

Research article

Open Access

Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region

Zhi J Jiang^{†1,2}, Todd A Castoe^{†3,4}, Christopher C Austin⁵, Frank T Burbrink^{1,6}, Matthew D Herron^{3,7}, Jimmy A McGuire^{5,8}, Christopher L Parkinson³ and David D Pollock^{*9}

Address: ¹Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, Baton Rouge, LA, USA, ²Computational Biology, Scripps Florida, Jupiter, FL, USA, ³Department of Biology, University of Central Florida, Orlando, FL, USA, ⁴Department of Biochemistry and Molecular Genetics, University of Colorado Health Sciences Center, Aurora, CO, USA, ⁵Museum of Natural Science, Louisiana State University, Baton Rouge, LA, USA, ⁶Department of Biology, City University of New York, Staten Island, USA, ⁷Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA, ⁸Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, USA and ⁹Department of Biochemistry and Molecular Genetics, University of Colorado Health Sciences Center, Aurora, CO, USA

Email: Zhi J Jiang - zjiang1@scripps.edu; Todd A Castoe - Todd.Castoe@uchsc.edu; Christopher C Austin - ccaustin@lsu.edu; Frank T Burbrink - burbrink@mail.csi.cuny.edu; Matthew D Herron - mherron@email.arizona.edu; Jimmy A McGuire - mcguirej@berkeley.edu; Christopher L Parkinson - cparkins@mail.ucf.edu; David D Pollock* - David.Pollock@uchsc.edu

* Corresponding author †Equal contributors

Published: 26 July 2007

Received: 6 September 2006

BMC Evolutionary Biology 2007, 7:123 doi:10.1186/1471-2148-7-123

Accepted: 26 July 2007

This article is available from: <http://www.biomedcentral.com/1471-2148/7/123>

© 2007 Jiang et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The mitochondrial genomes of snakes are characterized by an overall evolutionary rate that appears to be one of the most accelerated among vertebrates. They also possess other unusual features, including short tRNAs and other genes, and a duplicated control region that has been stably maintained since it originated more than 70 million years ago. Here, we provide a detailed analysis of evolutionary dynamics in snake mitochondrial genomes to better understand the basis of these extreme characteristics, and to explore the relationship between mitochondrial genome molecular evolution, genome architecture, and molecular function. We sequenced complete mitochondrial genomes from Slowinski's corn snake (*Pantherophis slowinskii*) and two cottonmouths (*Agkistrodon piscivorus*) to complement previously existing mitochondrial genomes, and to provide an improved comparative view of how genome architecture affects molecular evolution at contrasting levels of divergence.

Results: We present a Bayesian genetic approach that suggests that the duplicated control region can function as an additional origin of heavy strand replication. The two control regions also appear to have different intra-specific versus inter-specific evolutionary dynamics that may be associated with complex modes of concerted evolution. We find that different genomic regions have experienced substantial accelerated evolution along early branches in snakes, with different genes having experienced dramatic accelerations along specific branches. Some of these accelerations appear to coincide with, or subsequent to, the shortening of various mitochondrial genes and the duplication of the control region and flanking tRNAs.

Conclusion: Fluctuations in the strength and pattern of selection during snake evolution have had widely varying gene-specific effects on substitution rates, and these rate accelerations may have been functionally related to unusual changes in genomic architecture. The among-lineage and among-gene variation in rate dynamics observed in snakes is the most extreme thus far observed in animal genomes, and provides an important study system for further evaluating the biochemical and physiological basis of evolutionary pressures in vertebrate mitochondria.

Background

The vertebrate mitochondrial (mt) genome has been an important model system for studying molecular evolution, organismal phylogeny, and genome structure. Despite extensive molecular studies, little is known regarding the ways in which genome architecture might affect the various aspects of genome function and evolution (including replication, transcription, and RNA/protein function, as well as rates and patterns of nucleotide evolution). Nevertheless, patterns linking mt genome structure, function, and nucleotide evolution have begun to emerge [1-3].

Among the most direct demonstrated links among genome architecture, function and nucleotide evolution is that relating the asymmetrical genome replication process with gradients of transition substitutions in vertebrate mitochondrial genomes [1-3]. Gradients of transition mutations, arising from deamination mutations, are observed due to the differential time regions of the mt genome spend in an asymmetric mutagenic state during genome replication (T_{AMS} alternatively referred to as the time spent in a single-stranded state, D_{SSH} [4-6], but there is some controversy about this: see Additional file 1). Thus, gradients of transition biases are dependent upon the relative position of the functional origins of heavy and light strand replication. In vertebrate mt genomes, the origin of heavy strand replication (O_H) is thought to be within the control region (CR), and the origin of light strand replication (O_L) in the tRNA cluster referred to as the WANCY region (named for the five amino acids coded for by these five tRNAs). Among transition classes in vertebrate mt genomes, T→C light strand substitutions at degenerate 3rd codon positions increase linearly with increasing T_{AMS} and C/T nucleotide frequencies at degenerate 3rd positions are good predictors of T_{AMS} [4].

The mt genomes of snakes contain a number of characteristics that are unusual among vertebrates, and represent an ideal model for exploring potential links among genome structure, function, and evolution. Snake mitochondrial genomes appear to have the highest evolutionary rates among vertebrates and contain truncated tRNAs and other shortened genes [7,8]. All snake species sampled to date, except the scolecophidian snakes *Leptotyphlops dulcis*, *Ramphotyphlops australis*, and *Typhlops murius*, have a duplicated control region (CR2) between NADH dehydrogenase subunit 1 (ND1) and subunit 2 (ND2), in addition to a control region (CR1) adjacent to the 5'-end of the 12s rRNA as it is in other vertebrates [7-11]. These two control regions appear to undergo concerted evolution that acts to homogenize the nucleotide sequence of each duplicate copy within a given genome [7-9]. The functionality of these two control regions in transcription and initiation of heavy strand replication is not clear, but

given that the nucleotide sequence of each is nearly identical, any functional features that are not dependent on surrounding sequences should be similar. In contrast, recent evidence suggests that initiation of heavy strand replication may be distributed across a broad zone, including cytochrome b (CytB) and NADH dehydrogenase subunit 6 (ND6) [12], indicating that CR2 may not function as effectively in this role.

A number of interesting questions arise that might be addressed through comparative analysis, including: (1) does one or the other, or do both control regions function as origins of heavy strand DNA synthesis? (2) does the altered genome structure affect patterns of snake mt genome molecular evolution? (3) when during snake evolution did various features arise, and were any changes synchronous? (4) do patterns of mt molecular evolution vary at different depths of phylogeny? and (5) is there any evidence or plausible rationale for selection as a causative agent in generating differences in genomic structure and molecular evolutionary patterns?

To investigate outstanding questions regarding snake mitochondrial genome evolution, structure, and function, we analyzed a dataset consisting of three new complete snake mitochondrial genomes together with all eight previously published snake mitochondrial genomes that were available at the time of this study, and 42 other vertebrate mitochondrial genomes for comparative purposes. The new snake genomes were obtained from one *Pantherophis slowinskii* (Colubroidea: Colubridae; a corn snake from Louisiana; previously *Elaphe guttata*), and from two *Agkistrodon piscivorus* (Colubroidea: Viperidae; the cottonmouth or water moccasin; one specimen from Florida and the other from Louisiana).

Results

Brief summary of the new complete snake mitochondrial genomes

The gene contents of *A. piscivorus* and *P. slowinskii* mt genomes are very similar to other snakes (Figure 1; for detailed genome annotation see Additional file 2). As in all known alethinophidian snake mt genomes, these species have a presumably duplicated control region (CR2) between ND1 and ND2, in addition to the original control region (CR1) present in all vertebrates adjacent to the 5' end of the 12s rRNA gene [7-9]. These genomes also possess the translocated tRNA^{Leu} common to all alethinophidian snakes sampled thus far (3' of CR2). In addition to an intact tRNA^{Pro} between CytB and CR1, *P. slowinskii* has an apparent pseudo-tRNA^{Pro} gene (Ψ -tRNA^{Pro}) between ND1 and CR2 (as does the previously sequenced colubrid, *Dinodon semicarinatus*). In *P. slowinskii*, this Ψ -tRNA^{Pro} exactly matches the first 35 bases of tRNA^{Pro}. In contrast, the intact tRNA^{Pro} of *A. piscivorus*

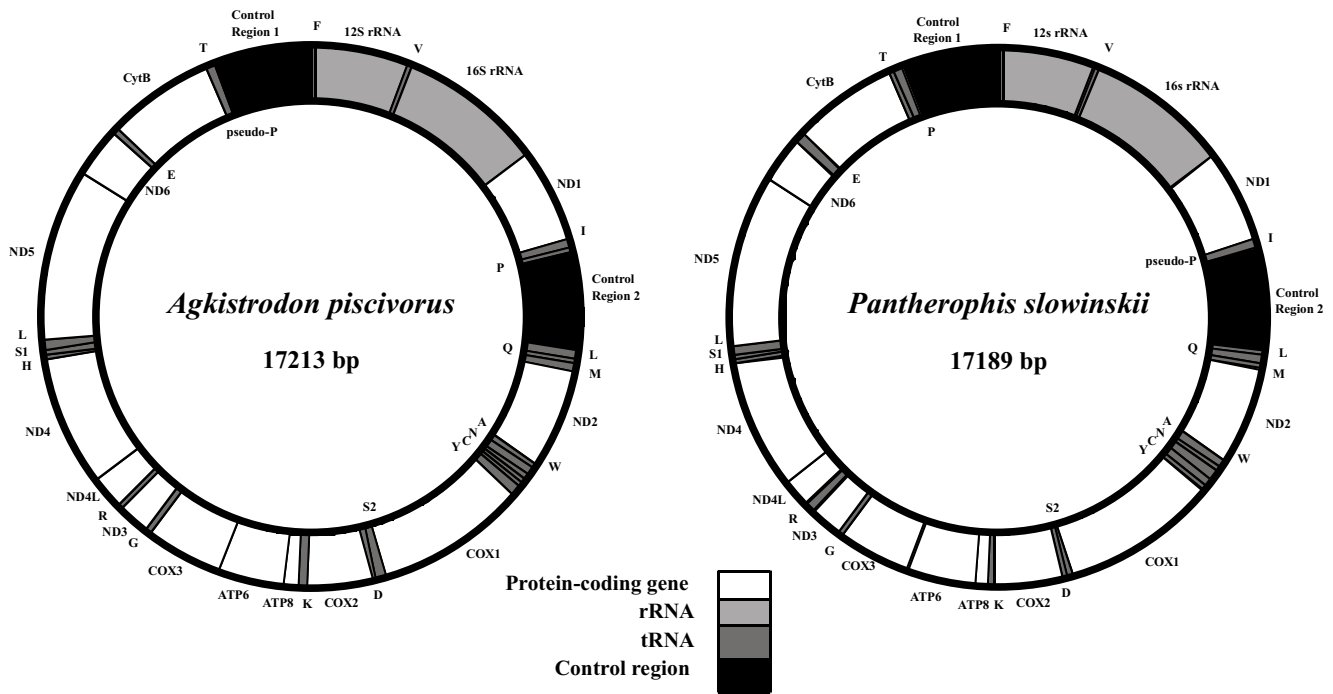


Figure 1
Annotated mitochondrial genome maps of *Agkistrodon piscivorus* and *Pantherophis slowinskii*. The two *A. piscivorus* samples (*Api1* and *Api2*) have identical annotations except for minor variations in gene length. Labels of genes outside the circle refer to genes transcribed from the light strand, and names within the circle represent genes transcribed from the heavy strand.

(and the only previously sequenced viperid, *Ovophis okinavensis*) is located between ND1 and CR2 (exactly the location of Ψ -tRNA^{Pro} in the colubrids), and there is a 31 bp non-coding fragment between tRNA^{Thr} and CR1, where tRNA^{Pro} is usually located. In *O. okinavensis*, this is clearly a Ψ -tRNA^{Pro}, since these 31 bp are an exact match to the CR1-proximal end of the complete tRNA^{Pro}, but in *A. piscivorus* the homology is much less clear. These alternative positions of tRNA^{Pro}, Ψ -tRNA^{Pro}, and a previously noted [7] duplication of tRNA^{Phe} in *O. okinavensis* are the only known mt genome gene rearrangements identified within alethinophidian snakes.

Within a mt genome, the two copies of the CR in each newly sequenced species are nearly identical (e.g., *Api1* CR1 and CR2), as is typical for alethinophidian snakes [7,8]. In *P. slowinskii* there is a single point mutation and four extra nucleotides at one end of CR1, in *Api1* there is one indel plus 14 extra nucleotides on one end of CR1, and in *Api2* there are seven indels and two base changes between the two control regions. Between *Api1* and *Api2*, CR1 differs by five indels and 19 point mutations, whereas CR2 differs by three indels (two at the 5' end) and 18 point mutations.

Comparison of *Agkistrodon piscivorus* genomes

Polymorphisms were observed between the two *A. piscivorus* genomes, *Api1* and *Api2*, for all protein and rRNA genes and for 14 of 22 tRNAs (see Additional file 2). The 12s and 16s rRNAs were the most conserved genes between the two *A. piscivorus* individuals, with 2% and 3% sequence divergence respectively (Figure 2A; Additional file 2). Protein-coding genes differed more, up to 6.2% for ND3 (Figure 2A; Additional file 2). Most differences occurred at 3rd codon positions (Figure 2A; Additional file 2), as expected under predominantly neutral patterns of divergence (for example, 57/58 substitutions in COX1 were at 3rd codon positions).

