RESEARCH ARTICLE SUMMARY

AVIAN GENOMICS

Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs

Richard E. Green,* Edward L. Braun, Joel Armstrong, Dent Earl, Ngan Nguyen, Glenn Hickey, Michael W. Vandewege, John A. St. John, Salvador Capella-Gutiérrez, Todd A. Castoe, Colin Kern, Matthew K. Fujita, Juan C. Opazo, Jerzy Jurka, Kenji K. Kojima, Juan Caballero, Robert M. Hubley, Arian F. Smit, Roy N. Platt, Christine A. Lavoie, Meganathan P. Ramakodi, John W. Finger Jr., Alexander Suh, Sally R. Isberg, Lee Miles, Amanda Y. Chong, Weerachai Jaratlerdsiri, Jaime Gongora, Christopher Moran, Andrés Iriarte, John McCormack, Shane C. Burgess, Scott V. Edwards, Eric Lyons, Christina Williams, Matthew Breen, Jason T. Howard, Cathy R. Gresham, Daniel G. Peterson, Jürgen Schmitz, David D. Pollock, David Haussler, Eric W. Triplett, Guojie Zhang, Naoki Irie, Erich D. Jarvis, Christopher A. Brochu, Carl J. Schmidt, Fiona M. McCarthy, Brant C. Faircloth, Federico G. Hoffmann, Travis C. Glenn, Toni Gabaldón, Benedict Paten, David A. Ray*

INTRODUCTION: Crocodilians and birds are the two extant clades of archosaurs, a group that includes the extinct dinosaurs and pterosaurs. Fossils suggest that living crocodilians (alligators, crocodiles, and gharials) have a most recent common ancestor 80 to 100 million years ago. Extant crocodilians are notable for their distinct morphology, limited intraspecific variation, and slow karyotype evolution. Despite their unique biology and phylogenetic position, little is known about genome evolution within crocodilians.

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.1254449

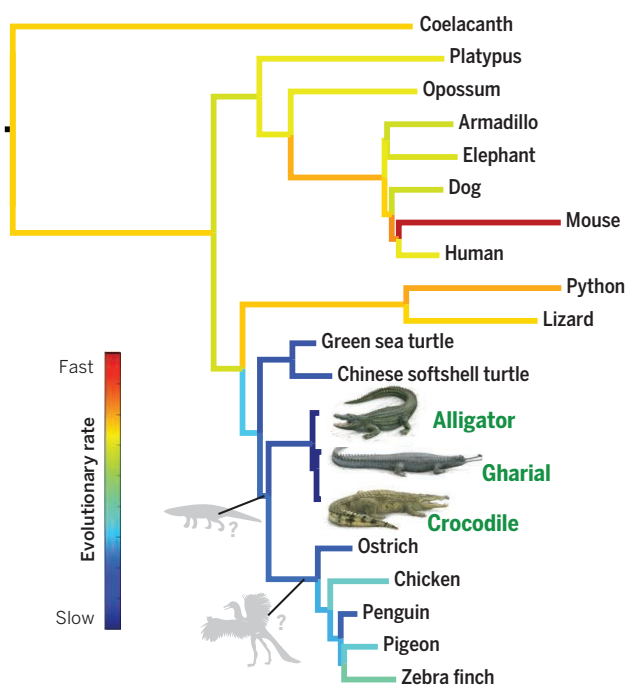
RATIONALE: Genome sequences for the American alligator, saltwater crocodile, and Indian gharial—representatives of all three extant crocodilian families—were obtained to facilitate better understanding of the unique biology of this group and provide a context for studying avian genome evolution. Sequence data from these three crocodilians and birds also allow reconstruction of the ancestral archosaurian genome.

RESULTS: We sequenced shotgun genomic libraries from each species and used a variety of assembly strategies to obtain draft genomes for these three crocodilians. The assembled scaffold N50 was highest for the alligator (508 kilobases). Using a panel of reptile genome sequences, we generated phylogenies that confirm the sister relationship between crocodiles and gharials, the relationship with birds as members of extant Archosauria, and the outgroup status of turtles relative to birds and crocodilians.

We also estimated evolutionary rates along branches of the tetrapod phylogeny using two approaches: ultraconserved element–anchored sequences and fourfold degenerate sites within stringently filtered orthologous gene alignments. Both analyses indicate that the rates of base substitution along the crocodilian and turtle lineages are extremely low. Supporting observations were made for transposable element content and for gene family evolution. Analysis of whole-genome alignments across a panel of reptiles and mammals showed that the rate of accumulation of microinsertions and microdeletions is proportionally lower in crocodilians, consistent with a single underlying cause of a reduced rate of evolutionary change rather than intrinsic differences in base-repair machinery. We hypothesize that this single cause may be a consistently longer generation time over the evolutionary history of Crocodylia.

Low heterozygosity was observed in each genome, consistent with previous analyses, including the Chinese alligator. Pairwise sequential Markov chain analysis of regional heterozygosity indicates that during glacial cycles of the Pleistocene, each species suffered reductions in effective population size. The reduction was especially strong for the American alligator, whose current range extends farthest into regions of temperate climates.

CONCLUSION: We used crocodilian, avian, and outgroup genomes to reconstruct 584 megabases of the archosaurian common ancestor genome and the genomes of key ancestral nodes. The estimated accuracy of the archosaurian genome reconstruction is 91% and is higher for conserved regions such as genes. The reconstructed genome can be improved by adding more crocodilian and avian genome assemblies and may provide a unique window to the genomes of extinct organisms such as dinosaurs and pterosaurs. ■



Evolutionary rates of tetrapods inferred from DNA sequences anchored by ultraconserved elements. Evolutionary rates among reptiles vary, with especially low rates among extant crocodilians but high rates among squamates. We have reconstructed the genomes of the common ancestor of birds and of all archosaurs (shown in gray silhouette, although the morphology of these species is uncertain).

A complete list of affiliations is available in the full article online.
*Corresponding author. E-mail: ed@soe.ucsc.edu (R.E.G.); david.a.ray@ttu.edu (D.A.R.)
Cite this article as R. E. Green *et al.*, *Science* **346**, 1254449 (2014). DOI: 10.1126/science.1254449

RESEARCH ARTICLE

Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs

Richard E. Green,[1]* Edward L. Braun,[2] Joel Armstrong,[1,3] Dent Earl,[1,3] Ngan Nguyen,[1,3] Glenn Hickey,[1,3] Michael W. Vandewege,[4] John A. St. John,[1]§ Salvador Capella-Gutiérrez,[5,6] Todd A. Castoe,[7,8] Colin Kern,[9] Matthew K. Fujita,[8] Juan C. Opazo,[10] Jerzy Jurka,[11]† Kenji K. Kojima,[11] Juan Caballero,[12] Robert M. Hubley,[12] Arian F. Smit,[12] Roy N. Platt,[4,13] Christine A. Lavoie,[4] Meganathan P. Ramakodi,[4,13]‡ John W. Finger Jr.,[14] Alexander Suh,[15,16] Sally R. Isberg,[17,18,19] Lee Miles,[18]# Amanda Y. Chong,[18] Weerachai Jaratlerdsiri,[18] Jaime Gongora,[18] Christopher Moran,[18] Andrés Iriarte,[20] John McCormack,[21] Shane C. Burgess,[22] Scott V. Edwards,[23] Eric Lyons,[24] Christina Williams,[25] Matthew Breen,[25] Jason T. Howard,[26] Cathy R. Gresham,[13] Daniel G. Peterson,[13,27] Jürgen Schmitz,[15] David D. Pollock,[7] David Haussler,[3,28] Eric W. Triplett,[29] Guojie Zhang,[30,31] Naoki Irie,[32] Erich D. Jarvis,[26] Christopher A. Brochu,[33] Carl J. Schmidt,[34] Fiona M. McCarthy,[35] Brant C. Faircloth,[36,37] Federico G. Hoffmann,[4,13] Travis C. Glenn,[14] Toni Gabaldón,[5,6,38] Benedict Paten,[3] David A. Ray[4,13,39]*

To provide context for the diversification of archosaurs—the group that includes crocodilians, dinosaurs, and birds—we generated draft genomes of three crocodilians: *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 crocodilians 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 comparatively rapid evolution is derived in birds. The data also provided the opportunity to analyze heterozygosity in crocodilians, which indicates a likely reduction in population size for all three taxa through the Pleistocene. Finally, these data combined with newly published bird genomes allowed us to reconstruct the partial genome of the common ancestor of archosaurs, thereby providing a tool to investigate the genetic starting material of crocodilians, birds, and dinosaurs.

