

Discovery of Highly Divergent Repeat Landscapes in Snake Genomes Using High-Throughput Sequencing

Todd A. Castoe¹, Kathryn T. Hall¹, Marcel L. Guibotsy Mboulas², Wanjun Gu†¹, A.P. Jason de Koning¹, Samuel E. Fox³, Alexander W. Poole¹, Vijetha Vemulapalli¹, Juan M. Daza⁴, Todd Mockler⁴, Eric N. Smith², Cédric Feschotte², and David D. Pollock*¹

¹Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine

²Department of Biology, University of Texas at Arlington

³Department of Botany and Plant Pathology, Oregon State University

⁴Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

†Present address: Key Laboratory of Child Development and Learning Science, Southeast University, Ministry of Education, Nanjing, People's Republic of China

*Corresponding author: E-mail: david.pollock@ucdenver.edu.

Accepted: 27 April 2011

Abstract

We conducted a comprehensive assessment of genomic repeat content in two snake genomes, the venomous copperhead (*Agkistrodon contortrix*) and the Burmese python (*Python molurus bivittatus*). These two genomes are both relatively small (~1.4 Gb) but have surprisingly extensive differences in the abundance and expansion histories of their repeat elements. In the python, the readily identifiable repeat element content is low (21%), similar to bird genomes, whereas that of the copperhead is higher (45%), similar to mammalian genomes. The copperhead's greater repeat content arises from the recent expansion of many different microsatellites and transposable element (TE) families, and the copperhead had 23-fold greater levels of TE-related transcripts than the python. This suggests the possibility that greater TE activity in the copperhead is ongoing. Expansion of CR1 LINEs in the copperhead genome has resulted in TE-mediated microsatellite expansion ("microsatellite seeding") at a scale several orders of magnitude greater than previously observed in vertebrates. Snakes also appear to be prone to horizontal transfer of TEs, particularly in the copperhead lineage. The reason that the copperhead has such a small genome in the face of so much recent expansion of repeat elements remains an open question, although selective pressure related to extreme metabolic performance is an obvious candidate. TE activity can affect gene regulation as well as rates of recombination and gene duplication, and it is therefore possible that TE activity played a role in the evolution of major adaptations in snakes; some evidence suggests this may include the evolution of venom repertoires.

Key words: Burmese python, copperhead, microsatellite seeding, non-avian reptile comparative genomics, transposable elements.

Introduction

Among vertebrates, the snake lineage represents an impressively speciose (~3100 sp.) and phenotypically diverse radiation, and as a result, snakes have become increasingly important model systems for diverse research areas. Snakes provide a unique model system for studying extreme physiological remodeling and metabolic cycling (Secor and Diamond 1995, 1998) and in venom-related research (Fry et al. 2006; Ikeda et al. 2010). Snakes have also become important models for developmental biology,

evolutionary ecology, and molecular evolution and adaptation (Cohn and Tickle 1999; Fry et al. 2006; Castoe et al. 2008, 2009b; Vonk et al. 2008). Despite the importance of snakes as models for basic and biomedical research, there is little known about the genomes of snakes and about reptile genomes in general (Shedlock et al. 2007; Janes et al. 2010).

Our aim here was thus to obtain comprehensive sequence-based comparative insight into snake genomic diversity, particularly the diversity and structure of the

© The Author(s) 2011. Published by Oxford University Press on behalf of the *Society for Molecular Biology and Evolution*.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

repetitive element landscape. Such insight is important because repetitive elements comprise major portions of vertebrate genomes and exert major influences over genome evolution. Among repetitive sequences, transposable elements (TEs) in particular have had a tremendous impact on the structural and functional evolution of genes and genomes. Numerous studies have documented how TE activity and ectopic recombination between TE copies promote small- and large-scale variation in the structure of genomes; such rearrangements provide a substrate for the emergence of new functional sequences, both coding and noncoding, including the birth of new protein-coding genes and the rewiring of regulatory networks (Feschotte 2008; Cordaux and Batzer 2009; Herpin et al. 2010). Despite a recent comprehensive review summarizing current knowledge about reptilian TEs (Kordis 2009), our understanding of vertebrate TE diversity and evolutionary dynamics remains largely dominated by perspectives from mammalian and to a lesser extent avian genomes.

The speciose nature and evolutionary age of the snake radiation make it an excellent amniote lineage for comparisons to mammals. Snakes and mammals share a common ancestor ~310 Ma and snakes diverged from other squamate reptiles about ~170 Ma (Castoe et al. 2009a), which slightly predates the estimated split of eutherian (placental) and metatherian (marsupial) mammals. In this study, we chose to focus on two fairly distantly related snake species, the Burmese python (*Python molurus bivittatus*) and the copperhead (*Agkistrodon contortrix*)—these two lineages share a common ancestor at about the same time as do all eutherian mammals, around 100 Ma (Castoe et al. 2009a). In comparison to mammalian genomes, snake genomes are generally small (Gregory et al. 2007), ranging from 1.3 to 3.8 Gbp and averaging 2.1 Gbp, and the two snakes chosen both have similarly sized genomes, on the small side of this range (~1.4 Gbp). Evidence from early DNA reassociation studies suggests that there may be extensive variation in the genomic repeat landscapes among snake species, particularly between pythons and colubroid species such as the copperhead (Olmo et al. 1981; Olmo 1984). These two snakes also represent important lineages for research. The Burmese python (*P. molurus bivittatus*) is an important model for physiological and metabolic adaptation and the copperhead (*A. contortrix*) is a model for metabolic adaptation and a viperid model for studies related to venom. Although distantly related, these two lineages (pythons and viperids) have convergently evolved extremely dynamic metabolisms to facilitate the infrequent consumption of large prey (Secor and Diamond 2000).

To gain insight into the repeat landscapes of these two genomes and the evolutionary processes that have shaped them, we obtained low-coverage 454 high-throughput sequencing data from genomic shotgun libraries, as well as 454 transcriptome sequence showing evidence of TE transcriptional activity

in both species. Using this data, we analyzed TE and simple sequence repeat (SSR) content, diversity, and origins. The results reveal extraordinary differences between these two snakes. They also contribute to a broader understanding of vertebrate genome evolution and diversity by beginning to show how snake genomes compare to one another and to other vertebrates.

Materials and Methods

Shotgun Library Creation and Sequencing

Whole genome random shotgun libraries were made from two snake species, *A. contortrix* (the copperhead; from a Texas population) and *P. molurus bivittatus* (the Burmese python; obtained from the pet trade); animals and tissues were procured and processed through the Amphibian and Reptile Diversity Research Center at The University of Texas Arlington. Total DNA was prepared from liquid-nitrogen snap-frozen liver tissue by standard phenol–chloroform–isoamyl alcohol extraction methods. 454 FLX-LR and 454 Titanium-XLR genomic shotgun libraries were prepared using the 454 shotgun library preparation kit and protocol (Roche). Libraries were sequenced on the Roche 454 FLX sequencing platform. From the *Agkistrodon* FLX-LR shotgun library, 60.3 Mb from 280,303 sequence reads were collected using Roche/454 FLX-LR sequencing kits, amounting to about 4.5% of the estimated 1.35 Gbp (Gregory et al. 2007) genome (supplementary tables S1, Supplementary Material online). Two libraries from the same individual (one FLX and one Titanium) were sequenced for the *Python*; from the FLX-LR library, we sequenced 61,256 reads, totaling 13.3 Mbp, and from the Titanium-XLR library, we sequenced 57,717 reads, totaling 15.2 Mbp. In sum, 28.5 Mbp of *Python* sequence from 118,973 reads were collected, representing ~2.0% of its estimated 1.42 Gbp genome (based on estimates of the related *Python reticulatus* genome, supplementary tables S1, Supplementary Material online). Comparisons of repeat annotations of FLX-LR and FLX-XLR data for *Python* indicated extremely little difference between the two data types (supplementary fig. S10, Supplementary Material online). Sequence reads with similarity to the mitochondrial genome were filtered out prior to analyses (see supplementary methods, Supplementary Material online).

