁸, Juan C Opazo¹⁰, Jerzy Jurka¹¹, Kenji K Kojima¹¹, Juan Caballero¹², Robert M Hubley¹², Arian F Smit¹², Roy N Platt^{4,13}, Christine A Lavoie⁴, Meganathan P Ramakodi^{4,13,14,15}, John W Finger Jr.¹⁶, Alexander Suh^{17,18}, Sally R Isberg^{19,20,21}, Lee Miles²⁰, Amanda Y Chong²⁰, Weerachai Jaratlerdsiri²⁰, Jaime Gongora²⁰, Christopher Moran²⁰, Andrés Iriarte²², John McCormack²³, Shane C Burgess²⁴, Scott V Edwards²⁵, Eric Lyons²⁶, Christina Williams²⁷, Matthew Breen²⁷, Jason T Howard²⁸, Cathy R Gresham¹³, Daniel G Peterson^{13,29}, Jürgen Schmitz¹⁷, David D Pollock⁸, David Haussler^{3,30}, Eric W Triplett³¹, Guojie Zhang^{32,33}, Naoki Irie³⁴, Erich D Jarvis²⁸, Christopher A Brochu³⁵, Carl J Schmidt³⁶, Fiona M McCarthy³⁷, Brant C Faircloth^{38,39}, Federico G Hoffmann^{4,13}, Travis C Glenn¹⁶, Toni Gabaldón^{5,6,40}, Benedict Paten³, and David A Ray^{4,13,41,*}

Affiliations

¹Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA

²Department of Biology and Genetics Institute, University of Florida, Gainesville, FL 32611, USA

³Center for Biomolecular Science and Engineering, University of California, Santa Cruz, California 95064, USA

⁴Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA

⁵Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG). Dr. Aiguader, 88. 08003 Barcelona, Spain

⁶Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain

⁷Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045, USA

⁸Department of Biology, University of Texas at Arlington, Arlington, TX 76019, USA

- ⁹Department of Computer and Information Sciences, University of Delaware, Newark, DE 19717, USA
- ¹⁰Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile
- ¹¹Genetic Information Research Institute, Mountain View, CA 94043, USA
- ¹²Institute for Systems Biology, Seattle, WA, 98109, USA
- ¹³Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA
- ¹⁶Department of Environmental Health Science, University of Georgia, Athens, GA 30602, USA
- ¹⁷Institute of Experimental Pathology (ZMBE), Von-Esmarch-Str. 56, D-48149 Münster, Germany
- ¹⁸Department of Evolutionary Biology (EBC), Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden
- ¹⁹Porosus Pty Ltd, Palmerston, NT 0831, Australia
- ²⁰Faculty of Veterinary Science, University of Sydney, Sydney, NSW 2006, Australia
- ²¹Centre for Crocodile Research, Noonamah, NT 0837, Australia
- ²²Departamento de Desarrollo Biotecnológico, Instituto de Higiene, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay
- ²³Moore Laboratory of Zoology, Occidental College, Los Angeles, CA 90041, USA
- ²⁴College of Agriculture and Life Sciences, University of Arizona, Tucson, AZ 85721, USA
- ²⁵Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
- ²⁶School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA
- ²⁷Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, NC 27607, USA
- ²⁸Howard Hughes Medical Institute, Duke University Medical Center, Department of Neurobiology, Durham, NC 27710, USA
- ²⁹Department of Plant and Soil Sciences, Mississippi State University, Mississippi State, MS 39762, USA
- ³⁰Howard Hughes Medical Institute, Bethesda, Maryland 20814, USA
- ³¹Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611 USA
- ³²China National GeneBank, BGI-Shenzhen, Shenzhen, China

³³Center for Social Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark

³⁴Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan

³⁵Department of Earth and Environmental Sciences, University of Iowa, Iowa City, IA 52242, USA

³⁶Department of Animal and Food Sciences, University of Delaware, Newark, DE 19717, USA

³⁷School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ 85721, USA

³⁸Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90019, USA

⁴⁰Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 arcelona, Spain

Acknowledgments

Genome drafts have been submitted to the National Center for Biotechnology Information repository under the following accession numbers: *Alligator mississippiensis*, AKHW00000000; *Crocodylus porosus*, JRXG00000000; *Gavialis gangeticus*, JRWT00000000. Supplemental files have been archived at GigaScience (<http://dx.doi.org/10.5524/100125>, <http://dx.doi.org/10.5524/100126>, <http://dx.doi.org/10.5524/100127>, <http://dx.doi.org/10.5524/100128>) and at crocgenomes.org. This project was conducted by the International Crocodylian Genomes Working Group (ICGWG; www.crocgenomes.org). Funding for this project was provided by the National Science Foundation (MCB-1052500 and DEB-1020865 to DAR; MCB-0841821 to DAR, DGP, FMM, and CJS; DUE-0920151 to ELB and EWT; DBI-0905714 to MKF; DEB-1242260 to BCF). DAR, FMM and DGP were also supported by the Institute for Genomics, Biocomputing and Biotechnology at Mississippi State University. SRI, LGM, JG, and CM were supported by Australian Rural Industries Research and Development Corporation grants (RIRDC PRJ-000549, RIRDC PRJ- 005355, RIRDC PRJ-002461). REG is a Searle Scholar, Sloan Fellow, and consultant for Dovetail Genomics.. EDJ was supported by the Howard Hughes Medical Institute and the National Institutes of Health. EL received support from the Gordon and Betty Moore Foundation (#3383). R. Elsey, S. Lance and T. Tuberville aided in collecting alligator samples. The following centers were vital in permitting the computational analyses required for this project: The High Performance Computing Collaborative (HPC2) at Mississippi State University, the High Performance Computing Center at Texas Tech University, the Georgia Advanced Computing Resource Center at the University of Georgia, the University of Florida High Performance Computing Center. The National Institutes of Health (NIH) provided funding for UCSC infrastructure used in computing whole genome alignments and ancestral genome reconstructions (1U41HG007234-01, 1U41HG006992-2, 5U01HG004695). Finally, we are grateful to K. Vliet and D. Barber for providing access to fresh gharial blood.