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

Crocodilians 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 females of all crocodilians (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 microchromosomes), 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 crocodilian lineages (*3*, *8*–*10*). These crocodilian genomes augment the list of assembled genomes from avian and nonavian reptiles (*11*–*16*), allowing us to probe the lineage-specific novelties in avian and crocodilian 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 to S3) from paired-end and mate-pair libraries from each species: alligator, crocodile, and gharial. The assembly strategy for each taxon differed because of varying legacy data and developments in library preparation methods during the course of the project (*17*). Genome scaffolding of the alligator (and, to a lesser extent, the crocodile) was aided by the availability

[1]Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA. [2]Department of Biology and Genetics Institute, University of Florida, Gainesville, FL 32611, USA. [3]Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA 95064, USA. [4]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA. [5]Bioinformatics and Genomics Programme, Centre for Genomic Regulation, 08003 Barcelona, Spain. [6]Universitat Pompeu Fabra, 08003 Barcelona, Spain. [7]Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045, USA. [8]Department of Biology, University of Texas, Arlington, TX 76019, USA. [9]Department of Computer and Information Sciences, University of Delaware, Newark, DE 19717, USA. [10]Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile. [11]Genetic Information Research Institute, Mountain View, CA 94043, USA. [12]Institute for Systems Biology, Seattle, WA 98109, USA. [13]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA. [14]Department of Environmental Health Science, University of Georgia, Athens, GA 30602, USA. [15]Institute of Experimental Pathology (ZMBE), University of Münster, D-48149 Münster, Germany. [16]Department of Evolutionary Biology (EBC), Uppsala University, SE-752 36 Uppsala, Sweden. [17]Porosus Pty. Ltd., Palmerston, NT 0831, Australia. [18]Faculty of Veterinary Science, University of Sydney, Sydney, NSW 2006, Australia. [19]Centre for Crocodile Research, Noonamah, NT 0837, Australia. [20]Departamento de Desarrollo Biotecnológico, Instituto de Higiene, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay. [21]Moore Laboratory of Zoology, Occidental College, Los Angeles, CA 90041, USA. [22]College of Agriculture and Life Sciences, University of Arizona, Tucson, AZ 85721, USA. [23]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [24]School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA. [25]Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, NC 27607, USA. [26]Howard Hughes Medical Institute, Department of Neurobiology, Duke University Medical Center, Durham, NC 27710, USA. [27]Department of Plant and Soil Sciences, Mississippi State University, Mississippi State, MS 39762, USA. [28]Howard Hughes Medical Institute, Bethesda, MD 20814, USA. [29]Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, USA. [30]China National GeneBank, BGI-Shenzhen, Shenzhen, China. [31]Center for Social Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [32]Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan. [33]Department of Earth and Environmental Sciences, University of Iowa, Iowa City, IA 52242, USA. [34]Department of Animal and Food Sciences, University of Delaware, Newark, DE 19717, USA. [35]School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ 85721, USA. [36]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90019, USA. [37]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA. [38]Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain. [39]Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA.

***Corresponding author. E-mail: ed@soe.ucsc.edu (R.E.G.); david.a.ray@ttu.edu (D.A.R.)** †Deceased. ‡Present address: Cancer Prevention and Control, Fox Chase Cancer Center, Philadelphia, PA 19111, USA, and Department of Biology, Temple University, Philadelphia, PA 19122, USA. §Present address: Driver Group LLC, 1700 Owens Street, Suite 200, San Francisco, CA 94158, USA. #Present address: FB Rice Patent & Trade Mark Attorneys, 44 Market Street, Sydney, NSW 2000, Australia.

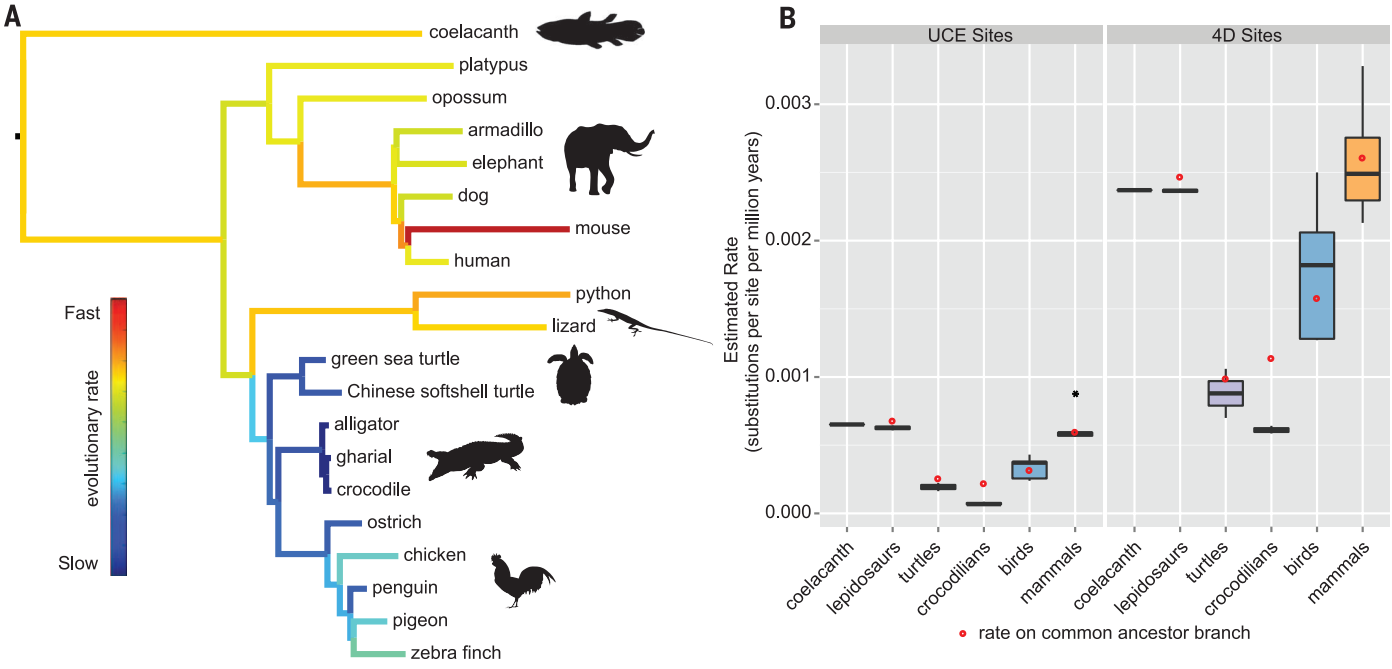


Fig. 1. Rates of substitution for ultraconserved elements (UCEs) and fourfold degenerate (4D) sites. (**A**) Inferred amniote phylogeny based on maximum likelihood analysis of partitioned UCE-anchored loci using RAxML v7.3.4 (*17*). All branches received 100% bootstrap support. Colors indicate the estimated rates, with cooler colors corresponding to lower rates of molecular evolution. (**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 and is indicated by a black dot.

of bacterial artificial chromosome (BAC) sequences and BAC end-sequence data. RNA-seq 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 to S4.

