Title: Molecular adaptations for sensing and securing prey and insight into amniote genome diversity from the garter snake genome

Blair W. Perry[1,#], Daren C. Card[1,#], Joel W. McGlothlin[2], Giulia I. M. Pasquesi[1], Richard H. Adams[1], Drew R. Schield[1], Nicole R. Hales[1], Andrew B. Corbin[1], Jeffery P. Demuth[1], Federico G. Hoffmann[3, 4], Michael W. Vandewege[5], Ryan K. Schott[6,7], Nihar Bhattacharyya[8], Belinda S.W. Chang[6], Nicholas R. Casewell[9], Gareth Whiteley[9], Jacobo Reyes-Velasco[1,10], Stephen P. Mackessy[11], Kenneth B. Storey[12], Kyle K. Biggar[12], Courtney N. Passow[13], Chih-Horng Kuo[14], Suzanne E. McGaugh[13], Anne M. Bronikowski[15], Jason de Koning[16], Scott V. Edwards[17], Michael E. Pfrender[18], Patrick Minx[19], Edmund D. Brodie III[20], Edmund D. Brodie, Jr.[21], Wesley C. Warren[19], and Todd A. Castoe[1,*]

[1]Department of Biology, University of Texas at Arlington, 501 S. Nedderman Drive, Arlington, TX, USA 76010

[2]Department of Biological Sciences, Virginia Tech, Derring Hall 2125, 926 West Campus Drive, Blacksburg, VA, USA 24061

[3]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Campus Mailstop Box 9655, Mississippi State, MS, USA 39762

[4]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, 2 Research Boulevard, Starkville, MS, USA 39759

[5]Department of Biology, Institute for Genomics and Evolutionary Medicine, Temple University, SERC Building, 1925 N. 12 Street, Philadelphia, Pennsylvania, USA 19122

[6]Department of Ecology and Evolutionary Biology, Department of Cell and Systems Biology, Centre for the Analysis of Genome Evolution & Function, University of Toronto, 25 Harbord Street, Toronto, ON, Canada M5S 3G5

[7]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, 10th and Constitution Ave, NW, Washington, DC 20560-0162

[8]Department of Cell and Systems Biology, University of Toronto, 25 Harbord St., Toronto, Ontario, Canada

[9]Alistair Reid Venom Research Unit, Parasitology Department, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK

[10]Department of Biology, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates

[11]School of Biological Sciences, University of Northern Colorado, 501 20th Street, Greeley, CO, USA 80639

[12]Institute of Biochemistry, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, Canada

[13]College of Biological Sciences, University of Minnesota, 1987 Upper Buford Circle, St. Paul, MN, USA 55108

[14]Institute of Plant and Microbial Biology, Academia Sinica, Taipei 11529, Taiwan

[15]Department of Ecology, Evolution, and Organismal Biology, 251 Bessey Hall, Iowa State University, Ames, IA, USA 50011

[16]Department of Biochemistry and Molecular Biology, Department of Medical Genetics, Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada T2N 1N4

[17]Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA, USA 02138

[18]Department of Biological Sciences and Environmental Change Initiative, University of Notre Dame, 109B Galvin Life Sciences, Notre Dame, IN, USA 46556

[19]The McDonnell Genome Institute, Washington University School of Medicine, 4444 Forest Park Avenue, St. Louis, MO, USA 63108

[20]Department of Biology, University of Virginia, P.O. Box 400328, Charlottesville, VA, USA 22904

[21]Department of Biology, Utah State University, 5305 Old Main Hill, Logan, UT, USA 84078

[#]These authors contributed equally.

*Author for Correspondence: Todd A. Castoe, Department of Biology, University of Texas at Arlington, Arlington, TX, USA 76010, Phone: 817-272-9084, Fax: 817-272-9615, Email: todd.castoe@uta.edu

ABSTRACT

Colubridae represents the most phenotypically diverse and speciose family of snakes, yet no well-assembled and annotated genome exists for this lineage. Here we report and analyze the genome of the garter snake, *Thamnophis sirtalis*, a colubrid snake that is an important model species for research in evolutionary biology, physiology, genomics, behavior and the evolution of toxin resistance. Using the garter snake genome, we show how snakes have evolved numerous adaptations for sensing and securing prey, and identify features of snake genome structure that provide insight into the evolution of amniote genomes. Analyses of the garter snake and other squamate reptile genomes highlight shifts in repeat element abundance and expansion within snakes, uncover evidence of genes under positive selection, and provide revised neutral substitution rate estimates for squamates. Our identification of Z and W sex chromosome-specific scaffolds provides evidence for multiple origins of sex chromosome systems in snakes and demonstrates the value of this genome for studying sex chromosome evolution. Analysis of gene duplication and loss in visual and olfactory gene families supports a dim-light ancestral condition in snakes and indicates that olfactory receptor repertoires underwent an expansion early in snake evolution. Additionally, we provide some of the first links between secreted venom proteins, the genes that encode them, and their evolutionary origins in a rear-fanged colubrid snake, together with new genomic insight into the coevolutionary arms race between garter snakes and highly toxic newt prey that led to toxin resistance in garter snakes.

Key words: *Thamnophis sirtalis*, co-evolution, venom, vision, sex chromosomes

INTRODUCTION

The order Squamata, comprised of lizards and snakes, represents a major portion of amniote vertebrate diversity with over 10,000 species that span ~200 million years of evolutionary history (Kumar, et al. 2017; Uetz and Etzold 1996). Among squamates, snakes exhibit many unique and often extreme adaptations, including potent venoms and limblessness, making them of particular interest in studies of amniote evolution and biology. Recent studies have further revealed interesting and divergent features of snake genomes, including unique genomic GC isochore structure (Castoe, et al. 2013), evidence for pervasive and rapid positive selection (Castoe, et al. 2013), multiple origins of sex chromosomes (Gamble, et al. 2017), and complex multigene family evolution in genes associated with sensory systems (Schott, et al. 2018; Vandewege, et al. 2016), venom (Vonk, et al. 2013; Yin, et al. 2016), and toxin resistance (McGlothlin, et al. 2014). Genomic resources are now available for several snake species, including the Burmese python (*Python molurus bivittatus*) (Castoe, et al. 2013), King cobra (*Ophiophagus hannah*) (Vonk, et al. 2013), *Boa constrictor* (Bradnam, et al. 2013), and the five-pace viper (*Deinagkistrodon acutus*) (Yin, et al. 2016). However, currently available genomic resources for snakes leave the single most diverse and speciose family of snakes, Colubridae, surprisingly underrepresented. Only one colubrid draft genome has been published to date (the corn snake, *Pantherophis gutattus*) (Ullate-Agote, et al. 2015), but the highly fragmented state of this assembly limits its use and value for comparative genomic analyses.

Among colubrids, garter snakes (genus *Thamnophis*) are arguably the most well-studied lineage and have emerged as an important vertebrate model for research spanning a broad range of

biological questions. For example, garter snakes have served as one of the only non-mammalian model systems for understanding the development of behavior, which has shown that snakes develop unique and consistent personalities that are shaped by experience with ecological factors including their interactions with predators, prey, and chemical stimuli (Herzog and Burghardt 1986; Herzog and Burghardt 1988). Garter snakes have also proved to be valuable models for a wide range of physiological questions, including those related to reproduction (Mendonca and Crews 1996), exercise (Jayne and Bennett 1989), and aging (Robert and Bronikowski 2010). They have also been important models for studying snake vision systems to examine the effects of fossorial ancestry on the morphology, physiology, and genetics of vision in snakes (Schott, et al. 2016; Schott, et al. 2018), and to study how decreased snake visual acuity may lead to enhanced chemosensation and pheromonal communication and behaviour (Crews, et al. 1984; Kubie, et al. 1978; Mason and Crews 1986).

Like many other colubrids, garter snakes are capable of producing a relatively mild venom that is used to subdue prey (Hayes and Hayes 1985; Vest 1981). In contrast to other venomous snake lineages such as cobras (Elapidae) and rattlesnakes (Viperidae) that produce extremely potent venom in specialized glands with delivery through hollow anterior fangs, colubrid venom is produced in a homologous gland (Duvernoy's gland) and delivered through grooved posterior maxillary teeth, which has led to the term "rear-fanged" for describing venomous colubrid snakes. Several studies have used available snake genomes to study the variation and evolution of venom gene families across snake species, but have been limited to either strictly non-venomous or highly venomous species (Reyes-Velasco, et al. 2014; Vonk, et al. 2013; Yin, et al. 2016). The genome of a rear-fanged venomous colubrid species such as the garter snake would provide a more complete representation of the diversity of venom systems in snakes, offering

increased potential to understand the evolution of this extreme and medically relevant feature of snake biology. In addition to producing mildly toxic venoms, some species of garter snakes are also remarkable for having evolved resistance to the toxic compounds of their prey, specifically in toxic newts of the genus *Taricha*. Given this unique adaptation, garter snakes have become an important model for studying the evolution of toxin resistance and coevolutionary dynamics of predator-prey arms races (Brodie and Brodie 1999; Brodie, et al. 2002; Feldman, et al. 2009; McGlothlin, et al. 2016).

Here we sequenced and annotated the genome of the garter snake (*Thamnophis sirtalis*) and used this genome and related resources to address broad questions related to vertebrate genome evolution and garter snake biology. We examined molecular evolution across gene families that underlie unique and interesting biological features of garter snakes related to toxin resistance and venom, as well as olfactory and visual systems. We also compared evolutionary patterns of genome composition and structure (repeat content, microRNAs, and sex chromosomes) in garter snakes in the context of other amniote genomes, identified genes that have likely been under positive selection in the garter snake and colubrids in general, and estimated neutral substitution rates across four-fold degenerate sites for the garter snake and other squamates reptiles.

MATERIALS AND METHODS

Genome sequencing and annotation. All animal procedures were conducted with registered IACUC protocols (see Supplementary Methods). Garter snake genome sequencing libraries were constructed from a single adult female *Thamnophis sirtalis* collected in Benton County, Oregon. A genome size of 1.9 Gbp, previously estimated using flow cytometry (Vinogradov 1998), was used for the construction of libraries. We created and sequenced multiple whole-genome shotgun

libraries on the Illumina platform, generating approximately 72x genome coverage, which was subsequently used for assembly with ALLPATHS-LG (Gnerre, et al. 2011), scaffolding with L_RNA_scaffolder (Xue, et al. 2013) and SSPACE (Boetzer, et al. 2010), and gapfilling using a custom script. This 6.0 version has been screened and cleaned of contaminating contigs and all contigs 200bp or smaller were removed. The genome assembly was annotated by NCBI using their automated Eukaryotic Genome Annotation Pipeline (Thibaud-Nissen, et al. 2013). To generate an updated estimate of genome size, kmers were extracted using jellyfish v2.2.3 (Marçais and Kingsford 2011) and genome size estimates based on both 19mers and 23mers were generated using GCE v1.0.0 (Liu, et al. 2013). The completeness of the *T. sirtalis* genome and 5 other squamate genomes (3 lizards: *Pogona vitticeps*, *Anolis carolinensis*, *Ophiosaurus gracilis*; and 2 snakes: *Python bivittatus*, *Ophiophagus hannah*) was assessed by comparing their annotated gene sets to benchmarking sets of universal single-copy vertebrate orthologs using BUSCO v1.1b1 (Simão, et al. 2015), which included 3,023 vertebrate genes. The NCBI-generated protein FASTA files for *T. sirtalis* and *P. bivittatus* contain both multiple isoforms and seemingly duplicated proteins (i.e. two or more proteins with identical sequence and description in protein FASTA file, identical coordinates in GFF file, but unique protein identifiers). We therefore filtered the protein FASTA files for these two species to contain only the longest isoform if multiple were present and only one protein entry in the case of duplicated entries, and used these filtered, non-redundant sets of proteins for BUSCO analysis. Genomic repeat elements were annotated using RepeatMasker (Smit, et al. 2015) using *de novo* repeat libraries derived from RepeatModeler (Smit, et al. 2014) analysis of the genome assembly, combined with repeat libraries from other snake genomes (Castoe, et al. 2013; Yin, et al. 2016). Distributions of

gene, exon, and intron number and length were generated using custom Python scripts. See Supplementary Methods 1.1-1.3 for additional details.

Analyses of positive selection on protein-coding genes. To infer patterns of positive selection in protein coding genes in squamate reptiles, including the branch leading to *T. sirtalis*, we generated orthologous coding sequence alignments for 27 species, including 18 reptiles and nine mammals (Supplementary Table S2), following a previous established protocol (McGaugh, et al. 2015). Notably, our alignments contained transcriptomes for several species from the Colubridae family of snakes: African house snake (*Lamprophis fuliginosus*) and two species of garter snakes *T. couchii* and *T. elegans* (the latter of which is represented by two separate transcriptomes for two distinct ecotypes), in addition to *T. sirtalis*. Alignments were used to identify evidence for positive selection in the codeml program in PAML v4.7 (Yang 2007). We used a likelihood-ratio test (LTR) and three separate branch-site models to test for evidence of positive selection. The three branch-site models varied in which branch was placed in the foreground and included: 1) the branch to the family Colubridae, 2) the branch to the *Thamnophis* genus, and 3) the branch to *T. sirtalis*. Functional characteristics of genes with significant evidence of positive selection were analyzed using Ingenuity Pathway Analysis (IPA) Comparison Analysis (Krämer, et al. 2013). Enrichment of functional categories was determined using Fisher's exact tests comparing genes with evidence of positive selection in each of the three branch site models to the IPA knowledge base, and was considered significant when $p < 0.05$. See Supplementary Methods 1.4 for additional details.