References

1. Janke A, Arnason U. The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent archosauria (birds and crocodiles). *Mol Biol Evol.* Dec.1997 14:1266. [PubMed: 9402737]
2. Sennikov AG. The first ctenosauriscid (Reptilia: Archosauromorpha) from the Lower Triassic Of Eastern Europe. *Paleontol J.* Sep 01.2012 46:499.
3. Brochu CA. Phylogenetic approaches toward crocodylian history. *Annual Review of Earth and Planetary Sciences.* 2003; 31:357.
4. Grigg, G.; Seebacher, F.; Franklin, CE., editors. *Crocodylian Biology and Evolution.* Surrey, Beatty and Sons, Chipping Norton; NSW: 2001.
5. Cohen MM, Gans C. The chromosomes of the order Crocodylia. *Cytogenetics.* 1970; 1970:81. [PubMed: 5461083]

6. Lang JW, Andrews HV. Temperature-dependent sex determination in crocodylians. *Journal of Experimental Zoology*. 1994; 270:28.
7. Olmo E. Evolution of genome size and DNA base composition in reptiles. *Genetica*. 1981; 57:39.
8. Harshman J, Huddleston CJ, Bollback JP, Parsons TJ, Braun MJ. True and false gharials: a nuclear gene phylogeny of crocodylia. *Syst Biol*. Jun.2003 52:386. [PubMed: 12775527]
9. Oaks JR. A Time-Calibrated Species Tree of Crocodylia Reveals a Recent Radiation of the True Crocodiles. *Evolution*. Nov.2011 65:3285. [PubMed: 22023592]
10. Janke A, Gullberg A, Hughes S, Aggarwal RK, Arnason U. Mitogenomic analyses place the gharial (*Gavialis gangeticus*) on the crocodile tree and provide pre-K/T divergence times for most crocodylians. *J Mol Evol*. Nov.2005 61:620. [PubMed: 16211427]
11. Alföldi J, et al. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature*. Sep 29.2011 477:587. [PubMed: 21881562]
12. Castoe TA, et al. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proceedings of the National Academy of Sciences of the United States of America*. Dec 17.2013 110:20645. [PubMed: 24297902]
13. Shaffer HB, et al. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol*. Mar 28.2013 14:R28. [PubMed: 23537068]
14. Vonk FJ, et al. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proceedings of the National Academy of Sciences of the United States of America*. Dec 17.2013 110:20651. [PubMed: 24297900]
15. Wan QH, et al. Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. *Cell Res*. Sep.2013 23:1091. [PubMed: 23917531]
16. Wang Z, et al. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat Genet*. Jun.2013 45:701. [PubMed: 23624526]
17. Green RE. Materials and methods are available as supplementary materials on Science Online.
18. Hillier LW, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. Dec 9.2004 432:695. [PubMed: 15592404]
19. Huerta-Cepas J, Capella-Gutierrez S, Pryszcz LP, Marcet-Houben M, Gabaldon T. PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Res*. Jan 1.2014 42:D897. [PubMed: 24275491]
20. Lander ES, et al. Initial sequencing and analysis of the human genome. *Nature*. Feb 15.2001 409:860. [PubMed: 11237011]
21. Warren WC, et al. The genome of a songbird. *Nature*. Apr 1.2010 464:757. [PubMed: 20360741]
22. Warren WC, et al. Genome analysis of the platypus reveals unique signatures of evolution. *Nature*. May 8.2008 453:175. [PubMed: 18464734]
23. Waterston RH, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*. Dec 5.2002 420:520. [PubMed: 12466850]
24. Bejerano G, et al. Ultraconserved elements in the human genome. *Science*. May 28.2004 304:1321. [PubMed: 15131266]
25. Faircloth BC, et al. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol*. Oct.2012 61:717. [PubMed: 22232343]
26. McCormack JE, et al. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res*. Feb 2.2012
27. Siepel A, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. Aug.2005 15:1034. [PubMed: 16024819]
28. Chiari Y, Cahais V, Galtier N, Delsuc F. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biol*. 2012; 10:65. [PubMed: 22839781]
29. Crawford NG, et al. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biol Lett*. Oct 23.2012 8:783. [PubMed: 22593086]

30. Tzika AC, Helaers R, Schramm G, Milinkovitch MC. Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent Sauropsida lineages and the phylogenetic position of turtles. *EvoDevo*. 2011; 2:19. [PubMed: 21943375]
31. Chamary JV, Parmley JL, Hurst LD. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature reviews Genetics*. Feb.2006 7:98.
32. Kunstner A, Nabholz B, Ellegren H. Significant selective constraint at 4-fold degenerate sites in the avian genome and its consequence for detection of positive selection. *Genome Biol Evol*. 2011; 3:1381. [PubMed: 22042333]
33. Brochu CA. Morphology, fossils, divergence timing, and the phylogenetic relationships of *Gavialis*. *Syst Biol*. Sep.1997 46:479. [PubMed: 11975331]
34. Glazko GV, Nei M. Estimation of divergence times for major lineages of primate species. *Mol Biol Evol*. Mar.2003 20:424. [PubMed: 12644563]
35. Cordaux R, Lee J, Dinoso L, Batzer MA. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. *Gene*. May 24.2006 373:138. [PubMed: 16527433]
36. Paten B, et al. Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res*. Nov.2008 18:1829. [PubMed: 18849525]
37. Blanchette M, Green ED, Miller W, Haussler D. Reconstructing large regions of an ancestral mammalian genome in silico. *Genome Res*. Dec.2004 14:2412. [PubMed: 15574820]
38. Niimura Y. Olfactory receptor multigene family in vertebrates: from the viewpoint of evolutionary genomics. *Curr Genomics*. Apr.2012 13:103. [PubMed: 23024602]
39. Huerta-Cepas J, Dopazo H, Dopazo J, Gabaldon T. The human phylome. *Genome Biol*. 2007; 8:R109. [PubMed: 17567924]
40. Niimura Y, Nei M. Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS one*. 2007; 2:e708. [PubMed: 17684554]
41. Stevenson, C.; Whitaker, R. Crocodiles. Status Survey and Conservation Action Plan. In: Manolis, SC.; Stevenson, C., editors. *Crocodile Specialist Group: Darwin*. 2010. p. 139-143.
42. Elsey, RM.; Woodward, AR. Crocodiles. Status Survey and Conservation Action Plan. In: Manolis, SC.; Stevenson, C., editors. *Crocodile Specialist Group: Darwin*. 2010. p. 1-4.
43. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. Jul 28.2011 475:493. [PubMed: 21753753]
44. van de Wal RSW, de Boer B, Lourens LJ, Kohler P, Bintanja R. Reconstruction of a continuous high-resolution CO2 record over the past 20 million years. *Clim Past*. 2011; 7:1459.
45. Zhang G, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*. 2014 TBD, TBD.
46. Wu CI, Li WH. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proceedings of the National Academy of Sciences of the United States of America*. Mar.1985 82:1741. [PubMed: 3856856]
47. Lin YH, Waddell PJ, Penny D. Pika and vole mitochondrial genomes increase support for both rodent monophyly and glires. *Gene*. Jul 10.2002 294:119. [PubMed: 12234673]
48. Martin AP, Palumbi SR. Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*. May 1.1993 90:4087. [PubMed: 8483925]
49. Gillooly JF, Allen AP, West GB, Brown JH. The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*. Jan 4.2005 102:140. [PubMed: 15618408]
50. Blueweiss L, et al. Relationships between body size and some life history parameters. *Oecologia*. 1978; 37:257.
51. Savage VM, Gillooly JF, Brown JH, West GB, Charnov EL. Effects of body size and temperature on population growth. *Am Nat*. Mar.2004 163:429. [PubMed: 15026978]
52. Lee MS, Cau A, Naish D, Dyke GJ. Dinosaur evolution. Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science*. Aug 1.2014 345:562. [PubMed: 25082702]