Gene annotation was accomplished using a combination of transcriptome sequencing (RNA-seq) data and homology-based analyses (*17*). We identified 23,323 protein-coding genes in the alligator, 13,321 in the crocodile, and 14,043 in the gharial (table S5). The unevenness likely reflects the larger overall scaffold size (N50) of the alligator genome assembly (table S4) and the predominance of alligator transcriptome data used to guide gene identification (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 crocodilian genes on the basis of orthology to vertebrates with existing standardized nomenclature (human, mouse, anole, chicken, and zebrafish). Between 60 and 70% of crocodilian proteins had conserved functional motifs on the basis of comparison to other vertebrates, and we provided 377,441 Gene Ontology (GO) annotations for 43,436 crocodilian proteins.

Transposable elements (TEs) were identified de novo in all three crocodilians, and analyses resulted in a library of 1269 different TEs (table S7)—a large number for a vertebrate. This high TE count in crocodilian genomes is attributable, at least in part, to the apparently low rate of base substitution in crocodilians, as discussed below. We find that ~37.5% of each crocodilian genome can be annotated as TEs (table S7), a value intermediate between mammals (40 to 60%) and birds (12 to 15%) (*18*–*23*).

Ultraconserved element phylogeny and molecular evolution

Ultraconserved elements (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 (Fig. 1A and figs. S1 to S7).

The phylogeny estimated using UCE-anchored data largely agrees with other studies (*8*, *10*, *28*, *29*). For example, we recovered Longirostres (crocodiles + gharials) within Crocodylia, found crocodilians to be the sister group of birds (supporting the clade Archosauria), and confirmed 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 (Fig. 1A), including the slowly evolving turtles. To explore the evolutionary tempo of crocodilians, we used divergence time estimates for critical nodes to derive estimates of absolute substitution rates across the tree (*17*). These estimates suggest that the molecular evolution of crocodilians is slower than that of all other lineages (figs. 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 important, the availability of multiple bird, crocodilian, and turtle genomes allows us to estimate the ancestral rates for these groups (Fig. 1B). Using a variety of calibration times for the TMRCA (time to most recent common ancestor) of birds, crocodilians, and archosaurs (fig. S8 and table S9), we find that the rate of UCE evolution for the avian stem lineage was similar to that of extant avian lineages (Fig. 1B and fig. S7). In contrast, the crocodilian 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

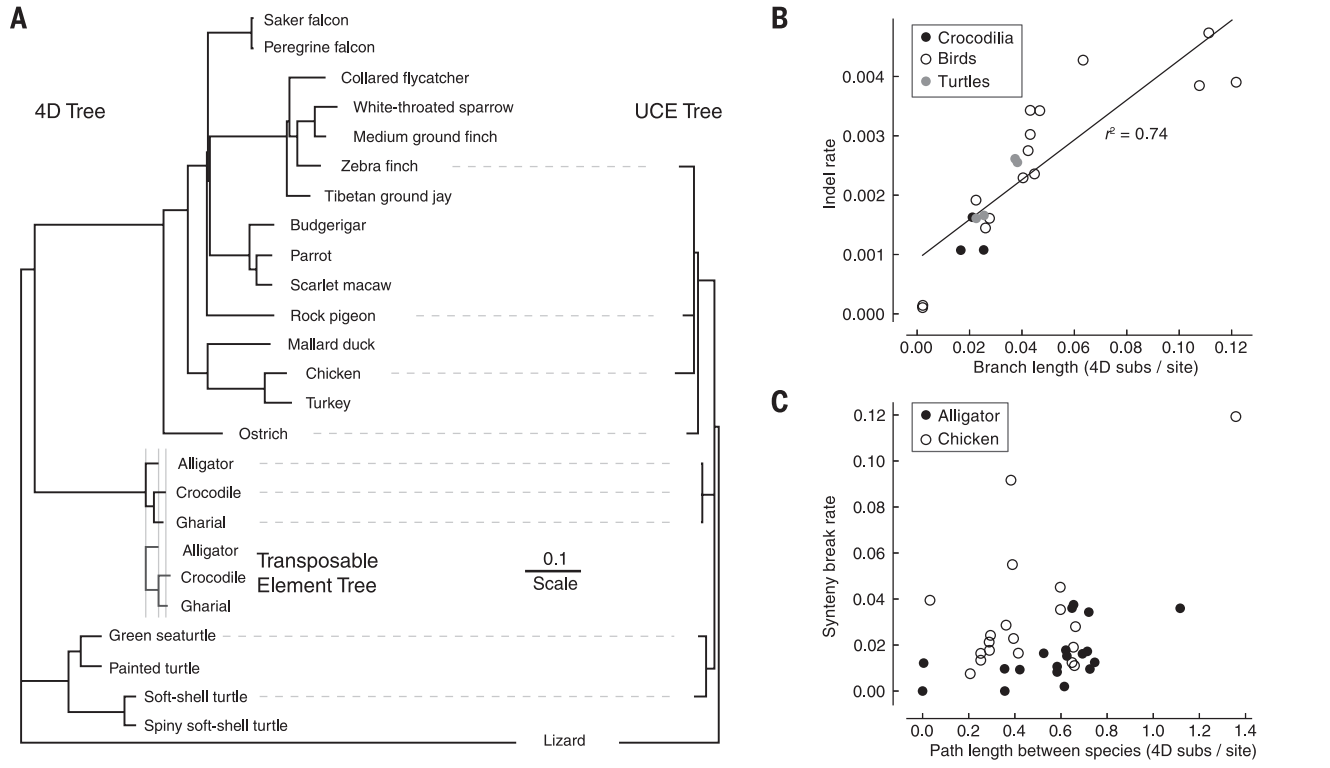


Fig. 2. Rates of substitution, micro-indels, and break-point evolution. (**A**) Rates of substitution at 4D sites, transposable elements (TEs), and, for comparison, UCE-anchored loci. Scale bar denotes substitutions per site. (**B**) Indel rate versus 4D substitutions per site for each extant lineage. (**C**) Gene synteny breakage rate versus 4D substitutions per site, each measured with respect to either alligator or chicken.

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 (fig. S10) produced a tree congruent with the UCE-based phylogeny shown in Fig. 1A and other amniote phylogenies (*28*, *30*). The concatenated alignment of orthologs was then further filtered to extract fourfold 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 (Fig. 1B and figs. S13 to S16).

A larger survey of aligned genes (without the single-copy orthology filters) found 9574 trees that suggested monophyly of birds, turtles, and crocodilians relative to squamates; the vast majority of those (6880; 72%) placed crocodilians and birds together in a clade. Only 28% of trees supported alternate topologies [birds + turtles or crocodilians + turtles (*17*)]. Although the placement of gharial within the crocodilian phylogeny has been contentious over the past several decades (*33*), a clear majority (78.4%) of protein-coding gene trees supported Longirostres (*8*).