Estimating neutral substitution rates for squamate reptiles. We inferred neutral substitution rates across squamate reptiles by estimating rates of substitution at four-fold degenerate third

codon positions of orthologous protein-coding genes. We used a Perl script provided in Castoe, et al. (2013) to identify and sample four-fold degenerate sites for 7 squamates (glass lizard, *Ophisaurus gracilis*; green anole, *Anolis carolinensis*; bearded dragon, *Pogona vitticeps*; king cobra, *Ophiophagus hannah*; garter snake, *Thamnophis sirtalis*; cottonmouth, *Agkistrodon piscivorus*; and Burmese python, *Python molurus*) and three mammal outgroups (chimpanzee, *Pan troglodytes*; human, *Homo sapiens*; and mouse, *Mus musculus*); these data were subsampled from the larger 1:1 orthologous gene alignments used for analyses of positive selection in this study (see Supplementary Methods 1.4 for details). We estimated branch-specific substitution rates using a relaxed log-normal clock model in BEAST v2.4.7 (Bouckaert, et al. 2014) by calibrating the divergence times of the 9 internal nodes (see Supplementary Table S4 for node calibrations and prior distributions) and constraining the topology to (((GlassLizard, (Anole,Pogona)),(((Cobra,Thamno),CottonMouth),Python)),((Chimp,Human),Mouse)). Two BEAST analyses were conducted for 100 million generations each, sampling parameter values every 10,000 generations. We discarded the first 25% of runs as burn-in (25 million) and confirmed convergence to the posterior by comparing the likelihood values and effective samples sizes (ESS > 1000 for all parameters).

Sex chromosomes. To identify sex chromosome-specific scaffolds in the *T. sirtalis* genome, we used BLAST to locate six gene fragments known to have both Z and W-specific alleles in snakes (Laopichienpong, et al. 2017; Matsubara, et al. 2006; Vicoso, et al. 2013). Maximum likelihood phylogenetic analyses on these gene fragments from several squamates were used to verify sex linkage, which would produce a pattern where Z and W alleles from distinct species cluster together phylogenetically. We also used sex-specific amplification of the *T. sirtalis* gene fragments by PCR to confirm that these scaffolds were sex-linked gametologous alleles and not

paralogs or misassemblies. Fluorescent *in situ* hybridization (FISH) of a *Bkm*-like repeat, $(GATA)_n$, was used to identify the W chromosome in the karyotype of a female *T. sirtalis*. See Supplementary Methods 1.5 for details.

Identification and analysis of microRNA and associated mRNA targets. The sequence-specific microRNA prediction (SMIRP) tool was used for the identification of microRNAs that are specific to *T. sirtalis* (Peace, et al. 2015), and these predictions were tested in a number of ways, as described in Supplementary Methods 1.6. Inferred mature microRNAs were putatively matched to target coding sequences in the *T. sirtalis* genome using miRanda v.3.3a (Enright, et al. 2003), using the parameters described in the Supplementary Methods 1.6.

Analyses of visual gene loss and opsin expression localization. To identify the full repertoire of visual genes in the garter snake genome, we used BLAST (Altschul, et al. 1990) to search the complete genome sequence, as well as the genomes of the king cobra (*Ophiophagus hannah*) and Burmese python (*Python molurus bivittatus*), using a library of visual and opsin genes compiled from previous studies of vertebrate vision (Schott, et al. 2017; Schott, et al. 2018). Genes that could not be recovered in any of the three snake genomes were presumed to be lost. Gene loss in snakes was compared to losses in mammals, geckos, and crocodilians, three groups with hypothesized nocturnal or dim-light ancestry. Under approved animal protocols at the University of Toronto, we conducted immunohistochemistry to identify the location of opsin gene expression across cells in the *T. sirtalis* retina based on previously published methods (described in Bhattacharyya, et al. (2017) and Schott, et al. (2016)). See Supplementary Methods 1.7 for additional experimental details.

Annotation and phylogenetic analysis of olfactory receptors. To infer patterns of expansion and loss in olfactory receptors (ORs) in snakes, we identified and analyzed OR genes from the garter snake and other reptile genomes. OR annotations from the green anole and Burmese python were collected from Vandewege, et al. (2016). In addition to the garter snake genome, three additional squamate genomes were included: the bearded dragon (*Pogona vitticeps*) (Georges, et al. 2015) and two snakes, the *Boa constrictor* (Bradnam, et al. 2013) and king cobra (*Ophiophagus hannah*) (Vonk, et al. 2013). Known OR amino acid sequences from the human, chicken, zebra fish, clawed frog, green anole, python, softshell turtle, painted turtle, saltwater crocodile, gharial and American alligator were used to query each genome using tBLASTn. Predicted OR genes were categorized into intact genes, pseudogenes and truncated genes as defined in Vandewege, et al. (2016). We aligned intact OR sequences and estimated phylogenetic trees for all OR sequences as described in Supplementary Methods 1.8.

Venom gene and protein analysis. To identify venom gene homologs in the *T. sirtalis* genome, we used tBLASTx based on a query set of known venom genes from diverse snake species, and secondarily tested orthology by inferring phylogenies of sets of *T. sirtalis* genes that showed homology to known venom genes. We then used RNAseq data from eight different *T. sirtalis* tissues (Duvernoy's gland, brain, kidney, liver, lung, ovary, upper digestive tract, and lower digestive tract) to estimate patterns of gene expression across tissues for venom gene orthologs. To characterize the protein composition of *Thamnophis* venom, and compare this to inferences based on RNAseq, we extracted and analyzed crude venom from *T. sirtalis* and *T. elegans* using SDS-PAGE and MALDI-TOF mass spectrometry, under approved animal protocols at the University of Northern Colorado (#9204). See Supplementary Methods 1.9 for additional details.

Analysis of tetrodotoxin resistance in voltage-gated sodium channels. Garter snakes are well known for their resistance to tetrodotoxin in their prey, which at the molecular level derives from resistant voltage-gated sodium channels. Because automated annotation approaches performed poorly on these genes, we manually predicted and extracted sequences of all sodium channel alpha subunit (*SCNA)* paralogs from the *T. sirtalis* genome and other published genomes, including three snakes (*Boa constrictor*, *Ophiophagus hannah*, and *Python molurus bivittatus*), two lizards (*Anolis carolinensis* and *Ophisaurus gracilis*), a bird (*Gallus gallus*), a turtle (*Chrysemys picta*), the human, and as an outgroup, a frog (*Xenopus tropicalis*). We focused on the evolution of three voltage-gated sodium channels previously undescribed in *T. sirtalis*: SCN5A ($Na_v1.5$), SCN10A ($Na_v1.8$), and SCN11A ($Na_v1.9$). We aligned these related paralogs using MUSCLE (Edgar 2004), built a gene tree using RAxML (Stamatakis 2014), and performed ancestral sequence reconstruction using the codeml program in PAML v4.8 (Yang 2007).

RESULTS AND DISCUSSION

Genome assembly and annotation

The garter snake represents the first fully annotated genome available for any colubrid snake and is available at NCBI: version Thamnophis_sirtalis-6.0, BioProject PRJNA294278, GenBank accession GCA_001077635.2. The scaffolded genome assembly is 1.42 Gbp in length (including gaps), with an N50 contig size of 10.4 kb and scaffold N50 size of 647.6 kb. The *T. sirtalis* genome annotation contains approximately 85% of BUSCO genes, making it more complete than the glass lizard (*Ophisaurus*; 84%) and king cobra (*Ophiophagus*; 71%) genomes, but less complete than the genomes of the python and other two lizards assessed (*Anolis* and *Pogona*), which each contained at least 95% of vertebrate BUSCOs (Supplementary Figure S1). Kmer-based estimation of genome size resulted in estimates of 1.19 Gbp and 1.27 Gbp (mean = 1.23

Gbp) using 19mers and 23mers, respectively. These estimates are substantially smaller than the estimate derived from flow cytometry (1.9 Gbp; (Vinogradov 1998)) but are similar to the total assembly length (1.42 Gbp), suggesting that kmer-based estimates are reasonable and likely more accurate than the original flow cytometry estimate.

Genome structure and evolution

Genomic repeat landscapes. The genomic landscapes of repeat elements in squamate reptiles show remarkable variation in repeat abundance among species and lineages when compared to other vertebrate genomes (Adams, et al. 2016; Castoe, et al. 2011; Castoe, et al. 2013). Based on limited studies, repetitive content appears to vary substantially between snake lineages in particular, despite low variation in genome size across snakes (Gregory 2001). We characterized the repeat landscape in the garter snake genome in relation to other squamate reptile genomes and estimate that readily identifiable repeat elements comprise 37.12% of the garter snake genome (ca. 417.1 Mbp), with TEs and microsatellites (SSRs) accounting for 32.9% and 2.8% of the assembled genome, respectively. Comparison of the garter snake repeat element landscape to those of other squamate reptiles reveals lineage- and species-specific patterns of repeat element evolution and expansion (Figure 1). Like other squamate genomes, the garter snake genome is dominated by LINE (non-LTR) elements of the CR1-like family (5.5% of the genome), as well as Gypsy LTR elements (1.37%), and hAT (4.6%) and Tc1-Mariner (TcMar; 4.7%) DNA transposons (Figure 1A). Overall, the genomes of colubroid snakes (*Deinagkistrodon*, *Thamnophis*, *Ophiophagus* in Figure 1) are similar in repeat element content and relative abundance, characterized by higher total repeat and TE content compared to the non-colubroid python genome, and all have notably higher CR1-L3 LINE and hAT and TcMar DNA transposon content (Figure 1A). The garter snake genome exhibits several unique repeat

landscape features, such as an expansion of SINEs and subfamilies of DNA transposons that are otherwise absent or very rare in snakes. Our analyses detected recent and highly lineage-specific expansion of Deu SINEs, which are absent in the anole and python and are 6- and 2-fold more abundant in the garter snake than in the viper and cobra, respectively. Additionally, Polinton DNA elements are detected only in the garter snake assembly, and Rolling Circle (RC) DNA elements show a recent expansion in the garter snake that is on average 5-fold greater than the other snake species analyzed (Figure 1B). Together, analyses of repeat element content highlight the highly active, diverse, and dynamic nature of snake repeat element landscapes, particularly in colubroid snakes.

Comparison of gene feature distributions. Because few direct comparisons of annotated gene features across amniotes are available, we used the garter snake annotation to compare major characteristics of gene annotations across major amniote lineages. The garter snake genome contains a similar number of genes (20,186) compared to the human (21,407), although the median gene length of the garter snake, 13,384 base pairs (bp), is substantially lower that of the human (23,247.50 bp; Supplementary Table S1). The chicken, with the smallest genome of the species analyzed, shows a relatively greater number of small genes (less than ~5,000 bp) and a relatively lower number of large genes (greater than ~20,000 bp) compared to the other three species (Supplementary Figure 2A). The human, with the largest genome of the species analyzed, possesses a greater number of genes greater than ~40,000 bp compared to the other three species (Supplementary Figure 2A). Despite the variation in genome size across these species, the distributions of exon length are remarkably similar, although humans appear to possess a slightly greater number of exons between ~1,000 and 3,000 bp (Supplementary Figure 2B). Furthermore, the median exon length is very similar in all four species (Supplementary Table S1).

Distributions of intron length are highly similar only for the garter snake and anole, whereas chicken shows a greater number of small introns (< ~2,000 bp) and lower number of larger introns (> ~3,000 bp) and the human shows a generally lower number of small introns (< ~3,000 bp) and slightly greater number of large introns (> ~6,5000; Supplementary Figure 2C). Accordingly, the garter snake and anole have similar median intron lengths (1,439 and 1,351 bp, respectively), whereas chicken introns are generally smaller (median = 777 bp) and human introns are generally larger (median = 1,638 bp; Supplementary Table S1).

Evidence of positive selection in the garter snake genome. Previous estimates of molecular evolution based on a small number of snake genomes showed that rates of protein evolution are accelerated in snake lineages (Castoe, et al. 2013). Specifically, these analyses identified many genes with evidence of positive selection, some of which included genes involved in the cardiovascular system, multiple signaling pathways, cell cycle control, and lipid and protein metabolism, among others. With the added sampling of the garter snake genome and other snake genomic data, we investigated whether genes with evidence of positive selection were enriched for particular functions at three levels of taxonomic organization, specifically along the branch leading to *Colubridae*, as well as branches leading to the genus *Thamnophis* and the branch leading to *Thamnophis sirtalis*. We identified four genes (*EMC7*, *WNT9A*, *STK40*, and *PICK1*) that were significant for the branch-site model when placing colubrids in the foreground (Supplementary Table S3). The last of these, *PICK1,* a synaptic protein, is notable for its interaction with both non-voltage gated sodium channels (BNC1) as well as acid-sensing ion channel (ASIC) (Hruska-Hageman, et al. 2002). We identified 10 genes that were significant when the branch to the *Thamnophis* genus was placed in the foreground. Mutations in many of these (e.g., *TBCD, SULF2, EXT2*) are associated with neurological developmental syndromes

(Supplementary Table S3) (Jaglin and Chelly 2009; Wade, et al. 2013; Wall, et al. 2009). Finally, we identified eight genes with significant evidence of positive selection when placing only *Thamnophis sirtalis* in the foreground (Supplementary Table S3).

