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

Patrick Minx
Washington University School of Medicine in St. Louis

Wesley C. Warren
Washington University School of Medicine in St. Louis

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation
https://digitalcommons.wustl.edu/open_access_pubs/7119

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact engeszer@wustl.edu.
Molecular Adaptations for Sensing and Securing Prey and Insight into Amniote Genome Diversity from the Garter Snake Genome

Blair W. Perry1,2, Daren C. Card1,†, Joel W. McGlotten2,†, Giulia I.M. Pasquesi3, Richard H. Adams1, Drew R. Schield1, Nicole R. Hales1, Andrew B. Corbin1, Jeffery P. Demuth1, Federico G. Hoffmann3,4, Michael W. Vandewege5, Ryan K. Schott6,7, Nihar Bhattacharyya8, Belinda S.W. Chang9, Nicholas R. Casewell9, Gareth Whiteley9, Jacobo Reyes-Velasco1,10, Stephen P. Mackessy11, Tony Gamble12,13, Kenneth B. Storey14, Kyle K. Biggar14, Courtney N. Passow15, Chih-Hong Kuo16, Suzanne E. McCaugh16, Anne M. Bronikowski17, A.P. Jason de Koning18, Scott V. Edwards19, Michael E. Pfrender20, Patrick Minx21, Edmund D. Brodie III22, Edmund D. Brodie Jr 23, Wesley C. Warren21, and Todd A. Castoe1,∗

1Department of Biology, University of Texas at Arlington, Arlington, Texas
2Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia
3Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State
4Institute for Genomics, Bioinformatics and Biotechnology, Mississippi State University, Starkville
5Department of Biology, Institute for Genomics and Evolutionary Medicine, Temple University
6Department of Ecology and Evolutionary Biology, Department of Cell and Systems Biology, Centre for the Analysis of Genome Evolution & Function, University of Toronto, Toronto, Ontario, Canada
7Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, District of Columbia
8Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada
9Alistair Reid Venom Research Unit, Parasitology Department, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, United Kingdom
10Department of Biology, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates
11School of Biological Sciences, University of Northern Colorado
12Department of Biological Sciences, Marquette University, Milwaukee, WI 53201, USA
13Bell Museum of Natural History, University of Minnesota, Saint Paul, MN, USA
14Institute of Biochemistry, Carleton University, Ottawa, Ontario, Canada
15College of Biological Sciences, University of Minnesota, St. Paul
16Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan
17Department of Ecology, Evolution, and Organismal Biology, Iowa State University
18Department of Biochemistry and Molecular Biology, Department of Medical Genetics, Alberta Children’s Hospital Research Institute, Cumming School of Medicine, University of Calgary, Alberta, Canada
19Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University
20Department of Biological Sciences and Environmental Change Initiative, University of Notre Dame
21The McDonnell Genome Institute, Washington University School of Medicine, St. Louis
22Department of Biology, University of Virginia
23Department of Biology, Utah State University

†These authors contributed equally to this work.
∗Corresponding author: E-mail: todd.castoe@uta.edu.

Accepted: July 26, 2018

Data deposition: This project has been deposited at NCBI GenBank under the accession GCA_001077635.2.
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 Myr of evolutionary history (Uetz and Etzold 1996; Kumar et al. 2017). 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 2013), evidence for pervasive and rapid positive selection (Castoe 2013), multiple origins of sex chromosomes (Gamble et al. 2017), and complex multigene family evolution in genes associated with sensory systems (Vandewege et al. 2016; Schott et al. 2018), 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 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 guttatus) (Ulate-Agote et al. 2014), 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 nonmammalian 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, 2010). Garter snakes have also proved to be valuable models for a wide range of physiological questions, including those related to reproduction (Mendonna 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, 2018), and to study how decreased snake visual acuity may lead to enhanced chemosensation and pheromonal communication and behavior (Kubie et al. 1978; Crews et al. 1984; Mason and Crews 1986).

