Title: Convergent Evolution of Pain-Inducing Defensive Venom Components in Spitting Cobras

Authors: T.D. Kazandjian[1†], D. Petras[2,3†], S.D. Robinson[4,5†], J. van Thiel[6], H.W. Greene[7], K. Arbuckle[8], A. Barlow[9,10], D.A. Carter[5], R.M. Wouters[6], G. Whiteley[1], S.C. Wagstaff[1,11], A.S. Arias[12], L-O. Albulescu[1], A. Plettenberg Laing[10], C. Hall[10], A. Heap[10], S. Penrhyn-Lowe[10], C.V. McCabe[13], S. Ainsworth[1], R.R. da Silva[2,14], P.C. Dorrestein[2], M.K. Richardson[6], J.M. Gutiérrez[12], J.J. Calvete[15], R.A. Harrison[1], I. Vetter[5,16], E.A.B. Undheim[4,5,17,18], W. Wüster[10], N.R. Casewell[1]*

Affiliations:

[1]Centre for Snakebite Research & Interventions, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK.

[2]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, United States.

[3]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, United States.

[4]Centre for Advanced Imaging, University of Queensland, St Lucia, QLD 4072, Australia.

[5]Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD 4072, Australia.

[6]Institute of Biology, University of Leiden, Leiden 2333BE, the Netherlands.

[7]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, United States.

[8]Department of Biosciences, College of Science, Swansea University, Swansea, SA2 8PP, UK.

[9]School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, UK

[10]Molecular Ecology and Fisheries Genetics Laboratory, School of Natural Sciences, Bangor University, Bangor, LL57 2UW, UK

[11]Research Computing Unit, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK.

[12]Instituto Clodomiro Picado, Facultad de Microbiología, Universidad de Costa Rica, San José 11501, Costa Rica.

[13]School of Earth Sciences, University of Bristol, Bristol, BS8 1RL, UK.

[14]NPPNS, Molecular Sciences Department, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil

[15]Evolutionary and Translational Venomics Laboratory, Consejo Superior de Investigaciones Científicas, Jaume Roig 11, 46010 Valencia, Spain

[16]School of Pharmacy, The University of Queensland, Woolloongabba, QLD 4102, Australia

[17]Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, 7491 Trondheim, Norway

[18]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, PO Box 1066 Blindern, 0316 Oslo, Norway

* Correspondence to: N.R. Casewell; Nicholas.casewell@lstmed.ac.uk

† These authors contributed equally.

Abstract: Convergent evolution provides insights into the selective drivers underlying evolutionary change. Snake venoms, with a direct genetic basis and clearly defined functional phenotype, provide a model system for exploring the repeated evolution of adaptations. While snakes use venom primarily for predation, and venom composition often reflects diet specificity, three lineages of cobras have independently evolved the ability to spit venom at adversaries. Using gene, protein and functional analyses, we show that the three spitting lineages possess venoms characterized by an upregulation of PLA_2 toxins, which potentiate the action of pre-existing venom cytotoxins to activate mammalian sensory neurons and cause enhanced pain. These repeated independent changes provide a fascinating example of convergent evolution across multiple phenotypic levels driven by selection for defense.

One Sentence Summary: Defensive venom spitting by snakes is underpinned by convergent increases in pain-enhancing toxins

Main Text:

Convergent evolution, the independent emergence of similar traits across taxa, is a pervasive characteristic of biodiversity and provides natural replicates to enable understanding of key evolutionary processes (*1*). Thanks to their discrete function and direct genotype-phenotype link, animal venoms are excellent systems to understand the driving forces and underlying genetic mechanisms of molecular adaptation. Snake venoms consist of variable mixtures of proteinaceous components causing potent hemotoxic, neurotoxic and/or cytotoxic pathologies in both prey and potential adversaries, including humans (*2*). Previous work suggests that venom variation is largely driven by dietary variation (*3*), but defensive drivers of snake venom evolution are rarely considered (although see (*4*)).

The evolution of venom projection or 'spitting' in cobras offers an ideal system for exploring the evolution of defensive toxins: this behavior plays no role in prey capture, targets specific sensory tissues, and is the only long-distance, injurious defensive adaptation among almost four thousand species of snakes. Remarkably, venom spitting evolved independently three times, all within a single clade of closely related elapid snakes (*5*, *6*): the African spitting cobras (*Naja*: subgenus *Afronaja*), Asian spitting cobras (*Naja*: subgenus *Naja*) and rinkhals (*Hemachatus*). All use fangs with modified orifices (*7*) to spray venom over distances of up to 2.5 m (*8*), targeting an aggressor's eyes (*9*) (Fig. S1). These behavioral, morphological, and biochemical traits result in intense ocular pain and inflammation, which can lead to the permanent loss of eyesight (*10*). The three origins of spitting, solely within a clade more generally characterized by the visually defensive behavior of hooding (*5*, *6*), allow us to test whether similar selective pressures have resulted in convergent changes in venom composition coevolving with morphological and behavioral adaptations.

We used a multi-disciplinary approach consisting of transcriptomic, proteomic, functional and phylogenetic comparisons of 17 widely distributed elapids: 14 *Naja* (true cobras), the rinkhals *Hemachatus haemachatus*, and two non-spitting immediate sister group species, *Walterinnesia aegyptia* and *Aspidelaps scutatus*, to investigate the evolution of venom spitting. First, we reconstructed the phylogeny of these snakes using a multilocus coalescent species tree approach based on two mitochondrial and five nuclear genes. Fossil-calibrated molecular dating suggests that spitting originated in African spitting cobras 6.7-10.7 million years ago (MYA), and around 4 million years later in the Asian spitting cobras (2.5-4.2 MYA) (Fig. 1A). The origin of spitting in *Hemachatus* could not be dated, beyond that it occurred <17 MYA, following divergence from true cobras (*Naja*) (Fig. 1A).

Next, we used a top-down proteomics approach underpinned by venom gland transcriptomic data (*11*) to characterize the venom composition of each species. All cobra venoms are dominated by three finger toxins (3FTX), while in many species phospholipases A_2 (PLA_2) are the second most abundant toxin family (Fig. 1A, Fig. S2). Principal coordinate analysis (PCoA, Bray-Curtis) of a proteomic data-derived venom composition matrix separated the spitting lineages into three clusters that are distinct from the homogeneous cluster of venoms from non-spitters (Fig. 1B). The sole exception, *N. philippinensis*, has a purely neurotoxic venom despite being able to spit (*12*), and its venom composition placed it alongside non-spitting species (see Fig. 1B). Nonetheless, these findings demonstrate that each spitting cobra lineage exhibits distinct venom compositions that collectively differ from those of non-spitting cobras – a finding consistent with differences in venom-induced pathology observed following bites to humans (*13*).

3FTXs are major venom components in many elapid snakes (*14*) (Table S1). They are encoded by a multilocus gene family, resulting in numerous functionally distinct isoforms, including neurotoxins and cytotoxins that disrupt cell membranes to cause cytotoxicity (*15*). Proteomic data revealed that cytotoxic 3FTXs (CTXs) are typically the most abundant toxins in *Naja* and *Hemachatus* venoms (mean 57.7% of all toxins), contrasting with the sister group species *W. aegyptia* and *A. scutatus* (*16*) and other elapids (Fig. 1A, Figs. S2, S3, Table S1). Although cytotoxins likely inflict defensive ocular pain, we found no significant difference in the abundances of CTXs between spitting and non-spitting species (PGLS; $t = -0.83$, $df = 15$, $p = 0.42$) (Table S2), and ancestral state estimations suggest that the origin of CTX-rich venom preceded that of venom spitting (Fig. S4) (*5*). Moreover, PCoA analysis of a CTX Euclidean distance matrix derived from venom proteomic data revealed that all highly abundant cobra CTXs cluster tightly together (Fig. 1C). Measures of irritation, stimulated by high doses (100 µg) of cobra venoms applied topically to non-sentient chick embryos (Tables S3, S4), also revealed no association between cytotoxicity and spitting (PGLS; $t = 1.08$, $df = 15$, $p = 0.30$) (Figs. S5, S6) – consistent with prior reports of comparable cytotoxicity to mammalian cells across all cobra venoms (*5*).

However, pain inflicted via slow-onset cytotoxicity may be less defensively relevant than rapid pain caused by direct algesic activity. To investigate venom-induced nociception, we assessed the activation of mammalian trigeminal neurons – sensory neurons derived from trigeminal ganglia that innervate the face and eyes. All cobra venoms activated sensory neurons, with the mechanism of action observed consistent with non-specific disruption of cell membranes, though activity was limited in African non-spitting cobras and the Asian non-spitter *Naja kaouthia* (Fig. S7). Half maximal effective concentrations (EC_{50}) of each venom in sensory

neuron-derived F11 cells demonstrated significantly higher activity in spitting cobra venoms (PGLS; $t = -4.48$, $df = 15$, $p = 0.0007$) (Fig. 2A, Fig. S8, Table S2). These findings support the hypothesis that venom spitting is associated with convergent elevations in venom-induced activation of mammalian sensory neurons, and that spitting cobra venoms are more effective in causing pain than their non-spitting counterparts.

To determine the toxins responsible for this effect, we repeated these experiments using fractionated venom from three representative spitting species (*N. nigricollis*, African; *N. siamensis*, Asian; *H. haemachatus*) (Fig. S9). For each species, only fractions corresponding to CTXs activated sensory neurons, while those corresponding to other toxins (e.g. neurotoxins, PLA_2s, etc) were inactive (Fig. S9). However, none of the CTX fractions completely recapitulated the effects of whole venom or re-pooled venom fractions, suggesting synergy between multiple venom components.

PLA_2s are nearly ubiquitous, typically enzymatic, multifunctional toxin components of snake venoms (*14*, *15*). As the hemolytic activity of CTXs was previously shown to be potentiated by PLA_2 toxins (*17*), we hypothesized that venom PLA_2s potentiate sensory neuron activation by CTXs. Consequently, we quantified the activation of sensory neurons stimulated by CTXs in the presence or absence of a corresponding PLA_2 fraction. The proportion of viable sensory neurons activated by each CTX fraction was significantly increased when combined with a PLA_2 fraction, and this result was consistent across representatives of the three spitting lineages (unpaired t-test; *N. nigricollis*, $t = 18.77$, $df = 2$, $p = 0.003$; *N. siamensis*, $t = 5.75$, $df = 4$, $p = 0.005$; *H. haemachatus*, $t = 4.18$, $df = 4$, $p = 0.01$) (Fig. 2B, Fig. S10). Moreover, significant reductions in sensory neuron activation occurred in the presence of the PLA_2 inhibitor

varespladib (unpaired t-test; t = 2.77, df = 14, p = 0.02) (Fig. 2C, Fig. S11), providing further compelling evidence that PLA_2s potentiate CTX effects on sensory neurons.

Consistent with the above findings, comparative analysis of the (i) proteomic abundance of PLA_2 toxins and (ii) enzymatic PLA_2 activity, determined via specific *in vitro* colorimetric assay, revealed that spitting cobra venoms have significantly higher PLA_2 abundance and activity than those of non-spitting species (PGLS; t = 4.24, df = 15, p = 0.0007 and t = 2.24, df = 15, p = 0.04, respectively) (Fig. 3A, 3B, Fig. S12, Table S2). Analysis of a PLA_2 Euclidean distance matrix derived from venom proteomic data revealed substantial differences between the different lineages of spitting and non-spitting cobras, particularly among the African species (Fig. 3C). Additionally, despite their divergence ~17 MYA, *H. haemachatus* PLA_2s clustered tightly with those from African spitting cobras, suggesting an element of molecular convergence. Interrogation of venom gland transcriptomic data revealed limited variation in PLA_2 gene number across the three spitting lineages, though all exhibited increased PLA_2 abundance compared with non-spitting cobras (Fig. S13A). Interestingly, phylogenetic analyses revealed a PLA_2 gene duplication event that coincided with the origin of venom spitting in the ancestor of African spitting cobras (Fig. S13B). These data, alongside the sensory neuron assays, demonstrate that independent evolution of spitting is tightly linked with convergent increases in PLA_2 toxins, which cause increased algesic activity. Our findings therefore imply evolution has funnelled defensive venom phenotypes along repeatable and predictable pathways, though different molecular mechanisms likely underpin this convergence.

To exclude the possibility that functional distinctions simply reflect general differences in venom potency (e.g., venom lethality for prey capture), we tested our venoms in murine lethality assays (Table S5) and found no significant differences between spitting and non-spitting species

(PGLS; t = 0.86, df = 15, p = 0.40) (Table S2). These results suggest that enhanced pain caused by spitting cobras is explicitly associated with defensive venom use, rather than being an evolutionary by-product of selection for prey subjugation. Moreover, reanalysis of our data with the Asian species *N. kaouthia* and *N. atra* scored as spitting cobras based on recent reports (*18*, *19*), did not alter our key findings relating to venom spitting and PLA_2-mediated enhanced activation of sensory neurons (Table S6).

Our results detail the molecular and functional correlates of the evolution of venom spitting and demonstrate that defense can be a major driver of snake venom composition. Spitting likely only evolved within a single, relatively small, clade of elapid snakes due to the integrated exaptation of a unique combination of pre-existing behaviors and cytotoxic venom activities. Early evolution of cytotoxic venom activity in cobras and near relatives (~26 MYA; Fig. S4) has previously been linked to defense, as cytotoxicity co-originates with 'hooding' (*5*), a long-distance visual aposematic display. This elevated posture directed towards predators, coupled with pre-emptive striking and occasional premature releases of venom, may have provided a behavioral precursor for evolution of more targeted venom spitting. Pre-existing CTXs, largely absent from the other elapid venoms (Table S1), likely provided the baseline ocular toxicity that favored inception and retention of spitting (*19*). Subsequent independent increases in PLA_2 toxins, which act in synergy with pre-existing CTXs, resulted in increased venom-induced activation of nociceptors (Fig. 2). This potentiating effect of PLA_2s may be crucial for causing immediate pain of sufficient intensity to rapidly deter aggressors, allowing the snake to escape.

Rare but repeatedly evolved adaptations likely result from similar ecological circumstances. Most discussions of defensive behavior involve potential predators, and certain

mammals and birds commonly eat snakes (*20*, *21*). However, predation on spitting cobras is evidently unremarkable in terms of evolutionary history and biogeography. Beyond predators, potential threats to snakes include inadvertent trampling and pre-emptive defensive killing. That spitting evolved to prevent snakes being trampled by ungulates in African savannas (*22*) does not explain the existence of primarily forest-dwelling Asian spitting cobras (*23*). Moreover, large ungulates typically have lateral eyes, making them unlikely to be especially vulnerable to spitting.

We believe several considerations make ancient hominins a plausible and compelling candidate for favoring repeated evolution of spitting in the Afro-Asian cobras: (i) growing evidence suggests that snakes have influenced primate neurobiology and behavior (*24*), and that interactions between these two lineages have been important throughout the 75 million year history of primates (*25*, *26*). (ii) Compared to carnivorous mammals, anthropoid primates as a clade are visually acute, cognitively complex, and culturally sophisticated (*24*). (iii) Diverse anthropoids mob snakes, with some distinguishing between harmless and dangerous species, killing the latter from a distance with clubs or projectiles (*25*–*27*). (iv) Characteristics (ii) and (iii) are enhanced among bipedal, larger-brained hominins (*24*, *26*), which thus could have posed a unique threat to snakes (*28*). The initial divergence of Africa spitting cobras as recently as 6.7 MYA (Fig. 1) occurred soon after the divergence of hominins from *Pan* (bonobos and chimpanzees) ~7 MYA (Fig. S14), coinciding with early evolution of bipedalism, enlarged brains, tool use, and occupation of savannas by the former (*29*). Likewise, the origin of the Asian spitter clade ~2.5 MYA is approximately contemporaneous with the arrival in Asia of *Homo erectus* (*30*, *31*) (Fig. S14). Though based on circumstantial evidence, additional fossils and

more finely-tuned dating of relevant cobra and primate divergences might allow further testing of this hypothesis.

In summary, spitting cobras highlight how similar selection pressures and shared exaptations can drive convergence at molecular, morphological, behavioral, and functional levels, resulting in the evolution of complex, integrated adaptations in ecologically important traits being funnelled down repeatable pathways.

References and Notes:

1. J. Losos, *Improbable Destinies: How predictable is Evolution?* (Penguin Random House, 2017).
2. J. M. Gutiérrez *et al.*, *Nat. Rev. Dis. Prim.* **3**, 17063 (2017).
3. J. C. Daltry, W. Wüster, R. S. Thorpe, *Nature*. **379**, 537–540 (1996).
4. H. Ward-Smith, K. Arbuckle, A. Naude, W. Wüster, *Toxins (Basel).* **12**, 1–20 (2020).
5. N. Panagides *et al.*, *Toxins (Basel).* **9**, 103 (2017).
6. W. Wüster *et al.*, *Mol. Phylogenet. Evol.* **45**, 437–453 (2007).
7. C. M. Bogert, *Bull. Am. Museum Nat. Hist.* **81**, 285–360 (1943).
8. S. Rasmussen, B. Young, H. Krimm, *J. Zool.* **237**, 27–35 (1995).
9. B. A. Westhoff, G., Boetig, M., Bleckmann, H., Young, *J. Exp. Biol.* **213**, 1797–1802 (2010).
10. D. A. Warrell, L. D. Ormerod, *Am. J. Trop. Med. Hyg.* **25**, 525–529 (1976).
11. S. Ainsworth *et al.*, *J. Proteomics*. **172**, 173–189 (2018).
12. G. Watt, L. Padre, M. L. Tuazon, R. D. G. Theakston, L. Laughlin, *Am. J. Trop. Med. Hyg.* **39**, 306–311 (1988).
13. J. Meier, J. White, *Handbook of Clinical Toxicology of Animal Venoms and Poisons* (Informa Healthcare USA Inc, New York, ed. 1st, 2008).
14. T. Tasoulis, G. K. Isbister, *Toxins (Basel).* **9**, 290- (2017).
15. C. R. Ferraz *et al.*, *Front. Ecol. Evol.* **7**, 1–19 (2019).
16. G. Whiteley *et al.*, *J. Proteomics*. **198**, 186–198 (2019).
17. E. Condrea, *Experientia*. **30**, 121–129 (1974).
18. A. Paterna, *Herpetol. Bull.*, **148**, 22–25 (2019).
19. V. Santra, W. Wüster, *Herpetol. Rev.* **48**, 455–456 (2017).
20. H. W. Greene, in *Biology of the Reptilia. Vol. 16: Ecology B: Defense and Life History*, C. G. and R. B. Huey, Ed. (Alan R. Liss, Inc., New York, ed. 1st, 1988), pp. 1–152.
21. B. Van Valkenburgh, R. K. Wayne, *Curr. Biol.* **20**, R915–R919 (2010).
22. T. Barbour, *Copeia*, **106**, 26–28 (1922).
23. E. R. Chu, S. A. Weinstein, J. White, D. A. Warrell, *Toxicon*. **56**, 259–272 (2010).
24. Lynne A. Isbell, *The Fruit, the Tree, and the Serpent: Why We See So Well* (Harvard University Press, Cambridge, 2009).

