Received: 16 June 2024 | Revised: 14 August 2024 | Accepted: 2 September 2024

DOI: 10.1111/1755-0998.14020

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

VenomCap: An exon-capture probe set for the targeted sequencing of snake venom genes

Scott L. Travers[1] | Carl R. Hutter[2] | Christopher C. Austin[2] | Stephen C. Donnellan[3,4] | Matthew D. Buehler[5] | Christopher E. Ellison[1] | Sara Ruane[6]

[1]Department of Genetics, Rutgers University, Piscataway, New Jersey, USA

[2]Museum of Natural Sciences and Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, USA

[3]South Australian Museum, Adelaide, Australia

[4]Australian Museum Research Institute, Australian Museum, Sydney, Australia

[5]Department of Biological Sciences and Auburn Museum of Natural History, Auburn University, Auburn, Alabama, USA

[6]Life Sciences Section, Negaunee Integrative Research Center, Field Museum, Chicago, Illinois, USA

Correspondence
Scott L. Travers, Department of Genetics, Rutgers University, Piscataway, NJ 08854, USA.
Email: scott.travers@rutgers.edu

Funding information
National Institutes of Health, Grant/Award Number: GM093854; National Science Foundation, Grant/Award Number: DEB-1146033, DEB-1926783, DEB-2224119 and DBI-2010988; National Geographic Society, Grant/Award Number: NGS-53506R-18; Coypu Foundation

Handling Editor: Andrew DeWoody

Abstract

Snake venoms are complex mixtures of toxic proteins that hold significant medical, pharmacological and evolutionary interest. To better understand the genetic diversity underlying snake venoms, we developed VenomCap, a novel exon-capture probe set targeting toxin-coding genes from a wide range of elapid snakes, with a particular focus on the ecologically diverse and medically important subfamily Hydrophiinae. We tested the capture success of VenomCap across 24 species, representing all major elapid lineages. We included snake phylogenomic probes in the VenomCap capture set, allowing us to compare capture performance between venom and phylogenomic loci and to infer elapid phylogenetic relationships. We demonstrated VenomCap's ability to recover exons from ~1500 target markers, representing a total of 24 known venom gene families, which includes the dominant gene families found in elapid venoms. We find that VenomCap's capture results are robust across all elapids sampled, and especially among hydrophiines, with respect to measures of target capture success (target loci matched, sensitivity, specificity and missing data). As a cost-effective and efficient alternative to full genome sequencing, VenomCap can dramatically accelerate the sequencing and analysis of venom gene families. Overall, our tool offers a model for genomic studies on snake venom gene diversity and evolution that can be expanded for comprehensive comparisons across the other families of venomous snakes.

KEYWORDS
animal venoms, Elapidae, Hydrophiinae, phylogenomics, snake genomics, venomics

1 | INTRODUCTION

Snake venoms are of broad interest to many fields of biology and human health. Medically, snakebite envenomation is a neglected human health crisis that annually contributes to over 100,000 human deaths and potentially half a million additional cases of venom-induced morbidity worldwide (Gutiérrez et al., 2017). Pharmacologically, snake venoms are a rich source of bioactive proteins, presenting further immense potential for drug discovery (Mohamed Abd El-Aziz et al., 2019). Evolutionarily and ecologically,

snake venoms offer a model system to study the origins of biochemical novelties, the molecular basis of adaptation and other facets of the relationship between genotype, phenotype, ecology and macroevolution (Almeida et al., 2021; Dowell et al., 2016; Holding et al., 2016; Holding et al., 2021; Mason et al., 2022).

Venoms have evolved independently in many organisms (e.g. bees, fish, mammals, reptiles, snails, spiders). Although snake venoms are among the better-characterized animal venoms due to their abundance and medical relevance, there are still major gaps in our understanding of their diversity, evolution and function. Venomous snakes inhabit every continent except Antarctica, as well as marine habitats in the Indian and Pacific oceans. The majority of medically significant species belong to the families Elapidae (e.g. cobras, mambas, kraits, coral snakes, Australo-Papuan elapids, sea snakes) and Viperidae (e.g. true vipers and pit vipers), which are diverse and globally distributed (405 elapid and 396 viperid species; Uetz, 2024). Both families possess efficient venom delivery systems, including fangs with hollow venom canals at the front of their mouths, ducts connecting the fangs to paired venom glands where venom is produced and stored, and associated venom delivery musculature (Mackessy, 2022). While elapid and viperid venoms are the most well studied, venoms are present across advanced snakes in the superfamily Colubroidea. Because snake venoms are used for both feeding and defence, they are subject to strong selection pressures, driving the evolution of a remarkable diversity of venom components and venom systems that we have yet to fully understand (Casewell et al., 2020; Tasoulis & Isbister, 2023).

Snake venoms are complex mixtures primarily comprising toxic proteins and peptides encoded by a variety of gene families. Over 60 protein families have been identified in snake venoms (Tasoulis & Isbister, 2017). Despite the complexity and importance of these venoms, our understanding of their genetic underpinnings remains incomplete. Estimates suggest that more than 20,000 different toxins may be found just among elapid and viperid venoms, yet only a few thousand have been characterized (Laustsen et al., 2016). Researchers have begun to utilize genomics-era technologies to better examine the genetic basis of snake venoms, which is revealing the diversity of toxin-coding genes, their evolution and their functions (Rao et al., 2022). However, studies of this nature are still limited.

Understanding the genetic basis of snake venom diversity remains challenging. While numerous studies have examined snake venoms through transcriptome sequencing of venom glands (Brahma et al., 2015; Modahl et al., 2020), these data often represent only a subset of the toxin-coding genes present in a snake's genome (Margres et al., 2021). For example, venom composition can vary within an individual or species due to a variety of factors, such as age or geographic location (Hogan et al., 2024; Smith et al., 2023). Defining venom toxins solely from venom gland transcriptomes is also difficult without comparative data from other tissue types (Reyes-Velasco et al., 2015). Further, transcriptomic research on snake venoms is limited by the availability of fresh, venom gland-derived RNA samples, which are hard to obtain due to species rarity, remote field collection sites, and the dangers involved in sampling. Consequently, these samples are not readily available for most venomous snake species.

Alternatively, genome-based sequencing approaches can uncover the full spectrum of toxin-coding genes in a snake's genome using non-venom gland derived tissue (e.g. Li et al., 2021; Margres et al., 2021). Genetic material from many venomous snake species is available from research collections for such studies. However, obtaining high-quality, full genome data remains relatively expensive and time-consuming to analyse, and thus, such data are currently available for only a few species of snakes (Pinto et al., 2023; Rao et al., 2022; Roberts et al., 2024; Tan, 2022).

Exon-capture is a targeted genome sequencing method that can significantly enhance our ability to uncover the genetic diversity underlying snake venoms. This approach uses probes (or 'baits') developed from existing genomic resources to selectively target and sequence specific loci from a wide range of species, offering a more cost-effective and faster alternative to full genome sequencing (Gnirke et al., 2009). Although various probe-based targeted sequencing methods have been developed for evolutionary genomic research on snakes, most have been designed primarily for phylogenomic purposes (e.g. Karin et al., 2020; Pyron et al., 2014; Singhal et al., 2017; Streicher & Wiens, 2016). Exon-capture techniques have been successfully employed to study the diversity of toxin-coding genes in other venomous animals, such as cone snails (Phuong & Mahardika, 2018). However, these approaches have not been widely applied to snake venoms; to our knowledge, only a single study has applied this approach to snakes, but the focus was on a single venom gene family from one pit viper species (Margres et al., 2017).

Here, we develop a novel exon-capture probe set designed to target toxin-coding genes from a diverse lineage of venomous snakes. This probe set aims to serve as a new sequencing toolkit for uncovering venom gene diversity, thereby enhancing our understanding of the genetic basis of snake venoms. Our design focuses on the elapid snake subfamily Hydrophiinae, which includes over 200 species of Australian and Melanesian terrestrial elapids (e.g. death adders, taipans, tiger snakes) and 'true' sea snakes, of which many are of medical importance and known for producing some of the deadliest toxins in the world (Sutherland & Tibbals, 2001). We focus on hydrophiines for several reasons: (1) hydrophiines are species-rich and ecologically diverse, representing over half of all elapid diversity (207 of 405 total elapid species; Uetz, 2024); (2) their venoms hold significant medical and pharmacological interest due to their implications in snakebite treatment and drug discovery (Earl et al., 2012; Johnston et al., 2017); (3) existing genomic and/or transcriptomic resources for a variety of hydrophiine species are available, which are essential for developing a comprehensive venom gene probe set; (4) the full genomic underpinnings of hydrophiine venom has been comprehensively analysed in only a few species (Li et al., 2021; Peng et al., 2020), and the venoms of many remain entirely unstudied. Although this probe set is developed with a focus on hydrophiines, we also test its capture success across a broad phylogenetic range of elapid lineages. We demonstrate the utility of this probe set for studying venom genes within the entire elapid family and discuss its

potential expansion to include other families of venomous snakes using the methodologies detailed here.

2 | MATERIALS AND METHODS

2.1 | Marker selection and probe design

To target genes that prior studies have shown to code for toxins in hydrophiine venoms, we used the following genome- and transcriptome-based approaches, detailed more thoroughly in the following paragraphs: (1) we collected all known hydrophiine venom toxin markers available on GenBank from traditional Sanger sequencing studies; (2) we used the venom gland transcriptome annotator, ToxCodAn, to predict toxins from hydrophiine venom gland transcriptomes following the published pipeline in Nachtigall et al. (2021); (3) we used the toxin protein sequence database from ToxCodAn to extract toxin markers from available hydrophiine genomes; and (4) we used the toxin gene annotations from a detailed study on sea snake genomes (Li et al., 2021) as an additional database to extract toxin markers from other genomes. We used this multifaceted approach to create a comprehensive database of hydrophiine venom toxin genes (and venom-related genes) to base our probe set on, by utilizing as many hydrophiine venom gland transcriptomes, full genomes and traditionally generated venom gene data as possible. We conducted all data processing and other analyses in R using customized scripts with the following R packages: GENOMICRANGES (Lawrence et al., 2013), SEQINR (Charif & Lobry, 2007) and APE (v5.0; Paradis & Schliep, 2019).

First, we collected all known hydrophiine toxin markers from GenBank derived from venom gene sequencing studies that utilized traditional cDNA cloning and Sanger sequencing techniques (e.g. Chatrath et al., 2011; Doley et al., 2008; Earl et al., 2006; Gong et al., 2000; Paiva et al., 2014; Richards et al., 2011; St Pierre et al., 2006; St Pierre et al., 2008; St Pierre, Birrell, et al., 2007; St Pierre, Fischer, et al., 2007; St Pierre, Masci, et al., 2005; St Pierre, Woods, et al., 2005; Welton & Burnell, 2005). All nucleotide sequences for each hydrophiine genus that matched the search terms, tissue type = 'venom gland' and molecule type = 'mRNA', were extracted and clustered together into groups of sequence identity 80% or higher using CD-HIT-EST (Li & Godzik, 2006). We then aligned the sequences on a marker-by-marker basis with the program MAFFT v7.312 (auto and default parameters; Katoh & Standley, 2013). Finally, we generated consensus sequences of these alignments to create a final set of markers to design the probes from.

For transcriptome-based venom marker selection, we used the ToxCodAn pipeline (Nachtigall et al., 2021) to predict toxins from assembled hydrophiine venom gland transcriptomes that were available on GenBank's Transcriptome Shotgun Assembly (TSA) database at the time (Table 1). This included venom gland transcriptomes from the following 12 species: *Acanthophis wellsi*, *Brachyurophis roperi*, *Cacophis squamulosus*, *Denisonia devisi*, *Echiopsis curta*, *Furina ornata*, *Hemiaspis signata*, *Hoplocephalus bungaroides*, *Pseudonaja modesta*, *P. textilis*, *Suta fasciata* and *Vermicella annulata*. The first eleven of these transcriptome assemblies were based on a single study, and from which only the assembled toxin transcripts they predicted were uploaded to the GenBank TSA database (Jackson et al., 2013). We processed these transcripts through the ToxCodAn pipeline to confirm toxin annotations, and toxin sequences from each species were retained. For *P. textilis*, the full transcriptome assembly was available on GenBank, which we processed through ToxCodAn, and retained sequences identified as toxins, putative toxins and uncharacterized transcripts (i.e. transcripts that did not correspond to any known proteins in the package's toxin or non-toxin databases, and retained for novel toxin discovery). Using the predictions from the transcriptomes, we aligned from each sample the same predicted toxins using MAFFT. We filtered out any taxa that were 40% or more divergent from the consensus sequence of the alignment, as these were likely misalignments or incorrectly assigned toxins.

