The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system

Freek J. Vonk[a,b,c,1], Nicholas R. Casewell[c,d,1], Christiaan V. Henkel[b,e], Alysha M. Heimberg[f], Hans J. Jansen[e], Ryan J. R. McCleary[g], Harald M. E. Kerkkamp[b], Rutger A. Vos[a], Isabel Guerreiro[h], Juan J. Calvete[i], Wolfgang Wüster[c], Anthony E. Woods[j], Jessica M. Logan[j], Robert A. Harrison[d], Todd A. Castoe[k,l], A. P. Jason de Koning[k,m], David D. Pollock[k], Mark Yandell[n], Diego Calderon[n], Camila Renjifo[d], Rachel B. Currier[d], David Salgado[f,o], Davinia Pla[i], Libia Sanz[i], Asad S. Hyder[b], José M. C. Ribeiro[p], Jan W. Arntzen[a], Guido E. E. J. M. van den Thillart[e], Marten Boetzer[q], Walter Pirovano[q], Ron P. Dirks[e], Herman P. Spaink[b,e], Denis Duboule[h], Edwina McGlinn[f], R. Manjunatha Kini[g], and Michael K. Richardson[b,2]

[a]Naturalis Biodiversity Center, 2333 CR, Leiden, The Netherlands; [b]Institute of Biology Leiden, Leiden University, Sylvius Laboratory, Sylviusweg 72, 2300 RA, Leiden, The Netherlands; [c]Molecular Ecology and Evolution Group, School of Biological Sciences, Bangor University, Bangor LL57 2UW, United Kingdom; [d]Alistair Reid Venom Research Unit, Liverpool School of Tropical Medicine, Liverpool L3 5QA, United Kingdom; [e]ZF-Screens B.V., Bio Partner Center, 2333 CH, Leiden, The Netherlands; [f]European Molecular Biology Laboratory Australia, Australian Regenerative Medicine Institute, Monash University, Clayton, 3800, Australia; [g]Department of Biological Sciences, National University of Singapore, Singapore 117543; [h]Department of Genetics and Evolution, University of Geneva, CH-1211 Geneva 4, Switzerland; [i]Instituto de Biomedicina de Valencia, Consejo Superior de Investigaciones Cientificas (Spain), 11 46010 Valencia, Spain; [j]School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, SA 5000, Australia; [k]Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045; [l]Department of Biology, University of Texas, Arlington, TX 76010; [m]Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Calgary and Alberta Children's Hospital Research Institute for Child and Maternal Health, Calgary, AB, Canada T2N 4N1; [n]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT; [o]Aix-Marseille Université, Inserm, GMGF UMR_S910, 13385 Marseille, France; [p]Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD 20892; and [q]BaseClear B.V., 2333 CC, Leiden, The Netherlands

Edited by David B. Wake, University of California, Berkeley, CA, and approved October 22, 2013 (received for review August 2, 2013)

Snakes are limbless predators, and many species use venom to help overpower relatively large, agile prey. Snake venoms are complex protein mixtures encoded by several multilocus gene families that function synergistically to cause incapacitation. To examine venom evolution, we sequenced and interrogated the genome of a venomous snake, the king cobra (*Ophiophagus hannah*), and compared it, together with our unique transcriptome, microRNA, and proteome datasets from this species, with data from other vertebrates. In contrast to the platypus, the only other venomous vertebrate with a sequenced genome, we find that snake toxin genes evolve through several distinct co-option mechanisms and exhibit surprisingly variable levels of gene duplication and directional selection that correlate with their functional importance in prey capture. The enigmatic accessory venom gland shows a very different pattern of toxin gene expression from the main venom gland and seems to have recruited toxin-like lectin genes repeatedly for new nontoxic functions. In addition, tissue-specific microRNA analyses suggested the co-option of core genetic regulatory components of the venom secretory system from a pancreatic origin. Although the king cobra is limbless, we recovered coding sequences for all *Hox* genes involved in amniote limb development, with the exception of *Hoxd12*. Our results provide a unique view of the origin and evolution of snake venom and reveal multiple genome-level adaptive responses to natural selection in this complex biological weapon system. More generally, they provide insight into mechanisms of protein evolution under strong selection.

genomics | phylogenetics | serpentes

Snake venom contains biologically active proteins (toxins) encoded by several multilocus gene families that each comprise several distinct isoforms (1, 2). Venom is produced in a postorbital venom gland (3) and associated in elapids (cobras and their relatives) and viperids (vipers and pit vipers) with a small downstream accessory gland of unknown function (Fig. 1). Understanding the origin and evolution of the snake venom system is not only of great intrinsic biological interest (3–5), but is also important for drug discovery (1, 2, 6), understanding vertebrate physiological pathways (7, 8), and addressing public health concerns about the enormous number of snake bites suffered in tropical countries (9, 10).

Significance

Snake venoms are toxic protein cocktails used for prey capture. To investigate the evolution of these complex biological weapon systems, we sequenced the genome of a venomous snake, the king cobra, and assessed the composition of venom gland expressed genes, small RNAs, and secreted venom proteins. We show that regulatory components of the venom secretory system may have evolved from a pancreatic origin and that venom toxin genes were co-opted by distinct genomic mechanisms. After co-option, toxin genes important for prey capture have massively expanded by gene duplication and evolved under positive selection, resulting in protein neofunctionalization. This diverse and dramatic venom-related genomic response seemingly occurs in response to a coevolutionary arms race between venomous snakes and their prey.