25. T. N. Headland, H. W. Greene, *Proc. Natl. Acad. Sci. USA*, **108**, E1470-E1474 (2011).
26. H. W. Greene, *Tracks and Shadows: Field Biology as Art* (University of California Press, Berkeley and Los Angeles, 2013).
27. H. W. Greene, *Am. Nat.* **190**, S69–S86 (2017).
28. S. Faurby, D. Silvestro, L. Werdelin, A. Antonelli, *Ecol. Lett.* **23**, 537–544 (2020).
29. L. Pozzi *et al.*, *Mol. Phylogenet. Evol.* **75**, 165–183 (2014).
30. S. Prat, *Comptes Rendus - Palevol*. **17**, 6–16 (2018).
31. F. Han *et al.*, *Quat. Int.* **434**, 75–83 (2017).
32. D. A. Dmitriev, R. A. Rakitov, *PLoS Comput. Biol.* **4**, e1000113 (2008).
33. M. Stephens, P. Scheet, *Am. J. Hum. Genet.* **76**, 449–462 (2005).
34. M. Stephens, N. J. Smith, P. Donnelly, *Am. J. Hum. Genet.* **68**, 978–989 (2001).
35. J. F. Flot, *Mol. Ecol. Resour.* **10**, 162–166 (2010).
36. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
37. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, *Syst. Biol.* **67**, 901–904 (2018).
38. J. J. Head, K. Mahlow, J. Müller, *Palaeontol. Electron.* **19**, 1–21 (2016).
39. D. Pla *et al.*, *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 814-823 (2017).
40. G. Whiteley *et al.*, *J. Proteomics*. **198**, 186–198 (2019).
41. J. Archer, G. Whiteley, N. R. Casewell, R. A. Harrison, S. C. Wagstaff, *BMC Bioinformatics*. **15**, 389 (2014).
42. A. Conesa *et al.*, *Bioinformatics*. **21**, 3674–3676 (2005).
43. S. Kumar, G. Stecher, K. Tamura, *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
44. R. C. Edgar, *Nucleic Acids Res.* **32**, 1792–1797 (2004).
45. D. Petras *et al.*, *J. Proteome Res.* **10**, 1266–1280 (2011).
46. D. Pla *et al.*, *J. Proteomics*. **174**, 71–84 (2018).
47. Q. Kou, L. Xun, X. Liu, *Bioinformatics*. **32**, 3495–3497 (2016).
48. L. J. Revell, *Methods Ecol. Evol.* **3**, 217–223 (2012).
49. D. Orme *et al.*, caper: Comparative Analyses of Phylogenetics and Evolution in R (2018), (available at https://cran.r-project.org/package=caper).
50. RStudio, RStudio: Integrated Development for R. (2016).
51. D. T. Jones, W. R. Taylor, J. M. Thornton, *Comput. Appl. Biosci.* **8**, 275–282 (1992).
52. M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, cluster: Cluster Analysis Basics and Extensions. (2019), (available at https://www.bibsonomy.org/bibtex/2153b1c71326b16e5d93b998666537ee8/enitsirhc).
53. A. S. Siang, R. Doley, F. J. Vonk, R. M. Kini, *BMC Mol. Biol.* **11**, 24 (2010).
54. Y. Jiang *et al.*, *BMC Genomics*. **12**, 1 (2011).
55. M. J. Margres, K. Aronow, J. Loyacano, D. R. Rokyta, *BMC Genomics*. **14**, 531 (2013).
56. Li,J., Zhang, J. H., Liu, K. Xu, *Biochem. J.* **398**, 233–242 (2006).
57. F. J. Vonk *et al.*, *Proc. Natl. Acad. Sci.* **110**, 20651–20656 (2013).
58. A. D. Hargreaves, M. T. Swain, D. W. Logan, J. F. Mulley, *Toxicon*. **92**, 140–156 (2014).
59. G. Singh *et al.*, *J. Mol. Biol.* **307**, 1049–1059 (2001).
60. G. Singh, *Protein Sci.* **14**, 395–400 (2005).
61. R. Yanoshita *et al.*, *Toxicon*. **47**, 416–424 (2006).
62. I. H. Tsai, H. Y. Tsai, A. Saha, A. Gomes, *FEBS J.* **274**, 512–525 (2007).
63. J.-M. Danse, *Nucleic Acids Res.* **18**, 4608 (1990).

64. J.-M. Danse, J.-M. Garnier, J. Kempf, *Nucleic Acids Res.* **18**, 4610 (1990).
65. C. Corrêa-Netto *et al.*, *J. Proteomics*. **74**, 1795–1809 (2011).
66. P. L. Ho, M. B. Soares, T. Yamane, I. Raw, *J. Toxicol. Toxin Rev.* **14**, 327–337 (1995).
67. C. Calcines-Cruz *et al.*, *Int. J. Biol. Macromol.* **108**, 826–836 (2018).
68. C. J. Bohlen *et al.*, *Nature*. **479**, 410–414 (2011).
69. L. Zeng *et al.*, *Toxicon*. **60**, 290–301 (2012).
70. Q.-Y. Wang, Y.-Y. Shu, M.-X. Zhuang, Z.-J. Lin, *Acta Biochim. Biophys. Sinica.* **33**, 340–344 (2001).
71. T. A. Castoe *et al.*, *Genome Biol.* **12**, 406 (2011).
72. J. Alföldi *et al.*, *Nature*. **477**, 587–591 (2011).
73. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, *Nat. Methods*. **9**, 772–772 (2012).
74. D. Posada, *Mol. Biol. Evol.* **25**, 1253–1256 (2008).
75. F. Ronquist *et al.*, *Syst. Biol.* **61**, 539–542 (2012).
76. M. A. Miller, W. Pfeiffer, T. Schwartz, *2010 Gatew. Comput. Environ. Work. GCE 2010*, 1-8 (2010).
77. N. P. Luepke, *Food Chem. Toxicol.* **23**, 287–291 (1985).
78. I. Vetter, R. J. Lewis, *Biochem. Pharmacol.* **79**, 908–920 (2010).
79. J. E. Hale, J. P. Butler, V. Gelfanova, J. S. You, M. D. Knierman, *Anal. Biochem.* **333**, 174–181 (2004).
80. C. Neumann *et al.*, *Toxicon*. **178**, 61–68 (2020).
81. R. D. G. Theakston, H. A. Reid, **61**, 949–956 (1983).
82. D. Petras, P. Heiss, R. D. Süssmuth, J. J. Calvete, *J. Proteome Res.* **14**, 2539–2556 (2015).
83. A. H. Laustsen *et al.*, *Toxicon*. **107**, 187–196 (2015).
84. A. M. F. Oh, C. H. Tan, G. C. Ariaranee, N. Quraishi, N. H. Tan, *J. Proteomics*. **164**, 1–18 (2017).
85. M. R. A. Rusmili, T. T. Yee, M. R. Mustafa, W. C. Hodgson, I. Othman, *J. Proteomics*. **110**, 129–144 (2014).
86. R. H. Ziganshin *et al.*, *Toxicon*. **107**, 197–209 (2015).
87. L. L. Shan *et al.*, *J. Proteomics*. **138**, 83–94 (2016).
88. A. M. F. Oh, C. H. Tan, K. Y. Tan, N. H. Quraishi, N. H. Tan, *J. Proteomics*. **193**, 243–254 (2019).
89. K. Y. Tan *et al.*, *J. Proteomics*. **192**, 246–257 (2019).
90. L. P. Lauridsen, A. H. Laustsen, B. Lomonte, J. M. Gutiérrez, *J. Proteomics*. **136**, 248–261 (2016).
91. A. H. Laustsen, B. Lomonte, B. Lohse, J. Fernández, J. M. Gutiérrez, *J. Proteomics*. **119**, 126–142 (2015).
92. C. H. Tan, K. Y. Tan, T. S. Ng, S. M. Sim, N. H. Tan, *Toxins (Basel).* **11**, 1–19 (2019).
93. J. J. Calvete *et al.*, *J. Proteomics*. **75**, 4091–4101 (2012).
94. B. Lomonte *et al.*, *J. Proteomics*. **103**, 137–152 (2014).
95. C. H. Tan, K. Y. Tan, S. E. Lim, N. H. Tan, *J. Proteomics*. **126**, 121–130 (2015).
96. C. H. Tan, K. Y. Wong, K. Y. Tan, N. H. Tan, *J. Proteomics*. **166**, 48–58 (2017).
97. O. Paiva *et al.*, *J. Proteomics*. **110**, 209–229 (2014).
98. J. Fernández *et al.*, *Toxicon*. **107**, 217–233 (2015).
99. S. D. Aird *et al.*, *Toxins (Basel).* **9**, 1–64 (2017).
100. P. Rey-Suárez, V. Núñez, J. Fernández, B. Lomonte, *J. Proteomics*. **136**, 262–273 (2016).
101. L. Sanz *et al.*, *Toxicon*. **166**, 39–45 (2019).

102. P. Rey-Suárez, V. Núñez, J. M. Gutiérrez, B. Lomonte, *J. Proteomics*. **75**, 655–667 (2011).
103. J. Fernández *et al.*, *J. Proteome Res.* **10**, 1816–1827 (2011).
104. T. Olamendi-Portugal *et al.*, *Toxicon*. **153**, 23–31 (2018).
105. L. Sanz *et al.*, *Toxins (Basel).* **8**, 9–12 (2016).
106. H.-W. Huang *et al.*, *J. Proteomics*. **128**, 92–104 (2015).
107. I. Malih *et al.*, *J. Proteomics*. **96**, 240–252 (2014).
108. K. Y. Tan, C. H. Tan, S. Y. Fung, N. H. Tan, *J. Proteomics*. **120**, 105–125 (2015).
109. N. Xu *et al.*, *J. Proteomics*. **159**, 19–31 (2017).
110. L. P. Lauridsen, A. H. Laustsen, B. Lomonte, J. M. Gutiérrez, *J. Proteomics*. **150**, 98–108 (2017).
111. S. Dutta *et al.*, *J. Proteomics*. **156**, 29–39 (2017).
112. K. Sintiprungrat *et al.*, *J. Proteomics*. **132**, 131–143 (2016).
113. C. H. Tan, K. Y. Wong, H. P. Chong, N. H. Tan, K. Y. Tan, *J. Proteomics*. **206**, 1–13 (2019).
114. N. H. Tan, K. Y. Wong, C. H. Tan, *J. Proteomics*. **157**, 18–32 (2017).
115. C. H. Tan, K. Y. Tan, N. H. Tan, *J. Proteomics*. **144**, 33–38 (2016).
116. C. H. Tan, K. Y. Tan, S. Y. Fung, N. H. Tan, *BMC Genomics*. **16**, 687 (2015).
117. M. Herrera *et al.*, *J. Proteomics*. **75**, 2128–2140 (2012).
118. D. Pla *et al.*, *J. Proteomics*. **150**, 201–215 (2017).
119. W. Wüster *et al.*, *Zootaxa*. **4455**, 68–98 (2018).
120. S. R. Palumbi, in *Molecular Systematics, 2nd edition*, D. M. Hillis, C. Moritz, B. K. Mable, Eds. (Sinauer, Sunderland, Massachusetts, 1996), pp. 205–247.
121. M. B. Harvey, D. G. Barker, L. K. Ammerman, P. T. Chippindale, *Herpetol. Monogr.*, **14**, 139–185 (2000).
122. F. T. Burbrink, R. Lawson, J. B. Slowinski, *Evolution (N. Y).* **54**, 2107–2118 (2000).
123. E. Arévalo, J. W. Sites, S. K. Davis, E. Arévalo, *Syst. Biol.* **43**, 387–418 (1994).
124. A. Plettenberg Laing, "A multilocus phylogeny of the cobra clade elapids". Thesis, Bangor University, UK (2019).
125. A. F. Hugall, M. S. Y. Lee, *Mol. Biol. Evol.* **21**, 2102–2110 (2004).
126. T. M. Townsend, R. E. Alegre, S. T. Kelley, J. J. Wiens, T. W. Reeder, *Mol. Phylogenet. Evol.* **47**, 129–142 (2008).
127. N. R. Casewell, S. C. Wagstaff, R. A. Harrison, W. Wüster, *Mol. Biol. Evol.* **28**, 1157–1172 (2011).

Acknowledgments: The authors thank Paul Rowley, Edouard Crittenden, Wendy Grail for technical support and NVIDIA Corporation for donation of the TITAN-X GPU used for BEAST analyses. **Funding:** This work was funded from a studentship supported by Elizabeth Artin Kazandjian to T.D.K., grant PE 2600/1 from the German Research Foundation (DFG) to D.P., grant OPUS 1354156 from the US National Science Foundation to H.W.G., grants FAPESP 2017/18922-2 and 2019/05026-4 from the São Paulo Research Foundation to R.R.d.S, grants RPG-2012-627 and RFG-10193 from the Leverhulme Trust to R.A.H. and W.W., grant MR/L01839X/1 from the UK Medical Research Council to J.M.G., R.A.H., J.J.C. and N.R.C., fellowship DE160101142 and grant DP160104025 from the Australian Research Council, and fellowship FRIPRO-YRT #287462 from the Research Council of Norway to E.A.B.U., and a Sir

Henry Dale Fellowship (200517/Z/16/Z) jointly funded by the Wellcome Trust and Royal Society to N.R.C. **Author contributions:** R.A.H., W.W. and N.R.C. conceived the research. T.D.K., D.P., K.A., M.K.R., R.A.H., J.J.C., I.V., E.A.B.U., W.W. and N.R.C. designed the research. T.D.K., D.P., S.R., J.vT., A.B., D.A.C., R.M.W., G.W., S.C.W., A.S.A., L-O.A., A.vP.L., C.H., A.H., S.P-L., C.V.M., S.A., R.R.d.S., P.C.D., J.M.G., J.J.C., I.V., E.A.B.U., W.W. and N.R.C. performed the research. T.D.K. and N.R.C. wrote the manuscript with major input from D.P., S.D.R., H.W.G., I.V., E.A.B.U. and W.W. All authors discussed and commented on the manuscript. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** The molecular data associated with species tree generation have been deposited to the nucleotide database of NCBI and accession numbers are displayed in Table S7. The transcriptome data have been deposited in the SRA and TSA databases of NCBI and are associated with the BioProject accession number PRJA506018. Mass spectrometry data and database search results for top-down and bottom-up proteomic experiments are publicly available in the MassIVE repository under accession number MSV000081885 and in proteomXchange with accession number PXD008597.

Supplementary Materials:

Materials and Methods

Figures S1-S14

Tables S1-S10

Data S1-S5

References (32-127)

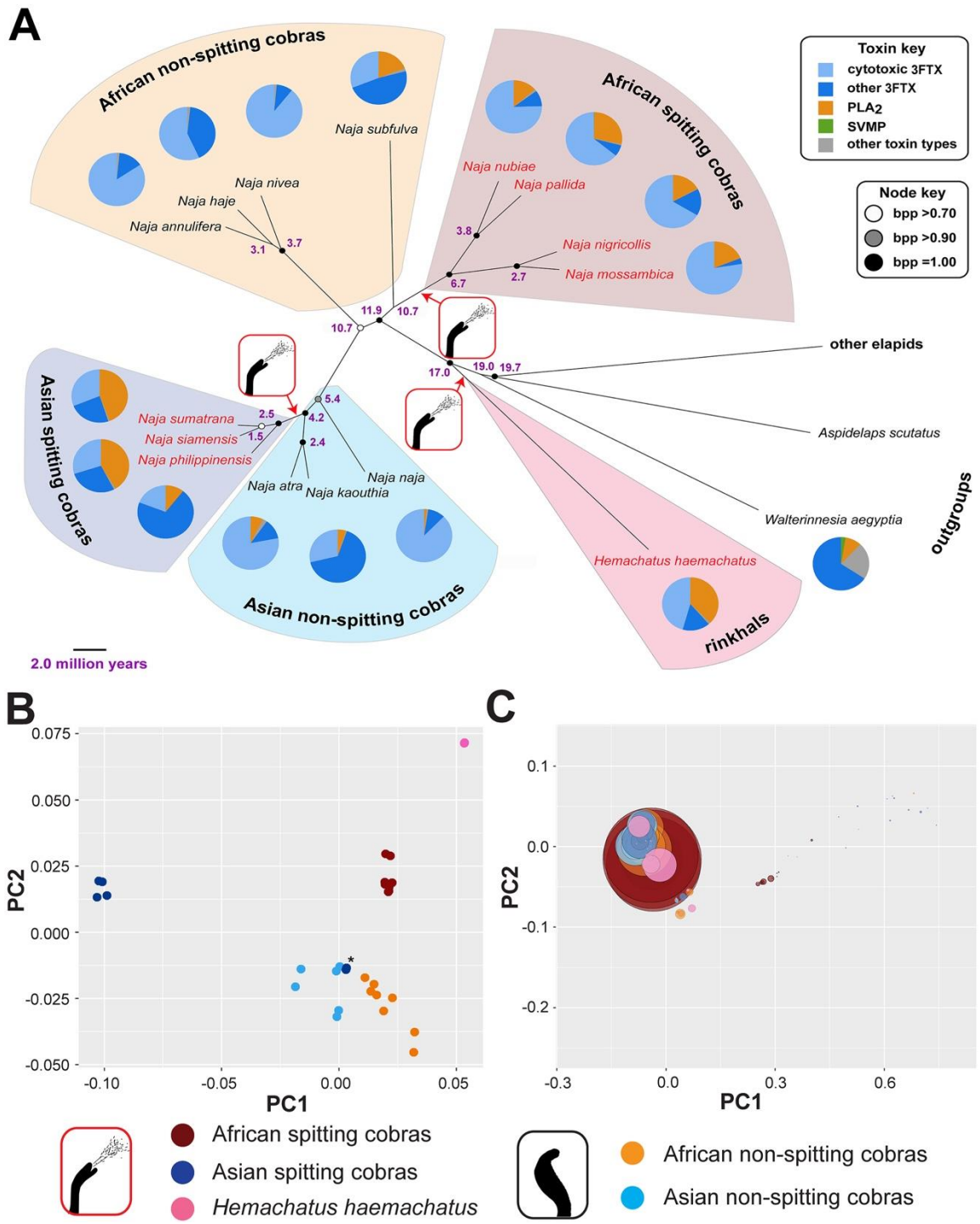


Fig. 1. Reconstruction of the evolutionary origin of venom spitting and comparative analysis of venom composition. (A) Multilocus-derived multispecies coalescent species tree, pruned to display the taxa whose venoms were analyzed in this study. Node support is indicated by colored circles, representing Bayesian posterior probabilities: black = 1.00, grey >0.90, white >0.70. Purple node labels indicate estimated divergence times (see Fig. S14 for credibility intervals). Spitting species are highlighted by red tip labels, and the three independent origins of

venom spitting are indicated by the red-boxed spitting images. Pie charts adjacent to tip labels represent proteomic toxin composition of each species as a percentage of total toxins. (B) Principal Coordinate Analysis (PCoA) of cobra (*Naja* spp.) and rinkhals (*H. haemachatus*) venom toxins reveal major distinctions between spitting and non-spitting lineages. Asterisk highlights the Asian spitting species *N. philippinensis*, which exhibits greater similarity to non-spitting species than to its nearest relatives. Note each species is represented by two, typically overlapping, data points, which represent technical proteomic duplicates. (C) PCoA of cobra (*Naja* spp.) and rinkhals (*H. haemachatus*) cytotoxic three-finger toxins (CTXs) derived from top-down venom proteomics reveals that the most abundant CTXs detected in venom exhibit little sequence diversity among spitting and non-spitting lineages. Circle sizes reflect relative abundances of CTXs detected.