Repeat Analyses

We used the current release of the Tetrapoda *RepBase* (version 12.12, 17 January 2008) as the repeat library for *RepeatMasker* (Smit et al. 2010) to identify known repeat elements in the snake genomes. For comparisons, we also ran *RepeatMasker* on ~50 Mbp of the *Anolis* genome (AnoCar1.0) from four combined genome scaffolds. For SSR analysis, we used a previously written Perl script (Castoe et al. 2010) that was modified to identify SSR loci

with repeated sequence motifs of 2–6 bases in length and a minimum of 12 bases in length (for 2–4mers) or containing 3 or more tandem repeats (for 5–6mers). We used the program *RepeatModeler* (Smit A, unpublished data) to identify de novo repeat sequences in our snake data sets, based on the run parameters suggested as defaults by the program. The approach essentially couples two de novo repeat finding methods, *RECON* (Bao and Eddy 2002) and *RepeatScout* (Price et al. 2005), together with Tandem Repeat Finder (Benson 1999). We modified *RepeatModeler's RepeatMasker* parameters to specify the Tetrapoda library. For all *RepeatModeler* analyses, we combined the new *Python* and *Agkistrodon* libraries into a single joint snake library to recover as many elements as possible and control for differences in sequencing depth. Consensus sequences from *RepeatModeler* were classified using *RepClass* (Feschotte et al. 2009). By identifying novel “families” that hit the same known TE family, we were able to reduce the original count of new family consensus sequences by 18.6% and 20.3% in *Python* and *Agkistrodon*, respectively (see supplementary methods and [supplementary tables S2 and S3, Supplementary Material](#) online). We also used the program *P-clouds* for identifying de novo repeats, with the following parameter settings: 2, 3, 6, 12, and 24 for low, core, step-1, step-2, and step-3 cutoffs (Gu et al. 2008). Details provided in supplementary methods ([Supplementary Material](#) online).

Consensus sequences for select TEs were assembled in an iterative fashion, using successive rounds of mapping reads to existing element sequences to extend element alignments and then taking the consensus of these alignments. Mapping was conducted using *gsmapper* software (Roche) and used to estimate read coverage across the length of elements. For estimation of sequence divergence from element consensus sequences, sequences were aligned using POA (Lee et al. 2002). In the case of LINEs, for which we expected multiple subfamilies to exist, sequence divergence was estimated excluding sites inferred to define element subfamilies (see supplementary methods, [Supplementary Material](#) online), similar to (Aleshin and Zhi 2010).

cDNA Sequencing and Analysis

RNA was extracted, poly-A enriched, and cDNA libraries were prepared from *A. contortrix* and *P. molurus* liver tissue samples using standard techniques. Libraries were bidirectionally sequenced using the FLX-LR reagents on the 454 FLX instrument. All steps were carried out on both samples, side-by-side, from RNA extraction through sequencing. Transcript contigs were assembled using the 454 *GSA assembler*, and we searched for TE sequences using our snake-specific libraries in *RepeatMasker*. Details provided in supplementary methods ([Supplementary Material](#) online).

Results

The 60.3 Mbp sequenced from the *A. contortrix* (copperhead hereafter, for readability in the text) random shotgun sequence library amounts to about 4.5% of its estimated 1.35 Gbp genome, whereas the 28.5 Mbp sequenced from the *P. molurus* (python hereafter) represents about 2.0% of its estimated 1.42 Gbp genome ([supplementary table S1, Supplementary Material](#) online; NCBI Sequence Read Archive accession SRA029568.1; also available at www.snakegenomics.org/SnakeGenomics/RawData.html). Distributions of base call quality scores are very similar for both species, facilitating direct comparisons between the data for both species ([supplementary fig. S1, Supplementary Material](#) online). To provide a baseline estimate for expected levels of neutral sequence divergence among the python, copperhead, and anole, we analyzed a previously published 12 nuclear protein-coding gene data set (Wiens et al. 2010). We estimated an average synonymous divergence (i.e., dS branch lengths; see supplementary methods, [Supplementary Material](#) online) of 0.709 substitutions/site between the anole and the snakes and 0.221 substitutions/site between the python and copperhead.

To obtain a preliminary understanding of repetitive content in these genomes, we first analyzed the frequencies of 15mers in 28 Mbp samples from both genomes (consisting of a random 28 Mbp sample from the copperhead and all of the python data). We chose 15mers because, by chance, any one 15mer should occur only about once in a genome of this size, making high-copy 15mers extremely infrequent by chance and thus indicative of repetitive elements (Gu et al. 2008). Both species contained similar amounts of single copy 15mers (15,364,028 in the python and 17,186,377 in the copperhead), but the python had more low-to-moderate copy number 15mers (i.e., 2–10 copies; [fig. 1A](#)). In contrast, the copperhead had more high-copy 15mers ([fig. 1A](#)), suggesting that it has more highly similar (recently expanded) repeat elements, whereas python repeat elements are comparatively older and/or fewer in number. Analysis of the anole lizard (*Anolis carolinensis*) genome revealed a 15mer profile similar to the copperhead ([fig. 1](#)), suggesting that it too has a substantial number of recently expanded repeats, and thus a repetitive landscape more similar to the copperhead than the python.

Identification of Interspersed and Tandem Repeats

The common method of identifying recognizable repeat elements is to scan sequences using a library of known repeat element consensus sequences (e.g., *RepBase*; Jurka 2000) with sequence similarity-based algorithms such as *RepeatMasker* (Smit et al. 2010). Such homology-based methods cannot recognize elements not in the database and have low power to identify repeat element fragments, even up to 200 bp, from moderately diverged repeat families (de

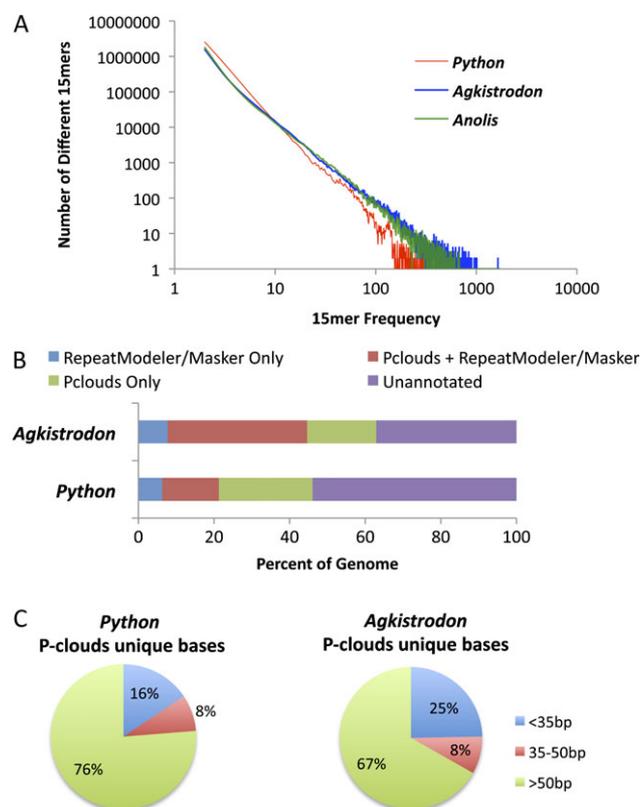


Fig. 1.—Comparison of repeat analyses. (A) The frequency of each different 15mer sequence was counted, and the number shown is the number of different 15mers having a particular count. Equal size samples of the genomes of the two snakes, copperhead (*Agkistrodon*) and python (*Python*), and the lizard (*Anolis*) were considered. SSRs were removed from the analyzed data using *RepeatMasker/RepBase*. (B) The repeat annotation methods (*P-clouds* and *RepeatModeler/RepeatMasker*) in the two snakes were compared with determine the percent of the genome (in nucleotides) that was masked by either method alone, both methods, or neither method (i.e., remained unannotated). (C) The size distribution of *P-clouds* results in the two snakes are shown.

