A Role for Genomics in Rattlesnake Research:
Current Knowledge and Future Potential

Drew R. Schield1, Darren C. Card1, Jacobo Reyes-Velasco1, Audra L. Andrew1, Cassandra A. Modahl2, Stephen P. Mackey2, David D. Pollock3, and Todd A. Castoe1

1Department of Biology
University of Texas at Arlington
Arlington, Texas 76019, USA.

2School of Biological Sciences
University of Northern Colorado
Greeley, Colorado 80638, USA.

3Department of Biochemistry and Molecular Genetics
University of Colorado at Denver
Aurora, Colorado 80045, USA.

Correspondence
Todd A. Castoe
E-mail: todd.castoe@uta.edu
Web: http://www.snakegenomics.org/CastoeLab/Castoe_Lab_Home.html

Rattlesnakes comprise two genera (Crotalus and Sistrurus) and have collectively become an important model system for an impressively broad array of research questions due predominantly to their broad distribution, diversity of phenotypes and natural histories, and the medical importance of their venom. Genomic information is rapidly accumulating for various rattlesnake species, and this growing foundation of rattlesnake genomic resources will expand the type and significantly increase the depth of questions we may address regarding rattlesnake biology and evolution. Here, we highlight what is currently known about the genomics of rattlesnakes and provide a brief and general introduction to practical issues involved in assembling genomes. We also identify a number of important outstanding questions and areas of research that we view as exciting frontiers that will soon be tractable to address given a greater understanding of rattlesnake genomes and genomic variation.
Relevance of genomics

Genomics is a rapidly evolving field that is capable of inferring the totality of an organism’s genetic material by sequencing its genome and computationally reassembling these data. Analysis and comparison with other sequenced genomes can facilitate interrogating inferences whereby evolutionary changes in an organism’s genome can be linked to differences in the physical attributes, or phenotype, between organisms. Increasing computational power and new computational methods, together with new sequencing technologies and massive decreases in DNA sequencing costs, have collectively made sequencing vertebrate genomes relatively affordable and tractable. This is particularly exciting for researchers whose interests fall outside the realm of traditional model organisms, because it is now feasible to generate genomic information at a massive scale for almost any species of interest. Furthermore, it is possible to use comparative genomic approaches to bridge biological information, including predictions of gene functions, for example, across different organisms to leverage existing information, including predictions of gene functions, from these genomes to facilitate such research are exciting. There are many questions that have yet to be addressed or that will benefit substantially from further understanding that the availability of genomic data can provide (Glossary and Box 1).

Here, we provide a summary of the current understanding of rattlesnake genomics, a discussion of areas of rattlesnake research that would likely benefit directly from genomic studies, and an argument for utilizing emerging sequencing advances to pursue rattlesnake genomic resources.

Snake genome size

Though genome size is not necessarily correlated with the complexity of an organism, it has been suggested that it is linked to aspects of life history at a number of levels. At the generic and cellular levels, repetitive genetic elements, nuclear volume, cell size, and cell physiology are related to the size of the genome. Similarly, at the organismal level, genome size has been correlated with longevity, metabolic rate, and development (Gregory, 2001).

Figure 1. Literature search results for rattlesnake publications since 1960. A. Number of Google Scholar hits by year, based on the combined search terms: “Crotalus,” “Sistrurus,” “Rattlesnake,” “Rattle- snake,” and “Rattle Snake.” B) PubMed search results by year using the same search criteria as in panel A.
chromosome. If two genomes share a high level of synteny, this would mean that their genes share the same order across a particular chromosome. Synteny:

Horizontal transfer:

transcripts in transcriptomic datasets.

Differential expression:

Transcriptomics:

alleles. This type of variation is typically the most commonly observed, and tends to occur more frequently in genomic regions that do not encode proteins or other functional molecules.

Single nucleotide polymorphism (SNP):

Genomic regions that have undergone convergence in distinct lineages have the ability to mislead phylogenetic inference and yield incorrect estimates of true relationships among lineages.

Molecular convergence:

birds). Similarly, molecular convergence involves the convergent evolution of DNA and amino acid sequences to the same sequence in two lineages.

Molecular convergence:

include the measurement of cell size and abundance, karyotyping, and genome size estimation.

Flow cytometry:

estimate a computational reconstruction of the genome.

