

# Multiple Lineages of Ancient CR1 Retroposons Shaped the Early Genome Evolution of Amniotes

Alexander Suh<sup>1,2,\*</sup>, Gennady Churakov<sup>1</sup>, Meganathan P. Ramakodi<sup>3,4,11,12</sup>, Roy N. Platt II<sup>3,4,5</sup>, Jerzy Jurka<sup>6,†</sup>, Kenji K. Kojima<sup>6</sup>, Juan Caballero<sup>7</sup>, Arian F. Smit<sup>7</sup>, Kent A. Vliet<sup>8</sup>, Federico G. Hoffmann<sup>3,4</sup>, Jürgen Brosius<sup>1</sup>, Richard E. Green<sup>9</sup>, Edward L. Braun<sup>10</sup>, David A. Ray<sup>3,4,5,‡</sup>, and Jürgen Schmitz<sup>1,‡</sup>

<sup>1</sup>Institute of Experimental Pathology (ZMBE), University of Münster, Germany

<sup>2</sup>Department of Evolutionary Biology (EBC), Uppsala University, Sweden

<sup>3</sup>Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University

<sup>4</sup>Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University

<sup>5</sup>Department of Biological Sciences, Texas Tech University

<sup>6</sup>Genetic Information Research Institute, Mountain View, California

<sup>7</sup>Institute for Systems Biology, Seattle, Washington

<sup>8</sup>Department of Biology, University of Florida

<sup>9</sup>Department of Biomolecular Engineering, University of California

<sup>10</sup>Department of Biology and Genetics Institute, University of Florida

<sup>11</sup>Present address: Cancer Prevention and Control Program, Fox Chase Cancer Center, Philadelphia, PA

<sup>12</sup>Present address: Department of Biology, Temple University, Philadelphia, PA

\*Corresponding author: E-mail: alexander.suh@ebc.uu.se.

†Deceased.

‡These authors contributed equally to this work.

Accepted: November 21, 2014

**Data deposition:** Raw survey sequence data have been deposited at Sequence Read Archive (SRA; accession number SRP048845). Sanger sequence data have been deposited at GenBank (accession numbers KP159624-KP159743).

## Abstract

Chicken repeat 1 (CR1) retroposons are long interspersed elements (LINEs) that are ubiquitous within amniote genomes and constitute the most abundant family of transposed elements in birds, crocodylians, turtles, and snakes. They are also present in mammalian genomes, where they reside as numerous relics of ancient retroposition events. Yet, despite their relevance for understanding amniote genome evolution, the diversity and evolution of CR1 elements has never been studied on an amniote-wide level. We reconstruct the temporal and quantitative activity of CR1 subfamilies via presence/absence analyses across crocodylian phylogeny and comparative analyses of 12 crocodylian genomes, revealing relative genomic stasis of retroposition during genome evolution of extant Crocodylia. Our large-scale phylogenetic analysis of amniote CR1 subfamilies suggests the presence of at least seven ancient CR1 lineages in the amniote ancestor; and amniote-wide analyses of CR1 successions and quantities reveal differential retention (presence of ancient relics or recent activity) of these CR1 lineages across amniote genome evolution. Interestingly, birds and lepidosaurs retained the fewest ancient CR1 lineages among amniotes and also exhibit smaller genome sizes. Our study is the first to analyze CR1 evolution in a genome-wide and amniote-wide context and the data strongly suggest that the ancestral amniote genome contained myriad CR1 elements from multiple ancient lineages, and remnants of these are still detectable in the relatively stable genomes of crocodylians and turtles. Early mammalian genome evolution was thus characterized by a drastic shift from CR1 prevalence to dominance and hyperactivity of L2 LINEs in monotremes and L1 LINEs in therians.

**Key words:** transposable elements, chicken repeat 1, phylogenomics, comparative genomics, crocodylian genomes, amniotes.

## Introduction

More than three decades ago, chicken repeat 1 (CR1) elements were the first transposable elements (TEs) to be identified in a genome of a nonmammalian land vertebrate (Stumph et al. 1981, 1984). CR1 elements are a family of long interspersed elements (LINEs) (Burch et al. 1993) that mobilize via an RNA intermediate and retropose via target-primed reverse transcription (Haas et al. 1997) similar to other LINEs (Luan et al. 1993). Full-length elements contain two open reading frames (ORFs) that encode the Gag-like ORF1p protein with a zinc finger-like motif (Haas et al. 1997, 2001; Kajikawa et al. 1997; Kapitonov and Jurka 2003), as well as the Pol-like ORF2p protein with endonuclease and reverse transcriptase (RT) domains (Burch et al. 1993; Haas et al. 1997, 2001; Kajikawa et al. 1997). The vast majority of CR1 insertions are heavily 5'-truncated elements (Vandergon and Reitman 1994; Hillier et al. 2004; Wicker et al. 2005) and thus functionally "dead on arrival" (Petrov and Hartl 1998), which hampers the reconstruction of full-length CR1 subfamilies and the confident determination of their exact 5'-UTR sequences (Kajikawa et al. 1997; Haas et al. 2001; Wicker et al. 2005). The extent of 5'-truncations appears to be more severe than in L1 LINEs and implies a lower processivity of CR1 reverse transcription (Hillier et al. 2004). On the other hand, the 3'-UTR of CR1 elements has been suggested to serve as a recognition site for the CR1-encoded RT (Kajikawa et al. 1997; Haas et al. 2001) and its hairpin structure and octamer microsatellite motif are highly conserved across various amniote CR1 subfamilies (Suh, Paus, et al. 2011).

CR1 retroposons are a ubiquitous genomic component that is present in all lineages of amniotes (Shedlock 2006; Suh, Paus et al. 2011), including mammals (Lovšin et al. 2001; Kapitonov and Jurka 2003; Shedlock 2006; Suh, Paus, et al. 2011) (but contra [Kordiš 2009]), and appear to constitute a Metazoa-specific, ancient clade of LINEs (Lovšin et al. 2001). In birds, many studies have shown that multiple lineages of CR1 elements were active in parallel, and some of them throughout long timespans of avian evolution (Hillier et al. 2004; Wicker et al. 2005; Kriegs et al. 2007; Abrusán et al. 2008; Liu et al. 2009; Suh, Paus, et al. 2011; Suh et al. 2012). On the other hand, the total CR1 element sequence from nonavian amniotes was only under 12 million basepair (Mb) of genomic data (Shedlock 2006; Shedlock et al. 2007), neglecting mammalian CR1 subfamilies. This paucity of comparative analyses is surprising, given that CR1 retroelements are "the major genome component" (Kordiš 2009) in birds (Hillier et al. 2004; Warren et al. 2010), crocodilians (Green et al. 2014), snakes (Castoe et al. 2013) and other lepidosaurs (Shedlock 2006), and turtles (Shaffer et al. 2013), and are thus crucial for understanding the genome evolution within Amniota.

Here, we present the first genome-wide study of CR1 retroelements across all major lineages of amniotes. First, we inferred the presence or absence of CR1 insertions across the crocodilian phylogeny. These data provided another important line of evidence that was able to resolve crocodilian phylogeny, a classic example of conflict between phylogenetic reconstructions using molecular versus morphological data (e.g., [Harshman et al. 2003; Janke et al. 2005]). Then we used whole-genome data to conduct a de novo characterization of crocodilian CR1 subfamilies in three genomes and complemented the whole-genome data by survey sequencing eight additional crocodilian species. Finally, phylogenetic reconstruction of the relationships among amniote CR1 lineages permitted us to infer the CR1 diversity in the ancestral amniote genome and reconstruct the subsequent events of CR1 expansion or inactivation that led to the pronounced differences among the repetitive landscapes of extant amniote genomes.