Like many other colubrids, garter snakes are capable of producing a relatively mild venom that is used to subdue prey (Vest 1981; Hayes and Hayes 1985). 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 nonvenomous or highly venomous species (Vonk et al. 2013; Reyes-Velasco et al. 2015; 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 nevts of the

 genus Tantch. Given this unique adaptation, garter snakes

 have become an important model for studying the evolution

 of toxin resistance and coevolutionary dynamics of predat-

 or–prey arms races (Brodie and Brodie 1999; Brodie et al.


 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 un-

 derlie unique and interesting biological features of garter

 snakes related to toxin resistance and venom, as well as of-

 factor 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 4-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, Supplementary

 Material online). Garter snake genome sequencing libraries

 were constructed from a single adult female T. sirtalis collected

 in Benton County, Oregon. A genome size of 1.9 Gbp, pre-

 viously estimated using flow cytometry (Vinogradov 1998),

 was used for the construction of libraries. We created and se-

 quenced multiple whole-genome shotgun libraries on the

 Illumina platform, generating ~72 x genome coverage, which

 was subsequently used for assembly with ALLPATHS-LG

 (Gnerre et al. 2011), scaffolding with L RNA-scaffold (Xue

 et al. 2013), and SSPACE (Boetzer et al. 2011), and gapfilling

 using a custom script. This 6.0 version has been screened and

 cleaned of contaminating contigs and all contigs 200 bp 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 com-

 pleteness of the T. sirtalis genome and five other squamate

 genomes (three lizards: Pogona vitticeps, Anolis carolinensis,

 Ophisaurus gracilis; and two snakes: Python bivittatus, O.

 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 mul-

 tiple were present and only one protein entry in the case of

 duplicated entries, and used these filtered, nonredundant

 sets of proteins for BUSCO analysis. Genomic repeat elements

 were annotated using RepeatMasker v4.0.6 (Smit et al. 2015)

 using de novo repeat libraries derived from RepeatModeler

 v1.0.9 (Smit et al. 2014) analysis of the genome assembly,

 combined with repeat libraries from other snake genomes

 (Castoe 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, 

 Supplementary Material online 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 (supple-

 mentary table S2, Supplementary Material online), 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

 (Lampropeltis tesselata) and two species of garter snakes

 Thamnophis couchii and Thamnophis 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.

 2014). 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, 

 Supplementary Material online for additional details.

 Estimating Neutral Substitution Rates for Squamate

 Reptiles

 We inferred neutral substitution rates across squamate rep-

 tiles by estimating rates of substitution at 4-fold degenerate

 third codon positions of orthologous protein-coding genes.

 We used a Perl script provided in Castoe (2013) to identify
and sample 4-fold degenerate sites for seven squamates (glass lizard, Ophisaurus gracilis; green anole, A. carolinensis; bearded dragon, P. vitticeps; king cobra, O. hannah; garter snake, T. sirtalis; cottonmouth, Agkistrodon piscivorus; and Burmese python, P. 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, Supplementary Material online 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, Supplementary Material online for node calibrations and prior distributions) and constraining the topology to (((GlassLizard, (Anole, Pogona)),((Cobra, Thamnos), 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 (Matsubara et al. 2006; Vicoso et al. 2013; Laopichienpong et al. 2017). 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, Supplementary Material online 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, Supplementary Material online. Inferred mature microRNAs were putatively matched to target coding sequences in the T. sirtalis genome using miRanda v3.3a (Enright et al. 2003), using the parameters described in the supplementary Methods 1.6, Supplementary Material online.

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 (O. hannah) and Burmese python (P. molurus bivittatus), using a library of visual and opsin genes compiled from previous studies of vertebrate vision (Schott et al. 2017, 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 with 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 Schott et al. 2016; Bhattacharyya et al. 2017). See supplementary Methods 1.7, Supplementary Material online for additional experimental details.