Sequencing reads:

Glossary

the conservation of blocks of physically linked genomic regions between species, resulting in loci from both species mapping to a common chromosome. If two genomes share a high level of synteny, this would mean that their genes share the same order across a particular chromosome.

Sequencing reads: raw data from sequencing machines are in the form of short (e.g., 100 nucleotides in length) sequencing reads that can be used to estimate a computational reconstruction of the genome.

Flow cytometry: a method for measuring the physical characteristics of particles using fluids and laser technology. Applications to molecular biology include the measurement of cell size and abundance, karyotyping, and genome size estimation.

Molecular convergence: convergent evolution results in the same trait occurring in multiple unrelated lineages (e.g., wings in bats, butterflies, and birds). Similarly, molecular convergence involves the convergent evolution of DNA and amino acid sequences to the same sequence in two lineages.

Single nucleotide polymorphism (SNP): a DNA sequence variation within a species in which a single nucleotide (A, T, C, or G) is different between alleles. This type of variation is typically the most commonly observed, and tends to occur more frequently in genomic regions that do not encode proteins or other functional molecules.

Differential expression: although the genome encodes a myriad of protein-coding genes, expression of these genes (as RNA transcripts) is not necessarily constant across tissues, individuals, and time. The observed differences in gene expression are often inferred from relative abundances of specific transcripts in transcriptomic datasets.

Transcriptomics: the study of RNA produced by a cell or tissue sample at a given time.

Horizontal transfer: the transmission of genetic material between organisms via mechanisms other than reproduction (vertical transfer). Genetic regions that have undergone convergence in distinct lineages have the ability to mislead phylogenetic inference and yield incorrect estimates of true relationships among lineages.

Horizontal transfer: the transmission of genetic material between organisms via mechanisms other than reproduction (vertical transfer).

Box 1. What is a genome?

Freely available complete genomes are important because they provide a central scaffold of knowledge on which to build a broader interpretation of phenotypic and genomic data, and they allow these data to be related to function, structure, and variation across individuals and species. For this reason, it is key to understand what a “complete genome” represents, and conversely, what it might not represent. A “complete genome sequence” for an organism essentially refers to the computational reconstruction (or estimate) of the actual genome as it exists in the organism. As with any estimate, it is only an approximation of the “true genome,” and different approximations may differ substantially in quality. Thus, the “completeness” of a genome is somewhat arbitrary, meaning that finished genomes can have a broad spectrum of quality and accuracy characteristics.

Practically speaking, genomes are constructed using relatively short sequencing reads (or small samples of genomic sequence) that are computationally aligned where they can be uniquely fitted together into “contigs,” which represent contiguous inferred pieces of a genome. Because vertebrate genomes contain large amounts of repetitive sequences, this aligning of unique sequences is difficult and often results in a highly fragmented set of tens of thousands of fragments (or more) of unconnected regions of the genome. These pieces are then “scaffolded” together into larger sections using various methods in which contigs are connected to other contigs, but often with some intervening sections of unknown sequence (typically these are repetitive regions). Based on this process, genomes or regions of a genome that contain a high level of repetitive DNA, and genomes that are larger in size, are more difficult computationally to piece back together well.

These “assembled genomes” often have limited intrinsic value and biological relevance prior to being annotated. Annotation is the process by which various elements of the genome are identified (e.g., exons, introns, and repeat elements). Annotation generally involves identification of such elements based on similarity to a reference genome (usually the closest related genome available), computational prediction of these elements, and often additional empirical data from studies identifying which genes are observed being transcribed in different tissues (i.e., transcriptome analysis). The diversity of tissues that are used in transcriptomic analysis will therefore strongly impact the quality of a genome annotation.

Three methods have been predominantly used to estimate genome size: Feulgen density (FD), static cell fluorometry (SCF), and flow cytometry (FC). When analyzed together, these data imply a high degree of variation in genome sizes across snake species. We believe, however, that this conclusion is most likely an artifact of inconsistencies and inaccuracies in particular methods. Genome size estimates from FD and SCF tend to be systematically higher, less precise, and often less accurate than estimates from flow cytometry for the same species, based on our analysis of data from snakes (Figure 2). In the presentation of snake genome size
estimates here, we have therefore separated estimates according to methods, with the expectation that estimates based on FC are likely the most accurate and precise (Figure 3).

