Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 362 (2005) 1 – 10
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Chompy: An infestation of MITE-like repetitive elements in the crocodilian genome

David A. Ray [a,b], Dale J. Hedges [a], Scott W. Herke [a], Justin D. Fowlkes [a], Erin W. Barnes [a], Daniel K. LaVie [a], Lindsey M. Goodwin [a], Llewellyn D. Densmore [c], Mark A. Batzer [a,*]

[a] *Department of Biological Sciences, Biological Computation and Visualization Center, Center for Bio-Modular Multiscale Systems, Louisiana State University, 202 Life Sciences Bldg., Baton Rouge, LA 70803, USA*
[b] *Department of Biology, West Virginia University, PO Box 6057, Morgantown, WV 26506, USA*
[c] *Department of Biological Sciences, Texas Tech University, Box 43131, Lubbock, TX 79409, United States*

Received 23 May 2005; received in revised form 1 July 2005; accepted 7 July 2005
Available online 23 September 2005
Received by W. Makalowski

Abstract

Interspersed repeats are a major component of most eukaryotic genomes and have an impact on genome size and stability, but the repetitive element landscape of crocodilian genomes has not yet been fully investigated. In this report, we provide the first detailed characterization of an interspersed repeat element in any crocodilian genome. *Chompy* is a putative miniature inverted-repeat transposable element (MITE) family initially recovered from the genome of *Alligator mississippiensis* (American alligator) but also present in the genomes of *Crocodylus moreletii* (Morelet's crocodile) and *Gavialis gangeticus* (Indian gharial). The element has all of the hallmarks of MITEs including terminal inverted repeats, possible target site duplications, and a tendency to form secondary structures. We estimate the copy number in the alligator genome to be ~46,000 copies. As a result of their size and unique properties, *Chompy* elements may provide a useful source of genomic variation for crocodilian comparative genomics.

Keywords: Archosaurian; MITE; Repetitive element; Transposable element

1. Introduction

The order Crocodylia is an excellent group for studying evolutionary patterns because it is not only ancient and ecologically diverse, but is also characterized by an easily manageable number of extant species (<25). The small number of taxa makes it relatively easy to handle data for comparative analyses. However, while the number of species is low, members of the order have colonized every tropical habitat and some temperate zones. Species occupy not only freshwater lakes and rivers, but can also be found miles offshore in the ocean and may traverse long distances over land (Ross, 1998). In addition, extant species provide examples of divergence times ranging from >60 my between *Alligator* and *Crocodylus* to <6 my between certain members of *Crocodylus* (Brochu, 2003, 2004).

Data on the genome structure of crocodilians is scarce. Genome size has been estimated for only five species representing only two of the three extant families (Alligatoridae, Crocodylidae, and Gavialidae) (Mirsky and Ris, 1951; Atkin et al., 1965; Olmo, 1976, 1981; De Smet, 1981; Capriglione et al., 1987). As with other reptiles, genome sizes are intermediate between birds and mammals. The largest observed genome sizes are found in *Crocodylus niloticus* (3.95) (Capriglione et al., 1987) and *Caiman latirostris* (3.90) (Olmo, 1981). The smallest genome observed belongs to *Alligator mississippiensis* (2.49) (Mirsky and Ris, 1951). Diploid chromosome numbers in crocodilians range from 30 to 42. However, the level of

Abbreviations: MITE, Miniature inverted-repeat transposable element; IE PCR, intra-element PCR; iPCR, inverse PCR.

* Corresponding author. Tel.: +1 225 578 7102; fax: +1 225 578 7113.
E-mail address: mbatzer@lsu.edu (M.A. Batzer).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.07.010

conservation of the genome is striking when it is noted that all species have a corrected number of arms of 42 and a fundamental number showing only 10% variation among all taxa. These observations contribute to the idea that the basic crocodilian karyotype is highly conserved and suggests relatively static genomes among members of the order (Cohen and Gans, 1970).

Our understanding of issues related to crocodilian biology and genome evolution may be improved by determining whether interspersed repeats have affected the relative stability of crocodilian genomes. For instance, the movement of these elements can change genomes by insertion mutagenesis (Deininger and Batzer, 1999; Kidwell and Lisch, 2001; Batzer and Deininger, 2002; Baust et al., 2002; Brouha et al., 2003; Ostertag et al., 2003), insertion-mediated genomic deletions (Kidwell and Lisch, 2001; Gilbert et al., 2002; Symer et al., 2002; Callinan et al., 2005), and transposition-mediated transduction (Goodier et al., 2000; Pickeral et al., 2000; Kidwell and Lisch, 2001; Jiang et al., 2004). Further, the presence of essentially identical fragments of DNA scattered throughout a genome promotes recombination, including both equal and unequal crossover events (Deininger and Batzer, 1999; Gebow et al., 2000; Kidwell and Lisch, 2001), segmental duplication (Bailey et al., 2003), gene conversion (Kass et al., 1995; Roy-Engel et al., 2002; Fischer et al., 2003), exon shuffling (Ejima and Yang, 2003; Jiang et al., 2004) and chromosomal rearrangements (Gray, 2000).

MITEs (miniature inverted-repeat transposable elements) are repetitive DNA elements typically ranging in size from ~80 to 500 bp in length but also reaching lengths of up to 1.6 kb (Feschotte et al., 2002). While these elements were first reported and have been best characterized in plants, several studies over the past decade have revealed their existence in other organisms including amphibians (Lepetit et al., 2000), fish (Izsvak et al., 1999), nematodes (Oosumi et al., 1995, 1996) mosquitoes (Tu, 1997, 2001), fungi (Yeadon and Catcheside, 1995), and humans (Morgan, 1995; Smit and Riggs, 1996).

These elements are characterized by the presence of terminal inverted repeats (TIR), a relatively small size, no coding potential and A+T richness. They are often found associated with genes and, along with other repetitive elements in the genome, are thought to contribute substantially to genome instability. Many families of MITEs have been identified, but classification has been difficult because they rarely share significant sequence identity (Feschotte et al., 2002) despite their common structural motifs.

We have identified a group of MITE-like families, dubbed *chompy*, in the genomes of three crocodilians. These elements share a common terminal inverted repeat motif, are highly variable in size, and can be grouped into several families. This study represents the first characterization of an interspersed repeat family in any crocodilian genome as well as the first steps in increasing our understanding of the repetitive landscape of crocodilian genomes.

2. Materials and methods

2.1. Sequence characterization of chompy elements

Computational analysis of available BAC sequences from the *A. mississippiensis* genome (~0.81 Mb, accession numbers AC148578,AC148923, AC148964, AC149025, AC149026, AC149028, and AC149029) using RECON (Bao and Eddy, 2002) revealed several potential repetitive DNA elements. The presence of conserved terminal inverted repeats in one of these potential interspersed repeat families prompted further investigation of an element we termed *chompy*.