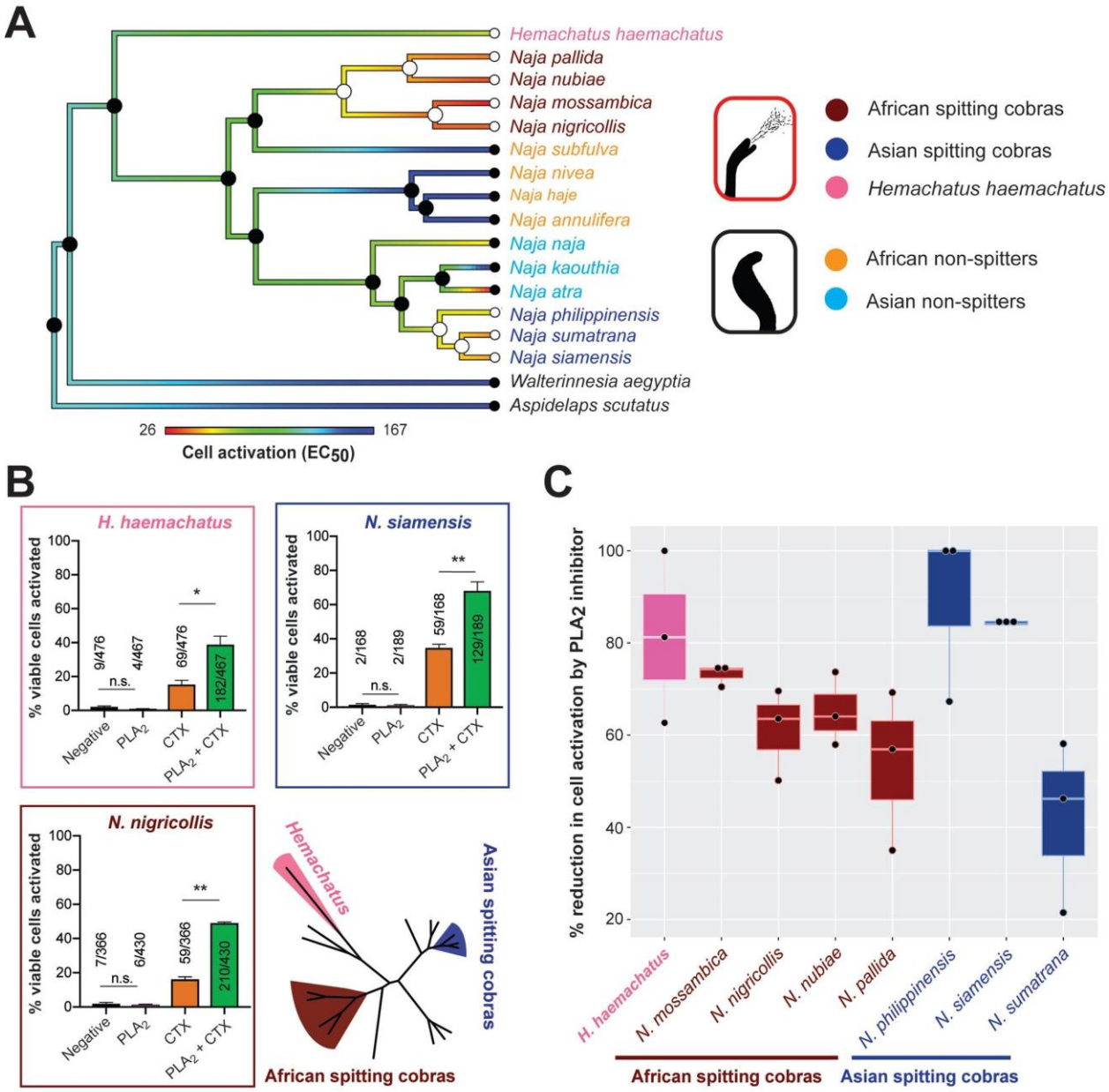


Fig. 2. Spitting cobra venoms cause significantly greater activation of sensory neurons than non-spitting cobras, mediated via potentiation by PLA_2 toxins. (A) Ancestral state estimation of half maximal effective concentrations (EC_{50}) of venom-induced activation of neuronal cells shows a significant association between increased potency and venom spitting (PGLS, t = -4.48, p = 0.0004). EC_{50} values are expressed as the mean of triplicate measurements and colored branches are scaled accordingly (red, low EC_{50} and thus high venom potency; blue, high EC_{50} and thus low venom potency). Filled or empty circles at nodes/tips represent estimated ancestral states of non-spitting or spitting, respectively, and colored tip labels correspond to the different lineages. (B) PLA_2 toxins in spitting cobra venoms potentiate the activating effect of CTXs on sensory neurons. A CTX fraction from each venom was added to dissociated mouse DRG neurons in the presence or absence of a corresponding PLA_2 fraction (added 1 min prior), neuronal activation (i.e. a rapid increase in $[Ca^{2+}]_i$) monitored, and data presented as mean ±

SEM of the resulting proportion of viable cells from 2-3 independent experiments. Statistical comparisons were performed using unpaired parametric *t*-tests; *, $p < 0.05$; **, $p < 0.01$. (C) The PLA_2 inhibitor varespladib reduces neuronal activation stimulated by spitting cobra venoms. Calcium influx in F11 cells was measured on a FLIPR instrument incubated in the presence of venom from spitting species (2.4 µg or 4.8 µg [in the case of *H. haemachatus* and *N. philippinensis*] venom) and in the presence or absence of varespladib (13 µM). The data displayed represents the percentage of venom only cell activation stimulated by treatment with venom and varespladib. Error bars represent SEM and box mid-lines median values.

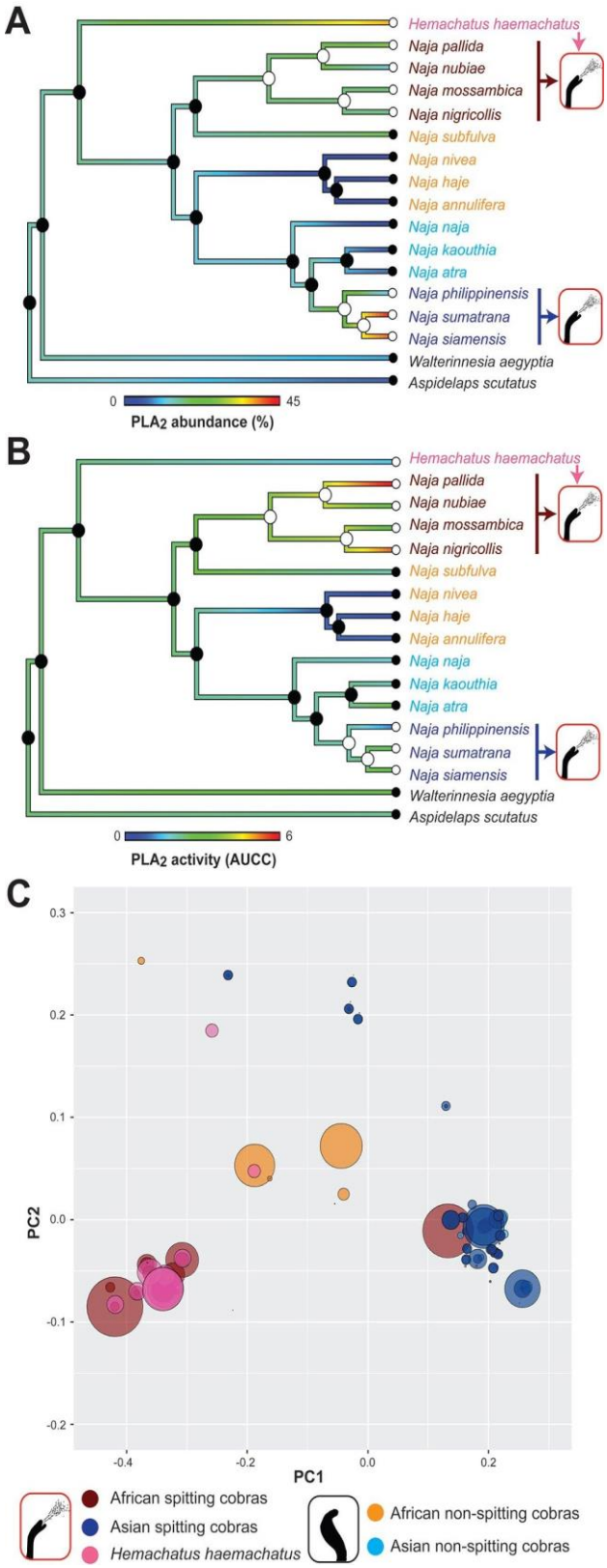


Fig. 3. The abundance, enzymatic activity and diversity of PLA_2 toxins are associated with convergent evolution of venom spitting. (A) Ancestral state estimation of PLA_2 proteomic

abundance, expressed as percentage of all toxins in venom proteomes, revealed a significant association with venom spitting (PGLS, $t = 4.24$, $p = 0.0007$). Colored branches are scaled according to PLA_2 abundance (blue, low abundance; red, high abundance), filled or empty circles at nodes/tips represent estimated ancestral states of non-spitting or spitting, respectively, and colored tip labels correspond to the different lineages. (B) Ancestral state estimation of enzymatic PLA_2 activity, expressed as area under the curve of concentration curves (AUCC) from kinetic *in vitro* colorimetric assay, revealed a significant association with venom spitting (PGLS, $t = 2.24$, $p = 0.04$). Colored branches are scaled according to PLA_2 activity (blue, low activity; red, high activity). Labels as in (A) and see Fig. S12 for PLA_2 activity concentration curves. (C) Principal Coordinate Analysis (PCoA) of cobra (*Naja* spp.) and rinkhals (*H. haemachatus*) PLA_2 toxins derived from top-down venom proteomics reveals major variation between African spitting and non-spitting lineages, but little variation between Asian cobras. Note the convergent placement of *Hemachatus* PLA_2 toxins with those of African spitting cobras. Circle sizes reflect relative abundances of PLA_2s detected in the venom proteomes.

Supplementary Materials for

Convergent Evolution of Defensive Venom Components in Spitting Cobras

T.D. Kazandjian, D. Petras, S.D. Robinson, J. van Thiel, H.W. Greene, K. Arbuckle, A. Barlow, D.A. Carter, R.M. Wouters, G. Whiteley, S.C. Wagstaff, A.S. Arias, L-O. Albulescu, A. Plettenberg Laing, C. Hall, A. Heap, S. Penrhyn-Lowe, C.V. McCabe, S. Ainsworth, R.R. da Silva, P.C. Dorrestein, M.K. Richardson, J.M. Gutiérrez, J.J. Calvete, R.A. Harrison, I. Vetter, E.A.B. Undheim, W. Wüster, N.R. Casewell

Correspondence to: Nicholas.casewell@lstmed.ac.uk

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Materials and Methods

Ethical approvals

Snakes held in the Hepetarium of the Liverpool School of Tropical Medicine were maintained in a temperature, humidity and light-controlled environment, and this facility and its protocols for expert animal husbandry is inspected and approved by the UK Home Office and the LSTM Animal Welfare and Ethical Review Board. Murine venom median lethal dose experiments were undertaken using protocols approved by the Institutional Committee for the Care and Use of Laboratory Animals (CICUA) of the University of Costa Rica (approval number CICUA 82-08). Sensory neuron experiments involving the use of dissected mouse tissue were performed under approval granted by The University of Queensland animal ethics committee. The hen's egg test-chorioallantoic membrane assay was performed in a licensed establishment for the breeding and use of experimental animals (Leiden University) and subject to internal regulations and guidelines, stating that advice is taken from the animal welfare body to minimise suffering for all experimental animals. Toxicology testing on non-self-feeding developmental stages is not considered an animal experiment under the Experiments on Animals Act, the applicable legislation in the Netherlands in accordance with the European guidelines (EU directive no. 2010/63/EU) regarding the protection of animals used for scientific purposes.

Venoms and venom glands

Venoms from cobras and related species were pooled from wild-caught or captive bred (CB) specimens maintained in the Liverpool School of Tropical Medicine Herpetarium. Crude venoms were lyophilized post-extraction and stored at 4°C until analysis. For venom gland transcriptomics, single specimens from each species were euthanized three days after venom

extraction via an overdose of Pentoject (pentobarbital sodium), and dissected venom glands were immediately flash frozen in liquid nitrogen and stored cryogenically until use. Details of the specimens used in this study are displayed in Table S8.

Species tree reconstruction

We reconstructed the phylogenetic relationships and divergence times of 46 species of cobra and related elapid taxa, representing almost all species of *Naja* and cobra-clade elapids under a multispecies coalescent model. This analysis was based on DNA sequence alignments of six genetic loci: mitochondrial DNA (partial CytB and ND4 gene sequences); and the phased partial exon sequences of the nuclear genes CMOS, NT3, PRLR, UBN1 and RAG1. Briefly, we obtained blood samples, ventral scale clips, shed skins or other tissue in the field or from captive collections. Whole genomic DNA was extracted using a Qiagen DNeasy^{TM} Tissue Kit (catalogue no. 69506 – www.qiagen.com), following the manufacturer's instructions, except for blood samples, where PBS buffer was not used and 200 µL of blood in Tris-EDTA buffer was added to 20 µL proteinase K. Where shed skins were used, the samples were left to lyse overnight. PCR was carried out in 15 µL volumes, using ThermoScientific 2X DreamTaq buffer. Primer sequences and thermocycling parameters are given in Table S9. Sanger dideoxy sequencing was carried out by Macrogen (dna.macrogen.com), using the forward primer for mitochondrial genes and forward and reverse primers for nuclear loci.

Sequences were checked, trimmed, translated into amino acid sequences and checked for unexpected indels or non-sense codons in CodonCode Aligner version 3.7.1 (www.codoncode.com). In nuclear locus sequences, heterozygous positions were replaced with the IUPAC uncertainty codes. The alleles of length heterozygotes were reconstructed using the

online tool Indelligent 1.2 (http://dmitriev.speciesfile.org/indel.asp) (*32*). Individual allele sequences (haplotypes) were then estimated from diploid nuclear loci using the software PHASE v. 2.1.1 (*33*, *34*) over 5000 iterations with a burn-in of 500 and a thinning interval of 10, after preparation of the sequence data using SEQPHASE (*35*). PHASE was run three times to confirm burn-in and convergence across multiple runs. PHASE was run separately by genus, or, in the case of *Naja* by subgenus.

Details of the final DNA sequence alignments were as follows: mitochondrial DNA, 1,316 bp, 281 sequences; CMOS, 628 bp, 206 sequences; NT3, 657 bp, 340 sequences; PRLR, 551 bp, 456 sequences; UBN1, 504 bp, 414 sequences; RAG1, 879 bp, 281 sequences. Sample details and accession numbers are shown in Table S7. For phylogenetic reconstruction, we used a multispecies coalescent model employed in *BEAST in BEAST v.1.8.2 (*36*). Our approach was to initially generate a species tree for the entire set of 46 taxa, including virtually all cobras and cobra-like elapids, and then to extract the tree of the focal 17 species from that tree. The multispecies coalescent model implemented in *BEAST coestimates individual gene (locus) trees embedded within the overall species tree while accounting for incomplete lineage sorting. It requires prior assignment of sequences to species, and that each of these taxa be represented by at least one sequence of each locus. We lacked CMOS and RAG 1 sequences for certain taxa (*N. christyi* and *N. sputatrix*), so a single "dummy" sequence consisting of Ns was inserted for each of these taxa into the respective alignments in order to initiate the analysis.

Preliminary *BEAST runs using node calibrations based on the fossil record to estimate divergence times on an unconstrained species tree failed to converge, indicating overparameterisation. To overcome this, we adopted a two-step approach. First, an uncalibrated analysis was carried out estimating the topology of the species tree with the substitution rate of

the mitochondrial alignment fixed at an arbitrary value and the substitution rate of the other gene alignments estimated relative to it. Then, a calibrated analysis was carried out to estimate divergence times by applying internal node calibrations based on the fossil record, with the monophyly of all species tree nodes that were well supported in the uncalibrated analysis (posterior probability > 0.99) constrained.