For data presented here, estimates of snake genome sizes were taken from the Animal Genome Size Database (Gregory, 2013). The average snake genome size estimated using flow cytometry is 1.9 billion bases, or Gbp (Figure 3a; n = 32, range = 1.5–3.0 Gbp). The estimated average viperid genome size is 2.06 Gbp (Figure 3b; n = 21, range = 1.3–3.06 Gbp) using estimates from all of the above-mentioned methods. These estimates are only slightly lower (1.81 Gbp; n = 4, range = 1.75–1.88 Gbp) when these statistics are calculated based solely on FC data (Figure 3a, b). There are currently only two genome size estimates for rattlesnakes. Using FD, Olmo (1981) provided the first rattlesnake (South American Rattlesnake, *Crotalus durissus terrificus*) genome size estimate at 1.32 Gbp (Figure 3c). Using FC, Tiersch et al. (1989) estimated the genome size of the Timber Rattlesnake, *Crotalus horridus* to be 1.75 Gbp (Figure 3c). Based on the apparent precision of flow cytometry across most snakes (Figure 2), and our estimated average from FC for viperids (Figure 3b, c), we suggest that rattlesnake genome sizes are likely most similar to the 1.75 Gbp of *C. horridus*.

Rattlesnake genome structure

**Mitochondrial genomes**

Animal genomes can be subdivided into nuclear and mitochondrial genomes. Genomic studies are primarily focused on the nuclear genome, though the organellar mitochondrial genome has been thoroughly studied in a multitude of species, including rattlesnakes. The mitochondrial genome, which is maternally inherited, contains genes that code for proteins involved in aerobic metabolism, as well as ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) that facilitate translation in the mitochondrion. The mitochondrial genome also contains a control region which is involved in both transcription and replication of the mitochondrial genome. The mitochondrial genomes of snakes are unusual because all alethinophidian snakes (i.e., all snakes except the basal blind snakes) possess a duplicated control region of the mitochondria between ND1 and ND2 protein-coding genes (Kumazawa et al., 1998). As a comparison of mitochondrial genome structure between alethinophidian snakes and most other vertebrates, the mitochondrial genome of the Cottonmouth (*Agkistrodon piscivorus*), a close relative of rattlesnake, is shown along with the Green Iguana (*Iguana iguana*) (Figure 4). The two control regions of snakes have nearly identical nucleotide sequences, and this sequence similarity is maintained by a poorly understood mechanism of concerted evolution (Jiang et al., 2007). Alethinophidian mitochondrial protein-coding and rRNA genes have been shortened over evolutionary time, possibly to partially make up for the added length from the two control regions (Jiang et al., 2007). There is also an alethinophidian snake-specific duplication of the tRNAPro and tRNALeu adjacent to the duplicated control region (Figure 4). Thus, while most vertebrate mitochondrial genomes are ~ 16.5 kb in length, alethinophidian snake mitochondrial genomes tend to be ~ 17.3 kb (Figure 4).
Mitochondrial genes have undergone extensive adaptation and paired to the duplicate control regions of the mitochondria, snake mitochondrial proteins may have also been applied to other squamate lineages, based on evidence for extensive molecular convergence to mitochondrial proteins, which implies that snakes may have evolved co-evolutionary changes in amino acid sequences of multiple mitochondrial proteins that has occurred between other vertebrates (Castoe et al., 2008). While snake metabolism is a highly unique and specialized aerobic metabolism relative to other vertebrates (Castoe et al., 2009b). In addition which are hypothesized to act as origins of genome replication and metabolic flux in almost any other vertebrate (Secor and Diamond, 2000). It is likely that the duplicate control regions of snakes, which are homologous in different families (Matsubara et al., 2006; O'Meally et al., 2011). The sex chromosomes of snakes were shown to be homologous across multiple mitochondrial proteins that has occurred between ancestral snake lineages and agamid lizards (Castoe et al., 2009a). It is known that several snake lineages, including rattlesnakes, undergo tremendous fluctuations in physiology after eating a large meal, which involves metabolic up-regulation that greatly exceeds the metabolic flux in almost any other vertebrate (Secor and Diamond, 2000). It is likely that the duplicate control regions of snakes, which are hypothesized to act as origins of genome replication and transcription, play a role in facilitating this exceptional metabolic up-regulation (Jiang et al., 2007; Castoe et al., 2009b). In addition to the duplicate control regions of the mitochondria, snake mitochondrial genes have undergone extensive adaptation and paired to the duplicate control regions of the mitochondria, snake mitochondrial proteins may have also been applied to other squamate lineages, based on evidence for extensive molecular convergence across multiple mitochondrial proteins that has occurred between ancestral snake lineages and agamid lizards (Castoe et al., 2009b).