## Materials and Methods

### TE Subfamily Prediction

As part of the collaborative efforts to annotate crocodilian genomes (St John et al. 2012; Green et al. 2014), consensus sequences of CR1 and other TE subfamilies from American alligator (*Alligator mississippiensis*), saltwater crocodile (*Crocodylus porosus*), and gharial (*Gavialis gangeticus*) were generated in the laboratories of DAR, JJ, and AFS. All subfamilies were predicted de novo using complementing methods implemented in RepeatModeler (<http://www.repeatmasker.org/RepeatModeler.html> last accessed January 13, 2015), followed by procedures that are described in detail elsewhere (Dasmahapatra et al. 2012; Green et al. 2014). Briefly, RepeatModeler was initially used to analyze the *A. mississippiensis* genome draft, after which manual work was necessary to confirm or extend the consensus sequences for each repeat by first querying the entire *A. mississippiensis* draft using BLAST (version 2.2.23 [Altschul et al. 1990]). Up to 50 of the top hits for each putative consensus were extracted along with up to 1,000 bp of flanking sequence. The extracted sequences were aligned with their respective RepeatModeler-generated partner using MUSCLE (version 4.0 [Edgar 2004]) and a majority-rule consensus sequence was created. To be considered "complete", a consensus sequence must exhibit highly variable flanking sequences at the 5'- and 3'-termini of the putative consensus, indicating insertion of an element at multiple distinct loci. If this condition was not met, the process was repeated by extending the flanking sequences. RepeatModeler analysis of *Cr. porosus* was followed by comparison to the resulting *A. mississippiensis* library to identify elements predicted from both genomes. Unique putative repeats from *C. porosus* were used to query the *C. porosus* assembly and the BLAST/extract/align process was repeated. Finally, the process was repeated once more for *G. gangeticus*

elements. After combining all three crocodylian repeat libraries, UCLUST (Edgar 2010) was used to group consensus sequences of potential subfamilies at 95% sequence identity thresholds. Consensus sequences with more than 95% identity were merged into a single subfamily consensus, whereas the remaining were defined as consensus sequences of separate subfamilies and named according to the UCLUST groupings.

### Presence/Absence Analyses

We extracted a total of approximately 12 million retroposon loci from the genome assemblies of saltwater crocodile, gharial, and American alligator, and selected approximately 30,000 cases with TE-free flanking sequences of 750 bp. These loci were BLASTn screened against a set of 26,637 short introns (<1.5 kb) from anole lizard and 28,713 short introns from chicken, yielding 122 TE candidate loci in alligator, 112 loci in crocodile, and 56 loci in gharial. Three-way alignments of the three crocodylians were compiled for all these loci in order not to bias the outcome of our experimental screening toward one of the two competing hypothesis regarding the position of the gharial. After we found no conflict among those three-way presence/absence patterns, we generated oligonucleotide primers (supplementary table S2, Supplementary Material online) for a total of 73 retroposon loci for subsequent experimental presence/absence screening.

We experimentally screened our marker candidate loci using standard procedures previously used in avian retroposed element (RE) presence/absence screenings (Suh, Paus, et al. 2011) across a taxon sampling comprising all crocodylian genera and the key species within the *Crocodylus* radiation sensu Oaks (2011). Briefly, we amplified all samples via touch-down polymerase chain reaction (PCR), followed by PCR product purification and direct sequencing (Suh, Paus, et al. 2011). All sequences were aligned per locus using MAFFT (Katoh and Toh 2008) (E-INS-I, version 6, <http://mafft.cbrc.jp/alignment/server/index.html> last accessed January 13, 2015), manually realigned, and presence/absence states carefully scored following the strict criteria of Suh, Paus, et al. (2011). That is, orthology of a phylogenetically informative retroposon insertion requires identity of RE target site, RE orientation, RE subtype, and (if present) target site duplication, as well as a clear absence (empty RE insertion site) in other species. In total, this was the case for 36 RE insertion loci (supplementary table S1, Supplementary Material online) that are all available as supplementary material (supplementary data S1, Supplementary Material online). As part of analyzing these loci, we also noted five non-TE indels that were phylogenetically informative and supplemented our phylogeny.

### Transposition in Transposition Analyses

We estimated chronologies of CR1 activity probabilities using the transposition in transposition (TinT) method (Kriegs et al.

2007; Churakov et al. 2010) (<http://www.compgen.uni-muenster.de/tools/tint/> last accessed January 13, 2015; default parameters for CR1 elements) on 6,752 nested CR1 in Chinese alligator, 8,816 nested CR1 in American alligator, 9,100 nested CR1 in saltwater crocodile, and 8,628 nested CR1 in gharial. The resultant graphs contain successions of probable retroposon activity periods on a relative timescale, where ovals represent 75%, vertical lines 95%, and horizontal lines 99% of the normal distribution.

### Survey Sequencing

Our survey sequencing sampling comprises *Alligator sinensis*, *Caiman latirostris*, *Caiman yacare*, *Crocodylus acutus*, *Crocodylus niloticus*, *Mecistops cataphractus*, *Melanosuchus niger*, *Osteolaemus tetraspis*, and *Tomistoma schlegelii*. We isolated DNA from blood and generated standard TruSeq Illumina libraries with insert sizes of 263 bp and bar codes for each taxon. All nine libraries were sequenced as 100-bp reads on a single lane of an Illumina GAIIx genetic analyzer and yielded an average of approximately 0.2× coverage per genome. We then applied the strategy from Diez et al. (2014) and conducted BLASTn searches of the resultant unassembled reads against a library of all crocodylian CR1 subfamilies. After selecting hits longer than 29 bp, cumulative quantities of CR1-derived reads were calculated for each CR1 subfamily and survey sequencing library. We then compared the CR1 representation in the Chinese alligator survey sequences with the recently published genome assembly (Wan et al. 2013) and, for each CR1 subfamily, derived coefficients for conversion of the remaining eight survey sequencing libraries into genome-wide estimates of CR1 quantities. This was possible because CR1 TinT patterns (fig. 2A) and CR1 landscapes (fig. 4) from the four genome assemblies (together spanning the breadth of crocodylian diversity) suggest similar age distributions (and quantities) for most CR1 subfamilies, which implies that old, diverged CR1 elements are probably equally abundant in the survey sequenced species due to subfamily activity in the last common ancestor of Crocodylia. Consequently, the detection of not only young (i.e., elements with near-identical sequence and thus high detectability), but also these older CR1 fragments is probably equivalent across the survey sequenced species. The resultant CR1 quantities (fig. 2B) support this assumption, as many CR1 subfamilies have nearly equal amounts of masked bases in all 12 sampled crocodylians.

### Phylogenetic Analyses

We compiled a sampling of 119 CR1 subfamilies from various amniote genomes in RepBase (Jurka et al. 2005) (<http://www.girinst.org/repbase/index.html> last accessed January 13, 2015), including the crocodylian CR1 consensus sequences generated for this study. As many of these subfamily predictions feature 5'-truncations, we focused on analyzing the 3'-part of the ORF2 RT domain together with the adjacent

3'-UTR. Note that we only used *Xenopus* CR1 subfamilies as our outgroup, because nontetrapod CR1 elements (e.g., from zebra fish) could not be aligned unambiguously at the nucleotide level. We also did not sample L2 LINEs, as they are closely related to nontetrapod CR1-like elements (Kapitonov and Jurka 2003) and likewise problematic for unambiguous alignment to amniote CR1 elements. All sampled sequences were first aligned using MAFFT (E-INS-I, version 6), manually realigned, and then ambiguously aligned sites at the boundary of ORF2 and 3'-UTR were excluded from the alignment. We then conducted maximum likelihood-based sequence analysis using RAxML (8.0.0 [Stamatakis et al. 2008], GTRCAT model, 1,000 bootstrap inferences) on the CIPRES Science Gateway (Miller et al. 2010) (<https://www.phylo.org/portal2/login!input.action> last accessed January 13, 2015). The full alignment is available as [supplementary material \(supplementary data S2, Supplementary Material online\)](#).

### Landscape Analyses

We created per-taxon custom CR1 libraries and used these for masking genomes of amniote representatives in a local installation of RepeatMasker (Smit et al. 1996–2014). Distances from the consensus sequence were calculated using the Kimura 2-parameter model (Kimura 1980) in the calcDivergenceFromAlign.pl script that is part of the RepeatMasker program package (Smit et al. 1996–2014). This Perl script removes hypermutable CpG sites during the calculation of Kimura 2-parameter distances between sequence pairs and converts the RepeatMasker “.align” output file (containing per-sequence pair transition–transversion ratios) into a table file (Pagán et al. 2010). We then generated CR1 landscape plots using the total base pairs of CR1-annotated sequence per CR1 group in bins of size 1% in the range of 0–50% divergence.