Annotation and Phylogenetic Analysis of Olfactory Receptors (ORs)

To infer patterns of expansion and loss in 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 (P. vitticeps) (Georges et al. 2015) and two snakes, the B. constrictor (Bradnam et al. 2013) and king cobra (O. 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, Supplementary Material online.

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, Supplementary Material online for additional details.

Analysis of Tetrodotoxin (TTX) Resistance in Voltage-gated Sodium Channels

Garter snakes are well known for their resistance to TTX 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 (B. constrictor, O. hannah, and P. molurus bivittatus), two lizards (A. carolinensis and O. 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: SCNSA (Na,1.5), SCN1A (Na,1.8), and SCN11A (Na,1.9). 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 ~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 fig. S1, Supplementary Material online). Kmer-based estimation of genome size resulted in estimates of 1.19 Gbp and 1.27 Gbp (mean = 1.23 Gbp) using 1mers 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 with other vertebrate genomes (Castoe et al. 2011; Castoe 2013; Adams 2016). On the basis of limited studies, repetitive content appears to vary substantially between snake lineages in particular, despite low variation in genome size across snakes (Gregory 2018). 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 (fig. 1). Like other squamate genomes, the garter snake genome is dominated by LINE (nonLTR) 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 (fig. 1A). Overall, the genomes of colubrid snakes (Deinagkistrodon, Thamnophis, Ophiophagus in fig. 1) are similar in repeat element content and relative abundance, characterized by higher total repeat and TE content compared with the noncolubrid python genome, and all have notably higher CR1-L3 LINE and hAT and TcMar DNA transposon content (fig. 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 (fig. 1B). Together, analyses of repeat element content highlight the highly active, diverse, and dynamic nature of snake repeat element landscapes, particularly in colubrid 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 with 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 (23,247.50bp; supplementary table S1, Supplementary Material online). The chicken, with the smallest
Fig. 1.—Comparison of genomic repeat element landscapes across squamate reptile 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 to 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, whereas elements with higher K-values represented more degenerated copies that were likely inserted in the past.

geneome of the species analyzed, shows a relatively greater number of small genes (< ~5,000 bp) and a relatively lower number of large genes (> ~20,000 bp) compared with the other three species (supplementary fig. S2A, Supplementary Material online). The human, with the largest genome of the species analyzed, possesses a greater number of genes > ~40,000 bp compared with the other three species (supplementary fig. S2A, Supplementary Material online). 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 exon between ~1,000 and 3,000 bp (supplementary fig. S2B, Supplementary
Material online). Furthermore, the median exon length is very similar in all four species (supplementary table S1, Supplementary Material online). 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 (>~6,000; supplementary fig. S2C, Supplementary Material online). 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, Supplementary Material online).

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 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 T. 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, Supplementary Material online). The last of these, PICK1, a synaptic protein, is notable for its interaction with both nonvoltage 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, Supplementary Material online; Jaglin and Chelly 2009; Wall et al. 2009; Wade et al. 2013). Finally, we identified eight genes with significant evidence of positive selection when placing only T. sirtalis in the foreground (supplementary table S3, Supplementary Material online).