In the IPA Comparison Analysis (Krämer, et al. 2013), focusing first on the genes with evidence of positive selection in colubrids, *PICK1* stood out as having functionality in cell morphology, organ morphology, and organismal development. *PICK1* also interacts with cation channels (e.g BNC1, ASIC) that are important in the central nervous system as well as peripheral sensory neurons (Hruska-Hageman, et al. 2002), and recent research identified an association between a snake venom toxin (a Texas coral snake toxin) and ASICs, specifically where ASIC1a and b had strong toxin-evoked responses (Bohlen, et al. 2011). Focusing in on the *Thamnophis* genus, we identified significant enrichment for genes involved in proline biosynthesis and in cancer (Fisher's exact test, $p < 0.05$) (Bernard, et al. 2001; Székely, et al. 2008). Finally, at the species level, *Thamnophis sirtalis* genes with evidence of positive selection were particularly prominent in cancer and cell death (apoptosis phenotypes). Snakes are of interest in cancer biology because reptiles, in general, have lower incidences of cancer than mammals (Effron, et al. 1977) and because within reptiles, snakes have the highest incidences of cancer (crocodilians and turtles have the lowest) (Garner, et al. 2004). We also identified genes, such as *ANO5*, that are associated with skeletal and cardiac muscle, specifically with diseases such as muscle dystrophy and gnathodiaphyseal dysplasia (i.e., bone calcification defect) (Tsutsumi, et al. 2004). Additionally, *PK3D* (phosphatidylinositol-4,5- bisphosphate 3-kinase catalytic subunit delta), which is associated with the venom production cycle in pit vipers (Kerchove, et al. 2008), showed evidence of positive selection on the branch leading to *T. sirtalis* specifically.

Neutral substitution rate estimates. Branch-specific estimates of neutral substitution rates across four-fold degenerate sites indicate that squamate reptiles have an average substitution rate (0.00076 per site per million year) that is similar to mammals (0.0008) (Figure 2). Consistent with other recent estimates (Green, et al. 2014), we find that the high substitution rate of the mouse represents an outlier among vertebrates. Snakes appear to have slightly lower substitution rates (Avg. = 0.00074) compared to lizards (Avg. = 0.00077), due largely to the python, which had the lowest estimated rate in our dataset. Notably, the garter snake had the highest estimated substitution rate among all squamate lineages (Figure 2). Overall, our estimates of substitution rates (for both squamates and mammals) using four-fold degenerate sites were generally similar to recent substitution rate estimates for these lineages (Green, et al. 2014). Variation in substitution rates has been linked to variation in recombination rate, efficacy of DNA repair, and life history traits such as body size or generation time (Bromham 2002; Bromham and Penny 2003; Martin and Palumbi 1993). Garter snakes are members of the most speciose family of snakes, the Colubridae, and future studies of substitution rates across this group may be useful in linking such rates with other features of snake biology and evolution (Tollis, et al. 2018).

Sex Chromosomes. Snakes are important model systems for studying the evolution of sex chromosomes, due to the variation in sex-chromosome differentiation across lineages (Matsubara, et al. 2006), and the recent discovery that sex chromosomes appear to have evolved multiple times and led to both XY and ZW systems (Gamble, et al. 2017). Caenophidian snakes (which includes the family Colubridae) possess ZZ/ZW sex chromosomes derived from the same ancestral autosomes (Matsubara, et al. 2006; Ohno 2013; Rovatsos, et al. 2015; Vicoso, et al. 2013), with males possessing two similar Z chromosomes and females possessing one Z and one W chromosome. Over time, limited recombination between the Z and W has led to degeneration

of the snake W, resulting in the loss of numerous functional genes, the evolution of W-specific alleles, and accumulation of an excess of repetitive sequences (Laopichienpong, et al. 2017; Matsubara, et al. 2006; O'Meally, et al. 2010; Rovatsos, et al. 2015; Vicoso, et al. 2013). Because a heterogametic female *T. sirtalis* was sequenced, we suspected that some genome scaffolds would be W-specific. BLAST of six genes with distinct Z and W alleles in snakes (Laopichienpong, et al. 2017; Matsubara, et al. 2006; Vicoso, et al. 2013) revealed Z and W-specific scaffolds for each gene. To confirm that these scaffolds were W-specific, we first aligned fragments of Z and W scaffolds identified in the *T. sirtalis* genome assembly with sex chromosome sequences from two additional snake species with homologous ZZ/ZW, the corn snake (*Pantherophis guttatus*) and timber rattlesnake (*Crotalus horridus*). We used two snake species with XX/XY sex chromosomes, *Boa constrictor* and *Python molurus bivittatus* (Gamble, et al. 2017), and several lizard species as outgroups. Phylogenetic analyses of the aligned sequences using maximum likelihood recovered distinct clades for caenophidian Z and W alleles (Figure 3A). We then confirmed that putative *T. sirtalis* W scaffolds were indeed sex-linked gametologous alleles and not paralogs or misassemblies by designing PCR primers from the W scaffolds. Primers from all six loci amplified in a sex-specific manner in two male and two female *T. sirtalis* (Supplementary Table S5, Figure 3B). Because a previously published *T. sirtalis* karyotype did not reveal heteromorphic sex chromosomes (Baker, et al. 1972), we investigated whether the *T. sirtalis* W chromosome could be identified cytogenetically by fluorescent labelling repetitive DNA sequences. The Y and W sex chromosomes of many plants and animals show accumulation of *Bkm*-like satellite DNA (Arnemann, et al. 1986; Gamble, et al. 2014; Jones and Singh 1985; O'Meally, et al. 2010; Parasnis, et al. 1999). Fluorescent *in situ* hybridization of a *Bkm*-like repeat, $(GATA)_n$, showed significant enrichment on a single

chromosome in the female karyotype, presumably the W (Figure 3C). Together, the identification of Z and W scaffolds from the genome assembly and the FISH analyses reveal *T. sirtalis* has distinct sex chromosomes with significant sequence differences between the Z and W. The assembly of both Z and W scaffolds in the *T. sirtalis* genome illustrate its high quality and provide a new resource for further investigation into snake sex chromosome evolution.

Annotation and analysis of microRNAs. MicroRNAs are short (18–23 nucleotide) non-coding RNAs that play central roles in the post-transcriptional regulation of mRNA transcripts (Ambros 2004; Bartel 2004). To establish a repertoire of small regulatory RNAs within the *T. sirtalis* genome, we identified 203 precursor (189 conserved and 14 novel; Supplementary Figure S3) and associated mature microRNAs from *T. sirtalis* (Supplementary File 1). Nucleotide frequencies within mature microRNAs were found to be highly similar between the garter snake and king cobra genomes when compared to human, with a high frequency of uracil base at position 9, just outside of the 2-8 seed region responsible for target binding (Supplementary Figure S4). A total of 576 mRNA transcripts were predicted to be targets of the novel *T. sirtalis* mature microRNAs (Supplementary File 1). Notably, the novel *T. sirtalis* microRNA, *tsi-miR-novel4*, was found to target several components of phosphoinositide metabolism; namely, phosphoinoside-3-kinase, regulatory subunit 5 (PIK3R5, also called p101), and inositol polyphosphate-4-phosphoatase (INPP4A). Interestingly, PI3K is responsible for the production of phosphatidylinositol 3,4,5-triphosphate (PIP_3) from the phosphorylation of phosphatidylinositol 4,5-bisphosphate (PIP_2). PIP_2 is a key lipid signaling molecule involved in the signaling cascade that is responsible for initiating the venom production cycle in snake secretory cells following the release and depletion of venom (Jansen and Foehring 1983; Kerchove, et al. 2008). We speculate that snake-specific microRNAs may play a role in venom

production cycle, perhaps by regulating the metabolism of PIP_2. As such, it is likely that *T. sirtalis*-specific microRNA are involved in specific cellular pathways that contribute to unique facets of its biology.

Evolution of genes related to vision, olfaction, and venom

Several multigene families are of particular interest in snakes due to their association with unique aspects of snake biology. Previous work in the python and cobra has shown significant expansions of gene families in terminal snake branches, including gene families involved in chemoreception (i.e. vomeronasal, olfactory, and ephrin-like receptors) and vision (i.e. opsin genes), which are associated with the utility of the snake tongue and the hypothesis of a fossorial ancestor to all snakes, respectively. Additionally, genes families involved in the production of snake venom proteins are important for understanding one of the most extreme adaptations of the snake lineage. Here, we characterize the size and composition of these gene families in the garter snake and provide additional insight into the evolution of visual, olfactory, and venom systems in snakes.

Evolution of visual systems and related genes in snakes. It has been proposed that a period of visual degeneration occurred during the origin of snakes due to a hypothesized fossorial phase, although alternate theories include marine or nocturnal ancestry (Davies, et al. 2009; Schott, et al. 2018; Simões, et al. 2015; Vidal and Hedges 2004). This period of visual degeneration is supported by many morphological changes to the snake eye and the loss of 11 opsin genes that are otherwise present in squamate reptiles (Castoe, et al. 2013; Schott, et al. 2017; Schott, et al. 2018). This loss is similar to mammals, which have independently also lost 11 opsin genes, presumably following the hypothesized nocturnal phase in early mammalian evolution

(Angielczyk and Schmitz 2014; Bickelmann, et al. 2015; Hall, et al. 2012). Recently, Schott et al. (2018) analyzed gene loss and patterns of selection in snake phototransduction genes and found evidence to support a dim-light, but not fossorially adapted, origin of snakes. However, numerous other genes are involved in visual perception in vertebrates and it is unknown how these genes were affected by the hypothesized visual degeneration of ancestral snakes. The increasing number of reptilian genomes allows us to reconstruct where in squamate evolution visual genes may have been lost. Here, we examined the losses of visual genes in snakes and compared those to losses in mammals and other vertebrate clades with nocturnal ancestry.

Snakes have lost 17 visual genes (20 including the single gene lost in all squamates and two genes lost in all reptiles), which is more than any other group with nocturnal or dim-light ancestry (Figure 4A, Supplementary Figure S5). In comparison, 11 genes have been lost in the lineage leading to mammals, all of which are opsins (Figure 4A). Placental mammals have lost three additional opsins, increasing the total number of lost opsin genes beyond that in snakes (Supplementary Figure S5). Geckos, another group with a hypothesized highly nocturnal ancestry (Gamble, et al. 2015; Walls 1942), have specifically lost 11 visual genes, 7 of which were, independently, also lost in snakes (Figure 4A). Crocodilians have lost the fewest visual genes (five), all of which were opsins (Figure 4A).

Our analyses of the visual gene complement in *T. sirtalis* and other snakes, when compared to other groups with nocturnal/dim-light ancestry, strongly supports dim-light ancestry in the lineage leading to snakes. Opsins appear to be highly susceptible to loss in nocturnal groups, with each group analyzed showing the loss of two visual opsins: a middle-wavelength opsin (*RH1* or *RH2*) and a short-wavelength opsin (either *SWS1* or *SWS2*). We did not observe the loss

of more than two visual opsins. Loss of nonvisual opsins was substantial in all groups, with many of the same genes being lost independently along different lineages. Loss of phototransduction genes was rare, and only occurred in snakes and geckos, while the loss of lens crystallin genes was largely unique to snakes, with the exception of a single loss in geckos. The snake lens differs substantially from other squamates in that it is rigid and spherical and is moved back and forth instead of being compressed during accommodation (Walls 1942). The loss of five lens crystallins genes may be related to these differences. The recruitment and evolution of new, snake-specific lens crystallins may also have occurred but has yet to be studied. Other than the increased loss of lens crystallins, gene loss in snakes closely resembles losses in mammals and geckos, which supports the hypothesis of dim-light ancestry in snakes.

Evidence of transmuted, cone-like rod photoreceptors in T. sirtalis. In the vertebrate eye, two types of photoreceptors are responsible for absorbing light for visual perception: cones, which function in bright light and for color vision, and rods, which function in dim-light. While most animals have both rods and cones, enabling them to see in a wide range of light conditions, some squamates, including snakes, have only one type of photoreceptor, leading to an all-cone or all-rod retina (Underwood 1970; Walls 1942). Recently, it has been shown that the apparent all-cone retina of a diurnal garter snake, *Thamnophis proximus,* evolved not through loss of the rods, but rather through transmutation (i.e., evolutionary modification) of the rods to resemble the appearance and function of cones (Schott, et al. 2016). This provided the first molecular support for Walls' (1942) transmutation theory proposed over 75 years ago. This has since been demonstrated in a second colubrid, the pine snake (*Pituophis melanoleucus*), with additional support for a cone-like function of rod opsin that is expressed in transmuted photoreceptors

(Bhattacharyya, et al. 2017). It is still unknown, however, how widespread this particular mechanism is in the development of all-cone retinas.