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

To explore patterns of molecular evolution across the genomes of crocodilians, we created a whole-genome alignment (WGA) (*17*) that included 23 reptile genomes, including the three crocodilians, 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 crocodilians (table S13); for example, the alligator and crocodile (which shared a common ancestor ~80 to 100 Ma; 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 Ma (*34*), indicating exceptionally low rates of evolution relative 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 (Fig. 2 and table S14) with those occurring within orthologous TE insertions that are shared among the three crocodilians (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 base pairs (bp) per event, filtered to avoid alignment errors] relative to substitutions. This ratio for crocodilians (0.064 micro-indels 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 microdeletions to microinsertions was similar across the tree (average ~1.94; table S16) and concordant with previous estimates from other taxa (*36*, *37*), with no apparent bias toward either category in crocodilians, birds, or turtles (table S20).

Finally, we 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 found high levels of gene order conservation between crocodilian genomes, similar to that between comparably separated bird genomes—a group marked by its extreme syntenic conservation relative to mammals (Fig. 2C). Thus, the low evolutionary rates observed in crocodilians are not specific to substitutions, but also include micro-indels and gene-level rearrangements.

Transposable elements evolve slowly in crocodilians

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

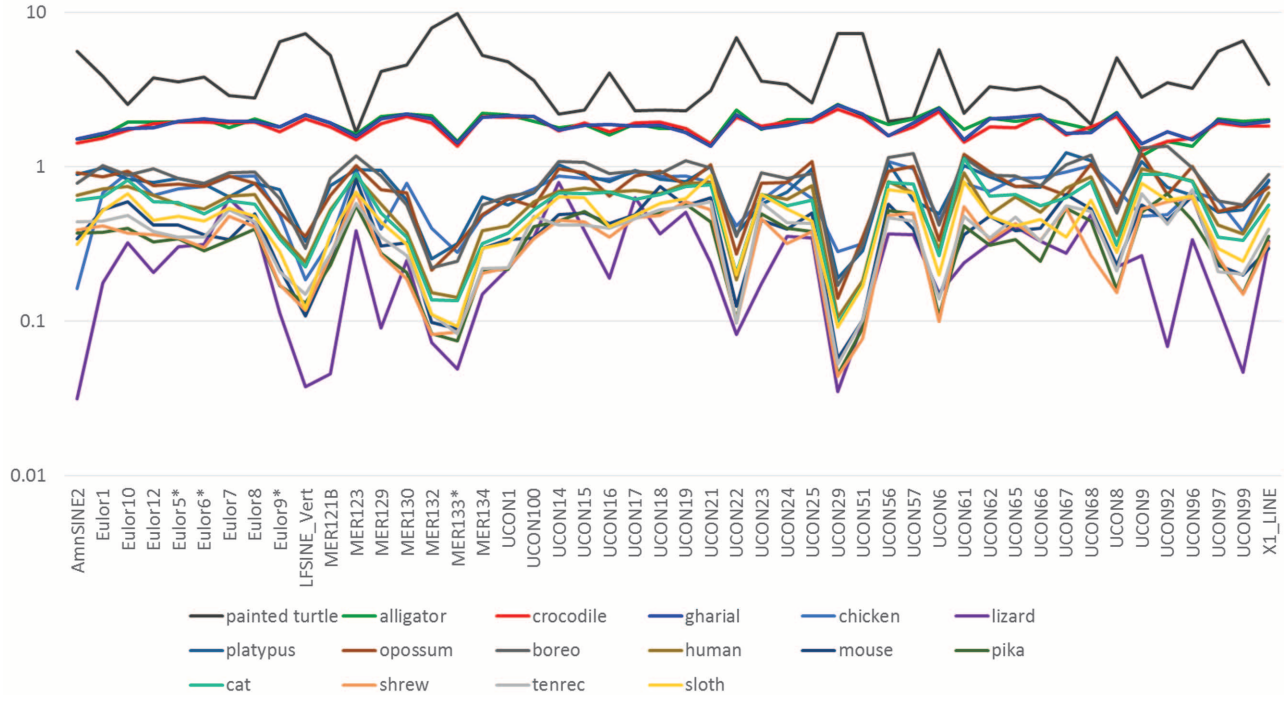


Fig. 3. Relative TE numbers among amniotes. Shown are TE copies that predate the speciation of crocodilians and mammals in 16 amniote genomes. The figure displays 55 unrelated TE families present in all amniote genomes. The numbers of bases, on a log scale, identified in each individual genome relative to the average in all 16 genomes are identified. An asterisk indicates that two or more subfamilies were combined to form a single category. See (*17*) for the full analysis encompassing all 74 TE families.

Longirostres (crocodile + gharial) from alligators, approximately 80 to 100 Ma (table S9). Given 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 crocodilians, with the exception of a minor burst of novel activity in the common ancestor of Longirostres (fig. S20). Indeed, in the ~235 million years between the mammal-crocodilian divergence and the origin of crown crocodilians, at least 823 TE families were active—a rate of around 3.5 TE families per million years. The rate has fallen below 1.0 in both crocodile and gharial since their divergence.

The "visibility" of TE copies introduced before the divergence of mammals and reptiles at ~310 to 330 Ma (table S9) (*17*) provides another line of evidence for the extraordinarily low rate of crocodilian genome evolution. Averaged over 74 unrelated families of such elements (*17*), crocodilian genomes contain a considerable amount of DNA that is recognizably derived from TEs: five times the amount in the typical mammalian genome, three times the amount in the reconstructed boreoeutherian (the mammalian clade comprising primates, rodents, carnivores, bats, and a number of additional orders; "boreo" in Fig. 3) genome (*36*), 3.8 times the identifiable amount in the chicken genome, and 15 times that in the anole genome. Surprisingly, relative to crocodilians, the painted turtle genome contains on average 2.3 times as many bases recognizably derived from each of these repeats (Fig. 3 and figs. S21 to 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 instead represent actual differences in mutation and deletion rates, and that crocodilians 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 crocodilian lineages

We used gene trees from the phylome analysis to search for gene families that underwent duplications within the crocodilian lineage (*17*). Olfactory receptors (ORs) constitute one of the largest vertebrate gene families; they are small, single-exon genes, making them relatively easy to investigate. ORs 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 overrepresented among duplicated genes in crocodilians. Crocodilians possess a diverse OR repertoire, and each species has about 1000 ORs, half of which are likely functional (table S21) (*17*); this 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 (Fig. 4 and fig. S24). In contrast, crocodilian OR repertoires almost exclusively reflect the retention of OR genes present in the common ancestor of the crown crocodilian, followed by a few gains or losses (Fig. 4 and fig. S24). This observation—many retained ancestral genes rather than independent expansion—suggests that crocodilians have achieved a diverse OR repertoire using a strategy of 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 crocodilian 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 relative to most mammalian and avian genomes (Fig. 5). Among the three crocodilian 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 crocodilians examined here is interesting given their recent or current status as endangered species. The gharial is critically endangered because of 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 the alligator, crocodile, and gharial (Fig. 5A) using the neutral mutation rate for crocodilians ($\mu = 7.9 \times 10^{-9}$ substitutions site^{-1} generation^{-1}) calculated from the pairwise divergence (*17*) between the alligator and the saltwater crocodile. We note that the alligator and crocodile were

wild caught and thus were likely to represent the genetic diversity of their respective species, whereas the gharial we sequenced was bred in captivity and of unknown recent ancestry.