53. Gnerre S, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America*. Dec 27.2011 108:1513. [PubMed: 21187386]
54. Shedlock AM, et al. Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc Natl Acad Sci U S A*. Feb 20.2007 104:2767. [PubMed: 17307883]
55. Li R, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res*. Feb.2010 20:265. [PubMed: 20019144]
56. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*. Mar 1.2008 24:637. [PubMed: 18218656]
57. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013; 14
58. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10:R25. [PubMed: 19261174]
59. Smit AFA, Hubley R. 2008–2010
60. Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. *Bioinformatics*. Jun.2005 21(Suppl 1):i152. [PubMed: 15961452]
61. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics*. Jun.2005 21(Suppl 1):i351. [PubMed: 15961478]
62. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*. Jan 14.2008 9
63. Stephen S, Pheasant M, Makunin IV, Mattick JS. Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. *Mol Biol Evol*. Feb.2008 25:402. [PubMed: 18056681]
64. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. Nov 1.2006 22:2688. [PubMed: 16928733]
65. Huerta-Cepas J, Bueno A, Dopazo JQ, Gabaldon T. PhylomeDB: a database for genome-wide collections of gene phylogenies. *Nucleic Acids Res*. Jan.2008 36:D491. [PubMed: 17962297]
66. Wallace MR, et al. A de novo Alu insertion results in neurofibromatosis type 1. *Nature*. Oct 31.1991 353:864. [PubMed: 1719426]
67. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. Aug 1.2009 25:1972. [PubMed: 19505945]
68. Guindon S, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. May.2010 59:307. [PubMed: 20525638]
69. Gabaldon T. Large-scale assignment of orthology: back to phylogenetics? *Genome Biol*. 2008; 9:235. [PubMed: 18983710]
70. Huerta-Cepas J, Dopazo J, Gabaldon T. ETE: a python Environment for Tree Exploration. *BMC Bioinformatics*. 2010; 11:24. [PubMed: 20070885]
71. Huerta-Cepas J, Gabaldon T. Assigning duplication events to relative temporal scales in genome-wide studies. *Bioinformatics*. Jan 1.2011 27:38. [PubMed: 21075746]
72. Wehe A, Bansal MS, Burleigh JG, Eulenstein O. DupTree: a program for large-scale phylogenetic analyses using gene tree parsimony. *Bioinformatics*. Jul 1.2008 24:1540. [PubMed: 18474508]
73. Tamura K, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. Oct.2011 28:2731. [PubMed: 21546353]
74. Steiger SS, Kuryshev VY, Stensmyr MC, Kempnaers B, Mueller JC. A comparison of reptilian and avian olfactory receptor gene repertoires: species-specific expansion of group gamma genes in birds. *BMC genomics*. 2009; 10:446. [PubMed: 19772566]
75. Paten B, et al. Cactus: Algorithms for genome multiple sequence alignment. *Genome Res*. Sep. 2011 21:1512. [PubMed: 21665927]
76. Jarvis ED, et al. Whole genome analyses resolve the early branches in the tree of life of modern birds. *Science*. 2014 TBD.

77. Barker FK, Cibois A, Schikler P, Feinstein J, Cracraft J. Phylogeny and diversification of the largest avian radiation. *Proceedings of the National Academy of Sciences of the United States of America*. Jul 27.2004 101:11040. [PubMed: 15263073]
78. Wright TF, et al. A Multilocus Molecular Phylogeny of the Parrots (Psittaciformes): Support for a Gondwanan Origin during the Cretaceous. *Mol Biol Evol*. Oct 1.2008 25:2141. [PubMed: 18653733]
79. Thomson RC, Shaffer HB. Sparse supermatrices for phylogenetic inference: taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Syst Biol*. Jan.2010 59:42. [PubMed: 20525619]
80. Nguyen N, et al. Building a pangenome reference for a population. *Research in Computational Molecular Biology*. 2014; 8349:207.
81. Hickey G, Paten B, Earl D, Zerbino D, Haussler D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics*. May 15.2013 29:1341. [PubMed: 23505295]
82. Hubisz MJ, Pollard KS, Siepel A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Briefings in bioinformatics*. Jan.2011 12:41. [PubMed: 21278375]
83. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. Jul 15.2009 25:1754. [PubMed: 19451168]
84. Jaratlerdsiri W, et al. Comparative analyses reveal adaptive MHC structure in the saltwater crocodile (*Crocodylus porosus*). *PloS one*. (In press).
85. Grabherr MG, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. Jul.2011 29:644. [PubMed: 21572440]
86. Bao Z, Eddy SR. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res*. Aug.2002 12:1269. [PubMed: 12176934]
87. Altschul S, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res*. Sep 1.1997 25:3389. [PubMed: 9254694]
88. Kohany O, Gentles A, Hankus L, Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RebaseSubmitter and Censor. *BMC Bioinformatics*. 2006; 25
89. Smit A, Hubley R, Green P. 1996–2010
90. Feschotte C, Keswani U, Ranganathan N, Guibotsy ML, Levine D. Exploring Repetitive DNA Landscapes Using REPCCLASS, a Tool That Automates the Classification of Transposable Elements in Eukaryotic Genomes. *Genome Biol Evol*. 2009; 1:205. [PubMed: 20333191]
91. Harris, RS. The Pennsylvania State University. 2007.
92. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 2005; 33:511. [PubMed: 15661851]
93. Lanfear R, Calcott B, Ho SY, Guindon S. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol*. Jun.2012 29:1695. [PubMed: 22319168]
94. Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC evolutionary biology*. 2014; 14:82. [PubMed: 24742000]
95. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*. Jan 22.2003 19:301. [PubMed: 12538260]
96. Benton MJ, Donoghue PCJ. Paleontological evidence to date the tree of life. *Mol Biol Evol*. Jan. 2007 24:26. [PubMed: 17047029]
97. Ksepka DT, Boyd CA. Quantifying historical trends in the completeness of the fossil record and the contributing factors: an example using Aves. *Paleobiology*. Win;2012 38:112.
98. Muller J, Reisz RR. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *Bioessays*. Oct.2005 27:1069. [PubMed: 16163732]
99. Parham JF, et al. Best practices for justifying fossil calibrations. *Syst Biol*. Mar.2012 61:346. [PubMed: 22105867]
100. Ksepka DT, et al. Synthesizing and databasing fossil calibrations: divergence dating and beyond. *Biol Lett*. Dec 23.2011 7:801. [PubMed: 21525049]
101. Brochu CA. A new Late Cretaceous gavialoid crocodylian from eastern North America and the phylogenetic relationships of thoracosaurus. *J Vert Paleontol*. Sep 10.2004 24:610.