Despite an apparent all-cone retina (Sillman, et al. 1997), *T. sirtalis* maintains all rod phototransduction genes within its genome, with the exception of rhodopsin kinase (*GRK1*), which was lost in all snakes, including those with retinas containing both rods and cones (e.g., *Python*). To determine if rod genes are actually expressed in the photoreceptors, we performed immunohistochemistry on frozen sections of *T. sirtalis* retina and visualized them with fluorescent confocal imaging. A small subset of photoreceptors in the all-cone retina of *T. sirtalis* show rhodopsin protein in the outer segment and rod transducin in the inner segment of the same photoreceptor (Figure 4B). The combination of the intact rhodopsin gene in the genome and the presence of rhodopsin protein in the outer segment of photoreceptors, along with a downstream member of the rod phototransduction cascade suggests that *T. sirtalis*, like *T. proximus,* has transmuted, cone-like rods (Schott, et al. 2016). These findings support the hypothesis that all-cone retinas in diurnal snakes evolved not through the loss of rods, but rather through evolutionary modification of the rods to resemble and function as cones. Recent work examining opsin expression in a variety of snakes suggests that this mechanism may be widespread (Simoes, et al. 2016), but additional work is needed to determine if this is the sole mechanism that produced all-cone retinas in snakes, and conversely, to determine the mechanisms behind the evolution of all-rod retinas.

Olfactory receptors. Snakes possess a greater reliance on olfaction than their more visually-oriented lizard ancestors (Castoe, et al. 2013; Dehara, et al. 2012; Mason 1992; Mason and Parker 2010; Vandewege, et al. 2016), which suggests that changes in olfactory receptor

repertoires between lizards and snakes may play an important role in the evolution and tuning of snake sensation. A comparison among six representative squamate reptile species identified several striking differences between the OR repertoires of snakes and lizards, most notably that the two lizards possess substantially fewer OR genes than snakes (Supplementary Table S6). The two lizards, the bearded dragon and anole, are visual insectivores and likely do not rely heavily on a sense of smell. In contrast, the high abundance of ORs in snakes is likely linked to the reliance of snakes on detecting odorants to find prey and mates (Mason and Parker 2010). Yet, despite the copy number variation observed between snakes and lizards, the overall composition of the OR repertoire was relatively similar (Figure 5A), with several exceptions. For example, there was a slightly higher abundance of Class I, subfamily 52 ORs in snakes (Figure 5A). Class I ORs are hypothesized to have a higher affinity for aquatic odorants and are typically uncommon in terrestrial vertebrates, excluding turtles (Wang, et al. 2013). To visualize the expansion and contraction history of ORs, we estimated the phylogeny of fully intact ORs for squamates in our analysis (Figure 5B-D). Within most subfamilies, there were clear snake specific expansions (Figure 5B-C), although no obvious expansions appear to take place specifically within the lineage leading to *Thamnophis* (Figure 5D). Additionally, *Thamnophis* had the smallest OR repertoire among snakes (Supplementary Table S6), with almost 180 fewer OR genes than the Burmese python. These results agree with a previous study (Vandewege, et al. 2016) that suggested significant expansions of ORs during snake evolution. However, here we provide new evidence that no major changes in snake OR repertoires appear to have occurred early in snake evolution, indicating that differential expansions of ORs do not seem to be associated with shifts in life history or diversity among snakes.

Venom and venom encoding genes. Venom is a key adaptation for many colubroid snakes that facilitates the capture of prey. While the majority of medically important venomous snakes are restricted to the families *Viperidae* and *Elapidae*, venom originated early in squamate evolution and, consequently, most advanced snakes are venomous to some degree (Fry, et al. 2008). This includes many rear-fanged venomous species of colubrid snakes, such as *T. sirtalis*. Although venom genes and gene families have been extensively studied in snakes, only a few studies have examined them in the context of a high-quality reference genome (Aird, et al. 2017; Reyes-Velasco, et al. 2014; Vonk, et al. 2013; Yin, et al. 2016). Additionally, the lack of genomic resources for rear-fanged venomous colubrid snake species has prevented detailed analysis of the genes involved in this venom system specifically. We were therefore motivated to examine a subset of known venom gene families in the garter snake and analyze the composition of its secreted venom to provide important context for the evolution of venom repertoires in rear-fanged snake species.

We analyzed the *T. sirtalis* genome for evidence of venom homologs related to the cysteine-rich secretory protein (CRISP), snake venom metalloproteinase (SVMP), cystatin, hyaluronidase, L-amino acid oxidase (LAAO), phospholipase A_2 (PLA_2), and snake venom serine protease (SVSP) toxin families. Phylogenetic reconstructions of genes identified in the *T. sirtalis* genome and related homologs from other snake and vertebrate species revealed a total of 15 garter snake genes nested within snake venom gene clades (Supplementary Figure S6-S12), indicative of a role in venom (based on forming a phylogenetic cluster with known snake venom toxins). We identified multiple paralogs for the SVMP (seven), CRISP (two) and LAAO (two) toxin families. In addition, we found single orthologous venom genes in the garter snake genome for cystatin, hyaluronidase, SVSP and PLA_2 toxin families. We investigated linkage among venom genes

based on our assembly for the garter snake and found that both CRISP genes are located on the same scaffold in the assembly and are separated by two long non-coding RNA genes, while all cystatins (including the venom paralog) are located on different scaffolds. Interestingly, we find evidence for duplications in two distinct clades of SVMP in *T. sirtalis* (Supplementary Figure S12), and three of the seven identified SVMPs are located on the same scaffold and in tandem, despite not all being members of the same clade. However, the annotation of one of these SVMPs (XM_014070785) is questionable because this gene and a second SVMP have exons that appear to be interspersed but not overlapping, potentially representing an erroneous assembly or mis-annotation of this region. A fourth gene (XM_014070779) in this region is also annotated as a SVMP, but this gene did not produce a significant BLAST hit to known SVMPs, and was therefore excluded from analyses. In sum, there remains some question as to the number of SVMP paralogs in the garter snake genome, which may be between six and eight; regardless, the SVMP toxin family encodes more paralogs in the *T. sirtalis* genome than any of the other detected toxin types. To summarize patterns of venom gene family evolution across squamate reptiles, we compared the total number of genes identified for the SVMP, CRISP, and cystatin toxin families in the garter snake genome to numbers identified previously in the genomes of two venomous snakes, two nonvenomous snakes, and a nonvenomous lizard (Yin, et al. 2016). Notably, we find that the genome of the garter snake, a lineage not often thought of as venomous, contains similar numbers of genes in these venom families as do snakes that are considered highly venomous (e.g., pitvipers; Supplementary Figure S13).

To assess whether these putative toxin genes contribute to the venom secretions of *T. sirtalis*, we compared their expression levels in the transcriptomes of various tissues, including the venom gland (Duvernoy's gland). There is growing evidence that many venom toxins have evolved

from genes that are co-expressed in at least a few different tissue types, and following their recruitment, exhibit significantly higher levels of transcription in the venom gland than other tissues (Hargreaves, et al. 2014; Junqueira-de-Azevedo, et al. 2015; Reyes-Velasco, et al. 2014; Vonk, et al. 2013). We compared the expression level of the venom homologs identified in the *T. sirtalis* genome in eight different tissues: the venom gland, brain, kidney, liver, lung, ovaries and the lower and upper segments of the digestive tract. Of the 15 genes encoding venom toxins found in the garter snake, 13 were expressed in the venom gland; however, only eight of those genes exhibited high levels of expression consistent with their use as venom toxins (>10,000 counts per million (CPM)) (Figure 6A). These eight toxins consisted of the seven SVMP genes (10,013-119,661 CPM) and one CRISP (157,558 CPM; shown in yellow in Figure 6A). In contrast, the remaining CRISP gene showed very low expression (0.42 CPM), comparable to expression levels in other tissues (0.07-0.66 CPM). Twelve of the 15 venom genes were expressed in more than two tissue types, with only hyaluronidase restricted to expression solely in the venom gland (albeit at low levels; 3.73 CPM; Figure 6A). In contrast, all eight of the toxins with high expression levels in the venom gland were also expressed in at least five of the seven other tissue types sampled, and in each case venom gland expression was >1,000 fold greater than elsewhere (Figure 6A). These findings of co-expression across tissue types, with high expression in the venom gland, are consistent with current hypotheses of venom evolution (Hargreaves, et al. 2014; Junqueira-de-Azevedo, et al. 2015; Reyes-Velasco, et al. 2014; Vonk, et al. 2013).

To link patterns of venom gene transcription with venom proteins, we analyzed venom protein content using MALDI-TOF-MS and RP-HPLC of crude venom, which demonstrated that the venom of *T. sirtalis* consists primarily (>80%) of two protein families: CRISPs and SVMPs

(Figure 6B,C). Consistent with transcriptomic data, CRISPs and SVMPs are highly abundant in the Duvernoy's venom gland (Figure 6A,B), and the SVMPs in particular appear to have multiple proteoforms that elute at different times following RP-HPLC (Figure 6C, Supplementary Figure S14). CRISPs and SVMPs (particularly PIII SVMPs) are common and often abundant components of rear-fanged snake venoms (Junqueira-de-Azevedo, et al. 2016; Mackessy 2002; Mackessy and Saviola 2016; Pla, et al. 2017). Many rear-fanged snake venom SVMPs have been shown to promote haemorrhage, suggesting an important role in feeding ecology, but the biological activities of most CRISPs are poorly defined at best, and their role in envenomation is unknown. Analysis of minor protein peaks by SDS-PAGE and MALDI-TOF-MS confirms the presence of numerous protein families (Figure 6B, Supplementary Figure S14), most of which contribute to total venom protein by less than 1% each (e.g., LAAO; Supplementary Figure S15) and are transcribed in the Duvernoy's venom gland of *T. sirtalis* at low to moderate levels (Figure 6A). Their contribution to envenomation sequelae, if any, is therefore likely to be minimal relative to the highly expressed SVMPs and CRISPs. Anecdotal support of this conclusion is provided by a case of human envenomation by *T. s. sirtalis*; following extended contact by the snake, the bite resulted in minor localized edema and ecchymosis, likely due to the effects of SVMPs, but resolved quickly (Hayes and Hayes 1985).

Evolution of tetrodotoxin resistance

Thamnophis sirtalis is among the few vertebrates capable of withstanding high concentrations of tetrodotoxin (TTX) found in newts of the genus *Taricha* (Brodie 1990; Brodie, et al. 2002; Feldman, et al. 2009). Populations of garter snakes and newts in western North America show a rough match between snake resistance and newt toxicity, a pattern indicative of an ongoing coevolutionary arms race (Brodie, et al. 2002; Hanifin and Brodie 2008). TTX binds to specific

amino acids in the four outer pore regions ("P-loops") of voltage-gated sodium channels (Na_v, encoded by genes in the family *SCNA*), blocking ion conduction and impairing physiological function (Fozzard and Lipkind 2010). Garter snake resistance to TTX is known to derive from evolutionary changes in these P-loops (Feldman, et al. 2009; Geffeney, et al. 2002; Geffeney, et al. 2005; McGlothlin, et al. 2014).

Previous work has shown that *T. sirtalis* possesses at least three TTX-resistant Na_v channels: the skeletal muscle channel $Na_v1.4$ (encoded by the gene *SCN4A*), and two peripheral nerve channels $Na_v1.6$ (*SCN8A*, which is also expressed in the brain) and $Na_v1.7$ (*SCN9A*) (Geffeney, et al. 2005; McGlothlin, et al. 2014). In contrast, three channels found only in the central nervous system ($Na_v1.1$, 1.2, and 1.3, encoded by *SCN1A*, *2A*, and *3A*, respectively) (Catterall, et al. 2005) and thus shielded by the blood-brain barrier (Zimmer 2010), do not appear to be TTX-resistant (McGlothlin, et al. 2014). Peripheral nerves evolved TTX resistance significantly earlier than did muscle, with resistant $Na_v1.7$ evolving at least 170 mya in non-snake squamates and resistant $Na_v1.6$ evolving ~40 mya in three different snake families (McGlothlin, et al. 2016). Resistance in $Na_v1.4$ evolved within *T. sirtalis* and is polymorphic, with more highly resistant alleles found in populations that overlap with more toxic newts (Feldman, et al. 2010; Geffeney, et al. 2005; Hague, et al. 2017). At least four other snake species have also evolved TTX resistance in $Na_v1.4$ due to interactions with toxic amphibians (Feldman, et al. 2009; Feldman, et al. 2012), all of which occur in lineages with resistant $Na_v1.6$ and 1.7 (McGlothlin, et al. 2016).