For the uncalibrated analysis, the precise model specification was determined by trialing more complex models initially, and then incrementally reducing model complexity (in terms of the number of free parameters) in order to achieve convergence. A Yule process was specified for the species tree with a piecewise constant population size model. A single GTR+G nucleotide substitution model with empirical base frequencies was specified for nuclear DNA, by linking the substitution model partitions of the respective five nuclear gene alignments. A separate GTR+G nucleotide substitution model with estimated base frequencies was specified for the mitochondrial alignment. A separate, unlinked strict molecular clock model was specified for each locus. The substitution rate of the mitochondrial locus was arbitrarily fixed to 1, and the substitution rates of the five nuclear loci estimated relative to it by assigning uninformative, uniform priors between 0 and 1. It was necessary to run the MCMC chain for hundreds of millions of generations, sampling every 1,000 generations, to achieve burn in and adequate sampling of parameters. This was run until the physical limits of the computer's hard drive were reached, at which time all parameters had burned in and achieved effective sample sizes > 200, except for mtDNA gene tree root height and the species tree root height, which were > 170, assessed using the program Tracer (*37*). Trees sampled during burn in were removed, and the posterior sample of trees downsampled to approximately 50,000 trees using the program

LogCombiner (*36*). The maximum clade credibility tree was then selected from these ~50,000 trees, with node heights centered on the median of the subsample, using TreeAnnotator (*36*).

We additionally investigated the sensitivity of our phylogenetic analysis to potential undersampling of nuclear loci using a jackknife resampling approach. This involved five replicated BEAST runs, each omitting one of the five nuclear loci, respectively. Other model specifications were as described above for the uncalibrated analysis. Each analysis ran for ~1 billion MCMC generations, sampling every 10,000 generations. All parameters of all runs were checked for convergence and adequate sampling using Tracer (all ESS > 200). MCC tree selection and annotation were as described for the uncalibrated analysis. Each jackknife replicate returned an almost identical MCC species tree topology (see Data S1). The only variable elements were the relationship between the four *Naja* subgenera, and the position of *Naja nivea* within the subgenus *Uraeus*. Critically, in none of these trees did the inferred number of origins for spitting behaviour and morphological adaptations change. We therefore concluded that the principal findings of our phylogenetic analysis would remain functionally unchanged if additional nuclear loci were included.

For the time-calibrated analysis, the model specification was identical except the monophyly of species tree nodes that were well supported in the uncalibrated analysis were enforced, and node calibrations based on the fossil record were used to estimate divergence timings. These were: an exponential prior with a 0.01 MY mean and 17.0 MY offset on the MRCA of *Naja*, *Hemachatus* and *Pseudohaje*, which effectively fixed the age of this node at 17 MYA (*38*); since fossil calibrations constrain the minimum but not the maximum age of a node, this represents the youngest plausible age of this node. We also applied a uniform prior between 0 and 10 MY on the MRCA of the Australian genera *Oxyuranus* and *Pseudechis*, effectively

placing a maximum divergence time on these genera of 10 MYA (*38*). The substitution rate of the mitochondrial locus was estimated based on these calibrations within an uninformative, uniform prior between 0 and 2% per million years. The substitution rates of the nuclear loci were estimated using uninformative uniform priors between 0 and 1.5% per million years. The MCMC chain was run for sufficient length to achieve burn in and effective sample sizes > 200 for all parameters. The maximum clade credibility tree was obtained as described above. We then pruned taxa not included in venom analyses from the tree while retaining the original node ages, and used this tree for all subsequent analyses. The uncalibrated and calibrated species trees, as well as the trees resulting from the jackknife analyses, are shown in Data S1.

Venom gland transcriptomics

Dissected venom glands were homogenized using a pestle and mortar, before RNA was extracted using the TRIzol Plus RNA Purification kit. RNA samples were then DNAse treated and mRNA selected for using the Dynabeads mRNA DIRECT purification kit protocol (Life Technologies). The RNA-Seq libraries were generated from 50 ng of each enriched sample using the TruSeq Stranded mRNA HT Sample Prep Kit (Illumina). As per the Illumina cBot User guide protocol, samples were then denatured and loaded at a concentration of 10 pM, and sequencing was carried out on multiple lanes (six samples multiplexed per lane) of an Illumina MiSeq with 2 × 250 bp paired-end sequencing (Centre for Genomic Research, University of Liverpool). The resulting reads were quality processed using CGR's standard protocols (*11*, *39*, *40*), and then assembled into contigs using the transcriptome assembly program VTBbuilder (*41*). The following parameters were used for assembly: minimum transcript length 150 bp, minimum read length 150 bp, minimum isoform similarity 96%. As part of this process, the VTBuilder

algorithm performs normalized read mapping approach on the assembled contigs to generate relative transcript expression data, which is expressed for each contig as a percentage of the total transcriptome expression (*41*). The resulting assembled contigs were then converted into six frame translations, to provide amino acid databases for later protein spectrum matching. Assembled DNA contigs were also annotated with BLAST2GO Pro v3 (*42*) using a BLASTx algorithm and a threshold of 1e-3 against the NCBI non-redundant protein database (GenBank release 219). Manual curation of BLAST2GO annotations were undertaken using the BLASTx annotations on all contigs with expression levels of 0.1% of the transcriptome or higher. Based on these annotations, contigs were classified into toxins (those showing homology with known snake venom toxins previously identified in the literature), non-toxins or 'unknown' (if no BLAST annotation or only hypothetical matches were detected). Toxins were further classified by toxin family and the contigs of all toxin families with >1% expression in at least one cobra species were subjected to downstream analysis in MEGA v7 (*43*). As a quality control step, contigs were removed from these datasets if any of the following conditions were not met: i) sequence length was less than 100 bp, ii) 50% or more of the sequence did not match the target toxin in a BLASTx search, iii) the first 50% of the sequence was interrupted by a stop codon, indicating the presence of a pseudogene or misassembly, or iv) the sequence was made up of two exonic regions interspersed by an intron, indicating genomic DNA contamination. For some sequences, it was apparent that 'underclustering' had occurred during the assembly process; a common by product of assembling isoform rich gene data (*41*), but more desirable than the alternative scenario of frequently producing chimeric contigs. To resolve this issue, contigs were merged if a $\geq$50 bp overlap resulted in $\geq$98% similarity. Sequences were then trimmed to the

open reading frame and aligned using the MUSCLE algorithm (*44*) in MEGA's amino acid space, before the alignments were visually inspected for errors, and edited manually if required.

Top-down venom proteomics

Denaturing top-down proteomic experiments were performed as described in previous papers (*45*, *46*). In short, venom samples were dissolved in ultrapure water to a final concentration of 10 mg/mL, and centrifuged at 12,000 x g for 5 min. For reduction of disulfide bonds, 10 µL of dissolved venom was mixed with 10 µL of 0.5 M tris(2-carboxyethyl) phosphine (TCEP), and 30 µL of 0.1 M citrate buffer (pH 3). After 30 min incubation at 65 °C, samples were mixed with 50 µL of acetonitrile/formic acid/H_2O (10:1:89, v/v/v) and centrifuged at 12,000 x g for 5 min. After centrifugation, 5 µL of supernatant of reduced samples were injected for LC-MS/MS analyses. LC-MS/MS experiments of two technical replicates were carried out on a Vanquish ultra-high-performance liquid chromatography (UHPLC) system coupled to a Q-Exactive hybrid quadrupole orbital ion trap (Thermo Fisher Scientific, Bremen, Germany). LC separation was performed on a Supelco Discovery Biowide C18 column (300Å pore size, 2 x 150 mm column size, 3 µm particle size) at a temperature of 30 °C. A flow rate of 0.5 mL/min was used and the samples were eluted with a gradient of water with 0.1% formic acid (FA) (solution A) and 0.1% FA in acetonitrile (ACN) (solution B). Gradient elution started with an isocratic phase with 5% B for 0.5 min, followed by a linear increase to 40% B for 50 min, and 40-70% B for 60 min. Finally, the column was washed with 70% B for 5 min and re-equilibrated at 5% B for 5 min. ESI settings of the mass spectrometer were adjusted to 53 L/min sheath gas, 18 L/min auxiliary gas, spray voltage 3.5 kV, capillary voltage 63 V, S lens RF level 90 V, and capillary temperature 350 °C. MS/MS spectra were obtained in data dependent acquisition (DDA) mode.

Mass spectra were acquired with 1 micro scan and 1000 ms maximal C-trap fill time. AGC targets were set to 1E6 for MS1 full scans and to 3E5 for MS/MS scans. Both, the survey scan and data dependent MS/MS scans were performed with a mass resolution (R) of 140,000 (at m/z 200). For MS/MS, the three most abundant ions of the survey scan with known charge were selected for Higher-energy C-trap dissociation (HCD) at the apex of a peak within 2 to 15 s from their first occurrence. Normalized collision energy was stepwise increased from 25% to 30% to 35%. The default charge state was set to z = 6, and the activation time to 30 ms. The mass window for precursor ion selection was set to 3 m/z. A window of 3 m/z was set for dynamic exclusion within 30 s. Ion species with unassigned charge states as well as isotope peaks were excluded from MS/MS experiments.

For data analysis, LC-MS/MS .raw files were converted to .mzXML file format using MSconvert of the ProteoWizard package (version 3.065.85). Extracted ion chromatograms (XICs) of intact proteins were generated of deconvoluted of multiple charged spectra with XTRACT of the Xcalibur Qual Browser version 2.2 (Thermo, Bremen, Germany). XICs of mono-isotopic deconvoluted LC-MS runs were performed with MZmine 2 (version 2.2). A 1.0E4 signal intensity threshold was used. The mass alignment for the creation of XICs was performed with a minimum peak width of 30 s, and 3.0E4 peak height. Mass tolerance was set to 10 ppm. For chromatographic deconvolution, the baseline cutoff algorithm with 1.0E4 signal threshold was used. Maximum peak width was set to 10 min. Feature alignment was performed with 10 ppm mass accuracy and 0.5 min retention time tolerance. For protein spectrum matching, multiple charged MS/MS spectra were then deconvoluted using MS-Deconv (version 0.8.0.7370). The maximum charge was set to 30, maximum mass was set to 50,000, signal-to-noise threshold was set to 2, and m/z tolerance was set to 10 ppm. Protein spectrum matching

was performed using TopPIC (http://proteomics.informatics.iupui.edu/software/toppic/) (*47*) (version 1.1.0) against the corresponding venom gland transcriptomic derived protein sequence database as well as all protein sequences from members of the genus *Naja* obtained from the NCBI protein database (Nov 2017), containing a total of 1,764 sequences. The search database also contained 116 protein sequences of known typical contaminants from the common Repository of Adventitious Proteins, (cRAP). TopPIC mass error tolerance was set to 10 ppm and a 1% false discovery rate (FDR, target-decoy) cut-off was used. The maximal allowed unexpected PTMs were set to one. Pairing of MS/MS derived protein ID from TopPic with MS1 XICs was performed by mass matching with an in-house R script available as a jupyter notebook at https://github.com/DorresteinLaboratory/match_tables_by_exact_mass. For relative quantification, area under the curve of XICs was normalized to total ion current (TIC) (see Data S2 and Fig. S2). Amino acid sequences of proteins were converted to FASTA format and imported into MEGA v7 (*43*) for alignment. Proteins were merged if found to be identical using MEGA's pairwise distance measurement.

Data analyses

Ancestral states for spitting (vs non-spitting) were estimated over the species tree using the rerooting method function of the R package "phytools" (*48*). After Log-likelihood (Log-lik) comparison of the all rates different (ARD), equal rates (ER) and symmetric rates (SYM) models, both ER and SYM (which had equal Log-lik values) were used to plot these discrete traits over the tree, with both resulting in visually-identical plots. Ancestral states for all continuous characters were also estimated via maximum likelihood on a pruned version of this species tree, using the contmap function of the phytools package under default settings. To test

for statistical support for differences between spitting and non-spitting cobras and the influences of geography at the proteomic and functional levels, we used the phylogenetic Generalized Least Squares (PGLS) approach using the pgls() function of the "caper" (*49*) package in RStudio (*50*), with the formula set as ([toxin characteristic, e.g. PLA_2 abundance] ~ grouping [spitter or non-spitter]) and lambda set to "Maximum Likelihood". Analyses were performed on the species tree pruned to those species whose venom was analysed in this study. To test for the clustering of spitting and non-spitting lineages based on the amino acid sequences of toxin proteins, we used Principal Coordinate Analysis (PCoA). To do this, a pairwise distance matrix was firstly conducted on the toxin amino acid sequences using the JTT matrix-based model (*51*) in MEGA v.7. The rate variation among sites was modelled with a gamma distribution (shape parameter = 0.8657) for PLA_2s and uniform rates for CTXs. A Euclidean dissimilarity matrix was generated from each of these matrices using the Daisy() function from the R package "Cluster" (*52*) and was followed by applying classical multidimensional scaling to the matrix using the cmdscale() function in R Studio.

Phylogenetic analyses

Phylogenetic analysis of three-finger toxin (3FTX) gene family was conducted using Bayesian inference on the aligned sequences derived from the venom gland transcriptomes. Construction of an initial 3FTX tree also used representative sequences from *Aspidelaps scutatus intermedius* (*40*), *Bungarus flaviceps* (*53*), *Bungarus multicinctus* (*54*), *Dendroaspis* spp. (*11*), *Micrurus fulvius* (*55*) and *Ophiophagus hannah* (*56*, *57*) for phylogenetic context, and a non-toxic *Python regius* 3FTX sequence (*58*) was used as the outgroup for rooting the tree. The final DNA alignment consisted of 56,781 bp with 285 sequences from 28 taxa.

Construction of a PLA_2 gene tree was also performed in a similar fashion, using a DNA alignment containing toxin PLA_2 sequences sourced from *Aspidelaps scutatus intermedius* (*40*), *Bungarus caeruleus* (*59*, *60*), *Bungarus candidus* (*61*), *Bungarus fasciatus* (*62*), *Bungarus flaviceps* (*53*, *61*), *Bungarus multicinctus* (*63*, *64*), *Micrurus altirostris* (*65*), *Micrurus corallinus* (*66*), *Micrurus fulvius* (*55*), *Micrurus laticollaris* (*67*), *Micrurus tener* (*68*) and *Ophiophagus hannah* (*57*, *69*, *70*), and non-toxin PLA_2 sequences from *Python bivittatus* (*71*) and *Anolis carolinensis* (*72*), with the latter used to root the tree. The final DNA alignment consisted of 86 sequences from 29 taxa, representing 35,268 bp.

The selected models for nucleotide sequence evolution, "GTR+G" for the 3FTX and "TIM3+G" for the PLA_2 dataset, were determined by jModelTest v2.1.6 based on the Akaike Information Criterion (*73*, *74*). Subsequently, Bayesian inference phylogenetic analysis was performed using MrBayes v3.2.6 (*75*) on the CIPRES Science Gateway (*76*). The analyses used four simultaneous runs with four chains (three hot, one cold) for 10×10^6 generations, sampling every 500th cycle from the chain and using default settings for priors. Burn-in was set at 25%; trees generated prior to this point were discarded, and a consensus tree constructed from the remaining 75%.

From the resulting initial 3FTX tree, a strongly supported CTX clade was identified via BLASTx searches on individual sequences and a separate pared-down CTX-specific phlyogenetic analysis was also performed to explore the evolutionary history of CTXs sequences in elapids. In addition to the CTXs identified in the original analysis, this second dataset included CTX sequences from *Ophiophagus hannah* (*56*, *57*) and used the non-CTX 3FTX from *Bungarus flaviceps* (GU190795.1) (*53*) as the outgroup for rooting. The resulting DNA alignment consisted of 18,098 bp with 80 sequences from 17 taxa, and phylogenetic analysis was

performed as described above, except using the selected "TVM+G" model of nucleotide substitution identified by jModelTest. The 3FTX sequence alignment is displayed in Data S3 and the CTX sequence alignment displayed in Data S4, with the resulting phylogeny of the CTX-specific analysis is displayed in Fig. S3. The PLA_2 sequence alignment is displayed in Data S5 and the resulting phylogeny in Fig. S13B.

Venom cytotoxicity via hen's egg test-chorioallantoic membrane (HET-CAM) assay

Lyophilized venom (Table S10) was dissolved in sterile, deionized water to make a stock solution. The protein concentration was determined using a NanoDrop 1000 UV/Vis Spectrophotometer (Thermo Fisher Scientific, Bleiswijk, the Netherlands) at an absorbance of 280 nm. The stock solution was snap-frozen using liquid nitrogen and stored at −80°C till further use. Just before the experiment the venom was prepared in Hanks' balanced salt solution (HBSS: Cat. H9394, Sigma Aldrich, Zwijndrecht, the Netherlands) to a final total protein concentration of 1 mg/mL. Unincubated hen's eggs were obtained from Drost Loosdrecht B.V. (Loosdrecht, the Netherlands). The eggs were incubated in a horizontal position at 38°C on stationary shelves in a humidified incubator for 10 days. Following incubation, eggs were candled and placed vertically with the blunt end upwards. A hole in the eggshell overlying the air sac was made with No. 3 watchmakers' forceps. The shell membrane was moistened with 200 µL HBSS, and then punctured and carefully peeled away using No. 5 watchmakers' forceps to reveal the chorioallantoic membrane (CAM) below. Excess HBSS was pipetted away. Venom or control solution (100 µL) was then pipetted onto the CAM and a timer started. Any hyperaemia, haemorrhage and/or coagulation was documented over a 5 min period. Photographs were taken at 15 s intervals for semi-quantitative analysis. Each venom and control sample was tested in

triplicate (technical replicates using three different eggs) in a single-blinded and randomized manner to avoid bias. The positive controls were 1% sodium dodecyl sulphate (Carl Roth, Karlsruhe, Germany) and 1 mg/mL of capsaicin (Sigma Aldrich, Zwijndrecht, the Netherlands), both known eye-irritants; and 10 mg/mL of histamine (Sigma Aldrich, Zwijndrecht, the Netherlands), a potent vasodilator. The negative control was HBSS. A semi-quantitative analysis was performed using the resulting photographs, where the severity of any hyperaemia, haemorrhage and/ or coagulation was manually scored at 0.5, 2.0 and 5.0 min using the method developed by Luepke (*77*) (Tables S3, S4). The average score of three replicates was taken and the cumulative score was matched with the corresponding 'irritation potential' previously developed (*77*) (Table S4).

