
3 Applications of Genomics and Related Technologies for Studying Reptile Venoms

Drew R. Schield, Blair W. Perry, Giulia I.M. Pasquesi, Richard W. Orton, Zachary L. Nikolakis, Aundrea K. Westfall and Todd A. Castoe

CONTENTS

3.1	Introduction	29
3.2	The Genomic Context for Reptile Venoms.....	31
3.2.1	What a “Complete Genome” Is, and Why They Are Not Created Equal	31
3.2.2	The Structural and Evolutionary Context for Venom in Squamate Reptile Genomes	32
3.2.3	Venom Gene Cluster Structure, and Why Venom Genes Are Difficult to Assemble and Identify Accurately	33
3.2.4	Where Are Venom Genes Located in the Genome?.....	35
3.3	Regulation of Venom Genes – The Black Box Linking the Genome With Venom	35
3.3.1	Towards an Understanding of Venom Regulation.....	35
3.3.2	Genomes Add Value to Transcriptomic and Proteomic Data	37
3.3.3	Understanding the Factors that Direct the Regulation of Venom Production	37
3.4	Population-Level Studies of Venom Variation and Evolution	37
3.4.1	Population-Level Sampling of Venomous Reptile Genomes.....	37
3.4.2	The Relevance of Hybrid Zones for Studying Venom.....	38
3.5	Conclusions.....	39
	References.....	39

Over the past decade, the increasing availability and decreasing cost of genome sequencing technologies have led to significant advancements in our understanding of the evolution, variation and function of reptile venoms. This includes multiple annotated genomes for venomous reptile species as well as countless public data sets of venom gene sequences, gene expression data, and data availability from a growing diversity of species. These genomic resources have provided new insight into reptile venoms, and the continuing development of additional genomic resources and related data sets holds promise for continued advancement of our understanding of reptile venom systems. Here, we present an overview of currently available genomic resources for the study of reptile venoms, discuss genomic approaches used to study reptile venoms, and highlight recent advances in our understanding of how venom evolves and functions in reptiles based on genomic data. We also discuss ongoing progress and future work that will be important for closing major gaps in our current understanding of the genomic basis of venom variation, evolution and regulation.

Key words: evolutionary genomics, genome quality, hybrid zones, transcriptomics, venom expression, venom genes

3.1 INTRODUCTION

In 2019, the World Health Organization (WHO) and Wellcome Trust announced a large-scale strategy for researching and advancing the treatment of snakebite, which the WHO defines as a major neglected tropical disease (Hand, 2019). A primary aim of this initiative is to develop more effective, sustainable and affordable treatments for snakebite – a goal that has historically proved challenging, due in part to variation in venom composition and biological activity within and among species. Leveraging recent advancements in genomic technologies and the decreasing cost of genome sequencing holds promise for a fast, economical and powerful means by which to understand better the variation in venom within and among venomous reptile species, as well as the factors that drive this variation. However, despite recent advancements in our understanding of the genomic basis of reptile venoms, there remain several key challenges in linking genomes to venoms. Perhaps the largest gap in our knowledge currently is precisely how genomic sequences direct the regulation of venom genes to produce varying amounts of distinct venom proteins that together constitute complex venom cocktails within a species, population or individual. Furthering our understanding

of the genomic basis and regulation of venom in venomous reptiles, and how this translates to medically relevant phenotypes, is thus a major priority for current and future research efforts. In this chapter, we highlight available venomous reptile genomes and key insights into venoms and venom evolution that these genomes have delivered, provide an overview of modern genomic technologies and their value for the study of reptile venoms and toxins, and describe a selection of major discoveries that reptile genomes have facilitated over the past decade that frame the future roles of genomics in understanding reptile venoms.

At the time of publication of the first edition of the *Handbook of Venoms and Toxins of Reptiles* (Mackessy, 2010a), no reptile genomes, venomous or otherwise, were available. While the gene and protein sequences of major venom components were known through traditional sequencing technologies and proteomics, the lack of any substantial genomic context for understanding reptile venoms meant that many questions regarding the evolution, diversity and function of venoms remained vastly unanswered and largely inaccessible. Over the past decade, the time and cost associated with sequencing, assembling and annotating genomes have decreased substantially with the development of more efficient and accessible genomic technologies. These advancements have made it feasible to sequence and assemble the genomes of nontraditional model organisms for a variety of research purposes (Ellegren, 2014), including the study of venoms and toxins. At the publication of this edition, around 28 squamate reptile genomes are publicly available, including 17 venomous reptiles (National Center for Biotechnology Information [NCBI] genome database; Figure 3.1). Studies leveraging these genomes have already greatly advanced our understanding of reptile venoms, providing insight into long-standing

questions about the evolution, function and diversity of reptile venoms and toxins, and raising even more questions for future study (e.g., Vonk et al., 2013; Yin et al., 2016; Shibata et al., 2018; Perry et al., 2019; Schield et al., 2019b).

The first non-avian reptile genome, that of the Green Anole (*Anolis carolinensis*), was published in 2011 by Alföldi et al. (2011). While this is not a venomous species, this genome provided the first broad insight into the structure, function and evolution of a reptile genome and effectively brought reptiles into the age of genomics. Two years later, the genomes of the King Cobra (*Ophiophagus hannah*; Vonk et al., 2013) and the Burmese Python (*Python bivittatus*; Castoe et al., 2013) were published, providing the first genomic perspectives on venom gene evolution in a venomous reptile, as well as the first non-venomous snake genome that could be used to compare and contrast genomic features to contextualize venom evolution (e.g., Reyes-Velasco et al., 2015). Over the next several years, additional genome assemblies of venomous reptiles were published, including that of the Five Pace Viper (*Deinagkistrodon acutus*; Yin et al., 2016), the Habu (*Protobothrops flavoviridis*; Shibata et al., 2018), the European Adder (*Vipera berus*; NCBI BioProject: PRJNA170536), the Komodo Dragon (*Varanus komodoensis*; Lind et al., 2019), and the rear-fanged Common Gartersnake (*Thamnophis sirtalis*; Perry et al., 2019). Genomes for a sampling of non-venomous squamates have also been sequenced, representing such divergent species as the Boa Constrictor (*Boa constrictor*; Bradnam et al., 2013), the Bearded Dragon (*Pogona vitticeps*; Georges et al., 2015), and Schlegel's Japanese Gecko (*Gekko japonicus*; Liu et al., 2015). While the availability of the first genomes of venomous squamates has proved valuable for initial genomic studies of reptile venoms and toxins, the utility of these resources has been limited, because these early genome

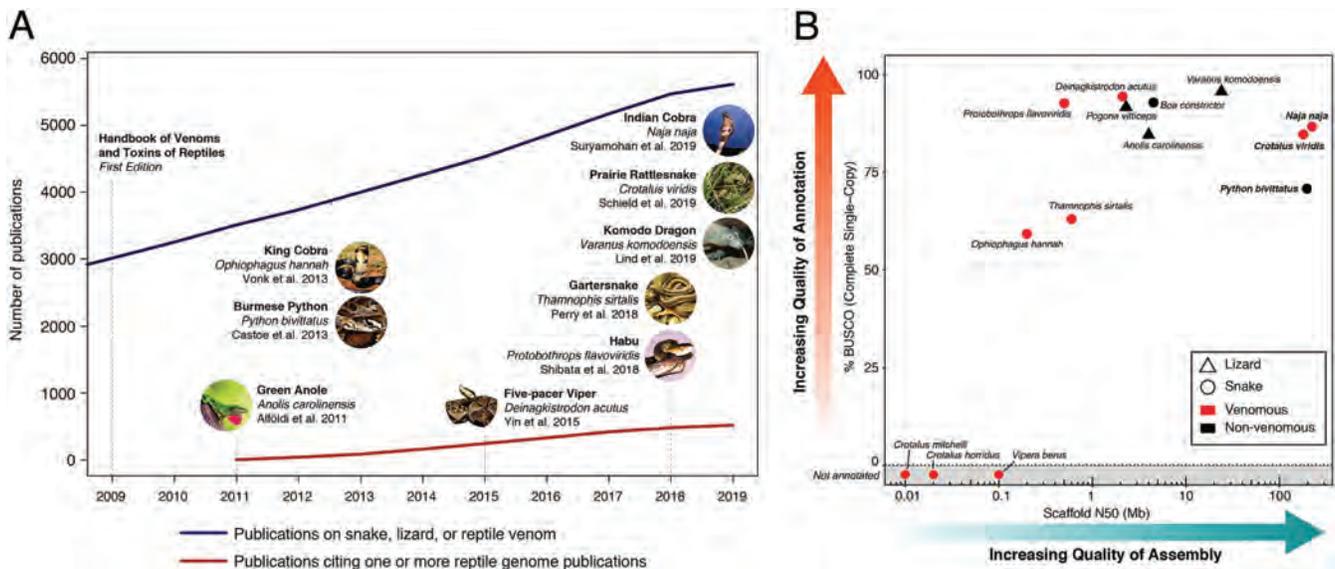


FIGURE 3.1 Overview of available squamate genome resources. (A) Number of publications found in Google Scholar searches (accessed September 15, 2019) including terms “snake venom,” “lizard venom,” or “reptile venom” over time (blue) along with the number of publications citing one or more reptile genome publications (red). Relevant reptile genomes are shown at their approximate date of publication. (B) Comparison of genome assembly and annotation quality of currently available reptile genomes. Chromosome-level assemblies are designated by bolded species names.

assemblies were only assembled into partial fragments rather than complete chromosomes and were in some cases relatively poorly annotated.

In 2019, the genome of the Prairie Rattlesnake (*Crotalus viridis*) became the first reptile genome to be assembled to the level of full chromosomes, providing the most complete resource for the genomic study of reptile venoms to date (Schield et al., 2019b). This genome was used to investigate the evolution of squamate genome structure at multiple scales (including chromosomes, GC content and repeat content), sex chromosome evolution and dosage compensation in snakes, and venom gene family evolution and regulation – several of these topics are discussed in detail later in this chapter. Soon after, a chromosome-level assembly was published for the Indian Cobra (*Naja naja*; Suryamohan et al., 2020), providing genome-wide insight into venom genes in an elapid and providing the first opportunities for powerful comparative studies of venom at a genome scale between divergent venomous snake lineages. As genomic technologies continue to advance in capabilities and decrease in cost, additional high-quality chromosome-level genome assemblies will no doubt become available for venomous and non-venomous reptiles, providing an unprecedented set of resources to catalyze progress towards understanding the genomic context, evolution and function of reptile venoms.