## Results and Discussion

### CR1 Markers Resolve the Early Branches of the Crocodylian Tree of Life

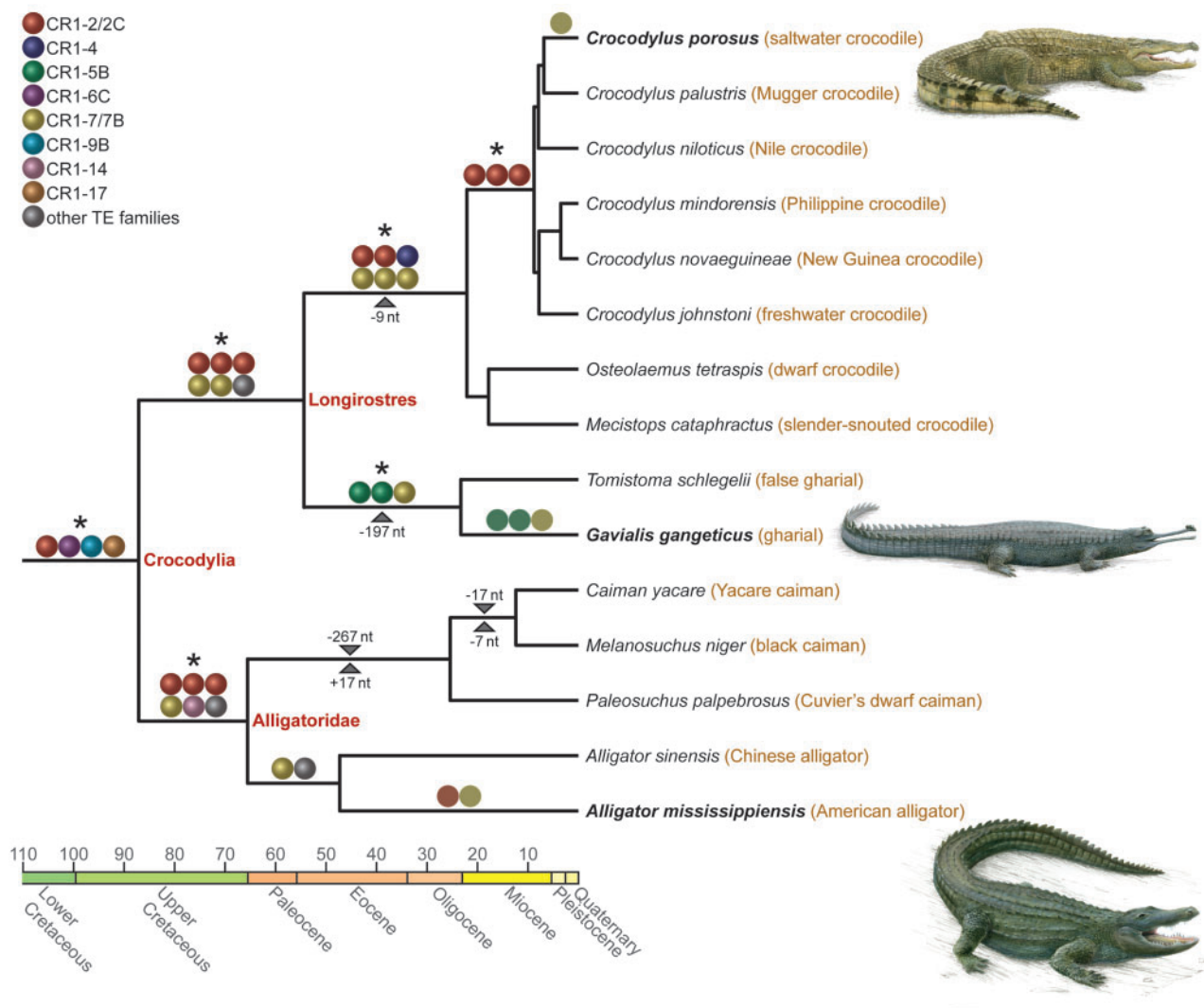
To reconstruct the temporal impact of CR1 LINEs on crocodylian genomes, we first studied CR1 presence/absence patterns among different crocodylian species. In addition to providing direct evidence for the timing of a TE insertion event, presence/absence patterns of CR1 and other REs constitute powerful, nearly homoplasy-free phylogenetic markers (Shedlock et al. 2004; Ray et al. 2006; Han et al. 2011). As they have been successfully used to resolve long-standing phylogenetic controversies among avian relationships (Kaiser et al. 2007; Kriegs et al. 2007; Suh, Kriegs, et al. 2011; Suh, Paus, et al. 2011; Haddrath and Baker 2012; Liu et al. 2012; Suh et al. 2012; Baker et al. 2014; Jarvis et al. 2014), they promised to be equally valuable markers for reconstructing the phylogeny of crocodylians, the extant sister taxon of birds. Using the

genome assemblies of saltwater crocodile (*C. porosus*), gharials (*G. gangeticus*), and American alligator (*A. mississippiensis*) as starting points for retroposon marker search (see Materials and Methods), we experimentally tested 73 loci via high-throughput PCR on a set of taxa comprising all extant crocodylian genera and the most distantly related species within *Crocodylus* (Oaks 2011). We sequenced all PCR amplicons of the sampled taxa and obtained 32 RE presence/absence patterns that span crocodylian phylogeny (fig. 1 and [supplementary table S1, Supplementary Material online](#)) and constitute the hitherto first RE presence/absence analysis in this amniote taxon. Most of these insertions (29 of 32; [supplementary table S1, Supplementary Material online](#)) correspond to CR1 insertions, although there were also two insertions of Penelope LINEs and one of an endogenous retrovirus. We also identified four CR1 markers that constitute unequivocal insertion events in the common ancestor of Crocodylia, as we could determine that the insertion was absent (i.e., empty insertion site) in the avian outgroup. Notably, all RE markers are fully congruent with each other and also conflict-free when compared with the sequence-based multilocus phylogeny of Oaks (2011). This suggests that, in sharp contrast to the situation for the early divergences of neoavian birds (Suh, Paus, et al. 2011; Jarvis et al. 2014) and placental mammals (Churakov et al. 2009; Nishihara et al. 2009), incomplete lineage sorting has not confounded the inference of crocodylian phylogeny.

We obtained three or more markers for most of the early branches in the crocodylian tree, thus constituting statistically significant support for each branch, respectively, according to the Waddell et al. (2001) likelihood ratio test for retroposon data. This includes six retroposon markers that provide a third, independent perspective on the long-standing “molecules versus morphology conflict” (Harshman et al. 2003) regarding the phylogenetic position of the gharial. In this conflict, virtually all morphological analyses place the gharial as the basal taxon to the remaining extant crocodylians (e.g., Gatesy et al. 2003; Delfino et al. 2008; Sereno and Larsson 2009; Holliday and Gardner 2012; Scheyer et al. 2013). In contrast, molecular sequence-based analyses suggest Alligatoridae (alligators and caimans) as sister to all other crocodylians and support a gharial + false gharial clade (e.g., Gatesy et al. 2003; Harshman et al. 2003; Janke et al. 2005; Meganathan et al. 2011; Oaks 2011). Our conflict-free retroposon markers add to the resolution of this controversy by unequivocally substantiating the latter hypothesis that includes the grouping of true crocodiles, gharial, and false gharial as “Longirostres” (Harshman et al. 2003).