Calcium imaging on mouse sensory neurons

Trigeminal ganglia (TG) or dorsal root ganglia (DRG) from 4-6 week old male C57BL/6 mice (Animal Resource Centre, WA, Australia), were dissociated and plated in DMEM (Gibco, MD, USA) containing 10% fetal bovine serum (FBS) (Assaymatrix, VIC, Australia) and penicillin/streptomycin (Gibco, MD, USA) on a 96-well poly-D-lysine-coated culture plate (Corning, ME, USA) and maintained overnight. Cells were loaded with Fluo-4 AM calcium indicator, according to the manufacturer's instructions (ThermoFisher Scientific, MA, USA). After loading (1 h), the dye-containing solution was replaced with assay solution (1x Hanks' balanced salt solution, 20 mM HEPES). Images were acquired at 20x objective at 1 frame per second (excitation 485 nm, emission 521 nm). Fluorescence corresponding to $[Ca^{2+}]_i$ of 100–150 cells per experiment was monitored in parallel using an Nikon Ti-E Deconvolution inverted

microscope, equipped with a Lumencor Spectra LED Lightsource. Baseline fluorescence was monitored for 30 s. At 30 s, assay solution was replaced with venom (100 ng/µL in assay solution) or venom fractions (estimated equivalent of amount for 100 ng/µL whole venom, in assay solution) and monitored for an additional 2 min.

FLIPR assays of F11 cells

F11 (Neuroblastoma X dorsal root ganglion (DRG) neuron hybrid) cells (catalog #08062601, European Collection of Authenticated Cell Cultures [ECACC]) were cultured as previously described (*78*). Briefly, cells were maintained on Ham's F12 media supplemented with 10 % fetal bovine serum, 100 µM hypoxanthine, 0.4 µM aminopterin, and 16 µM thymidine (Hybri-MaxTM, Sigma Aldrich). A 384-well imaging plate (Corning, Lowell, MA, USA) was seeded 48 h prior to calcium imaging resulting in 90 – 95% confluence at imaging. Cells were incubated for 30 min with Calcium 4 assay component A, according to manufacturer's instructions (Molecular Devices, Sunnyvale, CA) in physiological salt solution (PSS; composition in mM: 140 NaCl, 11.5 D-glucose, 5.9 KCl, 1.4 $MgCl_2$, 1.2 NaH_2PO_4, 5 $NaHCO_3$, 1.8 $CaCl_2$, 10 HEPES) at 37 °C. Ca^{2+} responses were measured using a FLIPRTETRA fluorescent plate reader equipped with a CCD camera (Ex: 470-490 nm, Em: 515-575 nM) (Molecular Devices, Sunnyvale, CA). Signals were read every second for 10 s before, and 300 s after the addition of crude venoms (final concentration range: 1.3 – 333 ng/µL) in PSS supplemented with 0.1% bovine serum albumin. All data represent the mean ± standard error of the mean (SEM) of a representative assay in triplicate unless otherwise stated. The mean fluorescence changes during venom addition was compared to a control solution addition, (0.1% BSA in PSS). The resulting maximum-minimum

florescence in the 300 s period after venom was added was recorded as the response. A four-parameter Hill equation (Variable slope, two-site) was fitted using Graphpad Prism8.

Venom fractionation and bottom-up sequencing

To identify the venom constituents responsible for sensory neuron activation, 500 µg of venoms from *N. siamensis, N. nigricollis* and *H. haemachatus* were each separated on a Phenomenex Gemini NX-C18 column (250 x 4.6 mm, 3 µm particle size, 110 Å pore size) using a gradient of 15–45% solvent B (90% ACN, 0.05% TFA) over 30 min at a flow rate of 1 mL min^{-1}. Fractions were collected on the basis of absorbance at 214 nm. Fractions were dried by vacuum concentration and each resuspended in 50 µL pure water from which 1 µL aliquots were used for calcium imaging experiments (diluted in 100 µL assay solution). Bottom-up proteomics was used to confirm the identity of the major component/s of each fraction. 20 µL of each resuspended fraction was dried by vacuum centrifugation. Each sample was reduced and alkylated with reagents in gas phase according to the protocol described by Hale et al. (*79*). Reduction/alkylation reagent (50% 0.1 M ammonium carbonate, 48.75% ACN, 0.25% triethylphosphine, 1% 2-iodoethanol (by volume)) was placed in the cap of each inverted tube and incubated at 37 °C for 60 min. Reduced and alkylated fractions were then digested by incubating with 20 ng/µL trypsin overnight at 37 °C according to the manufacturer's instructions (Sigma-Aldrich, MO, USA). Trypsin digestion was quenched by addition of FA to a final concentration of 0.5%. Each sample was then separated on a Shimadzu (Japan) Nexera uHPLC with an Agilent Zorbax stable-bond C18 column (2.1 mm × 100 mm, 1.8 µm particle size, 300 Å pore size), using a flow rate of 180 µL/min and a gradient of 1–40% solvent B (90 % ACN, 0.1% FA) in 0.1% FA over 30 min and analyzed on an AB Sciex 5600 TripleTOF mass spectrometer.

MS survey scans were acquired at 300–1800 m/z over 250 ms, and the 20 most intense ions with a charge of +2 to +5 and an intensity of at least 120 counts were selected for MS/MS. The unit mass precursor ion inclusion window mass ± 0.7 Da, and isotopes within ±2 Da were excluded from MS/MS, with scans acquired at 80–1400 m/z over 100 ms and optimized for high resolution. Using ProteinPilot version 4.0 (ABSciex), MS/MS spectra were searched against a database of the toxin sequences (translated open-reading frames) derived from the corresponding venom gland transcriptomes.

Phospholipase A_2 assay

We characterized the enzymatic PLA_2 activity of each venom using a recently described colorimetric assay (*80*). A PLA_2 reaction solution was made using 49.5 µM Cresol red dye, 0.875 mM Triton X-1007 M, and 1 mL of 5x Salt Mix (1mM Tris base pH 8.5, 500 mM sodium chloride, 500 mM potassium chloride and 50 mM calcium chloride). The pH of the reaction solution was checked using a pH strip and corrected (if necessary) with 1 M sodium hydroxide. To this solution 168.25 µL of the substrate (L-α-phosphatidylcholine; stock concentration 26 mM, SIGMA Life Sciences) was added. Venom samples were prepared in tenfold dilutions (0.1-100 µg/mL) of Tris buffer (pH 8.5), with 10 µL pipetted into the wells of a Greiner Bio-One clear 384-well microplate at volumes in quadruplicate (technical replicates) using a 230V Multidrop 384 (Labsystems), with control wells containing Tris buffer only. The plate was then overlaid with 40 µL of reaction solution in each well and read kinetically on a FLUOstar Omega (BMG LABTECH) spectrophotometer at 572 nm over 42 cycles (~46 s/cycle) at 25 °C. Concentration curves were generated for each venom by plotting the mean control AUC of absorbance at 572 nm, minus the mean area under the curve (AUC) for readings at each venom

concentration, using the values from the run period of 0-10 min (Fig. S12). The resulting mean area under the concentration curves (AUCC) were used for comparisons.

Venom median lethal doses (LD_{50})

We calculated the lethal dose 50 (LD_{50}) of each of the venoms using a previously established murine model of venom lethality (*81*). Mixed sex CD-1 mice (18-20 g, Animal Colony of Instituto Clodomiro Piacado, Costa Rica) were randomly allocated into groups (n=5), with each group receiving an intravenous (i.v.) injection, in the caudal vein, of various doses of each venom (Table S5), dissolved in 0.1 mL of 0.12 M NaCl, 0.04 M phosphates, pH 7.2 (PBS). Deaths were recorded at 24 h and the LD_{50}, as well as the 95% confidence limits, were estimated by probit analysis.

Retesting statistical associations with *N. atra* and *N. kaouthia* recoded as spitting species

While the Asian cobras *N. atra* and *N. kaouthia* have been treated as non-spitters throughout this study, recent reports suggest that certain individuals/populations of these two species show some spitting ability (*18, 19*). Consequently, we recoded these two species as spitting cobras and repeated our statistical analyses. The result of these analyses revealed only minor differences in statistical significance for the various analyses, suggesting that the coding of these two species does not impact upon our conclusions. The results of these analyses are displayed in Table S6.

CT imaging of snake fangs

Fangs were scanned in pairs on a Nikon XTH225ST CT scanner at the University of Bristol. All scans were performed on a rotating target with a tube voltage of 70 kV and beam current of 101

μA with a 1 s exposure per projection with a total of 3141 projections in 360° at a resolution of between 3.9 to 4.5 microns. The CT scan data were aligned, reconstructed and cropped to generate individual tiff image stacks for each fang using VG Studio MAX (Volume Graphics GmbH, Heidelberg, Germany). The tiff stacks were subsequently segmented to generate a 3D model of each fang in Avizo 9 (FEI, Hillsboro, Oregon, USA).

Supplementary Figures

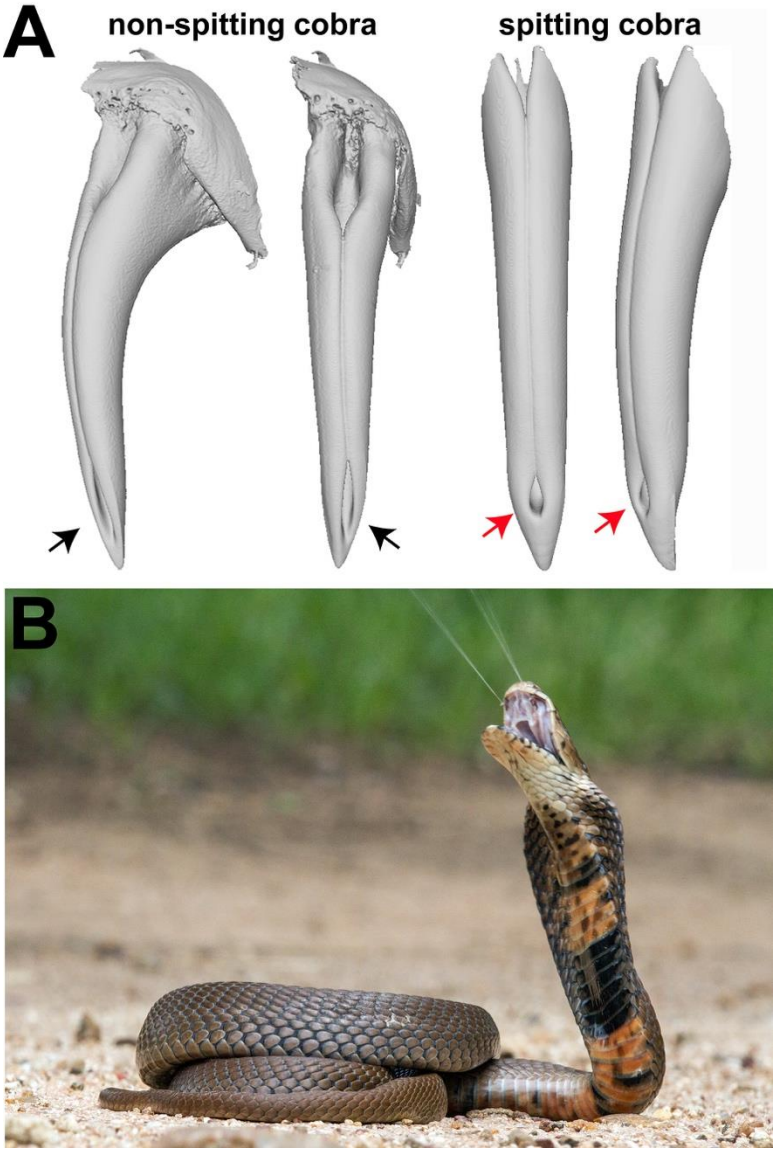


Fig. S1. The morphological adaptations associated with venom spitting. (A) MicroCT visualizations of fangs from representative non-spitting (*Naja nivea*) and spitting (*Naja nubiae*) cobra species. Fangs are visualized from the anterior perspective (central images) and rotated 45° (outer images). Black and red arrows highlight the distinct ejection orifices of non-spitting and spitting cobras, respectively. (B) Venom spitting displayed by a *Naja mossambica* from Limpopo Province, South Africa. Photograph by Wolfgang Wüster.

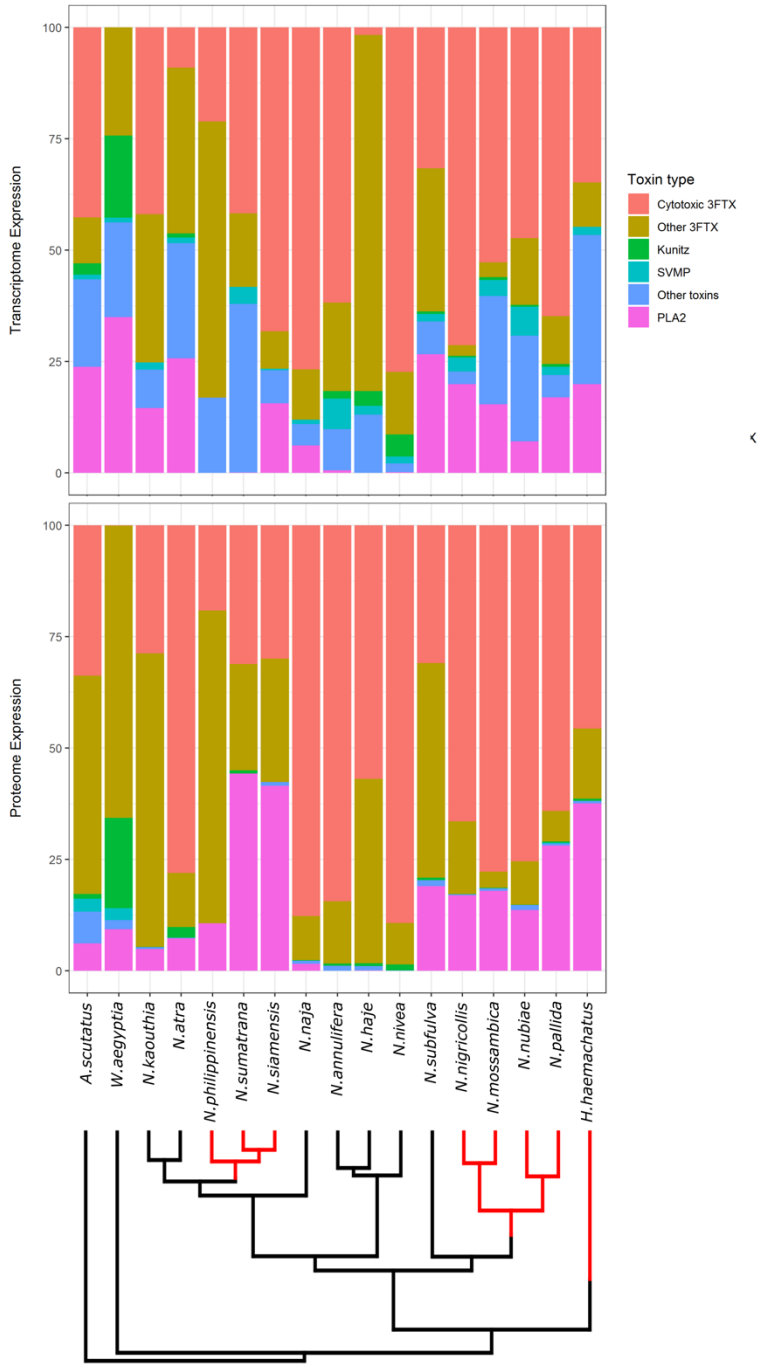


Fig. S2. Stacked bar charts representing the relative abundances of toxin families detected in the venom gland transcriptomes (top) and venom proteomes (bottom) of cobras and related species. The transcriptome data presented represents the percentage of all venom toxin-encoding gene expression detected in each venom gland transcriptome based on normalized read mapping to assembled contigs. The proteome data represents the percentage of all venom toxins detected in each venom proteome based on top-down proteomic analyses and quantification via area under the curve of extracted ion chromatograms (XICs) normalized to total ion current (TIC). The bars are plotted over the phylogenetic tree of the species used in this study. Branches of the tree

highlighted in red represent spitting lineages. Note that the transcriptomic and proteomic abundances for *A. scutatus* are taken from (*40*), and while the transcriptomic data is directly comparable, the proteomic analyses used a different measure of quantification than that applied for the remainder of the species investigated in this study.