TABLE 1 Genomic resources used for venom exon selection.

Species	Genomic resource type	NCBI accession
Emydocephalus ijimae	Genome	GCA_004319985.1
Hydrophis curtus	Genome	GCA_019472885.1
Hydrophis cyanocinctus	Genome	GCA_019473425.1
Hydrophis hardwickii	Genome	GCA_004023765.1
Hydrophis melanocephalus	Genome	GCA_004320005.1
Laticauda laticaudata	Genome	GCA_004320025.1
Notechis scutatus	Genome	GCA_900518725.1
Pseudonaja textilis	Genome	GCA_900518735.1
Acanthophis wellsi	Venom gland transcriptome	PRJNA189680
Brachyurophis roperi	Venom gland transcriptome	PRJNA189681
Cacophis squamulosus	Venom gland transcriptome	PRJNA189682
Denisonia devisi	Venom gland transcriptome	PRJNA189683
Echiopsis curta	Venom gland transcriptome	PRJNA189684
Furina ornata	Venom gland transcriptome	PRJNA189685
Hemiaspis signata	Venom gland transcriptome	PRJNA189686
Hoplocephalus bungaroides	Venom gland transcriptome	PRJNA189687
Pseudonaja modesta	Venom gland transcriptome	PRJNA189688
Pseudonaja textilis	Venom gland transcriptome	PRJNA274608
Suta fasciata	Venom gland transcriptome	PRJNA189689
Vermicella annulata	Venom gland transcriptome	PRJNA189690

Finally, we created a consensus sequence from each alignment to be used later for probe design.

For genome-based venom marker selection, we collected hydrophiine genomes with the goal of extracting venom toxin genes (and venom related genes) using a snake toxin protein sequence database. We downloaded all hydrophiine genomes that were available on GenBank at the time, with the addition of one *Laticauda* genome, the closely related, sister lineage to all hydrophiines (Table 1). In total, this included the following eight genome assemblies: *Emydocephalus ijimae*, *Hydrophis curtus*, *H. cyanocinctus*, *H. hardwickii*, *H. melanocephalus*, *L. laticaudata*, *Notechis scutatus* and *P. textilis*. To identify venom toxin genes in each of these genomes, we used the snake toxin protein sequence database from ToxCodAn. This database was developed from curated toxin protein sequences from Uniprot (https://www.uniprot.org/), and protein sequences from published and unpublished transcriptome assemblies from a wide range of venomous snake clades (Viperidae, Elapidae, Colubridae and Dipsadidae; Nachtigall et al., 2021). Using the database of proteins, we used BLAST to match the proteins to our database of genomes with the '-tblastn -task tblastn-fast' function for computational tractability. We filtered the BLAST results to keep matches that were 60% or greater percent identity, and 12 or more amino acid matches (36 nucleotides), and ensured each protein matched in length by at least 35% or greater to each genome target. Because the toxin database contained amino acid sequences for just the coding region of each gene (i.e. without introns), exons were identified by large gaps between separate BLAST matches for the same gene. We extracted the final matches from the genomes and created alignments with MAFFT. Finally, we generated consensus sequences to create a final set of markers that would be used for downstream probe design. We note that these markers have abundant paralogues, which is expected of toxin genes, and we did not distinguish them here as we desired to capture these paralogues and would sort through them bioinformatically after sequencing.

Finally, we used the venom-associated gene annotations (venom toxin genes and venom related genes) from a detailed genome study on two sea snake species (*Hydrophis curtus* and *H. cyanocinctus*; Li et al., 2021), which used a separate toxin protein database that they developed from the Swiss-Prot, InterPro and Pfam databases, as well as the manually reviewed sequences from the animal toxin annotation project (http://www.uniprot.org/program/Toxins). We first used the annotations from this study to extract the exons from the two genomes (*H. curtus* and *H. cyanocinctus*). Because the annotated gene/exon names were not consistent between species, to find the homologous sequences, we extracted the exons from one genome and used BLAST on the other genome ('with the -blastn -task dc-megablast' function). We filtered the BLAST results to keep matches that were 60% or greater percent identity, and 50 or more nucleotide matches, and ensured each protein matched in length by at least 35% or greater to each transcriptome or genome target. We kept the best remaining BLAST match for each exon, choosing the match with the highest bitscore, a metric that combines length and similarity of match, so the most similar and longest matches will be selected. We next used the genomic match coordinates and BEDTOOLS (Quinlan & Hall, 2010) *intersect* command to find overlapping regions between the opposing genome matches and their already existing toxin gene annotations (e.g. *H. curtus* matched to *H. cyanocinctus*, intersect *H. curtus* matches on *H. cyanocinctus* with *H. cyanocinctus* annotations) to ensure our exon matches fall within the gene annotations. We kept matches that overlapped with the toxin annotations and kept the unique matches between the two genomes. Finally, we again used BLAST ('with the -blastn -task dc-megablast' function) to match these sea snake venom-associated genes to the *P. textilis* genome and kept matches to be used for downstream probe design using the *P. textilis* sequence.

After we identified candidate toxin genes and exons from these different sources, we wanted to ensure that we did not have duplicates of the same genomic regions. To remove these duplicates, we concatenated all the final venom marker sets together and used CD-HIT-EST to cluster the markers and remove duplicates at a 95% similarity threshold. After these initial filtration steps of the matches, a final candidate venom marker set of 1975 exons remained (Figure 1a; Table 2).

We used the final sets of consensus sequences to design a myBaits custom bait library (Arbor Biosciences, Ann Arbor, MI), using 120mer baits to best capture sequences with greater than 5% divergence from the probes. Target exons <120 bp were expanded using flanking intron sequence from the *P. textilis* genome to match the bait length. We separated the target markers into individual 120 bp baits using 4X tiling which resulted in 17,181 baits. To increase capture success, we then filtered these baits by keeping only those with: (1) no repetitive sequences based on the RepeatMasker software package (http://www.repeatmasker.org); (2) no more than ten matches to genomic regions in the *P. textilis* genome; (3) at most, ten matches to genomic regions in the *P. textilis* genome with a calculated melting temperature of 62.5–65°C and four matches with a melting temperature above 65°C, and fewer than two passing baits on each flank; and (4) no matches with BLAST to other baits (using a 70% similarity criterion). After filtration, 16,107 baits remained.

To target additional markers for phylogenetic purposes, we combined our venom marker baits with baits from another sequence capture probe set that was specifically designed for snake phylogenomics. Full details on marker selection, probe design and capture success for this snake phylogenomic probe set is discussed elsewhere (Weinell et al., 2024). Briefly, this probe set targeted 3129 single-copy loci comprising three marker types: (1) 1894 exons, (2) 907 ultraconserved elements (UCEs) and (3) 328 double-digest RADseq-like loci (ddRAD-like). The exons were selected by matching available snake transcriptomes from a wide phylogenetic spectrum to the *Thamnophis sirtalis* genome to find shared orthologous, single-copy exons. The UCEs were selected from a subset of previously identified UCEs shared among a wide range of snake lineages (Streicher & Wiens, 2016). The ddRAD-like markers were designed from the *Thermophis baileyi* genome using an in silico

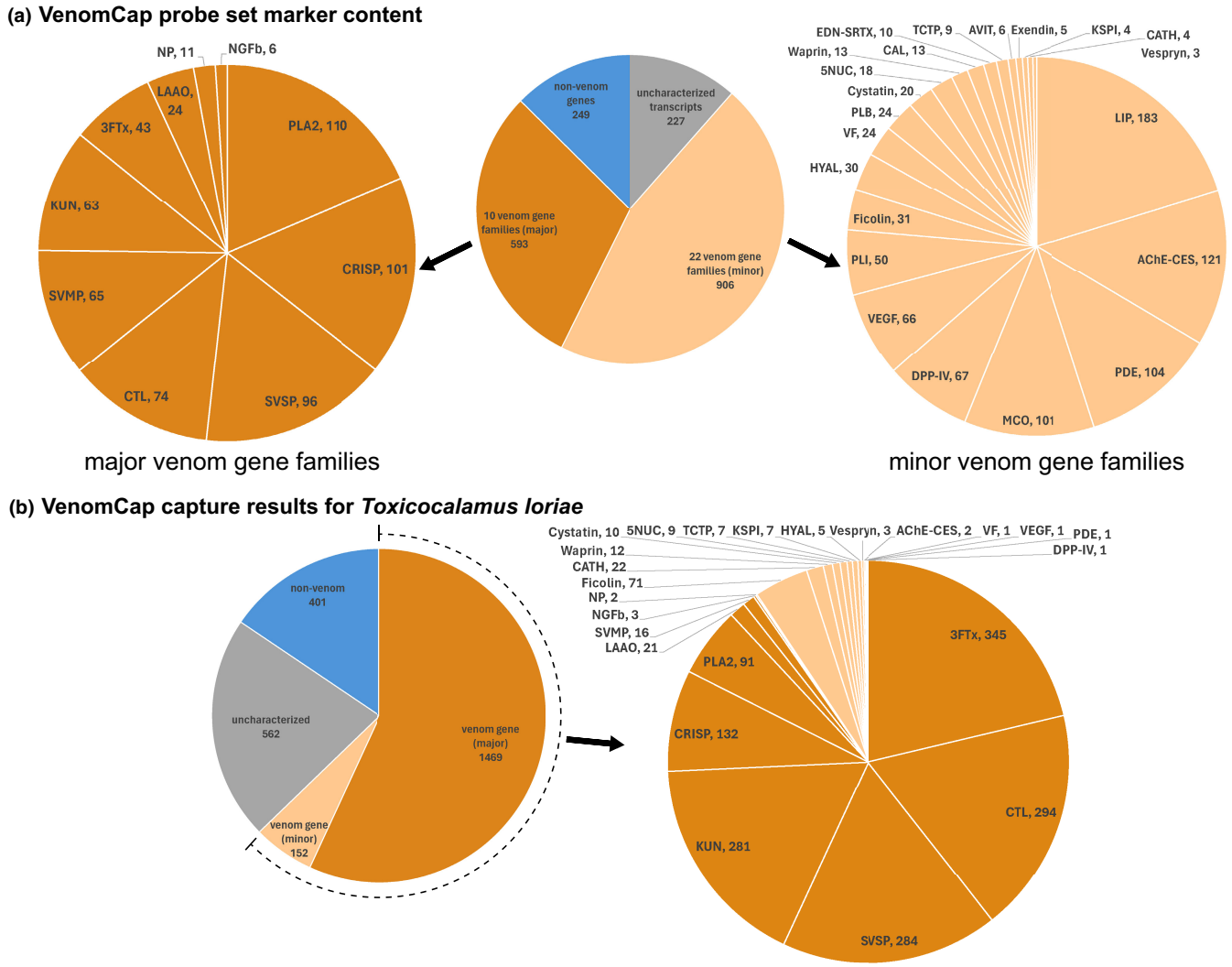

FIGURE 1 Graphical overview of the VenomCap probe set content and capture results for one species. (a) VenomCap target marker content broken down by marker type (middle), and the major (left) and minor (right) components of elapid venoms for the venom-related gene families included in the probe set. Number of exons targeted indicated next to each marker category or gene family abbreviation; 1975 exons targeted in total. (b) VenomCap capture results for one of the representative species we sequenced (*Toxicocalamus loriae*) lacking prior venom information. The number of exons captured is indicated next to each marker category or gene family abbreviation.

ddRAD approach to identify restriction-enzyme recognition sites for *SbfI* and *EcoRI* restriction enzymes, to select single-copy loci. In total, 20,020 baits were developed to target these 3129 phylogenomic markers. We combined the 20,020 phylogenomic marker baits with the newly designed 16,107 baits targeting the 1975 venom markers, which totaled 36,127 baits for 5104 markers.

2.2 | Sequencing

2.2.1 | Sampling

We sequenced 24 elapid species in total, which included 15 hydrophiine species (ingroup) and 9 non-hydrophiine elapid species (outgroup), representing a wide phylogenetic breadth of lineages from across the family. See Table 3 for a detailed list of samples.

2.2.2 | Library preparation and sequencing

Genomic DNA was extracted from liver or muscle tissue samples using a Qiagen® DNeasy Blood and Tissue Kit. The resultant DNA was quantified using a Qubit® fluorometer, and approximately 1000 ng total DNA per sample was shipped to Daicel Arbor Biosciences for library preparation, capture and sequencing. Prior to library preparation, the genomic DNA samples were quantified with fluorescence and up to 4 μg was then taken to sonication with a QSonica® Q800R instrument. After sonication and SPRI bead-based size-selection to modal lengths of roughly 300 bp, up to 500 ng of each sheared DNA sample were taken to Illumina Truseq-style sticky-end library preparation. Following adapter ligation and fill-in, each library was amplified for 6 cycles using unique combinations of i7 and i5 indexing primers, and then quantified with fluorescence. For each capture reaction, 125 ng of 8

TABLE 2 Venom-related gene families targeted by the VenomCap probe set. Categorizing the major components of elapid venoms follows Tasoulis and Isbister (2017).