Koning APJ, Gu W, Castoe TA, Pollock DD, unpublished data). For this reason, the identification of interspersed and tandem repeats in snakes is problematic because there are no closely related organisms in which repeat elements have been studied in-depth, and thus there are no snake-specific repeat libraries available. Because unassembled next-generation sequence reads likely contain many (perhaps mostly) TE sequence fragments, we utilized a three-pronged strategy to evaluate repeat content and maximize detection and classification of repeat elements: first, we applied the *P-clouds* method, which can estimate the repeat fraction without knowing a priori what the repeat elements are and which is relatively powerful at identifying short repeat fragments (Gu et al. 2008); second, we utilized the “Tetrapoda” repeat consensus library in *RepBase* to detect similar sequences by homology searching; and third, we used *RepeatModeler* (Smith A, unpublished data) to identify

new snake-specific repeat element clusters (families). We analyzed the output of *RepeatModeler* with *RepClass* (Feschotte et al. 2009), a tool that automates the classification of newly discovered TEs (supplementary figs. S2 and S3, Supplementary Material online). Repeat family consensus sequence libraries identified are available at www.snakegenomics.org/SnakeGenomics/Processed_Data.html.

Previous analyses using the *P-clouds* method have estimated genomic repetitive content in repeat-poor bird genomes at around 40% (Warren et al. 2010), in more repeat-rich genomes of the *Anolis* lizard at 73.7% (Warren et al. 2010), and in the human and panda genomes at about 70% (Li et al. 2010). In comparison, the *P-clouds* estimate of the repetitive content of the python genome is 39.8%, similar to bird genomes, whereas the predicted copperhead repetitive content is 55.2%, intermediate between birds and mammals/lizards (supplementary table S1, Supplementary Material online).

Only 4.48% of the python and 11.81% of the copperhead (supplementary table S1, Supplementary Material online) were identified as readily classifiable repeats (SSRs, tandem repeats, low-complexity sequences, and known TEs from *RepBase*). An additional 16.73% of the python and 32.77% of the copperhead were identified as repeat elements using the newly identified snake-specific repeat library from *RepeatModeler* (fig. 1B and supplementary figs. S2 and S3, Supplementary Material online), and about one-third of these new families identified in both the copperhead and the python were classified by *RepClass* into known TE classes (655/1996 and 203/571, respectively; supplementary tables S2–S7, Supplementary Material online). It is important to note that the same snake-specific repeat library was used to annotate both species and was derived from running *RepeatModeler* on data from each species independently and then combining the resulting libraries; this approach was designed to increase sensitivity and decrease bias due to different amounts of genomic sampling in the two species.

All methods agree that the copperhead has considerably more detectable repetitive sequence than the python, and a large part of this repetitive fraction in both species arises from recognizable TEs. Altogether about 45% of the copperhead was annotated by the homology-based methods compared with only 21% of the python genome; in contrast, the two estimates from *P-clouds* are 55% and 40%, respectively (fig. 1B and supplementary table S1, Supplementary Material online). We expect *P-clouds* to be more sensitive than homology-based methods for identifying more divergent and fragmented repeat elements, and the observation that the estimated python repetitive content is nearly twice as high based on *P-clouds* analysis indicates that much of its repetitive content may be older and/or more fragmented than that of the copperhead. There is substantial overlap between the methods (fig. 1B),

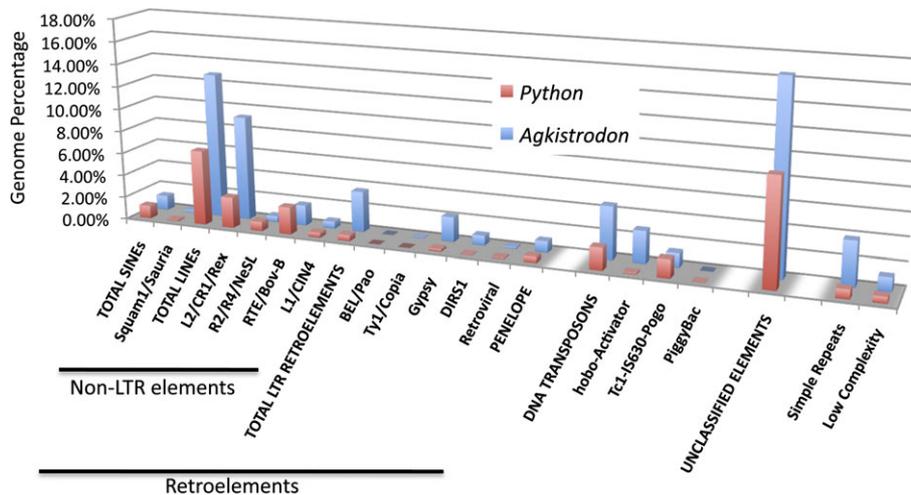


FIG. 2.—Comparison of the TE and simple repeat content in copperhead (*Agkistrodon*) and python (*Python*) genomes. TE families were determined based on the combined annotations of *Repbase*, *RepeatScout*, *RepeatModeler*, and *RepClass*, and coverage in the genome was annotated using *RepeatMasker*.

and most (~67–76%) of the regions uniquely annotated by *P-clouds* are fairly long (>50 bp; fig. 1C). The *P-clouds*-unique sequences may also include non-TE-derived sequences such as tandem duplications or large multigene families but given the level of sampling (<5% of each genome) and the makeup of known complete genomes, we expect that these types of repeats make up a fairly small percentage of the *P-clouds* annotation. These results support the conclusion from the 15mer profile analysis that the copperhead has many homogeneous TE sequences compared with the more diverged and/or lower frequency repeats in the python.

TE Landscapes

The joint annotation of previously known *RepBase* elements, together with newly identified elements classified using *RepClass* (Feschotte et al. 2009), revealed a substantial diversity of TEs in snake genomes (fig. 2 and supplementary fig. S1, Supplementary Material online). Almost all types of repeats appear to occur more frequently in the copperhead than the python (fig. 2 and supplementary tables S6 and S7, Supplementary Material online), but the breadth of diversity in each of the snakes was similar, with most subclasses and superfamilies (e.g., Bov-B LINES, DIRS1) found in the copperhead are also represented in the python (fig. 2). Although repetitive elements in the *Anolis* lizard genome have not yet been thoroughly annotated, the diversity of repeat types observed in the snakes was broadly similar to preliminary estimates of the diversity of repeat elements in the anole lizard genome (supplementary fig. S4 and table S8, Supplementary Material online). Comparing the two snakes, there are substantial differences in TE abundance (fig. 2; supplementary table S8, Supplementary Material online). The

greatest difference lies in the abundance of CR1 LINES that are more than four times as frequent in the copperhead, in contrast to Bov-B LINES, which are similarly abundant in both genomes (fig. 2).

For in-depth analysis of snake LINES, we manually assembled consensus sequences of Bov-B and CR1 LINES for each species (now available in *RepBase*). Analysis of the distribution of sequence divergence (from species-specific consensus sequences, excluding subfamily-defining sites) within these two LINE superfamilies reveals contrasting expansion histories both between LINES and between species (fig. 3A and B). Both snake lineages experienced recent (and likely independent) expansion of Bov-B LINES, indicated by the low sequence divergence among these LINES within each species (fig. 3A). The bulk of divergence within python Bov-B LINES appears to have occurred slightly more recently than for the copperhead. In contrast, CR1 accumulation appears to have occurred over an extended period in both lineages and likely in the ancestor of both of these species. We estimated 11% neutral divergence at synonymous sites of protein-coding genes above (0.221 substitutions per site pairwise difference between species, divided by 2). The finding that a notable proportion of CR1 elements in the snakes exceed 11% divergence suggests that CR1s have been active over a long time period in snake genomes, including being active in the ancestor of these two snakes. The extended time period of CR1 activity in both snakes, followed by a recent decrease in activity, contrasts sharply with the timing of Bov-B activity, which shows predominantly recent activity in both snakes (fig. 3). Also, despite a similar age distribution of CR1 elements in both snakes, the copperhead lineage accumulated many more CR1s than the python lineage in every time period (fig. 3B).