102. Gatesy J, Amato G, Norell M, DeSalle R, Hayashi C. Combined support for wholesale taxic atavism in gavialine crocodylians. *Syst Biol.* Jun.2003 52:403. [PubMed: 12775528]
103. Salisbury SW, Molnar RE, Frey E, Willis PMA. The origin of modern crocodyliforms: new evidence from the Cretaceous of Australia. *P Roy Soc B-Biol Sci.* Oct 7.2006 273:2439.
104. Salisbury SW, Frey E, Martill DM, Buchy MC. A new crocodylian from the Lower Cretaceous Crato Formation of north-eastern Brazil. *Palaeontogr Abt A.* Dec.2003 270:3.
105. Osi A, Clark JM, Weishampel DB. First report on a new basal eusuchian crocodyliform with multicusped teeth from the Upper Cretaceous (Santonian) of Hungary. *Neues Jahrb Geol P-A.* Feb.2007 243:169.
106. Pol D, Turner AH, Norell MA. Morphology of the Late Cretaceous Crocodylomorph *Shamosuchus Djadochtaensis* and a Discussion of Neosuchian Phylogeny as Related to the Origin of Eusuchia. *B Am Mus Nat Hist.* 2009; 1
107. Fortier DC, Schultz CL. A New Neosuchian Crocodylomorph (Crocodyliformes, Mesoeucrocodylia) from the Early Cretaceous of North-East Brazil. *Palaeontology.* Sep.2009 52:991.
108. Buscalioni AD, Piras P, Vullo R, Signore M, Barbera C. Early eusuchia crocodylomorpha from the vertebrate-rich Plattenkalk of Pietraroia (Lower Albian, southern Apennines, Italy). *Zool J Linn Soc.* Dec.2011 163:S199.
109. Holliday CM, Gardner NM. A New Eusuchian Crocodyliform with Novel Cranial Integument and Its Significance for the Origin and Evolution of Crocodylia. *PLoS one.* Jan 31.2012 7
110. Montefeltro FC, Larsson HCE, de Franca MAG, Langer MC. A new neosuchian with Asian affinities from the Jurassic of northeastern Brazil. *Naturwissenschaften.* Sep.2013 100:835. [PubMed: 23893176]
111. Brochu CA. A new gavialoid crocodylian from the Late Cretaceous of eastern North America and the phylogenetic relationships of thoracosauroids. *J Vert Paleontol.* 2004; 24:610.
112. Wu XC, Russell AP, Brinkman DB. A review of *Leidyosuchus canadensis* Lambe, 1907 (Archosauria: Crocodylia) and an assessment of cranial variation based upon new material. *Can J Earth Sci.* Dec.2001 38:1665.
113. Wu, XC. Dinosaur Provincial Park: A Spectacular Ancient Ecosystem Revealed. Currie, PJaK; EB, editors. University of Indiana Press; Bloomington, Indiana: 2005. p. 277-290.
114. Irmis, RB.; Hutchison, JH.; Sertich, JJW.; Titus, AL. At the Top of the Grand Staircase: the Late Cretaceous of Southern Utah. Titus, AL.; Loewen, MA., editors. Indiana University Press; Bloomington: 2013. p. 424-444.
115. Lillegraven JA. A new genus of therian mammal from the Late Cretaceous "El Gallo Formation," Baja California, Mexico. *J Paleontol.* 1976; 50:437.
116. Brochu CA, Langston W, An T. Rowe, phylogenetically significant alligatoroid from the Late Cretaceous (Campanian) of Mexico. *Journal of Vertebrate Paleontology* 33:94A–95A., A new, phylogenetically significant alligatoroid from the Late Cretaceous (Campanian) of Mexico. *J Vert Paleontol.* 2013; 33:94A.
117. Morris, WJ. A Guidebook to the Geology of Peninsular California. Gastil, G.; Lillegraven, JA., editors. American Association of Petroleum Geologists/Society of Economic Geologists; Pacific Section, Los Angeles: 1974. p. 60-66.
118. Renne PR, Fulford MM, Busbyspera C. High-Resolution Ar-40/Ar-39 Chronostratigraphy of the Late Cretaceous El Gallo Formation, Baja-California Del Norte, Mexico. *Geophys Res Lett.* Mar. 1991 18:459.
119. Roberts EM, Dieino AL, Chan MA. 40Ar/39Ar age of the Kaipirowits Formation, southern Utah, and correlation of contemporaneous Campanian strata and vertebrate faunas along the margin of the Western Interior Basin. *Cretaceous Res.* 2005; 26:307.
120. Eberth, DA. Dinosaur Provincial Park: A Spectacular Ancient Ecosystem Revealed. Currie, PJ.; Koppelhus, EB., editors. Indiana University Press; Bloomington: 2005. p. 54-82.
121. Colbert EH, Bird RT. A gigantic crocodile from the Upper Cretaceous beds of Texas. *American Museum Novitates.* 1954; 1688:1.
122. Schwimmer, DR. King of the Crocodylians: The Paleobiology of *Deinosuchus*. Indiana University Press; Bloomington: 2002. p. 220