Author contributions: F.J.V., N.R.C., and M.K.R. designed research; F.J.V. acquired samples for sequencing and estimated genome size; H.J.J., and R.P.D. prepared sequencing libraries; M.B., and W.P. developed assembly software; C.V.H. assembled the genome; H.J.J., M.Y., D.C., and H.P.S. annotated the genome; J.M.C.R. assembled and annotated RNA-seq libraries; F.J.V., N.R.C., H.M.E.K., and A.S.H. analyzed RNA-seq libraries; A.M.H., D.S., and E.M. annotated and analyzed small RNA libraries; H.M.E.K., I.G., H.P.S., and D.D. annotated and analyzed Hox genes; F.J.V., N.R.C., C.V.H., R.J.R.M., H.M.E.K., A.S.H., R.P.D., R.M.K., and M.K.R. annotated venom toxin genes and performed synteny analyses; N.R.C., R.A.V., and W.W. analyzed gene family evolution; A.M.H. performed miRNA in situ hybridization; A.E.W., and J.M.L. performed lectin in situ hybridization; N.R.C., R.J.R.M., J.J.C., R.A.H., C.R., R.B.C., D.P., L.S., and R.M.K. analyzed the venom proteome; T.A.C., A.P.J.d.K., and D.D.P. contributed Burmese python genome data and assisted with comparative analyses; H.J.J., J.W.A., G.E.E.J.M.v.d.T., R.P.D., H.P.S., and M.K.R. organized sequencing platforms and facilities; F.J.V., N.R.C., W.W., and M.K.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The king cobra genome assembly and reads reported in this paper have been deposited in the GenBank database (bioproject no. PRJNA201683). The transcriptome sequences reported in this paper have been deposited in the GenBank Short Read Archive database (bioproject no. PRJNA222479). The microRNA sequences reported in this paper have been deposited in miRBase, www.mirbase.org.

[1]F.J.V. and N.R.C. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: M.K.Richardson@biology.leidenuniv.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1314702110/-/DCSupplemental.

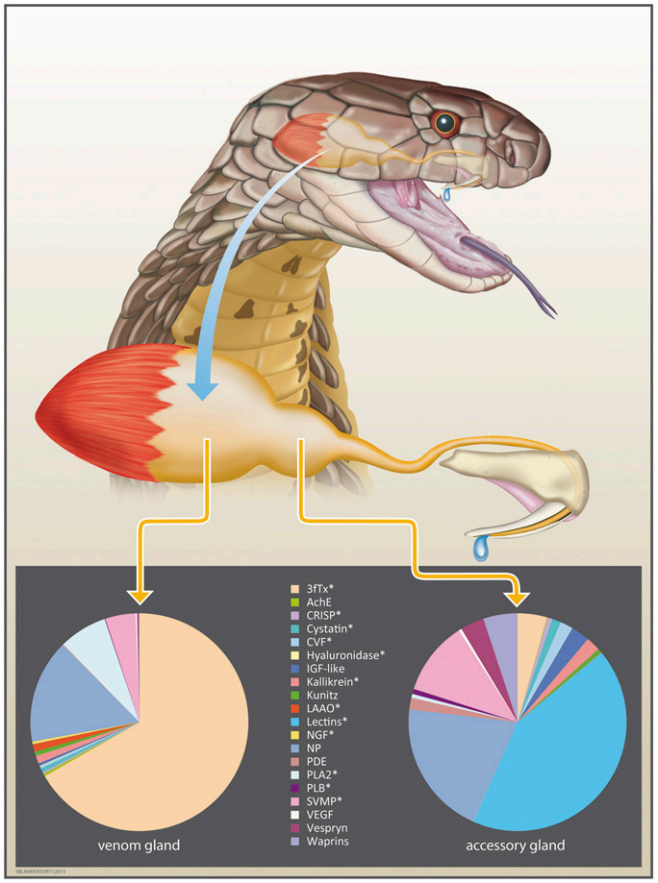

Fig. 1. The king cobra venom system with venom and accessory gland expression profiles. Pie charts display the normalized percentage abundance of toxin transcripts recovered from each tissue transcriptome. Three-finger toxins are the most abundant toxin family in the venom gland (66.73% of all toxin transcripts and 4.37% in the accessory gland), and they are represented in the genome by at least 21 loci. Lectins are the most abundant toxin family in the accessory gland (42.70% of all toxin transcripts and 0.03% in the venom gland), and they are represented in the genome by at least six loci. Asterisks indicate toxin gene families annotated in the genome. 3FTx, three-finger toxin; AchE, acetylcholinesterase; CRISP, cysteine-rich secretory protein; CVF, cobra venom factor; IGF-like, insulin-like growth factor; kallikrein, kallikrein serine proteases; kunitz, kunitz-type protease inhibitors; LAAO, L-amino acid oxidase; NGF, nerve growth factor; PDE, phosphodiesterase; PLA2, phospholipase A_2; PLB, phospholipase-B; SVMP, snake venom metalloproteinase. Drawing made based on a photo by F.J.V.

The birth and death model of gene evolution is the canonical framework used to explain the evolutionary origin of snake venom toxins. Drivers of toxin diversification may include (*i*) directional selection for toxins that facilitate prey capture, (*ii*) the need to target a diversity of receptors in different prey, and (*iii*) the concomitant evolution of venom resistance in some prey as part of an evolutionary arms race (2). The lack of genome sequences for any venomous snake and the consequent dependence on transcriptome data have hampered our understanding of not only the tempo and mode of venom toxin evolution but also, the genomic mechanisms that regulate toxin–gene expression.

To address these issues, we have produced a draft genome of a venomous snake—that of an adult male Indonesian king cobra (*Ophiophagus hannah*). This iconic species is the longest venomous snake in the world. Native to tropical Asia, it feeds on other snakes, and it is a member of the family Elapidae. We also deep-sequenced transcriptomes and small RNAs of the venom gland, the accessory gland, and a pooled, multitissue archive and characterized the king cobra venom proteome. These unique datasets provide an unprecedented insight into the evolution of venom.