3.2 THE GENOMIC CONTEXT FOR REPTILE VENOMS

3.2.1 WHAT A “COMPLETE GENOME” IS, AND WHY THEY ARE NOT CREATED EQUAL

As for other eukaryotes, reptile nuclear genomes consist of varying numbers of chromosomes, each of which consists of a single long molecule of DNA ranging from several million to a few hundred million nucleotides. Despite many advances, there exists no genome sequencing technology capable of sequencing an entire eukaryotic chromosome as a single continuous sequence. Instead, all available genome sequencing approaches entail sequencing much smaller pieces of fragmented chromosomal DNA and then computationally inferring how these pieces “fit back together” to reconstruct the entire genome sequence. This computational inference is made massively more challenging by the abundance of repetitive DNA, low complexity DNA, and duplicated DNA typical of eukaryotic genomes in general (Charlesworth et al., 1994) and reptilian genomes in particular (Organ et al., 2008; Castoe et al., 2011; Tollis and Boissinot, 2011; Pasquesi et al., 2018). Accordingly, there is surprisingly little universal meaning or value to simply having a “complete genome” – instead, each available genome is associated with its own degree of completeness and accuracy based on the quality of the computational reassembly of fragmented sequences, and the relative degree to which this reassembly was able to reconstruct an inference of the genome. In other words, no genomes are created equal, no “complete genomes” are in fact complete, and many “complete genomes” do not contain all the sequences

and genomic information that would be potentially relevant for studying venom and venom systems. In the following, we highlight some key characteristics that are useful for better understanding the relative variation in quality, completeness and utility of various reptile genomes available, and relate these features to what questions can, and cannot, be addressed effectively with genomes of varying quality.

Numerous features of genome quality, contiguity and annotation can be used to estimate the utility of a genome for a given set of questions. For example, relatively incomplete and highly fragmented genomes are typically more than sufficient to provide accurate estimates of genomic repeat content or nucleotide composition. Moderately complete and well-assembled genomes are typically more useful for analysis of gene content and coding sequences but often lack the information to study hard-to-annotate and/or multi-copy genes (i.e., most venom gene families) and regulatory elements (e.g., King Cobra genome; Vonk et al., 2013). However, analyses of multi-gene families, regulatory elements, and large-scale analyses of chromosome synteny (i.e., homology) typically require high-quality genome assemblies. Moreover, analyses of protein-coding genes require accurate genome annotation, and the quality and completeness of gene annotations depend on the quality of input gene models to predict genes, as well as the quality and contiguity of the genome assembly, leading to variation in gene annotation quality among genomes, in some cases independent of the quality of their assembly.

Available snake and lizard genomes vary considerably in genome completeness and genome assembly contiguity and thus, vary in utility for addressing various questions, especially questions relevant to venom gene content, structure, regulation, and evolution. As shown in Figure 3.1b, there are examples of squamate genomes that have high-quality and highly complete annotations based on BUSCO (Simão et al., 2015) single-copy gene completeness but that have highly fragmentary genome assemblies (e.g., *Protobothrops flavoviridis*; Shibata et al., 2018 and *Deinagkistrodon acutus*; Yin et al., 2016). These metrics suggest that these genomes contain a highly complete set of protein-coding genes, although the relative order, orientation and chromosomal location of clusters of genes are poorly known, and that regulatory sequences in intergenic regions are likely incomplete. Others have high assembly quality with intermediate annotation quality (e.g., *Crotalus viridis* (Schield et al., 2019b) and *Python bivittatus* (Castoe et al., 2013; Dudchenko et al., 2017, 2018) – highly contiguous genome assemblies typically do not have low-quality annotations.

Early snake genomes (e.g., *Ophiophagus hannah*; Vonk et al., 2013) were highly fragmentary, limiting our ability to study venom in a genomic context. Indeed, venom-encoding regions of most available snake genomes are very poorly assembled, due in part to the difficulty in reconstructing venom gene regions using short-read and mate-pair sequencing libraries. In recent years, sequencing and assembly technologies have made major strides to enable chromosome-level genome assemblies for venomous snakes, including the Prairie Rattlesnake (*Crotalus viridis*; Schield et al., 2019b) and the

Indian Cobra (*Naja naja*; Suryamohan et al., 2020). While the ability to develop such complete and powerful resources for venomous snakes is exciting in itself, the availability of these data has also begun to illustrate new and more complex questions that may now be addressed with access to high-quality chromosomal genomes, including questions related to the evolution and regulation of venom.

3.2.2 THE STRUCTURAL AND EVOLUTIONARY CONTEXT FOR VENOM IN SQUAMATE REPTILE GENOMES

In addition to their value for studying the evolution and regulation of venom, squamate reptile genomes also provide a rich comparative resource for understanding key features of vertebrate genome biology. A fundamental unit in genomics is genome size, or the amount of DNA in a cell, as genome size in part determines the amount of sequencing required to assemble a genome. Because genome size is impacted by various aspects of genome structure, it is also a useful coarse metric for comparing differences between lineages and for inferring the sources that drive these differences (e.g., segmental duplications and transposable elements). Most estimates of squamate genome sizes come from non-sequence-based estimation methods (e.g., Feulgen density, static flow cytometry, and flow cytometry, the last being the currently favored method; see Shaney et al., 2014 for a detailed review). The average haploid flow cytometry-based genome size estimate for squamates ($n = 90$) is 1.9 Gbp, and this is also the average estimate of genome sizes across snakes – these estimates place squamate genome sizes between the larger and more variable genome sizes of mammals and amphibians, and the consistently smaller genomes of birds (Janes et al., 2010).

Even prior to sequencing reptile genomes, karyotypes and chromosomal structure were fairly well characterized for a number of species, establishing that variation in chromosome number and arrangement is much greater in reptiles than in mammals (Organ et al., 2008). Squamates possess anywhere between 27 and 51 ($2n$) chromosomes, with lizards exhibiting the greatest variation in chromosome number (Olmo, 2005). Snakes have comparatively conserved karyotypes, with a typical karyotype of $2n = 36$ (8 macrochromosome pairs and 10 microchromosome pairs; Olmo, 2005; Matsubara et al., 2006; Srikulnath et al., 2009; Janes et al., 2010). Much of the variation in squamate genomes is manifested in the number of microchromosomes within the genome of a given species. Microchromosomes are comparatively smaller than macrochromosomes (e.g., less than 30 Mbp) but can be numerous, thereby increasing the number of independently sorting linkage groups within the genomes of reptiles. Microchromosomes also possess distinctive features across birds and squamates, including higher gene density, GC content, and recombination rates than macrochromosomes (McQueen et al., 1996, 1998; Hillier et al., 2004; Backström et al., 2010; Warren et al., 2010; Alfoldi et al., 2011; Schield et al., 2019b).

While characterizing variation in squamate chromosome number and arrangement has been feasible for some time using cytogenetic methods, the availability of chromosome-level

squamate genomes has allowed us only recently to explore chromosomal synteny (i.e., homology) at unprecedented resolution. Using a k-mer-based “chromosome painting” technique (McKenna et al., 2016), we explored nuclear chromosomal synteny between the Chicken (*Gallus gallus*), Green Anole (*Anolis carolinensis*), Burmese Python (*Python bivittatus*) and Prairie Rattlesnake (*Crotalus viridis*). In agreement with previous published studies (Alfoldi et al., 2011; Schield et al., 2019b), these analyses indicate that there are large regions of macrochromosomes that have been stable over hundreds of millions of years (Figure 3.2), especially among the squamate species. For example, differences in the number of macrochromosomes between lizard and snake species appear to be due to only a small number of fusion or fission events (e.g., *Anolis* Chromosome 3 is syntenic with contiguous regions of *Crotalus* chromosomes 4 and 5). This also emphasizes the remarkable degree of conserved synteny among snake macrochromosomes. Indeed, rearrangements between snake chromosomes are concentrated mainly within microchromosomes (Figure 3.2).

Squamate genomic composition is also dynamic at a finer scale, demonstrating major shifts in GC-isochore structure (Alfoldi et al., 2011; Fujita et al., 2011; Castoe et al., 2013; Perry et al., 2019; Schield et al. 2019b) and repeat element content across lineages (Castoe et al. 2013; Adams et al. 2016; Pasquesi et al. 2018). Vertebrate genomes typically contain relatively high fractions of repetitive elements, and squamate genomes stand out among vertebrates because they show extremely high variation in genomic repeat element composition among lineages and species. Additionally, in contrast to bird and mammal repeat element landscapes, which are dominated by a relatively narrow diversity of transposable element (TE) families and types that appear to be recently active in the genome (e.g., LINE1 and SINE in mammals and CR1 and ERV in birds), squamate genomes possess a more even and diverse representation of major TE families (Janes et al., 2010; Alfoldi et al., 2011; Castoe et al., 2013; Vonk et al., 2013; Perry et al., 2019; Schield et al., 2019a), most of which appear to be recently active (Ruggiero et al., 2017; Pasquesi et al., 2018; Figure 3.3). Comparative analyses have also demonstrated extensive variability in total genomic TE content even between closely related lizard and snake species, a level of variation that is unprecedented based on studies of mammals and birds (Pasquesi et al., 2018), suggesting that TE activity has played a particularly important role in shaping variation in genome structure and content among squamate lineages. Several studies have also shown that TEs further shape the structure of squamate genomes by sometimes harboring small microsatellite sequences on their tails, which some TE families may distribute and amplify across the genome as these TE families expand and replicate. Indeed, although such “microsatellite seeding” by specific TE types has in general been poorly characterized in vertebrates, snake genomes provide the most extreme example of large-scale recurrent microsatellite seeding by TEs of any vertebrates studied (Castoe et al., 2011; Pasquesi et al., 2018), begging the question of what the