Within *A. piscivorus*, the control regions (e.g., CR1 in *Api1* vs. CR1 in *Api2*) are as similar to each other as are the rRNA genes, and more similar than the protein coding genes (Figure 2A). This is in strong contrast to the normal pattern of divergence between vertebrate species, for which control region similarity is far less than that of protein-coding or rRNA genes, e.g., [13,14]. Between *A. piscivorus* and the other viperid, *O. okinavensis*, the control regions have 30% more differences (with indels included) than the rRNAs, and are on par with divergence in the pro-

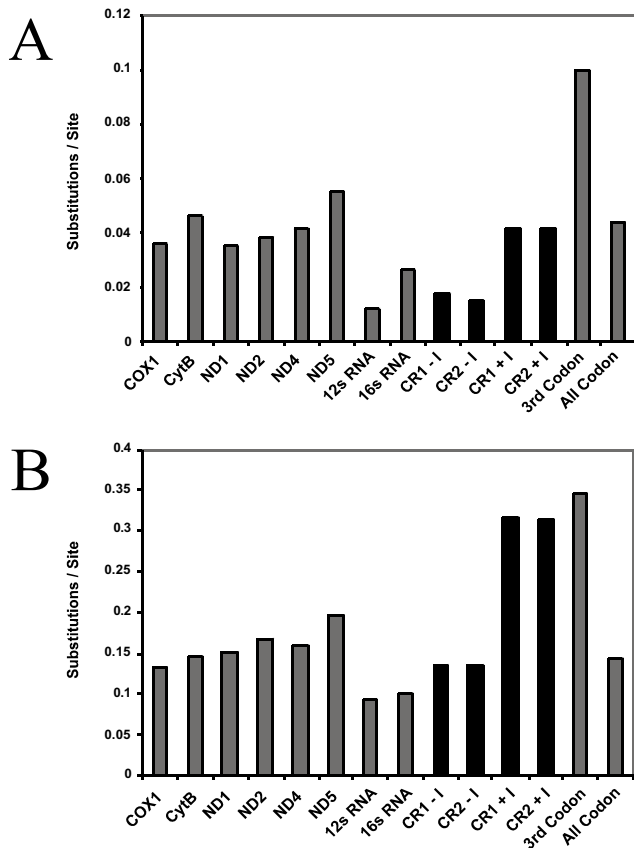


Figure 2
Differences per site for homologous genes or groups of sites in the two *Agkistrodon piscivorus* genomes and in the two viper genomes. The differences per site are shown for a comparison of *Api1* and *Api2* (A), and for *A. piscivorus* (mean of *Api1* and *Api2*) and *Ovophis okinavensis* (B). Differences are shown only for the longer protein-coding genes. For the control regions only (shaded black), differences are shown for each aligned site including indels (e.g., CR1+), or excluding indels (e.g., CR1-). For all other genes, indels are not included in the difference measure. The bars for 3rd codon positions (3rd Codon) and for all codon positions (All Codon) are summed over all protein-coding genes.

tein-coding genes (Figure 2B). If indels are included, the control regions of these two species are nearly as different as the average 3rd codon position (Figure 2B). The high degree of similarity (low divergence) observed between the CRs of the two *A. piscivorus* individuals is surprising, and contrasts sharply with the high relative divergence of CRs between *O. okinavensis* and *A. piscivorus* (Figure 2).

Phylogenetics

We present the phylogenetic tree estimate obtained by ML, with NJ bootstrap values (BS) and posterior probabilities (PP) for nodal support, which were generally high (Figure 3). Our phylogeny estimate provides a well-

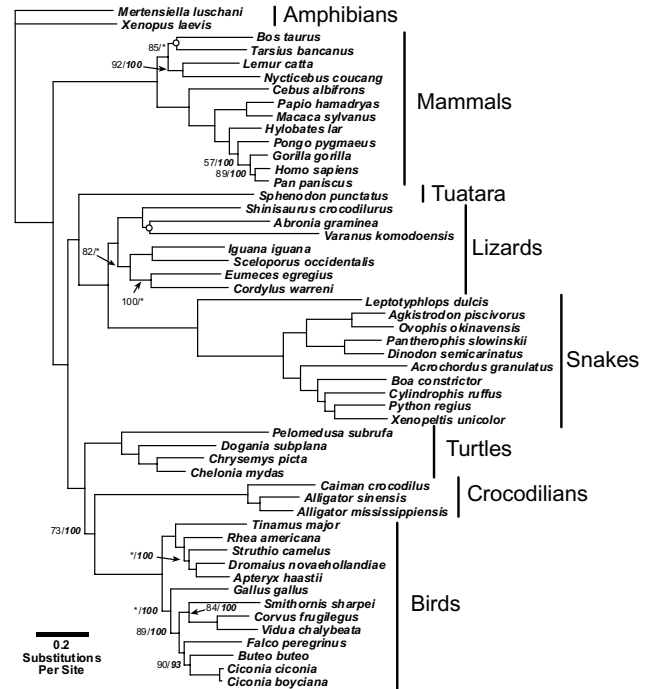
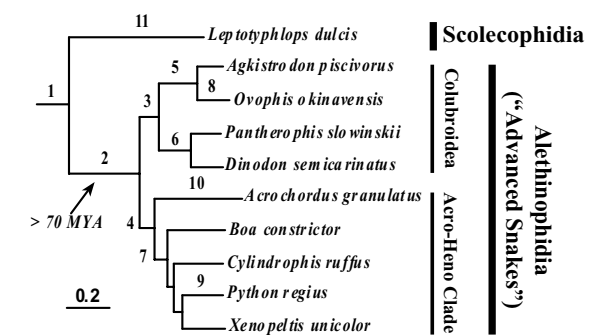


Figure 3
Maximum likelihood phylogeny for vertebrate taxa included in this study. This phylogeny is based on all protein-coding and rRNA genes. Most branches have greater than 95% support for both NJ ML distance bootstrap and Bayesian posterior probability support (see Methods), and are not annotated with support values. Where support from either measure is less than 95%, the support values are indicated by ratios, with the ML bootstrap support on top and the Bayesian posterior probability support below in italics, except for two nodes with less than 50% support by either measure, which are indicated by a hollow circle. Other than for these two nodes, support values less than 50% are indicated with an asterisk (*).

resolved and, in many cases, strongly-supported amniote phylogeny that is consistent with previous molecular studies. Differences between the ML topology (Figure 3), and the topology based on Bayesian analysis (not shown) were minor, and included an alternative placement of *Bos taurus* among mammals, and alternative placements of *Gallus gallus* and *Rhea americana* among birds. Additionally, relationships among lizard taxa varied, with *Cordylus warreni* estimated to be the sister lineage to all other lizards, and an alternative placement of *Varanus komodoensis*.

All phylogenetic estimates provided an identical, well-supported topology for relationships among snakes (Figure 3), and a summary of results concerning snake relationships is shown in Figure 4. The Scolecophidia (Typhlopoidea), represented here by *L. dulcis*, formed the sister group to all remaining snakes. Rather than finding



Branch	Major Genomic and Molecular Evolutionary Events
1	Length reduction in all genes; Simplification of the tRNA T-arms Acceleration of ATP6, ATP8, COX1, COX2, CytB, ND1, ND2, and ND5
2	Duplication of CR; Transposition of tRNA ^{Leu} Acceleration of ATP6, ATP8, COX1, COX2, CytB, and ND6
3	Duplication of tRNA ^{Pro} ; Length reduction in tRNA and rRNA genes Acceleration of ND5, ND6, and 12s, 16s rRNAs Rate of CR concerted evolution increases Increase in ND1 C/T ratio indicating CR2 function
4	Length increase in rRNA genes Acceleration of ATP6, COX3, ND3, ND4L, ND6, 16s rRNA
5	Degradation/loss of tRNA ^{Pro} duplicate (3' of CR1)
6	Degradation of tRNA ^{Pro} duplicate (3' of CR2)
7	Decrease in C/T ratio of ND1 indicating CR1 preference
8	Duplication/translocation of tRNA ^{Phe} Concerted evolution of tRNA ^{Phe} copies along with CRs Acceleration of 16s rRNA
9	Increase in C/T ratio of ND1 indicating CR2 preference
10	Acceleration of ATP6, ATP8, and COX2
11	Loss of light strand origin; Translocation of tRNA ^{Gln}

Figure 4
Hypotheses for the relative timing of alterations in mitochondrial genome architecture and molecular evolution throughout snake phylogeny. The topological relationships among snakes and branch lengths shown are the same as in Figure 3. Major groups of snakes are indicated along with the approximate diversification time of the Alethinophidia.

support for the Caenophidia (*Acrochordus* plus Colubroidea [7,15]), we find strong support for *Acrochordus* as the sister lineage to the Henophidia (including *Cylindrophis*). Hereafter we will therefore operationally refer to this clade including the Henophidia and *Acrochordus* as the "Acro-Heno" clade, and we will refer to its sister clade as the Colubroidea [16].

The snake and overall amniote phylogeny are strongly supported by our analysis of this dataset, and we henceforth treat this phylogeny as accurate. We wish to empha-

size, however, that the consistency of the phylogenetic results do not guarantee that they are, in fact, accurate. Although for simplicity we present a single nucleotide substitution model for the entire dataset we have analyzed an expanded version of this dataset (with additional unpublished snake and lizard mt genomes; with and without inclusion of the rRNA genes) using complex partitioned models for each gene and codon position. The results of this expanded (highly partitioned-model) phylogeny estimate (not shown) were essentially identical to those presented here in terms of the placement of snakes within squamates, and relationships among squamates. We provide evidence below for extremely complex non-stationary patterns of nucleotide substitution across branches and mt genome regions, and have previously identified asymmetric substitution gradients in mt genomes [4] that may vary among species (e.g., primates [3]). These latter patterns cannot be modeled using available phylogenetic programs (e.g., MrBayes [17]). We expect our phylogenetic estimates here to represent a good estimate of the relationships among mt genomes sampled, and if minor inaccuracies in the topology have occurred in our estimates, these changes should not substantially impact the qualitative conclusions of further analyses (e.g., sliding window analysis, SWA) because a majority of these later estimates are averaged over many branches of the tree, and the dynamics we concentrate on are quite dramatic and are likely to be obvious and qualitatively similar even with slight changes in the topology estimate.

Nucleotide frequencies and control region functionality

In *A. piscivorus* and *P. slowinskii* mt genomes, as in other vertebrates [5], nucleotides A and C are favored on the light strand, particularly at 3rd codon positions. This bias is probably related to elevated rates of deamination mutations on the heavy strand incurred during replication (see Background), and there is considerable variation in nucleotide content among individual mt genomes (see Additional file 2). Variation in snakes, even at 3rd codon positions, is not exceptional compared to other groups, and there is no clear snake-specific nucleotide bias evident (see Additional file 2) or strong branch-specific, or gene-specific nucleotide bias shifts across squamate mt genomes that would explain our findings of dramatic branch-specific and gene-specific rate dynamics.

Due to the simple linear relationship in most vertebrate mt genomes between C/T ratios and T_{AMS} predicted based on the location of the (functional) control region, it is of interest to determine whether there has been any clear genetic effect of the duplicated control region in alethinophidians. Exclusive use of one control region or the other would be most strongly observable in ND1 (the only protein-coding gene located between the two control

regions in alethinophidian snake mt genomes) because it is the only protein-coding gene that would spend a substantially different amount of time in the asymmetric mutagenic state (T_{AMS} see Additional file 2) depending on which control region is functional. Since the nucleotide sequence of duplicate control regions is nearly identical within each genome, however, it is also reasonable to consider the possibility that both control regions are functional.

To test these predictions about CR2 functionality, we applied our MCMC analysis [3] to fit alternative models of exclusive CR1 or CR2 usage, or mixed control region effect (Table 1). The Akaike weights for the alternative individual models provide a measure of the degree to which a control region is exclusively functional, while the weight parameter in the mixed model represents the time-averaged effect of mixed control region usage on the C/T ratios. There is evidence for at least mixed CR2 usage in all but one species (*Cylindrophis ruffus*). The evidence is good for exclusive or nearly exclusive CR2 functionality in two species (*Acrochordus granulatus* and *Python regius*), and for a strong CR2 preference in *A. piscivorus*. The patterns appear to be lineage-specific and evolutionarily labile (i.e., strong preferences for a particular control region are widely dispersed on the tree), which may indicate rapid evolution of the strength of the replication-associated substitution gradient (as suggested in primates [3]) or rapid evolution of differential usage of the two control regions. Species with ambiguous control region preferences may have mixed usage, may not have a strong enough gradient to differentiate, or may have previously switched usage and thus have not reached mutational equilibrium.

Gene length and stability of truncated tRNAs in snakes

In snakes, all mt protein-coding genes (except COX1), ribosomal RNAs, tRNAs, and individual CRs are shorter

than their counterparts in most lizards and most other vertebrates (see Additional file 3). An exception to this is *Sphenodon punctatus*, for which the control region, ATP8 (ATP synthase subunit 8) and the 12s rRNA are all shorter than in snakes. With the increased sampling in this study, it appears that while the tRNAs and proteins became shorter prior to the divergence of all snakes, the tRNAs became shorter still within the Colubroidea (Figures 4 and Additional file 3). Additionally, the rRNAs did not become shorter in *L. dulcis* or the Acro-Heno clade, but are dramatically shorter in the Colubroidea (Figures 4 and Additional file 3).

The shorter length of tRNAs in snakes results mainly from a truncated T-arm in the secondary structure (see also [8,9]). In some tRNAs, the D-arm is also shorter, but to a lesser extent than the T-arms. Although short tRNAs are typically less stable than long ones, there is only a minor effect of sequence length on secondary structure stability (ΔG) in snake tRNAs. The cloverleaf structures of most snake tRNAs are slightly less stable than their lizard counterparts (see Additional file 2), but two tRNAs (tRNA^{Ile}, tRNA^{Met}) are actually more structurally stable in snakes than in other squamates with longer tRNAs.