Results and Discussion

King cobra genome sequence data (*SI Appendix*, Table S1) were first assembled de novo into contigs, which were subsequently oriented and merged into scaffolds. Haploid genome size was estimated by flow cytometry to be 1.36–1.59 Gbp (*SI Appendix*, Fig. S1). The assembled draft genome has an N50 contig size of 3.98 Kbp and an N50 scaffold size of 226 Kbp. The total contig length is 1.45 Gbp, and the total scaffold length (which contains gaps) is 1.66 Gbp.

As a genome quality check, we examined the *Hox* cluster, because it is well-characterized in other vertebrates (11). We annotated all 39 *Hox* genes, which we found clustered at four genomic regions, like in other vertebrates. However, the gene clusters are substantially larger than the *Hox* clusters observed in mammals (*SI Appendix*, Fig. S2). Of special interest is the absence of *Hoxd12* from the king cobra, the Burmese python (*Python molurus bivittatus*) (12), and other snake genomes (13) (*SI Appendix*, Fig. S3). *Hoxd12* is important for limb development in tetrapods (11) and thus, may have been lost along with limbs before the snake diversification. We also mapped microRNAs that had been previously located within mammalian and avian *Hox* clusters (*SI Appendix*, Fig. S2 and Dataset S1).

We interrogated the king cobra genome and annotated the open reading frames of 12 venom toxin gene families (Fig. 1 and *SI Appendix*, Fig. S4). Venom toxins are thought to have been co-opted from gene homologs with nontoxic physiological functions that are expressed in tissues other than the venom gland (14, 15). Our analysis of tissue-specific transcriptomic data (12, 16–18) provides genome-scale confirmation that these venom genes have, indeed, been recruited from a wide variety of tissue types (*SI Appendix*, Table S2). Syntenic comparisons of king cobra genomic architecture with the genomes of other vertebrates revealed that toxin co-option has occurred by two distinct mechanisms: (*i*) gene hijacking/modification and (*ii*) duplication of nontoxin genes (*SI Appendix*, Fig. S5); they were followed in both cases by selective expression in the venom gland.

Sequencing and analysis of microRNA (miRNA) libraries made from a range of different tissues showed molecular similarities between the king cobra venom gland and known profiles of human and mouse pancreas (Fig. 2*A*). The most abundant miRNA in our venom gland library is *miR-375*, a canonical miRNA in the vertebrate pancreas. In the mouse, chicken, and zebrafish, *miR-375* expression is restricted to the pancreas and pituitary gland (19, 20). Here, we detected *miR-375* expression in the embryonic pancreas of the copperhead ratsnake (*Coelognathus radiatus*), the islet cell masses associated with the pancreas and spleen of the spitting cobra (*Naja siamensis*), and importantly, the venom gland of the king cobra (Fig. 2 *B–D* and *SI Appendix*, Fig. S6). In the past, it has been hypothesized that the snake venom gland evolved by evolutionary modification of the pancreatic system (21–23), although this hypothesis has since been abandoned, because little evidence exists that toxins expressed in the venom gland have been co-opted from related proteins expressed in the pancreas (14). However, our results are consistent with *miR-375* being part of a core genetic network regulating secretion that has been co-opted during the evolution of the snake venom gland from an ancestral role in the pancreas and foregut secretory cells (24); it highlights an inherent link between these two secretory tissues, a link which was first suggested by Kochva et al. (21–23).

We identified 20 toxin families in the king cobra venom gland transcriptome (Fig. 1 and Dataset S2), including all toxin families annotated in the genome. Of the transcriptome hits, 14 toxin families were identified in the venom proteome (*SI Appendix*,

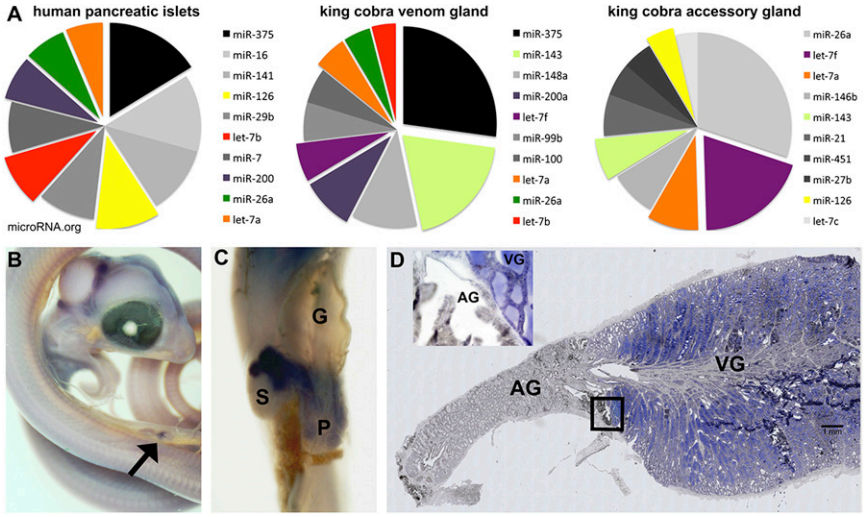

Fig. 2. MiRNA expression profiles of the king cobra venom gland and accessory gland and miRNA expression patterns by in situ hybridization. (*A*) The 10 most abundant miRNAs in the venom gland show similarities with the known expression profile of the vertebrate pancreas (shown here for human; microRNA.org). (*B*) In situ hybridization of *miR-375* in a *C. radiatus* embryo 27 d postoviposition with expression detected in the pancreas (arrow). (*C*) In situ hybridization of *miR-375* in an *N. siamensis* embryo 32 d postoviposition, showing expression in the islet cell masses of the pancreas and the intrasplenic islet tissue. (*D*) In situ hybridization of *miR-375* in a tissue section of the venom system of an adult *O. hannah* showing expression in the main venom gland. (*Inset*) Boundary of the venom gland (expression) and accessory gland (no expression) (*SI Appendix*, Fig. S6). AG, accessory gland; G, gallbladder; P, pancreas; S, spleen; VG, venom gland.