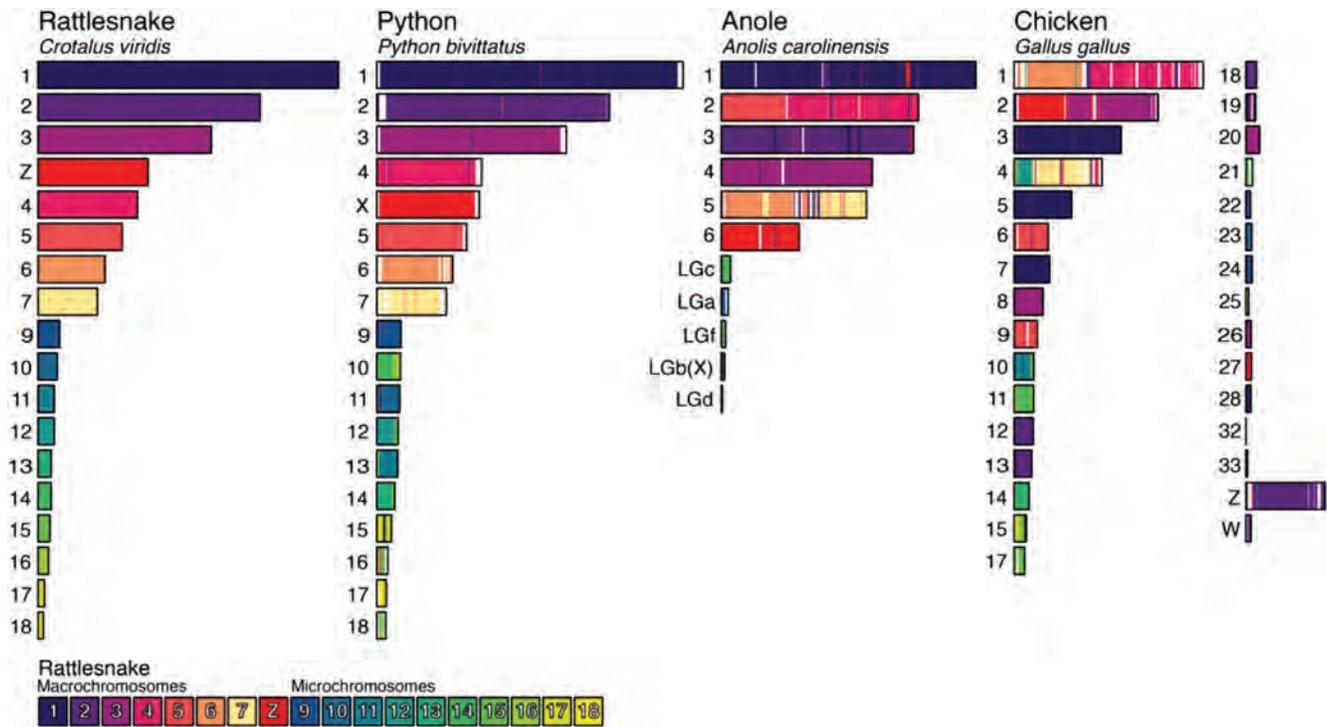


FIGURE 3.2 Chromosomal synteny among reptile genomes. Results of “chromosome painting” analysis to identify homologous stretches of chromosomes among reptile genomes. For each species, horizontal bars represent individual chromosomes (chromosome numbers are labeled at the left of each panel). The results shown here depict pairwise analyses between the Prairie Rattlesnake (*Crotalus viridis*) and the other three species. The legend at the bottom also shows the colors corresponding to rattlesnake chromosomes. Matching colors between the rattlesnake and other species represent genomic regions with common ancestry. (These results are redrawn from Schield et al. (2019a), with the addition of new analysis of the Burmese Python.)

functional and evolutionary consequences of such extreme microsatellite expansion may be in snake genomes.

3.2.3 VENOM GENE CLUSTER STRUCTURE, AND WHY VENOM GENES ARE DIFFICULT TO ASSEMBLE AND IDENTIFY ACCURATELY

One of the most confounding aspects of venom prior to the availability of genome resources was the arrangement of venom genes in the genome. Moreover, because venom gene families can include potentially numerous paralogs (i.e., multiple copies of a gene family), it is often difficult to infer the actual number of genes in the genome, even with data from transcriptomes or proteomes, due to many potential alternative splice variants of genes and potential post-translational modifications (Wong and Belov, 2012; Casewell et al., 2014). With advances in genomic resources for snakes, the prevailing body of evidence for the mechanism of evolution of venom gene families is that they tend to evolve by tandem duplication, whereby an initial duplication event of a “house-keeping” (non-venom) gene results in a novel gene copy under reduced selective constraint. This new copy may then undergo neofunctionalization or subfunctionalization and experience recruitment for venom gland gene expression, or it may fail to become functional and instead, become a pseudogene (Casewell et al., 2011, 2013; Wong and Belov, 2012; Vonk et al., 2013; Reyes-Velasco et al., 2015). Continued duplication

events of sub- or neofunctionalized gene paralogs may then result in the expansion of the venom gene family.

Venom gene accumulation through tandem duplication in discrete genomic regions (tandem arrays) has been shown previously for particular venom gene families (e.g., phospholipase A₂s: Lynch, 2007; Ikeda et al., 2010; Dowell et al., 2016 and snake venom metalloproteinases: Dowell et al., 2018). Based on the addition of recent chromosome-level snake genome assemblies, we now know that tandem venom gene arrays (and tandem duplication mechanisms) are essentially the rule for the evolution and expansion of venom gene families (Schield et al., 2019b; Suryamohan et al., 2020). These include several of the best-characterized venom gene families, which also make up the bulk of most snake venoms and include the best-characterized bioactive components (i.e., snake venom metalloproteinases, snake venom serine proteases, phospholipase A₂s, three-finger toxins, cysteine-rich secretory proteins, L-amino acid oxidases, C-type lectins and others). Indeed, there are relatively few venom gene families that are not organized into discrete tandem arrays (e.g., natriuretic peptides and hyaluronidases), and these components also do not generally contribute substantially to venom composition within venomous snakes, nor is their biological relevance in venom well understood (Mackessy, 2010a). The organization of venom gene families into tandem arrays has intriguing implications for understanding not only the evolutionary

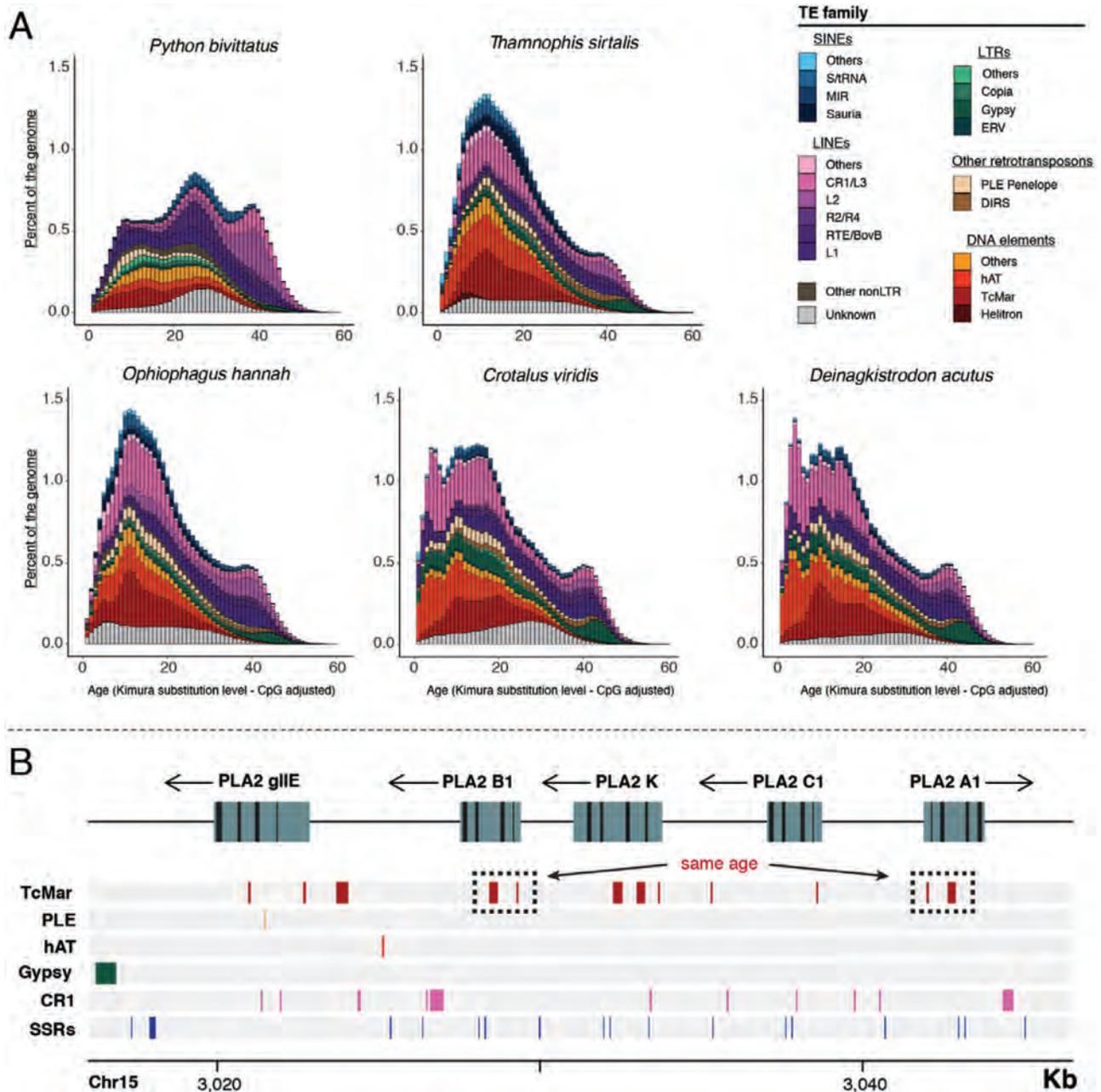


FIGURE 3.3 Repeat landscapes of squamate genomes and identification of transposable elements (TEs) in venom regions. (A) Results for TE copy divergence analyses showing profiles of genomic TE accumulation through time. Histograms report, for each species, the percent of the genome (y-axis) for each species corresponding to TEs belonging to different major families clustered according to the CpG-corrected Kimura 2-parameter distance (K -value from 0 to 60; x-axis) from each TE consensus sequence. Low K -values represent more recent transposition events, as inserted copies have higher similarity to their consensus, whereas higher K -values correspond to TE copies that through time have accumulated more mutations and are therefore symptomatic of past amplification instances. Analyses reveal that TE families (DNA elements, LINES, SINEs and LTRs) have been recently accumulating in venomous snakes, a pattern that sharply contrasts with mammal and bird genomes as well as with non-venomous snakes (e.g., *Python bivittatus*). (B) The phospholipase A_2 (PLA $_2$) region of the Prairie Rattlesnake genome, depicting PLA $_2$ gene structure, along with genomic tracks showing positions of annotated TEs. The dotted boxes emphasize the positions of the two Tc-Mariner elements that are very recently duplicated and are harbored within two genes of the PLA $_2$ gene family. SINEs = short interspersed nuclear elements, LINES = long interspersed nuclear elements, LTRs = long terminal repeats, SSRs = short sequence repeats.

origins of venom genes but also how the regulation of these genes has evolved.