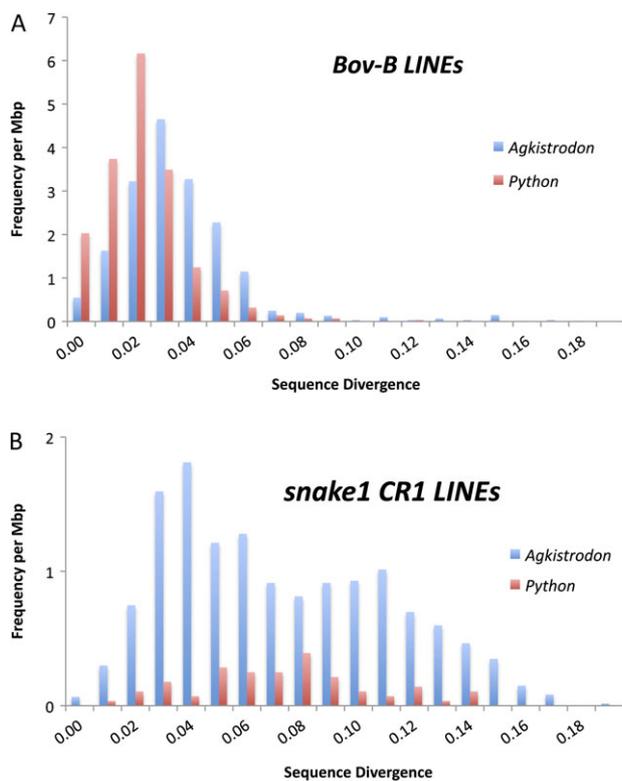


FIG. 3.—Sequence divergence of selected TEs. The species-specific consensus sequences were determined for (A) Bov-B LINEs and (B) CR1 LINEs, and the sequence divergence levels were calculated for all alignable sequences of these types, excluding sites that appeared to define subfamilies.

Evidence from LINEs suggests that there may be substantially different genomic processes at work on the two snake genomes that results in truncation and possible purging of longer repeat elements along the copperhead lineage. Whereas about half of Bov-B LINEs appear near full length in the python genome based on sequence coverage in reads mapped to the python Bov-B LINE consensus sequence (fig. 4A and supplementary fig. S10A, Supplementary Material online), a vast majority of Bov-B LINEs appear to be truncated in the copperhead genome (in which the 3-prime end of Bov-B is vastly overrepresented). Copperhead CR1 LINEs are truncated similarly to copperhead Bov-B LINEs, with the last 600–800 bp of the element greatly overrepresented in the sampled sequences, although the low copy number of CR1 LINEs in the python prevents meaningful comparisons between species (fig. 4B and supplementary fig. S10B, Supplementary Material online).

In addition to a greater abundance of LINEs, the copperhead also has a much greater abundance of both Gypsy-like (2.18% vs. 0.21%) and DIRS (0.84% vs. 0.03%) retrotransposons compared with the python; these families are also both more abundant in the python than in the anole lizard (fig. 2 and supplementary table S8, Supplementary Material

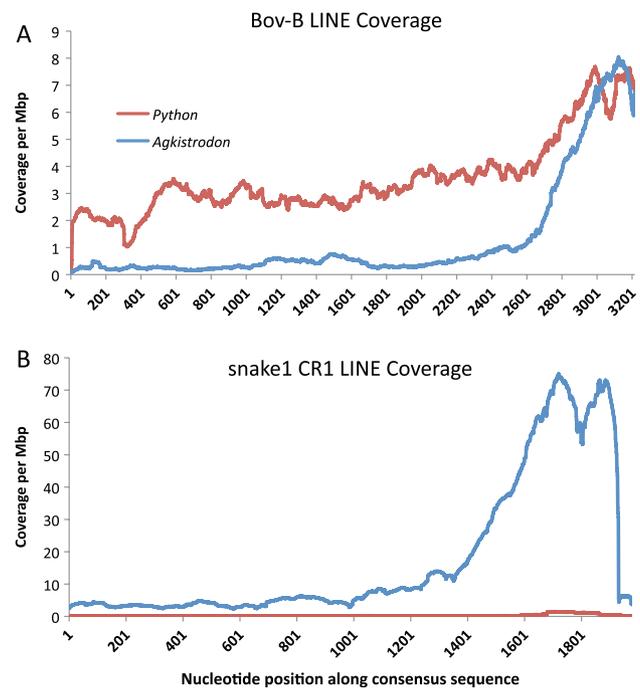


FIG. 4.—Depth of sequence coverage. The coverage depth per megabase of genomic sequence is shown for the 3' ends of (A) Bov-B and (B) snake1 CR1 LINEs for both the copperhead (*Agkistrodon*) and the python (*Python*) genome samples.

online). An increased abundance of DNA transposons in the copperhead (4.58% vs. 1.92% in the python) is also observed, primarily due to increases in hobo-Activator-Tam3 (hAT) transposons (fig. 2). The copperhead also experienced a notable expansion of SSRs and low-complexity regions relative to the python and lizard and contains more unclassified elements than the python (16.45% vs. 9.29%). The rarity of identified SINES may be an artifact because SINES are often difficult to classify due to their short length, rapid evolution, and turnover and because they do not encode proteins.

At the family level, over three times more new snake-specific families were identified (using *RepeatModeler*) in the copperhead (1,996) than the python (571). This de novo repeat identification method produces many potentially redundant family descriptions, but after collapsing redundant families (see supplementary methods, Supplementary Material online), the result holds; only 82 new collapsed families from the python were identified by *RepClass* compared with 243 in the copperhead (supplementary tables S2–S5, Supplementary Material online). Many more element families were identified in the copperhead compared with the python for numerous element types (supplementary table S4, Supplementary Material online), including CR1 and L1 LINEs, *penelope* retrotransposons, gypsy and DIRS retrotransposons, and hobo-Activator (hAt) and Mariner DNA transposons (supplementary table S4, Supplementary Material online). The familial diversity of DNA transposons (16

new python families, 48 new copperhead families), DIRS (0 new python families, 38 new copperhead families), and gypsy (8 new python families, 49 new copperhead families) are particularly skewed (supplementary table S4, Supplementary Material online). This indicates that the greater TE content in the copperhead compared with the python is based not only on greater numbers of elements but also on greater element diversity at the more fine-scale family level. Higher element abundance in the copperhead was not limited to a particular set of elements but rather distributed across a diverse set of elements. Furthermore, per element type, there tend to be more new element families/subfamilies identified in the copperhead; this is especially the case for more frequent element types.

Evidence for Horizontal Transfer of TEs

Previous studies have inferred horizontal transfer of Bov-B LINEs between mammals and snakes and/or squamate reptiles to explain the enigmatic distribution of these elements across amniote vertebrates (Kordis and Gubensek 1997, 1998b). Based on phylogenetic analysis of Bov-B sequences from available vertebrate genomes, we estimate that copperhead and anole Bov-B sequences are more closely related to each other than either sequence is to the python Bov-B sequences. In terms of organismal phylogeny, snakes are uniformly believed to form a monophyletic group to the exclusion of lizards (Townsend et al. 2004; Vidal and Hedges 2005; Castoe et al. 2009b; Wiens et al. 2010). Therefore, our inference that there are multiple lineages of snake Bov-B elements implies multiple episodes of horizontal transfer of Bov-B LINEs to or from squamate reptiles (supplementary fig. S5, Supplementary Material online). We also found traces of Maverick DNA transposons only in the python (supplementary tables S6 and S7, Supplementary Material online), and these are not otherwise known from squamate reptiles, although there is a report of one from the sister lineage of squamates, the tuatara (Pritham et al. 2007).