The three other members of the voltage-gated sodium channel, $Na_v1.5$ (found in cardiac muscle and encoded by the gene *SCN5A*), 1.8, and 1.9 (both found in small sensory neurons and encoded by *SCN10A* and *11A*, respectively) have not been previously described in *T. sirtalis*. These three

channels are closely related, arising from tandem gene duplications before the mammal-reptile split (Dib-Hajj, et al. 1999; Zakon, et al. 2010). In mammals, all three channels display TTX resistance due to the substitution of a non-aromatic residue for an aromatic one in the domain I (DI) P-loop (Akopian, et al. 1996; Backx, et al. 1992; Cummins, et al. 1999; Leffler, et al. 2005; Satin, et al. 1992). Previously, TTX resistance of these channels was thought to be ubiquitous in amniotes, with resistance arising before the gene duplication events, but recent work on $Na_v1.5$ has called this into question (Vornanen, et al. 2011). Therefore, reconstructing the history of these genes is important for a complete understanding of physiological resistance to TTX in *T. sirtalis.* Because each of these three channels is expressed in the periphery and thus potentially exposed to ingested TTX, we predict that all three paralogs should be resistant to TTX in garter snakes. This resistance may have either been inherited from an ancient common ancestor or evolved more recently in response to selection imposed by toxic prey.

From the current genome assembly, we obtained partial or complete coding sequences for all nine *SCNA* genes. Here, we focus on the three genes previously undescribed in *T. sirtalis*; additional detail on the location of annotated *SCNA* genes can be found in Supplementary Information 2.1. We found evidence suggesting that $Na_v1.5$, $Na_v1.8$, and $Na_v1.9$ are all either moderately or strongly resistant to TTX in *T. sirtalis* (Supplementary Table S7). $Na_v1.5$ has a cysteine in DI (Y371C) that is known to confer ~3,000-fold resistance to TTX in mammalian $Na_v1.5$ and pufferfish $Na_v1.4a$ (Backx, et al. 1992; Penzotti, et al. 1998; Perez-Garcia, et al. 1996; Satin, et al. 1992; Venkatesh, et al. 2005). (For each substitution we discuss, the ancestral amino acid refers to mammalian $Na_v1.4$, a TTX-sensitive channel, and the position number refers to the location for the specific *T. sirtalis* paralog.) DI also has an arginine (N374R) known to provide two-fold resistance to TTX (Terlau, et al. 1991). A substitution in domain III (DIII; D1425S)

occurs at a site known to influence TTX binding, but has unknown effects. Taken together, the presence of these amino acids suggests that the *T. sirtalis* $Na_v1.5$ should be blocked only at micromolar to millimolar concentrations of TTX (as opposed to nanomolar for TTX-sensitive channels).

Surprisingly, $Na_v1.8$ DI does not have Y348S as in mammals, but does possess N351R, homologous to the arginine in $Na_v1.5$ DI. DIII has a threonine (M1376T) known to provide ~15-fold resistance, which also occurs in several pufferfish channels (Jost, et al. 2008) and in the garter snake *T. couchii* $Na_v1.4$ (Feldman, et al. 2009). This suggests that *T. sirtalis* $Na_v1.8$ can probably function up to TTX concentrations of about 1 μM. $Na_v1.9$ is likely to be extremely TTX-resistant, due to Y362S (~7,000-fold) (Leffler, et al. 2005) and N365R in DI . A DII serine (E767S) and a DIV glutamate (G1479E) may provide additional TTX resistance, although these particular substitutions have not been tested. Thus, as in mammals, *T. sirtalis* $Na_v1.9$ is probably resistant to TTX in the millimolar range.

Ancestral reconstructions confirm that extreme resistance in $Na_v1.5$ and $Na_v1.9$ is ancient, dating to the common ancestor of mammals and reptiles (~320 mya; Figure 7) (Hedges, et al. 2015). All species we examined had a DI substitution providing micromolar resistance in $Na_v1.9$. In $Na_v1.5$, Y371C also dates to an ancient common ancestor, although some groups, including *Gallus* and *Anolis*, have lost this substitution. In contrast, ancestral $Na_v1.8$ shows only mild resistance via N351R. Although several lineages have independently evolved increased resistance in this channel, the M1376T found in *T. sirtalis* is unique. Based on our reconstruction, this substitution apparently evolved sometime within the last 50 my, after the split between colubrid and elapid snakes (Figure 7) (Zheng and Wiens 2016).

Combined with previous work, these results indicate that TTX resistance in *T. sirtalis* involves six out of the nine genes in the *SCNA* family, with only the brain channels $Na_v1.1$, $Na_v1.2$, and $Na_v1.3$ remaining completely TTX sensitive. Snakes were predisposed to evolve physiological TTX resistance because they inherited strong TTX resistance in three channels—$Na_v1.5$, $Na_v1.7$, and $Na_v1.9$—from their common ancestor. Later, $Na_v1.6$ evolved TTX resistance in three families, perhaps in response to eating amphibians, facilitating evolution of resistance in $Na_v1.4$ via co-evolutionary arms races with amphibians in five species (McGlothlin, et al. 2016). Our results show that a third channel, $Na_v1.8$, although mildly resistant to TTX for 300 my, has shown an evolutionary increase in TTX resistance within the last 50 million years, which may have also been selected for by an amphibian diet. Additional research is needed to determine the timing of this substitution and the selective factors involved.

CONCLUSION

Our analysis of the garter snake genome, together with the genomes of other amniotes, provides new insight into vertebrate genome structure, function, and evolution, and provides genomic context to key aspects of garter snake biology. Our analyses of genome evolution and structure reveal major shifts in repeat element landscape composition and expansion in the garter snake and other colubroid snakes, and identify both snake- and garter snake-specific microRNAs that may play a role in unique features of snake biology such as venom. We identified a small number of genes under positive selection in garter snakes and colubrids in general, and estimates of neutral substitution rates across four-fold degenerate sites suggest that the rate of substitutions in snakes is similar to that in mammals, and generally slightly lower than that of lizards. Identification of Z and W sex chromosome-specific scaffolds in the garter snake genome assembly highlight the value of this genome resource for future studies of sex chromosome

evolution. Evolutionary analyses of gene families involved in snake sensory systems reveal patterns of gene loss in visual gene families that strongly support dim-light ancestry in snakes, suggesting that the all-cone retina of *Thamnophis* evolved through transmutation of rods into cone-like cells. Additionally, these analyses provide evidence for a major expansion of olfactory receptor repertoires early in snake evolution, but with no further expansions associated with subsequent diversification or life history shifts in the snake lineage. Finally, we provide one of the first detailed analyses to characterize and link the protein composition of venom, the genes that encode these proteins, and their evolutionary origins in a rear-fanged venomous colubrid species, together with new insight into the evolution of TTX resistance in some garter snakes in response to toxic newt prey. Collectively, our analyses demonstrate a broad spectrum of genomic adaptations linked to the many extreme and unique features of snakes, while also highlighting the variation in genomic features among snakes that together demonstrate why snakes represent intriguing and valuable model systems for diverse research questions.

ACKNOWLEDGEMENTS

We thank the production sequencing group at the McDonnell Genome Institute for assistance with genome sequencing, assembly and annotation, and John Abramyan for assistance with SCNA annotations in the chicken and turtle. The work was supported by funding from the National Human Genome Research Institute at National Institutes of Health [5U54HG00307907 to W.C.W]; the National Science Foundation [DEB-1655571 and IOS-1655735 to T.A.C]; a Natural Sciences and Engineering Research Council of Canada Discovery Grant [to B.S.W.C., K.B.S., and K.K.B.]; an Academia Sinica award [to C.H.K.]; a Minnesota Supercomputing Institute award [to S.E.M.]; and a Grand Challenges in Biology Postdoctoral Fellowship [to C.N.P.].

REFERENCES

Adams RH, et al. 2016. Microsatellite landscape evolutionary dynamics across 450 million years of vertebrate genome evolution. Genome 59: 295-310.

Aird SD, et al. 2017. Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. Genome Biol. Evol. 9: 2640-2649.

Akopian AN, Sivilotti L, Wood JN 1996. A tetrodotoxin-resistant voltage-gated sodium channel expressed by sensory neurons. Nature 379: 257-262.

Altschul SF, et al. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Ambros V 2004. The functions of animal microRNAs. Nature 431: 350-355.

Angielczyk KD, Schmitz L 2014. Nocturnality in synapsids predates the origin of mammals by over 100 million years. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 281: 20141642.

Arnemann J, et al. 1986. Clustered GATA repeats (Bkm sequences) on the human Y chromosome. Hum. Genet. 73: 301-303.

Backx PH, et al. 1992. Molecular localization of an ion-binding site within the pore of mammalian sodium channels. Science 257: 248-252.

Baker RJ, Mengden GA, Bull JJ 1972. Karyotypic studies of thirty-eight species of North American snakes. Copeia: 257-265.

Bartel DP 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281-297.

Bernard MA, et al. 2001. Diminished levels of the putative tumor suppressor proteins EXT1 and EXT2 in exostosis chondrocytes. Cytoskeleton 48: 149-162.

Bhattacharyya N, et al. 2017. Cone-like rhodopsin expressed in the all cone retina of the colubrid pine snake as a potential adaptation to diurnality. J. Exp. Biol.: jeb. 156430.

Bickelmann C, et al. 2015. The molecular origin and evolution of dim‐light vision in mammals. Evolution 69: 2995-3003.

Boetzer M, et al. 2010. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27: 578-579.

Bohlen CJ, et al. 2011. A heteromeric Texas coral snake toxin targets acid-sensing ion channels to produce pain. Nature 479: 410.

Bouckaert R, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comp. Biol. 10: e1003537.

Bradnam KR, et al. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. GigaScience 2: 10.

Brodie ED 1990. Tetrodotoxin resistance in garter snakes: an evolutionary response of predators to dangerous prey. Evolution 44: 651-659.

Brodie ED, III, Brodie ED, Jr. 1999. Predator-prey arms races. Bioscience 49: 557-568.

Brodie ED, Jr., Ridenhour BJ, Brodie ED, III 2002. The evolutionary response of predators to dangerous prey: Hotspots and coldspots in the geographic mosaic of coevolution between newts and snakes. Evolution 56: 2067-2082.

Bromham L 2002. Molecular clocks in reptiles: life history influences rate of molecular evolution. Mol. Biol. Evol. 19: 302-309.

Bromham L, Penny D 2003. The modern molecular clock. Nature Reviews Genetics 4: 216.

Castoe TA, et al. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proceedings of the National Academy of Sciences 110: 20645-20650.

Castoe TA, et al. 2011. Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. Genome Biol. Evol. 3: 641-653.

Catterall WA, Perez-Reyes E, Snutch TP, Striessnig J 2005. Nomenclature and structure-function relationships of voltage-gated calcium channels, international union of pharmacology. XLVIII. Pharmacol. Rev. 57: 411-425.

Crews D, et al. 1984. Hormonal independence of courtship behavior in the male garter snake. Horm. Behav. 18: 29-41.

Cummins TR, et al. 1999. A novel persistent tetrodotoxin-resistant sodium current in SNS-null and wild-type small primary sensory neurons. J. Neurosci. 19: RC43-RC43.

Davies WL, et al. 2009. Shedding light on serpent sight: the visual pigments of henophidian snakes. J. Neurosci. 29: 7519-7525.

Dehara Y, et al. 2012. Characterization of squamate olfactory receptor genes and their transcripts by the high-throughput sequencing approach. Genome Biol. Evol. 4: 602-616.

Dib-Hajj SD, et al. 1999. Coding sequence, genomic organization, and conserved chromosomal localization of the mouse gene Scn11a encoding the sodium channel NaN. Genomics 59: 309-318.

Edgar RC 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792-1797.

Effron M, Griner L, Benirschke K 1977. Nature and rate of neoplasia found in captive wild mammals, birds, and reptiles at necropsy. J. Natl. Cancer Inst. 59: 185-198.

Enright AJ, et al. 2003. MicroRNA targets in Drosophila. Genome Biol. 5: R1.

Feldman CR, Brodie ED, Brodie ED, Pfrender ME 2009. The evolutionary origins of beneficial alleles during the repeated adaptation of garter snakes to deadly prey. Proc. Natl. Acad. Sci. USA 106: 13415-13420.

Feldman CR, Brodie ED, Pfrender ME 2010. Genetic architecture of a feeding adaptation: garter snake (Thamnophis) resistance to tetrodotoxin bearing prey. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 277: 3317-3325.

Feldman CR, Brodie Jr. ED, Brodie III ED, Pfrender ME 2012. Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. Proc. Natl. Acad. Sci. USA 109: 4556-4561.

Fozzard HA, Lipkind GM 2010. The tetrodotoxin binding site is within the outer vestibule of the sodium channel. Mar. Drugs 8: 219-234.

Fry BG, et al. 2008. Evolution of an arsenal structural and functional diversification of the venom system in the advanced snakes (Caenophidia). Mol. Cell. Proteomics 7: 215-246.

Gamble T, et al. 2017. The discovery of XY sex chromosomes in a boa and python. Curr. Biol. 27: 2148-2153. e2144.

Gamble T, Geneva AJ, Glor RE, Zarkower D 2014. Anolis sex chromosomes are derived from a single ancestral pair. Evolution 68: 1027-1041.

Gamble T, Greenbaum E, Jackman TR, Bauer AM 2015. Into the light: diurnality has evolved multiple times in geckos. Biol. J. Linn. Soc. 115: 896-910.