The crocodilians 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 sequential Markovian coalescent (PSMC) model (*43*) approach can probe population sizes further into the past than is possible for faster-evolving lineages. Using this model, we found that all three lineages experienced distinct changes in their estimated effective population size (N_e) over the past 7 million years (Fig. 5B and fig. S26). We also included estimates of air temperature data (*44*) to identify any potential relationship of demographic histories to climate change. The results indicate that the crocodile and gharial both maintained relatively stable population sizes through the Pleistocene and Pliocene but both experienced sharp declines during the last cooling cycle, between ~100,000 and 10,000 years before the present (Fig. 5, B and C). 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.

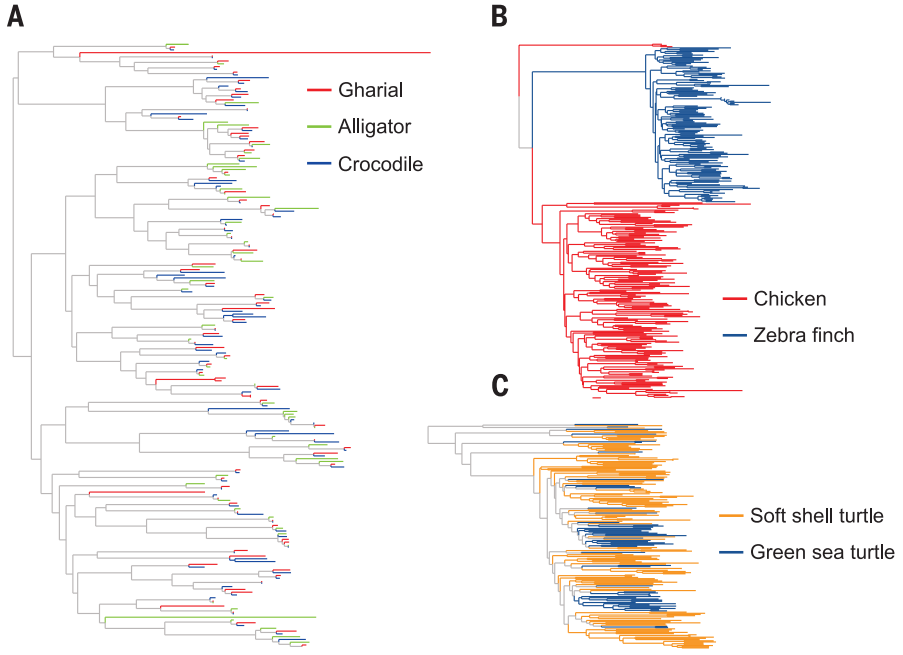


Fig. 4. OR expansions and contractions within archosaurs. Subtrees from neighbor-joining phylogenies of the intact crocodilian (**A**), avian (**B**), and testudine (**C**) OR repertoires. Crocodiles are represented by the gharial, American alligator, and saltwater crocodile; birds are represented by the chicken and zebra finch; and testudines are represented by softshell and green sea turtles. Note the paucity of lineage-specific (colored) clades among crocodilian ORs relative 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; 1000 bootstrap iterations were performed to evaluate support. See also fig. S24.

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. Because of the constant turnover of sequence during the ~300 million years since the divergence of birds and crocodilians 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. Using a standard continuous time substitution model to determine the nucleotide at each position in the

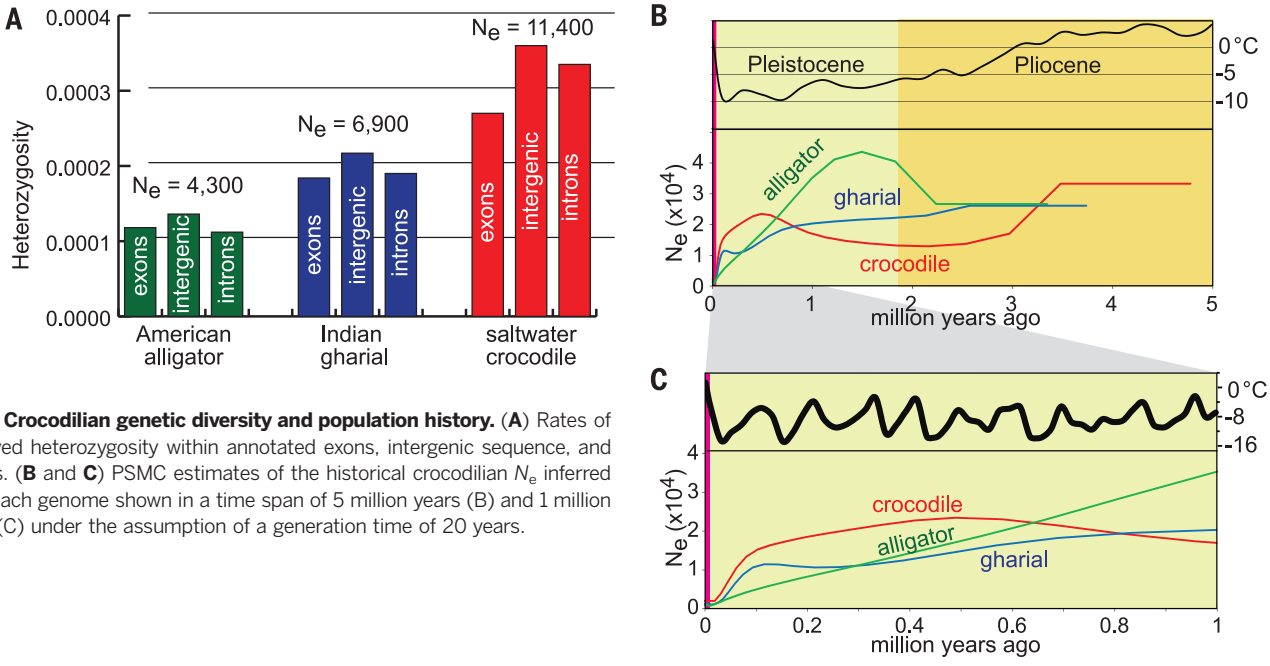


Fig. 5. Crocodilian genetic diversity and population history. (**A**) Rates of observed heterozygosity within annotated exons, intergenic sequence, and introns. (**B** and **C**) PSMC estimates of the historical crocodilian N_e inferred from each genome shown in a time span of 5 million years (B) and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

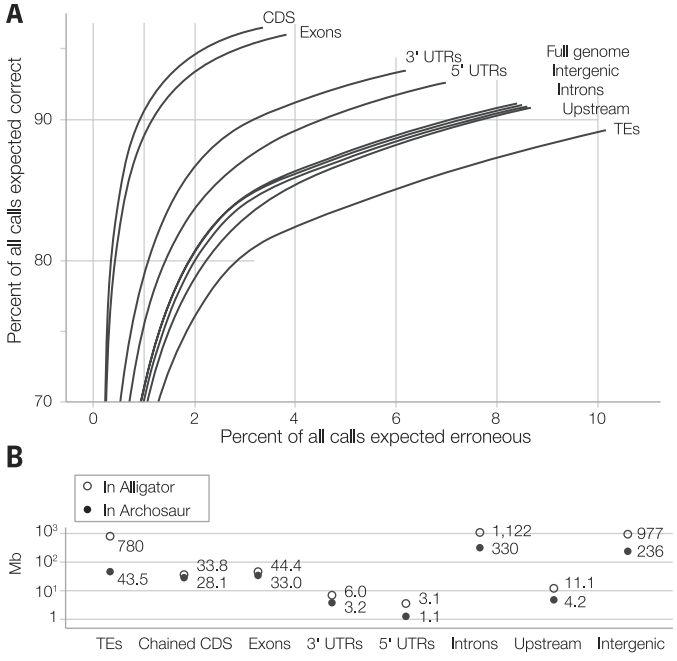


Fig. 6. Analyzing the archosaur assembly. (**A**) Expected base reconstruction accuracy. (**B**) Total archosaur bases assembled in several annotated functional classes and numbers of bases in each category from the alligator genome.

ancestral archosaur genome, the average expected reconstruction accuracy of archosaur bases is 91% (Fig. 6A and fig. 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 (CDSs), 3′ and 5′ untranslated regions (UTRs), exons, upstream sequences (defined by a 500-bp 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. Relative to putatively neutrally evolving elements such as 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; Fig. 6A). Concordantly, while on average only 26% of alligator bases had an aligned base in the archosaur reconstruction, the proportion of annotated bases mapping to archosaur was higher (Fig. 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, which suggests that intragene ordering constraints have helped to preserve sequence structure (figs. S17 to S19).