Hobo-Activator-Tam3 (hAT) DNA transposons are barely detectable in the python but comprise ~2% of the copperhead genome sample (fig. 2 and supplementary table S8, Supplementary Material online). Space invader (SPIN) elements, a type of hAT DNA transposon, are known to have been independently horizontally transferred into the genomes of multiple tetrapod lineages within the last 15–46 My, including that of the anole lizard (Pace et al. 2008; Novick et al. 2010). We found evidence of numerous SPINs in the genome of the copperhead (fig. 5) but found none in the python (corroborated by polymerase chain reaction [PCR] and Southern hybridizations; Feschotte C, unpublished data). In the copperhead, most SPIN-related sequences (1,142) found were MITEs (proximal ends) of SPIN elements representing deletion derivatives of longer and

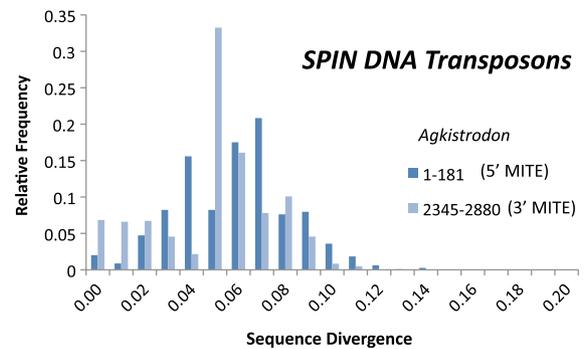


FIG. 5.—Sequence divergence between SPIN DNA transposon (MITE) sequences in the copperhead genome. Divergences from the consensus sequence were calculated as in figure 3.

presumably autonomous elements. An additional 19 reads mapped to (non-MITE) internal regions of the anole lizard SPIN transposon consensus sequence (Pace et al. 2008). SPIN MITE sequences from the copperhead display relatively low levels of sequence divergence (from a copperhead SPIN consensus), averaging around 6% (fig. 5). This is consistent with recent activity and invasion of SPINs into the copperhead genome at a similar time frame (<45 Ma) as SPINs appear to have invaded other tetrapod lineages, long after the ~100 Ma split between the python and copperhead (Castoe et al. 2009a), although the lack of a known neutral substitution rate for squamate genomes precludes precise dating.

SSR Structure

The frequency pattern of SSRs in snakes is similar to the frequency pattern of repetitive elements in that the copperhead has about four times the SSR content of the python and 2 or more times that of the anole lizard, which itself has substantially more than other reptiles or birds examined (fig. 6A). As with the TEs, it is surprising to observe such an expansion in a genome that is smaller than most other reptile genomes. Although the number of SSRs identified varies somewhat depending on the identification method (e.g., fig. 6A and supplementary fig. S5 and table S7, Supplementary Material online), the approximate proportions remain similar. The relative abundance of SSR loci in the copperhead, compared with the python and the anole, is also consistently higher across all repeat motif length classes, from 2mers to 6mers (supplementary fig. S6, Supplementary Material online). This points to a general expansion of SSR loci. The excessive relative enrichment of 4mers and 5mers in the copperhead, however, indicates a possible role for a motif-specific mechanism as well.

SSR motif sequence frequencies are quite similar between the python and anole lizard and surprisingly different in the copperhead, suggesting accelerated evolution of SSRs along the lineage leading to the copperhead (supplementary figs. S6 and S7, Supplementary Material online). Due to

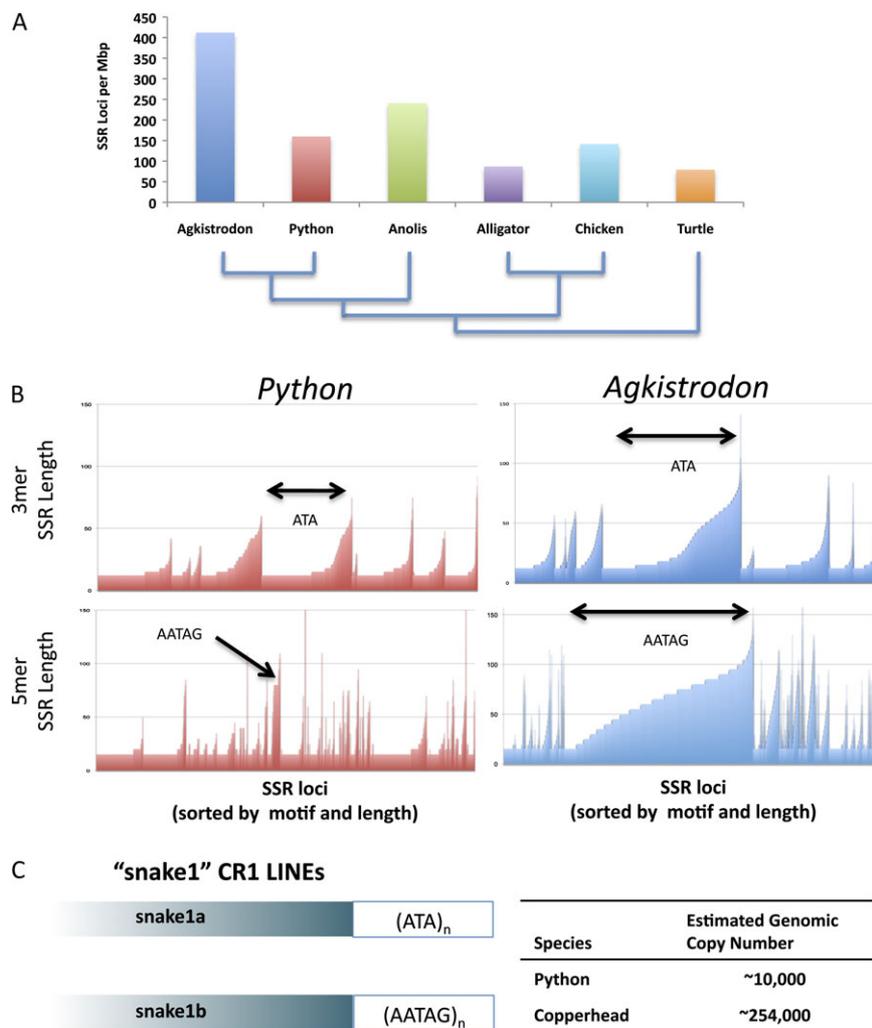


FIG. 6.—Expansion of specific SSR motifs in the copperhead genome. (A) The number of SSR loci per megabase for a sampling of amniote genomes is shown along with a phylogenetic tree of their relationships. Estimates for squamate reptiles based on *RepeatMasker* analyses; estimates for non-squamates taken from *Shedlock et al. (2007)*. (B) The 3mer and 5mer SSR loci of python (*Python*) and copperhead (*Agkistrodon*) are shown sorted first by SSR sequence motif and then by SSR length (in base pairs). The height of each bar corresponds to the length of each SSR (in base pairs), and the width is proportional to the identified number of sequences with a particular motif and length. The width of the portion of the graph devoted to each motif is proportional to the motif's relative abundance among SSRs (in terms of number of loci). The regions of the graph devoted to motifs ATA and AATAG are indicated with double arrows. (C) Two alternative SSR tails at the 5' ends of snake1 CR1 LINES are shown along with the estimated copy number of this LINE family in the two snakes.

common descent, the frequencies of SSR motifs in different species are expected to be correlated, and under the assumption of a constant rate of evolution (birth and death) of SSRs, the degree of correlation should decrease with the divergence time between species. The SSR motif-specific frequency profiles in the anole and python have a linear regression coefficient of $R^2 = 0.716$ compared with $R^2 = 0.405$ between the two snakes. The contrasting correlation strengths were particularly strong for comparisons of 2mers, 4mers, and 5mers (*supplementary fig. S6, Supplementary Material online*). These results are consistent with the hypothesis that SSR motifs are typically stable for long periods

of time (e.g., the time separating the python and the anole), but that the copperhead lineage has undergone an unusual amount of SSR turnover resulting in a major change in the SSR motif frequencies and overall abundances in the copperhead genome.