Garner MM, Hernandez-Divers SM, Raymond JT 2004. Reptile neoplasia: a retrospective study of case submissions to a specialty diagnostic service. Vet. Clin. North Am. Exot. Anim. Pract. 7: 653-671.

Geffeney S, Ruben PC, Brodie ED, Jr., Brodie ED, III 2002. Mechanisms of adaptation in a predator-prey arms race: TTX resistant sodium channels. Science 297: 1336-1339.

Geffeney SL, et al. 2005. Evolutionary diversification of TTX-resistant sodium channels in a predator-prey interaction. Nature 434: 759-763.

Georges A, et al. 2015. High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard Pogona vitticeps. Gigascience 4: 45.

Gnerre S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. USA 108: 1513-1518.

Green RE, et al. 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science 346: 1254449.

Gregory TR. 2001. Animal genome size database. In: TR Gregory.

Hague MTJ, Feldman CR, Brodie ED 2017. Convergent adaptation to dangerous prey proceeds through the same first‐step mutation in the garter snake Thamnophis sirtalis. Evolution 71: 1504-1518.

Hall MI, Kamilar JM, Kirk EC 2012. Eye shape and the nocturnal bottleneck of mammals. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 279: 4962-4968.

Hanifin CT, Brodie ED 2008. Phenotypic mismatches reveal escape from arms-race coevolution. PLoS Biol. 6: 471-482. doi: 10.1371/journal.pbio.0060060|ISSN 1544-9173

Hargreaves AD, et al. 2014. Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins. Genome Biol. Evol. 6: 2088-2095.

Hayes WK, Hayes FE 1985. Human envenomation from the bite of the eastern garter snake, Thamnophis s. sirtalis (Serpentes: Colubridae). Toxicon 23: 719-721.

Hedges SB, et al. 2015. Tree of life reveals clock-like speciation and diversification. Mol. Biol. Evol. 32: 835-845.

Herzog HA, Jr., Burghardt GM 1986. Development of antipredator responses in snakes: I. Defensive and open-field behaviors in newborns and adults of three species of garter snakes (*Thamnophis melanogaster, T. sirtalis, T. butleri*). J. Comp. Psychol. 100: 372-379.

Herzog J, H. A., Burghardt GM 1988. Development of antipredator responses in snakes III. Stability of individual and litter differences over the first year of life. Ethology 77: 250-258.

Hruska-Hageman AM, Wemmie JA, Price MP, Welsh MJ 2002. Interaction of the synaptic protein PICK1 (protein interacting with C kinase 1) with the non-voltage gated sodium channels BNC1 (brain Na+ channel 1) and ASIC (acid-sensing ion channel). Biochem. J. 361: 443.

Jaglin XH, Chelly J 2009. Tubulin-related cortical dysgeneses: microtubule dysfunction underlying neuronal migration defects. Trends Genet. 25: 555-566.

Jansen DW, Foehring RC 1983. The mechanism of venom secretion from Duvernoy's gland of the snake Thamnophis sirtalis. J. Morphol. 175: 271-277.

Jayne BC, Bennett AF 1989. The effect of tail morphology on locomotor performance of snakes: a comparison of experimental and correlative methods. J. Exp. Zool. 252: 126-133.

Jones KW, Singh L 1985. Snakes and the evolution of sex chromosomes. Trends Genet. 1: 55-61.

Jost MC, et al. 2008. Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. Mol. Biol. Evol. 25: 1016-1024.

Junqueira-de-Azevedo ILM, et al. 2015. Venom-related transcripts from Bothrops jararaca tissues provide novel molecular insights into the production and evolution of snake venom. Mol. Biol. Evol. 32: 754-766.

Junqueira-de-Azevedo ILM, Campos PF, Ching ATC, Mackessy SP 2016. Colubrid venom composition: An-omics perspective. Toxins 8: 230.

Kerchove CM, et al. 2008. α1-adrenoceptors trigger the snake venom production cycle in secretory cells by activating phosphatidylinositol 4, 5-bisphosphate hydrolysis and ERK signaling pathway. Comp. Biochem. Physiol., A: Mol. Integr. Physiol. 150: 431-437.

Krämer A, Green J, Pollard Jr J, Tugendreich S 2013. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics 30: 523-530.

Kubie JL, Vagvolgyi A, Halpern M 1978. Roles of the vomeronasal and olfactory systems in courtship behavior of male garter snakes. J. Comp. Physiol. Psychol. 92: 627.

Kumar S, Stecher G, Suleski M, Hedges SB 2017. TimeTree: a resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. 34: 1812-1819.

Laopichienpong N, et al. 2017. Evolutionary Dynamics of the Gametologous CTNNB1 Gene on the Z and W Chromosomes of Snakes. J. Hered. 108: 142-151.

Leffler A, et al. 2005. Pharmacological properties of neuronal TTX-resistant sodium channels and the role of a critical serine pore residue. Pflügers Archiv 451: 454-463.

Liu B, et al. 2013. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. arXiv preprint arXiv:1308.2012.

Mackessy SP 2002. Biochemistry and pharmacology of colubrid snake venoms. J. Toxicol.: Toxin Rev. 21: 43-83.

Mackessy SP, Saviola AJ 2016. Understanding biological roles of venoms among the Caenophidia: The importance of rear-fanged snakes. Integr. Comp. Biol. 56: 1004-1021.

Marçais G, Kingsford C 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27: 764-770.

Martin AP, Palumbi SR 1993. Body size, metabolic rate, generation time, and the molecular clock. Proceedings of the National Academy of Sciences 90: 4087-4091.

Mason RT. 1992. Reptilian Pheromones. In: Gans C, Crews D, editors. Hormones, Brain, and Behavior. Biology of the Reptilia. Physiology E, vol 18. Chicago: University of Chicago Press. p. 114-228.

Mason RT, Crews D. 1986. Pheromone mimicry in garter snakes. In: Duvall D, Muller-Schwarze D, Silverstein RM, editors. Chemical Signals In Vertebrates 4: Plenum Publishing Corporation.

Mason RT, Parker MR 2010. Social behavior and pheromonal communication in reptiles. J. Comp. Physiol., A 196: 729-749.

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. USA 103: 18190-18195. doi: 0605274103 [pii] 10.1073/pnas.0605274103

McGaugh SE, et al. 2015. Rapid molecular evolution across amniotes of the IIS/TOR network. Proc. Natl. Acad. Sci. USA 112: 7055-7060.

McGlothlin JW, et al. 2014. Parallel evolution of tetrodotoxin resistance in three voltage-gated sodium channel genes in the garter snake Thamnophis sirtalis. Mol. Biol. Evol. 31: 2836-2846.

McGlothlin JW, et al. 2016. Historical contingency in a multigene family facilitates adaptive evolution of toxin resistance. Curr. Biol. 26: 1616-1621.

Mendonca MT, Crews D 1996. Effects of ovariectomy and estrogen replacement on attractivity and receptivity in the red sided garter snake (Thamnophis sirtalis parietalis). J. Comp. Physiol., A 178: 373-381.

O'Meally D, et al. 2010. Non-homologous sex chromosomes of birds and snakes share repetitive sequences. Chromosome Res. 18: 787-800.

Ohno S. 2013. Sex chromosomes and sex-linked genes: Springer Science & Business Media.

Parasnis AS, et al. 1999. Microsatellite (GATA) n reveals sex-specific differences in papaya. Theor. Appl. Genet. 99: 1047-1052.

Peace RJ, Biggar KK, Storey KB, Green JR 2015. A framework for improving microRNA prediction in non-human genomes. Nucleic Acids Res. 43: e138-e138.

Penzotti JL, Fozzard HA, Lipkind GM, Dudley SC 1998. Differences in saxitoxin and tetrodotoxin binding revealed by mutagenesis of the Na+ channel outer vestibule. Biophys. J. 75: 2647-2657.

Perez-Garcia MT, Chiamvimonvat N, Marban E, Tomaselli GF 1996. Structure of the sodium channel pore revealed by serial cysteine mutagenesis. Proc. Natl. Acad. Sci. USA 93: 300-304.

Pla D, et al. 2017. What killed Karl Patterson Schmidt? Combined venom gland transcriptomic, venomic and antivenomic analysis of the South African green tree snake (the boomslang), Dispholidus typus. Biochimica et Biophysica Acta (BBA)-General Subjects 1861: 814-823.

Reyes-Velasco J, et al. 2014. Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. Mol. Biol. Evol. 32: 173-183.

Robert KA, Bronikowski AM 2010. Evolution of senescence in nature: physiological evolution in populations of garter snake with divergent life histories. Am. Nat. 175: 147-159.

Rovatsos M, Vukić J, Lymberakis P, Kratochvíl L editors. Proc. R. Soc. B. 2015.

Satin J, et al. 1992. A mutant of TTX-resistant cardiac sodium channels with TTX-sensitive properties. Science 256: 1202-1205.

Schott RK, et al. 2016. Evolutionary transformation of rod photoreceptors in the all-cone retina of a diurnal garter snake. Proc. Natl. Acad. Sci. USA 113: 356-361.

Schott RK, et al. 2017. Targeted capture of complete coding regions across divergent species. Genome Biol. Evol. 9: 398-414.

Schott RK, et al. 2018. Shifts in selective pressures on snake phototransduction genes associated with photoreceptor transmutation and dim-light ancestry. Mol. Biol. Evol. In Press.

Sillman AJ, et al. 1997. The photoreceptors and visual pigments of the garter snake (Thamnophis sirtalis): a microspectrophotometric, scanning electron microscopic and immunocytochemical study. J. Comp. Physiol., A 181: 89-101.

Simão FA, et al. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210-3212.

Simões BF, et al. 2015. Visual system evolution and the nature of the ancestral snake. J. Evol. Biol. 28: 1309-1320.

Simoes BF, et al. 2016. Multiple rod–cone and cone–rod photoreceptor transmutations in snakes: evidence from visual opsin gene expression. Proc. R. Soc. B 283: 20152624.

Smit AFA, Hubley R, Green P 2014. RepeatModeler Open-1.0. 2008-2010. Access date Dec.

Smit AFA, Hubley R, Green P 2015. RepeatMasker Open-4.0. 2013–2015. Institute for Systems Biology. http://reapeatmasker.org.

Stamatakis A 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313.

Székely G, et al. 2008. Duplicated P5CS genes of Arabidopsis play distinct roles in stress regulation and developmental control of proline biosynthesis. Plant J. 53: 11-28.

Terlau H, et al. 1991. Mapping the site of block by tetrodotoxin and saxitoxin of sodium channel II. FEBS Lett. 293: 93-96.

Thibaud-Nissen F, et al. 2013. Eukaryotic genome annotation pipeline.

Tollis M, et al. 2018. Comparative Genomics Reveals Accelerated Evolution in Conserved Pathways during the Diversification of Anole Lizards. Genome Biol. Evol.

Tsutsumi S, et al. 2004. The novel gene encoding a putative transmembrane protein is mutated in gnathodiaphyseal dysplasia (GDD). Am. J. Hum. Genet. 74: 1255-1261.

Uetz P, Etzold T 1996. The EMBL/EBI reptile database. Herpetol. Rev 27: 175.

Ullate-Agote A, Milinkovitch MC, Tzika AC 2015. The genome sequence of the corn snake (Pantherophis guttatus), a valuable resource for EvoDevo studies in squamates. Int. J. Dev. Biol. 58: 881-888.

Underwood G 1970. The Eye. In: Gans C, editor. Biology of the Reptilia. New York: Academic Press.: 1-97.

Vandewege MW, et al. 2016. Contrasting patterns of evolutionary diversification in the olfactory repertoires of reptile and bird genomes. Genome Biol. Evol. 8: 470-480.

Venkatesh B, et al. 2005. Genetic basis of tetrodotoxin resistance in pufferfishes. Curr. Biol. 15: 2069-2072.

Vest DK 1981. The toxic Duvernoy's secretion of the wandering garter snake, Thamnophis elegans vagrans. Toxicon 19: 831-839.

Vicoso B, et al. 2013. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. PLoS Biol. 11: e1001643.

Vidal N, Hedges SB 2004. Molecular evidence for a terrestrial origin of snakes. Proc. R. Soc. Lond., Ser. B: Biol. Sci. 271: S226-S229.

Vinogradov AE 1998. Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. Cytometry 31: 100-109.

Vonk FJ, et al. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proc. Natl. Acad. Sci. USA 110: 20651-20656. doi: 10.1073/pnas.1314702110

Vornanen M, Hassinen M, Haverinen J 2011. Tetrodotoxin sensitivity of the vertebrate cardiac Na+ current. Mar. Drugs 9: 2409-2422.

Wade A, et al. 2013. Proteoglycans and their roles in brain cancer. The FEBS journal 280: 2399-2417.

Wall DP, et al. 2009. Comparative analysis of neurological disorders focuses genome-wide search for autism genes. Genomics 93: 120-129.

Walls GL editor. Biol. Symposia. 1942.

Wang Z, et al. 2013. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. Nat. Genet. 45: 701-706.

Widmark J, Sundström G, Ocampo Daza D, Larhammar D 2010. Differential evolution of voltage-gated sodium channels in tetrapods and teleost fishes. Mol. Biol. Evol. 28: 859-871.