Gene family	Abbreviation	Exons targeted	Component of elapid venom
Three-finger toxin	3FTx	43	Major
Phospholipase A_2	PLA2	110	Major
Snake venom metalloproteinase	SVMP	65	Major
Snake venom serine protease	SVSP	96	Major
Cysteine-rich secretory protein	CRISP	101	Major
Flavin monoamine oxidase (L-amino acid oxidase)	LAAO	24	Major
Kunitz peptide	KUN	63	Major
C-type lectin	CTL	74	Major
Natriuretic peptide	NP	11	Major
β-nerve growth factor	NGFb	6	Major
Venom factor (venom complement C3 homologue)	VF	24	Minor
Platelet-derived growth factor/vascular endothelial growth factor	VEGF	66	Minor
Acetylcholinesterase/ type-B carboxylesterase	AChE	121	Minor
5′-nucleotidase	5NUC	18	Minor
Glycosyl hydrolase 56/ hyaluronidase (venom spreading factor)	HYAL	30	Minor
Nucleotide pyrophosphatase/ phosphodiesterase	PDE	104	Minor
Phospholipase B-like	PLB	24	Minor
Vespryn/ohanin	VESP	3	Minor
Cystatin	CYST	20	Minor
Translationally controlled tumour protein	TCTP	9	Minor
Snake waprin	WAP	13	Minor
Phospholipase A_2 inhibitor	PLI	50	Minor
Kazal-type serine protease inhibitor	KSPI	4	Minor
AVIT (prokineticin)	AVIT	6	Minor
Ficolin lectin/veficolin	FIC	31	Minor
AB hydrolase superfamily/ lipase family (lysosomal acid lipase and phospholipase A_1)	LIP	183	Minor
Multicopper oxidase family	MCO	101	Minor
Peptidase S9B family/ dipeptidyl peptidase IV subfamily	DPP-IV	67	Minor
Endothelin/sarafotoxin	EDN-SRTX	10	Minor
Glucagon (exendin)	Exendin	5	Minor
Cathelicidin	CATH	4	Minor
Calmodulin family/ calglandulin subfamily	CAL	13	Minor
non-venom genes		249	
uncharacterized transcripts		227	

libraries were pooled, and subsequently enriched for targets using the myBaits® v 3.1 protocol. Enrichment incubation times ranged from 18 to 21 h. Following enrichment, library pools were amplified for 10 cycles using universal primers and subsequently pooled in equimolar amounts for sequencing. Samples were sequenced on an Illumina HiSeq 3000 with 150 bp paired-end reads with 96 samples sequenced per lane of sequencing.

2.3 | Data processing

2.3.1 | Read processing

We used the R package PhyloProcessR (Hutter et al., 2022; https://github.com/chutter/PhyloProcessR) for filtering adapter contamination, assembling contigs and exporting alignments. Prior to processing raw reads, Illumina sequence data were de-multiplexed using the software bcl2fastq (Illumina Corp). Next, raw reads were cleaned of adapter contamination, low-complexity sequences and other sequencing artefacts using the program FASTP (modified settings: length_required = 30, complexity_threshold = 30; Chen et al., 2018). Adapter-cleaned reads were then matched to a database of common contaminants (bacterial, human skin, ultra-pure water contamination and other common bacteria; Laurence et al., 2014) and other genomes (*Caenorhabditis*, *Drosophila*) to ensure that no contamination persisted in our final dataset (for GenBank Accession Numbers of reference genomes, see tab S3 in Hutter et al., 2022). We decontaminated the adapter-cleaned reads with the program BWA (bwa-mem, default settings; Li & Durbin, 2009) by matching cleaned reads to each reference contaminant genome (reads removed if they matched >95% similarity). Next, we normalized reads using the program ORNA (Durai & Schulz, 2019). Prior to assembly, we merged paired-end reads using FASTP.

2.3.2 | Assembly

Merged singletons and paired-end reads were assembled de novo using the program SPADES v.3.15 (Bankevich et al., 2012), which runs BAYESHAMMER (Nikolenko et al., 2013) error correction on the reads internally. Data were assembled using several different

TABLE 3 Sample information and sequence capture statistics.

Species	Sample ID	Cleaned reads	Gigabases sequenced	Contigs assembled	Megabases assembled	Phylogenomic loci recovered	VenomCap targets matched	VenomCap contigs recovered
Acanthophis laevis	LSUMZ 98537	12,638,820	1.89	45,191	29.18	2806	1500	1860
Aipysurus duboisii	ABTC 72288	10,715,594	1.60	36,153	21.62	2765	1498	1611
Aspidomorphus muelleri	LSUMZ 91868	14,581,580	2.18	45,943	29.36	2737	1505	2688
Brachyurophis incinctus	ABTC 102798	8,718,004	1.31	34,156	20.85	2734	1499	2278
*Bungarus fasciatus**	LSUMZ 4845	9,615,564	1.44	22,830	15.11	2660	1463	2218
*Bungarus multicinctus**	LSUMZ 15945	11,095,224	1.66	35,891	23.95	2791	1500	2679
Cacophis churchilli	ABTC 82294	8,919,574	1.34	27,250	17.07	2703	1487	1820
*Calliophis bivirgatus**	LSUMZ 6522	11,206,454	1.68	38,508	25.31	2748	1490	2714
Demansia rimicola	ABTC 77027	12,779,464	1.92	46,774	29.98	2728	1513	2666
Demansia vestigiata	LSUMZ 98522	12,334,480	1.85	40,804	25.68	2716	1492	2496
*Dendroaspis polylepis**	LSUMZ 2734	10,012,962	1.50	30,051	20.26	2754	1459	2215
*Elapsoidea nigra**	LSUMZ 2381	9,390,080	1.41	30,334	20.17	2748	1483	1969
*Elapsoidea sundevallii**	LSUMZ 6687	10,466,572	1.57	32,206	21.02	2749	1510	2826
Hemiaspis signata	ABTC 6591	9,375,812	1.41	28,589	20.80	2724	1501	2066
Hydrophis macdowelli	ABTC 101326	11,992,898	1.79	44,404	28.04	2738	1493	1739
Laticauda cf. *colubrina**	CAS 236329	11,883,314	1.78	36,628	23.65	2750	1479	2016
*Micruroides euryxanthus**	LSUMZ 14737	10,775,862	1.62	30,542	19.93	2730	1484	2545
*Naja nigricollis**	LSUMZ 20269	8,577,756	1.29	27,340	19.30	2745	1462	2272
Notechis scutatus	MCZ 14229	9,420,308	1.41	26,569	17.33	2713	1503	2395
Oxyuranus scutellatus	LSUMZ 94345	12,838,540	1.92	41,460	27.17	2790	1493	1655
Pseudechis australis	MCZ R 195407	9,004,684	1.35	36,978	23.18	2738	1504	2652
Pseudonaja textilis	LSUMZ 90637	10,690,176	1.60	33,711	23.99	2764	1501	1720
Simoselaps anomalus	MCZ 14397	13,157,814	1.97	49,350	32.00	2776	1505	2046
Toxicocalamus loriae	CCA 17794	12,813,848	1.92	44,646	29.88	2783	1487	2584

Note: 'VenomCap targets matched' corresponds to the number of unique targets in the VenomCap probe set with at least one contig match (out of 1975 total target exons). 'VenomCap contigs recovered' corresponds to the total number of contigs that match VenomCap targets, which accounts for duplicate matches due to paralogues. Outgroup elapids denoted with asterisks next to species name, all others Hydrophiinae.

Abbreviations: ABTC, Australian Biological Tissue Collection; CAS, California Academy of Sciences; CCA, Chris Austin field catalogue (Louisiana State University); LSUMZ, Louisiana State University Museum of Natural Science; MCZ, Museum of Comparative Zoology.

k-mer values (21, 33, 55, 77, 99, 127), in which orthologous contigs resulting from the different k-mer assemblies were merged.

2.3.3 | Annotation of phylogenetic markers

The assembled contigs were matched against the reference marker sequences used to design the probes with BLAST (*dc-megablast*). Contigs were discarded if they failed to match ≥30% of the reference marker, and contig matches fewer than 50 bp were removed. Contig matches to reference markers were discarded if more than one contig matched to the marker and were overlapping. For non-overlapping matches to the same reference marker, we merged these contigs by joining them together (Ns inserted in matching positions). The final set of matching contigs was labelled with the name of each marker, followed by each sample's unique institutional identifier (i.e. the corresponding museum voucher catalogue number), and assembled in a single file to be parsed out separately for multiple sequence alignment in the next step.

2.3.4 | Annotation of VenomCap markers

To annotate VenomCap markers, which consist of paralogues, modifications were made to the annotation script above (option: retain-paralogues = TRUE). This option retains markers from samples that had duplicate matches to the VenomCap target marker set, to preserve the paralogues. Next, the longest match was retained for alignments for summary statistic collection. To collect summary statistics for gene families, the R function from PhyloProcessR 'geneFamilyStats' was used, where an input file with the gene family in the first column and the markers from the VenomCap set of markers

in the second column was created to associate the markers with their gene family. To create a match database, the VenomCap markers were BLASTed against each sample contigs. For each gene family, the number of copies was estimated by filtering out matches that were less than 50 bp long, less than 80% of the length of the marker, and had less than 75% match similarity. Markers were also discarded if they had more than 100 matches, as these were likely regions with substantial off-target capture. Finally, results were summarized on a per sample basis (number of gene families, mean number of copies, total targets matched, percent of total targets and base pair length summaries) and a per marker basis (unique matches to VenomCap markers ignoring duplicates, matches that include duplicates and base pair length summaries).

2.3.5 | Alignments for phylogenomics

For phylogenetic rooting purposes, we included data from an outgroup colubrid species (*Boiga irregularis*) that was sequenced for another project using just the phylogenomic probe set and processed the same as above. Next, the annotated phylogenomic markers were aligned on a marker-by-marker basis using MAFFT local pair alignment (settings: max iterations = 1000; ep = 0.123; op = 3; –adjustdirection). We screened each alignment for samples ≥40% divergent from consensus sequences, which were almost always incorrectly assigned contigs. Alignments were retained if they included 19 or more taxa (>75% of samples), had ≥500 bp length, and each sample covered the alignment by at least 35% of the bp. Alignments were externally trimmed resulting in alignments in which at least 50% of the samples have sequence data at both alignment ends. The alignments were next filtered by removing samples that consisted of 50% gaps or more and had less than 60 bp of data. Columns with 50% gaps or more were also removed. Finally, the gene datasets were internally trimmed using the program trimAl (automatic1 function; Capella-Gutierrez et al., 2009).

2.4 | Capture evaluation

2.4.1 | Sensitivity

Sample 'sensitivity' represents the proportion of bases in post-assembly contigs that overlap with the target markers in the probe set for a given sample. To determine sample sensitivity, we used BLAST to align the target markers for each probe set (VenomCap and phylogenetic markers) with contigs derived from the respective sample. Sensitivity was then calculated on a per-target-marker and per-sample basis by dividing the length of the target marker sequence by the extent of matching region within the sample contig, and then averaging across all markers for a sample. We used Welch's t-tests to compare sensitivity between: (1) the ingroup and outgroup samples for the VenomCap data only and (2) VenomCap versus phylogenetic probe sets.

2.4.2 | Specificity

Sample 'specificity' represents the proportion of cleaned reads that map to the target markers in the probe set for a given sample. To determine sample specificity, we generated an indexed reference based on the target markers for each probe set (VenomCap and phylogenetic markers) and aligned the processed reads from individual samples using BWAv0.712 (utilizing the functions bwa index and bwa mem; Li & Durbin, 2009). SAMTOOLS (Li et al., 2009) was used to facilitate conversion between different file formats (via functions view and fastq) and to quantify the count of reference-aligned processed reads, enabling specificity calculations. We used Welch's *t*-tests to compare specificity between (1) the ingroup and outgroup samples for the VenomCap data only and (2) the two probe sets.

2.4.3 | Missing data

We used two approaches to evaluate the extent of missing data in our capture results: (1) We assessed missing marker data by quantifying the percentage of markers absent for a sample across all alignments of target markers captured; and (2) We assessed missing base pair data by quantifying the percentage of base pairs absent for a sample (excluding indels) in the alignments trimmed to the target marker length and used the mean value of all the markers for a given sample. We used Welch's *t*-tests to compare missing data between: (1) the ingroup and outgroup samples for the VenomCap data only and (2) the two probe sets.