### Low Diversity of Active CR1 Elements during Lineage-Specific Crocodylian Evolution

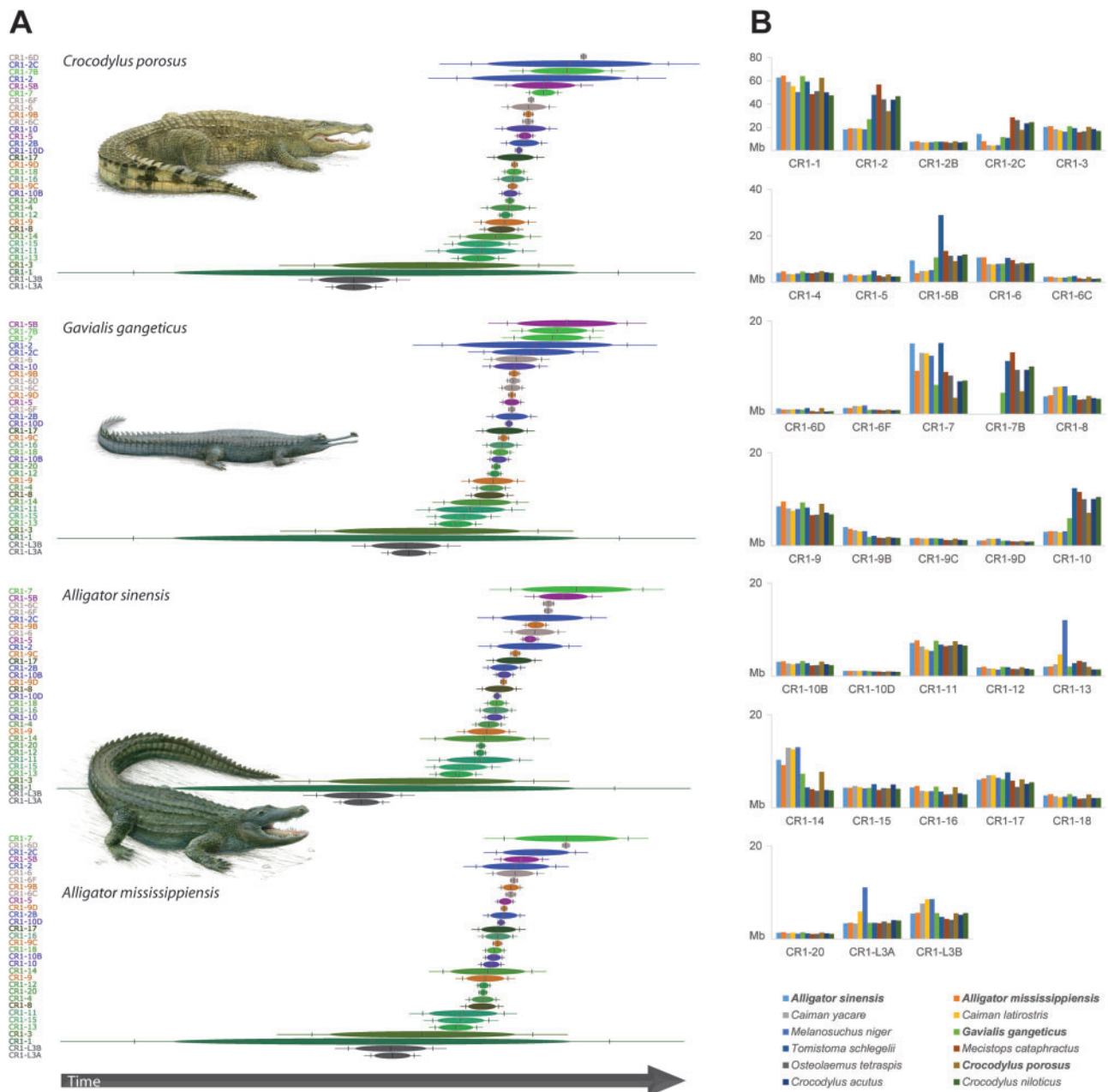
We assigned our CR1 markers to specific CR1 subfamilies sensu Green et al. (Green et al. 2014) and most belong to



**FIG. 1.**—CR1 retroposons resolve the early branches in the crocodylian tree of life. Retroposon markers (colored balls) and non-TE indel markers (triangles with number of inserted/deleted nucleotides) are mapped on Oaks's (Oaks 2011) dated tree of crocodylians (colors on the time axis correspond to the respective geological epoch in the International Stratigraphic Chart; <http://www.stratigraphy.org/ICChart/StratChart2010.pdf> last accessed January 13, 2015). Presence/absence states (supplementary table S1, Supplementary Material online) are based on per-locus alignments (supplementary data S1, Supplementary Material online) and species-specific autapomorphic retroposon insertions are depicted as colored circles. Statistically significant retroposon support sensu Waddell et al. (2001) is highlighted with asterisks. Since we did not observe any conflict, this corresponds to those branches with at least three insertions. Species names in bold letters are the three genome assemblies that were used as starting points for CR1 presence/absence screenings; names in red letters refer to higher-ranking taxa. Note that despite the fact that crocodylians diverged from their avian outgroup more than 219 Ma (Shedlock and Edwards 2009), making it difficult to align neutrally evolving DNA, we identified four CR1 markers for crocodylian monophyly. These could be aligned with chicken and zebra finch sequences, where they exhibit an orthologous insertion site without the CR1 insertion.

subfamilies CR1-2/2C, CR1-5B, and CR1-7/7B (fig. 1). This provides direct evidence for the activity of these CR1 subfamilies through large parts of crocodylian evolution, with CR1-7/7B activity having the widest temporal extent. This is because, in figure 1, CR1-7/7B activity can be identified in almost all branches since the last common ancestor of Crocodylia (>87 Ma [Oaks 2011]), including species-specific activity as recent as less than 9 Ma (Oaks 2011) in the saltwater crocodile. We independently verified these observations

with an estimation of CR1 activity in the TinT model that considers relative frequencies of insertions of different RE subtypes nested within each other (Kriegs et al. 2007; Churakov et al. 2010). The resulting TinT chronology of CR1 succession (fig. 2A and supplementary fig. S1, Supplementary Material online) corroborates that the aforementioned CR1 subfamilies exhibit a long period of activity, including recent activity in the genomes of saltwater crocodile, gharial, Chinese alligator, and American alligator. Furthermore, the TinT



**Fig. 2.**—Activity of CR1 elements in 12 crocodylian genomes. (A) Estimates of CR1 activity in genome assemblies of saltwater crocodile, gharial, Chinese alligator, and American alligator using the TinT model (Kriegs et al. 2007; Churakov et al. 2010). Congruent successions of predicted activities of merged CR1 subfamilies (see supplementary figure S1, Supplementary Material online, for unmerged TinT patterns) between the four genomes suggest that the core activity periods of most subfamilies lie in the common ancestor of crocodylians. Normal distributions of activity period estimations are shown as ovals (Churakov et al. 2010). (B) Comparison of CR1-derived base pair quantities in four assembled genomes (bold names) and eight unassembled survey sequence libraries. All CR1 subfamilies with the exception of CR1-7B appear to be present across the breadth of the sampled crocodylian species diversity. Nevertheless, quantitative differences within several subfamilies suggest differential activities in crocodylian genomes; for example, an extended activity of CR1-2 in Longirostres (all crocodylians except alligators and caimans (Harshman et al. 2003)) and of CR1-5B in *T. schlegelii*. Note that some of the lineage-specific differences in CR1 quantities might either be accounted to potential differences in rate of neutral sequence evolution or cryptic CR1 subfamilies that were not identified in our CR1 predictions (as they are based on the genome assemblies of saltwater crocodile, gharial, and American alligator). English names of the 12 genomes are (in the order of their appearance in the legend) Chinese alligator, American alligator, Yacare caiman, broad-snouted caiman, black caiman, gharial, false gharial, slender-snouted crocodile, dwarf crocodile, saltwater crocodile, American crocodile, and Nile crocodile.

patterns are largely congruent among these four genomes, suggesting activity of most CR1 subfamilies in the common ancestor of crocodylians, while only few subfamilies were active since the divergence of Alligatoridae and Longirostres. The same observation is evident when comparing the average intrasubfamily sequence divergence among CR1 copies ([supplementary fig. S2, Supplementary Material](#) online) among saltwater crocodile, gharial, Chinese alligator, and American alligator.

To test whether this trend of low lineage-specific CR1 diversity is a common feature of extant crocodylians, we conducted survey sequencing in Chinese alligator (*A. sinensis*) and eight additional species that span the breadth of crocodylian diversity, namely three caimans (*Ca. yacare*, *Ca. latirostris*, *Mel. niger*), false gharial (*T. schlegelii*), and four crocodiles (*C. acutus*, *C. niloticus*, *M. cataphractus*, *Osteolaemus tetraspis*). Given that TE identification in short sequences (e.g., survey data, see Materials and Methods) likely has a reduced reliability than in long sequences, we estimated the whole-genome CR1 content in these species ([fig. 2B](#)) by normalizing our unassembled survey sequencing libraries with a set of coefficients that was derived from comparing the representation of each CR1 subfamily in our own Chinese alligator survey sequences with the corresponding CR1 subfamily content of a recently published conspecific genome assembly (Wan et al. 2013). We emphasize that these are rough estimates compared with the CR1 quantities measured in the three assembled genomes. Nevertheless, direct comparison of the number of bases assigned to specific CR1 subfamilies in the total of 12 crocodylian genomes ([fig. 2B](#)) yields many subfamilies with similar amounts of annotated bases among all sampled species. This suggests that the survey sequences are comparable estimates of CR1 quantities and again reveals that most CR1 subfamilies were likely active before the divergence of extant lineages of Crocodylia. On the other hand, subfamilies such as CR1-2, CR1-7B, and CR1-10 appear to exhibit an increased or extended activity common to Longirostres (Crocodylidae + Gavialidae [Harshman et al. 2003]), and we even find evidence for lineage-specific expansion of a CR1 subfamily in the false gharial (CR1-5B), as well as in the black caiman (e.g., CR1-13). These lineage-specific CR1 activities in unassembled genomes might potentially constitute novel subfamilies, given that our *de novo* predictions of CR1 subfamilies are based on the assembled genomes of saltwater crocodile, gharial, and American alligator.