123. Wahl W, Hogbin J. Deinosuchus material from the Mesaerde Formation of Wyoming: filling in a gap. *J Vert Paleontol.* 2003; 23:107A.
124. Lucas SG, Spielmann JA, Sullivan RM, Lewis C. Late Cretaceous crocodylians from the San Juan Basin, New Mexico. *New Mexico Museum of Natural History and Science Bulletin.* 2006; 35:249.
125. Rivera-Sylva HE, Frey E. The first mandible fragment of Deinosuchus (Eusuchia: Alligatoroidea) discovered in Coahuila, Mexico. *Boletín de la Sociedad Geológica Mexicana.* 2011; 63:459.
126. Buscalioni AD, Sanz JL, Casanovas ML. A new species of the eusuchian crocodile *Diplocynodon* from the Eocene of Spain. *Neues Jahrbuch für Geologie und Paläontologie Abhandlungen.* 1992; 187:1.
127. Ginsburg L, Bulot C. Les *Diplocynodon* (Reptilia, Crocodylia) de l'Orléanien (Miocène inférieur à moyen) de France. *Geodiversitas.* 1997; 19:107.
128. Martin J, Gross M. Taxonomic clarification of *Diplocynodon* Pomel, 1847 (Crocodylia) from the Miocene of Styria, Austria. *Neues Jahrbuch für Geologie und Paläontologie Abhandlungen.* 2011; 261/262:177.
129. Delfino M, Smith T. Reappraisal of the Morphology and Phylogenetic Relationships of the Middle Eocene Alligatoroid *Diplocynodon Deponiae* (Frey, Laemmert, and Riess, 1987) Based on a Three-Dimensional Specimen. *J Vert Paleontol.* 2012; 32:1358.
130. Mook CC. A new crocodylian from the Lance Formation. *American Museum Novitates.* 1941; 1128:1.
131. Buscalioni AD, Ortega F, Vasse D. New crocodiles (Eusuchia: Alligatoroidea) from the Upper Cretaceous of southern Europe. *Comptes Rendus de l'Academie des Sciences de Paris, Sciences de la Terre et des Planètes.* 1997; 325:525.
132. Martin JE. New material of the Late Cretaceous globidontan *Acynodon iberoccitanus* (Crocodylia) from southern France. *J Vert Paleontol.* Jun 12.2007 27:362.
133. Martin JE. *Allodaposuchus Nopsca*, 1928 (Crocodylia, Eusuchia), from the Late Cretaceous of southern France and its relationships to Alligatoroidea. *J Vert Paleontol.* 2010; 30:756.
134. Delfino M, Martin JE, Buffetaut E. A new species of *Acynodon* (Crocodylia) from the Upper Cretaceous (Santonian-Campanian) of Villaggio del Pescatore, Italy. *Palaeontology.* Sep.2008 51:1091.
135. Martin JE, Buffetaut E. *Crocodylus affuvelensis* Matheron, 1869 from the Late Cretaceous of southern France: a reassessment. *Zool J Linn Soc.* Mar.2008 152:567.
136. Puertolas E, Canudo JI, Cruzado-Caballero P. A New Crocodylian from the Late Maastrichtian of Spain: Implications for the Initial Radiation of Crocodyloids. *PloS one.* Jun 8.2011 6
137. Buscalioni AD, Ortega F, Weishampel DB, Jianu CM. A revision of the crocodyliform *Allodaposuchus precedens* from the Upper Cretaceous of the Hateg Basin, Romania. Its relevance in the phylogeny of Eusuchia. *J Vert Paleontol.* Mar 26.2001 21:74.
138. Delfino M, et al. A complete skull of *Allodaposuchus precedens* Nopsca, 1928 (Eusuchia) and a reassessment of the morphology of the taxon based on the Romanian remains. *J Vert Paleontol.* Mar 12.2008 28:111.
139. Brochu CA, Parris DC, Grandstaff BS, Denton RK, Gallagher WB. A New Species of *Borealosuchus* (Crocodyliformes, Eusuchia) from the Late Cretaceous-Early Paleogene of New Jersey. *J Vert Paleontol.* 2012; 32:105.
140. Butler RJ, et al. The Sail-Backed Reptile *Ctenosauriscus* from the Latest Early Triassic of Germany and the Timing and Biogeography of the Early Archosaur Radiation. *PloS one.* Oct 14.2011 6
141. Nesbitt SJ. The Early Evolution of Archosaurs: Relationships and the Origin of Major Clades. *B Am Mus Nat Hist.* 2011; 1
142. Nesbitt SJ, Liu J, Li C. A sail-backed suchian from the Heshangou Formation (Early Triassic: Olenekian) of China. *Earth Env Sci T R So.* 2010; 101:271.
143. Brusatte SL, Niedzwiedzki G, Butler RJ. Footprints pull origin and diversification of dinosaur stem lineage deep into Early Triassic. *P Roy Soc B-Biol Sci.* Apr 7.2011 278:1107.
144. Lyson TR, Bever GS, Bhullar BAS, Joyce WG, Gauthier JA. Transitional fossils and the origin of turtles. *Biol Lett.* Dec 23.2010 6:830. [PubMed: 20534602]

145. Lyson TR, Bever GS, Scheyer TM, Hsiang AY, Gauthier JA. Evolutionary Origin of the Turtle Shell. *Curr Biol.* Jun 17.2013 23:1113. [PubMed: 23727095]
146. Gauthier J, Kluge AG, Rowe T. Amniote Phylogeny and the Importance of Fossils. *Cladistics-the International Journal of the Willi Hennig Society.* Jun.1988 4:105.
147. Lee MSY. Pareiasaur phylogeny and the origin of turtles. *Zool J Linn Soc.* Jul.1997 120:197.
148. Rieppel O, Reisz RR. The origin and early evolution of turtles. *Annu Rev Ecol Syst.* 1999; 30:1.
149. Carroll, RL. *Morphology and Evolution of Turtles.* Brinkman, DB.; Holroyd, PA.; Gardner, JD., editors. Springer; Dordrecht: 2013. p. 19-38.
150. Katsu Y, Braun EL, Guillette LJ, Iguchi T. From Reptilian Phylogenomics to Reptilian Genomes: Analyses of c-Jun and DJ-1 Proto-Oncogenes. *Cytogenet Genome Res.* 2009; 127:79. [PubMed: 20234127]
151. Field DJ, et al. Toward consilience in reptile phylogeny: miRNAs support an archosaur, not lepidosaur, affinity for turtles. *Evolution & development.* Jul-Aug;2014 16:189. [PubMed: 24798503]
152. Clarke JA, Tambussi CP, Noriega JI, Erickson GM, Ketchum RA. Definitive fossil evidence for the extant avian radiation in the Cretaceous. *Nature.* Jan 20.2005 433:305. [PubMed: 15662422]
153. Longrich NR, Tokaryk T, Field DJ. Mass extinction of birds at the Cretaceous-Paleogene (K-Pg) boundary. *Proceedings of the National Academy of Sciences of the United States of America.* Sep 13.2011 108:15253. [PubMed: 21914849]
154. Feduccia A. Avian extinction at the end of the Cretaceous: Assessing the magnitude and subsequent explosive radiation. *Cretaceous Res.* 2014; 50:1.
155. Lee MS, Cau A, Naish D, Dyke GJ. Morphological clocks in paleontology, and a mid-Cretaceous origin of crown Aves. *Syst Biol.* May.2014 63:442. [PubMed: 24449041]
156. Sicheritz-Ponten T, Andersson SG. A phylogenomic approach to microbial evolution. *Nucleic Acids Res.* Jan 15.2001 29:545. [PubMed: 11139625]
157. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol.* Mar 25.1981 147:195. [PubMed: 7265238]
158. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* Aug 19.2004 5:113. [PubMed: 15318951]
159. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in bioinformatics.* Jul.2008 9:286. [PubMed: 18372315]
160. Subramanian AR, Kaufmann M, Morgenstern B. DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. *Algorithms for molecular biology: AMB.* 2008; 3:6. [PubMed: 18505568]
161. Landan G, Graur D. Heads or tails: a simple reliability check for multiple sequence alignments. *Mol Biol Evol.* Jun.2007 24:1380. [PubMed: 17387100]
162. Wallace IM, O'Sullivan O, Higgins DG, Notredame C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.* 2006; 34:1692. [PubMed: 16556910]
163. Gascuel O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol.* Jul.1997 14:685. [PubMed: 9254330]
164. Akaike H. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control.* 1974; 19:716.
165. Huerta-Cepas J, et al. PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic Acids Res.* Jan.2011 39:D556. [PubMed: 21075798]
166. Shimodaira H, Hasegawa M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics.* Dec.2001 17:1246. [PubMed: 11751242]
167. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol.* Oct.2011 7:e1002195. [PubMed: 22039361]
168. Medina I, et al. Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res.* Jul.2010 38:W210. [PubMed: 20478823]

169. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one*. 2011; 6:e21800. [PubMed: 21789182]
170. Song S, Liu L, Edwards SV, Wu S. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences of the United States of America*. Sep 11.2012 109:14942. [PubMed: 22930817]
171. Teeling EC, Hedges SB. Making the impossible possible: rooting the tree of placental mammals. *Mol Biol Evol*. Sep.2013 30:1999. [PubMed: 23813980]
172. Nishihara H, Maruyama S, Okada N. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proceedings of the National Academy of Sciences of the United States of America*. Mar 31.2009 106:5235. [PubMed: 19286970]
173. Kurtz S, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004; 5:R12. [PubMed: 14759262]
174. Raney BJ, et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics*. Nov 30.2013
175. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. Mar 15.2010 26:841. [PubMed: 20110278]
176. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. Apr.2013 30:772. [PubMed: 23329690]
177. Sonnhammer EL, vonHeijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings/ International Conference on Intelligent Systems for Molecular Biology; ISMB International Conference on Intelligent Systems for Molecular Biology*. 1998; 6:175.
178. Fujita MK, Edwards SV, Ponting CP. The Anolis lizard genome: an amniote genome without isochores. *Genome Biol Evol*. 2011; 3:974. [PubMed: 21795750]
179. Fullerton SM, Bernardo Carvalho A, Clark AG. Local rates of recombination are positively correlated with GC content in the human genome. *Mol Biol Evol*. Jun.2001 18:1139. [PubMed: 11371603]
180. Hardison RC, et al. Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res*. Jan.2003 13:13. [PubMed: 12529302]
181. Mouchiroud D, et al. The distribution of genes in the human genome. *Gene*. Apr.1991 100:181. [PubMed: 2055469]
182. Boussau B, Gouy M. Efficient likelihood computations with nonreversible models of evolution. *Syst Biol*. Oct.2006 55:756. [PubMed: 17060197]
183. Romiguier J, Ranwez V, Douzery EJ, Galtier N. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. *Genome Res*. Aug.2010 20:1001. [PubMed: 20530252]
184. Duret L, Arndt PF. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet*. May.2008 4:e1000071. [PubMed: 18464896]
185. Benton MJ, Donoghue PCJ. Paleontological evidence to date the tree of life (vol 24, pg 26, 2007). *Mol Biol Evol*. Mar.2007 24:889.
186. dos Reis M, et al. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings Biological sciences/The Royal Society*. Sep 7.2012 279:3491. [PubMed: 22628470]
187. Reisz RR, Modesto SP, Scott DM. A new Early Permian reptile and its significance in early diapsid evolution. *Proceedings Biological sciences/The Royal Society*. Dec 22.2011 278:3731. [PubMed: 21525061]
188. Jones ME, et al. Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). *BMC evolutionary biology*. 2013; 13:208. [PubMed: 24063680]
189. Hirayama R. Oldest known sea turtle. *Nature*. Apr 16.1998 392:705.
190. Niimura Y, Nei M. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proceedings of the National Academy of Sciences of the United States of America*. Apr 26.2005 102:6039. [PubMed: 15824306]

191. Niimura Y. On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Biol Evol.* 2009; 1:34. [PubMed: 20333175]
192. Nei M, Niimura Y, Nozawa M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nature reviews Genetics.* Dec.2008 9:951.
193. Joanen T, Mcnease LL. Ecology and physiology of nesting and early development of the American alligator. *American Zoologist.* 1989; 29:987.
194. Ferguson, MWJ. *Biology of the Reptilia.* Gans, C.; Billett, F.; Maderson, PFA., editors. Vol. 14. John Wiley and Sons; New York: 1985. p. 329-492.
195. Klause, S. North Carolina State University. 1984.
196. Caldwell, J. *World trade in crocodylian skins 2008–2010.* UNEP-WCMC; Cambridge: 2012.
197. Dalrymple GH. Growth of American alligators in the Shark Valley region of Everglades National Park. *Copeia.* Feb 2.1996 212
198. Joanen T. Nesting ecology of alligators in Louisiana. *Proceedings of the Annual Conference of Southeastern Association of Game and Fish Commissioners.* 1969; 23:141.
199. Abercrombie, CL.; Hall, P. *Crocodiles: their ecology, management, and conservation.* International Union for the Conservation of Nature and Natural Resources; Gland, Switzerland: 1989. p. 1-16.
200. Webb GJW, Messel H, Crawford J, Yerbury MJ. Growth rates of *Crocodylus porosus* (Reptilia: Crocodylia) from Arnhem Land, Northern Australia. *Wildl Res.* 1978; 5:385.
201. Webb, GJW.; Manolis, SC.; Brien, ML. *Crocodiles Status Survey and Conservation Action Plan.* Manolis, SC.; Stevenson, C., editors. Darwin; 2010. p. 99-113.
202. Webb GJW, Messel H, Magnusson W. The nesting of *Crocodylus porosus* in Arnhem Land, Northern Territory. *Copeia.* 1977; 1977:238.
203. Stirrat SC, Lawson D, Freeland WJ, Morton R. Monitoring *Crocodylus porosus* populations in the Northern Territory of Australia: a retrospective power analysis. *Wildl Res.* 2001; 28:547.
204. Bustard HR, Maharana S. Size at first breeding in the gharial (*Gavialis gangeticus* (Gmelin)) (Reptilia: Crocodylia) in captivity. *J Bombay Nat Hist Soc.* 1982; 79:206.
205. Das, A.; Das, AK.; Sarma, PK.; Singh, AK. Final technical report. Aaranyak; 2011. An assessment of assisted recovery of gharial (*Gavialis gangeticus*) in river systems of Northeast India.
206. Whitaker, R. *Wildlife Management: Crocodiles and Alligators.* Webb, GJW.; Manolis, SC.; Whitehead, PJ., editors. Surrey Beatty & Sons; Australia: 1987. p. 63-72.
207. Hussain SA. Reproductive success, hatchling survival and rate of increase of gharial *Gavialis gangeticus* in National Chambal Sanctuary, India. *Biol Conserv.* Feb.1999 87:261.