2.4.4 | Depth of coverage

To calculate sequencing depth of coverage for each sample, we created a reference for the set of assembled contigs targeted by a probe set and mapped the cleaned reads to them using BWA ('bwa-mem' function). Exact duplicate reads were removed using PICARD (http://broadinstitute.github.io/picard/). We then computed the per-base overlap of reads with contig base pairs using SAMTOOLS ('depth' function). We standardized coverage values across samples by using reads per kilobase per million (RPKM), enabling comparisons among samples with varying sequencing efforts by scaling based on gene length and mapped reads. Coverage was calculated across all targeted markers for each sample, which exhibited positive skewness so median values were calculated. We used a Welch's *t*-test to compare coverage between both probe sets (VenomCap and phylogenetic).

2.4.5 | Genetic distance

To evaluate whether capture success was impacted by genetic distance from the probe set, we performed linear regressions between the measures of capture success (target loci matched, sensitivity,

specificity and missing data) and genetic distance. Genetic distance for each sample was calculated using uncorrected pairwise distance between a captured marker and the corresponding marker sequence the probes were created from, and the mean was then computed across markers for each sample. We performed two separate linear regressions for each metric of capture success to evaluate our results among: (1) all elapid samples and (2) the ingroup only (Hydrophiinae).

2.5 | Phylogenetics

Using just the phylogenomic markers, we concatenated the gene alignments described above into a single alignment for maximum likelihood (ML) phylogenetic analyses. We used IQ-Tree v.2.0 (Minh et al., 2020) to estimate phylogenetic trees from the concatenated alignment. We used models of molecular evolution identified via ModelFinder (Kalyaanamoorthy et al., 2017) built into IQ-Tree, which identified an optimal partitioning scheme and best model for each partition. We assessed support for the resulting topology using 1000 ultrafast bootstrap replicates; values above 95% were considered well supported (Minh et al., 2013).

3 | RESULTS

3.1 | Read and assembly statistics

Across the 24 genomic samples that were enriched and sequenced for the combined VenomCap and snake phylogenomic sequence capture probe sets, we recovered an average of 10,958,558 cleaned reads (range: 8,577,756–14,581,580 cleaned reads, Table 3) containing a mean of 1.64 Gb (Gigabases) sequenced in total (range: 1.29–2.18 Gb, Table 3) per sample. Assemblies resulted in an average of 36,096 contigs (range: 22,830–49,350 contigs, Table 3), with an average of 23.54 Mb (Megabases) assembled in total (range: 15.11–32 Mb, Table 3) per sample.

3.2 | Capture evaluation

3.2.1 | Target loci recovered

For the VenomCap probe set, we assessed the number of targets matched (number of unique targets in the probe set with at least one contig match) and total contigs recovered (total number of contigs that match VenomCap targets, which accounts for duplicate matches due to paralogues). Assembled contigs matched an average of 1492 (76%) of the 1975 exons that were targeted by the VenomCap probe set (range: 1459–1513 targets matched, Table 3). On average, we found that the ingroup samples (all hydrophiines) matched a significantly higher number of VenomCap target exons than the outgroup elapid samples (ingroup mean: 1499 targets matched, outgroup mean: 1481 targets matched; $t = 2.868$, $df = 9.641$, $p = .017$; Figure 2). Including duplicate matches due to paralogues, the total number of contigs recovered that matched VenomCap targets averaged 2239 across all samples (range: 1611–2826 contigs recovered, Table 3). We recovered representative exons from 24 of 32 venom gene families that were targeted by the VenomCap probe set (Table 4). For the snake phylogenetic probe set, we recovered an average of 2745 loci (88%) per sample of the 3129 loci that were targeted (range: 2660–2806 loci, Table 3).

3.2.2 | Sensitivity

The mean sensitivity (the proportion of bases in post-assembly contigs that overlap with the target markers in the probe set) was measured across all markers per sample for both probe sets. For the VenomCap

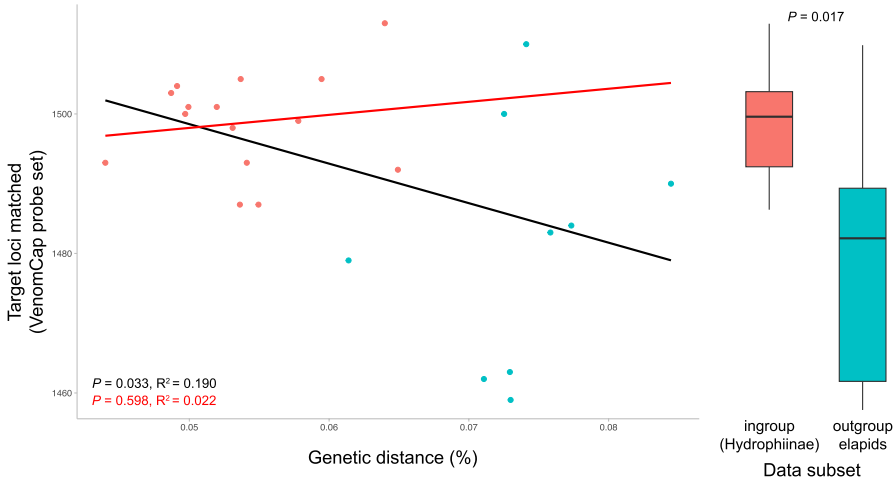

FIGURE 2 Effect of genetic distance on capture success for the VenomCap probe set. (left) Scatterplot of genetic distance from the probe set and total target loci matched. Points represent species, ingroup hydrophiines (red) and outgroup elapids (blue). Regression lines and corresponding p and R^2 values coloured by analysis (black = all species, red = ingroup only). (right) Boxplots comparing total target loci matched between ingroup and outgroup samples (colours and y-axis match the scatterplot, and p value indicated from a Welch's t-test).

TABLE 4 Total number of contigs recovered for each of the venom gene families captured using the VenomCap probe set.

Species	3FTx	PLA2	SVMP	SVSP	CRISP	CTL	KUN	LAAO	NP	NGFb	5NUC	AChE
Acanthophis laevis	96	125	37	296	61	197	165	41	3	2	9	2
Aipysurus duboisii	71	37	24	268	90	208	143	18	5	1	9	2
Aspidomorphus muelleri	82	186	98	367	355	409	327	17	4	2	9	4
Brachyurophis incinctus	109	114	38	432	90	305	219	36	3	4	7	2
Bungarus fasciatus	181	151	34	322	111	210	377	18	2	2	7	2
Bungarus multicinctus	231	72	45	353	78	551	357	22	1	2	5	1
Cacophis churchilli	124	60	24	306	49	219	188	16	2	2	9	0
Calliophis bivirgatus	263	146	53	329	93	190	699	25	0	4	4	2
Demansia rimicola	93	238	227	325	163	402	189	20	5	3	7	2
Demansia vestigiata	92	134	101	361	120	557	167	16	4	3	14	2
Dendroaspis polylepis	368	28	43	331	69	226	327	13	1	1	7	4
Elapsoidea nigra	82	129	39	305	138	193	268	23	0	2	6	0
Elapsoidea sundevallii	299	321	46	389	213	233	266	22	1	3	10	2
Hemiaspis signata	143	147	56	348	165	193	237	38	2	4	9	2
Hydrophis macdowelli	88	91	26	261	83	217	194	21	2	2	11	1
Laticauda cf. *colubrina*	233	125	17	275	63	176	368	16	1	2	5	2
Micruroides euryxanthus	284	212	58	391	352	270	199	15	0	2	5	1
Naja nigricollis	289	84	79	290	210	254	189	46	1	2	8	4
Notechis scutatus	102	340	37	306	221	281	236	42	6	2	6	2
Oxyuranus scutellatus	80	84	18	275	61	159	154	30	6	2	5	2
Pseudechis australis	67	507	83	310	87	314	306	62	7	2	11	4
Pseudonaja textilis	86	110	19	267	69	210	161	30	6	2	11	2
Simoselaps anomalus	99	82	51	443	75	209	149	27	7	2	8	2
Toxicocalamus loriae	345	91	16	284	132	294	281	21	2	3	9	2

Species	CATH	CYST	DPP-IV	FIC	HYAL	KSPI	PDE	TCTP	VEGF	VESP	VF	WAP
Acanthophis laevis	17	5	1	36	2	1	1	6	0	3	1	6
Aipysurus duboisii	19	7	1	34	4	1	1	4	1	3	1	3
Aspidomorphus muelleri	19	6	1	41	4	2	1	6	1	3	4	6
Brachyurophis incinctus	19	6	1	53	3	3	1	8	0	4	1	9
Bungarus fasciatus	17	7	1	65	3	2	1	4	1	3	4	18
Bungarus multicinctus	17	7	1	100	3	3	1	4	1	3	5	8
Cacophis churchilli	17	7	1	40	4	1	1	4	1	3	3	5
Calliophis bivirgatus	21	9	1	74	3	2	1	7	1	6	2	36
Demansia rimicola	28	9	2	83	2	2	1	4	1	6	2	9
Demansia vestigiata	12	9	1	55	3	6	1	5	1	3	3	7
Dendroaspis polylepis	16	7	1	50	2	2	1	9	1	3	3	6
Elapsoidea nigra	16	9	1	70	3	5	1	4	0	3	4	7
Elapsoidea sundevallii	24	10	1	102	3	3	1	6	1	4	3	9
Hemiaspis signata	16	13	1	39	5	1	1	8	0	6	3	6
Hydrophis macdowelli	13	6	1	24	6	1	1	4	1	3	1	6
Laticauda cf. *colubrina*	15	7	1	37	2	1	1	8	1	3	1	6
Micruroides euryxanthus	21	7	1	61	4	5	1	7	1	3	3	9
Naja nigricollis	27	9	1	56	4	1	1	4	1	3	5	6
Notechis scutatus	28	6	1	56	3	2	3	4	1	3	1	8
Oxyuranus scutellatus	13	4	1	31	3	2	1	4	1	3	1	5
Pseudechis australis	22	7	1	36	3	3	1	5	0	3	1	9
Pseudonaja textilis	15	4	1	39	4	2	1	4	1	3	2	5
Simoselaps anomalus	14	6	1	38	6	2	1	5	0	3	1	7
Toxicocalamus loriae	22	10	1	71	5	7	1	7	1	3	1	12

probe set, mean sample sensitivity was 92.6% (range: 91.5%–93.9%). We found no significant difference between mean sample sensitivity when comparing ingroup and outgroup samples for the VenomCap data ($t = 0.956$, $df = 17.281$, $p = 0.352$; Figure 3a). For the phylogenetic probe set, mean sample sensitivity was 96.5% (range: 95.6–97.0%). When comparing probe sets, we found that mean sample sensitivity was significantly higher in the phylogenetic probe set compared to VenomCap ($t = -25.244$, $df = 31.213$, $p < .001$; Figure 3a).

3.2.3 | Specificity

The mean specificity (the proportion of cleaned reads that map to the target markers in the probe set) was measured across all markers per sample for both probe sets. Overall, for each sample, approximately half of the reads mapped to the VenomCap markers, a quarter of the reads mapped to the phylogenomic markers, and a quarter of the reads were unmapped. For the VenomCap probe set, mean sample specificity was 52.8% (range: 47.1%–56.1%). We found no significant difference between mean sample specificity when comparing ingroup and outgroup samples for the VenomCap data ($t = 0.002$, $df = 18.99$, $p = .999$; Figure 3b). For the phylogenetic probe set, mean sample specificity was 26.6% (range: 22.3%–29.7%). When comparing probe sets, we found that mean sample specificity was significantly higher in VenomCap compared to the phylogenetic probe set ($t = 51.627$, $df = 45.602$, $p < .001$; Figure 3b). However, we note that our measures of specificity are not scaled by the number of baits used in each probe set, which differ, and thus are not fully comparable. Furthermore, for researchers interested in utilizing this probe set, we also note that specificity will likely vary depending on whether the VenomCap probe set is sequenced alone or combined with another probe set as we do here.