Discussion

The draft genome assemblies of these three crocodilian taxa add to the growing list of available reptilian genomes and allow a more comprehensive analysis of vertebrate genome evolution. Because crocodilians are the sister group of birds, these three genomes 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 crocodilian 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, such as 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 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; Fig. 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 crocodilians 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 crocodilian 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 crocodilians. The higher rate for stem birds, which is actually similar to that observed for extant birds, could indicate that this lineage had already decreased their 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 crocodilians and other lineages.

Materials and Methods

Sequencing and assembly

Genomic DNA was isolated using blood from four individuals: two *A. mississippiensis* and one each 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 88.7× 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 RNA-seq data, described below, were also used to aid the assembly. Crocodile data consisted of Illumina reads from three libraries ranging from 21.6× to 90.2× coverage. AllPaths-LG was used to assemble the raw data. As with the alligator genome draft, sequences from 360 major histocompatibility complex region BAC assemblies as well as RNA-seq 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 170× coverage. No legacy data were 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 RNA-seq. 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*). RNA-seq data from *A. mississippiensis* were aligned to the draft American alligator genome with 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 crocodilian assemblies with 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, GO, and pathway

information. Gene names were assigned on the basis of orthology or homology to species with a gene nomenclature project by transferring names to the crocodilian genes. GO was assigned to predicted proteins according to a combinatorial approach (*17*). Pathway information was assigned on the basis of 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).

TEs in the genomes were identified and annotated 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 crocodilians resulted in a library of 1269 different TEs. Full details of all sequence annotation protocols are in (*17*).

UCE identification and analysis

To create a large set of UCE loci (*17*), we combined two sets of UCEs (*25*, *63*) and kept unique and nonduplicate loci in the set (n = 8047 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. After 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, then reassembled 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 consequently contains fewer orthologous loci in a complete matrix.

Using the complete data matrices, we aligned FASTA data corresponding to each reassembled UCE locus for each taxon. After 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% confidence interval (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 with 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 ML gene trees for every gene encoded in each of the three crocodilian genomes (phylomes) were reconstructed using the phylomeDB pipeline (*17*, *65*). In brief, sequence searches were used to retrieve homologs (E-value 10^{-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 crocodilian 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 crocodilian 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 supertree was inferred from all trees in the three phylomes by means of 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 ORs, vomeronasal receptors (V1R and V2R), taste receptors (T1R and T2R), and trace amine-associated receptors (TAAR) of the three crocodilians 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 crocodilian genomes using known vertebrate ORs as queries, and the best nonoverlapping 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 crocodilian ORs relative to other vertebrates using a Poisson model of substitution and evaluated support for the nodes with 1000 pseudoreplicates. We compared the evolution of ORs for the three crocodilians, chicken, and zebra finch (*74*) as well as green sea turtle and Asian softshell turtle (*16*).

Genome alignments and ancestral genome reconstruction

The WGA 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 accompanying avian phylogeny 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 ML, using the general reversible continuous-time nucleotide substitution model. To parameterize the model and estimate branch lengths, we used phyloFit (*82*) on conserved 4D sites in alligator genes (*17*). A complete technical exposition of the alignment computation and statistics calculated is available in (*17*).

Mutation rate estimation

We used a phylogenetic approach to estimate the overall mutation rate μ along the crocodilian lineage. From both the WGA 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. To calculate a per-generation mutation rate, we used 90 Ma as the TMRCA of alligator and crocodile and an average generation time of 20 years (table S23) (*17*).

Heterozygosity and population history estimation

For each genome, we used BWA (*83*) to map paired-end genome reads from a single individual back to the final genome assembly (*17*). We used tools in the GATK package (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, and crocodile H = 0.000360. Using these values as an estimate for θ and the substitution rate m calculated above, we estimated the effective population size for each species as shown in Fig. 6.

To estimate historical population sizes, we called single-nucleotide polymorphisms with SAMtools using reads with a map score of >30 and base calls with a quality score of >20. We applied the PSMC (*43*) model using 20 years for the generation time (table S23). We used 90 Ma 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^{-1} site^{-1}. We conducted bootstrap tests for each of the three taxa by splitting the scaffolds into smaller segments and randomly sampling the segments with replacement (fig. 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 200,000-year bins. Climate oscillations over the past 1 million years were calculated in 20,000-year bins.