Xue W, et al. 2013. L_RNA_scaffolder: scaffolding genomes with transcripts. BMC Genomics 14: 604.

Yang Z 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24: 1586-1591.

Yin W, et al. 2016. Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. Nat. Commun. 7.

Zakon HH, Jost MC, Lu Y 2010. Expansion of voltage-dependent Na+ channel gene family in early tetrapods coincided with the emergence of terrestriality and increased brain complexity. Mol. Biol. Evol. 28: 1415-1424.

Zheng Y, Wiens JJ 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. Mol. Phylogen. Evol. 94: 537-547.

Zimmer T 2010. Effects of tetrodotoxin on the mammalian cardiovascular system. Mar. Drugs 8: 741-762.

Figure Legends

Fig. 1. Comparison of genomic repeat element landscapes across squamate reptiles genomes. A) Summary of the repeat element content of the genome of *Thamnophis sirtalis* in comparison to other squamate reptiles. Branches on the time-calibrated consensus phylogeny are colored according to the estimated rate of genomic transposable element evolution. The associated heatmap shows the total repeat element and transposable element genomic content (%) for each taxon, as well as the genome coverage (% masked) of major components of the repeat element landscape. B) Kimura 2-parameter distance-based TE copy divergence analysis. Genome abundance (% of genome; y-axis) is shown for each transposable element type and is clustered according to the CpG-corrected Kimura distance (*K*-value from 0-70; x-axis) from each type's consensus sequence. Elements with low *K*-values are the least divergent from their respective consensus sequences, and likely represent more recent transposition events, while elements with higher *K*-values represented more degenerated copies that were likely inserted in the past.

Fig. 2. Estimated neutral substitution rates of squamate reptiles. A) Branch-specific estimates of substitution rates at four-fold degenerate third-codon positions of coding genes. B) Jitter plot of four-fold degenerate site substitution rate estimates for mammals, lizards, and snakes, including all terminal and ancestral branches allocated to a clade. Terminal branches are shown as closed circles and are labeled with the corresponding species, while ancestral branches are shown as open circles. The horizontal lines denote the average substitution rate estimate for each clade.

Fig. 3. Identification of the W chromosome and sex-chromosome-linked genes in *Thamnophis sirtalis*. A) Maximum likelihood trees of fragments of six genes with Z and W alleles in *T. sirtalis*. Circles at nodes represent bootstrap values: black circles indicate bootstrap values >70; white circles indicate bootstrap values <70. B) Female-specific amplification of PCR primers designed from *T. sirtalis* W scaffolds in two male and two female *T. sirtalis* samples for six genes. Primer sequences are in Supplementary Table S5. C) Fluorescent *in situ* hybridization (FISH) of a *Bkm*-like repeat, $(GATA)_n$ in a female *T. sirtalis*. Sex-specific hybridization on the W chromosome is indicated by an arrow.

Fig. 4. Visual system gene loss across amniote vertebrates and rod-like characteristics of some cone cells in the all cone retina of *Thamnophis sirtalis*. A) Presence of 119 visual and opsin genes in snakes and other ancestrally nocturnal groups. B-G) Immunohistochemical staining of mouse (control, B-D) and *T. sirtalis* (E-G) 20um transverse retinal cryosections with rhodopsin (RET-P1) and rod-specific transducin (K20) antibodies. Rhodopsin is found in a small subset of cone-like cells mainly in the outersegment (OS) of the photoreceptor (E). Rod transducin (F) is also found in the inner segments (IS) of these same photoreceptors (G). The nuclei of the photoreceptor cells (CB) is stained in blue, rhodopsin (RET-P1) staining is shown in red, and rod-specific transducin (K20) staining is shown in green. Scale bar = 10um.

Fig. 5. Genomic content and evolutionary history of olfactory receptors in squamate reptiles. A) A heatmap representing the relative percentage of intact olfactory receptor sequences belonging to each subfamily for each species. B-D) Neighbor joining tree of 1872 OR amino acid sequences: B) Branches are colored by annotated subfamily; C) as either snake or lizard; D) garter snake terminal branches are contrasted from the remaining species.

Fig. 6. Transcriptomic and proteomic analysis of *Thamnophis* venom. A) Tissue-specific expression of venom gene homologs in *T. sirtalis*. IDs of orthologous venom genes are red and bolded. B) MALDI-TOF mass spectrum of *T. sirtalis parietalis* (top) and *T. elegans vagrans* (bottom) venom, using a mass window of 4-85 kDa. Note that only one minor peak, at 12.55 kDa (cystatin), and two major peaks, at 25 kDa (CRISP proteins) and 50 kDa (SVMP PIII), are observed. C) 12% acrylamide NuPAGE SDS-PAGE gel of major RP-HPLC fractions of 1.0 mg *T. s. parietalis* venom on a Jupiter C18 column, as well as total venom from *T. s. parietalis* and *T. e. vagrans*, indicating major protein families represented. See supplementary figure S12 for additional details. **LAAO**, L-amino acid oxidase; **SVMP**, snake venom metalloproteinase; **SP**, serine protease; **CRISP**, cysteine-rich secretory protein; **CTL**, C-type lectin; **PLA2**, phospholipase A_2; **3FTx**, three-finger toxins.

Fig. 7. Evolutionary reconstruction of TTX resistance in SCN5A ($Na_v1.5$), SCN10A ($Na_v1.8$), and SCN11A ($Na_v1.9$) across amniote vertebrates. Tips and branches are color-coded by predicted TTX resistance, with warmer colors representing stronger resistance (see Supplementary Table S7 for details). Resistance-conferring substitutions found in *T. sirtalis* are plotted at their origin, with warmer colors indicating stronger resistance (Table S7). A non-resistant SCN5A-like paralog from *Xenopus* is used as the outgroup.