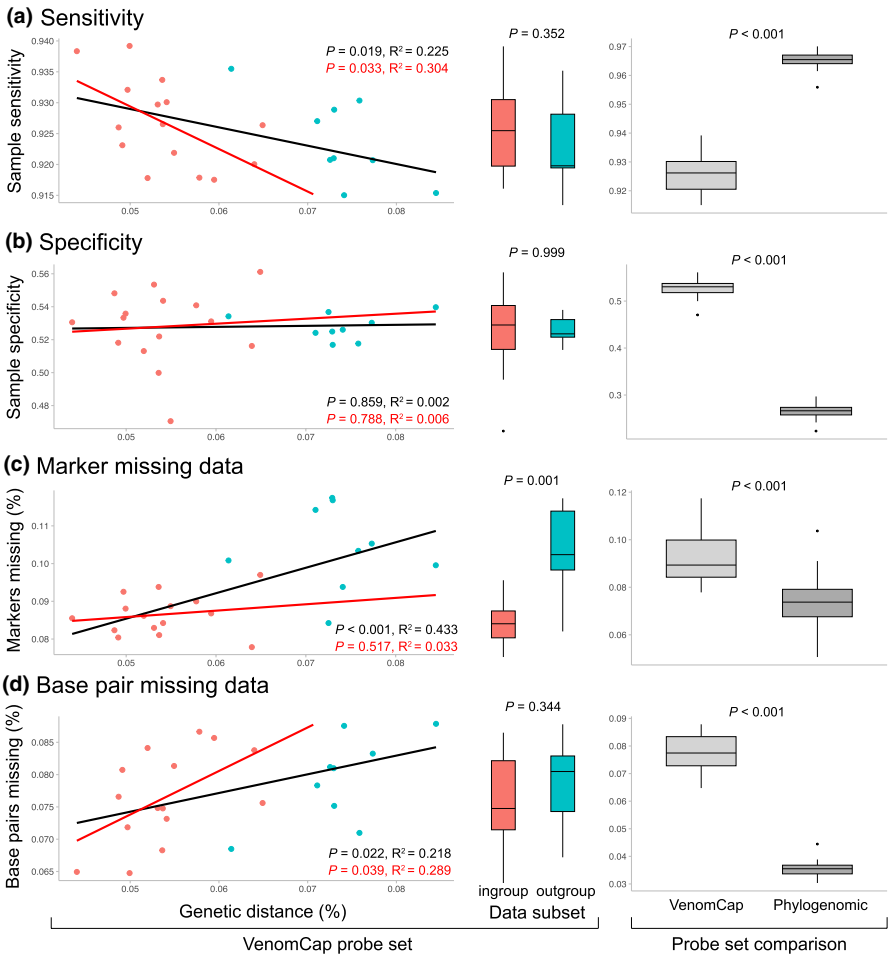

FIGURE 3 Sequence capture evaluation for the following metrics: (a) Sensitivity, the proportion of bases in assembled contigs that overlap with the target markers in the probe set. (b) Specificity, the proportion of cleaned reads that map to the target markers in the probe set. (c) Missing markers, the percentage of missing loci per sample across target marker alignments. (d) Missing bases, the percentage of missing base pairs per sample across all trimmed alignments of target markers captured. For the VenomCap probe set, we evaluated the effect of genetic distance (left scatterplots) and compared sampling groups (middle-coloured boxplots) for each metric of capture success. Scatterplots represent the metric's relationship with a sample's genetic distance from the probe set (red points: Ingroup Hydrophiinae; blue points: Outgroup elapids), with regression lines and corresponding p and R^2 values coloured by analysis (black: All species; red: Ingroup only). Middle boxplot colours correspond to scatterplot points, with the same y-axis; p value indicated from a Welch's t-test. Right boxplots (grayscale) compare the two probe sets using all samples; p value indicated from a Welch's t-test.

3.2.4 | Missing data

We measured the percentage of missing (1) markers and (2) base pairs per sample across all trimmed alignments of target markers captured for both probe sets. (1) For the VenomCap probe set, mean missing marker data per sample was 9.3% across all samples (range: 7.8%–11.7%). The outgroup elapid samples had a significantly higher percentage of missing markers compared to the ingroup hydrophiines for the VenomCap capture data (ingroup mean: 8.7%, outgroup mean: 10.4%; $t = -4.445$, $df = 10.278$, $p = .001$; Figure 3c). For the phylogenetic probe set, mean missing marker data per sample was 7.3% across all samples (range: 5.1%–10.4%). When comparing probe sets, we found that the mean proportion of missing markers from all alignments was significantly higher in the VenomCap data compared to the phylogenetic probe set ($t = 5.9296$, $df = 45.999$, $p < .001$; Figure 3c). (2) Missing base pair data in the VenomCap capture results averaged 7.8% across all samples (range: 6.5%–8.8%). We found no significant difference between mean missing base pair data when comparing ingroup and outgroup samples for the VenomCap data ($t = -0.973$, $df = 17.757$, $p = 0.344$; Figure 3d). For the phylogenetic probe set, mean missing base pair data per sample was 3.5% across all samples (range: 3.0%–4.4%). When comparing probe sets, we found that mean proportion of missing base pairs from all alignments was significantly higher in the VenomCap data compared to the phylogenetic probe set ($t = 27.035$, $df = 31.198$, $p < .001$; Figure 3d).

3.2.5 | Depth of coverage

We measured sequencing depth of coverage across targeted markers using RPKM to enable comparison across samples and probe sets (i.e. raw 'X' coverage is not directly comparable because it is not scaled by sequencing effort). For the VenomCap data, the median sequencing coverage across all samples was 1029.4 RPKM (range: 845.8–1215.2 RPKM; Table S1). The ingroup hydrophiines had significantly higher coverage compared to the outgroup elapid samples for the VenomCap capture data (ingroup mean: 1074.6 RPKM, outgroup mean: 954.0 RPKM; $t = 4.399$, $df = 18.542$, $p < .001$). For the phylogenetic capture results, the median sequencing coverage across all samples was 375.5 RPKM (range: 354.6–407.7 RPKM; Table S2). When comparing probe sets, we found that median coverage per sample was significantly higher in the VenomCap data compared to the phylogenetic data ($t = 35.832$, $df = 23.874$, $p < .001$).

3.2.6 | Genetic distance

For the VenomCap capture data, we assessed the relationship between capture results and genetic distance from the probe set to evaluate how the probe set performed across the phylogenetic breadth of our full dataset (all elapids) and only the samples from our ingroup clade of interest (Hydrophiinae). We found a significant negative relationship between the total target loci matched and genetic distance for the full dataset ($R^2 = 0.19$, $p = .033$; Figure 2), but a nonsignificant relationship between these two variables for the ingroup only ($R^2 = 0.022$, $p = .598$; Figure 2). We also found a significant negative relationship between sample sensitivity and genetic distance for both datasets (full dataset: $R^2 = 0.225$, $p = .019$; hydrophiine only: $R^2 = 0.304$, $p = .033$; Figure 3a). The relationship between sample specificity and genetic distance was nonsignificant for both datasets (full dataset: $R^2 = 0.002$, $p = .859$; hydrophiine only: $R^2 = 0.006$, $p = .788$; Figure 3b). We found a significant positive relationship between the percentage of missing marker data and genetic distance for the full dataset ($R^2 = 0.433$, $p < .001$; Figure 3c), but a nonsignificant relationship between these two variables for the ingroup only ($R^2 = 0.033$, $p = .517$; Figure 3c). We found a significant positive relationship between the percentage of missing base pair data and genetic distance for both datasets (full dataset: $R^2 = 0.218$, $p = .022$; hydrophiine only: $R^2 = 0.289$, $p = .039$; Figure 3d).

3.3 | Phylogenetics

After trimming and filtering the phylogenomic markers, a total of 2624 loci, with an average length of 1407 bp, were retained for phylogenetic analysis of the concatenated dataset. The resulting phylogeny was well-supported, with 100% bootstrap support values across all nodes (Figure 4). Evolutionary relationships among the sampled elapid species align well with previous phylogenomic studies that included major elapid lineages (Burbrink et al., 2020; Das et al., 2024). However, these studies did not comprehensively sample hydrophiines. Prior phylogenetic analyses focusing on hydrophiines were mostly based on Sanger sequencing data (Keogh et al., 1998; Lee et al., 2016; Sanders et al., 2008; Strickland et al., 2016). Our results are largely consistent with these earlier topologies, but our large data set provides much stronger support for the relationships among the hydrophiine lineages we sampled (Figure 4).

4 | DISCUSSION

This paper introduces the VenomCap probe set, a novel tool freely available for efficiently capturing and sequencing the exonic regions of venom genes across elapid snakes. We showcase the probe set's effectiveness in capturing exons from multiple venom gene families across 24 elapid species, including those lacking prior venom information (Figure 1b). Importantly, this method solely relies on genomic DNA from museum tissue samples, eliminating the need for costly and difficult to acquire venom gland-derived RNA. Below, we assess our sequence capture results in the context of overall capture success, capture success across our phylogenetic breadth of sampling, and a comparison between targeted marker types (VenomCap versus phylogenomic probe sets). While our primary focus during probe set design was on the Elapidae, particularly the diverse Hydrophiinae, we discuss its adaptability for expansion to include

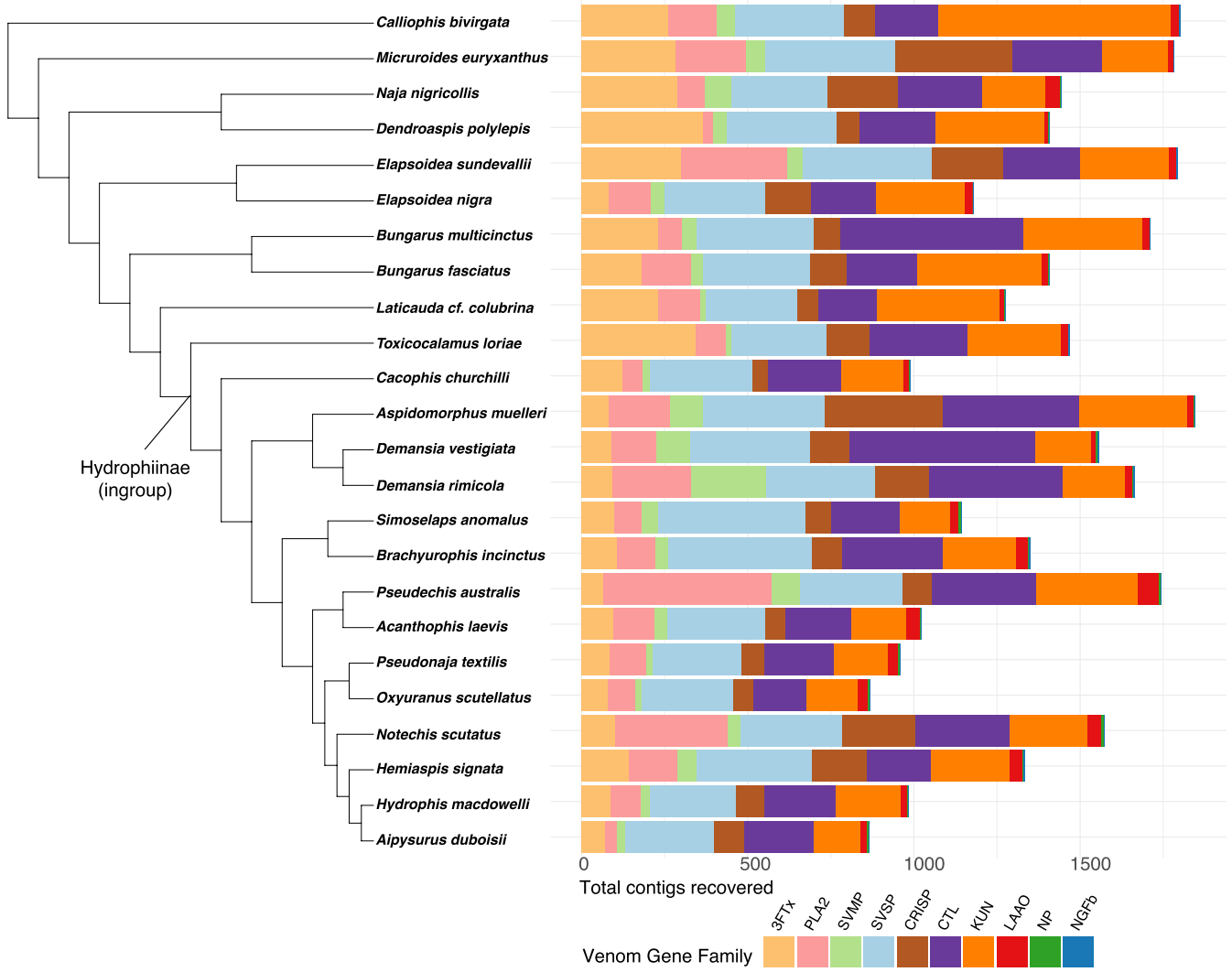

FIGURE 4 VenomCap capture results in a phylogenetic context. Ultrametric maximum likelihood phylogeny of elapid samples sequenced based on the concatenated alignment of 2624 loci captured by the phylogenomic probe set (excludes VenomCap loci). All nodes recovered with 100% ultrafast bootstrap support. Barplots correspond to VenomCap results, depicting the total contigs recovered for 10 of the major venom gene families in elapid venoms. Rightmost bars for NP and NGFb gene families are difficult to visualize as they recovered ~3 and ~2 loci, respectively, on average (see Table 4).

other snake lineages, newly identified venom genes, or customization for specific sequencing needs.