Microsatellite Seeding by CR1 LINES

Certain SSR sequence motifs were greatly expanded in the copperhead, including ATA, ATAG, AATAG (*fig. 6B* and *supplementary fig. S6, Supplementary Material online*), which are notably similar to one another. Analysis of the flanking sequences of these highly expanded copperhead SSR loci

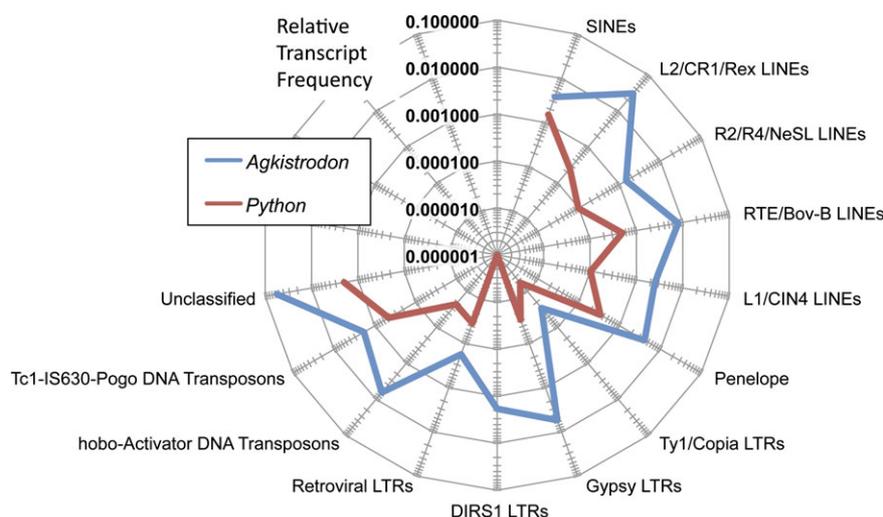


FIG. 7.—Relative frequencies of TEs in liver cDNA transcripts. Relative transcript frequencies in the two snakes are shown in a radar graph on a logarithmic scale. Sequences shown had long regions of high similarity to known TEs.

showed that the 3mer $(ATA)_n$ and 5mer $(AATAG)_n$ were associated with other non-SSR repetitive sequences: these SSR motifs are found in the 3' tails of a snake-specific family of CR1 LINEs that we refer to as snake1 CR1 LINEs (fig. 6C). Snake1 CR1s are found at low levels in the python sample (196 element fragments; ~ 7 element fragments/Mbp) but found ~ 25 times more often in the copperhead sample (11,324 element fragments; ~ 189 element fragments/Mbp); because the number of times an element was sampled multiple times by different fragments in these samples is expected to be negligible, we therefore estimate there to be $\sim 10,000$ snake1 CR1 LINEs in the python genome versus $\sim 254,000$ in the copperhead (fig. 6C). In the copperhead, 41.4% of all ATA SSR loci and 22.7% of all AATAG SSR loci were flanked by readily identifiable snake1 CR1 LINEs; only a small fraction of ATAG SSRs were flanked by snake1 CR1 LINEs (0.9%), and it is thus unclear whether LINEs directly seed these ATAG repeats or if they are mutated versions of related 3mers or 5mers. The most extreme perturbations in microsatellite frequencies between the two snakes thus seem to be due to the “microsatellite-seeding” (Arcot et al. 1995) activity of these snake1 CR1 LINEs.

The elements that we are calling snake1 CR1 LINEs have been identified previously (Nobuhisa et al. 1998; Fujimi et al. 2002; Ikeda et al. 2010) but were conflated with Bov-B LINEs because of a misannotation of a novel Bov-B LINE that was actually a Bov-B LINE flanked by two snake1 CR1 LINE fragments (supplementary fig. S9, Supplementary Material online). Snake1 CR1 LINEs are also notable because our data confirm previous speculation that they occur at high-frequency throughout phospholipase venom genes in viperid snakes (Ikeda et al. 2010), numerous other venom genes in viperids and elapids (based on Blast analysis; supplementary

table S10, Supplementary Material online), and in HOX gene clusters of colubrid snakes (Di-Poi et al. 2010). Although it would be ideal to test for significant enrichment of CR1 elements adjacent to venom genes, this is currently not possible because the lack of diversity of available sequences for venomous snake genomes that include both genic and intergenic annotated regions.

We also found that snake Bov-B LINEs tend to have a $(CAA)_n$ microsatellite repeat at their 3-prime end and thus appear to be capable of seeding $(CAA)_n$ microsatellites. Despite this, we do not find evidence for any substantial expansion of $(CAA)_n$ SSRs in the genomes of either snake species surveyed here (supplementary fig. S6, Supplementary Material online). It is also notable that even when all known LINE-associated SSR motifs are excluded from consideration, there are still large differences in SSR motif abundance between the snakes (e.g., supplementary figs. S6–S8, Supplementary Material online). This implies that in addition to the major impact of LINEs, other LINE-independent effects have altered SSR abundances in the lineage leading to the copperhead.

Evidence of TE Transcriptional Activity

Transcription levels in liver tissue samples from both species were evaluated to determine whether TE elements are actively transcribed in living snake tissues (NCBI Sequence Read Archive accession SRA029568.1; also available at www.snakegenomics.org/SnakeGenomics/Raw_Data.html). Transcripts with sequence homology to every TE class were more frequent in the copperhead liver transcriptome (fig. 7 and supplementary table S11, Supplementary Material online), with 23-fold greater overall levels of transcription of TEs in the

copperhead compared with the python, including many different TE classes that were inferred to be recently active from the genomic data. For example, LINEs in the copperhead represent ~4.6% of all transcripts (47-fold more than in the python). In addition, CR1s are particularly frequent in the copperhead, comprising ~3% of the copperhead transcriptome sample; this frequency is 122-fold greater than in the python transcripts (fig. 7 and [supplementary table S11, Supplementary Material](#) online). Bov-B LINEs were also observed in both species but were 16-fold more abundant in the copperhead (at 0.8% of transcripts; [supplementary table S11, Supplementary Material](#) online). From the genomic data, we inferred that Gypsy and DIRS1 LTR retroelements had expanded recently in the copperhead, and transcriptional data show moderately high levels of both of these in the copperhead (at 0.54% and 0.18% of transcript reads, respectively) yet these were either barely detected or not detected at all in the python transcriptome (fig. 7). We also found transcriptional evidence of hAT DNA transposon activity (which includes SPIN element activity) in the copperhead (0.64% of reads) at 280-fold greater levels than the python. The highest abundance of presumably TE-related transcripts was those that were “unclassified” TEs; we found 28-fold greater relative abundance of unclassified repeats in the copperhead (5.6% of reads) versus the python (0.2% of reads; fig. 7). Although we cannot interpret exactly what these unclassified elements represent, we expect this category to contain a substantial proportion of SINEs. Although the liver is not where TEs need to be expressed to make new inherited copies, and these transcripts do not necessarily arise from the TE’s own promoters, this data suggest the possibility that TE activity in the copperhead continues to be high compared with the python.