Spatio-temporal substitution rate dynamics across mt genes and genomic regions

To assess the difference in substitution rates among genes, we fixed the topology (to that in Figure 3) and calculated branch lengths based on rRNAs and on all protein-coding genes (Figure 5). For the rRNAs, most other major amniote groups have experienced similar amounts of total evolution from their common ancestor with the amphibians (i.e., the branch lengths from the root to the terminals are similar), whereas the snake lineages stand out as unusual in their apparently accelerated evolution (i.e., their exceptionally long root to tip branch lengths; Figure 5A). For protein-coding genes there is much more variation across

Table 1: Results of mitochondrial genome replication model analyses

Species	Individual model		Mixed model	
	O_H^{CR1}	O_H^{CR2}	$O_H^{CR1} + O_H^{CR2}$	% O_H^{CR2}
<i>Agkistrodon piscivorus</i>	1179.2 (18%)	1178.0 (60%)	1179.0 (22%)	99%
<i>Pantherophis slowinskii</i>	1164.6 (29%)	1164.1 (47%)	1164.8 (24%)	54%
<i>Dinodon semicarinatus</i>	1167.1 (21%)	1166.2 (57%)	1167.1 (22%)	78%
<i>Ovophis okinavensis</i>	1252.7 (38%)	1252.6 (45%)	1253.5 (17%)	59%
<i>Boa constrictor</i>	854.5 (29%)	853.9 (50%)	854.8 (21%)	64%
<i>Acrochordus granulatus</i>	1245.0 (2%)	1241.5 (72%)	1242.5 (26%)	100%
<i>Xenopeltis unicolor</i>	1159.4 (31%)	1159.0 (45%)	1159.6 (24%)	50%
<i>Python regius</i>	1133.0 (1%)	1128.9 (72%)	1130.0 (26%)	100%
<i>Cylindrophis ruffus</i>	1129.8 (70%)	1132.6 (4%)	1130.8 (26%)	< 1%

Negative log likelihood values and Akaike weights (in parentheses) for individual origin of replication models and the mixed model, along with the most likely CR2 preference parameter in the mixed model, for alethinophidian snakes.

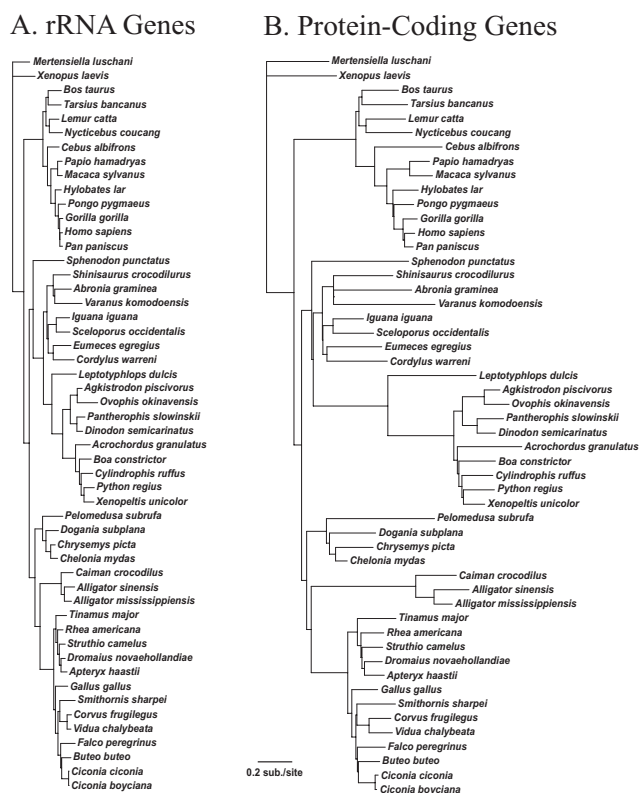


Figure 5
Phylograms based on the relative branch lengths for rRNA and protein-coding genes. Branch lengths were estimated on the topology of the ML phylogeny (Figure 3). Branch lengths on this constrained topology were estimated using all rRNA genes (A) or all protein-coding genes (B). The substitution rate scale is the same in both trees.

lineages (Figure 5B), although the snake clade has the longest branches of any tetrapod group. Certain snake branches (e.g., the branch leading to all snakes and to the Alethinophidia) are disproportionately long compared to branch lengths based on rRNAs (Figure 5). To evaluate this further, branch lengths were calculated for different genes and gene clusters and there was considerable variation among genes with respect to relative branch lengths in early snake lineages (data not shown).

To qualitatively elucidate the spatio-temporal dynamics in rates of substitution between gene regions that occur across branches, we plotted the branch lengths derived from rRNAs (which appear to have had little or no acceleration; e.g., Figure 5A) versus the branch lengths of various genes and gene clusters (Figure 6). All gene pairs generally appear to have highly correlated branch lengths (Figure 6), but some branches are outside the main distribution. Two branches consistently below the main distribution in most comparisons are the terminal branch

leading to *O. okinavensis* and the branch leading to the Acro-Heno clade (Figure 6); these two branches are also disproportionately longer in the rRNA trees than in the protein trees (Figure 5). These branches appear to have experienced acceleration of rRNA genes well beyond the mild accelerated evolution of rRNA that occurred along the lineages leading to all snakes and to the Alethinophidia.

To further evaluate the variation in spatio-temporal dynamics of relative rates of substitution across the mt genome, we used sliding window analyses of branch-specific and group-specific patterns of relative substitution rates. Only one of these comparisons, that of the Acro-Heno clade terminal branches, shows little variation of standardized relative substitution rates across the genome (Figure 7C). This suggests that the distribution of substitutions across the mt genome of terminal lineages within the Acro-Heno clade is nearly identical to the distribution across the mt genome of other tetrapods, and that these terminal snake lineages are not undergoing region or gene-specific selection. The plots for terminal colubroid branches are also fairly flat except for the downstream half of the 16s rRNA (Figure 7B), which may be entirely attributable to acceleration of the 16s rRNA in *O. okinavensis*, as discussed earlier.

Cross-referencing results from Figures 5, 6, 7, we can summarize the apparent nucleotide evolutionary rate dynamics in snake mt genomes as follows (see also Figure 4). The branch leading to all extant snakes appears to have experienced accelerated evolution in the region starting near the end of COX1 through COX2, ATP8, and somewhat into ATP6, and also in the region including the end of ND5, ND6, and CytB (and a rise in ND1). The COX1, COX2, ATP8, and ND6 accelerations increased and were stronger in the branch leading to Alethinophidia, while the ND5 acceleration decreased, and a notable acceleration of CytB also occurred. In the branch leading to the Colubroidea, only the ND6 acceleration continued, but new rate peaks arose in ND5, 12s rRNA, and the first part of the 16s rRNA, followed by a strong dropoff in all gene-specific acceleration in terminal colubroid lineages, except in the end of 16s rRNA in *O. okinavensis*. In the branch leading to the Acro-Heno clade, the accelerated rates of evolution (in COX1, COX2, ATP8, and ND5 genes) observed along the branch leading to the alethinophidians diminished (except for ND6 as in the Colubroidea), but new rate peaks arose in ATP6, COX3, ND3, ND4L, and the latter half of the 16s rRNA. These punctuated gene-specific accelerations were followed by the complete elimination of all gene-specific signals of atypical relative rate in terminal lineages within this Acro-Heno clade. We find no evidence for a constant accelerated rate of snake mt genome evolution. Instead, our analyses of

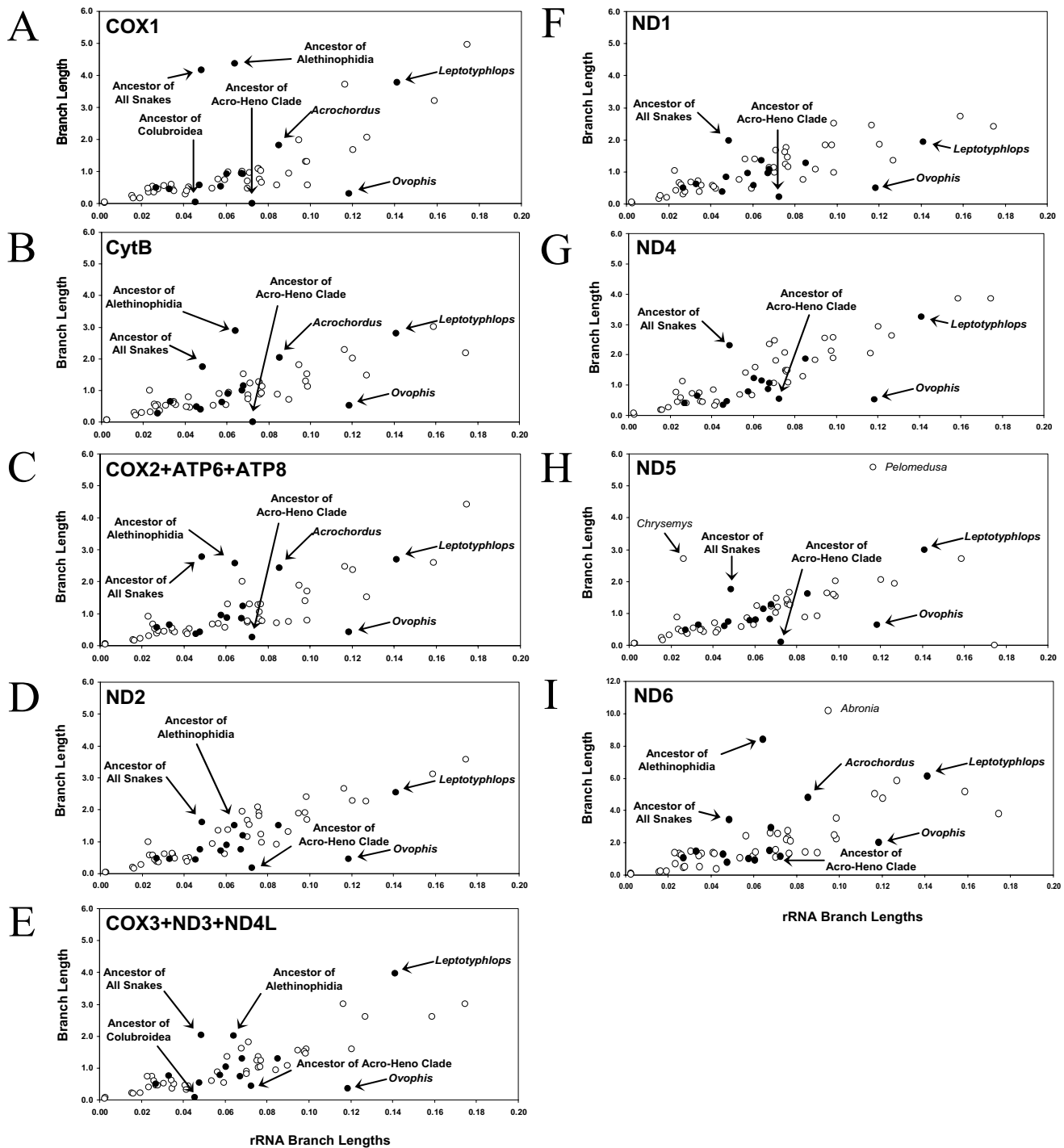


Figure 6
Plot of branch lengths obtained from rRNA versus various genes and gene clusters. Branches leading to the most recent common ancestor of a group are labeled e.g. "Ancestor of All Snakes". Snake branches are indicated with filled circles, and non-snake tetrapod branches are indicated with an unfilled circle. The locations of selected snake branches are labeled (in bold) with arrows. Outlying non-snake branches are indicated and labeled in normal type. Genes and gene clusters shown are (A) COX1, (B) CytB, (C) COX2 + ATP6 + ATP8, (D) ND2, and (E) COX3 + ND3 + ND4L, (F) ND1, (G) ND4, (H) ND5, (I) ND6.

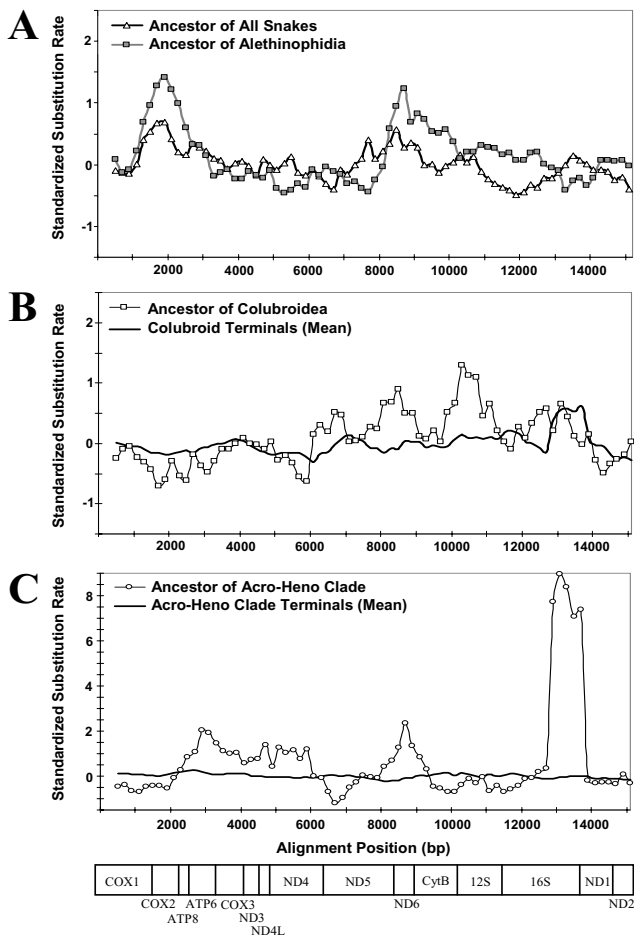


Figure 7
Standardized substitution rates across the mitochondrial genome for selected branches or clusters. For each 1000 bp window applied to a set of branches, standardized substitution rates were obtained by first dividing by the median window value for that branch, and then subtracting this value from the average across all non-snake branches. This helps to visualize regions of the genome that are evolving at slower or faster rates, with the average tetrapod relative rate being zero. Branches or branch sets shown are (A) the branch leading to the most recent common ancestor of all snakes and of the Alethinophidia; (B) the branch leading to the most recent common ancestor of the Colubroidea and the sum of all colubroid terminal branches; and (C) the branch leading to the most recent common ancestor of the Acro-Heno clade and the sum of all Acro-Heno clade terminal branches.

rates and patterns of substitution underscore both the spatial (gene-specific) and temporal (branch-specific) nature of molecular evolutionary relative rate dynamics in snake mt genomes.

Discussion

The three new complete snake mt genomes presented here, together with previously existing vertebrate mt genomes, provide a preliminary perspective on a complex history of potentially adaptive mt genomic change in snakes. Unusual changes in gene size and nucleotide substitution rates are associated with changes in mt genomic architecture (Figure 4). Nevertheless, the changes in substitution dynamics cannot be directly explained by the changes in mt genome architecture. Snake mt genome evolution is most consistent with some type of broad selective pressure on the efficiency and function of oxidative metabolism in snakes early in their evolutionary history.