Venom gene clusters bear the hallmarks of an evolutionary history that includes duplication events, rearrangements, and associations with other genetic elements (i.e., TEs and microsatellites). As described earlier, the genomes of venomous snakes are made up largely of TEs, and venom gene clusters are no exception. In fact, the snake venom metalloproteinase, serine protease and phospholipase A₂ regions of the Prairie Rattlesnake genome are enriched for the presence of recently active TEs. Specifically, the snake venom metalloproteinase gene cluster is made of roughly 56% TEs, while the immediate genomic background is made up of only 37% TEs. The phospholipase A₂ cluster shows an even more extreme increase in TE density compared with the immediate genomic background (i.e., 21.04% versus 8.91%). CR1 LINEs are among the most common TE families in all major venom gene regions of the Prairie Rattlesnake and have also contributed to TE-driven seeding of microsatellites within these regions. These elements have also been found in association with venom genes in several other snake species (Ikeda et al., 2010; Dowell et al., 2016), suggesting that TEs and associated microsatellites may play a role in shaping venom gene duplication and rearrangement, possibly by facilitating increased rates of local ectopic recombination (Castoe et al., 2011). Indeed, within each of the major venom gene clusters in the Prairie Rattlesnake genome, there are copies of TEs of the same length and relative age within or proximal to closely related venom gene homologs (e.g., PLA2-A and PLA2-B; Figure 3.3b), consistent with duplication events that have duplicated both venom genes and surrounding TEs together. Despite the apparent associations between venom genes, frequent tandem duplication of venom genes, and TEs, what specific roles TEs (and microsatellites) may have played in the evolution of venom clusters remains an open and intriguing question.

Despite the intriguing biology of venom gene clusters, the complexity of the organization and evolutionary history of tandem duplication of venom gene families, combined with close associations with repeat elements, leads to venom clusters being among the most difficult regions of the genome to assemble. For example, even in genomes that are reasonably well assembled for most sets of genes, venom gene clusters may remain highly fragmented and poorly assembled (e.g., Vonk et al., 2013). However, recent advances in genome sequencing and assembly technology have resulted in well-assembled and annotated venom gene regions for several venomous snakes, providing a jumping-off point for further investigation into the architecture of venom gene regions and their associations with TEs.

3.2.4 WHERE ARE VENOM GENES LOCATED IN THE GENOME?

Chromosome-level snake genome assemblies have also allowed us to identify accurately the genomic location of venom gene families to specific regions of chromosomes for the first time (Figure 3.4). In the Prairie Rattlesnake, these analyses indicate that venom genes are found on almost all autosomes (i.e., non-sex chromosomes) but are especially

enriched on very small (e.g., 5–22 Mbp) microchromosomes (Schield et al., 2019b). Similar analyses in the Habu (*Protobothrops flavoviridis*) genome identified a large number of microchromosome-linked venom genes (Shibata et al., 2018). Interestingly, available evidence suggests that there tends to be a bias in the characteristics of venom gene families and where they are located – venom gene families with numerous paralogs that contribute substantially to venom composition generally tend to be located on microchromosomes. For example, in the Prairie Rattlesnake, the snake venom metalloproteinase (SVMP), serine protease (SVSP) and phospholipase A₂ (PLA₂) gene families are each encoded within discrete regions of three distinct microchromosomes. These three families are therefore genetically unlinked from one another, meaning that SVMP, SVSP and PLA₂ alleles sort independently of one another in rattlesnake populations. Among the few contrary examples to this trend of high-copy venom gene families being located on microchromosomes, in the Indian Cobra, the major three-finger toxin gene cluster is localized to telomeric regions of macrochromosomes (mostly chromosome 3; Suryamohan et al., 2020).

The chromosomal location of venom gene regions is of particular interest because different regions of the genome have distinctive composition (i.e., gene content, repeat content, proportion GC bases, etc.) and possess fundamentally distinct evolutionary characteristics. As described earlier for avian and squamate microchromosomes, rattlesnake microchromosomes are known to have high gene density, GC content, genetic diversity, and recombination rates compared with macrochromosomes (Schield et al., 2019b, 2020). Telomeric regions of macrochromosomes also have high recombination rates. Because recombination acts to break up associations between alleles, higher recombination rates are predicted to allow natural selection to act more efficiently on genes (Otto and Barton, 1997; Cutter and Payseur, 2013). It is notable that several expanded venom gene families in these divergent venomous snake species are localized to genomic regions with high recombination, as such frequent recombination may in part contribute to the efficacy of natural selection that has been demonstrated in snake venom genes (e.g., Daltry et al., 1996; Juárez et al., 2008; Axel et al., 2009; Wong and Belov, 2012). Future studies using whole-genome data from venomous reptile populations will be useful for investigating the roles of recombination and selection in the evolution of snake venom.

3.3 REGULATION OF VENOM GENES – THE BLACK BOX LINKING THE GENOME WITH VENOM

3.3.1 TOWARDS AN UNDERSTANDING OF VENOM REGULATION

The majority of studies of reptile venoms have focused primarily at either the level of genes (i.e., what genes are present, how they have evolved, and at what levels they are expressed) or the secreted venom itself (i.e., what proteins make up

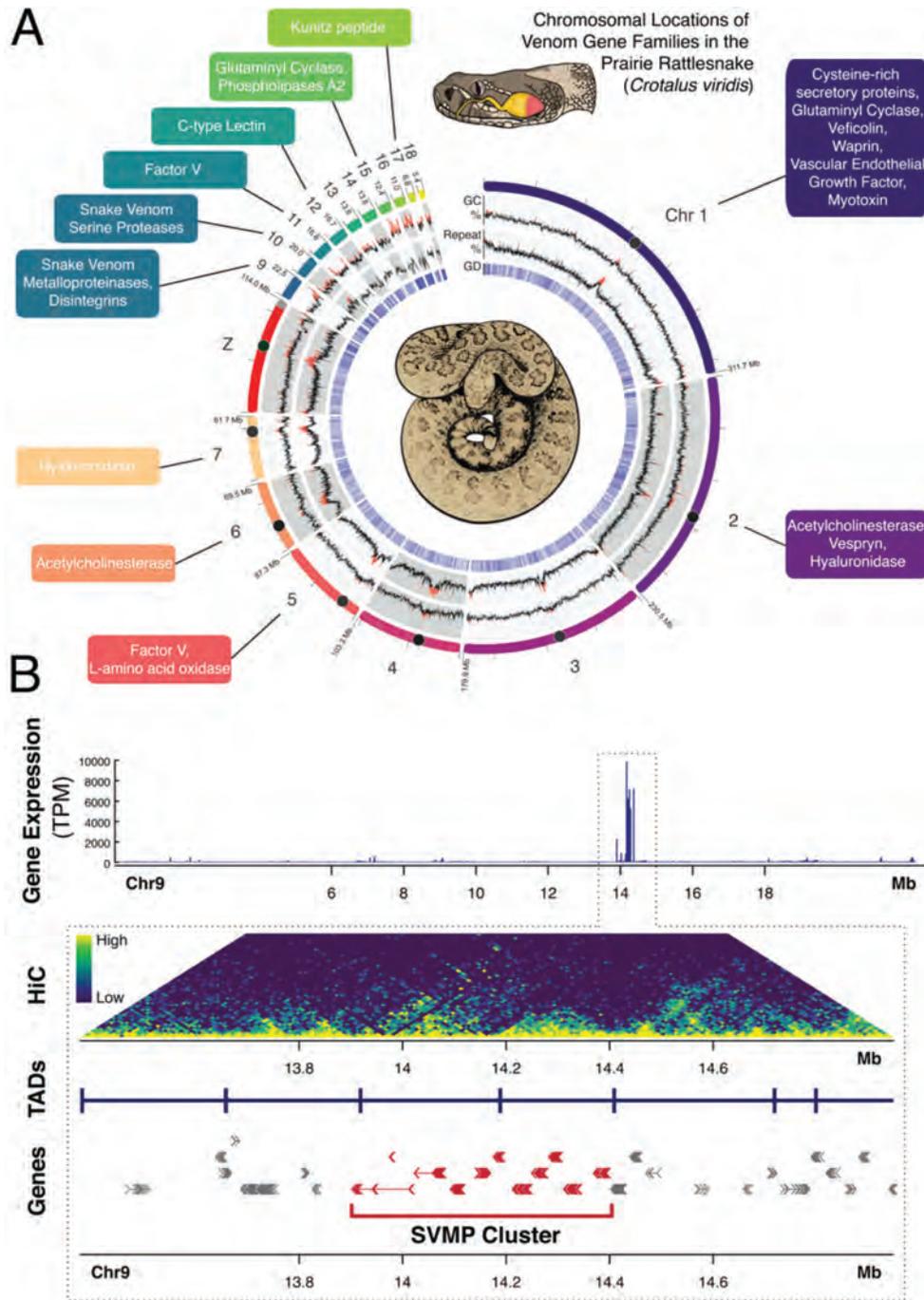


FIGURE 3.4 The genomic location and organization of snake venom genes. (A) Chromosomal locations of venom gene families in the Prairie Rattlesnake. Ideograms corresponding to individual chromosomes are arranged in a circle. For macrochromosomes, inferred relative positions of centromeres are shown as black circles. Inner rings within the circle correspond to the proportion of GC bases (GC%), proportion of bases made up by repeat elements (Repeat %), and the density of genes (GD) in 100 kb genomic windows. For GC% and Repeat%, regions with values greater than the genome-wide median are shown in red. For reference, microchromosomes are chromosomes 9–18. (B) Organization and regulation of the snake venom metalloproteinase (SVMP) gene family cluster. The top panel depicts gene expression in TPM across all of chromosome 9; the SVMP region is outlined by the dotted lines. The panels below are zoomed to the SVMP cluster and immediately surrounding regions of chromosome 9, showing the degree of Hi-C chromatin contacts (dark blue = low, yellow = high), the arrangement of topological association domains (TADs), and the arrangement of SVMP and non-venom genes within the region (SVMP genes are highlighted in red). (Panel (a) redrawn from Schield et al., 2019b.)

venoms, and what the biological activities of these proteins are). Remarkably few studies have investigated the mechanisms that regulate expression of venom genes and the degree to which this regulation contributes to the evolution and variation of venom phenotypes (e.g., Kerchova et al., 2004, 2008; Luna et al., 2009; Hargreaves et al., 2014; Schield et al., 2019a). The “black box” that is the regulation of venom therefore represents a major gap in our understanding of reptile venoms, and a better understanding of what is contained within this black box will allow us to link our understanding of genomic variation directly to venom variation.