In the IPA Comparison Analysis (Krämer et al. 2014), 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, T. 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) (Gamer et al. 2004). We also identified genes, such as ANOS, that are associated with skeletal and cardiac muscle, specifically with diseases such as muscle dystrophy and gynadiaphyseal 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 (Kerchov 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 4-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) (fig. 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 with lizards (Avg. = 0.00077), due largely to the python, which had the lowest estimated rate in our data set. Notably, the garter snake had the highest estimated substitution rate among all squamate lineages (fig. 2). Overall, our estimates of substitution rates (for both squamates and mammals) using 4-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 (Martin and Palumbi 1993; Bromham 2002; Bromham and Penny 2003). 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 ZZW sex chromosomes derived from the same ancestral autosomes (Matsubara et al. 2006; Ohno 1967; Vicoso et al. 2013; Rovatsos et al. 2015), 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 (Matsubara et al. 2006; O’Meally et al. 2010; Vicoso et al. 2013; Rovatsos et al. 2015; Laopichienpong et al. 2017). 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 (Matsubara et al. 2006; Vicoso et al. 2013; Laopichienpong et al. 2017) 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 ZZW, the corn snake (Pantherophis guttatus) and timber rattlesnake (Crotalus horridus). We used two snake species with XXXY sex chromosomes, B. constrictor and P. 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 (fig. 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, fig. S3B, Supplementary Material online). 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 (Jones and Singh 1985; Arnemann et al. 1986; Parasnis et al. 1999; O’Meally et al. 2010; Gamble et al. 2014). FISH of a Bkm-like repeat, (GATA)$_n$, showed significant enrichment on a single chromosome in the female karyotype, presumably the W (fig. 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 illustrates 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) noncoding RNAs that play central roles in the posttranscriptional 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 fig. S3, Supplementary Material online) and associated mature microRNAs from T. sirtalis.
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, Supplementary Material online. (C) Fluorescent in situ hybridization (FISH) of a BsmI-like repeat, (GATA)8, in a female T. sirtalis. Sex-specific hybridization on the W chromosome is indicated by an arrow.

(transcriptome file S1, Supplementary Material online). Nucleotide frequencies within mature microRNAs were found to be highly similar between the garter snake and king cobra genomes when compared with human, with a high frequency of uracil base at position 9, just outside of the 2–8 seed region responsible for target binding (supplementary fig. S4, Supplementary Material online). A total of 576 mRNA transcripts were predicted to be targets of the novel T. sirtalis mature microRNAs (supplementary file S1, Supplementary Material online). Notably, the novel T. sirtalis microRNA, ts-miR-novel4, was found to target several components of phosphoinositide metabolism; namely, phosphoinositide-3-kinase, regulatory subunit 5 (PIK3R5, also called p101), and inositol polyphosphate-4-phosphatase (INPP4A). Interestingly, PI3K
is responsible for the production of phosphatidylinositol 3, 4, 5-triphosphate (PIP3) from the phosphorylation of phosphatidylinositol 4, 5-bisphosphate (PIP2). PIP2 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; Kerchov et al. 2008). We speculate that snake-specific microRNAs may play a role in venom production cycle, perhaps by regulating the metabolism of PIP2. 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 (Vidal and Hedges 2004; Davies et al. 2009; Simões et al. 2015; Schott et al. 2018). 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 2013; Schott et al. 2017, 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 (Hall et al. 2012; Angielczyk and Schmitz 2014; Bickelmann et al. 2015). 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 (fig. 4A, supplementary fig. S5, Supplementary Material online). In comparison, 11 genes have been lost in the lineage leading to mammals, all of which are opsins (fig. 4A). Placental mammals have lost three additional opsins, increasing the total number of lost opsin genes beyond that in snakes (supplementary fig. S5, Supplementary Material online). Geckos, another group with a hypothesized highly nocturnal ancestry (Walls 1942; Gamble et al. 2015), have specifically lost 11 visual genes, 7 of which were, independently, also lost in snakes (fig. 4A). Crocodilians have lost the fewest visual genes (five), all of which were opsins (fig. 4A).

Our analyses of the visual gene complement in *T. sirtalis* and other snakes, when compared with 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 proteins: 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, whereas 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 (Walls 1942; Underwood 1970). 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 opsins 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 (Silman 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 (fig. 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.

**ORs**

Snakes possess a greater reliance on olfaction than their more visually oriented lizard ancestors (Mason 1992; Mason and Parker 2010; Dehara et al. 2012; Castoe 2013; Vandewege et al. 2016), which suggests that changes in OR 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, Supplementary Material online). 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 (fig. 5A), with several exceptions. For example, there was a slightly higher abundance of Class I, subfamily 52 ORs in snakes (fig. 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 (fig. 5B–D). Within most subfamilies, there were clear snake specific expansions (fig. 5B and C), although no obvious expansions appear to take place specifically within the lineage leading to *Thamnophis* (fig. 5D). Additionally, *Thamnophis* had the smallest OR repertoire among snakes ( supplementary table S6, Supplementary Material online), 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 is 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 (Vonk et al. 2013; Reyes-Velasco et al. 2015; Yin et al. 2016; Aird et al. 2017). 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 A2 (PLA2), 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 figs. S6–S12, Supplementary Material online), 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
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.