**Figure 4.** Mitochondrial genome annotation map for *Agkistrodon piscivorus* and *Iguana iguana* adapted from Jiang et al. (2007). Although the convention is to label each strand according to the strand from which the RNA is transcribed (i.e. heavy versus light strand), for simplicity we display control regions, ribosomal RNAs, and protein-coding genes on the outside and transfer RNAs (indicated by the letter representation of the amino acid they encode) primarily on the inside of the map. It is known that several snake lineages, including rattlesnakes, undergo tremendous fluctuations in physiology after eating a large meal, which involves metabolic up-regulation that greatly exceeds the metabolic flux in almost any other vertebrate (Secor and Diamond, 2000). It is likely that the duplicate control regions of snakes, which are hypothesized to act as origins of genome replication and transcription, play a role in facilitating this exceptional metabolic up-regulation (Jiang et al., 2007; Castoe et al., 2009b). In addition to the duplicate control regions of the mitochondria, snake mitochondrial genes have undergone extensive adaptation and paired to the duplicate control regions of the mitochondria, snake mitochondrial proteins may have also been applied to other squamate lineages, based on evidence for extensive molecular convergence across multiple mitochondrial proteins that has occurred between ancestral snake lineages and agamid lizards (Castoe et al., 2009b).

**Nuclear genome chromosomal structure and sex chromosomes**

Snake nuclear genomes are comprised of macrochromosomes and microchromosomes; the main difference between the two being size. Although microchromosomes are smaller than macrochromosomes, they tend to be richer in gene content (Smith et al., 2010). Relative to other tetrapod groups, chromosome number in snakes tends to be highly conserved, most species possess ~ 36 chromosomes, with ~ 16 macrochromosomes and ~ 20 microchromosomes (Organ et al., 2008). Karyotypes and sex chromosomes have also apparently been highly conserved during 280 million years of reptile evolution. For example, 19 out of 22 archived chicken chromosomes are syntenic to a single Anolis chromosome over their entire length (Alfoldi et al., 2011). The sex chromosomes of snakes were shown to be homologous in different families (Matsubara et al., 2006; O'Meally et al., 2010), and correspond to chromosome 6 of Reeves’ Butterfly Lizard, Lepidosaurus occidentalis (Sinha et al., 2009) and the Green Anole, Anolis carolinensis (Vicoso et al., 2011). All rattlesnakes, and pitvipers (Crotalinae) in general, are thought to possess a diploid number of 36 chromosomes (Zimmerman and Kilpatrick, 1973).

All snakes have ZW genetic sex determination (but see Booth and Schlueter, 2016), and their sex chromosomes reveal increased differentiation in a phylogenetic gradient from the morphologically “primitive” pythons (pythonids) to the more “advanced” colubrids, elapids, and vipers (Matsubara et al., 2006; Vicoso et al., 2013). Tetraptene analysis in both the Bushmaster (Lamaeuchis), pythons (Pythonidae) and pythons (Pythonidae) (highly differentiated sex chromosomes) and Pygmy Rattlesnake (Sistrurus) (highly differentiated sex chromosomes) indicates that, unlike mammals, heteromorphic ZW chromosomes in rattlesnakes lack chromosome-wide dosage compensation (Vicoso et al., 2013). Further knowledge of additional snake genomes may substantially contribute to our understanding of the origins and the sex determining genes for squamate reptiles. Also, knowledge of molecular markers that are sex chromosomes-specific might also contribute to our ability to evaluate the importance of sex-biased gene flow and dispersal in population genetic studies of rattlesnakes.