3.3.2 GENOMES ADD VALUE TO TRANSCRIPTOMIC AND PROTEOMIC DATA

Numerous studies of venom have leveraged transcriptomic data (Rokyta et al., 2011, 2012, 2013; Hargreaves et al., 2014), where all mRNA transcripts in a given tissue (e.g., venom gland) are sequenced to infer the protein-coding content of venom (and also of the genome to some extent) and the relative degree of expression of genes. Transcriptomic studies are far less expensive and computationally more feasible than whole-genome sequencing and assembly and are thus practical in systems that lack established genome resources. Accordingly, transcriptomic studies across a broad representation of venomous species have substantially contributed to our understanding of venom composition and tissue-specific gene expression (Casewell et al., 2009; Rokyta et al., 2012, 2013; Vonk et al., 2013).

While transcriptomic studies quantify the result of some aspects of venom gene regulation, they are unable to provide meaningful substantial insight into the mechanisms that drive this regulation, such as regulatory regions (i.e., promoters and enhancers) and features of chromatin structure and organization. The typical structure of a venom gene cluster can also make it difficult to infer information accurately, such as the number of paralogs of a given venom gene family, from transcriptome data. Based on transcriptome sequences, for example, it might be difficult to discern the distinction between multiple alleles at the same venom locus, alternative splice forms derived from the same venom gene, or transcripts derived from multiple recently duplicated (and thus similar) venom genes. Accordingly, in the absence of reference genomes for the same or closely related species, there is some ambiguity in translating venom protein diversity to transcript diversity and ultimately, to inferences of the diversity of venom genes that are encoded in the genome. However, as the cost of whole-genome sequencing continues to decrease, and new genomic resources are produced for a greater diversity of species, the value of transcriptomic data for studying aspects of venom expression and variation, by linking it to genome data, will undoubtedly increase in the future.

3.3.3 UNDERSTANDING THE FACTORS THAT DIRECT THE REGULATION OF VENOM PRODUCTION

The few existing experimental studies of venom gene regulation have provided evidence that venom production in

Bothrops jararaca is triggered by the stimulation of α - and β -adrenoceptors and the subsequent activation of the ERK signaling pathway (Kerchova et al., 2004, 2008). The transcription factors nuclear factor κ B (NF- κ B) and activating protein 1 (AP-1) have also been implicated as having roles in stimulating venom gene expression (Luna et al., 2009). Yet, it remains unknown whether these signaling mechanisms drive gene expression across multiple species or to what degree other mechanisms are involved. More recently, the transcription factors GRHL1, a component of the ERK signaling cascade, and multiple nuclear factor (NFI) transcription factors were implicated in venom gene regulation in the Prairie Rattlesnake based on the upregulation of these transcription factors during venom production and the presence of predicted binding sites for these transcription factors near a large number of venom genes (Schield et al., 2019b). Notably, however, the binding sites for these transcription factors were not unique or even statistically overrepresented in venom gene regions, suggesting that while they may be involved, the expression and activity of these transcription factors is likely not entirely sufficient to drive venom gene regulation, and that these represent only the first steps towards understanding mechanisms that control venom expression. Additionally, this study did not find evidence for upregulation of NF- κ B or AP-1 transcription factors in the Prairie Rattlesnake, which may indicate that distinct or perhaps nuanced regulatory mechanisms have evolved in divergent lineages of venomous reptiles.

Beyond transcription factors, recent studies also highlight the importance of chromatin structure – the open/closed state of regions of the genome together with its three-dimensional shape – in the regulation of venom (Schield et al., 2019b). This study found that venom gene clusters tend to inhabit distinct topologically associated domains (TADs), which are regions of the genome that exhibit an elevated degree of physical interaction within the region yet limited interaction outside the domain, that apparently regulate venom gene expression by isolating active gene expression to genes within the TAD (Figure 3.4b). These findings suggest that venom regulation is a tightly controlled process involving transcription factors and physical regulation of chromatin to direct transcription of venom genes precisely. Future studies to investigate further the role of three-dimensional chromatin structure may provide additional insight into the role of chromatin organization in driving venom gene regulation and aid in the discovery of other transcription factors and regulatory elements that are vital for better understanding the links between the genome and the regulation of venom production.

3.4 POPULATION-LEVEL STUDIES OF VENOM VARIATION AND EVOLUTION

3.4.1 POPULATION-LEVEL SAMPLING OF VENOMOUS REPTILE GENOMES

The increasing feasibility of genome sequencing for multiple individuals within a species to address questions related to the variation and evolution of venom represents an exciting

and emerging area of research. These population genomic–style datasets will, for the first time, enable the evaluation of long-standing hypotheses and predictions about venom evolution, finally to be tested with appropriate robust population-level sampling of genomic data. Examples of such prominent and long-standing questions are those related to the precise patterns of natural selection that may drive the diversity of venom gene alleles and how these patterns of natural selection may be shaped by environmental factors, local prey abundance, and the evolution of venom resistance in prey (Fry and Wuster, 2004; Casewell et al., 2012, 2013).

The precise patterns of natural selection acting on venom gene regions have yet to be thoroughly analyzed in a robust population genomic framework, although recent studies leveraging data from multiple venomous snakes have begun paving the way towards a more thorough understanding of venom evolution and have raised new questions regarding the evolutionary forces that drive venom variation. For example, Aird et al. (2017) analyzed whole-genome resequencing data for 20 Taiwan Habu (*Protobothrops mucrosquamatus*) individuals and a single Sakishima Habu (*P. elegans*) to estimate sorting and fixed genetic variation within and between these two species. They compared trends in coding variation in venom genes and housekeeping genes and found evidence for a greater proportion of amino acid substitutions fixed by natural selection and a reduction in selective constraint on venom genes. The authors did not, however, analyze signatures of selection at linked sites, perhaps because of the fragmentary nature of the *P. mucrosquamatus* genome, leaving open the questions of exactly what modes of selection operate on individual venom genes and how natural selection on venom influences broader patterns of genetic diversity (e.g., via selective sweeps that may reduce diversity in regions linked to venom loci). In a separate study on a closely related species, Shibata et al. (2018) also identified signatures of selection on venom genes in the *P. flavoviridis* genome based on an excess of non-synonymous variants in venom genes relative to their non-venom homologs. While their study did not include data from populations, a comparison between venom and non-venom genes with shared ancestry is notable, as non-venom homologs are by their nature on distinct evolutionary trajectories from venom genes.

Other studies have approached the question of natural selection on venom variation using population sampling in a geographic context, including studies focused on populations of the Mojave Rattlesnake (*Crotalus scutulatus*). Different Mojave Rattlesnake populations have strongly differentiated venoms – “type A” populations possess a potent neurotoxin encoded by two phospholipase A₂ subunits (Mojave toxin; MTX) and have relatively low metalloproteinase activity (Glenn et al., 1983; Mackessy, 2008). By contrast, “type B” populations have hemorrhagic venom with high metalloproteinase activity – type A and type B are also referred to as “type II” and “type I” venoms, respectively. Type A and B populations are scattered throughout the geographic distribution of the species (Strickland et al., 2018), and intermediate “A+B” phenotypes are sometimes observed in contact zones between type A and

B populations, making the Mojave Rattlesnake a particularly valuable system for investigating the evolutionary forces and selective pressures driving venom variation.

Two recent studies combined genomic, transcriptomic and proteomic data from Mojave Rattlesnake populations to investigate the geographic composition of venom in fine detail and also to identify the determinants of A and B phenotypes in nature. The first, by Strickland et al. (2018), explored venom variation throughout the entire distribution (including sampling of populations throughout Mexico). The authors characterized individuals as having type A, B or A+B venom phenotypes and then compared venom variation with an array of biotic and abiotic factors. They found that variation in the A versus B venom phenotypes occurs at a much finer evolutionary scale than previously identified, with the occurrence of type A, B and A+B venoms in each of the major clades identified by Schield et al. (2019b), with the exception of *C. s. salvini*, which only has type A venom. They also determined that type A+B individuals are comparatively rare, occurring in regions of contact between type A and B population clusters, and concluded that the fine scale of fixation for A or B venom phenotypes is a likely result of strong divergent local selection for alternative hemorrhagic or neurotoxic venom strategies. In the second study, Zancolli et al. (2019) focused on the lineage of Mojave Rattlesnakes that occupies regions of western New Mexico, southern Arizona and southern California. They inferred that venom variation is not associated with population structure or diet composition but instead with environmental heterogeneity, and most notably with temperature (this pattern was also observed by Strickland et al., 2018). They also concluded that the fine scale of geographic variation in venom is the likely result of strong differential selective regimes, and that the lack of evidence for an association between diet and venom may be due to more nuanced predatory–prey arms race dynamics (Zancolli et al., 2019).

3.4.2 THE RELEVANCE OF HYBRID ZONES FOR STUDYING VENOM

Hybrid zones provide a unique and valuable opportunity to study how the exchange of genetic material facilitates the inheritance of and interactions among traits (Barton and Hewitt, 1989), including venom. Hybridization between venomous snake lineages has been well documented, including between species with varying degrees of genetic divergence and remarkably different venom phenotypes (Zancolli et al., 2016; Schield et al., 2017, 2019b; Harrington et al., 2018). A major question relevant to studying hybrid zones, for example, is whether lineages may acquire new venom components or alternative venom phenotypes by hybridizing with lineages with distinct venom properties. Zancolli et al. (2016) addressed this by examining a narrow hybrid zone between type A Mojave Rattlesnakes (*Crotalus scutulatus*) and the Prairie Rattlesnake (*C. viridis*), which has hemorrhagic type B venom. They found hybrid individuals that possessed both hemotoxic venom components from the Prairie Rattlesnake and the neurotoxic MTX from the Mojave Rattlesnake.