Garter snake genome for cystatin, hyaluronidase, SVSP and PLA₂ 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 fig. S12, Supplementary Material online), 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 misannotation 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 fig. S13, Supplementary Material online).
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 (Vonk et al. 2013; Hargreaves et al. 2014; Junqueira-de-Azevedo et al. 2015; Reyes-Velasco et al. 2015). 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]) (fig. 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 fig. 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; fig. 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 (fig. 6A). These findings of co-expression across tissue types, with high expression in the venom gland, are consistent with current hypotheses of venom evolution (Vonk et al. 2013; Hargreaves et al. 2014; Junqueira-de-Azevedo et al. 2015; Reyes-Velasco et al. 2015).

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 (fig. 6B and C). Consistent with transcriptomic data, CRISPs and SVMPs are highly abundant in the Duvernoy’s venom gland (fig. 6A and B), and the SVMPs in particular appear to have multiple proteoforms that elute at different times following RP-HPLC (fig. 6C, supplementary fig. S14, Supplementary Material online). CRISPs and SVMPs (particularly PII SVMPs) are common and often abundant components of rear-fanged snake venoms (Mackessy 2002; Junqueira-de-Azevedo et al. 2016; Mackessy and Saviola 2016; Pia 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 (fig. 6B, supplementary fig. S14, Supplementary Material online), most of which contribute to total venom protein by <1% each (e.g., LAOO; supplementary fig. S15, Supplementary Material online) and are transcribed in the Duvernoy’s venom gland of *T. sirtalis* at low to moderate levels (fig. 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 TTX Resistance**

*Thamnophis sirtalis* is among the few vertebrates capable of withstanding high concentrations of TTX found in newts of the genus *Taricha* (Brodie and Brodie 1999; 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 et al. 2008). TTX binds to specific amino acids in the four outer pore regions (“P-loops”) of voltage-gated sodium channels (Na	extsubscript{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 (Geffeney et al. 2002, 2005; Feldman et al. 2009; McGlothlin et al. 2014).

Previous work has shown that *T. sirtalis* possesses at least three TTX-resistant Na	extsubscript{v} channels: The skeletal muscle channel Na	extsubscript{v}1.4 (encoded by the gene SCNA4), and two peripheral nerve channels Na	extsubscript{v}1.6 (SCNA8, which is also expressed in the brain) and Na	extsubscript{v}1.7 (SCNA9) (Geffeney et al. 2005; McGlothlin et al. 2014). In contrast, three channels found only in the central nervous system (Na	extsubscript{v}1.1, 1.2, and 1.3, encoded by SCNA1, 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	extsubscript{v}1.7 evolving at least 170 Ma in nonsnake squamates and resistant Na	extsubscript{v}1.6 evolving ~40 Ma in three different snake families (McGlothlin et al. 2016). Resistance in Na	extsubscript{v}1.4 evolved within *T. sirtalis* and is polymorphic, with more highly resistant alleles found in populations that overlap with more toxic newts (Geffeney et al. 2005; Feldman et al. 2010; Hague et al. 2017). At least four other snake species have also evolved TTX resistance in Na	extsubscript{v}1.4 due to interactions with toxic newts (Feldman et al. 2009, 2012), all of which occur in lineages with resistant Na	extsubscript{v}1.6 and 1.7 (McGlothlin et al. 2016).
The three other members of the voltage-gated sodium channel, Na,1.5 (found in cardiac muscle and encoded by the gene SCONA), 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. 2011; Widmark et al. 2011). 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 (Backx et al. 1992; Satin et al. 1992; Akopian et al. 1996; Cummins et al. 1999; Leffler et al. 2005). 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,1.5 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 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 SCN4A genes. Here, we focus on the three genes previously undescribed in T. sirtalis; additional detail on the location of annotated SCN4A genes can be found in Supplementary Information 2.1. We found evidence suggesting that Na,1.5, Na,1.8, and Na,1.9 are all either moderately or strongly resistant to TTX in T. sirtalis (supplementary table S7, Supplementary Material online). Na,1.5 has a cysteine in DI (Y371C) that is known to confer ~3,000-fold resistance to TTX in mammalian Na,1.5 and pufferfish.
Na,1.4a (Backx et al. 1992; Satin et al. 1992; Perez-Garcia et al. 1996; Perzotti et al. 1998; Venkatesh et al. 2005). For each substitution we discuss, the ancestral amino acid refers to mammalian Na,1.4, 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 2-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,1.5 should be blocked only at micromolar to millimolar concentrations of TTX (as opposed to nanomolar for TTX-sensitive channels).