Discussion

Comparison of two snake genomes, spanning ~100 My of snake evolution, revealed extensive differences in their genomic repeat landscapes. Although both snakes contain diverse sets of repeat elements distributed across most major element types and superfamilies, the copperhead genome contains more of essentially all of these repeats (occupying 45% of the copperhead genome vs. 21% of the python genome), and many repeats have expanded recently. In comparison, the largest known difference in genomic repeat content between placental mammalian genomes occurs between the human and the mouse (46% vs. 38%, respectively), which are separated by ~75 My (Waterston et al. 2002). Thus, for similar levels of temporal divergence, the difference in repeat content in these two snakes is exceptional. Furthermore, the greater repetitive content in the copperhead is not due to one or a few expanded repeat families but is distributed among a diversity of element families and subfamilies (243 collapsed TE families in the copperhead vs. 82 in the python).

TE-related transcripts appear to be expressed at much higher levels in copperhead tissue compared with python tissue, even when the greater genomic TE abundances in the copperhead genome are accounted for. Although we surveyed liver rather than gametic tissues for TE activity, the 23-fold greater overall levels of TE-related transcripts in the copperhead than in the python (fig. 7) suggests that TE transcription may be generally more active in the copperhead. This is true even in the case of CR1, for which there are 25 times more elements in the copperhead than in the python genome; there are 122 times more CR1-related transcripts in copperhead tissues than in python tissues (fig. 7 and [supplementary table S11, Supplementary Material](#) online). If transcription levels have also increased in germ line tissues, they may have contributed to increased genomic TE insertion activity. One hypothesis, to explain the observations that TEs have higher transcription levels and have been more active in the copperhead versus python genomes, is that mechanisms known to control TE proliferation (e.g., CpG methylation and chromatin structural regulation; Yoder et al. 1997; Lippman et al. 2004; Feschotte 2008) may be differentially effective in the two snakes. It is also possible that TEs may occur in greater proximity to transcriptional units in the copperhead genome, driving greater levels of read-through transcription. It is unclear, however, what mutational or selective force would have made TEs land and become fixed nearer transcriptional units in the copperhead than in the python. The increased transcription levels in the copperhead also suggest that TEs are more likely to influence flanking gene expression in the copperhead than in the python.

Prior to the present study, at least two plausible instances of horizontal transfer implicating snake TEs have been reported, involving Bov-B LINEs (Kordis and Gubensek 1998b) and Sauria SINEs (Piskurek and Okada 2007). This study provides novel evidence for additional horizontal transfer of TEs and by adding genomic data from two snake species in addition to the anole lizard, it provides the first large-scale comparative view into TE dynamics within squamate reptiles. Together, our sequence-based data and PCR-based confirmation of the absence of SPIN elements in the python (Feschotte C, unpublished data), in contrast to the abundance of recently inserted SPIN elements in the copperhead, provide compelling new evidence that, as with mammalian genomes (Pace et al. 2008; Gilbert et al. 2010), reptilian genomes have been differentially invaded by these elements. Although previous studies have already suggested horizontal transfer of Bov-B LINEs between squamate reptiles and mammals (Kordis and Gubensek 1997, 1998a, 1998b), our analysis suggests the possibility of multiple transfer events into and/or out of squamate genomes ([supplementary fig. S5, Supplementary Material](#) online). The previous report of a poxvirus-mediated transfer of Squam1 SINE elements

from viperid snakes to rodents demonstrates that viruses may sometimes mediate such horizontal transfer events (Piskurek and Okada 2007). This transfer is thought to be dependent on the enzymatic machinery of a Bov-B LINE (Piskurek and Okada 2007), and high transcript levels of Bov-B reverse transcriptase in snake tissues, such as those found in the copperhead, may thus increase the probability of horizontal transfer events.

The copperhead lineage also appears to have modified microsatellite evolutionary dynamics, including microsatellite seeding (Arcot et al. 1995; Tay et al. 2010) by a snake-specific CR1 LINE family (fig. 6). Our analyses show that a high percentage of expanded microsatellite motifs were adjacent to readily identifiable snake CR1 LINEs (41.4% of all ATA and 22.7% of all AATAG SSR loci). These findings suggest that microsatellite seeding by these LINEs in the copperhead has occurred at a scale that is several orders or magnitude greater than any other example that we are aware of (Nadir et al. 1996; Tay et al. 2010). The similar SSR motif frequencies between the python and anole lizard are consistent with previous suggestions that SSR evolution and turnover rates in non-avian reptiles are generally lower than in mammals (Matsubara et al. 2006; Shedlock et al. 2007). In contrast, the increase in SSR content and radically different motif frequencies in copperhead indicate that SSR turnover rates in squamates can evolve even more rapidly than what is known from mammalian genomes.

Despite its substantial and recently expanded repeat content, the copperhead has a genome size that is among the smallest of snakes. This is surprising, as it is reasonable to expect that small genomes should have low repetitive content, as is the case in pythons and birds. We suggest that unidentified processes must be acting differentially in the copperhead to remove genomic sequence, potentially due to mechanistic differences in the biology of the two snake lineages and/or differences in selection on mutations that alter genome size. Further evidence of differential processes operating in these two snake lineages comes from our observation of differences in the relative abundance of 3' truncated LINEs between species (fig. 4). The excess of short LINEs in the copperhead is consistent with pressure to limit genome expansion, although 3' LINE truncation has not been sufficient to balance the genome size equation (the total LINE element sequence is still considerably greater in the copperhead). The relative bias toward short elements in the copperhead could be caused by a greater tendency to generate shorter elements (at the time of insertion), a greater probability to fix shorter elements, and/or a greater probability to delete long elements at any time via ectopic recombination, which has been proposed to occur in the anole lizard genome (Novick et al. 2009). Selection could be more effective in the copperhead than the python due to a larger effective popu-

lation size rather than stronger selection against genome expansion, but this seems unlikely. It is expected that the Nearctic-distributed copperhead lineage suffered small effective population sizes due to bottlenecks during glacial cycles (Guiher and Burbrink 2008), and the likelihood of large and stable effective populations in tropical lineages such as the python also contraindicates a primary role for population size differences as a sole explanation.

Selection to maintain a smaller genome size has been hypothesized numerous times in relation to extreme metabolic demands in flighted birds (Hughes and Hughes 1995), although there is some controversy (Organ et al. 2007). Previous studies have suggested that extreme metabolic demand in snakes (Secor and Diamond 1995, 1998) has resulted in selection to decrease their mitochondrial genome size (Jiang et al. 2007), extensive evolutionary redesign (Castoe et al. 2008), and previously unprecedented molecular convergence in snake metabolic proteins (Castoe et al. 2009b). It is therefore plausible that selection related to metabolic demands could have shaped snake nuclear genomes. Broader understanding of genomic repeat landscapes in snakes may shed greater light on this question. There are a range of alternative theories about the evolution of genome size and complexity (Lynch and Conery 2003), and thus the role of selection in snake genome size and structure is a topic of considerable interest.

It is also an open question whether the biology of snake genomes may have contributed to the evolution of their extreme phenotypes and adaptations (Secor and Diamond 1995; Cohn and Tickle 1999; Fry et al. 2006; Castoe et al. 2008, 2009b; Vonk et al. 2008). Among the most conspicuous adaptations in snakes is that some lineages, including the ancestors of the copperhead, have evolved complex venom repertoires, largely by duplicating and repurposing existing genes to produce deadly toxins. Our evidence (supplementary table S10, Supplementary Material online), and that of others (Nobuhisa et al. 1998; Fujimi et al. 2002; Ikeda et al. 2010), shows a tentative association between CR1 LINEs and venom genes. Our genomic sampling suggests that CR1 LINEs (as well as SSRs and other TEs) have expanded substantially in the copperhead lineage, and this expansion might be expected to lead to increased rates of recombination, unequal crossing over, and gene conversion (Witherspoon et al. 2009; Stevison and Noor 2010). These events could have, at least in part, facilitated the expansion and regulatory rewiring of venom gene families in venomous snakes.

Supplementary Material

Supplementary methods, tables S1–S11, and figures S1–S10 are available at *Genome Biology and Evolution* online (http://www.oxfordjournals.org/our_journals/gbe/).