4.1 | Overall capture success

A high percentage of our sequencing reads were on target to the probe sets used here (overall mean specificity = 79.4%), indicating a higher target enrichment success than other studies that use similar approaches (Hutter et al., 2022; Phuong & Mahardika, 2018; Portik et al., 2016). This was largely driven by the high proportion of mapped reads to the VenomCap target markers (VenomCap mean specificity = 52.8%). We attribute this result to a combination of the following factors. First, we used a high tiling density (4x) for the VenomCap probe set. A higher tiling density increases the probability of capturing a target marker, and most phylogenomic probe sets typically use lower tiling schemes (e.g. 2x) as a trade-off to include more loci for a given bait kit size (Andermann et al., 2020; Tewhey et al., 2009). Next, we used a consensus sequence-based approach when designing the VenomCap probe set. Previous studies have shown that probes designed from consensus sequences, or similar approaches that incorporate sequence information from multiple lineages into probe design, can increase sample specificity and improve capture success across a wider phylogenetic breadth compared to probes designed from a single genome (Beaudry et al., 2021; Hugall et al., 2016; Hutter et al., 2022). Finally, given that venom genes can exhibit particularly high copy numbers (Margres et al., 2017), a higher proportion of mapped reads would be expected compared to phylogenomic probe sets based on single-copy loci.

In addition to the high proportion of mapped reads, a high percentage of bases in the assembled contigs overlapped with the VenomCap probe set as a whole (VenomCap mean

sensitivity = 92.6%). Thus, reads were not only on target, but they effectively captured most of the bases targeted by the VenomCap probe set. In total, these data contained representative exons from 24 of the 32 venom gene families that were targeted (Tables 2 and 4). An inspection of the eight venom gene families that were targeted but not captured revealed that many of the baits targeting these loci had suboptimal GC content (below 40% or above 60%) for target enrichment, which may have limited their capture success (Benjamini & Speed, 2012). Nevertheless, we were still able to capture the 10 most prevalent venom gene families found in elapid venoms (Tables 2 and 4; Figures 1 and 4; Tasoulis & Isbister, 2017). We also uncovered substantial variation in the number of contigs recovered for many of these gene families (Table 4; Figure 4), including those known to have high copy numbers within elapids (Li et al., 2021; Zhang et al., 2022). Although this is likely indicative of gene copy number variation, we urge readers not to over interpret these results until future studies can assess them in more detail.

4.2 | Genetic divergence from probes

VenomCap's success in capturing target loci across a diversity of elapid snake lineages sheds light on the utility of this tool for comparative genomic studies. Across all sequenced elapids, our findings reveal a significant negative relationship between the number of VenomCap target loci recovered and genetic distance from the probe set (Figure 2). This result is expected given that a species' genetic distance from the design markers is known to be one of the biggest limiting factors of sequence capture success (Andermann et al., 2020), and we included all the major elapid lineages, which are estimated to have initially diverged over 25 Ma (Burbrink et al., 2020; Lee et al., 2016). Nevertheless, the magnitude of this decay in loci recovered with increasing genetic distance was relatively minor—out of ~1500 targeted loci recovered, the more distantly related outgroup samples recovered only 18 fewer loci on average than the ingroup hydrophiine clade that was used for probe design. Furthermore, within the ingroup hydrophiines, which are estimated to have initially diverged in the early to mid-Miocene (Burbrink et al., 2020; Lee et al., 2016), genetic distance did not significantly predict the number of target loci recovered. This suggests a robustness of the VenomCap probe set within hydrophiines, which we attribute to the broad phylogenetic spectrum of hydrophiine genomic resources used for exon selection and our consensus sequence-based probe design. These results suggest that VenomCap's efficiency might vary depending on the evolutionary divergence of the targeted taxa from the hydrophiine clade, but overall, the probe set performs well at capturing the targeted markers across a broad phylogenetic spectrum of elapids.

4.3 | Molecular marker type

Comparing our capture results between the VenomCap and snake phylogenomic probe sets also offers insights into the impact of molecular marker type on capture efficiency. VenomCap is unique when compared to most phylogenomic target capture approaches, which are designed to capture single-copy loci that are sufficiently conserved to be recovered across many divergent taxa (Andermann et al., 2020). Paralogous loci and genes under strong diversifying selection are generally filtered out during marker selection and design of these probe sets. Venom genes, on the other hand, are known for high levels of duplication and loss (Dowell et al., 2016; Giorgianni et al., 2020; Zhang et al., 2022), as well as rapid molecular evolution due to strong selection pressures (Gibbs & Rossiter, 2008; Li et al., 2021; Shibata et al., 2018), which could reduce capture success. As expected, given these differences between the targeted marker types, our capture results indicate that the phylogenomic probe set performs significantly better than VenomCap in having lower levels of missing data across alignments (both missing markers and missing base pairs; Figure 3), and higher capture rates for the proportion of overall loci and bases targeted by the probe set (Table 3; Figure 3). This result is likely due to the higher turnover of venom genes leading to the absence of loci in certain taxa (i.e. gene loss or our inclusion of genes that are unique to select lineages used in marker selection) or strong positive selection on certain exons that would increase sequence divergence and decrease the likelihood of capturing it across divergent lineages. Nevertheless, the average amount of missing data across shared VenomCap markers is still quite low (<10% for both missing markers and base pairs across alignments), and we were also able to recover many loci from gene families that are noted for rapid molecular evolution (e.g. three-finger toxins; Li et al., 2021; Zhang et al., 2022).

4.4 | Limitations and probe set customization

Although VenomCap's ability to rapidly capture venom gene sequences across many taxa provides an important step forward for snake venom research, our approach does have its limitations. For example, as noted by Phuong and Mahardika (2018), the venom genes sequenced using an exon-capture approach will ultimately be limited to those that were targeted by the probe set. Thus, our approach will fail to capture completely novel venom genes, venom genes that were not available from the venomic resources used at the time of our marker selection, or venom genes that are too divergent from our baits. Fortunately, the ever-increasing availability of high-quality snake genomes (Pinto et al., 2023) and venom gland transcriptomes will continue to identify new markers that can be incorporated into the VenomCap probe set. Better genomic resources, coupled with new automated venom gene annotators such as ToxCodAn (Nachtigall et al., 2021) or ToxCodAn-Genome (Nachtigall et al., 2024), will facilitate easy customization of the VenomCap probe set to include new markers or markers for a specific snake clade of interest. Although we demonstrate that VenomCap successfully captures many of the targeted markers across Elapidae, for researchers interested in applying it to non-hydrophiine elapids, we recommend expanding the probe set by incorporating more genomic resources from their clade of interest to make it more

comprehensive. The same approach could easily be used to further expand VenomCap to include other venomous snake families.

Another limitation of VenomCap involves the annotation of venom genes from short-read, exon-capture data. First, for genes that have undergone extensive duplication, it will be difficult to assign exons to gene models without a reference genome, unless exons are physically linked on the same contig. Linked-read sequencing approaches (e.g. Marks et al., 2019) could prove useful for addressing this issue, provided that gene copies are far enough apart on a chromosome or on different chromosomes. However, duplicated venom genes are often arranged in tandem repeat clusters on a chromosome (Giorgianni et al., 2020). More likely, evolutionary inferences using VenomCap data will need to be carried out on an exon-by-exon basis in the absence of a close reference genome to accurately define full genes. Another issue will involve distinguishing toxin coding loci from non-toxin paralogues, along with assessing whether the captured exons are functional. In the absence of expression data, this will likely rely on homology- or phylogenetic-based approaches to identify sequence similarity with known toxin-coding genes, as well as verifying an open reading frame. This can still be limited by the high turnover of venom genes which frequently leads to the presence of pseudogenes and orphan exons, which lose function but retain high sequence identity with toxin-coding genes (Dowell et al., 2016; Giorgianni et al., 2020; Zhang et al., 2022). Ultimately, any venomic sequencing strategy will depend on the research question, and no single approach will fit all. We consider the wealth of new sequence data for comparative venom genomics, which VenomCap has the potential to rapidly generate, to far outweigh any of its limitations.

Finally, to guide researchers in planning studies using the VenomCap probe set, it is important to address cost considerations and synthesis options. Costs can vary widely based on factors such as the chosen provider, the number of samples, the bait kit size, and whether lab work is conducted in-house or outsourced. As a result, providing a precise cost estimate here would be challenging. However, we emphasize the accessibility of the markers and baits, which are freely available on the VenomCap GitHub page. This open-access resource allows researchers to synthesize the probe set with any provider that meets their needs, and it offers the flexibility to perform library preparation and capture in their own laboratories or outsource these tasks if preferred. Additionally, it offers the versatility to customize the bait kit to fit different objectives. Researchers can choose to synthesize only the venom markers, integrate additional probe sets like the snake phylogenomic markers used in this study, or modify the venom baits in other ways to align with their specific project goals.

5 | CONCLUSION

VenomCap is a new tool enabling targeted sequencing of snake venom genes across a diverse array of elapid species. It facilitates the rapid and cost-effective exploration of venom gene diversity, allowing for new lines of comparative genomic research on venom gene composition and molecular evolution. It is the first approach of its kind to be broadly tested across an entire venomous snake family. We demonstrate its target capture efficiency across the Hydrophiinae used to design the probe set, and the Elapidae as a whole. Currently, we are working on expanding our VenomCap sequencing effort across hydrophiines to test evolutionary hypotheses on the ecological drivers of venom gene evolution. However, continued refinement and customization of the probe set, with increases in genomic resources, hold promise for further expanding VenomCap to include more venomous snake lineages. With improved sequencing technologies, the true complexity of genetic diversity underlying snake venoms can now be revealed, and VenomCap offers a versatile tool to catalyse advances in this endeavour.

AUTHOR CONTRIBUTIONS

SLT conceptualized the study, with support of its design and execution from SR. SLT and CRH designed the probe set. SLT performed the laboratory work. CCA, SCD and MDB contributed tissue samples. CRH developed the bioinformatic pipelines for data processing and analysis, and SLT analysed the data. SLT wrote the manuscript, and all authors reviewed and revised the manuscript.

ACKNOWLEDGEMENTS

This work was supported by a National Institutes of Health IRACDA postdoctoral fellowship (NIH GM093854) to SLT. This work was also funded in part by grants from the Coypu Foundation, National Geographic (NGS-53506R-18), National Science Foundation DEB-1146033 and DEB-1926783 to CCA and National Science Foundation DEB-2224119 to SR. CRH was supported through an NSF Postdoctoral Fellowship in Biology (2010988). We thank the labs of SR and CEE at Rutgers University for input at various stages of this work, and especially Justin Bernstein for his assistance in the molecular lab. For processing tissue loans, we thank Sally South and Ralph Foster (Australian Biological Tissue Collection, South Australian Museum), David Boyd and Jackson Roberts (Louisiana State University Museum of Natural Science), Aaron Bauer (Villanova University), José Rosado and Breda Zimkus (Museum of Comparative Zoology, Harvard University) and Lauren Scheinberg (California Academy of Sciences). For the snake phylogenomic markers used here, we thank Jeff Weinell for sharing the probe sequences and Karen Olson for sharing unpublished sequence data for *Boiga irregularis*. Finally, we thank Daicel Arbor BioSciences and their staff scientists for assistance in designing the probes and sequencing, especially Jennifer Klunk, Brian Brunelle and Melissa Curtis.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The VenomCap probe set is freely available and can be found on GitHub (https://github.com/VenomCap). Raw sequence reads have

been deposited to the NCBI Sequence Read Archive (BioProject PRJNA1145850). Aligned sequence files are available through Dryad (https://doi.org/10.5061/dryad.7wm37pw2h). All data analysis scripts are available at GitHub (https://github.com/chutter/PhyloProcessR).

BENEFIT-SHARING STATEMENT

The research in this publication complies with relevant national laws implementing the Convention on Biological Diversity and Nagoya Protocol.