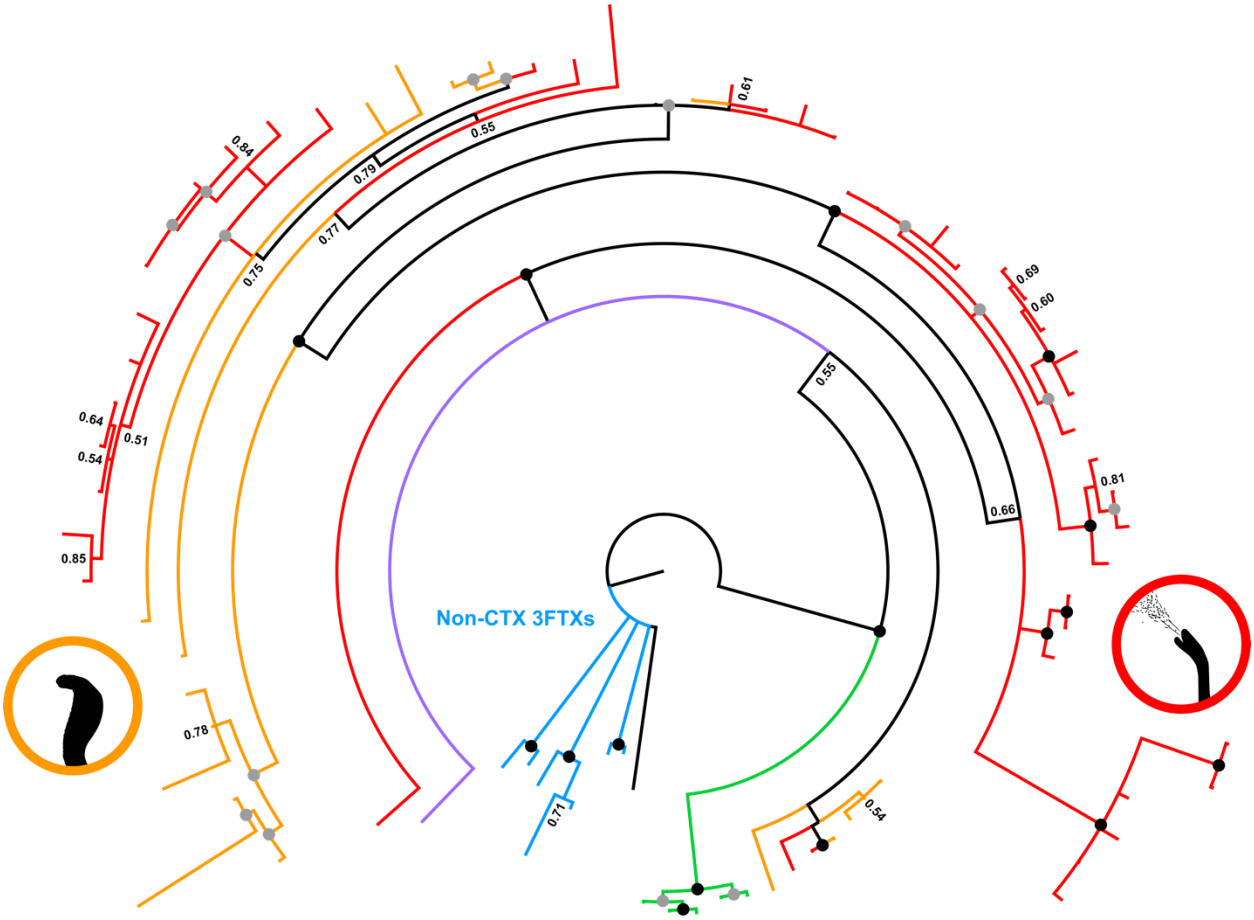


Fig. S3. Bayesian inference-derived DNA phylogeny of elapid three finger toxin cytotoxin (CTX) sequences demonstrates that the vast majority are from cobra species (genus *Naja*). Black-filled circles at each node represent Bayesian posterior probabilities of 1 and grey-filled circles represent values between 0.90 and 0.99. Tips labels have been removed to aid visualization. Colored branches represent non-CTX 3FTX outgroups (blue) and CTXs found in: the king cobra *Ophiophagus hannah* (green), *Aspidelaps scutatus* (purple), non-spitting cobras (orange), and spitting cobras (including *H. haemachatus*) (red). No bona fide CTX sequences were detected in the venom gland transcriptome of *Walterinnesia aegyptia*. The sequence alignment used for tree construction is displayed in Data S4.

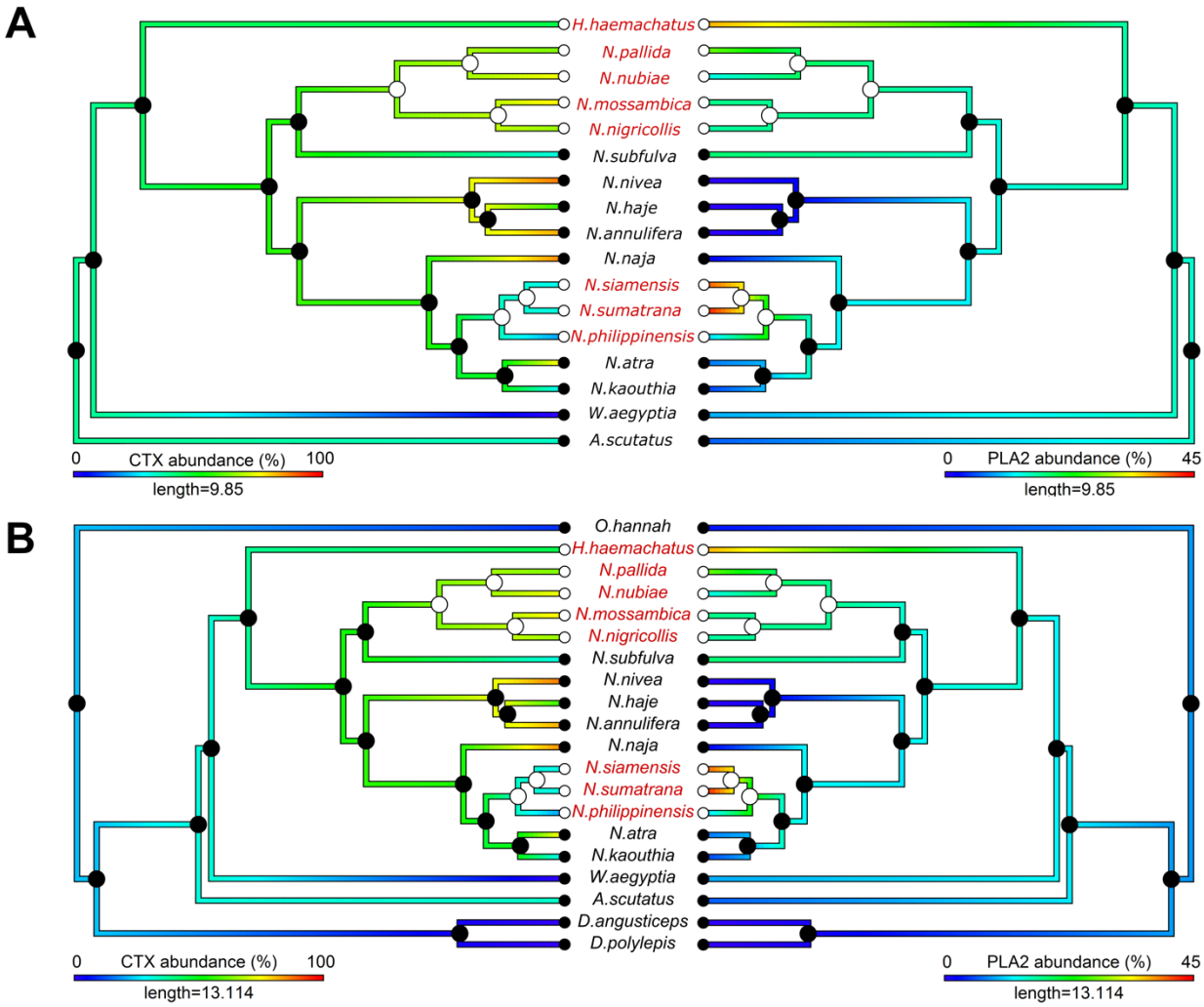
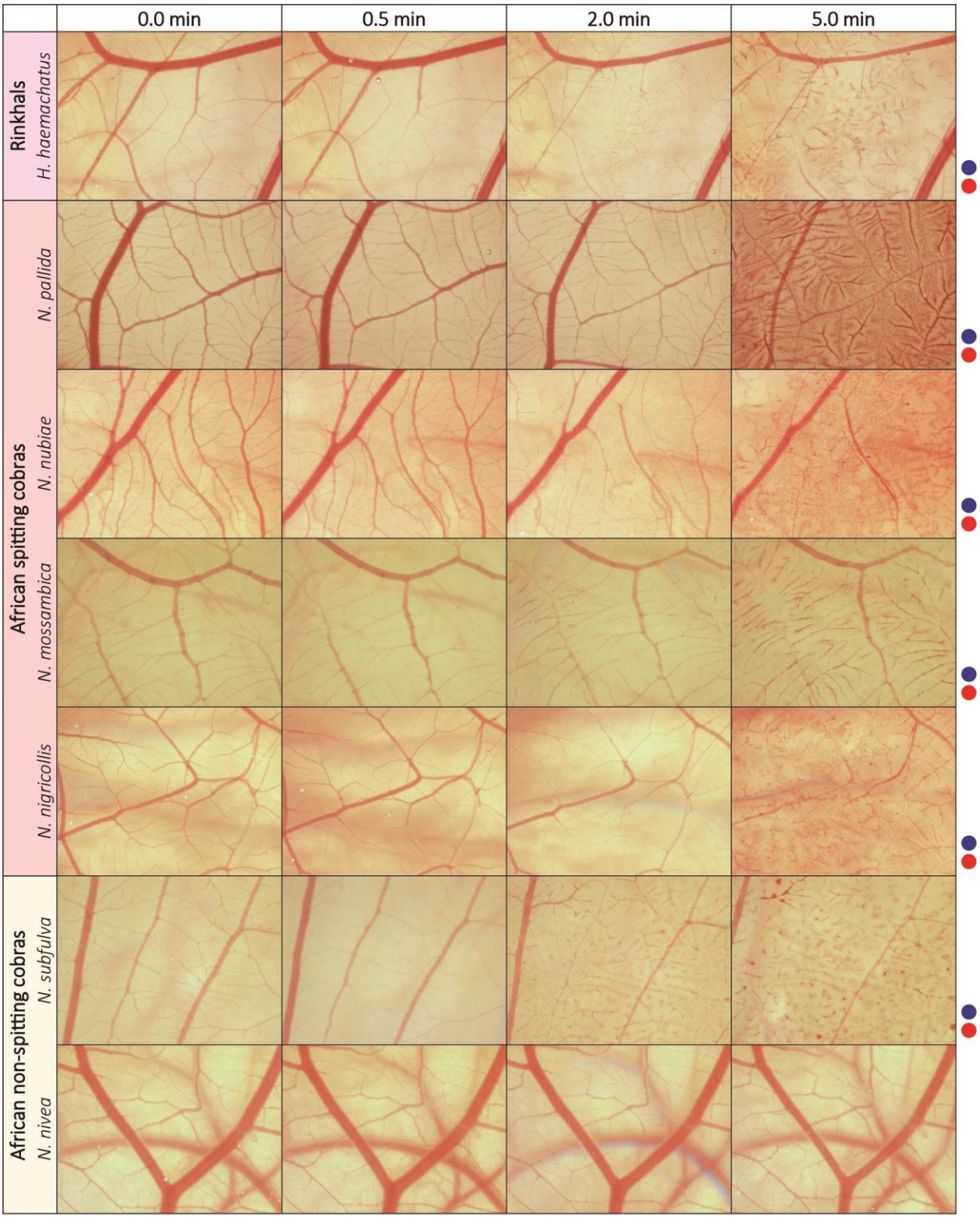
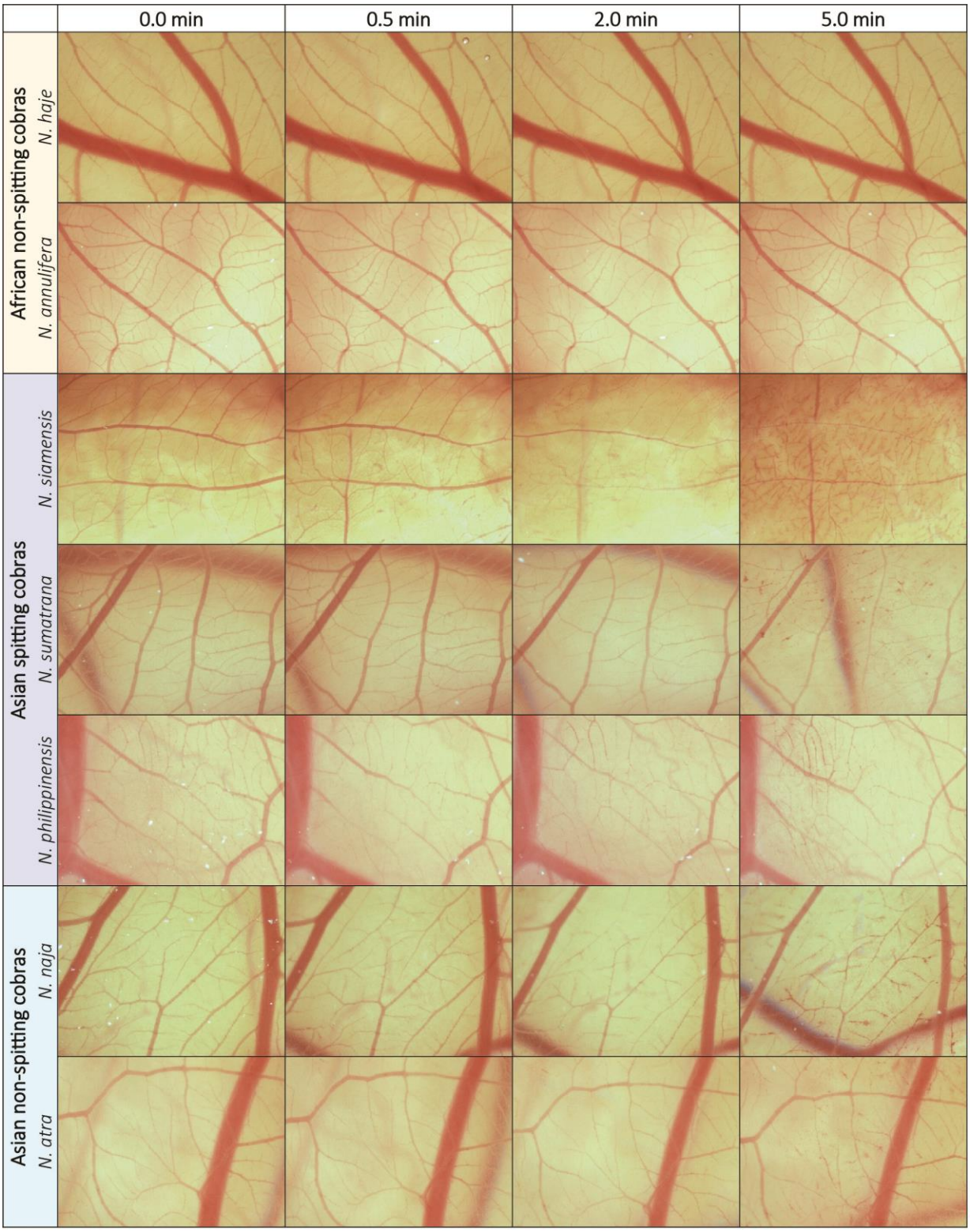


Fig. S4. Proteomic abundances of phospholipase A_2 (PLA_2) but not cytotoxic 3FTXs (CTXs) are significantly associated with the emergence of venom spitting. (A) Ancestral state estimations of the proteomic abundance (percentage of all toxins) of CTXs (left) and PLA_2s (right) mapped onto a species tree containing the cobras and representatives of the outgroup genera *Walterinnesia* and *Aspidelaps*. (B) Ancestral state estimations of the proteomic abundance of CTXs (left) and PLA_2s (right) mapped onto a species tree containing those taxa displayed in A, as well as the additional more distantly related elapid outgroups *Dendroaspis* (*D. angusticeps* and *D. polylepis*) and *Ophiophagus hannah*. Additional outgroups were incorporated due to the high reconstructed ancestral abundance of CTXs observed in A. For both (A) and (B) filled or empty circles at tips and nodes represent the observed or estimated ancestral state of non-spitting and spitting, respectively. Red tip labels are also used to highlight the spitting lineages. The Phylogenetic Generalized Least Squares (PGLS) statistics for each trait are as follows: CTX abundance with *Aspidelaps* as outgroup (A, left); t = -0.83, df = 15, p = 0.42, and with *Ophiophagus* as outgroup (B, left); t = -0.72, df = 18, p = 0.48. PLA_2 abundance with *Aspidelaps* as outgroup (A, right); t = 4.27, df = 15, p = 0.0007, and with *Ophiophagus* as outgroup (B, right); t = 5.12, df = 18, p = 0.00007. Proteomic abundances for *Aspidelaps scutatus*, *Dendroaspis* spp. and *Ophiophagus hannah* were taken from (*11*, *40*, *82*).