**Figure 4.** Mitochondrial genome annotation map for *Agkistrodon piscivorus* and *Iguana iguana* adapted from Jiang et al. (2007). Although the convention is to label each strand according to the strand from which the RNA is transcribed (i.e. heavy versus light strand), for simplicity we display control regions, ribosomal RNAs, and protein-coding genes on the outside and transfer RNAs (indicated by the letter representation of the amino acid they encode) primarily on the inside of the map. It is known that several snake lineages, including rattlesnakes, undergo tremendous fluctuations in physiology after eating a large meal, which involves metabolic up-regulation that greatly exceeds the metabolic flux in almost any other vertebrate (Secor and Diamond, 2000). It is likely that the duplicate control regions of snakes, which are hypothesized to act as origins of genome replication and transcription, play a role in facilitating this exceptional metabolic up-regulation (Jiang et al., 2007; Castoe et al., 2009b). In addition to the duplicate control regions of the mitochondria, snake mitochondrial genes have undergone extensive adaptation and paired to the duplicate control regions of the mitochondria, snake mitochondrial proteins may have also been applied to other squamate lineages, based on evidence for extensive molecular convergence across multiple mitochondrial proteins that has occurred between ancestral snake lineages and agamid lizards (Castoe et al., 2009b).
Molecular evolutionary rates in snake genomes
Rates of molecular evolution also differ substantially across reptile lineages. While turtle genomes evolve remarkably slowly compared to other sequenced amniotes (Shaffer et al., 2013), snake genome sequences are apparently among the fastest to evolve (Castoe et al., 2013). This trend holds when considering all sites in protein-coding genes together, as well as for synonymous neutrally evolving third codon positions (Castoe et al., 2013). Analyses of more than 40 nuclear genomes for over 150 squamate reptiles indicate a trend of accelerated neutral evolution in the ancestral lineages of squamate reptiles, snakes, and crocodyliform snakes that include the rattle snakes (Castoe et al., 2013). Furthermore, rates of evolution for Z-linked genes in snakes, including rattle snakes, are increased relative to their pseudogenome homologs, both among synonymous and amino acid sites (Vicoso et al., 2013). These findings collectively suggest that rates of evolution in snakes are exceptionally high (Castoe et al., 2013), and that mutation rates may be male-biased as they are in other animals (Vicoso et al., 2013).

Repeat element landscapes of rattlesnake genomes
Much of our perspective on vertebrate genome structure and diversity stems from sequenced mammalian genomes, though new sequence-based information on reptilian genome structure and content is emerging rapidly (Shedlock et al., 2007; Kordis, 2009; Vonk et al., 2013). Repeat elements are ubiquitous among vertebrate genomes and large portions of squamate genomes are composed of repeat elements, similar to patterns in mammals. The small number of squamate genomes sequenced indicates a highly diverse repertoire of repeat element types (Figure 5) (Shedlock et al., 2007; Castoe et al., 2011a, b, 2013). Relative to the genomes of mammals and birds, high quality annotated squamate genome assemblies exist for Anolis (Alfoldi et al., 2011), and in snakes, for the Burmese Python (Python molurus bivittatus; Castoe et al., 2013) and King Cobra (Ophiophagus hanah; Vonk et al., 2013). Low-coverage (1-fold coverage) partial genome samples of several other squamate genomes have also been published and analyzed, including the close relative of rattlesnakes, the Copperhead (Agkistrodon contortrix) and the Western Diamond-backed Rattlesnake (Crotalus atrox). Genomic sample-sequencing and analysis of unassembled random genomic “shotgun” sequencing libraries from two snake species (Python molurus bivittatus and Agkistrodon contortrix) determined that while genome size does not vary much across snakes (Figure 2), repeat element relative abundances can vary widely (Figure 5). Most of the differences in repeat content between snake species apparently stem from differences in abundance of many different repeat element types and families, instead of simple expansion or contraction of one or few repeat element groups (Castoe et al., 2011b, 2013).

Two groups of non-LTR retrotransposons, CR1 LINEs and Bov-B LINES, are highly abundant and apparently active in snake genomes. Across major lineages of snakes, the advanced snakes (colubroids, including rattlesnakes) have some of the highest percentages of genomic repeat elements (Castoe et al., 2011b, 2013), with Python having among the lowest. This is interesting because it appears that these major changes in repetitive content have occurred despite a low variance in genome size across snakes (Figure 2). In some cases, the abundance of certain repeat element groups in snake genomes also appears highly variable over relatively short evolutionary distances. For example, A. contortrix and C. atrox have been estimated to have a common ancestor approximately 13.63 million years ago (Reyes-Velasco et al., 2013), yet C. atrox shows notably higher levels of Bov-B LINES and lower levels of L2/C1/Rex LINEs, Gypsy/DIRs1 DNA transposons, and unclassified elements than does A. contortrix (Figure 5).