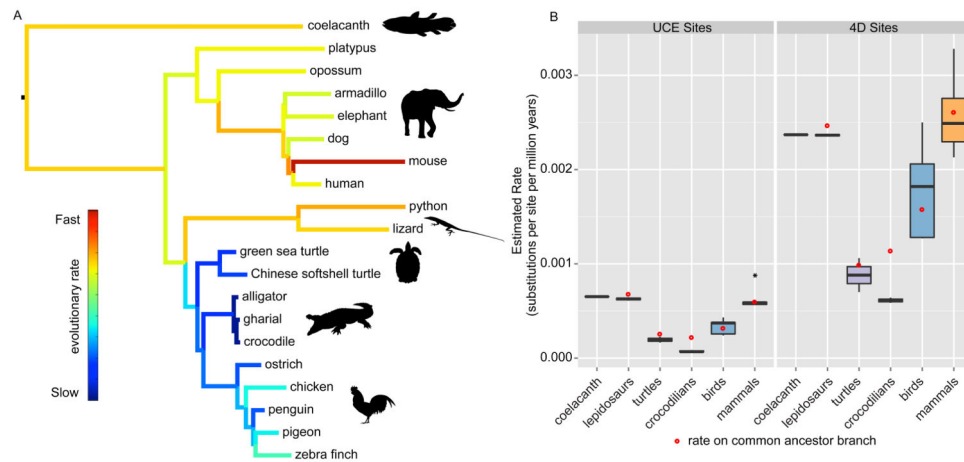


Figure 1. Rates of substitution for ultraconserved elements (UCEs) and fourfold degenerate (4D) sites

(A) Inferred amniote phylogeny based on maximum likelihood analysis of partitioned UCE-anchored loci using RAxML v7.3.4 (17). All branches received 100% bootstrap support. Colors indicate the estimated rates, with cooler colors corresponding to lower rates of molecular evolution (key presented as an insert). (B) Estimated rates of molecular evolution for UCE-anchored loci (left) and 4D sites (right). Red dots indicate the estimated rate for the branch ancestral to the group of interest. The UCE rate for mouse is an outlier; it is indicated by the black dot.

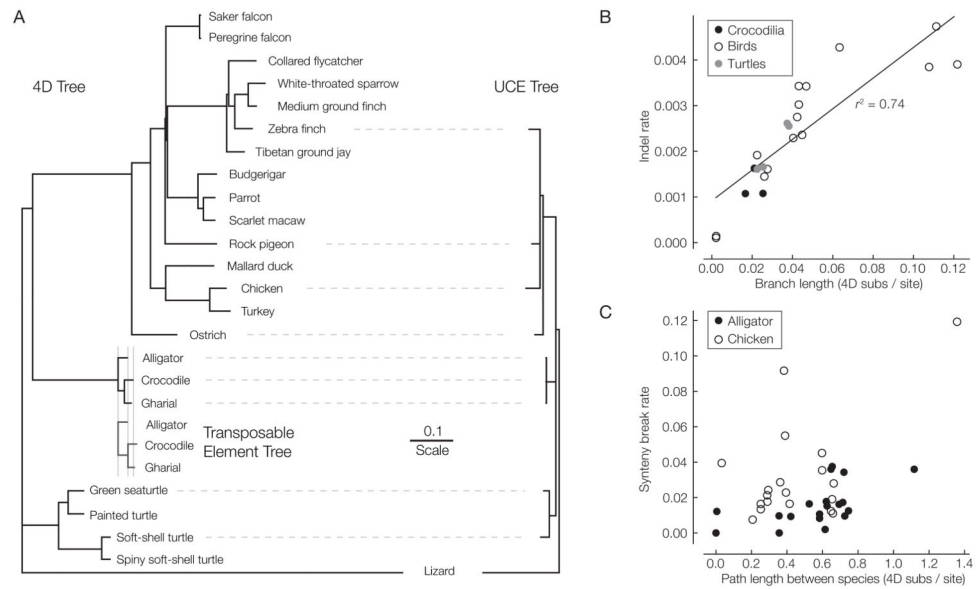


Figure 2. Rates of substitution, micro-indel and break-point evolution
 (A) 4D, transposable element and, for comparison, UCE rates, measured as subs/site. (B) Indel rate vs. 4D subs/site for each extant lineage. (C) Gene syteny breakage rate vs. 4D subs/site distance from the gene annotated species (either alligator or chicken).

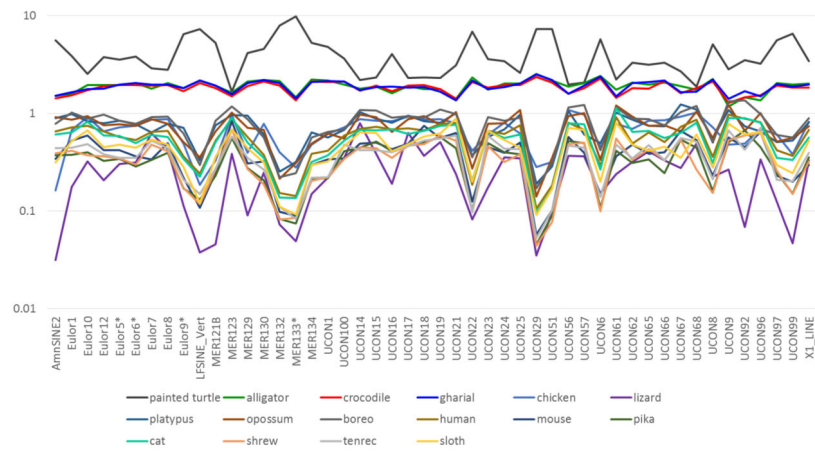


Figure 3. Relative TE numbers among aminotes

Shown are TE copies that predate the speciation of crocodylians and mammals in 16 amniote genomes. The figure displays, on a log scale, 55 unrelated TE families present in all amniote genomes, the bases identified in each individual genome relative to the average identified in all 16 genomes. A * indicates that two or more subfamilies were combined to form a single category. See the Supplemental Materials (17) for the full analysis encompassing all 74 TE families.