In mt genomes (particularly in vertebrates), the processes of replication and transcription are not entirely functionally independent, and genome structural organization plays a prominent role in both processes. The CR acts as the origin of heavy strand replication, in addition to its role as the promoter for both heavy and light strand transcription [18]. Genome replication also depends on the processing of light strand transcripts to produce short primers required for heavy strand initiation of genome replication (originating from the CR [19]). The regular distribution of the tRNA genes throughout the mt genome is functionally significant, and these play an important role in RNA processing of polycistrons to yield mature RNAs, transcription initiation and termination, as well as initiation of light strand replication [18]. Collectively, many functional ramifications are linked tightly to mt genome architecture in vertebrates.

Mitochondrial genome size reduction due to gene shortening in alethinophidians is more than offset by the retention of their duplicate control regions. If size reduction is caused by selective pressure, the long term retention of dual CRs suggests that having both copies provide some selective advantage. Although the duplicate control region appears to function in heavy strand replication in at least some snakes, there is considerable variation in CR usage across snake lineages (Table 1). Thus, if the duplication has been maintained by selection, control of replication may not be the singular or primary selective driving force.

The possession of two functional control regions in most snake mt genomes might be advantageous by increasing the rate at which genome replication proceeds, and/or increasing the overall number of genome copies per mitochondrion. Since the dual CRs essentially flank the rRNA genes, they (along with adjacent tRNAs) could also plausibly function to independently control rates of protein-coding and rRNA gene transcription. Across snake species, variation in the tRNAs flanking the CRs includes the translocation of tRNA^{Leu} (3' of CR2) and the duplication/trans-

location/truncation of tRNA^{Pro}. In vertebrates, tRNA^{Leu} has been shown to decouple rates of rRNA and mRNA transcription by acting as a terminator of ~95% of heavy strand transcripts (leading to ~20-fold higher rRNA vs. mRNA levels; [18]). Considering the ectothermy of snakes, transcriptional decoupling via independent control regions could provide a more direct means of countering thermodynamic depression of enzymatic rates at low temperatures.

Independent CR duplications have also been identified in eels [20], frogs [21], birds [22,23], and lizards [24,25]. Our results (and additional unpublished data) suggest that the dramatic shifts in rates and patterns of molecular evolution in snakes represent a unique phenomenon that we do not expect to be necessarily associated with CR duplication, but rather more likely associated with selection for mitochondrial function. Nevertheless, these independent duplications may be useful to test the consequences of duplication on mutational processes.

Concerted evolution in and around the duplicate control regions

The two control regions clearly undergo concerted evolution to maintain reciprocal homogeneity between control regions within a genome [7-9], presumably through gene conversion. Interestingly, an apparently nonfunctional partial (or pseudo) proline tRNA (Ψ -tRNA^{Pro}) in colubrid mt genomes also appears to be maintained by concerted evolution (Figure 1). The gene conversion process that homogenizes the control region may also occasionally pick up extra DNA, making tRNA^{Pro}, or part of it, prone to duplication at this location. The existence of a duplicate tRNA^{Phe} between CR2 and tRNA^{Leu} in the viperid *O. okinavensis* [7] suggests that frequent gene duplication adjacent to the CRs may occur (these two tRNA^{Phe} differ by only 3 of 64 bp; implying either concerted evolution or recent duplication). The concerted evolution of these tRNAs could be explained by a tendency for gene conversion events involving the duplicate control regions to extend into the homologous flanking tRNA regions.

Another point of interest concerning gene conversion that arises from this study is a preliminary indication of differential evolutionary processes operating on the CRs within versus between species. Vertebrate mitochondrial control regions typically evolve very rapidly, and this is the case in a comparison of the two viperid species (*O. okinavensis* and *A. piscivorus*) in which CRs from these species are (on average) approximately as divergent as the fastest evolving positions within the mt genome, third codon positions (Figure 2B). In contrast, the two *A. piscivorus* genomes, *Api1* and *Api2*, have surprisingly similar CRs between individuals (Figure 2A; Additional file 2), comparable to the similarity between rRNA genes, among the slowest

evolving regions in the mt genome. A previous study on viperid snakes also showed slow within-species CR evolutionary rates [26], and other studies have demonstrated particularly slow intra-species rates and differential rates of CR evolution operating within versus between species in birds [27] and fish [28].

In this study we have found a great deal of rate heterogeneity among genes, so it is certainly possible that the normally unconserved control regions have become suddenly critical and conserved in *A. piscivorus*. Alternatively, it is plausible that the complex (and poorly understood) process of gene conversion of CRs within a genome may also alter rates of CR evolution within species through a yet unknown process of gene conversion that may involve intragenomic (or even intergenomic) recombination.

Comparative rates of molecular evolution

Previous studies have suggested that snake mt genomes have an accelerated rate of evolution [7,8]. Our results suggest this general conclusion is an oversimplification of a much more complex scenario, and that rates of snake mt genome evolution incorporate broad temporal (branch-specific) and spatial (gene and gene region-specific) dynamics. Branches early in snake evolution appear to be associated with dramatically elevated evolutionary rates and extreme relative rate dynamics across the mt genome (Figure 4). In contrast, terminal branches appear to have patterns of mt genome evolution that are strikingly similar to other (non-snake) vertebrates.

In support of a hypothesis involving selection for overall oxidative metabolic function, the accelerated rates of molecular evolution in snakes appear to depend greatly on gene function, with most ND subunits accelerating only slightly and occasionally, while COX, ATP, CytB, and rRNA evolutionary accelerations are dramatic and punctuated. The roles of these proteins (and the mitochondria in general) in energetics via oxidative phosphorylation are well known, and it may be that a single causative agent accompanying the diversification of snakes that dramatically altered metabolic demand, or led to a fluctuation in metabolic demand, was responsible for large-scale changes in selective pressure on these proteins.

Conclusion

Snake mitochondrial genomes present a rare opportunity to investigate the evolutionary interactions and ramifications that link genome architecture, molecular evolution, and multi-level molecular function. Available evidence points to selective pressures acting at many hierarchical levels within snake mt genomes, and at different times during snake evolution, leading to diverse, dramatic, and broad-scale changes in the genome. Interestingly, some consequences of this adaptive shift appear to have dimin-

ished over time (e.g., accelerated evolutionary rates of COX and other genes), whereas others appear to continue in extant snakes (i.e., the effects of control region duplication on mutation gradients, replication, and potentially transcription, and remnant functional consequences of short and highly substituted genes). Although the precise causes are unknown, this outstanding example of an apparent punctuated adaptive shift involving multiple aspects of genome architecture evolution provides an important comparative tool for the study of vertebrate mt genome evolution.

Methods

Sampling, sequencing and annotation

DNA was extracted from vouchered specimens available at the Louisiana State University Museum of Natural Science (LSUMZ) and the University of Central Florida (CLP). The *Agkistrodon piscivorus* (cottonmouth or water moccasin; Viperidae) specimens were from Louisiana, USA (LSUMZ-17943) and from Florida, USA (CLP-73). We refer to these as *Api1* (Louisiana specimen) and *Api2* (Florida specimen). The *Pantherophis slowinskii* (corn snake; Colubridae) specimen was from Louisiana, USA (LSUMZ- H-2036). The genus *Pantherophis* [29] was recently erected to contain a clade of species formerly allocated to *Elaphe*. The species *P. slowinskii* was formerly considered *Pantherophis (Elaphe)guttatus*, and was recently recognized as a distinct species [30]. Details of molecular laboratory methods (e.g., PCR, cloning, sequencing), genome annotation [31], and accession numbers are provided in Additional files (see Additional files 2 and 4).

Phylogenetic and sliding-window analyses

In addition to the three new snake mt genome sequences, the sequence dataset used included all eight snake mt genomes available at the time of the study, and 42 additional taxa for comparative purposes, including heavy sampling of birds, mammals (mostly primates), and lizards (species scientific names and access numbers are given in Additional file 2). Sequences of protein-coding and rRNA genes were aligned using ClustalX [32], followed by manual adjustment. Protein-coding genes were first aligned at the amino acid level, and then the nucleotide sequences were aligned according to the corresponding amino acid alignment. The alignment of rRNAs contained a small number of sites (corresponding to the loop-forming structures of the rRNAs) with somewhat ambiguous alignments only among major tetrapod lineages. Since we wanted to compare estimates of mitochondrial gene evolutionary rates and patterns, we chose not to exclude any sites of the alignment. This was also justified by preliminary phylogenetic estimates that suggested the incorporation of these few potentially ambiguous sites did not affect phylogenetic results. The main phylogeny presented here was inferred using the concatenated

nucleotide sequence of all 13 protein-coding and two rRNA genes by maximum-likelihood (ML) analysis in PAUP 4.0 beta10 [33]. This analysis used the GTR + Γ + I model of evolution, the best-fit model under all criteria in ModelTest [34].

Support for this topology was evaluated in two ways: (1) based on 1000 NJ bootstraps (in PAUP) with ML distances calculated under the same model as above, but with down-weighted synonymous sites to avoid saturation problems (rRNAs relative weight = 5 and 1st, 2nd, and 3rd codon positions relative weights = 4, 5, and 1) and (2) based on Bayesian posterior probability support estimated by conducting two simultaneous independent MCMC runs conducted for 10⁶ generations (with the first 400,000 generations of each run discarded as burn-in) using a GTR + Γ + I model of evolution (in MrBayes 3.1 [17]). The burnin period was determined by visual assessment of stationarity and convergence of likelihood values between the chains. To analyze nucleotide substitution rate variation in different lineages and different genes, branch length estimates were separately calculated under the GTR + Γ + I model for different genes (COX1, ND1, ND2, ND4, ND5, CytB) and gene clusters (COX2 + ATP8 + ATP6, and COX3 + ND3 + ND4L; each comprising groups of individually short genes adjacent along the mt genome) using the ML topology and PAML [35]. To further analyze fluctuations in nucleotide substitution rates, we conducted sliding window analyses (SWA) on the phylogenetic dataset. The program Hyphy [36] was used to estimate branch lengths (estimated numbers of substitutions) for 1000 bp windows. SWA was conducted using the GTR model with global parameter estimation and topological relationships specified based on the ML tree estimate, with a window slide of 200 bp. Based on preliminary trials, the size of the window and slide length were chosen to minimize noise observed with shorter windows, but to allow differentiation of patterns in different regions. To compare patterns of substitution across the mitochondrial genome for select branches or groups of branches, we first divided substitution estimates for each window by the median substitution rate across all windows. Since branch lengths are estimates of $\delta_b t_b$ (the branch-specific substitution rate times divergence time) this procedure estimates a ratio of substitution rates, $\delta_b^w / \delta_b^\xi$, where δ_b^w is the branch- and window-specific substitution rate, and δ_b^ξ is the branch-specific substitution rate in the median window. To evaluate whether the

windows had relative rates that were slower or faster than expected, we took the substitution rate ratio from the set of all branches in the non-snakes (NS) as a standard. This was then subtracted from the branch-specific ratio to obtain a "standardized substitution rate", $\delta_b^w / \delta_b^z - \delta_{NS}^w / \delta_{NS}^z$. When relative rates of substitution are distributed similarly across the mt genome, in comparison with NS, this standardized rate comparison approaches zero.

tRNA structure

The secondary structures of squamate tRNAs were determined under the guidance of the mammalian tRNA cloverleaf structures [37] and the tRNAscan program [38], and then used to modify tRNA alignments by hand (tRNA^{Ser} [AGY] was not included in these analyses because it does not form a cloverleaf structure). To determine the relative stabilities of the tRNA secondary structures, we calculated the energy (ΔG) of the cloverleaf structure using the Vienna Package version 1.4 [39].

Analysis of control region functionality

The calculation of T_{AMS} differs depending on whether CR1 or CR2 is functional, but only for the genes that are positioned between the two control regions, the two rRNAs and ND1 (see Additional file 2). Based on previous work, the light strand C/T ratio at synonymous two-fold and fourfold redundant 3rd codon positions is expected to increase linearly with T_{AMS} , so we used this prediction to determine whether there was any evidence for activity of CR1 or CR2 in initiating heavy strand replication. We implemented a slightly modified version of the MCMC approach in [3] to estimate the most likely slope and intercept of the C/T ratio gradient depending on the calculated T_{AMS} at every site. We applied these calculations using T_{AMS} from CR1 and CR2, and also separately calculated the slope and intercept for the most likely weighted average T_{AMS} for the two control regions. Other than the addition of the weighting parameter, all details of the Markov chain were as in [3]. Relative support for alternative hypotheses was determined using Akaike Information Criterion (AIC) and Akaike weights [40,41].

Abbreviations

rRNA, tRNA : ribosomal RNA, transfer RNA

mt: mitochondrial

O_H: origin of heavy strand replication

CR, CR1, CR2: control region, control region 1, control region 2

O_L: origin of light strand replication

ND#: NADH dehydrogenase subunit #

COX#: Cytochrome C oxidase subunit #

D_{ssH}: Duration of time spent single-stranded by the heavy strand during replication

T_{AMS}: Time spent in an asymmetric mutagenic state during replication

C, T, A, G: cytosine, thymine, adenine, guanine

CytB: cytochrome b

ATP#: ATP synthase subunit #

Ile, Met, Pro, Thr, Leu, Phe, Ser: isoleucine, methionine, proline, threonine, leucine, phenylalanine, serine

SWA: sliding window analysis

MYA: million years ago

LSUMZ: Louisiana State University Museum of Natural Science specimen tag

CLP: University of Central Florida specimen tag

Api1, Api2: *Agkistrodon piscivorus* specimen #

NS: non-snakes

Authors' contributions

ZJJ co-wrote the manuscript, performed much of the data analysis, and participated in sequencing of *P. slowinskii* and Api1. TAC co-wrote the manuscript, performed much of the data analysis, and participated in sequencing of Api2. CCA helped manage the project and assisted in writing and editing the manuscript. FTB performed the primary sequencing of *P. slowinskii* and Api1 and edited the manuscript. MDH performed the preliminary sequencing of Api2 and edited the manuscript. JAM contributed to the design and conception of the project and edited the manuscript. CLP supervised the sequencing of Api2 and edited the manuscript. DDP co-wrote the manuscript, designed and conceived the project, and supervised the sequencing of *P. slowinskii* and Api1 and the analysis of the data.