Transposable elements occasionally contain microsatellite or simple sequence repeat (SSR) on their tails, making them capable of seeding new microsatellite repeat loci in the genome. Snake genomes have emerged as the most extreme example of this phenomenon among vertebrates (Castoe et al., 2011b). Analyses of P. bivittatus and A. contortrix genomes highlight a conspicuous increase in SSR and low-complexity content in snake genomes, indicating a putative increase in genomic SSR evolution and turnover in snakes (Castoe et al., 2011b). More intriguingly, this change must have occurred subsequent to a lull in SSR evolution and turnover earlier in the reptilian lineage (Shedlock et al., 2007). Comparisons

Figure 5. Comparison of the readily-identifiable genome repeat content among squamate reptiles, including the lizard Anolis carolinensis, and the snakes Python molurus bivittatus, Agkistrodon contortrix, and Crotalus atrox. Data based on analysis of complete assembled genomes for Anolis (Alfoldi et al., 2011) and Python (Castoe et al., 2013), and unassembled genomic sequence datasets for A. contortrix (Castoe et al., 2011) and C. atrox (Castoe et al., 2013). Select repeat element families, and broader membership of these in major repeat element types, are shown on the horizontal axis. The vertical axis indicates the proportion of the total genome, or genome sample, comprised by various repetitive element types. Results are based on analysis using Repeat-Marker with custom snake-specific repeat element libraries (Castoe et al., 2011b, 2013).
of the Copperhead (A. contortrix) with the python indicate the rela-
tive abundance of Snakes. C1 LINEs have increased dramatically, a
result mirrored in C. atrox and likely other colubrids (Caoose et al.,
2003). Specifically, in a majority of all SSFs in the Copperhead across
one of closely related sequence motifs (AGA, AGAT, or AGATA),
which represent the microsatellite tails of Snake1 C1 LINEs.

Microsatellites are known to alter genomic recombination struc-
ture and rates, potentially facilitating unequal crossing-over events
and leading to tandem duplication of segments of the genome.
Most venom genes are derived from non-tissue gene families that
experienced gene duplication (Caswell et al., 2012), and the cur-
rent model for the evolution of venom toxins (at least in snakes)
includes tandem duplication of genes (Ikeda et al., 2010). Snake1
C1 LINEs, which seed these microsatellites, have also been dem-
onstrated to occur at high frequency throughout phospholipase ven-
om genes in viperid snakes (Ikeda et al., 2010), in numerous other
venom genes in viperid and elapid snakes (Caoose et al., 2011b),
and in HOM genes in colubrid snakes (Du-Pot et al., 2010). This
provides evidence that transposable elements, and the microsatellites
they seed, may have played a role in the evolution and expansion
of venom loci in snakes. Emerging sequence data from rattlesnakes,
including C. atrox and Crotalus cerastes, will likely provide further evi-
dence necessary to confirm or refute this emerging model of venom
gene and loci evolution. Indeed, given the phylogenetic diversity, widespread
occurrence necessary to confirm or refute this emerging model of venom
gene and loci evolution. Indeed, given the phylogenetic diversity, widespread
occurrence necessary to confirm or refute this emerging model of venom
gene and loci evolution. Indeed, given the phylogenetic diversity, widespread
occurrence necessary to confirm or refute this emerging model of venom
gene and loci evolution. Indeed, given the phylogenetic diversity, widespread
occurrence necessary to confirm or refute this emerging model of venom
gene and loci evolution. Indeed, given the phylogenetic diversity, widespread
occurrence.
order to digest these large prey items adequately, their bodies undergo massive swings in oxidative metabolism, as has been demonstrated in the Sidewinder Rattlesnake, *Crotalus cerastes* (Secor et al., 1994) and Burmese Python, *P. m. bivittatus* (Secor and Diamond, 1998). In addition to exceptional metabolic fluctuations, many organs increase extensively in mass, including the heart, liver, kidneys, and intestine, increasing 35-100% within only a few days after feeding (Secor and Diamond, 1998). Because genomic methods can reveal differences in gene-specific expression patterns that may underlie these tremendous physiological and phenotypic swings, rattlesnakess and other snakes hold great potential for understanding how these intriguing (and medically-relevant) fears may be accomplished in a vertebrate system. With the availability of the *P. m. bivittatus* genome, we can more closely study the regulation and physiological pathways.