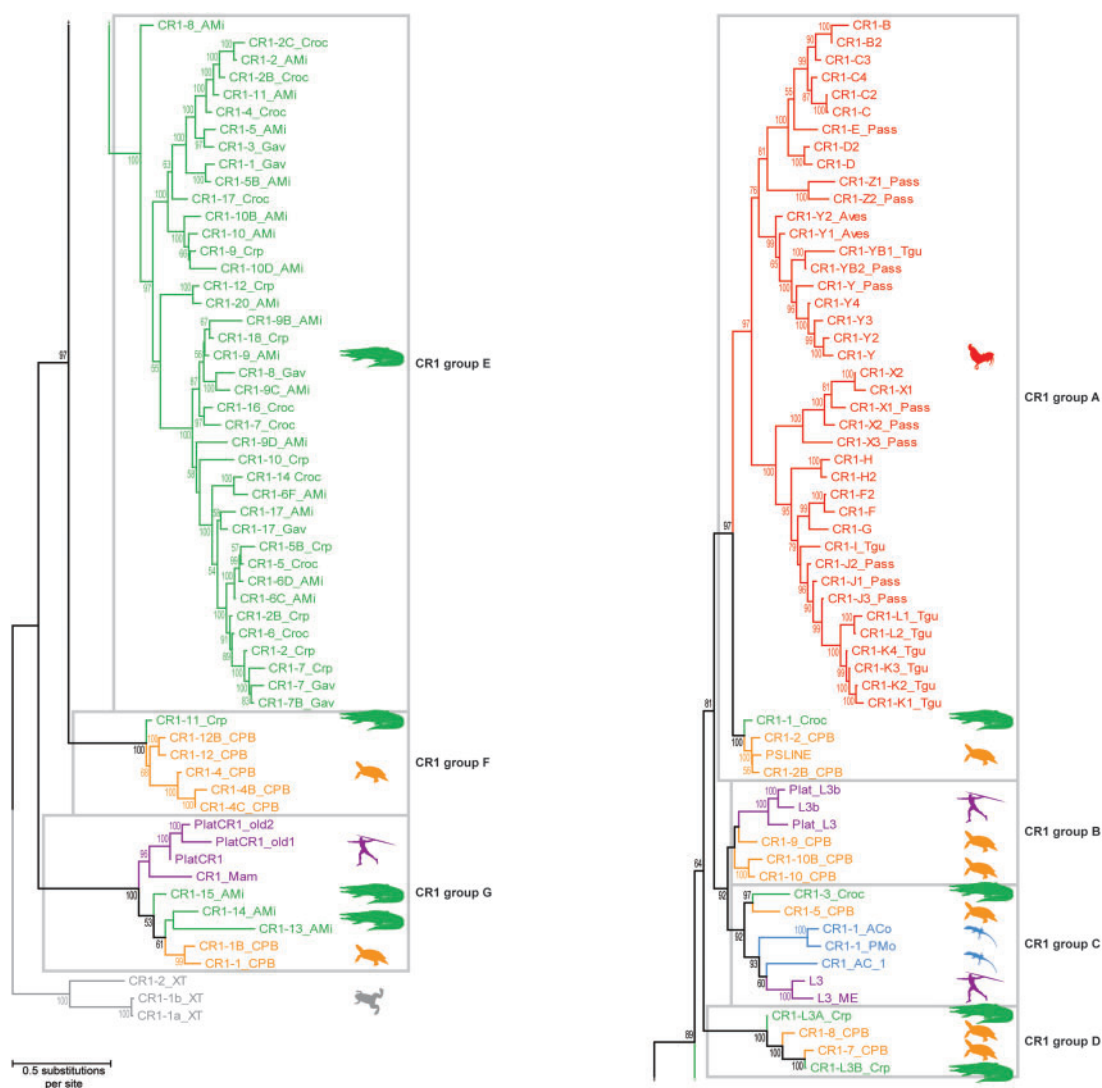
In addition to our findings, reduced TE diversity within Crocodylia since the divergence of Longirostres from Alligatoridae (~87 Ma [Oaks 2011]) is also suggested by other collaborative efforts within the International Crocodylian Genomes Working Group. Green et al. (2014) estimated that only approximately 5% of all TE copies were deposited in crocodylian genomes within that timeframe, suggesting an overall decline of both the rate as well as the diversity of TE activity. This applies not only to CR1 activity as

described in this study, but also to the activity of DNA transposons that has declined to an even more extreme degree (Green et al. 2014). Instead, retrovirus-like elements constitute about two-thirds of the younger TE-derived DNA in crocodylian genomes (Green et al. 2014) and Chong et al. (2014) suggest that this is the result of multiple infection events of various unrelated retroviral lineages.

### Phylogeny of CR1 Elements Suggests Multiple Ancient Lineages within Amniotes

We reconstructed the phylogenetic relationships of crocodylian CR1 subfamilies based on maximum likelihood analyses of part of the ORF2 RT domain + 3'-UTR of all amniote CR1 subfamily consensus sequences available in RepBase (Jurka et al. 2005). These consensus sequences represent an approximation of the (most often long extinct) master genes that gave rise to the paralogous TE copies visible in genomes. The resultant CR1 tree ([fig. 3](#)) was rooted to an amphibian outgroup and exhibits a topology with crocodylian, mammalian, and turtle CR1 subfamilies grouping not according to their hosts, but with many highly diverged species groups which are dispersed among multiple CR1 lineages. Considering this topology together with the phylogenetic relationships among amniotes (Shedlock and Edwards 2009; Shaffer et al. 2013; Green et al. 2014) and assuming vertical transmission as the usual mode of RE inheritance among hosts, the most parsimonious explanation for this is that at least seven CR1 lineages were present in the common ancestor of amniotes (CR1 groups A–G), although it is possible that this cautious estimate could expand with further sampling of host genomes and CR1 subfamilies. It is likely that the complex branching pattern of crocodylian and turtle CR1 subfamilies within CR1 groups D and G is the result of ancient activity in their common ancestor, a hypothesis consistent with their high intrasubfamily divergence levels, the highest among CR1 elements in crocodylian genomes ([supplementary fig. S2, Supplementary Material](#) online). On the other hand, the topology within CR1 group C suggests that L3 elements of placental and marsupial mammals (therians) are most closely related to lepidosaurian CR1 subfamilies or even nested within these, which might be the result of multiple independent extinctions of multiple ancient CR1 lineages in most amniote genomes.

Our study is the first genome-scale exploration of CR1 diversity across the breadth of amniote phylogeny. The phylogenetic relationships among amniote CR1 lineages have been previously studied only by Shedlock (2006) and Shedlock et al. (2007) who analyzed 1.5–3.7 Mb of genomic sequences per bacterial artificial chromosome (BAC) library from three birds, one crocodylian, and two lepidosaurs. Those studies suggested that the genomes of nonmammalian amniotes exhibit several ancient CR1 lineages. Our analyses suggest that the same is also the case for mammalian genomes, while our CR1 phylogeny ([fig. 3](#)) contains no evidence for nonmonophyly of



**FIG. 3.**—Phylogenetic relationships among amniote CR1 elements. Maximum likelihood nucleotide sequence analysis (RAxML 8.0.0 (Stamatakis et al. 2008), GTRCAT model, 1,000 bootstrap inferences) of part of the ORF2 RT domain + 3'-UTR of all CR1 subfamily consensus sequences available in RepBase (Jurka et al. 2005) suggests a complex tree topology of avian (red), crocodilian (green), turtle (orange), lepidosaurian (blue) and mammalian (purple) CR1 elements. Given this topology and the parsimonious assumption that CR1 retroposons are inherited vertically, we infer a minimum number of seven ancient CR1 lineages (groups A–G; see also [supplementary table S3, Supplementary Material](#) online, for a group classification of all analyzed CR1 subfamilies) that were present in the common ancestor of amniotes. The tree was rooted using amphibian CR1 subfamilies from *Xenopus* as outgroup, as they were the only nonamniote CR1 elements alignable to ingroup CR1s on the nucleotide level. Note that, in contrast to the merged CR1 subfamily definitions of figures 1 and 2, we included all CR1 consensus sequences generated in the individual TE annotations of the saltwater crocodile, gharial, and American alligator genomes. Given the short internodes and high sequence similarity among the unmerged crocodilian CR1 subfamilies of group E, we refrain from renaming these CR1 subfamilies according to the topology of the present tree, until additional crocodilian genome assemblies are available. The present nomenclature of crocodilian CR1 subfamilies is therefore solely based on the UCLUST grouping of their consensus sequences (see Materials and Methods). Unlabeled nodes received a bootstrap support of less than 50%.

detectable CR1 subfamilies in birds, which is contrary to BAC-scale analyses that sampled about two dozens of CR1 copies per amniote genome (Shedlock 2006; Shedlock et al. 2007). This is striking because the avian CR1 consensus sequences sampled herein were predicted in the two independent, high-quality TE annotations of chicken (Hillier et al. 2004) and zebra

finch (Warren et al. 2010), yet all form a monophyletic clade within the CR1 group A. To explain this discrepancy, we assumed that CR1 elements unrelated to group A have extremely low copy numbers in birds and were thus not detected as distinct TE subfamilies in the aforementioned TE annotations. To test this, we conducted BLASTn (Altschul et al.