Figs. S7–S9 and Tables S3 and S4 and Dataset S3), and nerve growth factor, phospholipase-B, and cobra venom factor have not previously been reported in king cobra venom. We also identified a unique snake venom protein, insulin-like growth factor, which we found selectively expressed in the venom gland and the venom proteome (*SI Appendix*, Fig. S10 and Table S4). Recent findings have shown adaptive evolution in insulin-like growth factor genes in snakes, although the site of their expression was unknown (25). Evidence of selective venom gland expression combined with adaptive evolution is consistent with a function of these proteins as venom toxins. In addition, we discovered a unique independent recruitment event of L-amino acid oxidase into king cobra venom (*SI Appendix*, Fig. S11).

Comparisons of toxin expression in the venom gland, accessory gland, and pooled multitissue archive revealed that most toxins are expressed at high levels only in the venom gland (Fig. 1 and *SI Appendix*, Fig. S10). Our results indicate that toxin gene transcription in the venom gland is regulated independently from its expression in those other tissues. Most toxins observed in the venom gland transcriptome are expressed at low levels in the accessory gland. One exception was the lectin toxin family, with expression that was at least 40 times higher in the accessory gland (*SI Appendix*, Fig. S10). Our evolutionary analysis of the lectins shows that they have been recruited to the oral secretory glands before the radiation of the advanced snakes, followed by expansion of the gene family (*SI Appendix*, Figs. S12 and S13). Our king cobra data suggest a model in which venom-like lectin paralogs have then repeatedly become transcriptionally activated in the accessory gland and deactivated in the venom gland (*SI Appendix*, Figs. S12 and S13).

In situ hybridization showed that the expression of these recruited lectins is concentrated in the serous cells located in the proximal region (26) of the accessory gland (Fig. 3 and *SI Appendix*, Fig. S14). No lectins were detected in the king cobra venom proteome (*SI Appendix*, Figs. S7–S9), consistent with their low transcript abundance in the venom gland. These results suggest that lectins do not contribute to king cobra envenoming, which is in contrast to many other venomous snakes (1, 27), and that their repeated recruitment to the accessory gland is associated with the subsequent evolution of unidentified, nontoxic functions (15).

The venom gland transcriptome and venom proteome revealed multiple related venom isoforms for many different toxin families. To investigate the role of gene duplication in driving the genomic expansion of venom genes, we examined the evolutionary history of nine different toxin families by comparing gene orthologs and paralogs from other venomous snakes and the Burmese python, and green anole lizard (*Anolis carolinensis*) genomes and tissue transcriptomes (12, 16–18) (*SI Appendix*, Figs. S11 and S15–S22). We then used these data to perform tests of directional selection. Our results reveal multiple distinct patterns of gene duplication and sequence evolution under positive selection in different protein-coding gene families both before and after their recruitment into venom-producing pathways (Fig. 4 *A* and *B* and *SI Appendix*, Table S5). Significantly, we found evidence of higher rates of duplication and selection in the most highly expressed,

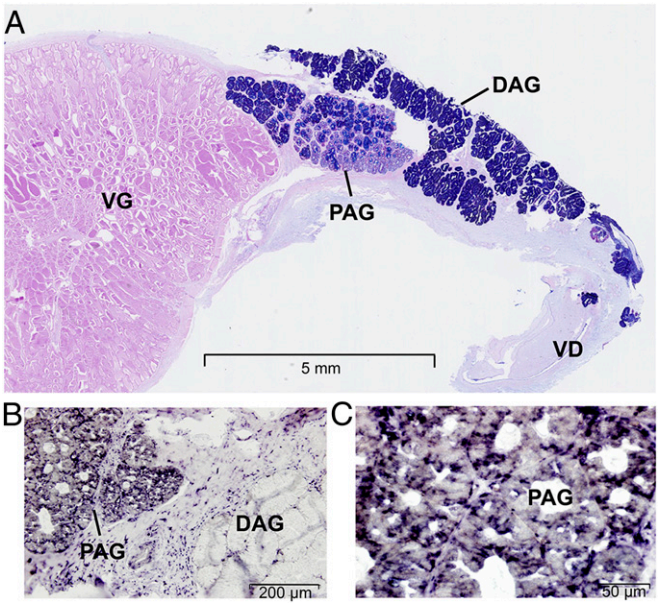

Fig. 3. Histological section of the complete venom apparatus of the king cobra and spatial expression of lectin genes in the accessory gland. (*A*) Longitudinal section of the venom system reveals the two regions of the accessory gland: the proximal portion (PAG) and the distal portion (DAG; consistent with a previous morphological study) (26). The venom system is stained by alcian blue and periodic acid–Schiff, in which the secretory epithelial cells and secretion of the venom gland are periodic acid–Schiff-positive and the seromucous acini of the PAG and the mucous acini comprising the DAG are stained with alcian blue. (*B*) In situ hybridization of lectin gene Oh-516 (genome ID s8808 gene 2) shows that lectin expression is restricted to the PAG. DAG shows no staining. (*C*) Detail of the PAG shown in *B* showing strong granular staining in the epithelium of the PAG (*SI Appendix*, Fig. S14). VD, venom duct; VG, venom gland.

proteomically abundant, and functionally important (28) gene families analyzed. The major lethal toxin family of the king cobra, the three-finger toxins (28, 29), is the most abundantly represented and isomerically diverse toxin family found in the venom gland transcriptome and venom proteome (Fig. 1 and *SI Appendix*, Fig. S7). This family has undergone massive expansion and shows high levels of positive selection and gene duplication (Fig. 4 *B* and *C*). In addition, phospholipase A_2, snake venom metalloproteinase, and kallikrein toxin families also exhibit substantial gene duplication (Fig. 4*D*), and evidence of positive selection was identified in two of these gene families (Fig. 4*B*).