Acknowledgments

We acknowledge the support of the National Institutes of Health (NIH; GM083127 to D.D.P., GM77582 to C.F.); NIH Training Grant (LM009451 to T.A.C.); National Science Foundation Support (DEB-0416160 to E.N.S.). We acknowledge Roche 454 for contributing the sequencing of the cDNA libraries based on a proof of principle grant to D.P. and T.C. We thank Carl Franklin from the Amphibian and Reptile Diversity Research Center at the University of Texas at Arlington for assistance obtaining snake samples.

Literature Cited

- Aleshin A, Zhi D. 2010. Recombination-associated sequence homogenization of neighboring Alu elements: signature of nonallelic gene conversion. *Mol Biol Evol.* 27:2300–2311.
- Arcot SS, et al. 1995. Alu repeats: a source for the genesis of primate microsatellites. *Genomics* 29:136–144.
- Bao Z, Eddy SR. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* 12:1269–1276.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–580.
- Castoe TA, et al. 2008. Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS One.* 3:e2201.
- Castoe TA, et al. 2009a. Dynamic nucleotide mutation gradients and control region usage in squamate reptile mitochondrial genomes. *Cytogenet Genome Res.* 127:112–127.
- Castoe TA, et al. 2009b. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc Natl Acad Sci U S A.* 106:8986–8991.
- Castoe TA, et al. 2010. Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Mol Ecol Resour.* 10:341–347.
- Cohn MJ, Tickle C. 1999. Developmental basis of limblessness and axial patterning in snakes. *Nature* 399:474–479.
- Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. *Nat Rev Genet.* 10:691–703.
- Di-Poi N, et al. 2010. Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature* 464:99–103.
- Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet.* 9:397–405.
- Feschotte C, et al. 2009. Exploring repetitive DNA landscapes using REPCLASS, a tool that automates the classification of transposable elements in eukaryotic genomes. *Genome Biol Evol.* 2009:205–220.
- Fry BG, et al. 2006. Early evolution of the venom system in lizards and snakes. *Nature* 439:584–588.
- Fujimi TJ, Tsuchiya T, Tamiya T. 2002. A comparative analysis of invaded sequences from group IA phospholipase A(2) genes provides evidence about the divergence period of genes groups and snake families. *Toxicon.* 40:873–884.
- Gilbert C, et al. 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464:1347–1350.
- Gregory TR, et al. 2007. Eukaryotic genome size databases. *Nucleic Acids Res.* 35:D332–D338.
- Gu W, et al. 2008. Identification of repeat structure in large genomes using repeat probability clouds. *Anal Biochem.* 380:77–83.
- Guiher TJ, Burbrink FT. 2008. Demographic and phylogeographic histories of two venomous North American snakes of the genus *Agkistrodon*. *Mol Phylogenet Evol.* 48:543–553.
- Herpin A, et al. 2010. Transcriptional rewiring of the sex determining *dmrt1* gene duplicate by transposable elements. *PLoS Genet.* 6:e1000844.
- Hughes AL, Hughes MK. 1995. Small genomes for better flyers. *Nature* 377:391.
- Ikeda N, et al. 2010. Unique structural characteristics and evolution of a cluster of venom phospholipase A2 isozyme genes of *Protobothrops flavoviridis* snake. *Gene.* 461:15–25.
- Janes DE, et al. 2010. Genome evolution in Reptilia, the sister group of mammals. *Annu Rev Genomics Hum Genet.* 11:239–264.
- Jiang ZJ, et al. 2007. Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. *BMC Evol Biol.* 7:123.
- Jurka J. 2000. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* 16:418–420.
- Kordis D. 2009. Transposable elements in reptilian and avian (sauropsida) genomes. *Cytogenet Genome Res.* 127:94–111.
- Kordis D, Gubensek F. 1997. Bov-B long interspersed repeated DNA (LINE) sequences are present in *Vipera ammodytes* phospholipase A2 genes and in genomes of Viperidae snakes. *Eur J Biochem.* 246:772–779.
- Kordis D, Gubensek F. 1998a. The Bov-B lines found in *Vipera ammodytes* toxic PLA2 genes are widespread in snake genomes. *Toxicon.* 36:1585–1590.
- Kordis D, Gubensek F. 1998b. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. *Proc Natl Acad Sci U S A.* 95:10704–10709.
- Lee C, Grasso C, Sharlow MF. 2002. Multiple sequence alignment using partial order graphs. *Bioinformatics* 18:452–464.
- Li R, et al. 2010. The sequence and de novo assembly of the giant panda genome. *Nature* 463:311–317.
- Lippman Z, et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430:471–476.
- Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302:1401–1404.
- Matsubara K, et al. 2006. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *Proc Natl Acad Sci U S A.* 103:18190–18195.
- Nadir E, Margalit H, Gallily T, Ben-Sasson SA. 1996. Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. *Proc Natl Acad Sci U S A.* 93:6470–6475.
- Nobuhisa I, et al. 1998. Retrotransposable CR1-like elements in crotalinae snake genomes. *Toxicon.* 36:915–920.
- Novick PA, et al. 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.
- Novick P, Smith J, Ray D, Boissinot S. 2010. Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. *Gene.* 449:85–94.
- Olmo E. 1984. Genomic composition of reptiles: evolutionary perspectives. *J Herpetol.* 18:20–32.
- Olmo E, Stingo V, Odierna G, Cobror O. 1981. Variations in the repetitive DNA and evolution in reptiles. *Comp Biochem Physiol.* 69:687–691.
- Organ CL, et al. 2007. Origin of avian genome size and structure in non-avian dinosaurs. *Nature* 446:180–184.

- Pace JK 2nd, Gilbert C, Clark MS, Feschotte C. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc Natl Acad Sci U S A*. 105:17023–17028.
- Piskurek O, Okada N. 2007. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. *Proc Natl Acad Sci U S A*. 104:12046–12051.
- Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics*. 21:i351–i358.
- Pritham EJ, Putliwala T, Feschotte C. 2007. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene*. 390:3–17.
- Secor SM, Diamond J. 1995. Adaptive responses to feeding in Burmese pythons: pay before pumping. *J Exp Biol*. 198:1313–1325.
- Secor SM, Diamond J. 1998. A vertebrate model of extreme physiological regulation. *Nature* 395:659–662.
- Secor SM, Diamond JM. 2000. Evolution of regulatory responses to feeding in snakes. *Physiol Biochem Zool*. 73:123–141.
- 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.
- Smit AFA, Hubley R, Green P. 2010. RepeatMasker open-3.0. Available from: <http://repeatmasker.org>.
- Stevison LS, Noor MA. 2010. Genetic and evolutionary correlates of fine-scale recombination rate variation in *Drosophila persimilis*. *J Mol Evol*. 71:332–345.
- Tay WT, Behere GT, Batterham P, Heckel DG. 2010. Generation of microsatellite repeat families by RTE retrotransposons in lepidopteran genomes. *BMC Evol Biol*. 10:144.
- Townsend TM, Larson A, Louis E, Macey JR. 2004. Molecular phylogenetics of squamata: the position of snakes, amphisbaenians, and dibamids, and the root of the squamate tree. *Syst Biol*. 53:735–757.
- Vidal N, Hedges SB. 2005. The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. *C R Biol*. 328:1000–1008.
- Vonk FJ, et al. 2008. Evolutionary origin and development of snake fangs. *Nature* 454:630–633.
- Warren WC, et al. 2010. The genome of a songbird. *Nature* 464:757–762.
- Waterston RH, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562.
- Wiens JJ, et al. 2010. Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: molecular data change the placement of fossil taxa. *Syst Biol*. 59:674–688.
- Witherspoon DJ, et al. 2009. Alu repeats increase local recombination rates. *BMC Genomics*. 10:530.
- Yoder JA, Walsh CP, Bestor TH. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet*. 13:335–340.

Associate editor: Richard Cordaux