Functional constraints of important conserved developmental genes involved in extreme remodeling of these genes in fasted and post-fed states in order to identify protein pathways that are involved in extreme remodeling that comes with the life-history trait of infrequent feeding. Many of the genes in pythons with significant expression level changes upon feeding are homologs to human genes that are associated with development, diseases, and metabolism (Casas et al., 2013). Though collecting thousands of genome-wide SNPs per species. Though collecting thousands of genome-wide markers and enables one to survey the genome for information, which could be used to develop such a large panel of nuclear loci.

In contrast to the scarcity of rattlesnake population genetic research, there have been many studies on the phylogeny and systematics of rattlesnakes. Most such studies have used only a small and select set of mitochondiral and nuclear loci to discern inter- and intra-specific relationships (e.g., Poole et al., 2000; Casas et al., 2007; Bryson et al., 2011a, b). Matrilineal mitochondrial protein-coding loci have been useful markers for understanding phylogenetic relationships, but may underestimate or misrepresent the full scope of population genetic structure for multiple reasons (e.g., they do not capture patterns of male-biased gene flow; Meik et al., 2012). The small number of nuclear loci used to date, however, have proven to be of relatively little utility in resolving rattlesnake relationships (Reyes-Velasco et al., 2013; see Wüster, this volume, Phylogeny). Despite the number of publications focused on rattlesnake systematics, the phylogenetic affinities of several rattlesnake lineages are still unresolved. Incomplete lineage sorting, low phylogenetic signal of chosen loci, and problematic and mislabeled sequences on GenBank (Reyes-Velasco et al., 2015) have contributed to difficulties in resolving rattlesnake relationships. Thus, incorporating a larger number of nuclear genetic markers sampled from throughout the genome would likely greatly clarify questions of rattlesnake relationships and phylogeography. Continued efforts to establish complete genomic and transcriptomic information for multiple lineages of rattlesnakes would provide a wealth of much-needed information, which could be used to develop such a large panel of nuclear loci.

**Population genetics, phylogeography, and systematics**

Knowledge of population genetic structure and demography within a species provides an important context for understanding and making inferences about adaptation, speciation, population ecology, and historical processes which have shaped the evolution of a lineage. Relatively few population genetic studies, however, have been conducted on rattlesnake populations, and those that have been conducted have only focused on two species: the Timber Rattlesnake (*Crotalus horridus*; Villarreal et al., 1996) and the Eastern Massasauga Rattlesnake (*Sistrurus catenatus catenatus*; Gibbs et al., 1997). These studies used microsatellite markers to infer population genetic patterns and structure, but their resolution was limited due to the availability of relatively few microsatellite loci (<5). The detection, amplification, and sequencing of microsatellites has proven to be both expensive and time consuming, but recently-developed technologies that utilize cost- and time-efficient next-generation sequencing to develop thousands of microsatellites for snake species have solved the problem of having few markers to choose from (Casas et al., 2010, 2012, 2013; Oylee-McCance et al., 2013). The development of restriction site-associated DNA (RAD) sequencing further utilizes genome-wide markers and enables one to survey the genome for SNPs (Peterson et al., 2012). A number of groups are currently and successfully using these markers in snakes, including *Crotalus* species. Though sequencing thousands of genome-wide SNPs per individual a powerful method, RADseq analysis is more robust when targeted to a reference genome, thus highlighting the further utility of rattlesnake reference genomes.