1990) searches of representative crocodylian or turtle CR1 consensus sequences from groups B to G against the chicken and zebra finch genomes and analyzed the phylogenetic affinities of all sequence hits in the framework of the CR1 tree of figure 3. The resultant tree ([supplementary fig. S3, Supplementary Material online](#)) groups most of these hits within CR1 group A, yet three hits from chicken and three hits from zebra finch cluster within CR1 group F. We therefore propose that, while the sampled bird genomes contain only CR1 group A elements in large copy numbers, there are a handful of CR1 group F elements ([supplementary fig. S3, Supplementary Material online](#)) still detectable as remnants of ancient group F activity in the last common ancestor of Archosauria. The high intrasubfamily divergence of CR1 group F copies in crocodylian genomes ([supplementary fig. S2, Supplementary Material online](#)) suggests that these elements might have been inactive in Crocodylia for an equally long time, but many copies are still discernible due to the much slower rate of molecular evolution in crocodylians compared with birds (Shaffer et al. 2013; Green et al. 2014).

#### Ancient CR1 Lineages Were Differentially Retained within Amniotes

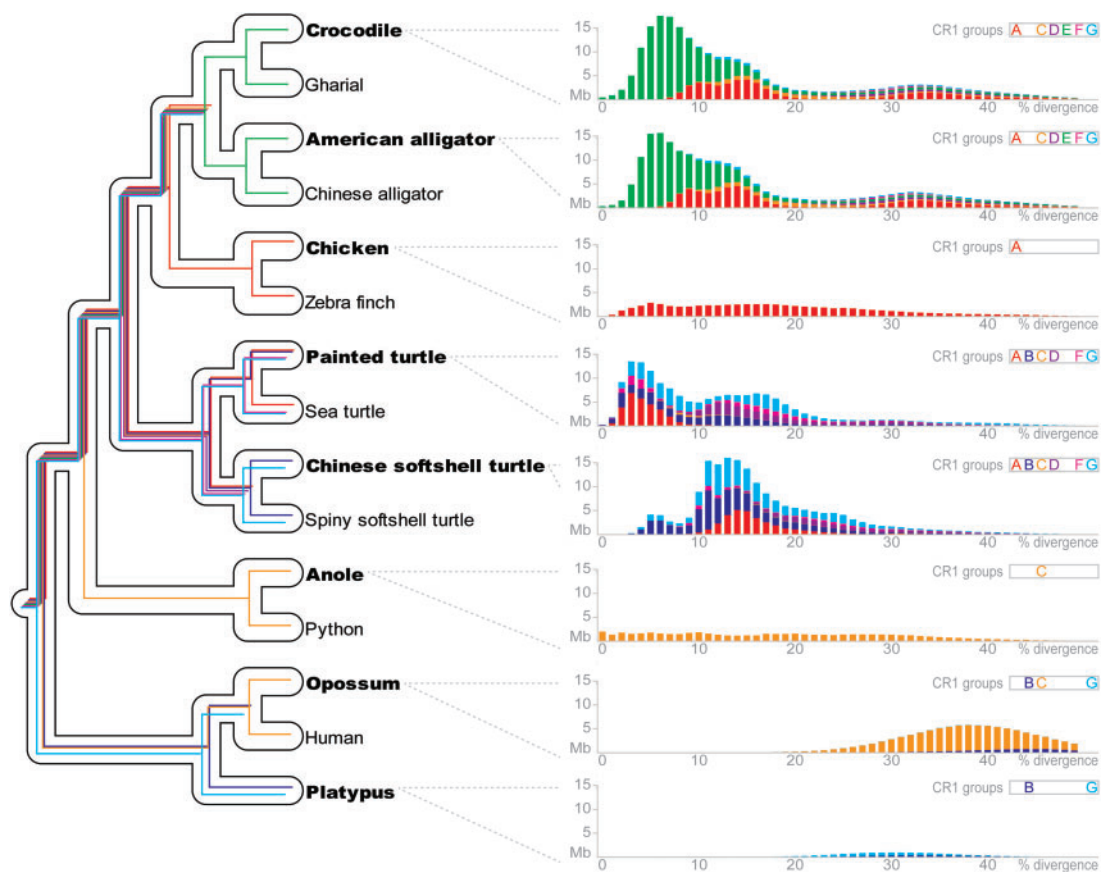
Our classification of amniote CR1 elements into groups A–G according to their aforementioned phylogenetic relationships permitted us to study the impact of these ancient CR1 lineages during amniote evolution. This was accomplished by analyzing their temporal successions and quantitative distributions in the genomes of four crocodylians (Wan et al. 2013; Green et al. 2014), two birds (Hillier et al. 2004; Warren et al. 2010), four turtles (Shaffer et al. 2013; Wang et al. 2013), three lepidosaurs (Alföldi et al. 2011; Castoe et al. 2013; Vonk et al. 2013), and three mammals (Lander et al. 2001; Mikkelsen et al. 2007; Warren et al. 2008). The resultant CR1 landscape plots (fig. 4, right panel; [supplementary fig. S4, Supplementary Material online](#)) illustrate the cumulative quantities of CR1 bases plotted against the level of divergence to their respective consensus sequences, which roughly corresponds to a relative time axis, and suggests differential retention of CR1 lineages throughout early amniote evolution (fig. 4, left panel). The four crocodylian genomes have retained all ancient CR1 lineages except group B and the crocodylian CR1 landscapes are almost identical, which suggests that only CR1 group E was active after the onset of the diversification of extant Crocodylia. This is in line with the large amount of closely related CR1 subfamilies within group E, and the fact that all of our CR1 markers for crocodylian phylogeny belong to group E (fig. 1 and [supplementary table S1, Supplementary Material online](#)). On the other hand, the avian sister group of crocodylians retained only CR1 group A activity (but see [supplementary fig. S3, Supplementary Material online](#), for low-copy relics of CR1 group F). Similar to the situation in crocodylians, turtle genomes appear to exhibit all CR1 lineages

but group E, yet their genomes show a diversity of several ancient CR1 lineages that were active during the evolution of extant turtles. Specifically, CR1 groups A, B, and G were active in the painted turtle and sea turtle lineages, whereas groups B and G were recently active in softshell turtles. Notably, the CR1 landscapes of lepidosaurs contain only CR1 group C elements, and thus bear resemblance to therian mammal genomes that exhibit a predominance of group C elements and low levels of group B and G activity. In contrast to this, CR1 retention in monotreme mammals only comprises detectable elements from CR1 groups B and G.

Our CR1 landscape analyses illustrate contrasting fates of CR1 activity among amniote lineages (fig. 4 and [supplementary fig. S4, Supplementary Material online](#)). Although CR1 elements have probably been long extinct since the early evolution of mammals, several of the nonmammalian genomes studied herein suggest ongoing, very recent CR1 activity in some amniote lineages. Evidence for this has so far been limited to the anole lizard (Novick et al. 2009; Alföldi et al. 2011) and one bird lineage, grebes (Suh et al. 2012). In-depth studies of chicken CR1 retroposons suggest that the lineage leading to the chicken recently lost CR1 activity (Hillier et al. 2004; Wicker et al. 2005; Abrusán et al. 2008), which is corroborated by our chicken CR1 landscape (fig. 4) and is similar to the fate of CR1 activity in the lineage leading to the zebra finch ([supplementary fig. S4, Supplementary Material online](#)). Our amniote-wide genome analyses suggest extant, “ongoing” CR1 activity in several lineages, as we detected the presence of more than 0.1 Mb of very young CR1 elements (i.e., no sequence divergence from the consensus) in the genomes of anole lizard, painted turtle, sea turtle, and four crocodylians, respectively. Notably, the gharial lineage exhibits the highest extant activity of crocodylian CR1 retroposons with a total of approximately 2 Mb of virtually identical CR1-derived sequences in the gharial genome ([supplementary fig. S4, Supplementary Material online](#)). This could be a major part of the explanation of why the gharial genome assembly exhibits the lowest scaffold N50 value of all four assembled crocodylian genomes (Wan et al. 2013; Green et al. 2014). Thus, availability of the gharial genome promises to provide full-length, intact CR1 retroposon sequences for future in vitro studies of the mechanism of amniote CR1 proliferation.