0.0 min
0.5 min
2.0 min
5.0 min
Rinkhals
H. haemachatus
African spitting cobras
N. pallida
N. nubiae
N. mossambica
N. nigricollis
African non-spitting cobras
N. subfulva
N. nivea


0.0 min
0.5 min
2.0 min
5.0 min
African non-spitting cobras
N. haje
N. annulifera
Asian spitting cobras
N. siamensis
N. sumatrana
N. philippinensis
Asian non-spitting cobras
N. naja
N. atra

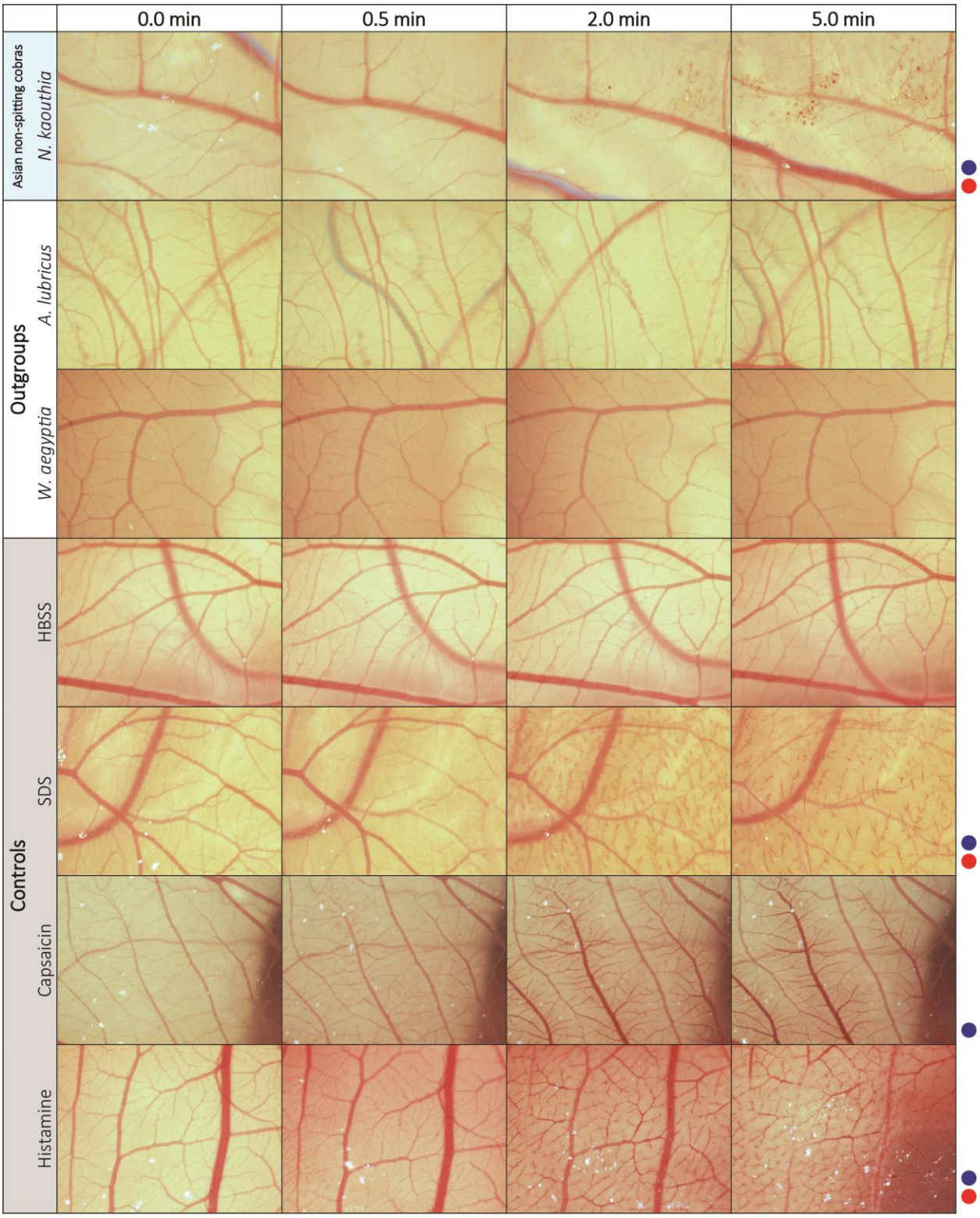


Fig. S5. The vascular pathologies induced by spitting cobra venoms. The eye-irritation potential of snake venoms and controls were assessed using the *in vivo* hen's egg test-chorioallantoic membrane (HET-CAM) assay (*77*). The vascular effects observed are indicated

on the right of each series if present: hyperaemia (●) and coagulation (●). All series were also assessed for evidence of haemorrhage, but this pathology was not observed in any sample. Key: HBSS, Hanks' balanced salt solution; SDS, sodium dodecyl sulphate.

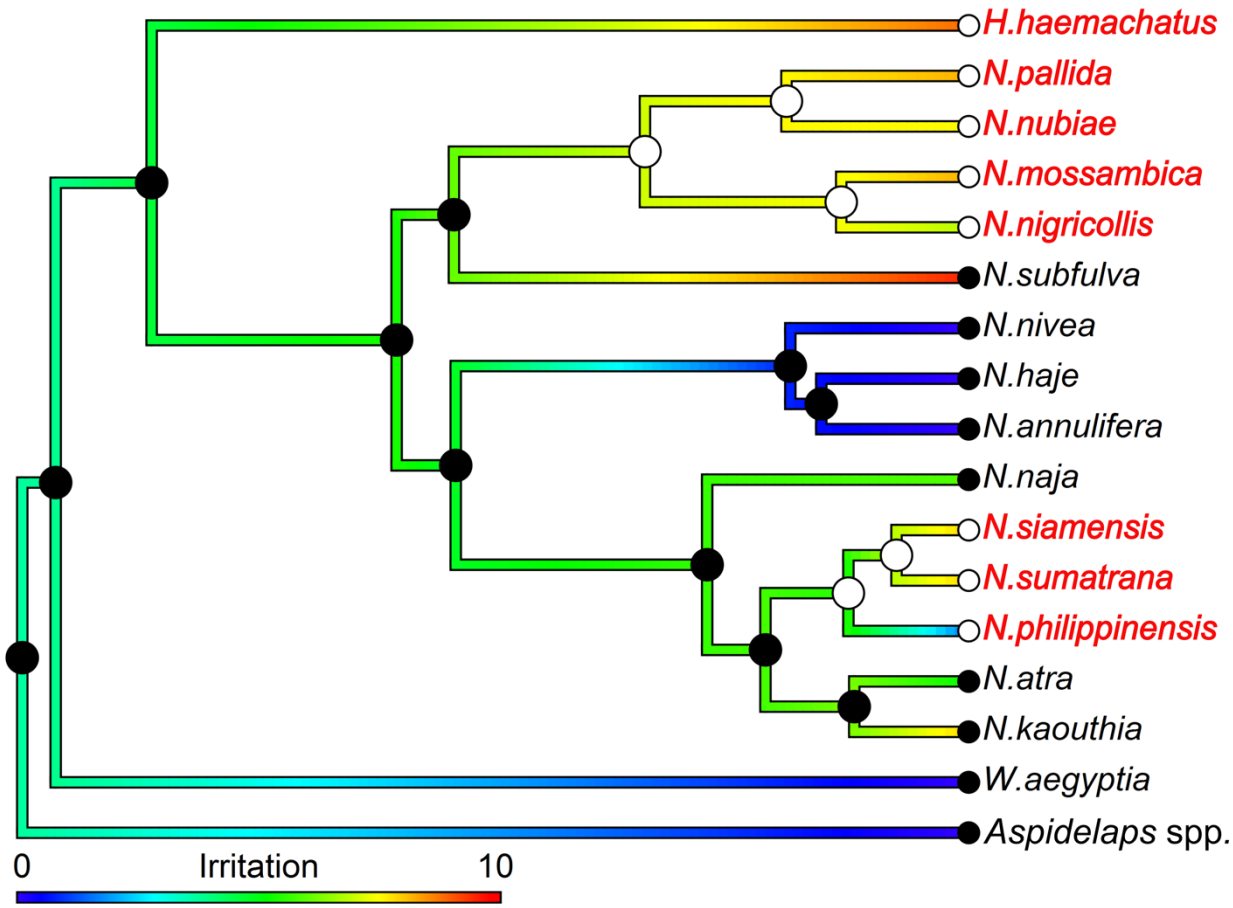


Fig. S6. Ancestral state estimation of eye-irritation scores, determined in the hen's egg test-chorioallantoic membrane assay, demonstrate no significant association with the emergence of venom spitting. The Phylogenetic Generalized Least Squares (PGLS) analysis revealed no association between the extent of venom-induced eye irritation and the emergence of spitting ($t = 1.08$, $df = 15$, $p = 0.30$). Colored branches are scaled according to extent of irritation (blue, low abundance; red, high abundance). At each node and tip of the tree, filled or empty circles represent the estimated ancestral state of non-spitting or spitting, respectively, and red tip labels are used to highlight spitting lineages.

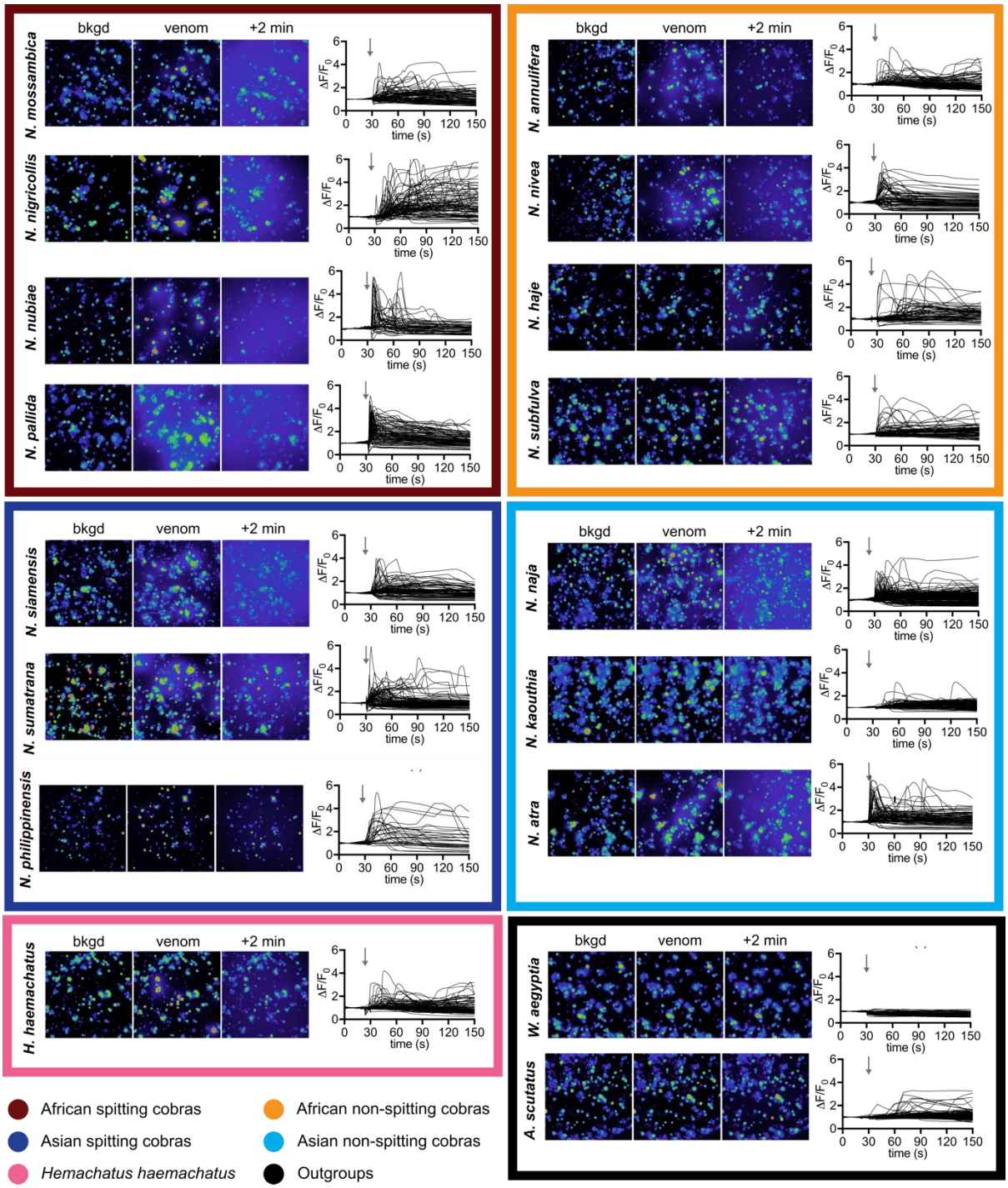


Fig. S7. *Naja* and *Hemachatus* venoms activate sensory neurons *in vitro*. Application of *Naja* and *Hemachatus* venoms (100 ng/µL) to dissociated mouse trigeminal ganglion cells caused, over the 2 min recording, increases in $[Ca^{2+}]_i$ in both neurons and non-neuronal cells, which was followed by a decrease in $[Ca^{2+}]_i$ reflecting dye leakage from the cells into the media. These data indicated that components of the venom permeate cell membranes allowing a massive influx of Ca^{2+} (and likely other ions) followed by leakage of cellular contents. Such an action on trigeminal ganglion neurons is consistent with the painful effects reported for envenomation.

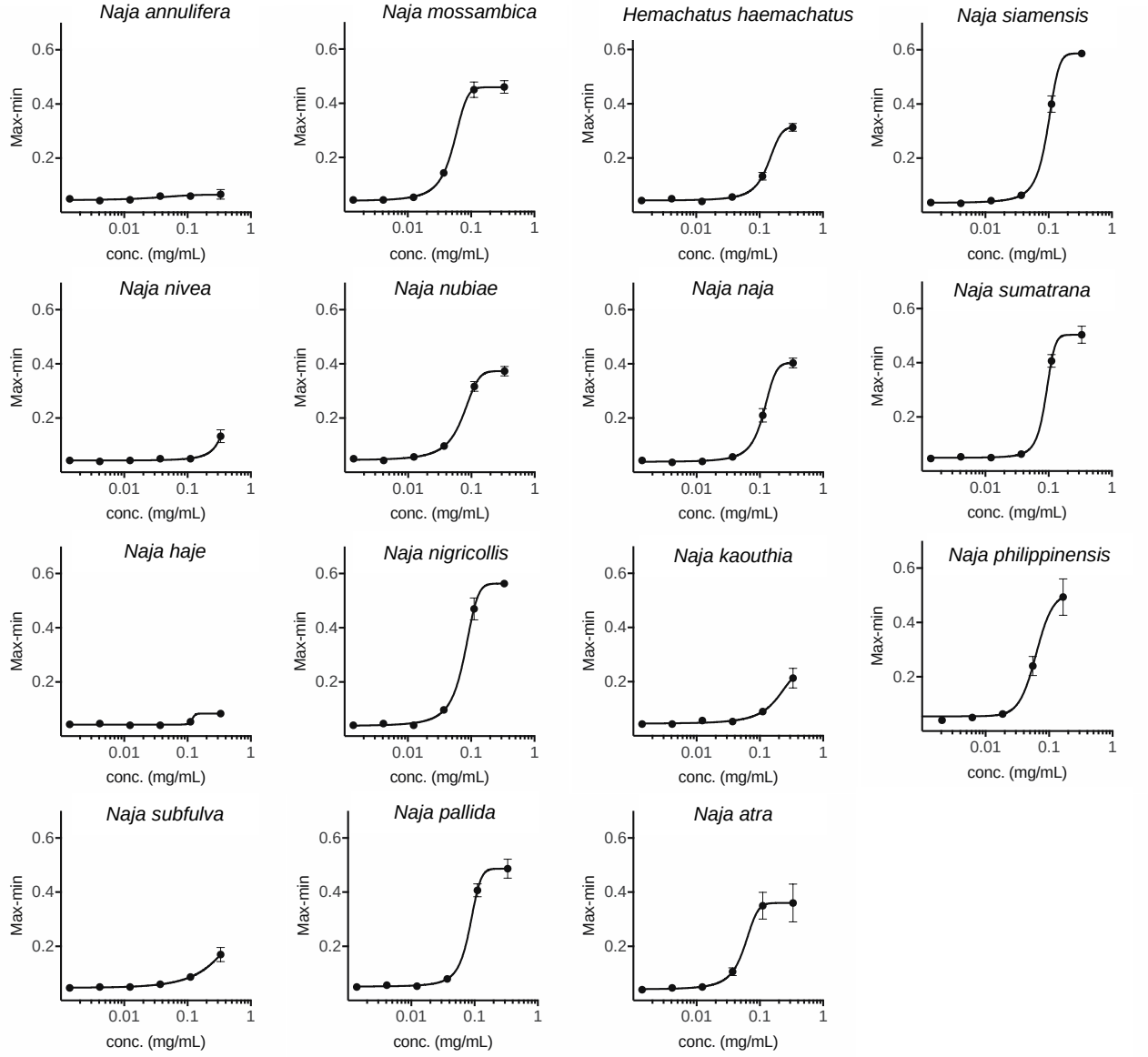


Fig. S8. Potency of *Naja* and *Hemachatus* venoms on activation of sensory neuron-derived F11 cells. Values are max-min (n = 3), and error bars represent standard error of the mean (SEM). Curves were fitted using a four-parameter Hill equation (Variable slope) in Graphpad Prism (version 8.02).

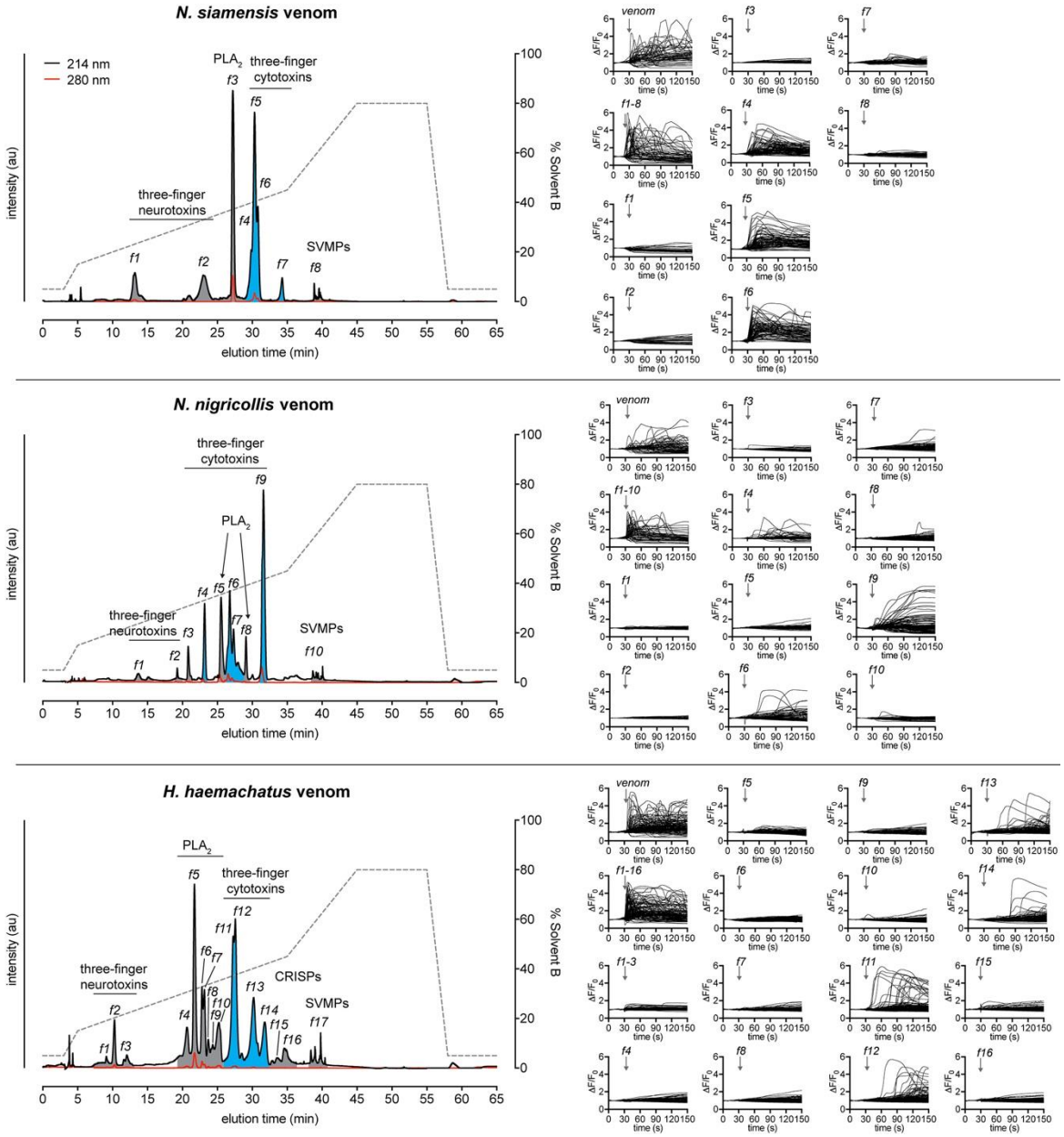


Fig. S9. Separation of *Naja siamensis, Naja nigricollis* and *Hemachatus haemachatus* venoms and activity of fractions on sensory neurons. *N. siamensis*, *N. nigricollis* and *H. haemachatus* venoms (500 µg) were each separated by RP-HPLC using a Phenomenex Gemini NX-C18 column (250 x 4.6 mm, 3 µm particle size, 110 Å pore size; gradient of 15–45% solvent B (90% ACN, 0.05% TFA) over 30 min; flow rate of 1 mL min^{-1}). Fractions were collected according to absorbance at 214 nm. Bottom-up proteomics was used to confirm the identity of the major component/s of each fraction. Individual fractions were assessed for activity on sensory neurons (mouse DRG). Active fractions are highlighted in blue, inactive in grey. Each set of traces represents all cells of one experiment. Only the pooled fractions and those individual fractions containing three-finger cytotoxins were able to activate sensory neurons.