Gene duplication coupled with positive selection is the mechanism underlying venom protein neofunctionalization (30–34). Our results are, therefore, consistent with a prominent role for prey-driven natural selection in generating the genetic diversity of the most pathogenic toxin families (28). By contrast, toxin families with ancillary functions show lower levels of gene expression, little to no evidence of gene duplication, and no evidence of directional selection (Fig. 4 *B* and *E*). For example, hyaluronidase, which possibly functions to break down prey tissue at the envenomation site (35), is not under positive selection. These results suggest that ancillary venom genes are less likely to generate resistance in prey and therefore, likely to experience lower selection pressures. These gene families likely have conserved functional activities and do not participate in the evolutionary arms race seen in the more toxic venom protein families.

In conclusion, this study highlights the diversity of genomic responses to extrinsic selective factors (the imperative to overpower prey quickly). These responses include function-modulated patterns of transcript abundance, gene duplication, and protein evolution in different toxin families. In contrast with the only other venomous vertebrate genome sequenced to date [the platypus (36, 37)], gene duplication is apparently of fundamental importance in the adaptive evolution of the king cobra venom system. This distinction likely reflects the differences in selective pressures relating to the very different biological role of venom in these organisms. Platypus venom is implicated in male–male combat, and its evolution is driven by sexual selection, whereas snake venom is primarily used for predatory purposes. The requirement of snake venom to rapidly immobilize prey coupled with the concomitant evolution of resistance in some prey species apparently results in an evolutionary arms race that drives a diverse and dramatic genomic response in venomous snakes. Our study provides unique genome-wide perspectives on the

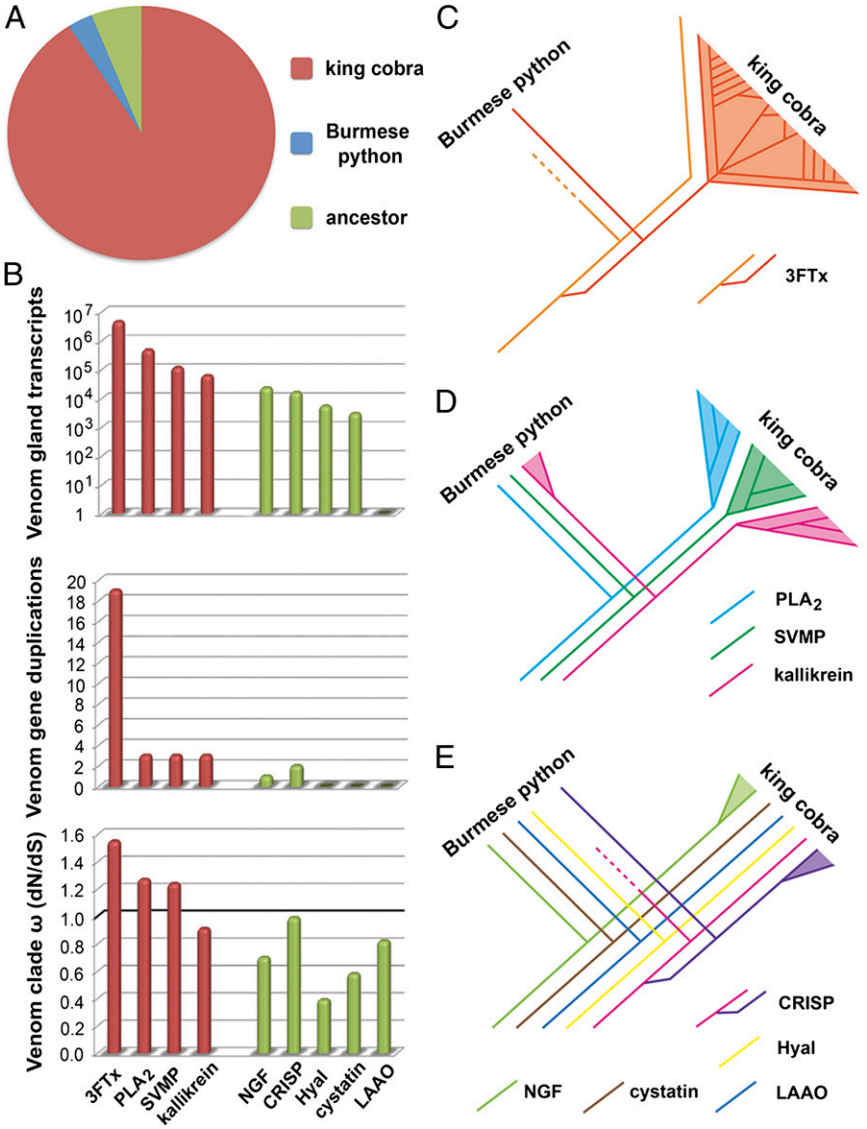

Fig. 4. Contrasting evolutionary histories of king cobra toxin gene families. (*A*) The vast majority of toxin family gene duplication events occurred in the king cobra lineage compared with the Burmese python and their common ancestor. (*B*) Comparisons of venom gland expression, venom-related gene duplication events, and rate of evolution of main toxin families (red) and ancillary toxin families (green). (*C*) Massive expansion of the three-finger toxin gene family and (*D*) moderate expansion of other pathogenic toxin families by duplication of venom-expressed genes after the split of the Burmese python from the advanced snakes. (*E*) Ancillary toxin families show reduced evidence of gene duplication. Colored lines indicate gene loci, with line splits representing gene duplication events and dotted lines indicating gene loss. Venom gene duplications are defined as duplications that occurred after the split of the Burmese python from the advanced snakes (king cobra). ω represents the dN/dS ratio identified for venomous gene clades. The boundary for directional selection is indicated by a bold line. Note the logarithmic scale in the normalized venom gland expression graph. 3FTx, three-finger toxin; CRISP, cysteine-rich secretory protein; Hyal, hyaluronidase; kallikrein, kallikrein serine proteases; LAAO, L-amino acid oxidase; NGF, nerve growth factor; PLA2, phospholipase A_2; SVMP, snake venom metalloproteinase.

adaptive evolution of such venom systems as well as to protein evolution in general, and thus, it contributes an essential foundation for understanding and comparing evolutionary genomic processes in venomous organisms.