ORCID

Scott L. Travers https://orcid.org/0000-0003-4656-4613
Carl R. Hutter https://orcid.org/0000-0001-6381-6339
Christopher C. Austin https://orcid.org/0000-0003-3890-418X
Stephen C. Donnellan https://orcid.org/0000-0002-5448-3226
Matthew D. Buehler https://orcid.org/0000-0001-5460-7770
Christopher E. Ellison https://orcid.org/0000-0002-0350-0962
Sara Ruane https://orcid.org/0000-0002-9543-1297

REFERENCES

Almeida, D. D., Viala, V. L., Nachtigall, P. G., Broe, M., Gibbs, H. L., Serrano, S. M. D. T., Moura-da-Silva, A. M., Ho, P. L., Nishiyama-Jr, M. Y., & Junqueira-de-Azevedo, I. L. (2021). Tracking the recruitment and evolution of snake toxins using the evolutionary context provided by the *Bothrops jararaca* genome. *Proceedings of the National Academy of Sciences*, *118*(20), e2015159118. https://doi.org/10.1073/pnas.2015159118

Andermann, T., Torres Jiménez, M. F., Matos-Maraví, P., Batista, R., Blanco-Pastor, J. L., Gustafsson, A. L. S., Kistler, L., Liberal, I. M., Oxelman, B., Bacon, C. D., & Antonelli, A. (2020). A guide to carrying out a phylogenomic target sequence capture project. *Frontiers in Genetics*, *10*, 1407. https://doi.org/10.3389/fgene.2019.01407

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Beaudry, M. S., Wang, J., Kieran, T. J., Thomas, J., Bayona-Vásquez, N. J., Gao, B., Devault, A., Brunelle, B., Lu, K., Wang, J.-S., Rhodes, O. E., & Glenn, T. C. (2021). Improved microbial community characterization of 16S rRNA via metagenome hybridization capture enrichment. *Frontiers in Microbiology*, *12*, 644662. https://doi.org/10.3389/fmicb.2021.644662

Benjamini, Y., & Speed, T. P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Research*, *40*(10), e72. https://doi.org/10.1093/nar/gks001

Brahma, R. K., McCleary, R. J., Kini, R. M., & Doley, R. (2015). Venom gland transcriptomics for identifying, cataloging, and characterizing venom proteins in snakes. *Toxicon*, *93*, 1–10. https://doi.org/10.1016/j.toxicon.2014.10.022

Burbrink, F. T., Grazziotin, F. G., Pyron, R. A., Cundall, D., Donnellan, S., Irish, F., Keogh, J. S., Kraus, F., Murphy, R. W., Noonan, B., Raxworthy, C. J., Ruane, S., Lemmon, A. R., Lemmon, E. M., & Zaher, H. (2020). Interrogating genomic-scale data for Squamata (lizards, snakes, and amphisbaenians) shows no support for key traditional morphological relationships. *Systematic Biology*, *69*(3), 502–520. https://doi.org/10.1093/sysbio/syz062

Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Casewell, N. R., Jackson, T. N., Laustsen, A. H., & Sunagar, K. (2020). Causes and consequences of snake venom variation. *Trends in Pharmacological Sciences*, *41*(8), 570–581. https://doi.org/10.1016/j.tips.2020.05.006

Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural Approaches to Sequence Evolution* (Vol. *3*, pp. 207–232). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-35306-5_10

Chatrath, S. T., Chapeaurouge, A., Lin, Q., Lim, T. K., Dunstan, N., Mirtschin, P., Kumar, P. P., & Kini, R. M. (2011). Identification of novel proteins from the venom of a cryptic snake *Drysdalia coronoides* by a combined transcriptomics and proteomics approach. *Journal of Proteome Research*, *10*(2), 739–750. https://doi.org/10.1021/pr1008916

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Das, S., Greenbaum, E., Brecko, J., Pauwels, O. S., Ruane, S., Pirro, S., & Merilä, J. (2024). Phylogenomics of *Psammodynastes* and *Buhoma* (Elapoidea: Serpentes), with the description of a new Asian snake family. *Scientific Reports*, *14*(1), 9489. https://doi.org/10.1038/s41598-024-60215-2

Doley, R., Tram, N. N. B., Reza, M. A., & Kini, R. M. (2008). Unusual accelerated rate of deletions and insertions in toxin genes in the venom glands of the pygmy copperhead (*Austrelaps labialis*) from kangaroo Island. *BMC Evolutionary Biology*, *8*, 70. https://doi.org/10.1186/1471-2148-8-70

Dowell, N. L., Giorgianni, M. W., Kassner, V. A., Selegue, J. E., Sanchez, E. E., & Carroll, S. B. (2016). The deep origin and recent loss of venom toxin genes in rattlesnakes. *Current Biology*, *26*(18), 2434–2445. https://doi.org/10.1016/j.cub.2016.07.038

Durai, D. A., & Schulz, M. H. (2019). Improving in-silico normalization using read weights. *Scientific Reports*, *9*, 5133. https://doi.org/10.1038/s41598-019-41502-9

Earl, S. T., Birrell, G. W., Wallis, T. P., St Pierre, L. D., Masci, P. P., de Jersey, J., Gorman, J. J., & Lavin, M. F. (2006). Post-translational modification accounts for the presence of varied forms of nerve growth factor in Australian elapid snake venoms. *Proteomics*, *6*(24), 6554–6565. https://doi.org/10.1002/pmic.200600263

Earl, S. T., Masci, P. P., de Jersey, J., Lavin, M. F., & Dixon, J. (2012). Drug development from Australian elapid snake venoms and the Venomics pipeline of candidates for haemostasis: Textilinin-1 (Q8008), Haempatch™(Q8009) and CoVase™(V0801). *Toxicon*, *59*(4), 456–463. https://doi.org/10.1016/j.toxicon.2010.12.010

Gibbs, H. L., & Rossiter, W. (2008). Rapid evolution by positive selection and gene gain and loss: PLA_2 venom genes in closely related *Sistrurus* rattlesnakes with divergent diets. *Journal of Molecular Evolution*, *66*, 151–166. https://doi.org/10.1007/s00239-008-9067-7

Giorgianni, M. W., Dowell, N. L., Griffin, S., Kassner, V. A., Selegue, J. E., & Carroll, S. B. (2020). The origin and diversification of a novel protein family in venomous snakes. *Proceedings of the National Academy of Sciences*, *117*(20), 10911–10920. https://doi.org/10.1073/pnas.1920011117

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., & Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, *27*(2), 182–189. https://doi.org/10.1038/nbt.1523

Gong, N., Armugam, A., & Jeyaseelan, K. (2000). Molecular cloning, characterization and evolution of the gene encoding a new group of short-chain α-neurotoxins in an Australian elapid, *Pseudonaja textilis*. *FEBS Letters*, *473*(3), 303–310. https://doi.org/10.1016/S0014-5793(00)01549-0

Gutiérrez, J. M., Calvete, J. J., Habib, A. G., Harrison, R. A., Williams, D. J., & Warrell, D. A. (2017). Snakebite envenoming. *Nature Reviews Disease Primers*, *3*(1), 17063. https://doi.org/10.1038/nrdp.2017.63

Hogan, M. P., Holding, M. L., Nystrom, G. S., Colston, T. J., Bartlett, D. A., Mason, A. J., Ellsworth, S. A., Rautsaw, R. M., Lawrence, K. C., Strickland, J. L., He, B., Fraser, P., Margres, M. J., Gilbert, D. M., Gibbs, H. L., Parkinson, C. L., & Rokyta, D. R. (2024). The genetic regulatory architecture and epigenomic basis for age-related changes in rattlesnake venom. *Proceedings of the National Academy of Sciences*, *121*(16), e2313440121. https://doi.org/10.1073/pnas.2313440121

Holding, M. L., Biardi, J. E., & Gibbs, H. L. (2016). Coevolution of venom function and venom resistance in a rattlesnake predator and its squirrel prey. *Proceedings of the Royal Society B: Biological Sciences*, *283*, 20152841. https://doi.org/10.1098/rspb.2015.2841

Holding, M. L., Strickland, J. L., Rautsaw, R. M., Hofmann, E. P., Mason, A. J., Hogan, M. P., Nystrom, G. S., Ellsworth, S. A., Colston, T. J., Borja, M., Castañeda-Gaytán, G., Grünwald, C. I., Jones, J. M., Freitas-de-Sousa, L. A., Viala, V. L., Margres, M. J., Hingst-Zaher, E., Junqueira-de-Azevedo, I. L. M., Moura-da-Silva, A. M., … Parkinson, C. L. (2021). Phylogenetically diverse diets favor more complex venoms in north American pitvipers. *Proceedings of the National Academy of Sciences*, *118*(17), e2015579118. https://doi.org/10.1073/pnas.2015579118

Hugall, A. F., O'Hara, T. D., Hunjan, S., Nilsen, R., & Moussalli, A. (2016). An exon-capture system for the entire class Ophiuroidea. *Molecular Biology and Evolution*, *33*, 281–294. https://doi.org/10.1093/molbev/msv216

Hutter, C. R., Cobb, K. A., Portik, D. M., Travers, S. L., Wood, P. L., Jr., & Brown, R. M. (2022). FrogCap: A modular sequence capture probe-set for phylogenomics and population genetics for all frogs, assessed across multiple phylogenetic scales. *Molecular Ecology Resources*, *22*(3), 1100–1119. https://doi.org/10.1111/1755-0998.13517

Jackson, T. N., Sunagar, K., Undheim, E. A., Koludarov, I., Chan, A. H., Sanders, K., Ali, S. A., Hendrikx, I., Dunstan, N., & Fry, B. G. (2013). Venom down under: Dynamic evolution of Australian elapid snake toxins. *Toxins*, *5*(12), 2621–2655. https://doi.org/10.3390/toxins5122621

Johnston, C. I., Ryan, N. M., Page, C. B., Buckley, N. A., Brown, S. G., O'Leary, M. A., & Isbister, G. K. (2017). The Australian snakebite project, 2005–2015 (ASP-20). *Medical Journal of Australia*, *207*(3), 119–125. https://doi.org/10.5694/mja17.00094

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Karin, B. R., Gamble, T., & Jackman, T. R. (2020). Optimizing phylogenomics with rapidly evolving long exons: Comparison with anchored hybrid enrichment and ultraconserved elements. *Molecular Biology and Evolution*, *37*(3), 904–922. https://doi.org/10.1093/molbev/msz263

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Keogh, J. S., Shine, R., & Donnellan, S. (1998). Phylogenetic relationships of terrestrial Australo-Papuan elapid snakes (subfamily Hydrophiinae) based on cytochrome b and 16S rRNA sequences. *Molecular Phylogenetics and Evolution*, *10*(1), 67–81. https://doi.org/10.1006/mpev.1997.0471

Laurence, M., Hatzis, C., & Brash, D. E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low abundance microbes. *PLoS One*, *9*(5), e97876–e97878. https://doi.org/10.1371/journal.pone.0097876

Laustsen, A. H., Engmark, M., Milbo, C., Johannesen, J., Lomonte, B., Maria Gutierrez, J., & Lohse, B. (2016). From fangs to pharmacology: The future of snakebite envenoming therapy. *Current Pharmaceutical Design*, *22*(34), 5270–5293. https://doi.org/10.2174/1381612822666160623073438

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Lee, M. S., Sanders, K. L., King, B., & Palci, A. (2016). Diversification rates and phenotypic evolution in venomous snakes (Elapidae). *Royal Society Open Science*, *3*(1), 150277. https://doi.org/10.1098/rsos.150277

Li, A., Wang, J., Sun, K., Wang, S., Zhao, X., Wang, T., Xiong, L., Xu, W., Qiu, L., Shang, Y., Liu, R., Wang, S., & Lu, Y. (2021). Two reference-quality sea snake genomes reveal their divergent evolution of adaptive traits and venom systems. *Molecular Biology and Evolution*, *38*(11), 4867–4883. https://doi.org/10.1093/molbev/msab212

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Mackessy, S. P. (2022). Venom production and secretion in reptiles. *Journal of Experimental Biology*, *225*(7), jeb227348. https://doi.org/10.1242/jeb.227348

Margres, M. J., Bigelow, A. T., Lemmon, E. M., Lemmon, A. R., & Rokyta, D. R. (2017). Selection to increase expression, not sequence diversity, precedes gene family origin and expansion in rattlesnake venom. *Genetics*, *206*(3), 1569–1580. https://doi.org/10.1534/genetics.117.202655

Margres, M. J., Rautsaw, R. M., Strickland, J. L., Mason, A. J., Schramer, T. D., Hofmann, E. P., Stiers, E., Ellsworth, S. A., Nystrom, G. S., Hogan, M. P., Bartlett, D. A., Colston, T. J., Gilbert, D. M., Rokyta, D. R., & Parkinson, C. L. (2021). The Tiger rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *Proceedings of the National Academy of Sciences*, *118*(4), e2014634118. https://doi.org/10.1073/pnas.2014634118

Marks, P., Garcia, S., Barrio, A. M., Belhocine, K., Bernate, J., Bharadwaj, R., Bjornson, K., Catalanotti, C., Delaney, J., Fehr, A., Fiddes, I. T., Galvin, B., Heaton, H., Herschleb, J., Hindson, C., Holt, E., Jabara, C. B., Jett, S., Keivanfar, N., … Church, D. M. (2019). Resolving the full spectrum of human genome variation using linked-reads. *Genome Research*, *29*(4), 635–645. https://doi.org/10.1101/gr.234443.118