**Interaction of genomics and venomology**

The use of high-throughput technologies for genomic, transcriptomic, and proteomic analysis has greatly improved our understanding of rattlesnake venoms, and snake venoms in general (Pahari et al., 2007; Gibbs and Madsen, 2009; Voerk et al., 2013; see Madsen and Casas, this volume, Venom). Addressing questions regarding the evolutionary origins of venom genes will allow a deep understanding of their structure and context within the genome. Unfortunately, what is currently known about rattlesnake venom genes largely lacks genomic context because it is based on transcripts from venom glands. In this regard, only provides information about the transcribed exonic and adjacent untranslated transcribed regions, making it difficult to relate extents of mRNA transcripts directly to venomous nature. Genomic sequence will provide information about intronic regions and venom gene promoters. Promoter sequences for venom genes have demonstrated unique cis-elements that have been proposed to be responsible for the changes in gene expression during gene recruitment in the venom gland; however, this has been shown not to be the case for all venomous snakes and has yet to be explored in rattlesnakes (Toonen et al., 2009; Voerk et al., 2013). Rattlesnake venom gene promoter sequences would provide needed insight into venom gene expression and regulation. Rattlesnake venom gene intronic regions have been shown to be highly conserved, while exonic regions have demonstrated diversification and accelerated rates of evolution (Doley et al., 2008). Even though intron sequences show greater conservation, the number of venom gene introns has been found to differ from their non-toxin protein homologs and could be linked to venom gland recruitment and venom gene duplication events (San et al., 2012).
the absence of well-assembled and annotated genomes. In many forms, different alleles, and different yet similar loci is difficult. For these reasons, the distinction between different splice can make translating information from transcriptomic data dif
utes to the diversity of venom proteins (Siigur et al., 2001) and genomic data sets is that because of the propensity for multi
ple gene duplication in several venom protein families (Heyborne and Mackessy, 2013), important functional differences between
nearly identical gene products can be virtually impossible to dis
and genomic data sources. Including generation of more rattlesnake genomic re
a strong justification for continued investments in these model
resources for rattlesnakes and their close relatives are currently
system, and provides a critical foundation for further research.
imals highlights their impact and significance as a model
questions about venom evolution and variation, gene evolution,
vestibular system, and provides an important framework for fur
ss, and its close relatives are currently minimal, this area is growing rapidly. While there is much for us to
learn about rattlesnake genomes, what is currently known is quite cxisting (e.g., mitochondrial genome structure, sex cons
omes, represent element content and diversity) and provides a strong justifi
ation for continued investments in these model
salus viridis viridis
rat snakes (Casewell et al., 2012).
Venom genes often occur in duplicated tandem arrays, form
large multigene families and multiple protein isoforms (Pa
hart et al., 2007; Ikeda et al., 2010). These are thought to be the result of evolutionary duplication of ancestral non-venom protein coding genes that have been subfunctionalized and or nonfunc
ionalized into venom toxins (Casewell et al., 2012). Gene dupli
ation allows for a selective advantage and flexibility over the op
imization of just a single protein product (Casewell et al., 2012).
Multiple venom genes also contribute to gene dosing effects, which has been observed for the protein conantre in the South
American Rattle	ance (Crotalus durissus) with multiple gene cop
ies correlated with venom conantre concentrations (Ogaiura et al., 2009). Identification of venom polypeptides, genes that have resulted from gene duplication only to become nonfunctional due to mutations, in the rattlesnake genome would provide more support of a birth-and-death model of venom protein evolution (Fry et al., 2003). However, a limitation to both transcriptomic and genomic data sets is that because of the propensity for multi
ple gene duplication in multiple venom protein families (Heyborne and Mackey, 2013), important functional differences between
nearly identical gene products can be virtually impossible to dis
cern.
Alternative splicing of venom gene transcripts further contrib
utes to the diversity of venom proteins (Stiger et al., 2001) and can make translating information from transcriptomic data diff
icult for these reasons, the distinction between different splice forms, different alleles, and different yet similar loci is difficult in the absence of well-assembled and annotated genomes. In many ways, the anticipated availability of well-assembled reference ge
omes for rattlesnakes will provide a valuable context for linking genetic variation with venom variation, and will facilitate accu
rate links between transcripts and translated venom toxins.

Conclusions

Rattlesnake genes are transcriptionally and translationally highly variable model systems, including generation of more rattlesnake genomic re
sources for rattlesnakes will provide a valuable context for linking genetic variation with venom variation, and will facilitate accu
rate links between transcripts and translated venom toxins.

References

celis signal molecule modulates venom on the silicochemistry and intracellular survival of Zigothamnus crowii. Parasitology 138: 46–58.
Bakken, G. S., and A. R. Reichwald. 2007. The imaging properties and sensi
venom proteins in squamate reptiles. Nature Communications 3: 1,046.