Surprisingly, Na,1.8 DI does not have Y348S as in mammals, but does possess N351R, homologous to the arginine in Na,1.5 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,1.4 (Feldman et al. 2009). This suggests that T. sirtalis Na,1.8 can probably function up to TTX concentrations of about 1 μM. Na,1.9 is likely to be extremely TTX-resistant, due to Y362S (~7,000-fold; Leffler et al. 2005) and N365R in DI. A DI 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,1.9 is probably resistant to TTX in the millimolar range.

Ancestral reconstructions confirm that extreme resistance in Na,1.5 and Na,1.9 is ancient, dating to the common ancestor of mammals and reptiles (~320Ma; fig. 7; Hedges et al. 2015). All species we examined had a DI substitution providing micromolar resistance in Na,1.9. In Na,1.5, Y371C also dates to an ancient common ancestor, although some groups, including Gallus and Anolis, have lost this substitution. In contrast, ancestral Na,1.8 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. On the basis of our reconstruction, this substitution apparently evolved sometime within the last 50 my, after the split between colubrid and elapid snakes (fig. 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,1.1, Na,1.2, and Na,1.3 remaining completely TTX sensitive. Snakes were predisposed to evolve physiological TTX resistance because they inherited strong TTX resistance in three channels—Na,1.5, Na,1.7, and Na,1.9—from their common ancestor. Later, Na,1.6 evolved TTX resistance in three families, perhaps in response to eating amphibians, facilitating evolution of resistance in Na,1.4 via co-evolutionary arms races with amphibians in five species (McGlothlin et al. 2016). Our results show that a third channel, Na,1.8, although mildly resistant to TTX for 300 Myr, has shown an evolutionary increase in TTX resistance within the last 50 Myr, 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 colubrid 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 4-fold degenerate sites suggest that the rate of substitutions in snakes is similar.

---

**Fig. 7.**—Evolutionary reconstruction of TTX resistance in SCNA5 (Na,1.5), SCNA10A (Na,1.8), and SCNA11A (Na,1.9) across amniote vertebrates. Tips and branches are color-coded by predicted TTX resistance, with warmer colors representing stronger resistance (see supplementary table S7, Supplementary Material online for details). Resistance-conferring substitutions found in Thamnophis sirtalis are plotted at their origin, with warmer colors indicating stronger resistance (supplementary table S7, Supplementary Material online). A nonresistant SCNA5-like paralog from Xenopus is used as the outgroup.
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 OR 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 venemous 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.

**Supplementary Material**

Supplementary data are available at Genome Biology and Evolution online.

**Acknowledgments**

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 [SUS4HG00307907 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.].

**Literature Cited**


Fozard HA, Lipkind GM. 2010. The tetradotoxin binding site is within the outer vestibule of the sodium channel. Mar Drugs. 8(2):219–234.


Associate editor: Mary O’Connell