Mason, A. J., Holding, M. L., Rautsaw, R. M., Rokyta, D. R., Parkinson, C. L., & Gibbs, H. L. (2022). Venom gene sequence diversity and expression jointly shape diet adaptation in pitvipers. *Molecular Biology and Evolution*, *39*(4), msac082. https://doi.org/10.1093/molbev/msac082

Minh, B. Q., Nguyen, M. A. T., & Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic

era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Modahl, C. M., Brahma, R. K., Koh, C. Y., Shioi, N., & Kini, R. M. (2020). Omics technologies for profiling toxin diversity and evolution in snake venom: Impacts on the discovery of therapeutic and diagnostic agents. *Annual Review of Animal Biosciences*, *8*, 91–116. https://doi.org/10.1146/annurev-animal-021419-083626

Mohamed Abd El-Aziz, T., Soares, A. G., & Stockand, J. D. (2019). Snake venoms in drug discovery: Valuable therapeutic tools for life saving. *Toxins*, *11*(10), 564. https://doi.org/10.3390/toxins11100564

Nachtigall, P. G., Durham, A. M., Rokyta, D. R., & Junqueira-de-Azevedo, I. L. (2024). ToxCodAn-genome: An automated pipeline for toxin-gene annotation in genome assembly of venomous lineages. *GigaScience*, *13*, giad116. https://doi.org/10.1093/gigascience/giad116

Nachtigall, P. G., Rautsaw, R. M., Ellsworth, S. A., Mason, A. J., Rokyta, D. R., Parkinson, C. L., & Junqueira-de-Azevedo, I. L. (2021). ToxCodAn: A new toxin annotator and guide to venom gland transcriptomics. *Briefings in Bioinformatics*, *22*(5), bbab095. https://doi.org/10.1093/bib/bbab095

Nikolenko, S. I., Korobeynikov, A. I., & Alekseyev, M. A. (2013). BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics*, *14*(Suppl 1), S7. https://doi.org/10.1186/1471-2164-14-S1-S7

Paiva, O., Pla, D., Wright, C. E., Beutler, M., Sanz, L., Gutiérrez, J. M., Williams, D. J., & Calvete, J. J. (2014). Combined venom gland cDNA sequencing and venomics of the New Guinea small-eyed snake, *Micropechis ikaheka*. *Journal of Proteomics*, *110*, 209–229. https://doi.org/10.1016/j.jprot.2014.07.019

Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Peng, C., Ren, J. L., Deng, C., Jiang, D., Wang, J., Qu, J., Chang, J., Yan, C., Jiang, K., Murphy, R. W., Wu, D. D., & Li, J. T. (2020). The genome of Shaw's sea snake (*Hydrophis curtus*) reveals secondary adaptation to its marine environment. *Molecular Biology and Evolution*, *37*(6), 1744–1760. https://doi.org/10.1093/molbev/msaa043

Phuong, M. A., & Mahardika, G. N. (2018). Targeted sequencing of venom genes from cone snail genomes improves understanding of conotoxin molecular evolution. *Molecular Biology and Evolution*, *35*(5), 1210–1224. https://doi.org/10.1093/molbev/msy034

Pinto, B. J., Gamble, T., Smith, C. H., & Wilson, M. A. (2023). A lizard is never late: Squamate genomics as a recent catalyst for understanding sex chromosome and microchromosome evolution. *Journal of Heredity*, *114*(5), 445–458. https://doi.org/10.1093/jhered/esad023

Portik, D. M., Smith, L. L., & Bi, K. (2016). An evaluation of transcriptome based exon capture for frog phylogenomics across multiple scales of divergence (class: Amphibia, order: Anura). *Molecular Ecology Resources*, *16*(5), 1069–1083. https://doi.org/10.1111/1755-0998.12541

Pyron, R. A., Hendry, C. R., Chou, V. M., Lemmon, E. M., Lemmon, A. R., & Burbrink, F. T. (2014). Effectiveness of phylogenomic data and coalescent species-tree methods for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes: Caenophidia). *Molecular Phylogenetics and Evolution*, *81*, 221–231. https://doi.org/10.1016/j.ympev.2014.08.023

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rao, W. Q., Kalogeropoulos, K., Allentoft, M. E., Gopalakrishnan, S., Zhao, W. N., Workman, C. T., Knudsen, C., Jiménez-Mena, B., Seneci, L., Mousavi-Derazmahalleh, M., Jenkins, T. P., Rivera-de-Torre, E., Liu, S. Q., & Laustsen, A. H. (2022). The rise of genomics in snake venom research: Recent advances and future perspectives. *GigaScience*, *11*, giac024. https://doi.org/10.1093/gigascience/giac024

Reyes-Velasco, J., Card, D. C., Andrew, A. L., Shaney, K. J., Adams, R. H., Schield, D. R., Casewell, N. R., Mackessy, S. P., & Castoe, T. A. (2015). Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. *Molecular Biology and Evolution*, *32*(1), 173–183. https://doi.org/10.1093/molbev/msu294

Richards, R., St Pierre, L., Trabi, M., Johnson, L. A., de Jersey, J., Masci, P. P., & Lavin, M. F. (2011). Cloning and characterisation of novel cystatins from elapid snake venom glands. *Biochimie*, *93*(4), 659–668. https://doi.org/10.1016/j.biochi.2010.12.008

Roberts, J. R., Bernstein, J. M., Austin, C. C., Hains, T., Mata, J., Kieras, M., Pirro, S., & Ruane, S. (2024). Whole snake genomes from eighteen families of snakes (Serpentes: Caenophidia) and their applications to systematics. *Journal of Heredity*, *esae026*, 487–497. https://doi.org/10.1093/jhered/esae026

Sanders, K. L., Lee, M. S., Leys, R., Foster, R., & Keogh, J. S. (2008). Molecular phylogeny and divergence dates for Australasian elapids and sea snakes (Hydrophiinae): Evidence from seven genes for rapid evolutionary radiations. *Journal of Evolutionary Biology*, *21*(3), 682–695. https://doi.org/10.1111/j.1420-9101.2008.01525.x

Shibata, H., Chijiwa, T., Oda-Ueda, N., Nakamura, H., Yamaguchi, K., Hattori, S., Matsubara, K., Matsuda, Y., Yamashita, A., Isomoto, A., Mori, K., Tashiro, K., Kuhara, S., Yamasaki, S., Fujie, M., Goto, H., Koyanagi, R., Takeuchi, T., Fukumaki, Y., … Ogawa, T. (2018). The habu genome reveals accelerated evolution of venom protein genes. *Scientific Reports*, *8*(1), 11300. https://doi.org/10.1038/s41598-018-28749-4

Singhal, S., Grundler, M., Colli, G., & Rabosky, D. L. (2017). Squamate conserved loci (SqCL): A unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. *Molecular Ecology Resources*, *17*(6), e12–e24. https://doi.org/10.1111/1755-0998.12681

Smith, C. F., Nikolakis, Z. L., Ivey, K., Perry, B. W., Schield, D. R., Balchan, N. R., Parker, J., Hansen, K. C., Saviola, A. J., Castoe, T. A., & Mackessy, S. P. (2023). Snakes on a plain: Biotic and abiotic factors determine venom compositional variation in a wide-ranging generalist rattlesnake. *BMC Biology*, *21*(1), 136. https://doi.org/10.1186/s12915-023-01626-x

St Pierre, L., Birrell, G. W., Earl, S. T., Wallis, T. P., Gorman, J. J., de Jersey, J., Masci, P. P., & Lavin, M. F. (2007). Diversity of toxic components from the venom of the evolutionarily distinct black whip snake, *Demansia vestigiata*. *Journal of Proteome Research*, *6*(8), 3093–3107. https://doi.org/10.1021/pr0701613

St Pierre, L., Earl, S. T., Filippovich, I., Sorokina, N., Masci, P. P., de Jersey, J., & Lavin, M. F. (2008). Common evolution of waprin and kunitz-like toxin families in Australian venomous snakes. *Cellular and Molecular Life Sciences*, *65*, 4039–4054. https://doi.org/10.1007/s00018-008-8573-5

St Pierre, L., Fischer, H., Adams, D. J., Schenning, M., Lavidis, N., de Jersey, J., Masci, P. P., & Lavin, M. F. (2007). Distinct activities of novel neurotoxins from Australian venomous snakes for nicotinic acetylcholine receptors. *Cellular and Molecular Life Sciences*, *64*, 2829–2840. https://doi.org/10.1007/s00018-007-7352-z

St Pierre, L., Flight, S., Masci, P. P., Hanchard, K. J., Lewis, R. J., Alewood, P. F., de Jersey, J., & Lavin, M. F. (2006). Cloning and characterisation of natriuretic peptides from the venom glands of Australian elapids. *Biochimie*, *88*(12), 1923–1931. https://doi.org/10.1016/j.biochi.2006.06.014

St Pierre, L., Masci, P. P., Filippovich, I., Sorokina, N., Marsh, N., Miller, D. J., & Lavin, M. F. (2005). Comparative analysis of prothrombin activators from the venom of Australian elapids. *Molecular Biology and Evolution*, *22*(9), 1853–1864. https://doi.org/10.1093/molbev/msi181

St Pierre, L., Woods, R., Earl, S., Masci, P. P., & Lavin, M. F. (2005). Identification and analysis of venom gland-specific genes from the coastal taipan (*Oxyuranus scutellatus*) and related species. *Cellular*

and Molecular Life Sciences, *62*, 2679–2693. https://doi.org/10.1007/s00018-005-5384-9

Streicher, J. W., & Wiens, J. J. (2016). Phylogenomic analyses reveal novel relationships among snake families. *Molecular Phylogenetics and Evolution*, *100*, 160–169. https://doi.org/10.1016/j.ympev.2016.04.015

Strickland, J. L., Carter, S., Kraus, F., & Parkinson, C. L. (2016). Snake evolution in Melanesia: Origin of the Hydrophiinae (Serpentes, Elapidae), and the evolutionary history of the enigmatic new Guinean elapid *Toxicocalamus*. *Zoological Journal of the Linnean Society*, *178*(3), 663–678. https://doi.org/10.1111/zoj.12423

Sutherland, S. K., & Tibbals, J. (2001). *Australian animal toxins: The creatures, their toxins and care of the poisoned patient*. Oxford University Press.

Tan, C. H. (2022). Snake venomics: Fundamentals, recent updates, and a look to the next decade. *Toxins*, *14*(4), 247. https://doi.org/10.3390/toxins14040247

Tasoulis, T., & Isbister, G. K. (2017). A review and database of snake venom proteomes. *Toxins*, *9*(9), 290. https://doi.org/10.3390/toxins9090290

Tasoulis, T., & Isbister, G. K. (2023). A current perspective on snake venom composition and constituent protein families. *Archives of Toxicology*, *97*(1), 133–153. https://doi.org/10.1007/s00204-022-03420-0

Tewhey, R., Nakano, M., Wang, X., Pabón-Peña, C., Novak, B., Giuffre, A., Lin, E., Happe, S., Roberts, D. N., LeProust, E. M., Topol, E. J., Harismendy, O., & Frazer, K. A. (2009). Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biology*, *10*(10), R116. https://doi.org/10.1186/gb-2009-10-10-r116

Uetz, P. (2024). The reptile database. http://reptile-database.reptarium.cz

Welton, R. E., & Burnell, J. N. (2005). Full length nucleotide sequence of a factor V-like subunit of oscutarin from *Oxyuranus scutellatus scutellatus* (coastal taipan). *Toxicon*, *46*(3), 328–336. https://doi.org/10.1016/j.toxicon.2005.05.001

Weinell, J. L., Burbrink, F. T., Das, S., & Brown, R. M. (2024). Novel phylogenomic inference and 'Out of Asia' biogeography of cobras, coral snakes and their allies. *Royal Society Open Science*, *11*, 240064. http://doi.org/10.1098/rsos.240064

Zhang, Z. Y., Lv, Y., Wu, W., Yan, C., Tang, C. Y., Peng, C., & Li, J. T. (2022). The structural and functional divergence of a neglected three-finger toxin subfamily in lethal elapids. *Cell Reports*, *40*(2), 111079. https://doi.org/10.1016/j.celrep.2022.111079

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Travers, S. L., Hutter, C. R., Austin, C. C., Donnellan, S. C., Buehler, M. D., Ellison, C. E., & Ruane, S. (2024). VenomCap: An exon-capture probe set for the targeted sequencing of snake venom genes. *Molecular Ecology Resources*, *24*, e14020. https://doi.org/10.1111/1755-0998.14020