Methods

SI Appendix, SI Materials and Methods has additional information relating to the methodologies described below.

Tissue Acquisition and Processing. All animal procedures complied with local ethical guidelines. Genome sequencing was undertaken on a blood sample obtained from an adult male king cobra that originated in Bali, Indonesia. Venom was extracted, and 4 d later (to maximize mRNA production), the venom gland, accessory gland, and other tissue samples were sourced from a second Indonesian adult male specimen and stored in RNAlater.

Genome Sequencing. We used a whole-genome shotgun sequencing strategy and Illumina sequencing technology. Genomic DNA was isolated from blood using the Qiagen Blood and Tissue DNeasyKit and paired-end libraries prepared from 5 μg isolated gDNA using the Illumina Paired-End Sequencing Sample Prep Kit. Either a 200- or 500-bp band was cut from the gel (library PE200 or PE500, respectively) (*SI Appendix*, Table S1). Similarly, mate pair libraries were prepared from 10 μg isolated gDNA using the Illumina Mate Pair 2–5 Kb Sample Prep Kit and bands from 2 to 15 Kbp cut from the gel (MP2K, MP7K, MP10K, and MP15K libraries) (*SI Appendix*, Table S1). After circularization, shearing, isolation of biotinylated fragments, and amplification, the 400- to 600-bp fraction of the resulting fragments was isolated from the gel. Genomic libraries were paired-end sequenced with a read length of 36–151 nt on an Illumina GAIIx instrument.

Genome Assembly. For genome assembly, we largely followed the strategy pioneered in the work by Li et al. (38) for the assembly of the giant panda genome. Sequencing reads from both paired-end libraries were first used for building initial contigs. Both sets were preprocessed to eliminate low-quality reads and nucleotides as well as adapter contamination. For initial contig assembly, we used the CLC Assembly Cell De Novo Assembler (version 3.2; CLC Bio, Aarhus, Denmark), which implements a De Bruijn graph-based assembler. A run with a minimum-required contig size of 100 bp and a k-mer length of 31 nt resulted in an assembly with a total length of 1.45 Gbp and a contig N50 of 3,982 bp [i.e., 50% of the assembly (725 Mbp) is in contigs of at least this length]. Initial contigs were subsequently oriented into larger supercontigs (scaffolds) using SSPACE (39). SSPACE aligns paired reads to the contigs using Bowtie (40). SSPACE was used to scaffold contigs in a hierarchical fashion using first links obtained from the PE500 library to generate intermediate supercontigs, which were then used as the input for subsequent runs, with links from individual mate-pair libraries increasing in size. At each stage, a minimum of three nonredundant links was required to join two contigs. This procedure resulted in a final scaffold set with a total length of 1.66 Gbp and an N50 of 225,511 bp.

Genome Annotation. Automated gene prediction was undertaken using the automated annotation pipeline MAKER (41, 42). Gene annotations were made using a protein database combining the Uniprot/Swiss-Prot protein database and all king cobra and green anole (*A. carolinensis*) sequences from the National Center for Biotechnology Information protein database. Ab initio gene predictions were created by MAKER using the programs SNAP (43) and Augustus (44). Gene models were further improved by providing MAKER with all king cobra mRNAseq data generated in this study, which were combined to generate a joint assembly of transcripts using Trinity (45). A total of three iterative runs of MAKER was used to produce the final gene set. Additional extensive manual annotation was performed to establish the intron–exon boundaries of members of venom toxin gene families.

mRNA-Seq and Small RNA Libraries. King cobra tissue sequencing libraries were prepared for the venom gland, accessory gland, and a pooled multitissue archive (heart, lung, spleen, brain, testes, gall bladder, pancreas, small intestine, kidney, liver, eye, tongue, and stomach). Total RNA was isolated from each tissue using the Qiagen miRNeasy Kit. Transcriptome libraries were subsequently prepared from 10 μg total RNA (using equal amounts of RNA isolated from each tissue for the pooled multitissue archive) using the Illumina mRNA-Seq Sample Preparation Kit. Total RNA from the same samples was used to prepare the small RNA libraries using the Illumina small RNA v1.5 Sample Preparation Kit. RNAseq and small RNA libraries were sequenced on the Illumina GAIIx sequencing platform.

Transcriptome Assembly. Reads for the venom gland, accessory gland, and pooled multitissue archive were coassembled with Abyss (46, 47) with various k values (every even number from 50 to 96). The resulting assemblies were joined by an iterative BLAST and cap3 assembler (48). Coding sequences were extracted using an automated pipeline based on similarities to known proteins or by obtaining coding sequences from the larger ORF of the contigs containing a signal peptide. To map the raw Illumina reads to the coding sequences and determine their tissue bias, raw reads from each library were blasted to the coding sequences using blastn with a word size of 25 (–W 25 switch) and allowing recovery of up to three matches. The three matches were used if they had less than two gaps and their scores were equal to the best score. The resulting blast file was used to compile the number of reads each coding DNA sequence received from each library.