#### Conclusions

This study is the first to infer crocodylian RE presence/absence patterns and demonstrates that these cladistic markers provide a conflict-free resolution of deep crocodylian phylogeny, including the unambiguous grouping of gharial and false gharial and the placement of that clade sister to the crocodiles. We provide comparative genomic evidence from 12 crocodylian genomes that, while there was some degree of recent CR1 diversification and succession of activity of CR1 subfamilies throughout crocodylian evolution despite their relative



**Fig. 4.**—Differential retention of ancient CR1 lineages during amniote evolution. Considering the consensus phylogeny of amniotes (Shedlock and Edwards 2009; Shaffer et al. 2013; Green et al. 2014) (left panel), comparison of divergence landscape plots (right panel; species designations are in bold letters on the left panel) of CR1 groups A–G from figure 3 (see also [supplementary table S3](#), [Supplementary Material](#) online) suggests retention of ancient CR1 diversity in crocodylian and turtle genomes, while most CR1 lineages became extinct in the ancestors of other amniote taxa, in particular birds and Lepidosauria (anole and python). We used the distribution of relics of CR1 groups among amniote genomes to parsimoniously infer durations of CR1 lineage retention (branches not to scale). These approximate durations are illustrated using colored lines corresponding to the respective CR1 group (from the right panel). Divergences of CR1 copies to their respective consensus sequences were calculated excluding CpG dinucleotide sites. See [supplementary figure S4](#), [Supplementary Material](#) online, for CR1 landscape plots of those species not shown here.

genomic stability (Green et al. 2014), most CR1 diversity and activity was present before the diversification of extant Crocodylia. Our genome-wide analyses of CR1 retroposons across amniote phylogeny revealed that both crocodylians and turtles contain a rich repertoire of ancient CR1 integrations that provide unique insights into the early genome evolution of amniotes. We conducted the hitherto first amniote-wide analyses of CR1 subfamilies and inferred that the genome of the amniote ancestor was impacted by the activity of at least seven CR1 subfamilies that subsequently gave rise to the CR1 groups that are detectable in extant amniotes. Both crocodylians and turtles have retained six of these ancient, mostly long-extinct CR1 groups, respectively. This probably reflects the exceptional genome stability and slow molecular evolution in turtles and crocodylians (Shaffer et al. 2013; Green et al. 2014) that makes these and other

nonfunctional sequences (e.g., endogenous hepadnaviruses [Suh et al. 2014]) recognizable even after >200 Myr of neutral decay.

CR1 retroelements are the most abundant, dominant group of TEs in some of the major lineages of amniotes (Shedlock et al. 2007). This applies to the genomes of birds (Hillier et al. 2004; Warren et al. 2010), crocodylians (Green et al. 2014), turtles (Shaffer et al. 2013), and snakes (Castoe et al. 2013), whereas mammalian genomes exhibit dominance of L1 or L2 LINEs (Lander et al. 2001; Mikkelsen et al. 2007; Warren et al. 2008) and the anole lizard genome contains various nearly equally dominant TEs (i.e., CR1 LINEs, L1 LINEs, L2 LINEs, LTR retroposons, DNA transposons) (Novick et al. 2009; Alföldi et al. 2011; Tollis and Boissinot 2011). Considering the phylogenetic relationships among CR1 lineages as well as their abundance in most amniote genomes, it is parsimonious to assume that dominant CR1 activity already

existed in the ancestral amniote genome (Janes et al. 2010) (i.e., >320 Ma [Shedlock and Edwards 2009]) and subsequently persisted throughout early amniote evolution. We hypothesize that the diversity and copy numbers of ancient CR1 lineages in the genomes of crocodylians and turtles reflect this ancestral genome organization which therefore must have been preserved in their common ancestor that lived more than 230 Ma (Shedlock and Edwards 2009). This is congruent with studies on amniote genome size evolution that reconstruct an ancestral amniote genome size comparable to that of crocodylians and turtles (Organ et al. 2007, 2011). Consequently, we propose that the smaller genomes of birds and lepidosaurs are the result of genome size reduction in their respective common ancestor via purging of ancient TE copies through rapid molecular evolution, accompanied by reduced TE expansion via inactivation of all but one of the multiple ancestral CR1 lineages. On the other hand, during early evolution of mammals, CR1 activity was replaced by a massive expansion of L2 LINES in monotremes (Warren et al. 2008) and L1 LINES in therians (Kordiš et al. 2006), which led to a drastic change in noncoding genome organization and a slight increase in genome size as the result of accumulation of hundreds of thousands of L1/L2 LINES and L1/L2-mobilized SINES after decay of ancient TE copies. We therefore conclude that the stable genomes of crocodylians and turtles constitute unique windows into the distant past of early amniote genome evolution and the processes that gave rise to the dissimilar genomic landscapes of mammalian and nonmammalian TEs.

## Supplementary Material

Supplementary data S1 and S2, tables S1–S3, figures S1–S4 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

J.J., coauthor of the present study, passed away shortly before manuscript submission. This article is dedicated to him and his scientific legacy. The authors thank Linnéa Smeds for helpful discussions and Jón Baldur Hlíðberg for providing the crocodylian paintings. The authors are grateful to the St. Augustine Alligator Farm Zoological Park (St. Augustine, FL) for samples, and to Ellen Pritham and two anonymous reviewers for providing valuable comments. This study was financially supported by the Deutsche Forschungsgemeinschaft (SCHM1469/3-2 to J.S.). D.A.R., E.L.B., and the International Crocodylian Genomes Working Group ([www.crocgenomes.org](http://www.crocgenomes.org)) were supported by the National Science Foundation (MCB-1052500, MCB-0841821, DEB-1020865, DUE-0920151) and funds from the Institute for Genomics, Biocomputing and Biotechnology at Mississippi State University. F.G.H. was partially funded by the National

Science Foundation (EPS-0903787). A.S. performed some of the computations on resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b2012135. Access to these resources was kindly made possible by Hans Ellegren who was supported by an Advanced Investigator Grant (NEXTGENMOLECOL) from the European Research Council, a Wallenberg Scholar Award from the Knut and Alice Wallenberg Foundation and grants from the Swedish Research Council (2007-8731 and 2010-5650).

## Literature Cited

- Abrusán G, Krambeck H-J, Junier T, Giordano J, Warburton PE. 2008. Biased distributions and decay of long interspersed nuclear elements in the chicken genome. *Genetics* 178:573–581.
- Alföldi J, et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 477:587–591.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410.
- Baker AJ, Haddrath O, McPherson JD, Cloutier A. 2014. Genomic support for a moa-tinamou clade and adaptive morphological convergence in flightless ratites. *Mol Biol Evol*. 31:1686–1696.
- Burch JBE, Davis DL, Haas NB. 1993. Chicken repeat 1 elements contain a *po*-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. *Proc Natl Acad Sci U S A*. 90: 8199–8203.
- Castoe TA, et al. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc Natl Acad Sci U S A*. 110: 20645–20650.
- Chong AY, et al. 2014. Evolution and gene capture in ancient endogenous retroviruses - insights from the crocodylian genomes. *Retrovirology* 11:71.
- Churakov G, et al. 2009. Mosaic retroposon insertion patterns in placental mammals. *Genome Res*. 19:868–875.
- Churakov G, et al. 2010. A novel web-based TinT application and the chronology of the primate *Alu* retroposon activity. *BMC Evol Biol*. 10:376.
- Dasmahapatra KK, et al. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487:94–98.
- Delfino M, et al. 2008. A complete skull of *Allodaposuchus precedens* Nopcsa, 1928 (Eusuchia) and a reassessment of the morphology of the taxon based on the Romanian remains. *J Vertebr Paleontol*. 28: 111–122.
- Diez CM, Meca E, Tenaillon MI, Gaut BS. 2014. Three groups of transposable elements with contrasting copy number dynamics and host responses in the maize (*Zea mays* ssp. *mays*) genome. *PLoS Genet*. 10: e1004298.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.
- Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461.
- Gatesy J, Amato G, Norell M, DeSalle R, Hayashi C. 2003. Combined support for wholesale taxic atavism in gavialine crocodylians. *Syst Biol*. 52:403–422.
- Green RE, et al. 2014. Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. *Science* 346:1335.
- Haas NB, et al. 2001. Subfamilies of CR1 non-LTR retrotransposons have different 5'-UTR sequences but are otherwise conserved. *Gene* 265: 175–183.
- Haas NB, Grabowski JM, Sivitz AB, Burch JBE. 1997. Chicken repeat 1 (CR1) elements, which define an ancient family of vertebrate non-LTR

