Blair W. Perry[1]

Abstract

Colubridae represents the most phenotypically diverse and speciose family of snakes, yet no well-assembled and annotated genome

compounds of their prey, specifically in toxic newts of the genus *Ta c a*. 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 (*T a*) 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 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

and sample 4-fold degenerate sites for seven squamates (glass lizard, *O.* ; green anole, *A.* ; bearded dragon, *P.* ; king cobra, *O.* ; garter snake, *T.* ; cottonmouth, *A.* ; and Burmese python, *P.*) and three mammal outgroups (chimpanzee, *Pa.* ; human, *H.* ; and mouse, *M.*); 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, 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.* 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.* gene fragments by PCR to confirm that these scaffolds were sex-linked gametologous alleles and not paralogs or misassemblies. Fluorescent hybridization (FISH) of a *B* -like repeat, (GATA) , was used to identify the W chromosome in the karyotype of a female *T.* . 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.* (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.* genome using miRanda v.3.3a (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.*) and Burmese python (*P.*), 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.* 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 e c4cle(166())]T

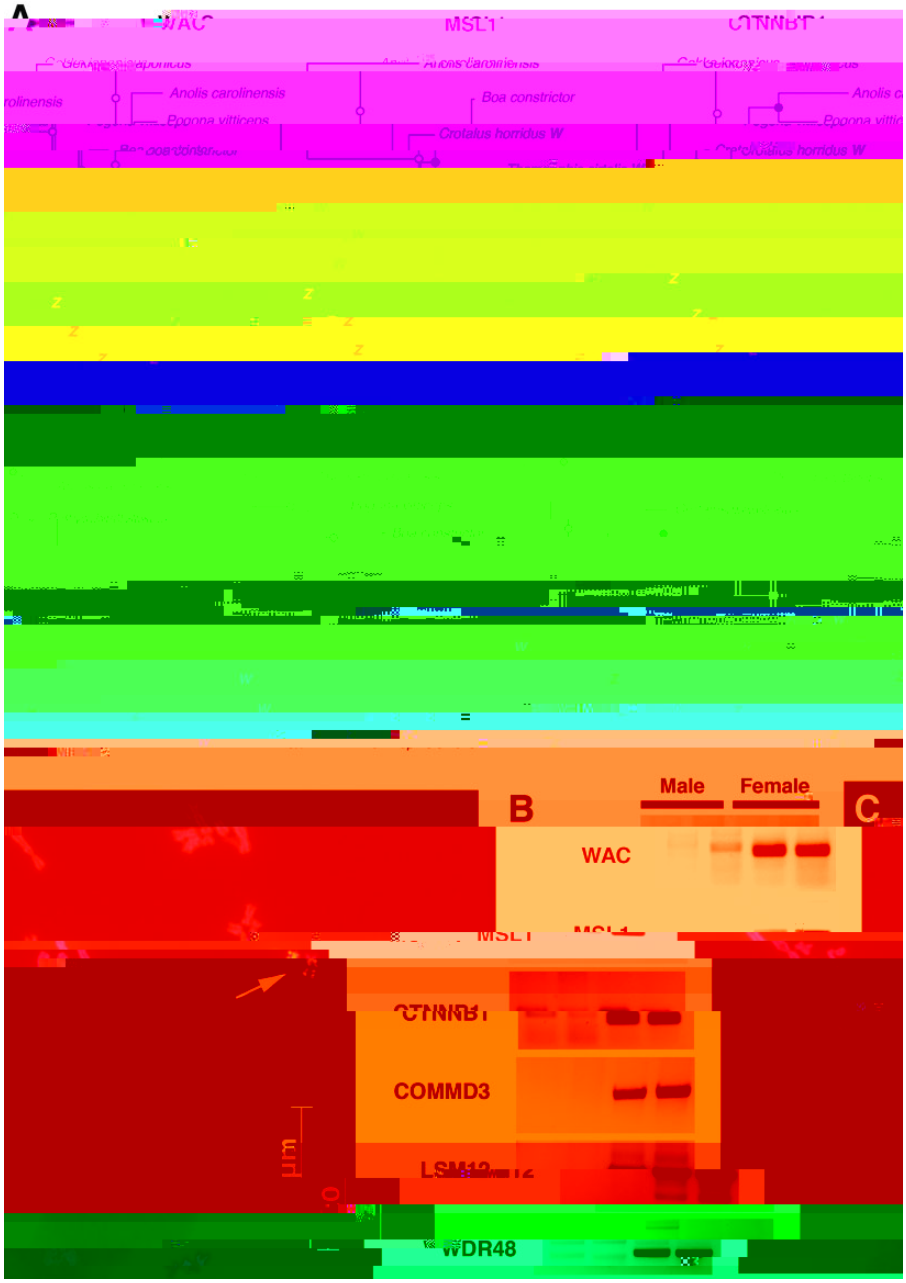
T a venom, and compare this to inferences based on RNAseq, we extracted and analyzed crude venom from *T.* and *T.* *a* using SDS-PAGE and MALDI-TOF mass spectrometry, under approved animal protocols at the University of Northern Colorado (#9204). See supplementary Methods 1.9, Supplementary Material 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 subuni5(p)ubun1(e)14.3(p)-6(e)1.7tttia296TJ.66uT19.6(6)-22ir0(e)4(t(T)4.[(ua)-28i4 0 TD[(us)18(3.5627J/F5 1 ge)l)5(nc)-3bunme(e)1.73-2