To experimentally determine the boundaries of the *chompy* element, we implemented a linker PCR protocol described previously (Roy et al., 1999; Ray et al., 2005). To summarize, forward and reverse primers were designed to amplify in both directions from within the element to oligonucleotide linkers ligated to genomic DNA previously digested with *Avr*II. Because we had limited access to *A. mississippiensis* DNA, we used *Crocodylus moreletii* (Morelet's crocodile) genomic DNA for the linker ligation protocol. Oligonucleotide primer and linker sequences are listed in Table 1. Digestions, ligation reactions, PCR amplifications, cloning reactions and sequencing of cloned fragments were performed as previously described (Ray et al., 2005).

Endpoints of the representative *chompy* sequences were first determined manually, using sequence alignments of 109 linker PCR fragments. The endpoints were confirmed by using BioEdit v7.0 (Hall, 1999) to generate entropy plots, which measure the entropic variability at any given position in an alignment. Fig. 1 illustrates the increase in entropy at the ends of the 5′ and 3′ conserved sequences relative to flanking sequence for *chompy* elements.

Once the boundaries of the *chompy* elements had been determined, a consensus sequence for the 5′ and 3′ ends was constructed (see Fig. 2 for a diagram illustrating sequence relationships). The original BAC sequence data plus an additional ~1.17 Mb of data not available during the initial RECON analysis of *A. mississippiensis* sequences (AC154087, AC154169,AC154170, AC154945, AC155799, AC155800, AC155801, and AC155802) were masked for the ends of the element using a local version of repeatmasker (A.F.A. Smit and P. Green, unpublished data Current Version: 20040306-web; http://www.repeatmasker.org). The recovered ends and the intervening sequences were aligned in BioEdit.

We designed oligonucleotide primers at the extreme 5′ and 3′ ends of the *chompy* consensus sequence to amplify full-length representatives from *A. mississippiensis*, *C. moreletii*, and *Gavialis gangeticus* (Indian gharial) through intra-element PCR. Twenty-five μl PCR amplifications were performed under the following conditions: 10–50 ng template DNA, 7 pM each of primers *chompy*-F and *chompy*-R (Table 1), 200 mM dNTPs, 50 mM KCl, 10 mM

Table 1
Oligonucleotide sequences used for the characterization of *chompy* elements using intra-element PCR (IE PCR), linker PCR, iPCR (shaded), and population and phylogenetic analyses

	Primer	Sequence (5′–3′)
IE PCR	*chompy*-F	GGCTAGGGACAGAHATTACACATAAACYNG
	chompy-R	GGCTAGGGACAGACATTCAAAAAGCC
Linker protocol	*chompy*7-F	CAGACTTAACTGATTTGGCTCAAACCGG
	*chompy*7-R	CCGGTTTGAGCCAAATCAGTTAAGTCTG
	*chompy*8-F	ATTAGCTAGCCGACCACATGCTGGC
	*chompy*8-R	GCCAGCATGTGGTCGGCTAGCTAAT
	*chompy*7B-F	CTTCTAGCCTGAGCCACTACAGG
	*chompy*7B-R	CCTGTAGTGGCTCAGGCTAGAAG
	*chompy*8B-F	CCTGACCAGGTTAGACTAACCTGC
	*chompy*8B-R	GCAGGTTAGTCTAACCTGGTCAGG
	*chompy*8(C)-F	CAGATGTTCAGTGCACATAAACCAG
	*chompy*8(C)-R	CTGGTTTATGTGCACTGAACATCTG
	Avr-top	CCTAGGAAGGAGAGGACGCTGTCTGTCGAAGG
	Avr-bottom	GAGCGAATTCGTCAACATAGCATTTCTGTCCTCTCCTTC
	LNP	GAATTCGTCAACATAGCATTTCT
Population and phylogeny panels	Cmor_iPCR_4_14-F	CCCGTCCTCGGTGCCTGATAC
	Cmor_iPCR_4_14-R	CTGTCAGCCCAGTCTTACCTTGGTC
	Cmor_iPCR_4_16-F	GGTCAACATCCAAGTGGAGCCAG
	Cmor_iPCR_4_16-R	GGCTCTGTCCCTGCACTCCTG
	Cmor_iPCR_4_4-F	GCATTGAAACTTTTGCCACTCCCTCTC
	Cmor_iPCR_4_4-R	CTGCCCATTGTGCTAGGAAGTGTAC
	Cmor_iPCR_3_20-F	GACAGGTGCATAAAACGAGGA
	Cmor_iPCR_3_20-R	GCACAACATTTTGGGTGAAA
	Cmor_iPCR_3_27-F	CATGGGTAGGAGGACCAGAA
	Cmor_iPCR_3_27-R	TGTCCAAGAACAATTTAGGATTTG
	Cmor_iPCR_3_30-F	CTATGTGGGTGGTTGGAAGC
	Cmor_iPCR_3_30-R	CAGAGCCCTGACCTTACTGG
	Cmor_iPCR_4_30-F	CTCTCAGGGAGCAGAACAGG
	Cmor_iPCR_4_30-R	GGCCTGTCAATGACCAAGAT
	Cmor_iPCR_4_11-F	CACCAACCCAGGCACAACTGG
	Cmor_iPCR_4_11-R	AAAGACATCTATGTGTAGATGCATGAGGC
	Cmor_iPCR_4_22-F	CATCCTGCCCTGCTTATGTA
	Cmor_iPCR_4_22-R	GTGGCAGTAGGAGCACCTTT
	Cmor_iPCR_3_21-F	GCTCTCTGAACAGGACATGCT
	Cmor_iPCR_3_21-R	GTCTCAGGGTTTGGGCTTAC
	Cmor_iPCR_3_31-F	TGCCTTTTTCCTTTCAGTTCC
	Cmor_iPCR_3_31-R	GGCACAAAAAGCACTGGAAT
	Cmor_iPCR_3_33-F	CAGGCATGACCACAGAGTCA
	Cmor_iPCR_3_33-R	CACTTTGTAGCTAGCAGGGTCA

Tris–HCl (pH 8.4), 2.0 mM $MgCl_2$ and Taq DNA polymerase (2.5 units). An initial denaturation at 94 °C (120 s) was followed by 32 cycles of 94 °C (15 s), 58 °C (15 s), and 72 °C (70 s). A final incubation at 72 °C (3 min) prepared the fragments for TA cloning.