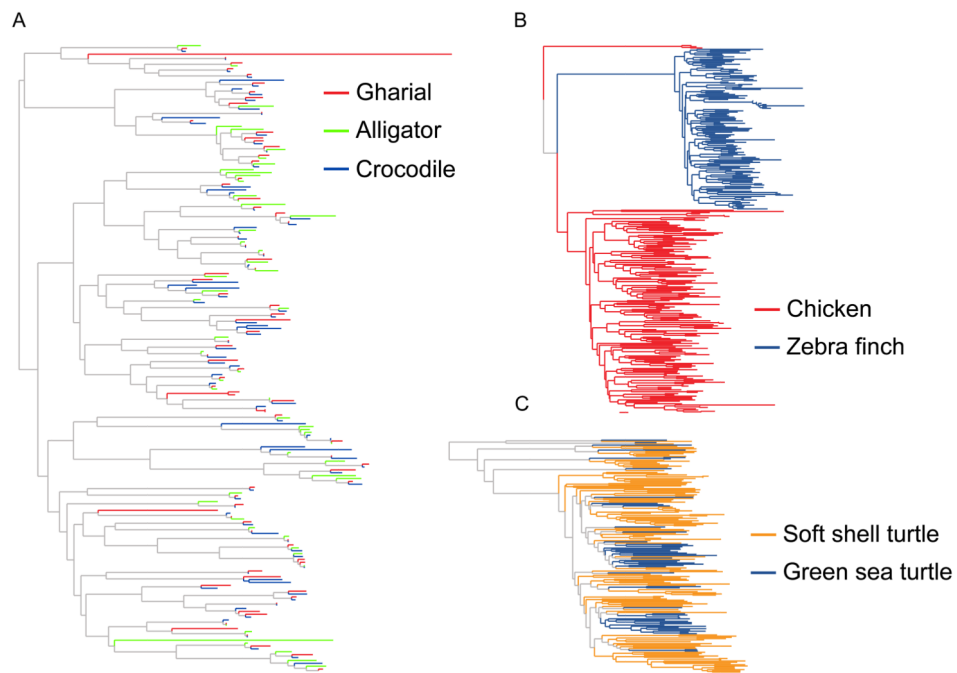


Figure 4. OR expansions/contractions within archosaurs

Subtrees from neighbor-joining phylogenies of the intact crocodilian (A), avian (B) and testudine (C) OR repertoires. Crocodiles are represented by the saltwater crocodile, gharial and American alligator; birds are represented by chicken and zebra finch; and testudines are represented by the soft shell and green sea turtle. Note the paucity of lineage-specific clades, in color, among crocodilian ORs, as opposed to avian and testudine ORs. Most crocodilian ORs are outparalogs, groups of paralogous genes that emerged prior to the divergence of the species analyzed, whereas the vast majority of avian and testudine ORs fall on monophyletic groups of inparalogs, groups of paralogous genes the emerged after the divergence of the species analyzed. Neighbor-joining trees were inferred using MEGA v5, a Poisson model of substitution, and 1000 bootstrap iterations were performed to evaluate support. Also see Figure S24

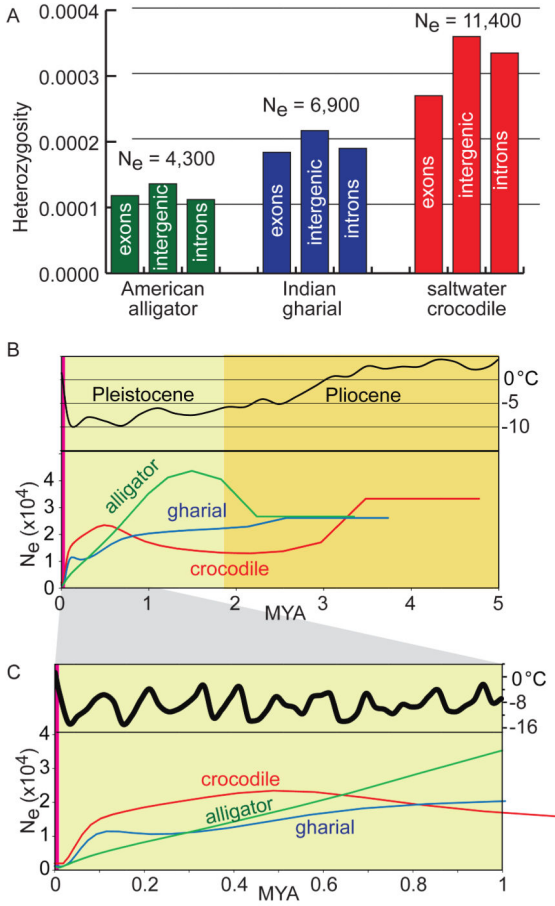


Figure 5. Crocodilian genetic diversity and population history

A. The rate of observed heterozygosity within annotated exons, intergenic sequence, and introns. B. PSMC estimates of the historical crocodilian N_e inferred from each genome shown in a time span of (B) 5 million years and (C) 1 million years under the assumption of a generation time of 20 years.

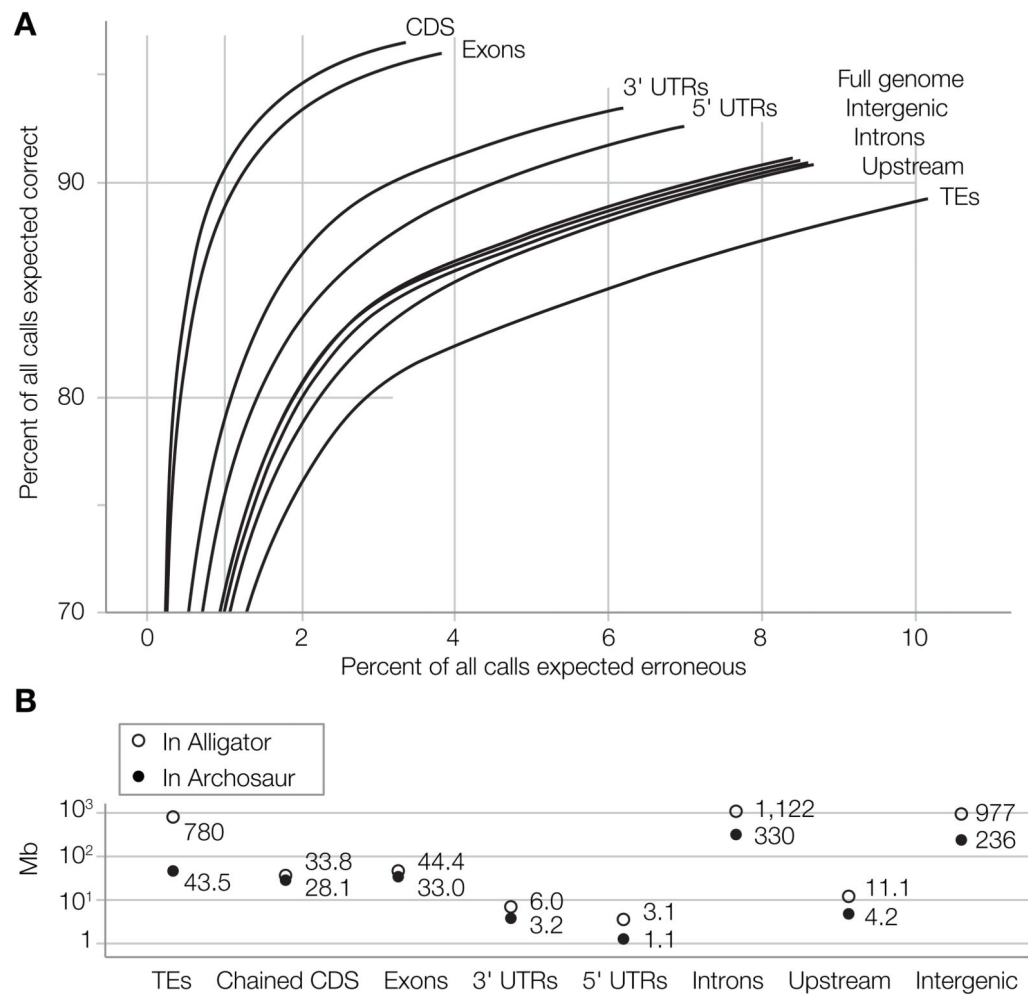


Figure 6. Analyzing the archosaur assembly

A. Expected base reconstruction accuracy B. Total archosaur bases assembled in several annotated functional classes and the number of bases in each category from the alligator genome.