Interestingly, they found no evidence of MTX in the parental Prairie Rattlesnake population and proposed that introgression of this potent neurotoxin may not be advantageous (i.e., may be selected against). In contrast to these findings, Dowell et al. (2018) found evidence that phospholipase A₂ and metalloproteinase gene regions of neurotoxic Southern Pacific Rattlesnakes (*C. oreganus helleri*) were likely acquired via hybridization with Mojave Rattlesnakes. It is unclear, however, how relevant the potential for introgression between these species is to the venom composition of natural populations, as most populations of *C. o. helleri* possess multiple expressed metalloproteinases in their venoms (Mackessy, 2010b; Sunagar et al., 2014).

Lab-based studies have also demonstrated expression of mixed venom phenotypes within hybrid offspring. In one such example, Smith and Mackessy (2016) crossed a male type I (B) Southern Pacific Rattlesnake with a type II (A) female Mojave Rattlesnake and quantified venom expression over time, finding that hybrids expressed venom proteins apparently derived from parental venom cocktails, with the male offspring providing the strongest evidence of co-expressed type I and II venoms, while the female offspring exhibited decreasing levels of metalloproteinase activity over time. These patterns could suggest sex-biased effects and/or ontogenetic shifts in venom throughout the lifespan of hybrid offspring and may also indicate variation in the inheritance of venom components and their coevolved mechanisms of regulation that depend on genome-wide contexts and dosage effects.

An intriguing pattern from the previously mentioned studies examining natural hybrid zones is the relatively low number of individuals expressing both parental venom phenotypes and the lack of penetrance of the presumably adaptive neurotoxin in adjacent populations (e.g., Zancolli et al., 2016; Strickland et al., 2018). This pattern is notable, because it would seem that expressing both venom phenotypes would confer an evolutionary advantage for securing diverse prey and/or securing prey more effectively. One explanation may be that there are constraints on venom composition imposed by coevolutionary interactions with other regions of the genome that regulate venom or prevent auto-toxicity. Indeed, hybrid fitness and venom phenotypes may depend greatly on the co-inheritance of venom alleles with these factors, leading to the comparatively rare cases of hybrids that are able to express both parental phenotypes. For example, in the case of neurotoxic Southern Pacific Rattlesnakes (Dowell et al., 2018), genomic regions that have coevolved with the neurotoxic phospholipase A₂ haplotype were likely also inherited during hybridization with the Mojave Rattlesnake.

Together, access to hybrid zones between parental lineages with divergent venoms, recent methodological advances for studying hybrid zones (e.g., Gompert and Buerkle, 2011; Derryberry et al., 2014; Schumer et al., 2018; Martin et al., 2019), and the overall increase in genomic sequencing capabilities enabling population genomic sampling collectively hold promise for illuminating new paradigms and surprising interactions that dictate the inheritance, expression, and evolution of venom.

3.5 CONCLUSIONS

Despite substantial progress in forging connections between the genome, venom genes, their transcripts, and venom proteins and their effects, there are still substantial advances to be made with the availability of genomic resources for venomous reptiles. One fundamental step forward would be the availability of chromosome-level and well-assembled genomes for multiple diverse venomous squamates to provide multiple complete genome references in which venom genes, along with their genomic context, can be directly linked to venom gene transcripts, venom proteins, and patterns of evolution that drive difference in venom composition. Accordingly, the writing of this chapter corresponds with a major transition in the roles of genomics in studying venom, as the field moves from gaining the first glimpses of venomous reptile genomes toward having access to far more complete and useful genomes and new capabilities to sample genomes at the scale of populations. The impacts of this transition will undoubtedly represent a fundamental shift in our abilities to address long-standing questions about venoms in ways we have outlined in this chapter and probably even more ways that are difficult to envision creatively at this point.

REFERENCES

- Adams, R.H., H. Blackmon, J. Reyes-Velasco, D.R. Schield, D.C. Card, A.L. Andrew, N. Waynewood, T.A. Castoe. 2016. Microsatellite landscape evolutionary dynamics across 450 million years of vertebrate genome evolution. *Genome* 59:295–310.
- Aird, S.D., J. Arora, A. Barua, L. Qiu, K. Terada, A.S. Mikheyev. 2017. Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. *Genome Biol. Evol.* 9:2640–49.
- Alfoldi, J., F. Di Palma, M. Grabherr, C. Williams, L. Kong, E. Mauceli, P. Russell, C.B. Lowe, R.E. Glor, J.D. Jaffe, D.A. Ray. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 477:587–91.
- Axel, B., C.E. Pook, R.A. Harrison, W. Wüster. 2009. Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. *Proc. Biol. Sci.* 276:2443–9.
- Backström, N., W. Forstmeier, H. Schielzeth, H. Mellenius, K. Nam, E. Bolund, M.T. Webster, T. Öst, M. Schneider, B. Kempnaers, H. Ellegren. 2010. The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Res.* 20(4):485–95.
- Barton, N.H., G.M. Hewitt. 1989. Adaptation, speciation and hybrid zones. *Nature* 341:497–503.
- Bradnam, K.R., J.N. Fass, A. Alexandrov, P. Baranay, M. Bechner, I. Birol, S. Boisvert, J.A. Chapman, G. Chapuis, R. Chikhi, H. Chitsaz, W.C. Chou, J. Corbeil, C. Del Fabbro, T.R. Docking, R. Durbin, D. Earl, S. Emrich, P. Fedotov, N.A. Fonseca, G. Ganapathy, R.A. Gibbs, S. Gnerre, E. Godzaridis, S. Goldstein, M. Haimel, G. Hall, D. Haussler, J.B. Hiatt, I.Y. Ho, J. Howard, M. Hunt, S.D. Jackman, D.B. Jaffe, E.D. Jarvis, H. Jiang, S. Kazakov, P.J. Kersey, J.O. Kitzman, J.R. Knight, S. Koren, T.W. Lam, D. Lavenier, F. Laviolette, Y. Li, Z. Li, B. Liu, Y. Liu, R. Luo, I. Maccallum, M.D. Macmanes, N. Maillet, S. Melnikov, D. Naquin, Z. Ning, T.D. Otto, B. Paten, O.S. Paulo, A.M. Phillippy, F. Pina-Martins, M. Place,

- D. Przybylski, X. Qin, C. Qu, F.J. Ribeiro, S. Richards, D.S. Rokhsar, J.G. Ruby, S. Scalabrin, M.C. Schatz, D.C. Schwartz, A. Sergushichev, T. Sharpe, T.I. Shaw, J. Shendure, Y. Shi, J.T. Simpson, H. Song, F. Tsarev, F. Vezzi, R. Vicedomini, B.M. Vieira, J. Wang, K.C. Worley, S. Yin, S.M. Yiu, J. Yuan, G. Zhang, H. Zhang, S. Zhou, I.F. Korf. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience* 2:10.
- Casewell, N.R., G.A. Huttley, W. Wüster. 2012. Dynamic evolution of venom proteins in squamate reptiles. *Nat. Commun.* 3:1066.
- Casewell, N.R., R.A. Harrison, W. Wüster, S.C. Wagstaff. 2009. Comparative venom gland transcriptome surveys of the saw-scaled vipers (Viperidae: *Echis*) reveal substantial intra-family gene diversity and novel venom transcripts. *BMC Genomics* 10:564.
- Casewell, N.R., S.C. Wagstaff, R.A. Harrison, C. Renjifo, W. Wüster. 2011. Domain loss facilitates accelerated evolution and non-functionalization of duplicate snake venom metalloproteinase toxin genes. *Mol. Biol. Evol.* 28:2637–49.
- Casewell, N.R., S.C. Wagstaff, W. Wüster, D.A. Cook, F.M. Bolton, S.I. King, D. Pla, L. Sanz, J.J. Calvete, R.A. Harrison. 2014. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. *PNAS* 111:9205–10.
- Casewell, N.R., W. Wüster, F.J. Vonk, R.A. Harrison, B.G. Fry. 2013. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol. Evol.* 28:219–29.
- Castoe, T.A., A.P.J. de Koning, K.T. Hall, D.C. Card, D.R. Schield, M.K. Fujita, R.P. Ruggiero, J.F. Degner, J.M. Daza, W.J. Gu, J. Reyes-Velasco, K.J. Shaney, J.M. Castoe, S.E. Fox, A.W. Poole, D. Polanco, J. Dobry, M.W. Vandewege, Q. Li, R. Schott, A. Kapusta, P. Minx, C. Feschotte, P. Uetz, D.A. Ray, F.G. Hoffmann, R. Bogden, E.N. Smith, B.S.W. Chang, F. Vonk, N.R. Casewell, C.V. Henkel, M.K. Richardson, S.P. Mackessy, A.M. Bronikowski, M. Yandell, W.C. Warren, S.M. Secor, D.D. Pollock. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *PNAS* 110:20645–50.
- Castoe, T.A., K.T. Hall, M.L. Guibotsy Mboulas, W. Gu, A.P. de Koning, S.E. Fox, A.W. Poole, V. Vemulapalli, J.M. Daza, T. Mockler, E.N. Smith, C. Feschotte, D.D. Pollock. 2011. Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. *Genome Biol. Evol.* 3:641–53.
- Charlesworth, B., P. Sniegowski, W. Stephan. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371:215–20.
- Cutter, A.D., B.A. Payseur. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat. Rev. Genet.* 14:262.
- Daltry, J.C., W. Wüster, R.S. Thorpe. 1996. Diet and snake venom evolution. *Nature* 379:537–40.
- Derryberry, E.P., G.E. Derryberry, J.M. Maley, R.T. Brumfield. 2014. HZAR: hybrid zone analysis using an R software package. *Mol. Ecol. Resour.* 14:652–63.
- Dowell, N.L., M.W. Giorgianni, S. Griffin, V.A. Kassner, J.E. Selegue, E.E. Sanchez, S.B. Carroll. 2018. Extremely divergent haplotypes in two toxin gene complexes encode alternative venom types within rattlesnake species. *Curr. Biol.* 28:1016–26.
- Dowell, N.L., M.W. Giorgianni, V.A. Kassner, J.E. Selegue, E.E. Sanchez, S.B. Carroll. 2016. The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr. Biol.* 26:2434–45.
- Dudchenko, O., M.S. Shamim, S. Batra, N.C. Durand, N.T. Musial, R. Mostofa, M. Pham, B.G. St Hilaire, W. Yao, E. Stamenova, M. Hoeger, S.K. Nyquist, V. Korchina, K. Pletch, J.P. Flanagan, A. Tomaszewicz, D. McAloose, C. Pérez Estrada, B.J. Novak, A.D. Omer, E. Lieberman Aiden. 2018. The Juicebox Assembly tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. *bioRxiv*:254797.
- Dudchenko, O., S.S. Batra, A.D. Omer, S.K. Nyquist, M. Hoeger, N.C. Durand, M.S. Shamim, I. Machol, E.S. Lander, A. Presser Aiden, E. Lieberman Aiden. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356:92–95.
- Ellegren, H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* 29:51–63.
- Fry, B.G., W. Wüster. 2004. Assembling an arsenal: Origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Mol. Biol. Evol.* 21:870–83.
- Fujita, M.K., S.V. Edwards, C.P. Ponting. 2011. The *Anolis* lizard genome: an amniote genome without isochores. *Genome Biol. Evol.* 3:974–84.
- Georges, A., Q. Li, J. Lian, D. O’Meally, J. Deakin, Z. Wang, P. Zhang, M.K. Fujita, H.R. Patel, C.E. Holleley, Y. Zhou, X. Zhang, K. Matsubara, P. Waters, J.A. Marshall Graves, S.D. Sarre, G. Zhang. 2015. High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*. *Gigascience* 4:45.
- Glenn, J.L., R.C. Straight, M.C. Wolfe, D.L. Hardy. 1983. Geographical variation in *Crotalus scutulatus scutulatus* (Mojave rattlesnake) venom properties. *Toxicon* 21:119–30.
- Gompert, Z., C.A. Buerkle. 2011. Bayesian estimation of genomic clines. *Mol. Ecol.* 20:2111–27.
- Hand, E. 2019. Snakebites, a globally neglected killer, get a “transformational” injection of research funds. *Science*. doi:10.1126/science.aay0508.
- Hargreaves, A.D., M.T. Swain, M.J. Hegarty, D.W. Logan, J.F. Mulley. 2014. Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins. *Genome Biol. Evol.* 6:2088–95.
- Harrington, S.M., B.D. Hollingsworth, T.E. Higham, T.W. Reeder. 2018. Pleistocene climatic fluctuations drive isolation and secondary contact in the red diamond rattlesnake (*Crotalus ruber*) in Baja California. *J. Biogeogr.* 45:64–75.
- Hillier, L.W., W. Miller, E. Birney, W. Warren, R.C. Hardison, C.P. Ponting, P. Bork, D.W. Burt, M.A.M. Groenen, M.E. Delany, J.B. Dodgson, International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716.
- Ikeda, N., T. Chijiwa, K. Matsubara, N. Oda-Ueda, S. Hattori, Y. Matsuda, M. Ohno. 2010. Unique structural characteristics and evolution of a cluster of venom phospholipase A 2 isozyme genes of *Protobothrops flavoviridis* snake. *Gene* 461:15–25.
- Janes, D.E., C.L. Organ, M.K. Fujita, A.M. Shedlock, S.V. Edwards. 2010. Genome evolution in reptilia, the sister group of mammals. *Annu. Rev. Genomics Hum. Genet.* 11:239–64.
- Juárez, P., I. Comas, F. González-Candelas, J.J. Calvete. 2008. Evolution of snake venom disintegrins by positive darwinian selection. *Mol. Biol. Evol.* 25:2391–407.
- Kerchove, C.M., M.S.A. Luna, M.B. Zablith, M.F.M. Lazari, S.S. Smaili, N. Yamanouye. 2008. α 1-adrenoceptors trigger the snake venom production cycle in secretory cells by activating phosphatidylinositol 4, 5-bisphosphate hydrolysis and ERK signaling pathway. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* 150:431–7.