miRNA Profiles and in Situ Hybridizations. The small RNA sequences were analyzed using CLC Bio Genome Workbench. Briefly, small RNA sequences were filtered for quality and size, and reads of low quality and lengths less than 17 or greater than 26 nt were discarded. The remaining pool of small RNAs was compared with miRBase release 18 (http://www.miRBase.org) to extract orthologous mature miRNA sequences from each king cobra RNA sample. These miRNAs were subsequently mapped to the king cobra genome, with 70 bp upstream and downstream of the mature sequence extracted as the potential precursor miRNA sequence using PHP scripts and blast (49) 2.2.26+. The expression level of each miRNA was assessed using CLC Bio and compared with data available at the miRNA targets and expression database (http://www.microRNA.org; release August 2010) for the expression profiles of orthologous miRNA genes in mouse and human (e.g., *miR-375*). Whole-mount in situ hybridizations for *miR-375* detection were performed using 5′ digoxigenin-labeled locked nucleic acid (LNA; Exiqon) probes following the protocol in the work by Darnell et al. (19). The standard tissue section in situ protocol in the work by Jostarndt et al. (50) for paraffin-embedded tissues was followed for *miR-375* detection in the adult king cobra venom gland. For whole-mount in situ hybridizations in late-stage snake embryos (27 d postoviposition or older), embryos were skinned, and the abdominal wall was cut open followed by an extended probe hybridization for ~36 h. All *miR-375* LNA in situ hybridizations were carried out at 57 °C (22°C below the calculated probe melting temperature of 79°C) along with a no-probe control. *miR-196* LNA in situ was carried out at 47 °C as an additional negative control in the adult venom gland.

Venom Proteomics. We used king cobra venom extracted from the same animal used for transcriptomics. The venom was reduced, alkylated, digested with trypsin, separated by column chromatography, and analyzed by ESI-ion trap tandem MS. The peptide fragments created by collision-induced dissociation were compared against the assembled king cobra venom gland and accessory gland transcriptomes and a Lepidosaurian (National Center for Biotechnology Information) database using Sequest and Mascot software with a false discovery rate of 0.01.

Evolutionary Analyses. King cobra sequences exhibiting homology to toxin families were identified through (*i*) annotation in the genome or transcriptome and (*ii*) blast searching the king cobra genome and transcriptome datasets in CLC Main Workbench with representative templates of toxin and nontoxin gene homologs. Coding regions of identified toxin gene loci were aligned using the MUSCLE algorithm (51) with putative paralogs and orthologs from selected vertebrates, including other venomous snakes and the *P. molurus bivittatus* and *A. carolinensis* genomes and transcriptomes (12, 16–18). These sequences were obtained by mining GenBank for blast hits and using the datasets in work by Casewell et al. (15).

DNA gene trees for each toxin family were reconstructed using Bayesian inference in MrBayes v3.2 (52) incorporating optimized models of sequence evolution selected by MrModelTest v2.3 (53). Each dataset was run in duplicate using four chains for 5×10^6 generations, sampling every 500th cycle from the chain, and using default settings in regards to priors. Tracer v1.4 (54) was used to estimate effective sample sizes for all parameters and verify the point of convergence (burnin), with trees generated before the completion of burnin discarded. The locations of gene expression of snake sequences determined by transcriptomics were mapped on the gene trees to visualize relative expression in different tissue types. Toxin family gene duplication events were inferred by pruning the gene trees to only contain king cobra and Burmese python genes along with a single outgroup sequence. The ensuing gene trees were analyzed using the duplication and loss criterion in iGTP (55) with the following species tree: [outgroup (king cobra, Burmese python)]. For tests of directional selection, we inferred fully resolved maximum likelihood trees from each of the toxin family datasets using the BEST tree-searching algorithm in PHYML (56). The most parsimonious

points of recruitment into venom-producing pathways were then reconstructed on these trees, thereby classifying tree branches into venomous and nonvenomous. The method of Yang and Nielsen (57) was implemented in the PAML software package to estimate $\omega_{venomous}$ and $\omega_{nonvenomous}$ for each toxin family.

ACKNOWLEDGMENTS. We thank the following persons who helped us or contributed material used in this study: Austin Hughes, Nathan Dunstan, Daniëlle de Wijze, and Youri Lammers. We thank Bas Blankevoort for constructing Fig. 1. This work received funding from the following sources: internal funding from the Naturalis Biodiversity Center (F.J.V., J.W.A., and M.K.R.), a Rubicon Grant from the Netherlands Organization for Scientific Research (to F.J.V.), a research fellowship from the United Kingdom Natural Environment Research Council (to N.R.C.), an Netherlands Organization for Scientific Research Visitor's Travel Grant from Nederlandse Organisatie voor Wetenschappelijk Onderzoek (to R.J.R.M., R.M.K., and M.K.R.), a studentship from the United Kingdom Biotechnology and Biological Sciences Research Council (to R.B.C.), and a Smart Mix Grant from the Dutch Government (to M.K.R.).