The resulting smeared PCR products were excised from agarose gels as three fragment ranges (200–400, 400–700, and 700+ bp) to limit preferential insertion of smaller fragments into the cloning vector. Products were purified using the Wizard PCR gel purification kit (Promega) and cloned using the TOPO-TA cloning kit (Invitrogen, Carlsbad, CA). For each of the three taxa, 96 colonies (32 colonies per fragment range) were randomly selected for sequencing using chain termination sequencing (Sanger et al., 1977) on an ABI 3100 Genetic Analyzer. Duplicated sequences were removed and the remaining unique sequences were organized into groups representing families of similar sequences and lengths. Consensus sequences for these families of *chompy* elements were generated in BioEdit v7.0 using a simple "majority rules" approach for families with ≥ 4 representatives. Standard IUBMB codes were used when there was no majority at a particular site.

2.2. Inverse PCR

We used genomic DNA from *C. moreletii* to perform inverse-PCR (iPCR) in an effort to identify full-length elements with flanking sequences. First, genomic DNA (5 μg) was digested with the restriction endonuclease *Pvu*II, using the manufacturer's instructions. Afterwards, the digested DNA (500 ng) was self-ligated in a volume of 500 μl; the circularized products were ethanol precipitated

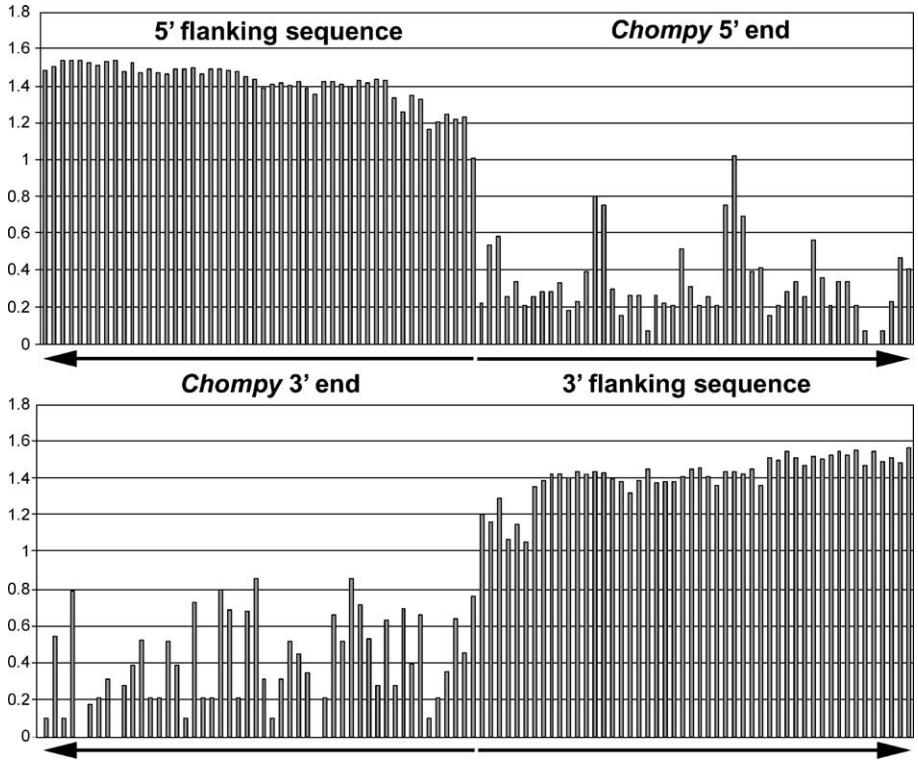

Fig. 1. Entropy plots spanning 100 bp around the 5′ and 3′ ends of *chompy* elements.

using a glycogen carrier and resuspended in 50 μl TLE (10 mM Tris/0.1 mM EDTA). The circularized products (5 μl) served as template for a 50 μl iPCR reaction using the primers *chompy*7B-F and *chompy*8(C)-R (Table 1). Reaction conditions were as described above. PCR products were purified, cloned and sequenced. Oligonucleotide primers flanking the insertions were designed using Primer3 (Rozen and Skaletsky, 1998) and tested on *C. moreletii* DNA. Positive results were cloned and sequenced as described above to verify the presence of the *chompy* insertion. All of the alignments mentioned above are available as supplementary files from our website, http://batzerlab.lsu.edu, under publications. Sequences generated for this project have been deposited in Genbank under accession numbers DQ063271–DQ063547.

2.3. Chompy element distribution

We were able to test twelve of the iPCR-generated loci (Table 1) for evidence of recent mobilization within *C. moreletii* by amplifying each locus from a panel of fifteen individuals. We also tested a crocodilian DNA panel (*C. moreletii*, *C. acutus*, *C. intermedius*, *C. niloticus*, *C. porosus*, *Osteolaemus tetraspis*, *G. gangeticus*, and *A. mississippiensis*) to determine when these elements had inserted relative to the various speciation events.

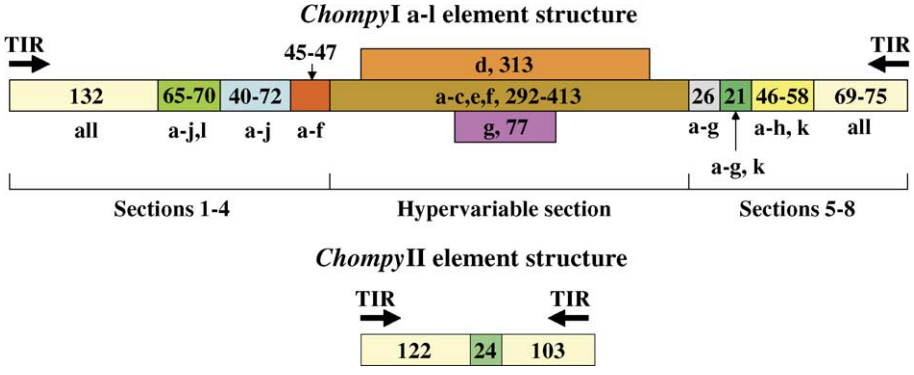

Fig. 2. Diagram representing the structure of *chompy*I and II elements. Color coded regions indicate highly similar sequence among family consensus sequences. Numbers indicate approximate sizes for the regions. Lower case letters indicate the family(s) of *chompy*I elements bearing the sequence motifs. Arrows indicate the terminal inverted repeat (TIR) sequences.