Additional material

Additional file 1

Mitochondrial genome replication and substitution gradients background. Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-123-S1.pdf]

Additional file 2*Laboratory and genome annotation methods.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-7-123-S2.pdf]

Additional file 3*Comparison of gene lengths in snakes and other squamates.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-7-123-S3.pdf]

Additional file 4*All supplementary tables.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-7-123-S4.pdf]

Acknowledgements

We thank Sameer Raina for modifying his program on Bayesian analysis of mitochondrial genome gradients to apply to this project, Jeremiah Faith for help with the preliminary annotations, Wanjun Gu for running some gradient analyses for us, and Judith Beekman for critical comments on the manuscript. This work was primarily supported by grants to D.D.P. from the State of Louisiana Board of Regents (Research Competitiveness Subprogram LEQSF (2001-04)-RD-A-08) and to C.L.P. from a UCF startup package and a National Science Foundation grant (DEB-0416000). It was also in part supported by grants to D.D.P. from the National Institutes of Health (GM065612-01 and GM065580-01), the National Science Foundation through Louisiana EPSCOR and the Center for Biomolecular Multi-scale Systems, and the State of Louisiana Board of Regents Millennium Research Program's Biological Computation and Visualization Center and Governor's Biotechnology Initiative.

References

- Krishnan NM, Raina SZ, Pollock DD: **Analysis of among-site variation in substitution patterns.** *Biol Proced Online* 2004, **6**:180-188.
- Krishnan NM, Seligmann H, Raina SZ, Pollock DD: **Detecting gradients of asymmetry in site-specific substitutions in mitochondrial genomes.** *DNA Cell Biol* 2004, **23**(10):707-714.
- Raina SZ, Faith JJ, Disotell TR, Seligmann H, Stewart CB, Pollock DD: **Evolution of base-substitution gradients in primate mitochondrial genomes.** *Genome Res* 2005, **15**(5):665-673.
- Faith JJ, Pollock DD: **Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes.** *Genetics* 2003, **165**(2):735-745.
- Reyes A, Gissi C, Pesole G, Saccone C: **Asymmetrical directional mutation pressure in the mitochondrial genome of mammals.** *Mol Biol Evol* 1998, **15**(8):957-966.
- Tanaka M, Ozawa T: **Strand asymmetry in human mitochondrial-DNA mutations.** *Genomics* 1994, **22**(2):327-335.
- Dong S, Kumazawa Y: **Complete mitochondrial DNA sequences of six snakes: Phylogenetic relationships and molecular evolution of genomic features.** *J Mol Evol* 2005, **61**(1):12-22.
- Kumazawa Y, Ota H, Nishida M, Ozawa T: **The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions.** *Genetics* 1998, **150**(1):313-329.
- Kumazawa Y, Ota H, Nishida M, Ozawa T: **Gene rearrangements in snake mitochondrial genomes: Highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster.** *Mol Biol Evol* 1996, **13**(9):1242-1254.
- Kumazawa Y: **Mitochondrial DNA sequences of five squamates: phylogenetic affiliation of snakes.** *DNA Res* 2004, **11**(2):137-144.
- Douglas D, Janke A, Arnason U: **A mitogenomic study on the phylogenetic position of snakes.** *Zool Scr* 2006, **35**(6):545-558.
- Reyes A, Yang MY, Bowmaker M, Holt IJ: **Bidirectional replication initiates at sites throughout the mitochondrial genome of birds.** *J Biol Chem* 2005, **280**(5):3242-3250.
- Pesole G, Gissi C, De Chirico A, Saccone C: **Nucleotide substitution rate of mammalian mitochondrial genomes.** *J Mol Evol* 1999, **48**(4):427-434.
- Ingman M, Kaessmann H, Paabo S, Gyllenstein U: **Mitochondrial genome variation and the origin of modern humans.** *Nature* 2000, **408**(6813):708-713.
- Gower DJ, Vidal N, Spinks JN, McCarthy CJ: **The phylogenetic position of Anomochilidae (Reptilia: Serpentes): first evidence from DNA sequences.** *Journal of Zoological Systematics and Evolutionary Research* 2005, **43**(4):315-320.
- Lawson R, Slowinski JB, Crother BI, Burbrink FT: **Phylogeny of the Colubroidea (Serpentes): New evidence from mitochondrial and nuclear genes.** *Mol Phylogenet Evol* 2005, **37**(2):581-601.
- Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**(12):1572-1574.
- Fernandez-Silva P, Enriquez JA, Montoya J: **Replication and transcription of mammalian mitochondrial DNA.** *Exp Physiol* 2003, **88**(1):41-56.
- Clayton DA: **Replication of animal mitochondrial DNA.** *Cell* 1982, **28**(4):693-705.
- Inoue JG, Miya M, Tsukamoto K, Nishida M: **Evolution of the deep-sea gulper eel mitochondrial genomes: Large-scale gene rearrangements originated within the eels.** *Mol Biol Evol* 2003, **20**(11):1917-1924.
- Sano N, Kurabayashi A, Fujii T, Yonekawa H, Sumida M: **Complete nucleotide sequence of the mitochondrial genome of Schlegel's tree frog *Rhacophorus schlegelii* (family Rhacophoridae): duplicated control regions and gene rearrangements.** *Genes Genet Syst* 2005, **80**(3):213-224.
- Abbott CL, Double MC, Trueman JWH, Robinson A, Cockburn A: **An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in *Thalassarche albatrosses*.** *Mol Ecol* 2005, **14**(11):3605-3613.
- Eberhard JR, Wright TF, Bermingham E: **Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona*.** *Mol Biol Evol* 2001, **18**(7):1330-1342.
- Amer SAM, Kumazawa Y: **Mitochondrial genome of *Pogona vitticeps* (Reptilia; Agamidae): control region duplication and the origin of Australasian agamids.** *Gene* 2005, **346**:249-256.
- Kumazawa Y, Endo H: **Mitochondrial genome of the Komodo dragon: Efficient sequencing method with reptile-oriented primers and novel gene rearrangements.** *DNA Res* 2004, **11**(2):115-125.
- Ashton KG, de Queiroz A: **Molecular systematics of the western rattlesnake, *Crotalus viridis* (Viperidae), with comments on the utility of the D-loop in phylogenetic studies of snakes.** *Mol Phylogenet Evol* 2001, **21**(2):176-189.
- Crochet PA, Desmarais E: **Slow rate of evolution in the mitochondrial control region of gulls (Aves: Laridae).** *Mol Biol Evol* 2000, **17**(12):1797-1806.
- Tang Q, Liu H, Mayden R, Xiong B: **Comparison of evolutionary rates in the mitochondrial DNA cytochrome b gene and control region and their implications for phylogeny of the Cobitoidea (Teleostei: Cypriniformes).** *Mol Phylogenet Evol* 2006, **39**(2):347-57. Epub 2005 Oct 4.
- Utiger U, Helfenberger N, Schättli B, Schmidt C, Ruf M, Ziswiler V: **Molecular systematics and phylogeny of Old World and New World ratsnakes, *Elaphe* Auct., and related genera (Reptilia, Squamata, Colubridae).** *Russian Journal of Herpetology* 2002, **9**(2):105-124.
- Burbrink FT: **Phylogeographic analysis of the cornsnake (*Elaphe guttata*) complex as inferred from maximum likelihood and Bayesian analyses.** *Mol Phylogenet Evol* 2002, **25**(3):465-476.
- Slack KE, Janke A, Penny D, Arnason U: **Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features.** *Gene* 2003, **302**(1-2):43-52.

32. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25(24)**:4876-4882.
33. Swofford DL: **PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods).** Sunderland, Massachusetts, Sinauer Associate; 1997.
34. Posada D, Crandall KA: **ModelTest: testing the model of DNA substitution.** *Bioinformatics* 1998, **14(9)**:817-818.
35. Yang ZH: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13(5)**:555-556.
36. Pond SLK, Frost SDW, Muse SV: **HyPhy: hypothesis testing using phylogenies.** *Bioinformatics* 2005, **21(5)**:676-679.
37. Helm M, Brule H, Friede D, Giege R, Putz D, Florentz C: **Search for characteristic structural features of mammalian mitochondrial tRNAs.** *RNA* 2000, **6(10)**:1356-1379.
38. Lowe TM, Eddy SR: **tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 1997, **25(5)**:955-964.
39. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P: **Fast folding and comparison of RNA secondary structures.** *Monatshefte Fur Chemie* 1994, **125(2)**:167-188.
40. Akaike H: **Information theory and an extension of the maximum likelihood principle.** In *Second International Symposium on Information Theory* Edited by: Petrov BN, Csake F. Budapest, Akademia Kiado; 1973:673-681.
41. Akaike H: **Information measures and model selection.** *Int Stat Inst* 1983, **22**:277-291.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



Additional File 2 (Supplementary Background)

Supplementary Background. Mitochondrial Replication and Transition Biases.

The mt genome has long been believed to replicate asymmetrically [1], which strongly contributes to the substantial difference in mutation rates and nucleotide composition biases between strands [2-6]. During replication under the classical model, the synthesis of the nascent heavy strand initiates at the origin of heavy strand replication (O_H), within the control region (CR). This has been reviewed extensively elsewhere [2, 7], but in brief, after two thirds of the nascent heavy strand is synthesized, the synthesis of the nascent light strand starts at the origin of light strand replication (O_L), a short secondary-structure-forming segment located within the tRNA cluster (the WANCY region) between the NADH dehydrogenase subunit 2 (ND2) and Cytochrome C oxidase subunit 1 (COX1) genes. The strand-asymmetric replication mechanism has been thought to expose different regions of the parental heavy strand to varying amounts of time in the single-stranded state during replication (D_{ssH} ; [6]), depending on the distances of the regions from the O_H and O_L . Variation in this strand-asymmetric replication process appears to have contributed substantially to variation in substitution rates among genes [2, 7, 8].

Controversy has arisen recently over the classical mt genome replication mechanism, mostly concerning the asymmetry of the process, the role of the putative origin of light strand replication, and whether the replicating DNA spends substantial amounts of time single-stranded [9-11]. Although newly proposed models of replication that embody these concerns are directly at odds with the genetic data, one of us has hypothesized

(Pollock, *in review*) that most of the biochemical and genetic data are compatible with a reconciled model of mt genome replication, which retains most critical features of the classical model except for single strandedness. Regardless of the final reconciliation, here we take a neutral position on the biochemical issue of single-strandedness by referring to the time that a gene or nucleotide is predicted to spend in an asymmetric mutagenic state (T_{AMS}), rather than the predicted duration of time that the heavy strand spends single-stranded (D_{SSH}); the calculation of T_{AMS} is, however, identical to that for D_{SSH} [6, 7, 12].

Cytosine → Uracil deaminations are common in single-stranded DNA, while Adenine → Hypoxanthine deaminations are less common [13, 14]. These two deaminations lead to mutations (Cytosine → Thymine and Adenine → Guanine, or C → T and A → G) that appear to account for most of the asymmetry in synonymous substitutions found in vertebrate mt genome [7, 8, 12, 15-19]. C → T and A → G mutations on the heavy strand during replication apparently lead respectively to G → A and T → C substitutions (and G and T deficiencies) on the light strand. Most protein-coding genes (all but ND6) use the heavy strand as a template; thus, the mutation biases observed in the light strand parallel the biases in most protein-coding gene transcripts. Faith and Pollock [7] found that, in vertebrates, T → C light strand substitutions at four-fold and two-fold redundant 3rd codon positions increase linearly with increasing T_{AMS} . In contrast, G → A light strand substitutions increase rapidly but quickly reach a maximal level. Consequently, T → C substitutions and the resultant C/T nucleotide frequency gradient are good predictors of T_{AMS} .

References

1. Clayton DA: **Replication of animal mitochondrial DNA.** *Cell* 1982, **28**(4):693-705.
2. Bielawski JP, Gold JR: **Mutation patterns of mitochondrial H- and L-Strand DNA in closely related cyprinid fishes.** *Genetics* 2002, **161**(4):1589-1597.
3. Jermiin LS, Graur D, Crozier RH: **Evidence from analyses of intergenic regions for strand-specific directional mutation pressure in metazoan mitochondrial DNA.** *Mol Biol Evol* 1995, **12**(4):558-563.
4. Perna NT, Kocher TD: **Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes.** *J Mol Evol* 1995, **41**(3):353-358.
5. Perna NT, Kocher TD: **Unequal base frequencies and estimation of substitution rates.** *Mol Biol Evol* 1995, **12**(2):359-361.
6. Tanaka M, Ozawa T: **Strand asymmetry in human mitochondrial-DNA mutations.** *Genomics* 1994, **22**(2):327-335.
7. Faith JJ, Pollock DD: **Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes.** *Genetics* 2003, **165**(2):735-745.
8. Raina SZ, Faith JJ, Disotell TR, Seligmann H, Stewart CB, Pollock DD: **Evolution of base-substitution gradients in primate mitochondrial genomes.** *Genome Res* 2005, **15**(5):665-673.
9. Reyes A, Yang MY, Bowmaker M, Holt IJ: **Bidirectional replication initiates at sites throughout the mitochondrial genome of birds.** *J Biol Chem* 2005, **280**(5):3242-3250.
10. Yang MY, Bowmaker M, Reyes A, Vergani L, Angeli P, Gringeri E, Jacobs HT, Holt IJ: **Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication.** *Cell* 2002, **111**(4):495-505.
11. Yasukawa T, Yang MY, Jacobs HT, Holt IJ: **A bidirectional origin of replication maps to the major noncoding region of human mitochondrial DNA.** *Mol Cell* 2005, **18**(6):651-662.
12. Reyes A, Gissi C, Pesole G, Saccone C: **Asymmetrical directional mutation pressure in the mitochondrial genome of mammals.** *Mol Biol Evol* 1998, **15**(8):957-966.
13. Frederico LA, Kunkel TA, Shaw BR: **A sensitive genetic assay for the detection of cytosine deamination - determination of rate constants and the activation energy.** *Biochemistry (Mosc)* 1990, **29**(10):2532-2537.
14. Impellizzeri KJ, Anderson B, Burgers PMJ: **The spectrum of spontaneous mutations in a *Saccharomyces cerevisiae* Uracil-DNA-Glycosylase mutant limits the function of this enzyme to cytosine deamination repair.** *J Bacteriol* 1991, **173**(21):6807-6810.
15. Bielawski JP, Gold JR: **Unequal synonymous substitution rates within and between two protein-coding mitochondrial genes.** *Mol Biol Evol* 1996, **13**(6):889-892.