Material online). Furthermore, the median exon length is very similar in all four species (

chromosome differentiation across lineages (



. . **3.**—Identification of the W chromosome and sex-chromosome-linked genes in *T a* , ... (*A*) Maximum likelihood trees of fragments of six genes with Z and W alleles in *T.* .

(supplementary 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.* . mature microRNAs (supplementary file S1, Supplementary Material online). Notably, the novel *T.* . microRNA, - *R-* , *4*, 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.* [illegible] -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

that the apparent all-cone retina of a diurnal garter snake, *T a* , 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

garter snake genome for cystatin, hyaluronidase, SVSP and

To assess whether these putative toxin genes contribute to the venom secretions of *T. [illegible]*, 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. [illegible]* 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 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.* [illegible]. 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 ubiqui-

$Na_v1.4a$ (Backx et al. 1992; Satin et al. 1992; Perez-Garcia et al. 1996; Penzotti et al. 1998; 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.*

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 *T a* , ●. 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 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.

Supplementary Material

Supplementary data are available at *G* , *B* , , *a d* *E* , ; 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 [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

Mackessy SP. 2002. Biochemistry and pharmacology of colubrid snake venoms. J Toxicol: Toxin Rev. 21(1–2):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(5):1004–1021.
Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27(6):764–770.
Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci U S A. 90(9):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. New York, NY: Plenum Publishing Corporation.
Mason RT, Parker MR. 2010. Social behavior and pheromonal communication in reptiles. J Comp Physiol A. 196(10):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(48): 18190–18195.
McGaugh SE, et al. 2015. Rapid molecular evolution across amniotes of the IIS/TOR network. Proc Natl Acad Sci U S A. 112(22):7055–7060.
McGlothlin JW, et al. 2014. Parallel evolution of tetrodotoxin resistance in three voltage-gated sodium channel genes in the garter snake *T a* . Mol Biol Evol. 31(11):2836–2846.
McGlothlin JW, et al. 2016. Historical contingency in a multigene family facilitates adaptive evolution of toxin resistance. Curr Biol. 26(12):1616–1621.
Mendonca MT, Crews D. 1996. Effects of ovariectomy and estrogen replacement on attractivity and receptivity in the red sided garter snake (*T a* parietalis). J Comp Physiol A. 178(3):373–381.
O'Meally D, et al. 2010. Non-homologous sex chromosomes of birds and snakes share repetitive sequences. Chromosome Res. 18(7): 787–800.
Ohno S. 1967. Sex chromosomes and sex-linked genes. Berlin: Springer.
Parasnis AS, et al. 1999. Microsatellite (GATA) n reveals sex-specific differences in papaya. Theor Appl Genet. 99(6):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.
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(6): 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 U S A. 93(1):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), *D* . Biochim Biophys Acta (BBA)-Gener Subj. 1861(4):814–823.