Table 2
Potential target site duplications and occurrences recovered for *chompy* elements

TSD sequence	Count	Sequence ID
AA	1	Cmor_iPCR_3_29
AAA	3	Cmor_iPCR_3_20, Cmor_iPCR_4_23, AC154088.1_Al_C_fam7_con_25
ACT	1	Cmor_iPCR_3_30
AG	1	Cmor_iPCR_4_30
AGR	1	Cmor_iPCR_3_36
AR	1	Cmor_iPCR_4_25
ATA	1	Cmor_iPCR_4_33
ATG	1	AC154088.1_Al_C_fam7_con_26
CA	2	AC155801.1_Al_C_fam7_con_46, AC154170.1_Al_C_fam7_con_32
CAA	1	Cmor_iPCR_3_23
CAG	1	AC155799.1_Al_+_fam7_con_42
CCT	1	Cmor_iPCR_3_21
CT	2	Cmor_iPCR_3_19, AC154170.1_Al_+_fam7_con_29
CTG	1	Cmor_iPCR_4_17
CYA	1	Cmor_iPCR_4_21
CYT	1	Cmor_iPCR_3_28
TA	1	AC154170.1_Al_C_fam7_con_30
TAA	1	Cmor_iPCR_3_27
TAG	3	Cmor_iPCR_3_32, Cmor_iPCR_4_28, Cmor_iPCR_4_42
TAT	1	Cmor_iPCR_3_25
TGA	1	Cmor_iPCR_3_17
TR	1	Cmor_iPCR_3_33
TTA	1	Cmor_iPCR_3_35
TTR	2	Cmor_iPCR_4_29, AC148578.2_Al_+_fam7_con_6
TTT	1	Cmor_iPCR_3_26
WAG	1	Cmor_iPCR_3_22
YA	1	Cmor_iPCR_4_20
YAG	1	Cmor_iPCR_3_31
YTA	1	Cmor_iPCR_4_22

Sequence IDs refer to sequences alignments available as supplemental data.

To investigate the broader phylogenetic distribution of *chompy*, consensus sequences from each family were subjected to BLAT (UCSC Genome Browser: http://genome.ucsc.edu/, (Kent et al., 2002) searches of all complete genome sequences available at the time of writing (human, chimpanzee, chicken, mouse, rat, opossum, cow, dog, zebrafish, *Fugu*, *Tetraodon*, *Xenopus*, *Caenorhabditis intestinalis*, and *Drosphila* spp.). BLAST searches of the NCBI non-redundant (nr) and high-throughput genomic sequence (htgs) databases were also performed.

2.4. Copy number estimation and coding potential

The copy number for *chompy* elements in the *A. mississippiensis* genome was estimated by counting occurrences of the 5′ and 3′ conserved ends of the element separately in the ~2 Mb available from the NIH sequence database. We then extrapolated each of these numbers to the whole genome. Representative sequences from each *chompy* family were submitted to tBLASTX (translated query vs. translated database) and BLASTX (translated query vs. protein database) searches to investigate similarity to known protein sequences.

3. Results

3.1. Chompy element structure

All *chompy* elements share conserved 5′ and 3′ end sequences. The first 47 bp of all *chompy* elements are nearly perfectly conserved, and this region contains a perfect TIR with the sequence GGCTAGGGACAGACATT (Fig. 2). Based on length and sequence identity, there are at least two distinct groups of MITE-like families (*chompy*I and *chompy*II).

The *chompy*I group of MITE-like families is very diverse, with consensus sequences ranging from 277 to 894 bp (Fig. 2) and considerable variation in length among its members (a–l). The conserved 5′ end of *chompy*I elements consists of a 132 bp fragment that is shared among all families. The 3′ end is also highly conserved and consists of a 69–75 bp fragment. Between these ends, the elements demonstrate a hierarchical nature with some sections being shared among particular families and others being unique to one family or another. For example, the most common family (*chompy*I-g) is 503 bp long and contains three major parts: sections 1 (132 bp), 2 (69 bp) and 3 (45 bp); a 77 bp family-specific region; and sections 5 (26 bp), 6 (21 bp), 7 (58 bp) and 8 (75 bp). Unlike elements of families I-a to I-f, *chompy*I-g elements lack section 4. This pattern suggests that each family may have originated from a distinct progenitor while relying on the same source of transposase for expansion.

With 59 members, the *chompy*I-g family was represented most frequently in the various sequence alignments (17 was the next highest representation). Family sequence alignments are available as supplementary files on our website, http://batzerlab.lsu.edu.

Only six *Chompy*II elements were identified, and their consensus sequence of 249 bp (Fig. 1b) is considerably shorter than that of most *chompy*I elements. These elements consist mainly of the conserved 5′ and 3′ ends with a short G+C-rich central section. Although the 5′ end shares high sequence similarity with the first 122 bp of *chompy*I elements, it remains distinct. This is demonstrated by the 3.9% mean within group sequence divergence (K2P) for all *chompy*I elements along the first 122 bp versus the 20% average divergence between the *chompy*II 5′ consensus and the 5′ end of *chompy*I elements. A similar pattern is found at the 3′ end. The final 103 bp of *chompy*II elements is very similar to a portion of section 7 and all of section 8 of *chompy*I elements. However, the mean divergence (K2P) between the final 103 bp of *chompy*I elements is 2.7%, while the average divergence between *chompy*I and *chompy*II elements over the same region is 10 times higher.

The A+T content of *chompy* elements tends to decrease as the total length of the elements increases. Nevertheless, *chompy* elements can be subdivided into A+T and G+C rich zones. For instance, the conserved 5′ and 3′ ends (i.e., the core components) of *chompy* elements are A+T rich (~58%). By contrast, the variable central sections (2–7, and H) found in the *chompy*I families are so C+G rich that the longest elements (*chompy*I-a) have an overall A+T content of only ~47%, despite the A+T rich 5′ and 3′ends.

To examine whether the *chompy* elements encode any proteins, representative sequences from each *chompy* family were submitted to tBLASTX (translated query vs. translated database) and BLASTX (translated query vs. protein database) searches search to investigate similarity to known protein sequences. No potential coding sequences were identified among the *chompy* consensus sequences that were examined. This is not unexpected given that MITEs are typically non-coding entities (Feschotte et al., 2002).

3.2. Evidence of previous mobilization

MITEs tend to have short (2–3 bp) A+T rich target site duplications (TSD) (Feschotte et al., 2002). From the 37 iPCR unique sequences generated for *chompy* elements, we found 28 with potential TSDs (80%, Table 2); of the eleven full-length elements recovered from the *Alligator* BAC sequences, eight contained potential TSDs (73%). TSDs were either two or three bases long and tended to be A+T rich (74% of the total non-ambiguous bases). No single TSD sequence dominated. Examination of the first three bases adjacent to either end of the element compared to 100 bases flanking either end showed a small, but not significant, increase in average A+T content. Given the lack of homogeneity in the sequences of these potential TSDs it is premature to definitively identify them as authentic target site duplications. Future characterization of orthologous "filled" and "empty" sites to verify the sequence duplication will serve to confirm this hypothesis.