Figure 1

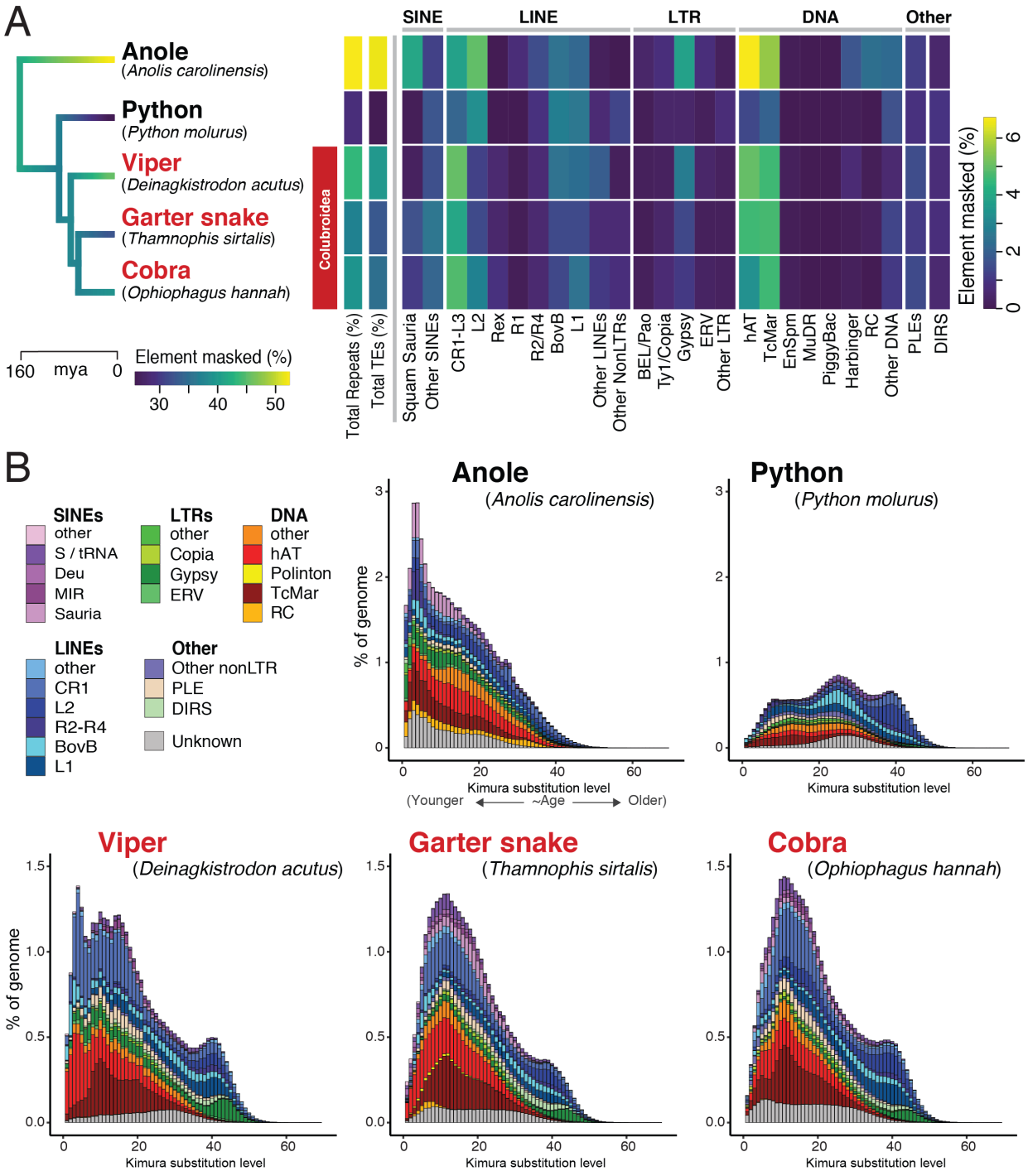

Figure 2

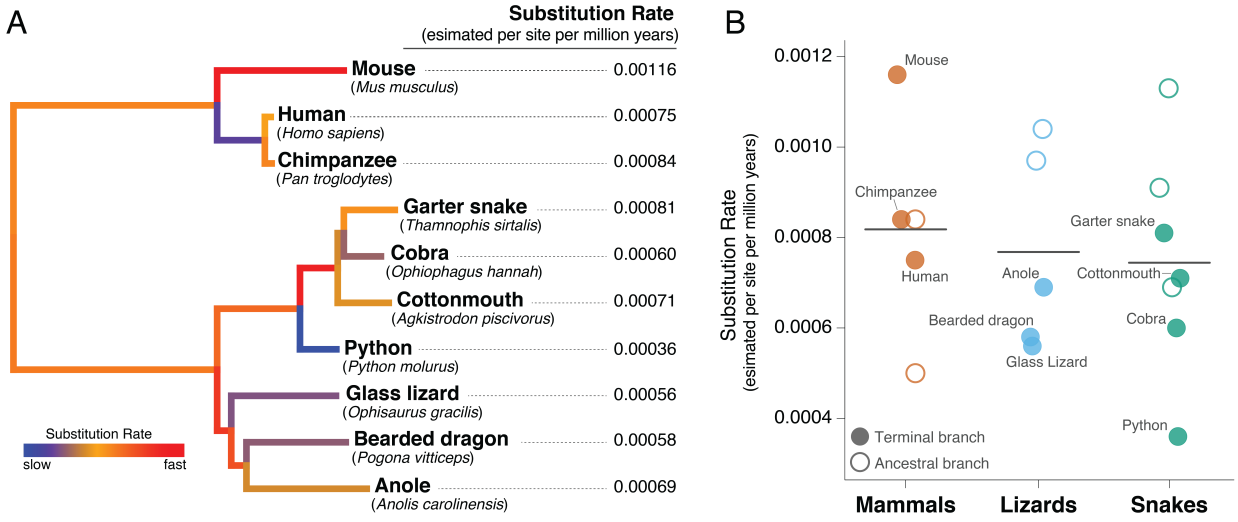

Figure 3

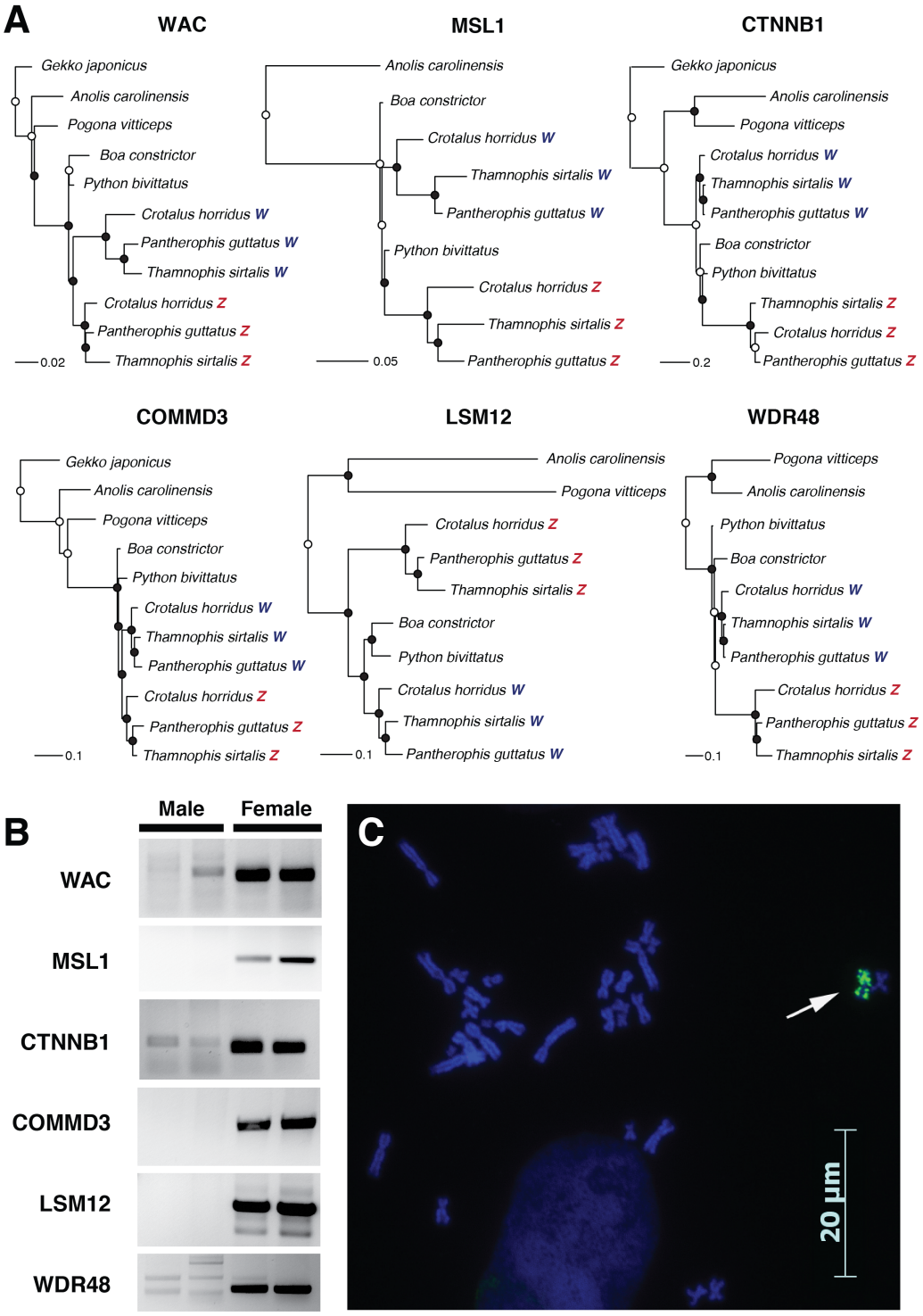

Figure 4

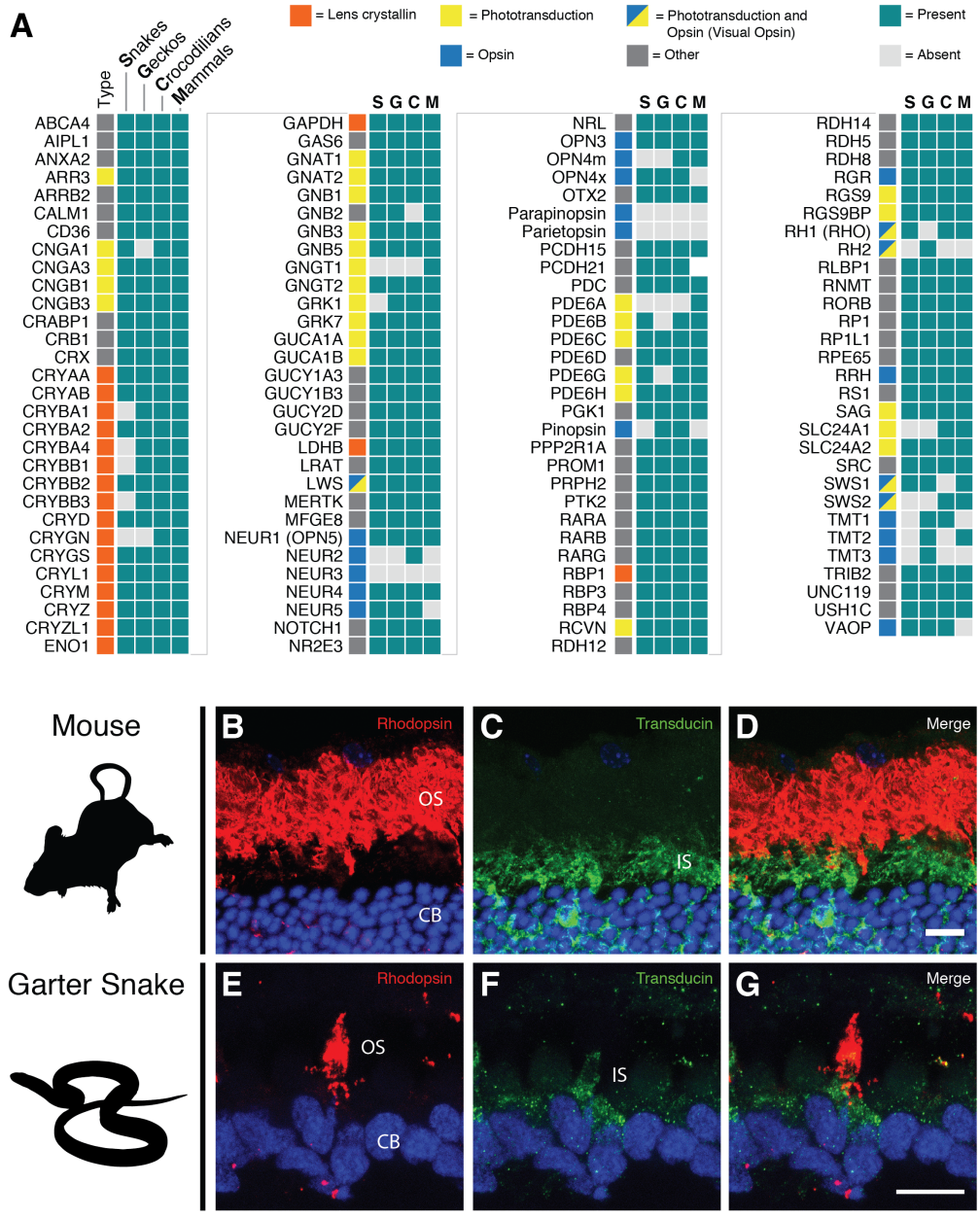

Figure 5

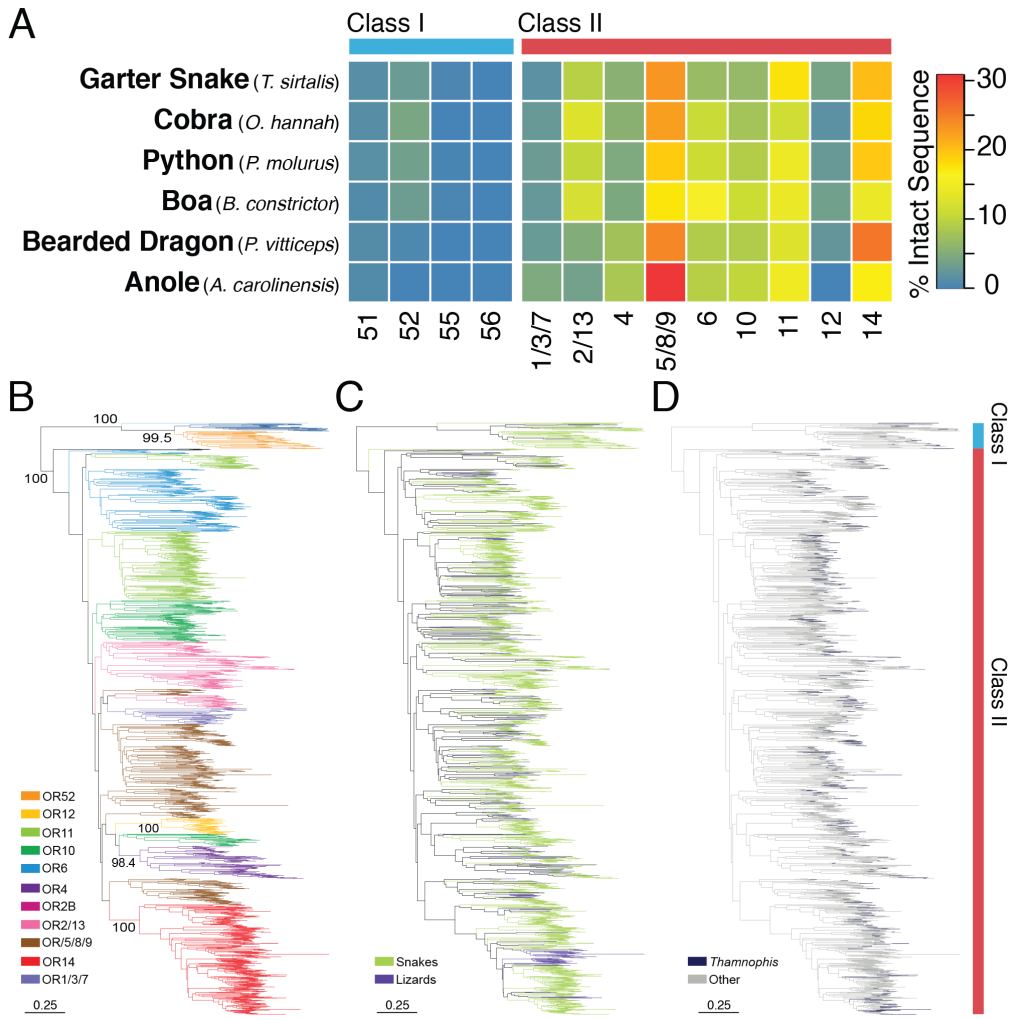

Figure 6

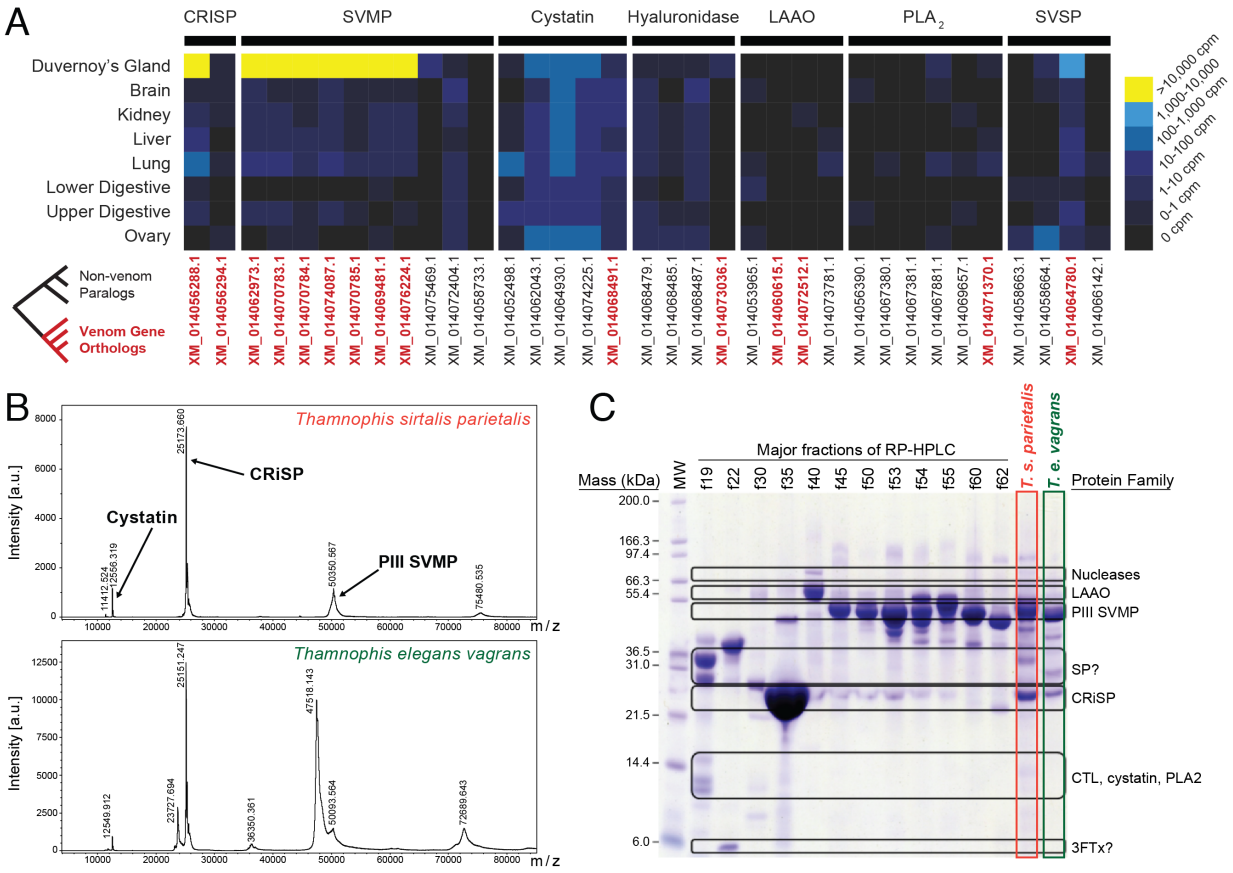

Figure 7