REFERENCES AND NOTES

1. A. Janke, U. Arnason, The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent archosauria (birds and crocodiles). *Mol. Biol. Evol.* **14**, 1266–1272 (1997). doi: 10.1093/oxfordjournals.molbev.a025736; pmid: 9402737
2. A. G. Sennikov, The first ctenosauriscid (Reptilia: Archosauromorpha) from the Lower Triassic of eastern Europe. *Paleontol. J.* **46**, 499–511 (2012). doi: 10.1134/S0031030112050097
3. C. A. Brochu, Phylogenetic approaches toward crocodylian history. *Annu. Rev. Earth Planet. Sci.* **31**, 357–397 (2003). doi: 10.1146/annurev.earth.31.100901.141308
4. G. Grigg, F. Seebacher, C. E. Franklin, Eds., *Crocodilian Biology and Evolution* (Surrey Beatty, Chipping Norton, NSW, Australia, 2001).
5. M. M. Cohen, C. Gans, The chromosomes of the order Crocodilia. *Cytogenetics* **9**, 81–105 (1970). doi: 10.1159/000130080; pmid: 5461083
6. J. W. Lang, H. V. Andrews, Temperature-dependent sex determination in crocodilians. *J. Exp. Zool.* **270**, 28–44 (1994). doi: 10.1002/jez.1402700105
7. E. Olmo, Evolution of genome size and DNA base composition in reptiles. *Genetica* **57**, 39–50 (1981). doi: 10.1007/BF00057541
8. J. Harshman, C. J. Huddleston, J. P. Bollback, T. J. Parsons, M. J. Braun, True and false gharials: A nuclear gene phylogeny of crocodylia. *Syst. Biol.* **52**, 386–402 (2003). doi: 10.1080/10635150390197028; pmid: 12775527
9. J. R. Oaks, A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles. *Evolution* **65**, 3285–3297 (2011). doi: 10.1111/j.1558-5646.2011.01373.x; pmid: 22023592
10. A. Janke, A. Gullberg, S. Hughes, R. K. Aggarwal, U. Arnason, Mitogenomic analyses place the gharial (*Gavialis gangeticus*) on the crocodile tree and provide pre-K/T divergence times for most crocodilians. *J. Mol. Evol.* **61**, 620–626 (2005). doi: 10.1007/s00239-004-0336-9; pmid: 16211427
11. J. Alföldi *et al*., The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591 (2011). doi: 10.1038/nature10390; pmid: 21881562
12. T. A. Castoe *et al*., The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20645–20650 (2013). pmid: 24297902
13. H. B. Shaffer *et al*., The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14**, R28 (2013). doi: 10.1186/gb-2013-14-3-r28; pmid: 23537068
14. F. J. Vonk *et al*., The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20651–20656 (2013). doi: 10.1073/pnas.1314702110; pmid: 24297900
15. Q. H. Wan *et al*., Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. *Cell Res.* **23**, 1091–1105 (2013). doi: 10.1038/cr.2013.104; pmid: 23917531
16. Z. Wang *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.* **45**, 701–706 (2013). doi: 10.1038/ng.2615; pmid: 23624526
17. See supplementary materials on *Science* Online.
18. L. W. Hillier *et al*., Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004). doi: 10.1038/nature03154; pmid: 15592404
19. J. Huerta-Cepas, S. Capella-Gutiérrez, L. P. Pryszcz, M. Marcet-Houben, T. Gabaldón, PhylomeDB v4: Zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Res.* **42**, D897–D902 (2014). doi: 10.1093/nar/gkt1177; pmid: 24275491
20. E. S. Lander *et al*., Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001). doi: 10.1038/35057062; pmid: 11237011
21. W. C. Warren *et al*., The genome of a songbird. *Nature* **464**, 757–762 (2010). doi: 10.1038/nature08819; pmid: 20360741
22. W. C. Warren *et al*., Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175–183 (2008). doi: 10.1038/nature06936; pmid: 18464734
23. A. T. Chinwalla *et al*., Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002). doi: 10.1038/nature01262; pmid: 12466850
24. G. Bejerano *et al*., Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004). doi: 10.1126/science.1098119; pmid: 15131266
25. B. C. Faircloth *et al*., Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* **61**, 717–726 (2012). doi: 10.1093/sysbio/sys004; pmid: 22232343
26. J. E. McCormack *et al*., Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res.* **22**, 746–754 (2012). doi: 10.1101/gr.125864.111; pmid: 22207614
27. A. Siepel *et al*., Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005). doi: 10.1101/gr.3715005; pmid: 16024819
28. Y. Chiari, V. Cahais, N. Galtier, F. Delsuc, Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biol.* **10**, 65 (2012). doi: 10.1186/1741-7007-10-65; pmid: 22839781
29. N. G. Crawford *et al*., More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biol. Lett.* **8**, 783–786 (2012). doi: 10.1098/rsbl.2012.0331; pmid: 22593086
30. A. C. Tzika, R. Helaers, G. Schramm, M. C. Milinkovitch, Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent Sauropsida lineages and the phylogenetic position of turtles. *EvoDevo* **2**, 19 (2011). doi: 10.1186/2041-9139-2-19; pmid: 21943375
31. J. V. Chamary, J. L. Parmley, L. D. Hurst, Hearing silence: Non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* **7**, 98–108 (2006). doi: 10.1038/nrg1770; pmid: 16418745
32. A. Künstner, B. Nabholz, H. Ellegren, Significant selective constraint at 4-fold degenerate sites in the avian genome and its consequence for detection of positive selection. *Genome Biol. Evol.* **3**, 1381–1389 (2011). doi: 10.1093/gbe/evr112; pmid: 22042333
33. C. A. Brochu, Morphology, fossils, divergence timing, and the phylogenetic relationships of Gavialis. *Syst. Biol.* **46**, 479–522 (1997). doi: 10.1093/sysbio/46.3.479; pmid: 11975331
34. G. V. Glazko, M. Nei, Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* **20**, 424–434 (2003). doi: 10.1093/molbev/msg050; pmid: 12644563
35. R. Cordaux, J. Lee, L. Dinoso, M. A. Batzer, Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. *Gene* **373**, 138–144 (2006). doi: 10.1016/j.gene.2006.01.020; pmid: 16527433
36. B. Paten *et al*., Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.* **18**, 1829–1843 (2008). doi: 10.1101/gr.076521.108; pmid: 18849525
37. M. Blanchette, E. D. Green, W. Miller, D. Haussler, Reconstructing large regions of an ancestral mammalian genome in silico. *Genome Res.* **14**, 2412–2423 (2004). doi: 10.1101/gr.2800104; pmid: 15574820
38. Y. Niimura, Olfactory receptor multigene family in vertebrates: From the viewpoint of evolutionary genomics. *Curr. Genomics* **13**, 103–114 (2012). doi: 10.2174/138920212799860706; pmid: 23024602
39. J. Huerta-Cepas, H. Dopazo, J. Dopazo, T. Gabaldón, The human phylome. *Genome Biol.* **8**, R109 (2007). pmid: 17567924
40. Y. Niimura, M. Nei, Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLOS ONE* **2**, e708 (2007). pmid: 17684554
41. C. Stevenson, R. Whitaker, in *Crocodiles. Status Survey and Conservation Action Plan*, S. C. Manolis, C. Stevenson, Eds. (Crocodile Specialist Group, Darwin, Australia, 2010), pp. 139–143.
42. R. M. Elsey, A. R. Woodward, in *Crocodiles. Status Survey and Conservation Action Plan*, S. C. Manolis, C. Stevenson, Eds. (Crocodile Specialist Group, Darwin, Australia, 2010), pp. 1–4.
43. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011). doi: 10.1038/nature10231; pmid: 21753753
44. R. S. W. van de Wal, B. de Boer, L. J. Lourens, P. Kohler, R. Bintanja, Reconstruction of a continuous high-resolution CO_2 record over the past 20 million years. *Clim. Past* **7**, 1459–1469 (2011). doi: 10.5194/cp-7-1459-2011
45. G. Zhang *et al*., Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
46. C. I. Wu, W. H. Li, Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1741–1745 (1985). doi: 10.1073/pnas.82.6.1741; pmid: 3856856
47. Y. H. Lin, P. J. Waddell, D. Penny, Pika and vole mitochondrial genomes increase support for both rodent monophyly and glires. *Gene* **294**, 119–129 (2002). doi: 10.1016/S0378-1119(02)00695-9; pmid: 12234673
48. A. P. Martin, S. R. Palumbi, Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4087–4091 (1993). doi: 10.1073/pnas.90.9.4087; pmid: 8483925
49. J. F. Gillooly, A. P. Allen, G. B. West, J. H. Brown, The rate of DNA evolution: Effects of body size and temperature on the molecular clock. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 140–145 (2005). doi: 10.1073/pnas.0407735101; pmid: 15618408
50. L. Blueweiss *et al*., Relationships between body size and some life history parameters. *Oecologia* **37**, 257–272 (1978). doi: 10.1007/BF00344996
51. V. M. Savage, J. F. Gillooly, J. H. Brown, G. B. West, E. L. Charnov, Effects of body size and temperature on population growth. *Am. Nat.* **163**, 429–441 (2004). doi: 10.1086/381872; pmid: 15026978
52. M. S. Lee, A. Cau, D. Naish, G. J. Dyke, Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562–566 (2014). doi: 10.1126/science.1252243; pmid: 25082702
53. S. Gnerre *et al*., High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1513–1518 (2011). doi: 10.1073/pnas.1017351108; pmid: 21187386
54. A. M. Shedlock *et al*., Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2767–2772 (2007). doi: 10.1073/pnas.0606204104; pmid: 17307883
55. R. Li *et al*., De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010). doi: 10.1101/gr.097261.109; pmid: 20019144
56. M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008). doi: 10.1093/bioinformatics/btn013; pmid: 18218656
57. D. Kim *et al*., TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013). doi: 10.1186/gb-2013-14-4-r36; pmid: 23618408
58. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). doi: 10.1186/gb-2009-10-3-r25; pmid: 19261174
59. A. F. A. Smit, R. Hubley, RepeatModeler Open-1.0; www.repeatmasker.org.
60. R. C. Edgar, E. W. Myers, PILER: Identification and classification of genomic repeats. *Bioinformatics* **21** (suppl. 1), i152–i158 (2005). doi: 10.1093/bioinformatics/bti1003; pmid: 15961452
61. A. L. Price, N. C. Jones, P. A. Pevzner, De novo identification of repeat families in large genomes. *Bioinformatics* **21** (suppl. 1), i351–i358 (2005). doi: 10.1093/bioinformatics/bti1018; pmid: 15961478