- Kerchova, C.M., S.M. Carneiro, R.P. Markus, N. Yamanouye. 2004. Stimulation of the α -adrenoceptor triggers the venom production cycle in the venom gland of *Bothrops jararaca*. *J. Exp. Biol.* 207:411–16.
- Lind, A.L., Y.Y.Y. Lai, Y. Mostovoy, A.K. Holloway, A. Iannucci, A.C.Y. Mak, M. Fondi, V. Orlandini, W.L. Eckalbar, M. Milan, M. Rovatsos, I.G. Kichigin, A.I. Makunin, M. Johnson Pokorná, M. Altmanová, V.A. Trifonov, E. Schijlen, L. Kratochvíl, R. Fani, P. Velensky, I. Reháč, T. Patarnello, T.S. Jessop, J.W. Hicks, O.A. Ryder, J.R. Mendelson III, C. Ciofi, P. Kwok, K.S. Pollard, B.G. Bruneau. 2019. Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. *Nat. Ecol. Evol.* 3:1241–52.
- Liu, Y., Q. Zhou, Y. Wang, L. Luo, J. Yang, L. Yang, M. Liu, Y. Li, T. Qian, Y. Zheng. 2015. *Gekko japonicus* genome reveals evolution of adhesive toe pads and tail regeneration. *Nat. Commun.* 6:10033.
- Luna, M.S.A., T.M.A. Hortencio, Z.S. Ferreira, N. Yamanouye. 2009. Sympathetic outflow activates the venom gland of the snake *Bothrops jararaca* by regulating the activation of transcription factors and the synthesis of venom gland proteins. *J. Exp. Biol.* 212:1535–43.
- Lynch, V.J. 2007. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A₂ genes. *BMC Evol. Biol.* 7:2.
- Mackessy, S.P. 2008. Venom composition in rattlesnakes: trends and biological significance. In: Hayes W.K., K.R. Beaman, M.D. Cardwell, S.P. Bush, editors. *The Biology of Rattlesnakes*. Loma Linda, CA: Loma Linda University Press.
- Mackessy, S.P. 2010a. *Handbook of venoms and toxins of reptiles*. New York, NY: CRC press, 521pp.
- Mackessy, S.P. 2010b. Evolutionary trends in venom composition in the Western Rattlesnakes (*Crotalus viridis sensu lato*): Toxicity vs. tenderizers. *Toxicon* 55:1463–74.
- Martin, S.H., J.W. Davey, C. Salazar, C.D. Jiggins. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol.* 17:e2006288.
- Matsubara, K., H. Tarui, M. Toriba, K. Yamada, C. Nishida-Umehara, K. Agata, Y. Matsuda. 2006. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *PNAS* 103:18190–95.
- McKenna, D.D., E.D. Scully, Y. Pauchet, K. Hoover, R. Kirsch, S.M. Geib, R.F. Mitchell, R.M. Waterhouse, S.-J. Ahn, D. Arsala, J.B. Benoit, H. Blackmon, T. Bledsoe, J.H. Bowsher, A. Busch, B. Calla, H. Chao, A.K. Childers, C. Childers, D.J. Clarke, L. Cohen, J.P. Demuth, H. Dinh, H. Doddapaneni, A. Dolan, J.J. Duan, S. Dugan, M. Friedrich, K.M. Glastad, M.A.D. Goodisman, S. Haddad, Y. Han, D.S.T. Hughes, P. Ioannidis, J.S. Johnson, J.W. Jones, L.A. Kuhn, D.R. Lance, C.-Y. Lee, S.L. Lee, H. Lin, J.A. Lynch, A.P. Moczek, S.C. Murali, D.M. Muzny, D.R. Nelson, S.R. Palli, K.A. Panfilio, D. Pers, M.F. Poelchau, H. Quan, J. Qu, A.M. Ray, J.P. Rinehart, H.M. Robertson, R. Roehrdanz, A.J. Rosendale, S. Shin, C. Silva, A.S. Torsion, I.M. Vargas Jentsch, J.H. Werren, K.C. Worley, G. Yocum, E.M. Zdobnov, R.A. Gibbs, S. Richards. 2016. Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. *Genome Biol.* 17:227.
- McQueen, H.A., G. Siriaco, A.P. Bird. 1998. Chicken microchromosomes are hyperacetylated, early replicating, and gene rich. *Genome Res.* 8:621–30.
- McQueen, H.A., J. Fantes, S.H. Cross, V.H. Clark, A.L. Archibald, A.P. Bird. 1996. CpG islands of chicken are concentrated on microchromosomes. *Nat. Genet.* 12:321–4.
- Olmo, E. 2005. Rate of chromosome changes and speciation in reptiles. *Genetica* 125:185–203.
- Organ, C.L., R.G. Moreno, S.V. Edwards. 2008. Three tiers of genome evolution in reptiles. *Integr. Comp. Biol.* 48:494–504.
- Otto, S.P., N.H. Barton. 1997. The evolution of recombination: Removing the limits to natural selection. *Genetics* 147:879–906.
- Pasquesi, G.I.M., R.H. Adams, D.C. Card, D.R. Schield, A.B. Corbin, B.W. Perry, J. Reyes-Velasco, R.P. Ruggiero, M.W. Vandewege, J.A. Shortt, T.A. Castoe. 2018. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat. Commun.* 9:2774.
- Perry, B.W., D.C. Card, J.W. McGlothlin, G.I.M. Pasquesi, R.H. Adams, D.R. Schield, N.R. Hales, A.B. Corbin, J.P. Demuth, F.G. Hoffman, M.W. Vandewege, R.K. Schott, N. Bhattacharyya, B.S.W. Chang, N.R. Casewell, G. Whiteley, J. Reyes-Velasco, S.P. Mackessy, K.B. Storey, K.K. Biggar, C.N. Passow, C. Kuo, S.E. McGaugh, A.M. Bronikowski, J. deKoning, S.V. Edwards, M.E. Pfreder, P. Minx, E.D. Brodie III, E.D. Brodie, Jr., W.C. Warren, T.A. Castoe. 2019. Multi-species comparisons of snakes identify coordinated signaling networks underlying post-feeding intestinal regeneration. *Proc. R. Soc. B.* 286:20190910.
- Reyes-Velasco, J., D.C. Card, A.L. Andrew, K.J. Shaney, R.H. Adams, D.R. Schield, N.R. Casewell, S.P. Mackessy, T.A. Castoe. 2015. Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. *Mol. Biol. Evol.* 32:173–83.
- Rokyta, D.R., A.R. Lemmon, M.J. Margres, K. Aronow. 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics* 13:312.
- Rokyta, D.R., K.P. Wray, A.R. Lemmon, E.M. Lemmon, S.B. Caudle. 2011. A high-throughput venom-gland transcriptome for the eastern diamondback rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. *Toxicon* 57:657–71.
- Rokyta, D.R., K.P. Wray, M.J. Margres. 2013. The genesis of an exceptionally lethal venom in the timber rattlesnake (*Crotalus horridus*) revealed through comparative venom-gland transcriptomics. *BMC Genomics* 14:394.
- Ruggiero, R.P., Y. Bourgeois, S. Boissinot. 2017. LINE insertion polymorphisms are abundant but at low frequencies across populations of *Anolis carolinensis*. *Front. Genet.* 8:44.
- Schild, D.R., B.W. Perry, R.H. Adams, D.C. Card, T. Jezkova, G.I.M. Pasquesi, Z.L. Nikolakis, K. Row, J.M. Meik, C.F. Smith, S.P. Mackessy, T.A. Castoe. 2019a. Allopatric divergence and secondary contact with gene flow: a recurring theme in rattlesnake speciation. *Biol. J. Linn. Soc.* 128:149–69.
- Schild, D.R., D.C. Card, N.R. Hales, B.W. Perry, G.I.M. Pasquesi, H. Blackmon, R.H. Adams, A.B. Corbin, C.F. Smith, B. Ramesh, J.P. Demuth, E. Betrán, M. Tollis, J.M. Meik, S.P. Mackessy, T.A. Castoe. 2019b. The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Res.* 29:590–601.
- Schild, D.R., D.C. Card, R.H. Adams, A.B. Corbin, T. Jezkova, N.R. Hales, J.M. Meik, C.L. Spencer, L.L. Smith, G. Campillo Garcia, N.M. Bouzid, J.L. Strickland, R.W. Bryson Jr., C.L. Parkinson, O. Flores-Viella, S.P. Mackessy, T.A. Castoe. 2018. Cryptic genetic diversity, population structure, and gene flow in the Mojave Rattlesnake (*Crotalus scutulatus*). *Mol. Phylogenet. Evol.* 127:669–81.