Table 3
CpG to non-CpG mutation density ratios in the members of various *chompy* families

Element type	Sample size	CpG substitution density	Non-CpG substitution density	CpG : non-CpG mutation ratio
*chompy*I-a	17	0.29	0.19	1.47
*chompy*I-b	16	0.26	0.11	2.44
*chompy*I-c	11	0.24	0.11	2.21
*chompy*I-d	9	0.31	0.10	3.07
*chompy*I-e	9	0.35	0.09	3.76
*chompy*I-f	13	0.24	0.10	2.48
*chompy*I-g	59	0.67	0.12	3.13
*chompy*I-h	4	0.27	0.09	2.87
*chompy*I-i	6	0.00	0.12	0.00
*chompy*I-j	6	0.25	0.10	2.51
*chompy*I-k	6	0.25	0.13	1.85
*chompy*I-l	8	0.27	0.18	1.52
*chompy*II	6	0.10	0.20	0.50

3.3. Copy number and genomic representation

A. mississippiensis has an estimated genome size of ~2.5 Gb (Mirsky and Ris, 1951; Gregory, 2005). Thirty-seven instances each of the 5′ and the 3′ conserved ends of *chompy* were recovered from the ~2 Mb of sequence data using a local version of repeatmasker (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker). The resulting calculations suggest a copy number of around 46,000 full-length elements in the *A. mississippiensis* genome, which is within the range found for other MITEs in vertebrate genomes (Tu, 1997, 2001; Izsvak et al., 1999; Feschotte et al., 2002; Gaffney et al., 2003). Using this copy number and an average length of 503 bp (the length of the most common consensus sequence) we estimate that ~23 Mb (0.92%) of the *A. mississippiensis* genome is made up of *chompy* elements. It should be noted, however, that MITEs display a strong tendency to insert in genic regions (Tu, 1997; Feschotte and Mouches, 2000). This may influence their distribution and future estimates of copy number.

3.4. Nucleotide diversity

During the course of our sequence analysis, we noted what appeared to be relatively high rates of mutation at CpG dinucleotides in *chompy* elements. To investigate this further, we used a perl script originally designed to estimate CpG dinucleotide mutation rates in primate *Alu* elements (Xing et al., 2004). Briefly, this script identifies CpG dinucleotides in the consensus sequence, and then compares the number of substitutions at these sites to the number of mutations at non-CpG sites. These data were used to determine the ratio of CpG to non-CpG mutations, which in *chompy* elements averaged 2.14 (Table 3).

3.5. Chompy element distribution

Within the *C. moreletii* population panel, no size variation was observed at any of the iPCR-derived loci. When these same loci were tested for members of more divergent taxa representing Alligatoridae, Gavialidae, and Crocodylidae, amplification was often difficult to obtain. However, when successful amplification was achieved in multiple taxa from all three families, fragments of identical or near-identical sizes were always obtained. Thus, at least at these loci, the *chompy* elements were inserted prior to the divergence of the three crocodilian families.

Chompy-like elements were not identified in BLAT searches of all genome sequences available through the UCSC Genome Bioinformatics database. No matches of greater than 34 bp of contiguous sequence were found. Thus, this family of MITEs appears to be specific to crocodilians. Searches of the nr and htgs databases at NCBI resulted only in matches to the *Alligator* sequences that had been used in the original RECON and Repeatmasker searches.

4. Discussion

Crocodilians represent a fascinating group for evolutionary biologists for multiple reasons. One feature of crocodilian biology that has been of interest is the apparently high potential for morphological convergence and its role in confounding certain aspects of phylogenetic reconstruction in the group (Langston, 1973; Gatesy et al., 2003). Superficially, working from the neck backwards, one crocodilian is much like any other. While there are some differences in body musculature that have been investigated (Frey et al., 1989), the general body plan has been very successful and has been retained since at least the late Cretaceous.

The apparent lack of reproductive isolation among members of genus *Crocodylus* is another intriguing aspect of crocodilian evolutionary biology. Although considered valid species, it is well-known among persons farming crocodilians that all species from the genus have the potential to hybridize and produce fertile offspring. Examples of such events have been reported between both sympatric (Ramos et al., 1994) and allopatric species (Fitzsimmons et al., 2002). Thus, the traditional biological species concept (Mayr, 1942) does not seem to apply to the genus. This has complicated efforts to reintroduce captive individuals of *C. siamensis* to the wild (Fitzsimmons et al., 2002) and attempts to resolve population structure in *C. moreletii* (Ray et al., 2004).

It might be suggested that the relative stability of crocodilian genomes as identified by Cohen and Gans (1970) could contribute to both of the above issues. For example, a relatively static genome may result in morphological stasis. It may also delay the development of post-zygotic reproductive isolation mechanisms as is exemplified by *Crocodylus* spp. The influence of repetitive elements on genome structure and stability has been well characterized in insects (Deininger and Roy-Engel, 2002; Labrador and Corces, 2002; Nouaud et al., 2003), rodents (Waterston et al., 2002) and humans (Kazazian and Moran, 1998; Deininger and Batzer, 1999; Lander et al., 2001; Kazazian and Goodier, 2002). Thus, the study of repetitive DNA elements in the genomes of crocodilians may be critical to our understanding of these and other aspects of their evolution.

Unfortunately, investigations of interspersed repeats in crocodilians have been limited to only a few studies. For example, L2 elements were not found in *Caiman* or *Alligator* (Lovsin et al., 2001), while CR1 elements were found in a *Caiman* genome (Ohshima et al., 1996). The current study represents the first characterization of an interspersed repeat family in any crocodilian genome.

When we were able to amplify full-length iPCR-recovered *chompy* loci in representatives of all three families in our crocodilian panel, amplicon sizes showed no evidence that the element was missing in any taxa. This observation suggests that *chompy* elements are evolutionarily relatively old and probably expanded during a short period before the divergence of the three extant families. It should be noted, however, that we were only able to test a limited number of loci that may have been subject to sampling error. For example, when examining the *Alu* family of SINEs in humans, over 80% of the elements inserted before the divergence of Old and New World primates ~35 mya (Shen et al., 1991; Deininger and Batzer, 1999). Thus, a phylogenetic analysis of primates using a small number of randomly chosen *Alu* insertion loci would be unlikely to show any evidence of polymorphism. Nevertheless, we know that examination of large numbers of carefully chosen *Alu* loci are informative at all levels of primate phylogeny (Salem et al., 2003b; Singer et al., 2003; Roos et al., 2004; Ray et al., 2005; Xing et al., in press) and even within populations (Batzer and Deininger, 1991, 2002; Perna et al., 1992; Batzer et al., 1994; Stoneking et al., 1997; Salem et al., 2003a; Watkins et al., 2003; Xing et al., 2003; Carter et al., 2004; Otieno et al., 2004; Ray et al., 2005). The observation that the average sequence divergence at the 5′ and 3′ ends of *chompy*I elements is between 2% and 4% suggests that some of these elements may actually be relatively young.