62. D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008). doi: 10.1186/1471-2105-9-18; pmid: 18194517
63. S. Stephen, M. Pheasant, I. V. Makunin, J. S. Mattick, Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. *Mol. Biol. Evol.* **25**, 402–408 (2008). doi: 10.1093/molbev/msm268; pmid: 18056681
64. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006). doi: 10.1093/bioinformatics/btl446; pmid: 16928733
65. J. Huerta-Cepas, A. Bueno, J. Dopazo, T. Gabaldón, PhylomeDB: A database for genome-wide collections of gene phylogenies. *Nucleic Acids Res.* **36**, D491–D496 (2008). doi: 10.1093/nar/gkm899; pmid: 17962297
66. M. R. Wallace *et al*., A de novo Alu insertion results in neurofibromatosis type 1. *Nature* **353**, 864–866 (1991). doi: 10.1038/353864a0; pmid: 1719426
67. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009). doi: 10.1093/bioinformatics/btp348; pmid: 19505945
68. S. Guindon *et al*., New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010). doi: 10.1093/sysbio/syq010; pmid: 20525638
69. T. Gabaldón, Large-scale assignment of orthology: Back to phylogenetics? *Genome Biol.* **9**, 235 (2008). doi: 10.1186/gb-2008-9-10-235; pmid: 18983710
70. J. Huerta-Cepas, J. Dopazo, T. Gabaldón, ETE: A python environment for tree exploration. *BMC Bioinformatics* **11**, 24 (2010). doi: 10.1186/1471-2105-11-24; pmid: 20070885
71. J. Huerta-Cepas, T. Gabaldón, Assigning duplication events to relative temporal scales in genome-wide studies. *Bioinformatics* **27**, 38–45 (2011). doi: 10.1093/bioinformatics/btq609; pmid: 21075746
72. A. Wehe, M. S. Bansal, J. G. Burleigh, O. Eulenstein, DupTree: A program for large-scale phylogenetic analyses using gene tree parsimony. *Bioinformatics* **24**, 1540–1541 (2008). doi: 10.1093/bioinformatics/btn230; pmid: 18474508
73. K. Tamura *et al*., MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011). doi: 10.1093/molbev/msr121; pmid: 21546353
74. S. S. Steiger, V. Y. Kuryshev, M. C. Stensmyr, B. Kempenaers, J. C. Mueller, A comparison of reptilian and avian olfactory receptor gene repertoires: Species-specific expansion of group gamma genes in birds. *BMC Genomics* **10**, 446 (2009). doi: 10.1186/1471-2164-10-446; pmid: 19772566
75. B. Paten *et al*., Cactus: Algorithms for genome multiple sequence alignment. *Genome Res.* **21**, 1512–1528 (2011). doi: 10.1101/gr.123356.111; pmid: 21665927
76. E. D. Jarvis *et al*., Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
77. F. K. Barker, A. Cibois, P. Schikler, J. Feinstein, J. Cracraft, Phylogeny and diversification of the largest avian radiation. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11040–11045 (2004). doi: 10.1073/pnas.0401892101; pmid: 15263073
78. T. F. Wright *et al*., A multilocus molecular phylogeny of the parrots (Psittaciformes): Support for a Gondwanan origin during the Cretaceous. *Mol. Biol. Evol.* **25**, 2141 (2008). doi: 10.1093/molbev/msn160
79. R. C. Thomson, H. B. Shaffer, Sparse supermatrices for phylogenetic inference: Taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Syst. Biol.* **59**, 42–58 (2010). doi: 10.1093/sysbio/syp075; pmid: 20525619
80. N. Nguyen *et al*., Building a pangenome reference for a population. *Res. Comput. Mol. Biol.* **8394**, 207–221 (2014). doi: 10.1007/978-3-319-05269-4_17
81. G. Hickey, B. Paten, D. Earl, D. Zerbino, D. Haussler, HAL: A hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341–1342 (2013). doi: 10.1093/bioinformatics/btt128; pmid: 23505295
82. M. J. Hubisz, K. S. Pollard, A. Siepel, PHAST and RPHAST: Phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011). doi: 10.1093/bib/bbq072; pmid: 21278375
83. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi: 10.1093/bioinformatics/btp324; pmid: 19451168

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 (DOIs: 10.5524/100125, 10.5524/100126, 10.5524/100127, and 10.5524/100128) and at crocgenomes.org. This project was conducted by the International Crocodilian Genomes Working Group (ICGWG; www.crocgenomes.org). Supported by NSF grants MCB-1052500 and DEB-1020865 (D.A.R.), MCB-0841821 (D.A.R., D.G.P., F.M.M., C.J.S.), DUE-0920151 (E.L.B., E.W.T.), DBI-0905714 (M.K.F.), and DEB-1242260 (B.C.F.). D.A.R., F.M.M., and D.G.P. were also supported by the Institute for Genomics, Biocomputing and Biotechnology at Mississippi State University. S.R.I., L.G.M., J.G., and C.M. were supported by Australian Rural Industries Research and Development Corporation grants (RIRDC PRJ-000549, RIRDC PRJ- 005355, RIRDC PRJ-002461). R.E.G. is a Searle Scholar, Sloan Fellow, and consultant for Dovetail Genomics.. E.D.J. was supported by the Howard Hughes Medical Institute and the National Institutes of Health. E.L. 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 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.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1254449/suppl/DC1
Materials and Methods
Figs. S1 to S29
Tables S1 to S24
References (*84*–*206*)

7 April 2014; accepted 6 November 2014
10.1126/science.1254449

Science

Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs

Richard E. Green, Edward L. Braun, Joel Armstrong, Dent Earl, Ngan Nguyen, Glenn Hickey, Michael W. Vandewege, John A. St. John, Salvador Capella-Gutiérrez, Todd A. Castoe, Colin Kern, Matthew K. Fujita, Juan C. Opazo, Jerzy Jurka, Kenji K. Kojima, Juan Caballero, Robert M. Hubley, Arian F. Smit, Roy N. Platt, Christine A. Lavoie, Meganathan P. Ramakodi, John W. Finger Jr., Alexander Suh, Sally R. Isberg, Lee Miles, Amanda Y. Chong, Weerachai Jaratlerdsiri, Jaime Gongora, Christopher Moran, Andrés Iriarte, John McCormack, Shane C. Burgess, Scott V. Edwards, Eric Lyons, Christina Williams, Matthew Breen, Jason T. Howard, Cathy R. Gresham, Daniel G. Peterson, Jürgen Schmitz, David D. Pollock, David Haussler, Eric W. Triplett, Guojie Zhang, Naoki Irie, Erich D. Jarvis, Christopher A. Brochu, Carl J. Schmidt, Fiona M. McCarthy, Brant C. Faircloth, Federico G. Hoffmann, Travis C. Glenn, Toni Gabaldón, Benedict Paten and David A. Ray

Science **346** (6215), 1254449.
DOI: 10.1126/science.1254449