- retrotransposons, contain two closely spaced open reading frames. *Gene* 197:305–309.
- Haddrath O, Baker AJ. 2012. Multiple nuclear genes and retroposons support vicariance and dispersal of the palaeognaths, and an Early Cretaceous origin of modern birds. *Proc R Soc B Biol Sci.* 279: 4617–4625.
- Han K-L, et al. 2011. Are transposable element insertions homoplasy free?: an examination using the avian tree of life. *Syst Biol.* 60: 375–386.
- Harshman J, Huddleston CJ, Bollback JP, Parsons TJ, Braun MJ. 2003. True and false gharials: a nuclear gene phylogeny of Crocodylia. *Syst Biol.* 52:386–402.
- Hillier LW, et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716.
- Holliday CM, Gardner NM. 2012. A new eusuchian crocodyliform with novel cranial integument and its significance for the origin and evolution of Crocodylia. *PLoS One* 7:e30471.
- Janes DE, Organ CL, Fujita MK, Shedlock AM, Edwards SV. 2010. Genome evolution in reptilia, the sister group of mammals. *Annu Rev Genomics Hum Genet.* 11:239–264.
- Janke A, Gullberg A, Hughes S, Aggarwal R, Arnason U. 2005. Mitogenomic analyses place the gharial (*Gavialis gangeticus*) on the crocodile tree and provide pre-K/T divergence times for most crocodylians. *J Mol Evol.* 61:620–626.
- Jarvis ED, et al. 2014. Whole genome analyses resolve the early branches in the tree of life of modern birds. *Science* 346:1320–1331.
- Jurka J, et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 110:462–467.
- Kaiser VB, van Tuinen M, Ellegren H. 2007. Insertion events of CR1 retrotransposable elements elucidate the phylogenetic branching order in galliform birds. *Mol Biol Evol.* 24:338–347.
- Kajikawa M, Ohshima K, Okada N. 1997. Determination of the entire sequence of turtle CR1: the first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. *Mol Biol Evol.* 14:1206–1217.
- Kapitonov VV, Jurka J. 2003. The esterase and PHD domains in CR1-like non-LTR retrotransposons. *Mol Biol Evol.* 20:38–46.
- Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform.* 9:286–298.
- Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 16:111–120.
- Kordiš D. 2009. Transposable elements in reptilian and avian (Sauropsida) genomes. *Cytogenet Genome Res.* 127:94–111.
- Kordiš D, Lovšin N, Gubenšek F. 2006. Phylogenomic analysis of the L1 retrotransposons in Deuterostomia. *Syst Biol.* 55:886–901.
- Kriegs JO, et al. 2007. Waves of genomic hitchhikers shed light on the evolution of gamebirds (Aves: Galliformes). *BMC Evol Biol.* 7:190.
- Lander ES, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
- Liu GE, Jiang L, Tian F, Zhu B, Song J. 2009. Calibration of mutation rates reveals diverse subfamily structure of galliform CR1 repeats. *Genome Biol Evol.* 1:119–130.
- Liu Z, He L, Yuan H, Yue B, Li J. 2012. CR1 retroposons provide a new insight into the phylogeny of Phasianidae species (Aves: Galliformes). *Gene* 502:125–132.
- Lovšin N, Gubenšek F, Kordiš D. 2001. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia. *Mol Biol Evol.* 18: 2213–2224.
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72: 595–605.
- Meganathan PR, Dubey B, Batzer MA, Ray DA, Haque I. 2011. Complete mitochondrial genome sequences of three *Crocodylus* species and their comparison within the Order Crocodylia. *Gene* 478:35–41.
- Mikkelsen TS, et al. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447:167–177.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans (LA): Institute of Electrical and Electronics Engineers (IEEE). p. 1–8.
- Nishihara H, Maruyama S, Okada N. 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proc Natl Acad Sci U S A.* 106:5235–5240.
- Novick PA, Basta H, Floumanhaft M, McClure MA, Boissinot S. 2009. The evolutionary dynamics of autonomous non-LTR retrotransposons in the lizard *Anolis carolinensis* shows more similarity to fish than mammals. *Mol Biol Evol.* 26:1811–1822.
- Oaks JR. 2011. A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles. *Evolution* 65:3285–3297.
- Organ CL, Canoville A, Reisz RR, Laurin M. 2011. Paleogenomic data suggest mammal-like genome size in the ancestral amniote and derived large genome size in amphibians. *J Evol Biol.* 24:372–380.
- Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV. 2007. Origin of avian genome size and structure in non-avian dinosaurs. *Nature* 446: 180–184.
- Pagán HJT, Smith JD, Hubley RM, Ray DA. 2010. *PiggyBac*-ing on a primate genome: novel elements, recent activity and horizontal transfer. *Genome Biol Evol.* 2:293–303.
- Petrov DA, Hartl DL. 1998. High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. *Mol Biol Evol.* 15: 293–302.
- Ray DA, Xing J, Salem A-H, Batzer MA. 2006. SINEs of a nearly perfect character. *Syst Biol.* 55:928–935.
- Scheyer TM, et al. 2013. Crocodylian diversity peak and extinction in the late Cenozoic of the northern Neotropics. *Nat Commun.* 4:1907.
- Sereno P, Larsson H. 2009. Cretaceous crocodyliforms from the Sahara. *ZooKeys* 28:1–143.
- Shaffer HB, et al. 2013. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* 14:R28.
- Shedlock AM. 2006. Phylogenomic investigation of CR1 LINE diversity in reptiles. *Syst Biol.* 55:902–911.
- Shedlock AM, Edwards SV. 2009. Amniotes (Amniota). In: Hedges SB, Kumar S, editors. *The timetree of life*. New York: Oxford University Press. p. 375–379.
- Shedlock AM, et al. 2007. Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc Natl Acad Sci U S A.* 104:2767–2772.
- Shedlock AM, Takahashi K, Okada N. 2004. SINEs of speciation: tracking lineages with retroposons. *Trends Ecol Evol.* 19:545–553.
- Smit A, Hubley R, Green P. 1996-2014. RepeatMasker Open-3.3.0. Available from: <http://www.repeatmasker.org>.
- St John J, et al. 2012. Sequencing three crocodylian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biol.* 13:415.
- Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAXML web servers. *Syst Biol.* 57:758–771.
- Stumph WE, Hodgson CP, Tsai MJ, O'Malley BW. 1984. Genomic structure and possible retroviral origin of the chicken CR1 repetitive DNA sequence family. *Proc Natl Acad Sci U S A.* 81:6667–6671.
- Stumph WE, Kristo P, Tsai M-J, O'Malley BW. 1981. A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. *Nucleic Acids Res.* 9:5383–5398.

- Suh A, et al. 2014. Early Mesozoic coexistence of amniotes and Hepadnaviridae. *PLoS Genet.* 10:e1004559.
- Suh A, Kriegs JO, Brosius J, Schmitz J. 2011. Retroposon insertions and the chronology of avian sex chromosome evolution. *Mol Biol Evol.* 28: 2993–2997.
- Suh A, Kriegs JO, Donnellan S, Brosius J, Schmitz J. 2012. A universal method for the study of CR1 retroposons in non-model bird genomes. *Mol Biol Evol.* 29:2899–2903.
- Suh A, Paus M, et al. 2011. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat Commun.* 2:443.
- Tollis M, Boissinot S. 2011. The transposable element profile of the *Anolis* genome: how a lizard can provide insights into the evolution of vertebrate genome size and structure. *Mob Genet Elements.* 1: 107–111.
- Vandergon TL, Reitman M. 1994. Evolution of chicken repeat 1 (CR1) elements: evidence for ancient subfamilies and multiple progenitors. *Mol Biol Evol.* 11:886–898.
- Vonk FJ, et al. 2013. 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.
- Waddell PJ, Kishino H, Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Inform.* 12:141–154.
- Wan Q-H, et al. 2013. Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. *Cell Res.* 23:1091–1105.
- Wang Z, et al. 2013. 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.
- Warren WC, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453:175–183.
- Warren WC, et al. 2010. The genome of a songbird. *Nature* 464:757–762.
- Wicker T, et al. 2005. The repetitive landscape of the chicken genome. *Genome Res.* 15:126–136.

**Associate editor:** Ellen Pritham