- Schild, D.R., R.H. Adams, D.C. Card, B.W. Perry, G.I.M. Pasquesi, T. Jezkova, D.M. Portik, A.L. Andrew, C.L. Spencer, E.E. Sanchez, M.K. Fujita, S.P. Mackessy, T.A. Castoe. 2017. Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. *Ecol. Evol.* 7:3951–66.
- Schild, D.R., G.I.M. Pasquesi, B.W. Perry, R.H. Adams, Z.L. Nikolakis, A.K. Westfall, R.W. Orton, J.M. Meik, S.P. Mackessy, T.A. Castoe. 2020. Snake recombination landscapes are concentrated in functional regions despite PRDM9. *Mol. Biol. Evol.* 37:1272–94.
- Schumer, M., C. Xu, D.L. Powell, A. Durvasula, L. Skov, C. Holland, J.C. Blazier, S. Sankararaman, P. Andolfatto, G.G. Rosenthal. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* 360:656–60.
- Shaney, K.J., D.C. Card, D.R. Schild, R.P. Ruggiero, D.D. Pollock, S.P. Mackessy, T.A. Castoe. 2014. Squamate reptile genomics and evolution. In Gopalakrishnakone, P. (ed.), *Handbooks of Toxicology. Venom Genomics and Proteomics*, 18 p. Springer Science, Dordrecht.
- Shibata, H., T. Chijiwa, N. Oda-Ueda, H. Nakamura, K. Yamaguchi, S. Hattori, K. Matsubara, Y. Matsuda, A. Yamashita, A. Isomoto, K. Mori, K. Tashiro, S. Kuhara, S. Yamasaki, M. Fujie, H. Goto, R. Koyanagi, T. Takeuchi, Y. Fukumaki, M. Ohno, E. Shoguchi, K. Hisata, N. Satoh, T. Ogawa. 2018. The habu genome reveals accelerated evolution of venom protein genes. *Sci. Rep.* 8:11300.
- Simão, F.A., R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–12.
- Smith, C.F., S.P. Mackessy. 2016. The effects of hybridization on divergent venom phenotypes: Characterization of venom from *Crotalus scutulatus scutulatus* x *Crotalus oreganus helleri* hybrids. *Toxicon* 120:110–23.
- Srikulnath, K., C. Nishida, K. Matsubara, Y. Uno, A. Thongpan, S. Suputtitada, S. Apisitwanich, Y. Matsuda. 2009. Karyotypic evolution in squamate reptiles: comparative gene mapping revealed highly conserved linkage homology between the butterfly lizard (*Leiolepis reevesii rubritaeniata*, Agamidae, Lacertilia) and the Japanese four-striped rat snake (*Elaphe quadrivirgata*, Colubridae, Serpentes). *Chromosom. Res.* 17:975–86.
- Strickland, J.L., C.F. Smith, A.J. Mason, D.R. Schild, M. Borja, G. Casteñeda-Gaytán, C.L. Spencer, L.L. Smith, A. Trápaga, N.M. Bouzid, G. Campillo-García, O.A. Flores-Villela, D. Antonio-Rangel, S.P. Mackessy, T.A. Castoe, D.R. Rokyta, and C.L. Parkinson. 2018. Evidence for divergent patterns of local selection driving venom variation in Mojave rattlesnakes (*Crotalus scutulatus*). *Sci. Rep.* 8:17622.
- Sunagar, K., E.A.B. Undheim, H. Scheib, E.C.K. Gren, C. Cochran, C.E. Person, I. Koludarov, W. Kelln, W.K. Hayes, G.F. King, A. Antunes, B.G. Fry. 2014. Intraspecific venom variation in the medically significant Southern Pacific Rattlesnake (*Crotalus oreganus helleri*): Biodiscovery, clinical and evolutionary implications. *J. Proteomics* 99:68–83.
- Suryamohan, K., S.P. Krishnankutty, J. Guillory, M. Jevit, M.S. Schröder, M. Wu, B. Kuriakose, O.K. Mathew, R.C. Perumal, I. Koludarov, L.D. Goldstein, K. Senger, M.D. Dixon, D. Velayutham, D. Vargas, S. Chaudhuri, M. Muraleedharan, R. Goel, Y.J. Chen, A. Ratan, P. Liu, B. Faherty, G. de la Rosa, H. Shibata, M. Baca, M. Sagolla, J. Ziai, G.A. Wright, D. Vucic, S. Mohan, A. Antony, J. Stinson, D.S. Kirkpatrick, R.N. Hannoush, S. Durinck, Z. Modrusan, E.W. Stawiski, K. Wiley, T. Raudsepp, R.M. Kini, A. Zachariah, S. Seshagiri. 2020. The Indian cobra genome and transcriptome enables comprehensive identification of venom toxins. *Nat. Genet.* 52:106–17.
- Tollis, M., S. Boissinot. 2011. The transposable element profile of the *Anolis* genome: how a lizard can provide insights into the evolution of vertebrate genome size and structure. *Mob. Genet. Elements* 1:107–11.
- Vonk, F.J., N.R. Casewell, C.V. Henkel, A. Heimberg, H.J. Jansen, R.J. R. McCleary, H.M.E. Kerckamp, R. Vos, I. Guerreiro, J.J. Calvete, W. Wüster, A.E. Woods, J.M. Logan, R.A. Harrison, T.A. Castoe, A.P.J. de Koning, D.D. Pollock, C. Renjifo, R.B. Currier, D. Salgado, D. Pla, L. Sanz, A.S. Hyder, J.M.C. Ribeiro, J.W. Arntzen, G.E.E.J.M. van den Thillart, M. Boetzer, W. Pirovano, R.P. Dirks, H.P. Spaink, D. Duboule, E. McGlinn, R.M. Kini, M.K. Richardson. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Nat. Acad. Sci. USA* 110:20651–6.
- Warren, W.C., D.F. Clayton, H. Ellegren, A.P. Arnold, L.W. Hillier, A. Künstner, S. Searle, S. White, A.J. Vilella, S. Fairley, A. Heger, L. Kong, C.P. Ponting, E.D. Jarvis, C.V. Mello, P. Minx, P. Lovell, T.A. Velho, M. Ferris, C.N. Balakrishnan, S. Sinha, C. Blatti, S.E. London, Y. Li, Y.C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backström, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A.R. Pfenning, J. Howard, M. Völker, B.M. Skinner, D.K. Griffin, L. Ye, W.M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X.S. Puente, T. Olender, D. Lancet, A.F. Smit, R. Hubley, M.K. Konkel, J.A. Walker, M.A. Batzer, W. Gu, D.D. Pollock, L. Chen, Z. Cheng, E.E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S.V. Edwards, S.P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E.R. Mardis, R.K. Wilson. 2010. The genome of a songbird. *Nature* 464:757–62.
- Wong, E.S.W., K. Belov. 2012. Venom evolution through gene duplications. *Gene* 496:1–7.
- Yin, W., Z.J. Wang, Q.Y. Li, J.M. Lian, Y. Zhou, B.Z. Lu, L.J. Jin, P.X. Qiu, P. Zhang, W.B. Zhu, B. Wen, Y.-J. Huang, Z.-L. Lin, B.-T. Qiu, X.-W. Su, H.-M. Yang, G.-J. Zhang, G.-M. Yan, Q. Zhou. 2016. Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. *Nat Commun.* 7:13107.
- Zancolli, G., T.G. Baker, A. Barlow, R.K. Bradley, J.J. Calvete, K.C. Carter, K. de Jager, J.B. Owens, J.F. Price, L. Sanz, A. Scholes-Higham, L. Shier, L. Wood, C.E. Wüster, W. Wüster. 2016. Is hybridization a source of adaptive venom variation in rattlesnakes? A test, using a *Crotalus scutulatus* x *viridis* hybrid zone in southwestern New Mexico. *Toxins* 8:188.
- Zancolli, G., J.J. Calvete, M.D. Cardwell, H.W. Greene, W.K. Hayes, M.J. Hegarty, H.W. Herrmann, A.T. Holycross, D.I. Lannutti, J.F. Mulley, L. Sanz, Z.D. Travis, J.R. Whorley, C.E. Wüster, W. Wüster. 2019. When one phenotype is not enough: divergent evolutionary trajectories govern venom variation in a widespread rattlesnake species. *Proc. R. Soc. B* 286:20182735.