16. Frank AC, Lobry JR: **Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms.** *Gene* 1999, **238**(1):65-77.
17. Krishnan NM, Raina SZ, Pollock DD: **Analysis of among-site variation in substitution patterns.** *Biological Procedures Online* 2004, **6**:180-188.
18. Krishnan NM, Seligmann H, Raina SZ, Pollock DD: **Detecting gradients of asymmetry in site-specific substitutions in mitochondrial genomes.** *DNA Cell Biol* 2004, **23**(10):707-714.
19. Rand DM, Kann LM: **Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA.** *Genetica* 1998, **103**:393-407.

Additional File 2 (Supplementary Tables)

Supplementary Table S1. Detailed genome annotation of *Agkistrodon piscivorus*.

Supplementary Table S2. Detailed genome annotation of *Pantherophis slowinskii*.

Supplementary Table S3. Gene-specific polymorphisms observed between the two *Agkistrodon piscivorus* genomes (*Api1* and *Api2*)

Supplementary Table S4. Polymorphisms observed in tRNA genes between *Agkistrodon piscivorus* genomes (*Api1* and *Api2*).

Supplementary Table S5. Nucleotide frequencies of mitochondrial genome regions (rRNA and protein-coding genes) used in phylogenetic and molecular evolutionary analyses in this study.

Supplementary Table S6. Estimated T_{AMS} values of genes for squamates. Two T_{AMS} values are given for each species of alethinophidian snakes; T_{AMS}^1 is estimated based on the assumption of exclusive CR1 usage, whereas T_{AMS}^2 is estimated based on exclusive CR2 usage. Genes that have alternative T_{AMS} estimates under different CR usage scenarios in alethinophidian mtDNAs are indicated in bold.

Supplementary Table S7. Energy (*kcal/mol*) of the cloverleaf structures of tRNAs in snakes and lizards.

Supplementary Table S8. Complete mitochondrial genomes used in this study, and associated Genbank accession numbers.

Supplementary Table S9. Primer sets used to amplify mitochondrial genome fragments in this study.

Supplementary Table S1. Detailed genome annotation of *Agkistrodon piscivorus*.

	From	To	Size	Strand	Codon	StartCodon	StopCodon
Phe	1	65	65	L	TTC		
12sRNA	62	976	915	-			
Val	977	1040	64	L	GTA		
16sRNA	1041	2527	1487	-			
ND1	2528	3488	961	L		ATC	T
Ile	3489	3556	68	L	ATC		
Pro	3560	3622	63	H	CCA		
CR1	3623	4642	1020	-			
Leu	4643	4715	73	L	TTA		
Gln	4716	4785	70	H	CAA		
Met	4786	4848	63	L	ATG		
ND2	4849	5878	1030	L		ATA	T
Trp	5879	5944	66	L	TGA		
Ala	5945	6009	65	H	GCA		
Asn	6010	6081	72	H	AAC		
O_L	6084	6117	34	-			
Cys	6116	6175	60	H	TGC		
Tyr	6176	6236	61	H	TAC		
COX1	6238	7839	1602	L		GTG	AGA
Ser4	7830	7897	68	H	TCA		
Asp	7898	7960	63	L	GAC		
COX2	7962	8646	685	L		ATG	T
Lys	8647	8710	64	L	AAA		
ATP8	8711	8875	165	L		ATG	TAA
ATP6	8866	9546	681	L		ATG	TAA
COX3	9546	10329	784	L		ATG	T
Gly	10330	10390	61	L	GGA		
ND3	10391	10733	343	L		ATC	T
Arg	10734	10797	64	L	CGA		
ND4L	10798	11087	290	L		ATG	TA
ND4	11088	12425	1338	L		ATG	AGA
His	12426	12487	62	L	CAC		
Ser2	12488	12542	55	L	AGC		
Leu4	12543	12614	72	L	CTA		
ND5	12616	14403	1788	L		ATG	TAA
ND6	14399	14908	510	H		GTG	AGG
Glu	14918	14980	63	H	GAA		
CytB	14981	16094	1114	L		ATG	T
Thr	16095	16159	65	L	ACA		
Pseudo-Pro	16160	16190	31	-			
CR2	16191	17213	1019	-			

*standard amino acid abbreviations represent the tRNA that codes for that amino acid

Supplementary Table S2. Detailed genome annotation of *Pantherophis slowinskii*.

	From	To	Size (bp)	Strand	Codon	StartCodon	StopCodon
Phe*	1	60	60	L	TTC		
12sRNA	59	991	933	-			
Val	992	1054	63	L	GTA		
16sRNA	1055	2531	1477	-			
ND1	2532	3495	964	L		ATA	T
Ile	3496	3561	66	L	ATC		
Pseudo-Pro	3558	3592	35				
CR1	3593	4613	1021	-			
Leu2	4614	4686	73	L	TTA		
Gln	4689	4759	71	H	CAA		
Met	4761	4822	62	L	ATG		
ND2	4823	5852	1030	L		ATT	T
Trp	5853	5917	65	L	TGA		
Ala	5919	5981	63	H	GCA		
Asn	5983	6055	73	H	AAC		
O_L	6058	6093	36	-			
Cys	6092	6152	61	H	TGC		
Tyr	6153	6214	62	H	TAC		
COX1	6216	7817	1602	L		GTG	AGA
Ser4	7808	7874	67	H	TCA		
Asp	7875	7938	64	L	GAC		
COX2	7940	8624	685	L		ATG	T
Lys	8625	8688	64	L	AAA		
ATP8	8690	8848	159	L		ATG	TAA
ATP6	8839	9519	681	L		ATG	TAA
COX3	9519	10302	784	L		ATG	T
Gly	10303	10363	61	L	GGA		
ND3	10364	10706	343	L		GTG	T
Arg	10707	10771	65	L	CGA		
ND4L	10772	11061	290	L		ATG	TA
ND4	11062	12399	1338	L		ATG	TAA
His	12400	12464	65	L	CAC		
Ser2	12465	12521	57	L	AGC		
Leu4	12519	12589	71	L	CTA		
ND5	12590	14536	1947	L		ATG	ATT
ND6	14353	14853	501	H		ATG	TAG
Glu	14863	14924	62	H	GAA		
CytB	14923	16039	1117	L		ATG	T
Thr	16040	16103	64	L	ACA		
Pro	16104	16164	61	H	CCA		
CR2	16165	17189	1025	-			

*standard amino acid abbreviations represent the tRNA that codes for that amino acid

Supplementary Table S3. Gene-specific polymorphisms observed between the two *Agkistrodon piscivorus* genomes (*Api1* and *Api2*)

Genes	Length	Similarity	Substitutions				
			all	1st	2nd	3rd	AA
12s RNA	915	98.80%	11	-	-	-	-
16s RNA	1487	97.40%	39	-	-	-	-
ATP6	681	95.00%	32	5	2	25	4
ATP8	165	93.94%	11	3	1	7	3
COX1	1602	96.38%	58	0	1	57	2
COX2	685	96.50%	24	6	0	18	3
COX3	786	96.40%	28	6	1	21	5
CytB	1114	95.33%	52	10	3	39	10
ND1	960	96.46%	34	8	1	25	3
ND2	1030	96.12%	40	6	4	30	8
ND3	343	93.88%	21	2	6	20	8
ND4	1338	95.81%	56	9	3	44	5
ND4L	290	97.93%	6	2	0	4	2
ND5	1788	94.46%	96	21	9	69	28
ND6	510	95.00%	26	3	4	19	5
CR1	1021	98.20%	19	-	-	-	-
CR2	1022	98.40%	18	-	-	-	-

Supplementary Table S4. Polymorphisms observed in tRNA genes between *Agkistrodon piscivorus* genomes (*Api1* and *Api2*).

tRNA	Length	Similarity	Substitution location
<i>Phe</i>	65	96.92%	deletion of g in D-Loop and t-c in T-loop
<i>Val</i>	64	98%	t-c in T-Loop
<i>Ile</i>	68	92.65%	a-g g-a,c-t,t-c in T-Loop, and a-g in stem
<i>Pro</i>	63	100%	
<i>Leu</i>	73	100%	
<i>Gln</i>	70	100%	
<i>Met</i>	63	100%	deletion of a in D-arm
<i>Trp</i>	66	95.45%	g-a and a-g in anticodon arm, and g-t in T-Loop
<i>Ala</i>	65	98.46%	c-t in variable loop
<i>Asn</i>	72	100%	
<i>Cys</i>	60	96.67%	c-t in stem, t-c in T-Loop
<i>Tyr</i>	61	100%	
<i>Ser4</i>	68	98.53%	t-g in D-Loop
<i>Asp</i>	63	100%	
<i>Lys</i>	64	98.44%	deletion of t in T-Loop
<i>Gly</i>	61	100%	deletion of a in D-arm
<i>Arg</i>	64	98.44%	a-g in stem
<i>His</i>	62	98.39%	c-t in stem
<i>Ser2</i>	55	98.18%	t-g in D-Loop
<i>Leu4</i>	72	94.44%	c-t in stem, insertion of c in variable loop, a-g in anticodon stem, a-t in T-Loop
<i>Glu</i>	63	93.65%	t-g in D-stem, a-t, t-a and deletion of g in T-Loop
<i>Thr</i>	65	100%	

Supplementary Table S5. Nucleotide frequencies of mitochondrial genome regions (rRNA and protein-coding genes) used in phylogenetic and molecular evolutionary analyses in this study.

	Species	Entire Alignment				rRNA Genes				Protein-Coding Genes				
		A	C	G	T	A	C	G	T	A	C	G	T	
Amphibians	<i>Mertensiella luschani</i>	0.31	0.24	0.15	0.3	0.39	0.19	0.17	0.25	0.3	0.25	0.15	0.31	
	<i>Xenopus laevis</i>	0.31	0.24	0.14	0.31	0.35	0.23	0.18	0.24	0.31	0.24	0.13	0.32	
Mammals	<i>Bos taurus</i>	0.33	0.26	0.14	0.28	0.37	0.22	0.18	0.23	0.31	0.27	0.13	0.29	
	<i>Cebus albifrons</i>	0.32	0.26	0.13	0.29	0.35	0.24	0.17	0.24	0.31	0.27	0.12	0.3	
	<i>Gorilla gorilla</i>	0.3	0.3	0.14	0.26	0.35	0.26	0.18	0.22	0.29	0.31	0.13	0.27	
	<i>Homo sapiens</i>	0.3	0.31	0.14	0.25	0.34	0.26	0.18	0.22	0.29	0.32	0.13	0.26	
	<i>Hylobates lar</i>	0.3	0.31	0.14	0.25	0.34	0.27	0.18	0.21	0.29	0.32	0.14	0.25	
	<i>Lemur catta</i>	0.32	0.25	0.13	0.3	0.36	0.23	0.17	0.24	0.31	0.25	0.12	0.31	
	<i>Macaca sylvanus</i>	0.31	0.3	0.14	0.26	0.35	0.25	0.18	0.22	0.3	0.31	0.13	0.27	
	<i>Nycticebus coucang</i>	0.31	0.27	0.15	0.28	0.36	0.23	0.18	0.23	0.3	0.27	0.14	0.29	
	<i>Pan paniscus</i>	0.3	0.3	0.13	0.26	0.35	0.26	0.18	0.22	0.29	0.31	0.13	0.27	
	<i>Papio hamadryas</i>	0.3	0.3	0.14	0.26	0.34	0.25	0.18	0.23	0.3	0.31	0.13	0.26	
	<i>Pongo pygmaeus</i>	0.3	0.32	0.14	0.24	0.34	0.26	0.19	0.21	0.29	0.33	0.13	0.25	
	<i>Tarsius bancanus</i>	0.32	0.26	0.13	0.29	0.37	0.23	0.17	0.24	0.31	0.27	0.12	0.3	
Tuatara	<i>Sphenodon punctatus</i>	0.32	0.26	0.15	0.27	0.37	0.23	0.18	0.22	0.31	0.26	0.14	0.29	
Lizards	<i>Abronia graminea</i>	0.33	0.27	0.14	0.27	0.38	0.24	0.17	0.21	0.31	0.27	0.13	0.29	
	<i>Cordylus warreni</i>	0.3	0.3	0.14	0.25	0.35	0.26	0.18	0.21	0.29	0.31	0.14	0.27	
	<i>Eumeces egregius</i>	0.29	0.28	0.16	0.26	0.35	0.25	0.19	0.21	0.28	0.29	0.16	0.27	
	<i>Iguana iguana</i>	0.3	0.32	0.14	0.23	0.35	0.28	0.18	0.19	0.29	0.33	0.14	0.24	
	<i>Sceloporus occidentalis</i>	0.32	0.28	0.15	0.25	0.37	0.25	0.18	0.21	0.31	0.29	0.14	0.26	
	<i>Shinisaurus crocodilurus</i>	0.31	0.27	0.14	0.28	0.37	0.24	0.17	0.23	0.29	0.28	0.14	0.29	
	<i>Varanus komodoensis</i>	0.29	0.31	0.14	0.27	0.34	0.29	0.17	0.2	0.28	0.31	0.13	0.28	
	Snakes	<i>Acrochordus granulatus</i>	0.35	0.25	0.12	0.28	0.39	0.24	0.15	0.21	0.34	0.25	0.11	0.3
		<i>Agkistrodon piscivorus (Api1)</i>	0.32	0.29	0.14	0.25	0.37	0.25	0.17	0.21	0.31	0.3	0.13	0.26
		<i>Agkistrodon piscivorus (Api2)</i>	0.32	0.29	0.14	0.25	0.37	0.24	0.17	0.21	0.31	0.31	0.13	0.25
<i>Boa constrictor</i>		0.36	0.26	0.14	0.24	0.41	0.23	0.17	0.19	0.35	0.27	0.13	0.25	
<i>Cylindrophis ruffus</i>		0.35	0.27	0.13	0.25	0.39	0.25	0.16	0.2	0.34	0.27	0.12	0.27	
<i>Dinodon semicarinatus</i>		0.35	0.27	0.13	0.26	0.39	0.25	0.16	0.2	0.34	0.27	0.12	0.27	
<i>Leptotyphlops dulcis</i>		0.34	0.31	0.13	0.21	0.36	0.29	0.18	0.17	0.33	0.32	0.12	0.22	
<i>Ovophis okinavensis</i>		0.32	0.29	0.13	0.26	0.38	0.24	0.16	0.22	0.31	0.3	0.13	0.26	
<i>Pantherophis slowinskii</i>		0.35	0.25	0.13	0.26	0.39	0.23	0.16	0.21	0.34	0.26	0.13	0.28	
<i>Python regius</i>		0.33	0.29	0.13	0.25	0.38	0.26	0.16	0.2	0.32	0.3	0.13	0.26	
Turtles	<i>Xenopeltis unicolor</i>	0.34	0.27	0.13	0.26	0.38	0.24	0.16	0.21	0.33	0.27	0.13	0.27	
	<i>Chelonia mydas</i>	0.34	0.28	0.12	0.26	0.38	0.23	0.17	0.22	0.33	0.29	0.12	0.27	
	<i>Chrysemys picta</i>	0.33	0.26	0.13	0.27	0.39	0.23	0.17	0.21	0.32	0.27	0.13	0.28	
	<i>Dogania subplana</i>	0.34	0.26	0.13	0.26	0.39	0.23	0.16	0.22	0.33	0.27	0.12	0.27	
	<i>Pelomedusa subrufa</i>	0.32	0.27	0.13	0.28	0.36	0.24	0.17	0.23	0.32	0.28	0.12	0.29	
Crocodilians	<i>Alligator mississippiensis</i>	0.3	0.29	0.14	0.26	0.35	0.26	0.18	0.21	0.29	0.3	0.13	0.27	
	<i>Alligator sinensis</i>	0.28	0.31	0.16	0.25	0.33	0.27	0.2	0.2	0.27	0.32	0.15	0.26	
	<i>Caiman crocodilus</i>	0.31	0.31	0.14	0.25	0.34	0.27	0.19	0.2	0.3	0.32	0.13	0.26	
Birds	<i>Apteryx haastii</i>	0.3	0.3	0.14	0.26	0.35	0.25	0.18	0.22	0.29	0.31	0.13	0.27	
	<i>Buteo buteo</i>	0.29	0.33	0.14	0.24	0.33	0.28	0.19	0.21	0.28	0.34	0.13	0.24	
	<i>Ciconia boyciana</i>	0.29	0.33	0.15	0.23	0.33	0.28	0.19	0.2	0.28	0.34	0.14	0.24	
	<i>Ciconia ciconia</i>	0.29	0.33	0.15	0.23	0.33	0.28	0.19	0.2	0.28	0.34	0.14	0.24	
	<i>Corvus frugilegus</i>	0.3	0.3	0.15	0.25	0.33	0.24	0.21	0.22	0.29	0.31	0.14	0.26	
	<i>Dromaius novaehollandiae</i>	0.3	0.29	0.14	0.27	0.35	0.24	0.19	0.22	0.29	0.3	0.13	0.28	
	<i>Falco peregrinus</i>	0.3	0.32	0.14	0.24	0.34	0.28	0.19	0.19	0.29	0.33	0.13	0.25	
	<i>Gallus gallus</i>	0.29	0.33	0.14	0.24	0.33	0.28	0.18	0.2	0.28	0.34	0.13	0.25	
	<i>Rhea americana</i>	0.27	0.33	0.15	0.25	0.33	0.27	0.19	0.21	0.26	0.34	0.15	0.26	
	<i>Smithornis sharpei</i>	0.29	0.33	0.14	0.25	0.33	0.29	0.18	0.2	0.28	0.33	0.13	0.26	
	<i>Struthio camelus</i>	0.29	0.3	0.15	0.26	0.34	0.26	0.18	0.21	0.28	0.31	0.14	0.27	
	<i>Tinamus major</i>	0.29	0.3	0.14	0.27	0.34	0.26	0.18	0.22	0.28	0.31	0.13	0.28	
	<i>Vidua chalybeata</i>	0.3	0.31	0.16	0.23	0.33	0.25	0.21	0.21	0.3	0.32	0.15	0.24	
	Mean	0.31	0.29	0.14	0.26	0.36	0.25	0.18	0.21	0.3	0.3	0.13	0.27	