Regardless of whether *chompy* elements are polymorphic between or within species, the presence of these elements has prompted certain observations that have contributed to our nascent understanding of crocodilian genomics. One example involves an increased rate of CpG dinucleotide mutations. Methylation of deoxycytosine to create 5-methylcytosine (5mC) is a common epigenetic modification within vertebrate genomes. Methylation has been associated with several important cellular functions including control of gene expression, organismal development and repression of transposable element activity (Batzer and Deininger, 1991; Bestor et al., 1994; Bestor, 1998; Schmid, 1998; Paulsen and Ferguson-Smith, 2005). In vertebrate genomes, 60–90% of the CpG dinucleotides contain 5mC (Tweedie et al., 1997; Hendrich and Tweedie, 2003). The 5mC at CpG sites mutates unidirectionally to thymine by spontaneous deamination at a much higher transition rate compared to non-CpG dinucleotides (Coulondre et al., 1978), leading to a rapid decay of CpG sites. This process is in turn thought to cause the observed deficiency in CpG dinucleotides and the corresponding increase in TpG and CpA dinucleotide frequency in vertebrate genomes. In the human genome, for example, CpG dinucleotides are present only at about 20% of their expected frequency (Lander et al., 2001).

While the mutation density difference between CpG and non-CpG dinucleotides in the *A. mississippiensis* sequences we obtained was not as high as the 6-fold increase described for primate genomes (Xing et al., 2004), there was a consistently higher mutation rate for these sites when compared to non-CpG sites. Furthermore, the true value may actually be higher than what we have reported. In the process of generating consensus sequences, there were YpG or CpR locations where the conversion of a C to a T or a G

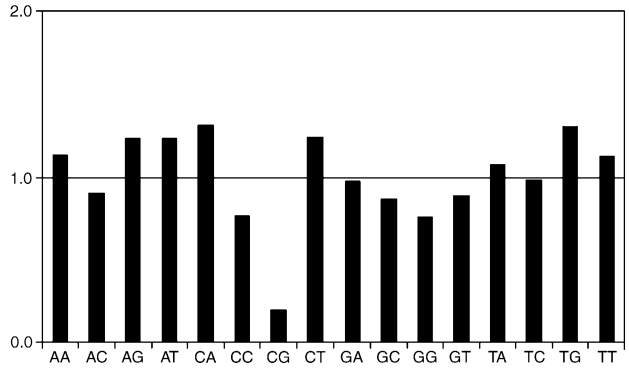

Fig. 3. Observed proportions of each possible dinucleotide recovered for ~2 Mb of sequence data from the genome of *A. mississippiensis*. The line at 1.0 represents the expected proportions if each dinucleotide was distributed randomly in the genome.

to an A appeared saturated (assuming the original sequence was CpG). This gave rise to several sites where the consensus was TpG or CpA when the original site may have been a CpG dinucleotide, and thus the actual number of CpG sites may have been underestimated. This suggestion is further supported by the observation that the frequency of CpG dinucleotides in the ~2 Mb of *A. mississippiensis* sequence data examined is similar to the observed human distribution of ~20% of the expected value (Fig. 3).

There were two exceptions to the general rule of higher CpG mutation rates. Comparisons between *chompy*I-i and *chompy*II consensus sequences and the representatives recovered ($n=6$ in both cases) revealed comparatively low CpG to non-CpG mutation rate ratios. This result may simply be due to sampling issues. However, it may also be indicative of recent insertion activity if *chompy*I-i elements have inserted so recently that the rate of CpG mutation has not yet overtaken the non-CpG mutation rate. Unfortunately, none of the loci ascertained by iPCR and amplifiable in multiple taxa belonged to the *chompy*I-i or *chompy*II families. Thus, further studies will be needed to fully explore this hypothesis.

Evidence suggests that MITEs are related in sequence to transposons and likely use the transposase encoded by these partner elements to mobilize (Feschotte et al., 2002). In addition, biochemical evidence of interactions between MITEs and *mariner*-like transposases was recently reported in rice (Feschotte et al., 2005). Thus, it appears that one potential next step in our studies of *chompy* elements will be to investigate the identity of the transposon partner for these putative MITEs. The identification of an active partner transposon will confirm the identity of these elements as MITEs and may also explain some of their differences when compared to other miniature inverted terminal repeat elements (e.g. non-conserved TSDs).

The characterization of *chompy* elements provides evidence that crocodilians are not an exception to the rule regarding the presence of interspersed repeats in eukaryotic genomes. This point is supported by a query of repeat-masker using the BAC sequences from *A. mississippiensis* which identified large numbers of CR1 elements (data not shown). Similarly, we subjected the *A. mississippiensis* BAC sequences to a BLAST search using MER consensus sequences and several were identified (data not shown). While our limited attempts to locate *chompy* loci that would be informative at the population and species levels were inconclusive, further work will be necessary to verify whether or not these MITEs have recently been active. In addition, CR1 repeats as well as other interspersed repeats may have had recent bursts of activity, and investigations of their activity should also be pursued. Future research should focus on all of these elements to determine whether additional and active repetitive element families are present in crocodilian genomes. In addition, identifying rates of genome-changing events such as non-homologous recombination, segmental duplications and chromosomal rearrangements that are associated with interspersed repeats in crocodilian genomes will either support or refute some of our speculations on the relationship between repetitive elements and genome stability in this order.

Acknowledgements

T. C. Glenn and G. Clark provided blood and/or DNA for this project. J. Xing and M. Konkel provided comments on early versions of the manuscript. This research was supported by the Louisiana Board of Regents Governor's Biotechnology Initiative GBI (2002-005) (M.A.B.), National Science Foundation EPS-0346411 (M.A.B.) the State of Louisiana Board of Regents Support Fund (M.A.B.), and National Geographic Grants 6529-99 and 7007-01 (L.D.D.).

References

Atkin, N.B., Mattinson, G., Becak, W., Ohno, S., 1965. The comparative DNA content of 19 species of placental mammals, reptiles and birds. Chromosoma 17, 1–10.

Bailey, J.A., Liu, G., Eichler, E.E., 2003. An Alu transposition model for the origin and expansion of human segmental duplications. Am. J. Hum. Genet. 73, 823–834.

Bao, Z., Eddy, S.R., 2002. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 12, 1269–1276.