1. Vonk FJ, et al. (2011) Snake venom: From fieldwork to the clinic: Recent insights into snake biology, together with new technology allowing high-throughput screening of venom, bring new hope for drug discovery. *Bioessays* 33(4):269–279.
2. Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG (2013) Complex cocktails: The evolutionary novelty of venoms. *Trends Ecol Evol* 28(4):219–229.
3. Vonk FJ, et al. (2008) Evolutionary origin and development of snake fangs. *Nature* 454(7204):630–633.
4. Fry BG, et al. (2006) Early evolution of the venom system in lizards and snakes. *Nature* 439(7076):584–588.
5. Saviola AJ, Chiszar D, Busch C, Mackessy SP (2013) Molecular basis for prey relocation in viperid snakes. *BMC Biol* 11:20.
6. Lewis RJ, Garcia ML (2003) Therapeutic potential of venom peptides. *Nat Rev Drug Discov* 2(10):790–802.
7. Bohlen CJ, et al. (2011) A heteromeric Texas coral snake toxin targets acid-sensing ion channels to produce pain. *Nature* 479(7373):410–414.
8. Diochot S, et al. (2012) Black mamba venom peptides target acid-sensing ion channels to abolish pain. *Nature* 490(7421):552–555.
9. Kasturiratne A, et al. (2008) The global burden of snakebite: A literature analysis and modelling based on regional estimates of envenoming and deaths. *PLoS Med* 5(11): e218.
10. Mohapatra B, et al. (2011) Snakebite mortality in India: A nationally representative mortality survey. *PLoS Negl Trop Dis* 5(4):e1018.
11. Zákány J, Kmita M, Duboule D (2004) A dual role for Hox genes in limb anterior-posterior asymmetry. *Science* 304(5677):1669–1672.
12. Castoe TA, et al. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc Natl Acad Sci USA* 110:20645–20650.
13. Di-Poï N, et al. (2010) Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature* 464(7285):99–103.
14. Fry BG (2005) From genome to "venome:" Molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res* 15(3):403–420.
15. Casewell NR, Huttley GA, Wüster W (2012) Dynamic evolution of venom proteins in squamate reptiles. *Nat Commun* 3:1066.
16. Alföldi J, et al. (2011) The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 477(7366):587–591.
17. Castoe TA, et al. (2011) Sequencing the genome of the Burmese python (*Python molurus bivittatus*) as a model for studying extreme adaptations in snakes. *Genome Biol* 12(7):406.
18. Eckalbar WL, et al. (2013) Genome reannotation of the lizard *Anolis carolinensis* based on 14 adult and embryonic deep transcriptomes. *BMC Genomics* 14:49.
19. Darnell DK, et al. (2006) MicroRNA expression during chick embryo development. *Dev Dyn* 235(11):3156–3165.
20. Lynn FC, et al. (2007) MicroRNA expression is required for pancreatic islet cell genesis in the mouse. *Diabetes* 56(12):2938–2945.
21. Kochva E (1978) *Biology of the Reptilia*, eds Gans C, Gans KA (Academic, London).
22. Kochva E, Nakar O, Ovadia M (1983) Venom toxins: Plausible evolution from digestive enzymes. *Amer Zool* 23(2):427–430.
23. Kochva E (1987) The origin of snakes and evolution of the venom apparatus. *Toxicon* 25(1):65–106.
24. Christodoulou F, et al. (2010) Ancient animal microRNAs and the evolution of tissue identity. *Nature* 463(7284):1084–1088.
25. Sparkman AM, et al. (2012) Rates of molecular evolution vary in vertebrates for insulin-like growth factor-1 (IGF-1), a pleiotropic locus that regulates life history traits. *Gen Comp Endocrinol* 178(1):164–173.
26. Mackessy SP (1991) Morphology and ultrastructure of the venom glands of the northern pacific rattlesnake *Crotalus viridis oreganus*. *J Morphol* 208:109–128.
27. Morita T (2005) Structures and functions of snake venom CLPs (C-type lectin-like proteins) with anticoagulant-, procoagulant-, and platelet-modulating activities. *Toxicon* 45(8):1099–1114.
28. Mebs D, Claus I (1991) *Snake Toxins*, ed Harvey AL (Pergamon, New York), pp 425–447.
29. Kini RM, Doley R (2010) Structure, function and evolution of three-finger toxins: Mini proteins with multiple targets. *Toxicon* 56(6):855–867.
30. Kini RM, Chan YM (1999) Accelerated evolution and molecular surface of venom phospholipase A_2 enzymes. *J Mol Evol* 48(2):125–132.
31. Fry BG, et al. (2003) Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *J Mol Evol* 57(1):110–129.
32. Lynch VJ (2007) Inventing an arsenal: Adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evol Biol* 7:2.
33. Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wüster W (2011) Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Mol Biol Evol* 28(9):2637–2649.
34. Sunagar K, Johnson WE, O'Brien SJ, Vasconcelos V, Antunes A (2012) Evolution of CRISPs associated with toxicoferan-reptilian venom and mammalian reproduction. *Mol Biol Evol* 29(7):1807–1822.
35. Fox JW (2013) A brief review of the scientific history of several lesser-known snake venom proteins: l-Amino acid oxidases, hyaluronidases and phosphodiesterases. *Toxicon* 62:75–82.
36. Warren WC, et al. (2008) Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453(7192):175–183.
37. Wong ES, Papenfuss AT, Whittington CM, Warren WC, Belov K (2012) A limited role for gene duplications in the evolution of platypus venom. *Mol Biol Evol* 29(1): 167–177.
38. Li R, et al. (2010) The sequence and *de novo* assembly of the giant panda genome. *Nature* 463(7279):311–317.
39. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding preassembled contigs using SSPACE. *Bioinformatics* 27(4):578–579.
40. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
41. Cantarel BL, et al. (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 18(1):188–196.
42. Holt C, Yandell M (2011) MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12: 491–505.
43. Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* 5:59–68.
44. Stanke M, Schöffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7:62.
45. Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644–652.
46. Birol I, et al. (2009) De novo transcriptome assembly with ABySS. *Bioinformatics* 25(21):2872–2877.
47. Simpson JT, et al. (2009) ABySS: A parallel assembler for short read sequence data. *Genome Res* 19(6):1117–1123.
48. Karim S, Singh P, Ribeiro JM (2011) A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. *PLoS One* 6(12):e28525.
49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
50. Jostarndt K, Puntschart A, Hoppeler H, Billeter R (1994) The use of 33P-labelled riboprobes for in situ hybridizations: Localization of myosin alkali light-chain mRNAs in adult human skeletal muscle. *Histochem J* 26(1):32–40.
51. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
52. Ronquist F, et al. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3):539–542.
53. Nylander JAA (2004) *MrModeltest v2* (Evolutionary Biology Centre, Uppsala University, Finland).
54. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
55. Chaudhary R, Bansal MS, Wehe A, Fernández-Baca D, Eulenstein O (2010) iGTP: A software package for large-scale gene tree parsimony analysis. *BMC Bioinformatics* 11:574.
56. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5):696–704.
57. Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19(6):908–917.