Supplementary Table S6. Estimated T_{AMS} values of genes for squamates. Two T_{AMS} values are given for each species of alethinophidian snakes; T_{AMS}^1 is estimated based on the assumption of exclusive CR1 usage, whereas T_{AMS}^2 is estimated based on exclusive CR2 usage. Genes that have alternative T_{AMS} estimates under different CR usage scenarios in alethinophidian mtDNAs are indicated in bold.

Genes	Snakes																				Lizards					
	<i>Agkistrodon</i>		<i>Ovophis</i>		<i>Pantherophis</i>		<i>Dinodon</i>		<i>Acrochordus</i>		<i>Boa</i>		<i>Cylindrophis</i>		<i>Python</i>		<i>Xenopeltis</i>		<i>Leptotyphlops</i>	<i>Iguana</i>	<i>Eumeces</i>	<i>Sceloporus</i>	<i>Cordylus</i>	<i>Abronia</i>	<i>Shinisaurus</i>	
	T_{AMS}^1	T_{AMS}^2	T_{AMS}^1	T_{AMS}^2	T_{AMS}^1	T_{AMS}^2	T_{AMS}^1	T_{AMS}^2	T_{AMS}^1	T_{AMS}^2	T_{AMS}^1	T_{AMS}^2	T_{AMS}^1	T_{AMS}^2	T_{AMS}^1	T_{AMS}^2	T_{AMS}^1	T_{AMS}^2	T_{AMS}	T_{AMS}	T_{AMS}	T_{AMS}	T_{AMS}	T_{AMS}		
12s	0.35	1.36	0.34	1.34	0.35	1.35	0.35	1.35	0.35	1.35	0.33	1.33	0.35	1.35	0.36	1.36	0.32	1.32	0.45	0.44	0.47	0.46	0.47	0.43	0.45	
16s	0.50	1.51	0.48	1.48	0.50	1.50	0.50	1.49	0.50	1.49	0.47	1.46	0.50	1.49	0.51	1.50	0.46	1.45	0.61	0.60	0.62	0.62	0.62	0.59	0.60	
ATP6	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.35	0.35	0.33	0.33	0.35	0.35	0.36	0.36	0.33	0.33	0.39	0.37	0.35	0.36	0.36	0.39	0.37
ATP8	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.29	0.29	0.34	0.32	0.31	0.31	0.31	0.31	0.33	0.32
COX1	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.10	0.10	0.12	0.11	0.11	0.11	0.11	0.12	0.11	
COX2	0.26	0.26	0.25	0.25	0.26	0.26	0.26	0.26	0.25	0.25	0.23	0.23	0.25	0.25	0.26	0.26	0.23	0.23	0.28	0.26	0.25	0.26	0.25	0.27	0.26	
COX3	0.45	0.45	0.44	0.44	0.45	0.45	0.45	0.45	0.44	0.44	0.41	0.41	0.44	0.44	0.45	0.45	0.41	0.41	0.48	0.46	0.44	0.45	0.45	0.48	0.46	
CytB	1.10	1.10	1.08	1.08	1.09	1.09	1.09	1.09	1.07	1.07	1.00	1.00	1.08	1.08	1.10	1.10	1.01	1.01	1.17	1.15	1.10	1.12	1.11	1.19	1.15	
ND1	0.64	1.65	0.62	1.62	0.64	1.64	0.64	1.64	1.63	0.64	1.64	0.60	1.60	0.64	1.64	0.66	1.66	0.59	1.59	0.77	0.76	0.78	0.77	0.77	0.76	0.76
ND2	0.91	0.91	0.92	0.92	0.91	0.91	0.91	0.91	0.92	0.92	0.92	0.92	0.92	0.92	0.91	0.91	0.92	0.92	0.91	0.91	0.91	0.91	0.91	0.91	0.91	
ND3	0.52	0.52	0.51	0.51	0.52	0.52	0.52	0.52	0.51	0.51	0.47	0.47	0.51	0.51	0.52	0.52	0.47	0.47	0.55	0.54	0.51	0.52	0.52	0.56	0.54	
ND4	0.66	0.66	0.65	0.65	0.66	0.66	0.66	0.66	0.64	0.64	0.60	0.60	0.65	0.65	0.66	0.66	0.60	0.60	0.70	0.68	0.65	0.67	0.66	0.71	0.68	
ND4L	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.55	0.55	0.51	0.51	0.55	0.55	0.56	0.56	0.52	0.52	0.60	0.58	0.56	0.57	0.56	0.60	0.58	
ND5	0.86	0.86	0.85	0.85	0.86	0.86	0.86	0.86	0.84	0.84	0.79	0.79	0.85	0.85	0.86	0.86	0.79	0.79	0.92	0.90	0.86	0.88	0.87	0.93	0.90	
ND6	0.99	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.97	0.97	0.91	0.91	0.98	0.98	1.00	1.00	0.91	0.91	1.06	1.04	0.99	1.01	1.01	1.08	1.04	

Supplementary Table S7. Energy (*kcal/mol*) of the cloverleaf structures of tRNAs in snakes and lizards.

	Snakes										Lizards and Tuatara								
	<i>Agkistrodon piscivorus</i>	<i>Pantherophis guttatus</i>	<i>Dinodon semicarinatus</i>	<i>Acrochordus granulatus</i>	<i>Boa constrictor</i>	<i>Ovophis okinavensis</i>	<i>Cylindrophis ruffus</i>	<i>Python regius</i>	<i>Xenopeltis unicolor</i>	<i>Leptotyphlops dulcis</i>	<i>Iguana iguana</i>	<i>Eumeces egregius</i>	<i>Sceloporus occidentalis</i>	<i>Cordylus warreni</i>	<i>Abronia graminea</i>	<i>Shinisaurus crocodilurus</i>	<i>Varanus komodoensis</i>	<i>Sphenodon punctatus</i>	
Ala	-10	-4.8	-8.9	-6.6	-13	-10	-11	-9.5	-8.5	-15	-12	-7.3	-11	-8.5	-8.1	-8.2	-8.2	-7.6	
Arg	-7.3	-3.5	-7	-14	-11	-9.6	-10	-11	-14	-5.4	-18	-6.5	-17	-12	-15	-12	-19	-11	
Asn	-14	-21	-12	-13	-20	-14	-16	-13	-20	-17	-13	-15	-20	-18	-14	-15	-19	-15	
Asp	-16	-13	-14	-9.3	-27	-19	-20	-13	-11	-8.9	-9	-16	-15	-16	-15	-13	-6.9	-7.2	
Cys	-19	-18	-18	-18	-24	-19	-15	-24	-24	-14	-19	-19	-20	-13	-16	-17	-19	-3.8	
Gln	-10	-10	-8.9	-7.4	-13	-10	-12	-15	-13	-13	-14	-16	-14	-8.6	-13	-13	-2.5	-14	
Glu	-12	-11	-8.7	-9	-11	-13	-13	-15	-11	-11	-12	-11	-9.4	-11	-16	-13	-12	-14	
Gly	-6.9	-4.9	-10	-7.4	-9.2	-7	-11	-12	-11	-7.2	-11	-15	-9	-14	-12	-12	-17	-9.3	
His	-5.8	-5.2	-9.4	-4.2	-7.6	-3.5	-6.2	-8.6	-7.6	-8.8	-9.3	-11	-8.6	-9.6	-1.2	-7.9	-9.4	N/A	
Ile	-15	-15	-17	-14	-15	-11	-16	-15	-15	-8.6	-7.9	-14	-13	-17	-13	-10	-15	-13	
Leu2	-10	-8.7	-10	6.1	-9	-11	-8.3	-12	-9	-18	-18	-10	-14	-13	-12	-13	-9.3	-12	
Leu4	-20	-15	-16	-16	-17	-16	-14	-15	-16	-13	-15	-17	-14	-12	-16	-13	-15	-25	
Lys	-11	-15	-13	-12	-14	-13	-11	-11	-10	-19	-19	-17	-19	-16	-16	-15	-14	-16	
Met	-14	-16	-16	14.4	-14	-11	-13	-14	-14	-13	-9.3	-9	-9.1	N/a	-9	-9.4	-16	-9.9	
Phe	-3.8	-9.6	-0.3	-8.7	-12	-11	-7.5	-12	-7.6	-15	-17	-17	-12	-8.1	-11	-15	-11	-9.9	
Pro	-13	-8.5	-8.7	-6.5	-8.2	-14	-6.6	-9.8	-5.8	-11	-15	-13	-15	-11	-14	-6.6	-9.6	-19	
Ser4	-14	-13	-15	-13	-17	-10	-16	-11	-18	-19	-14	-23	-14	-16	-16	-14	-14	-8.6	
Thr	-7.5	-10	-4.6	-11	-8.2	-8.7	-10	-11	-6.5	-12	-14	-8	-14	-13	-12	-18	-15	N/A	
Trp	-8.3	-7.7	-6.6	-5.2	-5.3	-8.5	-5.4	-7.4	-4.9	-7.4	-8.6	-21	-22	-13	-17	-5.5	-5.9	-16	
Tyr	-15	-11	-12	-16	-16	-15	-14	-18	-11	-15	-16	-29	-16	-21	-16	-14	-14	-14	
Val	-7.4	-8.8	-6.7	-7.1	-7.6	-7.5	-7.2	-7.5	-8.3	-3.7	-9.2	-6.4	-6.9	-7.1	-5.6	-6.9	-10	-11	

Supplementary Table S8. Complete mitochondrial genomes used in this study, and associated Genbank accession numbers.