Batzer, M.A., Deininger, P.L., 1991. A human-specific subfamily of Alu sequences. Genomics 9, 481–487.

Batzer, M.A., Deininger, P.L., 2002. Alu repeats and human genomic diversity. Nat. Rev., Genet. 3, 370–379.

Batzer, M.A., et al., 1994. African origin of human-specific polymorphic Alu insertions. Proc. Natl. Acad. Sci. U. S. A. 91, 12288–12292.

Baust, C., Baillie, G.J., Mager, D.L., 2002. Insertional polymorphisms of ETn retrotransposons include a disruption of the *wiz* gene in C57BL/6 mice. Mamm. Genome 13, 423–428.

Bestor, T.H., 1998. Gene silencing. Methylation meets acetylation. Nature 393, 311–312.

Bestor, T.H., Chandler, V.L., Feinberg, A.P., 1994. Epigenetic effects in eukaryotic gene expression. Dev. Genet. 15, 458–462.

Brochu, C.A., 2003. Phylogenetic approaches toward crocodilian history. Annu. Rev. Earth Planet. Sci. 31, 357–397.
Brochu, C.A., 2004. Calibration age and quartet divergence date estimation. Evolution 58, 1375–1382.
Brouha, B., et al., 2003. Hot L1s account for the bulk of retrotransposition in the human population. Proc. Natl. Acad. Sci. U. S. A. 100, 5280–5285.
Callinan, P.A., Wang, J., Herke, S.W., Garber, R.K., Liang, P., Batzer, M.A, 2005. Alu retrotransposition-mediated deletion. J. Mol. Biol. 348, 791–800.
Capriglione, T., Olmo, E., Odierna, G., Improta, B., Morescalchi, A., 1987. Cytofluorometric DNA-base determination in vertebrate species with different genome sizes. Basic Appl. Histochem. 31, 119–126.
Carter, A.B., et al., 2004. Genome wide analysis of the human Alu Yb lineage. Hum. Genomics 1, 167–178.
Cohen, M.M., Gans, C., 1970. The chromosomes of the order Crocodylia. Cytogenetics 9, 81–105.
Coulondre, C., Miller, J.H., Farabaugh, P.J., Gilbert, W., 1978. Molecular basis of base substitution hotspots in *Escherichia coli*. Nature 274, 775–780.
Deininger, P.L., Batzer, M.A., 1999. Alu repeats and human disease. Mol. Genet. Metab. 67, 183–193.
Deininger, P.L., Roy-Engel, A.M., 2002. Mobile elements in animal and plant genomes. In: Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A.M. (Eds.), Mobile DNA II. ASM Press, Washington, DC.
De Smet, W.H.O., 1981. The nuclear Feulgen-DNA content of the vertebrates (especially reptiles), as measured by fluorescence cytophotometry, with notes on the cell and chromosome size. Acta Zool. Pathol. Antverp. 76, 119–167.
Ejima, Y., Yang, L., 2003. Trans mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling. Hum. Mol. Genet. 12, 1321–1328.
Feschotte, C., Mouches, C., 2000. Recent amplification of miniature inverted-repeat transposable elements in the vector mosquito *Culex pipiens*: characterization of the Mimo family. Gene 250, 109–116.
Feschotte, C., Zhang, X., Wessler, S.R., 2002. Miniature inverted-repeat transposable elements and their relationship to established DNA transposons. In: Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A.M. (Eds.), Mobile DNA II. ASM Press, Washington, DC, pp. 1147–1158.
Feschotte, C., Osterlund, M.T., Peeler, R., Wessler, S.R., 2005. DNA-binding specificity of rice mariner-like transposases and interactions with stowaway MITEs. Nucleic Acids Res. 33, 2153–2165.
Fischer, S.E., Wienholds, E., Plasterk, R.H., 2003. Continuous exchange of sequence information between dispersed Tc1 transposons in the *Caenorhabditis elegans* genome. Genetics 164, 127–134.
Fitzsimmons, N.N., et al., 2002. Identification of purebred *Crocodylus siamensis* for reintroduction in Vietnam. J. Exp. Zool. 294, 373–381.
Frey, E., Riess, J., Tarsitano, S.F., 1989. The axial tail musculature of recent crocodiles and it phyletic implications. Am. Zool. 29, 857–862.
Gaffney, P.M., Pierce, J.C., Mackinley, A.G., Titchen, D.A., Glenn, W.K., 2003. Pearl, a novel family of putative transposable elements in bivalve mollusks. J. Mol. Evol. 56, 308–316.
Gatesy, J., Amato, G., Norell, M., DeSalle, R., Hayashi, C., 2003. Combined support for wholesale taxic atavism in gavialine crocodilians. Syst. Biol. 52, 403–422.
Gebow, D., Miselis, N., Liber, H.L., 2000. Homologous and nonhomologous recombination resulting in deletion: effects of p53 status, microhomology, and repetitive DNA length and orientation. Mol. Cell. Biol. 20, 4028–4035.
Gilbert, N., Lutz-Prigge, S., Moran, J.V., 2002. Genomic deletions created upon LINE-1 retrotransposition. Cell 110, 315–325.
Goodier, J.L., Ostertag, E.M., Kazazian Jr., H.H., 2000. Transduction of 3′-flanking sequences is common in L1 retrotransposition. Hum. Mol. Genet. 9, 653–657.
Gray, Y.H., 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet. 16, 461–468.
Gregory, T.R., 2005. Animal Genome Size Database.
Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
Hendrich, B., Tweedie, S., 2003. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet. 19, 269–277.
Izsvak, Z., Ivics, Z., Shimoda, N., Mohn, D., Okamoto, H., Hackett, P.B., 1999. Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. J. Mol. Evol. 48, 13–21.
Jiang, N., Bao, Z., Zhang, X., Eddy, S.R., Wessler, S.R., 2004. Pack-MULE transposable elements mediate gene evolution in plants. Nature 431, 569–573.
Kass, D.H., Batzer, M.A., Deininger, P.L., 1995. Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. Mol. Cell. Biol. 15, 19–25.
Kazazian Jr., H.H., Goodier, J.L., 2002. LINE drive. Retrotransposition and genome instability. Cell 110, 277–280.
Kazazian Jr., H.H., Moran, J.V., 1998. The impact of L1 retrotransposons on the human genome. Nat. Genet. 19, 19–24.
Kent, W.J., et al., 2002. The human genome browser at UCSC. Genome Res. 12, 996–1006.
Kidwell, M.G., Lisch, D.R., 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 55, 1–24.
Labrador, M., Corces, V.G., 2002. Interactions between transposable elements and the host genome. In: Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A.M. (Eds.), Mobile DNA II. ASM Press, Washington, DC.
Lander, E.S., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.
Langston, W., 1973. The crocodilian skull in historical perspective. In: Gans, C., Parsons, T.H. (Eds.), Biology of the Reptilia. Academic Press, New York, pp. 263–284.
Lepetit, D., Pasquet, S., Olive, M., Theze, N., Thiebaud, P., 2000. Glider and Vision: two new families of miniature inverted-repeat transposable elements in *Xenopus laevis* genome. Genetica 108, 163–169.
Lovsin, N., Gubensek, F., Kordi, D., 2001. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia. Mol. Biol. Evol. 18, 2213–2224.
Mayr, E., 1942. Systematics and the Origin of Species. Columbia University Press, New York.
Mirsky, A.E., Ris, H., 1951. The desoxyribonucleic acid content of animal cells and its evolutionary significance. J. Gen. Physiol. 34, 451–462.
Morgan, G.T., 1995. Identification in the human genome of mobile elements spread by DNA-mediated transposition. J. Mol. Biol. 254, 1–5.
Nouaud, D., Quesneville, H., Anxolabehere, D., 2003. Recurrent exon shuffling between distant P-element families. Mol. Biol. Evol. 20, 190–199.
Ohshima, K., Hamada, M., Terai, Y., Okada, N., 1996. The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. Mol. Cell. Biol. 16, 3756–3764.
Olmo, E., 1976. Genome size in some reptiles. J. Exp. Zool. 195, 305–310.
Olmo, E., 1981. Evolution of genome size and DNA base composition in reptiles. Genetica 57, 39–50.
Oosumi, T., Garlick, B., Belknap, W.R., 1995. Identification and characterization of putative transposable DNA elements in solanaceous plants and *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U. S. A. 92, 8886–8890.
Oosumi, T., Garlick, B., Belknap, W.R., 1996. Identification of putative nonautonomous transposable elements associated with several transposon families in *Caenorhabditis elegans*. J. Mol. Evol. 43, 11–18.
Ostertag, E.M., Goodier, J.L., Zhang, Y., Kazazian Jr., H.H., 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. Am. J. Hum. Genet. 73, 1444–1451.