Vertebrate Group	Genbank Accession	Taxon	Vertebrate Group	Genbank Accession	Taxon
Amphibians	NC_002756	<i>Mertensiella luschani</i>	Birds	NC_002782	<i>Apteryx haastii</i>
	NC_001573	<i>Xenopus laevis</i>		NC_003128	<i>Buteo buteo</i>
Turtles	NC_000886	<i>Chelonia mydas</i>		NC_002196	<i>Ciconia boyciana</i>
	NC_002073	<i>Chrysemys picta</i>		NC_002197	<i>Ciconia ciconia</i>
	NC_002780	<i>Dogania subplana</i>		NC_002069	<i>Corvus frugilegus</i>
	NC_001947	<i>Pelomedusa subrufa</i>		NC_002784	<i>Dromaius novaehollandiae</i>
Tuatara	NC_004815	<i>Sphenodon punctatus</i>		NC_000878	<i>Falco peregrinus</i>
Lizards	NC_005958	<i>Abronia graminea</i>		NC_001323	<i>Gallus gallus</i>
	NC_005962	<i>Cordylus warreni</i>		NC_000846	<i>Rhea americana</i>
	NC_000888	<i>Eumeces egregius</i>		NC_000879	<i>Smithornis sharpei</i>
	NC_002793	<i>Iguana iguana</i>		NC_002785	<i>Struthio camelus</i>
	NC_005960	<i>Sceloporus occidentalis</i>		NC_002781	<i>Tinamus major</i>
	NC_005959	<i>Shinisaurus crocodilurus</i>		NC_000880	<i>Vidua chalybeata</i>
	AB080275-6	<i>Varanus komodoensis</i>	Mammals	NC_001567	<i>Bos taurus</i>
Snakes	NC_007400	<i>Acrochordus granulatus</i>		NC_002763	<i>Cebus albifrons</i>
	DQ523162	<i>Agkistrodon piscivorus (Api1)</i>		NC_002082	<i>Hylobates lar</i>
	EF669477	<i>Agkistrodon piscivorus (Api2)</i>		NC_001646	<i>Pongo pygmaeus</i>
	NC_007398	<i>Boa constrictor</i>		NC_001644	<i>Pan paniscus</i>
	NC_007401	<i>Cylindrophis ruffus</i>		NC_001645	<i>Gorilla gorilla</i>
	NC_001945	<i>Dinodon semicarinatus</i>		NC_001807	<i>Homo sapiens</i>
	NC_005961	<i>Leptotyphlops dulcis</i>		NC_001992	<i>Papio hamadryas</i>
	NC_007397	<i>Ovophis okinavensis</i>		NC_002764	<i>Macaca sylvanus</i>
	DQ523161	<i>Pantherophis slowinskii</i>		NC_002811	<i>Tarsius bancanus</i>
	NC_007399	<i>Python regius</i>		NC_004025	<i>Lemur catta</i>
	NC_007402	<i>Xenopeltis unicolor</i>		NC_002765	<i>Nycticebus coucang</i>

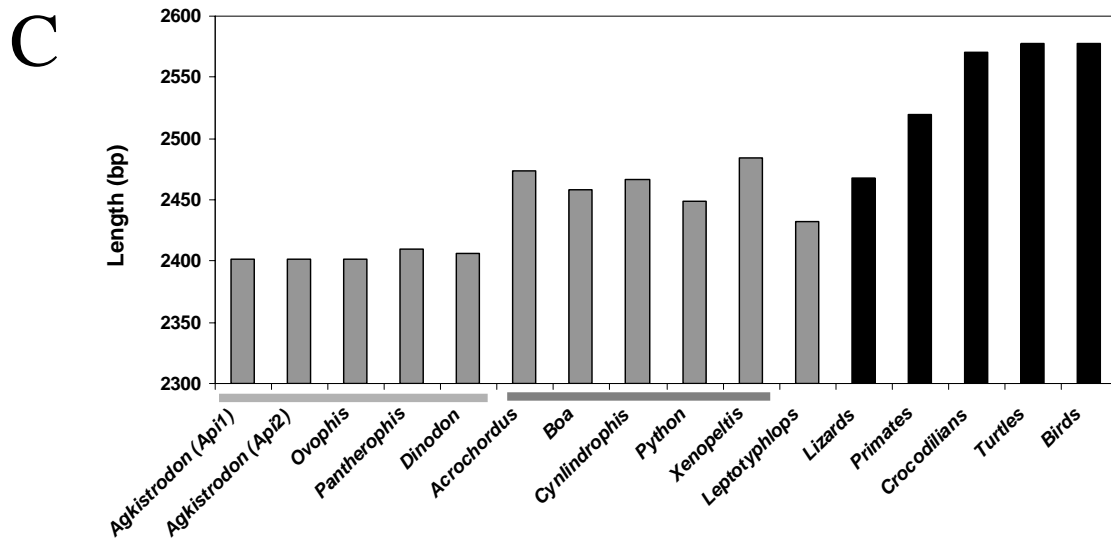
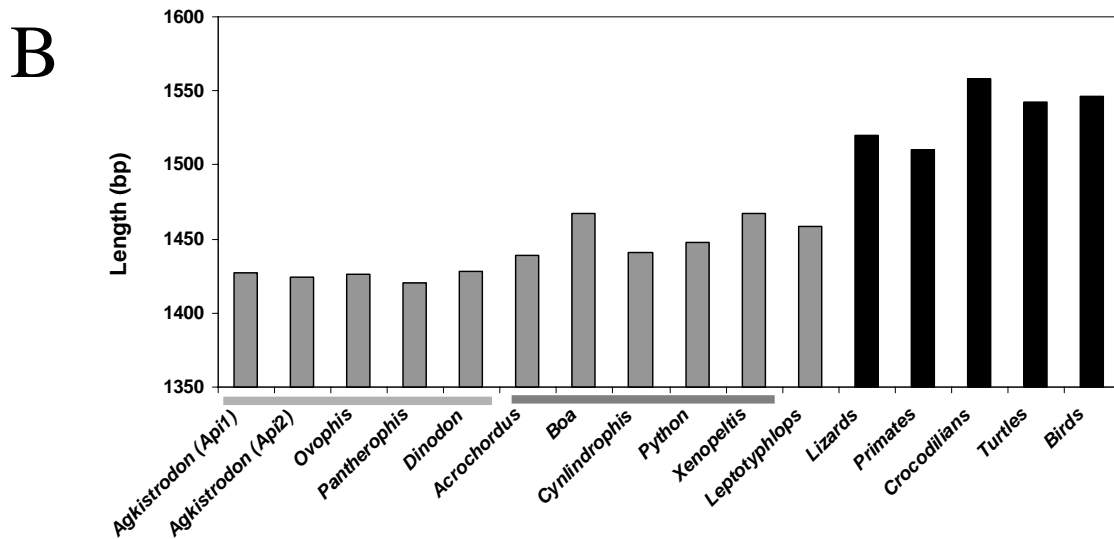
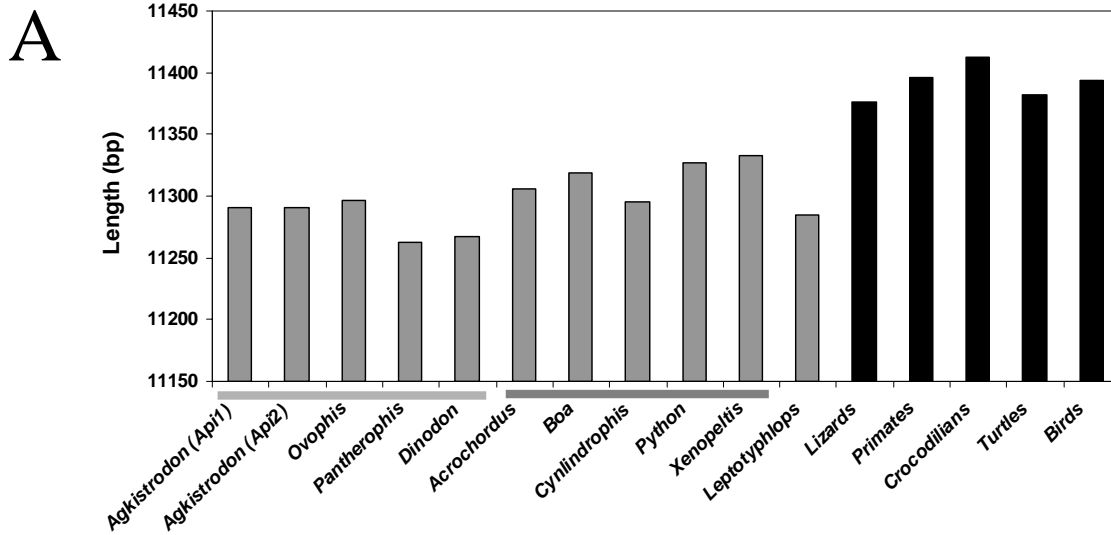
Supplementary Table S9. Primer sets used to amplify mitochondrial genome fragments in this study.

Primer Name	Primer sequence (5' – 3')	Source
<i>Agkistrodon piscivorus</i> - <i>Api2</i> amplification primers		
L2932	MYTGGTGCCAGCCGCCGCGG	This study
tRNATrpR	GGCTTTGAAGGCTMCTAGTTT	R. Lawson, unpub.
ND1L	CTATCCCCATCATAGCMC	This study
ND2H	TCGGGGTATGGGCCCCG	This study
LRattle	ACTCTAACGCTCCTAACCTGAC	K. Zamudio, unpub.
Leu	CCAACACCTVTTCTGATT	Arévalo et al. 1994
L6929	CCAACACCTVTTCTGATT	This study
ND4CP200	ARATTGYRGCTRCTACTARGCC	This study
ND4	CACCTATGACTACCAAAAGCTCATGTAGAAGC	Arévalo et al. 1994
AtrCB3	TGAGAAGTTTTTCYGGGTCRRT	Parkinson et al. 2002
Gludg	TGACTTGAARAACCAAYCGTTG	Parkinson et al. 2002
H3059	CCGGTCTGAACTCAGATCACGT	This study
<i>Agkistrodon piscivorus</i> - <i>Api1</i> amplification primers		
DPFB002R	AGTGGTCAWGGGCTKGGGACTA	This study
DPFB0013F	CGGCCGCGGTATYCTAACCGTGCAAAG	This study
DPFB001F	TAGTAGACCCMAGCCCWTGACCACT	This study
DPFB0021R	CTGATCCAACATCGAGGTCGTAAACC	This study
<i>Pantherophis slowinskii</i> amplification primers		
DPAL007	CTACGTGATCTGAGTTCAGACC	This study
DPFB007	CTCAGAAKGATATYTGTCYCATGG	This study
DPFB006	CCATGRGGACARATATCMTTCTGAG	This study
DPAL006	CTCCGGTCTGAACTCAGATCAC	This study

Additional File 3 (Supplementary Figure)

Supplementary Figure S1 - Comparison of mitochondrial gene lengths among snakes and other tetrapod groups.

The total length (b.p.) is shown for all protein coding regions (A), tRNAs (B), and rRNAs (C). All snakes are in gray, while other squamates (lizards) are in black; light gray and dark gray bars are drawn under snakes species to indicate membership in the Colubroidea or Acro-Heno Clade, respectively.



Additional File 4 (Supplementary Methods)

Supplementary Methods. Laboratory and genome annotation methods.

Total DNA was isolated from frozen (-80C) liver tissue of *Api2* using the Qiagen DNeasy extraction kit and protocol (Qiagen Inc.). Using the Expand Long Template PCR system (Roche Molecular Biochemicals), the mitochondrial genome was amplified in six overlapping fragments with 12 primers (Table S9). In addition, several smaller fragments were also amplified using the BIO-X-ACT Short PCR kit (Bioline) to fill-in otherwise inadequately sequenced regions. Cycling conditions followed the manufacturers' suggestions, with annealing temperatures between 50°C and 55°C, and for 35 cycles.

Positive PCR products were electrophoretically separated and excised from agarose gels, followed by purification using the GeneCleanIII kit (BIO101). Purified PCR products were cloned using either the TopoTA or TopoXL cloning kits (Invitrogen). Plasmids containing amplification fragments were isolated and purified using QIAprep Spin Miniprep kits (Qiagen) and sequenced using M13 primers (flanking the cloning site in the Topo vectors), an array of internal primers (details available upon request), and the CEQ Dye Terminator Cycle Sequencing Quick Start Kit (Beckman-Coulter), and were run on a Beckman CEQ8000 automated sequencer according to the manufacturers' protocols.

Total DNA was extracted from *Api1* using a High Pure PCR Template Preparation Kit (Roche), and amplified into two long overlapping fragments, 8kb and 9kb, using the Expand Long Template PCR Amplification System (Roche) and 4 primers (Table S9). These two fragments overlap in the 16s RNA and COIII genes. Conditions followed the manufacturer's recommendations, with annealing temperatures of 58.4°C (9kb fragment), and 52.2°C (8kb fragment). After electrophoresis as above, PCR products were purified

using the Agarose Gel DNA Purification kit (Mo Bio Laboratory), followed by end phosphorylation, ligation, and shearing in a nebulizer (Invitrogen). Fragments ranging from 1.5-3kb were purified from 0.8% agarose gels using QIAquick Gel Extraction Kit (Qiagen), cloned into pPCR-Script Amp SK(+) vector (Stratagene PCR-Script Amp Cloning Kit), and transformed into XL-10 Gold Kan ultracompetent cells (Stratagene). Bacterial clones containing plasmids with snake mitochondrial inserts were amplified using M13 primers, and the products were purified by QIAquick PCR Purification Kit and sequenced using T3 primer and Big Dye Terminator Sequence Master (PE Biosystems) using standard protocols. The reactions were purified on DyeEx columns (Qiagen), and the DNA sequence was determined using an ABI 3700 automated sequencer.

Total DNA from *Pantherophis* was extracted and amplified using the same protocol and reagents as for *Api1*, but with a different set of four primers (Table S1) yielding 12.5 Kb and 4.5 Kb fragments. These two fragments overlap in the CytB and 16s rRNA genes, and were sequenced following the same protocol as used for *Api1*, with additional internal primers.

Most tRNAs in the raw genome sequences were detected using tRNAscan (Lowe et al. 1997), followed by manual verification. The tRNAs not identified by tRNAscan were identified by their position in the genome and folded manually based on homology. The tRNAs were then used to identify approximate boundaries of protein coding genes, control region, and ribosomal RNAs. Final boundaries of protein coding genes were set based on position of the most plausible first start and last stop codons in each region, including non-canonical signal codons known to operate in vertebrate mitochondrial

genome (Slack et al. 2003). Proteins were also translated to their amino acid sequence, and all amino acid and DNA sequences were compared to the corresponding genes or regions from published snake genomes to verify the annotation.