Otieno, A.C., et al., 2004. Analysis of the human Alu Ya-lineage. J. Mol. Biol. 342, 109–118.

Paulsen, M., Ferguson-Smith, A.C., 2005. DNA methylation in genomic imprinting, development, and disease. J. Pathol. 195, 97–110.

Perna, N.T., Batzer, M.A., Deininger, P.L., Stoneking, M., 1992. Alu insertion polymorphism: a new type of marker for human population studies. Hum. Biol. 64, 641–648.

Pickeral, O.K., Makalowski, W., Boguski, M.S., Boeke, J.D., 2000. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res. 10, 411–415.

Ramos, R., de Buffrenil, V., Ross, J.P., 1994. Current status of the Cuban crocodile, *Crocodylus rhombifer*, in the wild. 12th Working Meeting of the Crocodile Specialist Group. IUCN, pp. 113–140.

Ray, D.A., et al., 2004. Low levels of nucleotide diversity in *Crocodylus moreletii* and evidence of hybridization with *C. acutus*. Conserv. Genet. 5, 449–462.

Ray, D.A., et al., 2005. Alu insertion polymorphisms and platyrrhine primate phylogenetic relationships. Mol. Phylogenet. Evol. 35, 117–126.

Roos, C., Schmitz, J., Zischler, H., 2004. Primate jumping genes elucidate strepsirrhine phylogeny. Proc. Natl. Acad. Sci. U. S. A. 101, 10650–10664.

Ross, J.P. Crocodiles, 1998. Status Survey and Conservation Action Plan. 2nd Edition. IUCN Crocodile Specialist Group, Gland, Switzerland, pp. viii+96 pp.

Roy, A.M., et al., 1999. Recently integrated human Alu repeats: finding needles in the haystack. Genetica 107, 149–161.

Roy-Engel, A.M., et al., 2002. Non-traditional Alu evolution and primate genomic diversity. J. Mol. Biol. 316, 1033–1040.

Rozen, S., Skaletsky, H.J., 1998. Primer3. (http://www-genome.wi.mit.edu/genome_software/other/primer3.html).

Salem, A.H., Kilroy, G.E., Watkins, W.S., Jorde, L.B., Batzer, M.A., 2003a. Recently integrated Alu elements and human genomic diversity. Mol. Biol. Evol. 20, 1349–1361.

Salem, A.H., et al., 2003b. Alu elements and hominid phylogenetics. Proc. Natl. Acad. Sci. U. S. A. 100, 12787–12791.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467.

Schmid, C.W., 1998. Does SINE evolution preclude Alu function? Nucleic Acids Res. 26, 4541–4550.

Shen, M.R., Batzer, M.A., Deininger, P.L., 1991. Evolution of the master Alu gene(s). J. Mol. Evol. 33, 311–320.

Singer, S.S., Schmitz, J., Schwiegk, C., Zischler, H., 2003. Molecular cladistic markers in New World monkey phylogeny (Platyrrhini, Primates). Mol. Phylogenet. Evol. 26, 490–501.

Smit, A.F., Riggs, A.D., 1996. Triggers and DNA transposon fossils in the human genome. Proc. Natl. Acad. Sci. U. S. A. 93, 1443–1448.

Stoneking, M., et al., 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res. 7, 1061–1071.

Symer, D.E., et al., 2002. Human l1 retrotransposition is associated with genetic instability in vivo. Cell 110, 327–338.

Tu, Z., 1997. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. Proc. Natl. Acad. Sci. U. S. A. 94, 7475–7480.

Tu, Z., 2001. Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. Proc. Natl. Acad. Sci. U. S. A. 98, 1699–1704.

Tweedie, S., Charlton, J., Clark, V., Bird, A., 1997. Methylation of genomes and genes at the invertebrate–vertebrate boundary. Mol. Cell. Biol. 17, 1469–1475.

Waterston, R.H., et al., 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562.

Watkins, W.S., et al., 2003. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. Genome Res. 13, 1607–1618.

Xing, J.C., et al., 2003. Comprehensive analysis of two Alu Yd subfamilies. J. Mol. Evol. 57, S76–S89.

Xing, J., Hedges, D.J., Han, K., Wang, H., Cordaux, R., Batzer, M.A., 2004. Alu element mutation spectra: molecular clocks and the effect of DNA methylation. J. Mol. Biol. 344, 675–682.

Xing, J., et al., in press. A mobile element based phylogeny of old world monkeys. Molecular Phylogenetics and Evolution.

Yeadon, P.J., Catcheside, D.E., 1995. Guest: a 98 bp inverted repeat transposable element in *Neurospora crassa*. Mol. Gen. Genet. 247, 105–109.