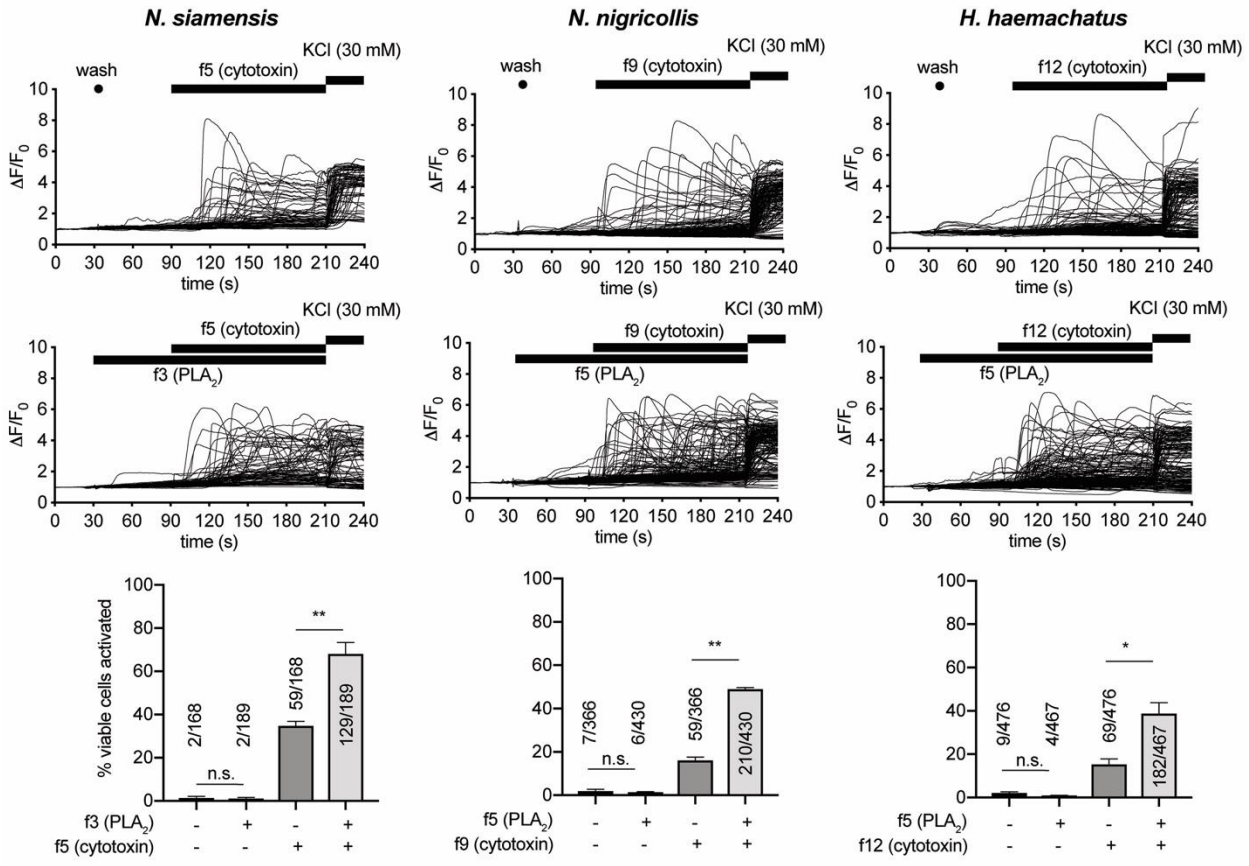


Fig. S10. Activation of sensory neurons by spitting cobra three-finger cytotoxins is potentiated by PLA_2s. For all venoms tested, there was no significant difference in cell (mouse DRG) activation between buffer/wash and a PLA_2 fraction alone, while cytotoxin fraction activation was potentiated in the presence of a corresponding PLA_2 fraction. Each set of traces represents all cells of one representative experiment. Data are derived from 2-3 independent experiments. Statistical comparisons were made using unpaired parametric *t*-tests in Graphpad Prism (version 8.02). *, $P < 0.05$; **, $P < 0.01$.

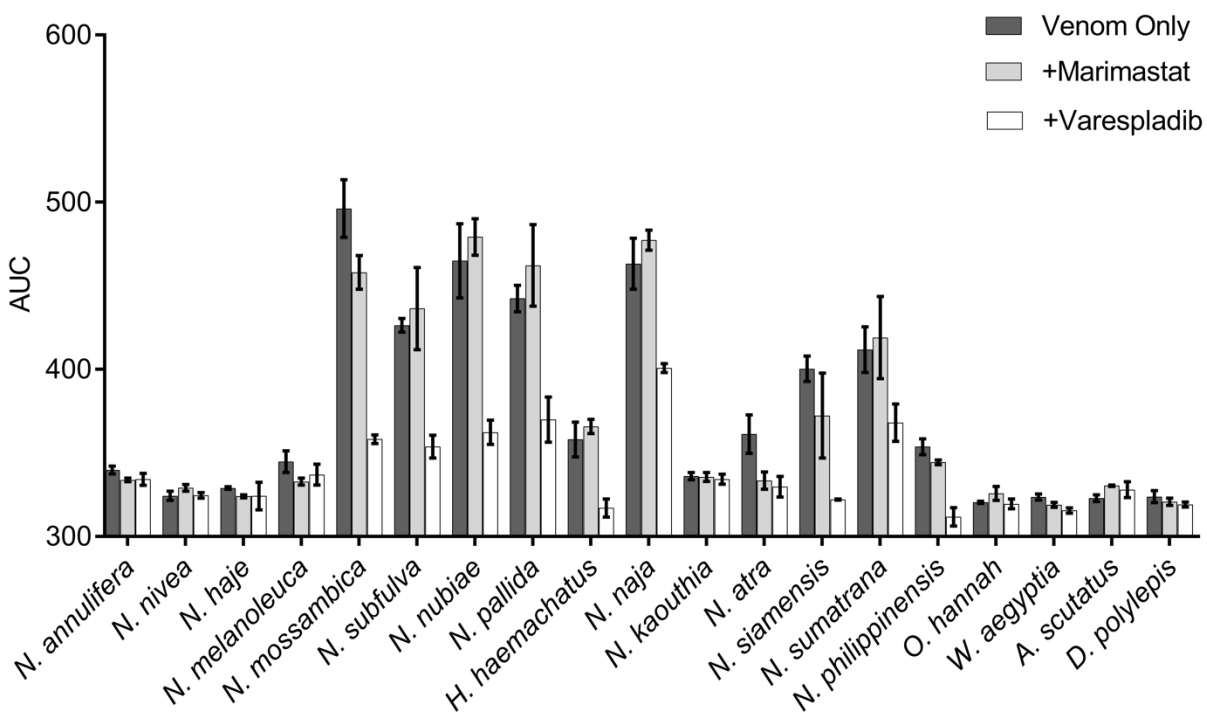


Fig. S11. The PLA_2 inhibitor varespladib reduces the activation of sensory neurons caused by spitting cobra venoms. 80 µg/mL of venoms (or 160 µg/mL in the case of *H. haemachatus* and *N. philippinensis* venom) were added to F11 cells. Values are area-under-curve (n = 3), and error bars represent standard error of the mean (SEM), with either venom only (negative control), the PLA_2 inhibitor varespladib (13 µM) or the metalloprotease inhibitor marimastat (15 µM).

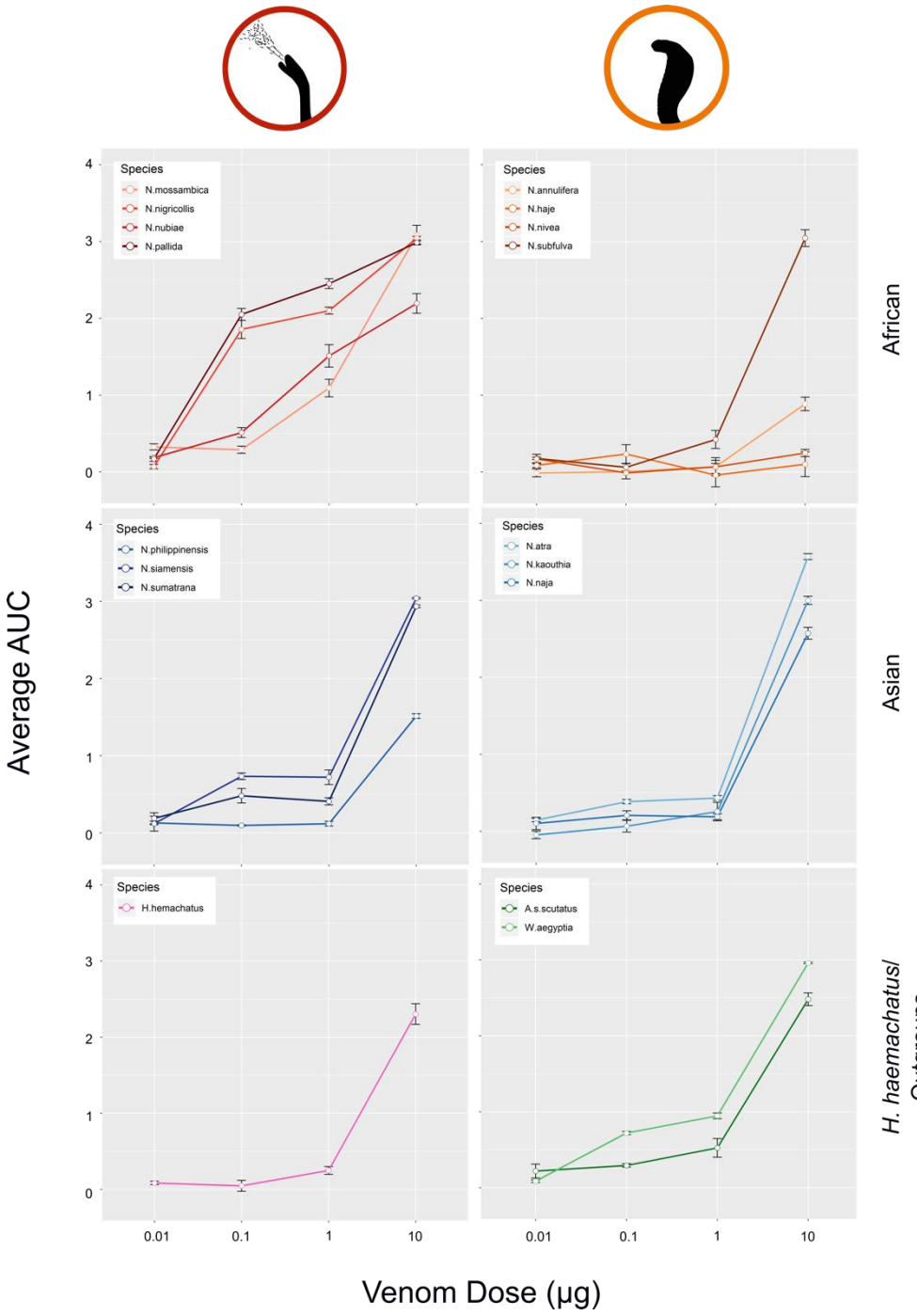


Fig. S12. Concentration curves displaying the enzymatic PLA_2 activity of the various cobra venoms. The PLA_2 activity of each venom was characterized using kinetic measurement of an *in vitro* colorimetric assay (*80*). Venom doses of 0.01 µg, 0.1 µg, 1 µg and 10 µg were quantified in quadruplicate alongside Tris Buffer negative controls. The mean area under the curve (AUC) of the resulting control profile at 10 min was calculated from the four replicates and used to subtract from each individual venom measurement. The mean of the resulting venom data was then plotted, with error bars representing the standard error of the mean of the AUC values at each dose tested.

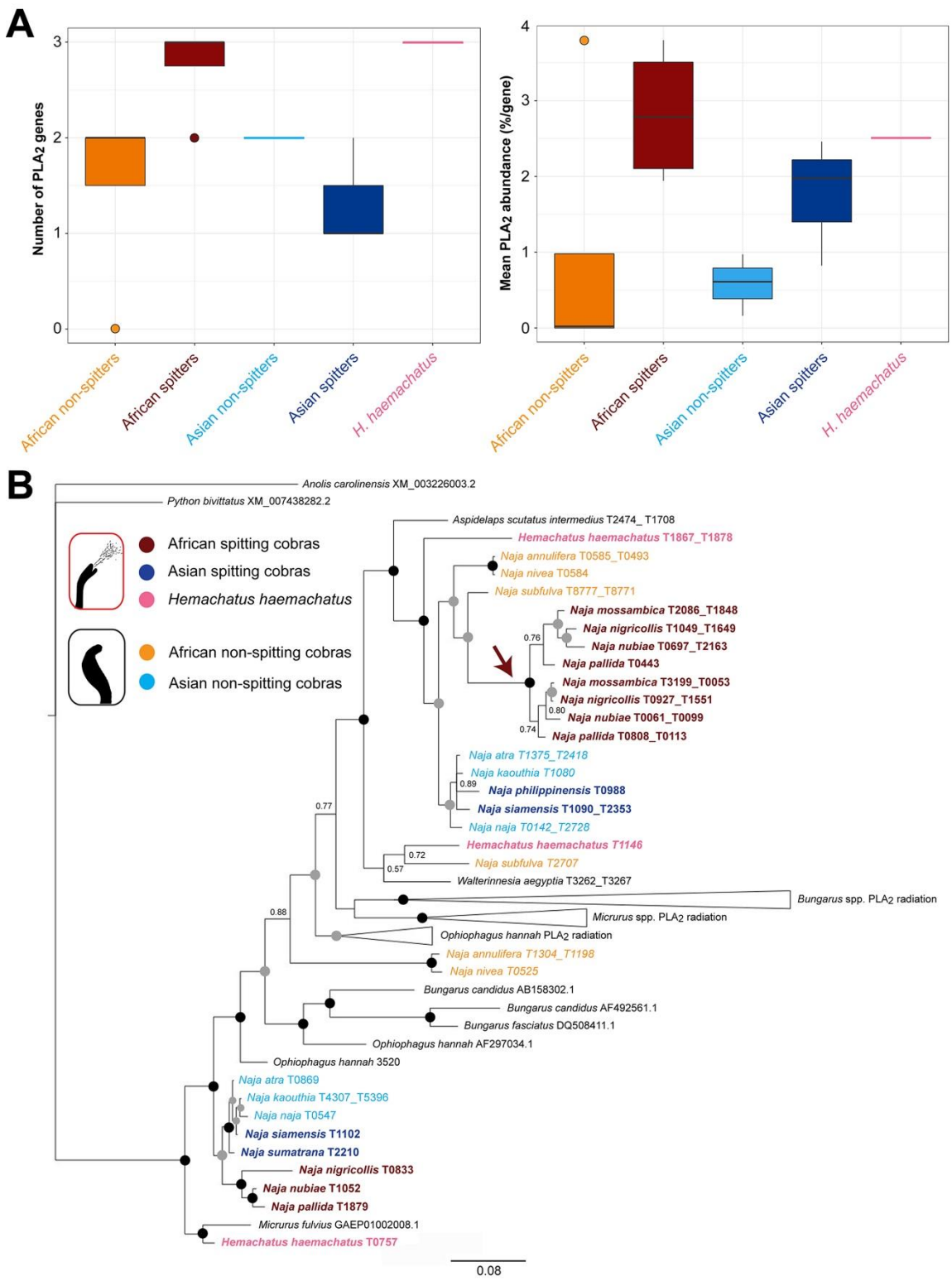


Fig. S13. Molecular analysis of cobra PLA$_2$ toxins reveals different molecular mechanisms underpin their convergent upregulation associated with venom spitting. (A) Box plots of the number of PLA$_2$ encoding genes recovered from the venom gland (left) and their mean proteomic abundance, displayed as PLA$_2$s as percentage of total venom toxins per species

divided by PLA_2 gene number (right). These analyses reveal minor variations in gene number, with African spitting cobras and *H. haemachatus* generally showing an increase in PLA_2 number compared with other species, though this increase is not observed in Asian spitting cobras (left). Irrespective of gene number, each of the three spitting cobra lineages exhibit increased mean (per gene) PLA_2 abundances compared with non-spitting cobras (right). Boxes represent the interquartile range. Midline represents the median. Outliers are indicated by dots. (B) Bayesian inference-derived DNA phylogeny of elapid PLA_2 sequences demonstrates two major clades of cobra PLA_2s, and evidence of a lineage-specific gene duplication event that occurred prior to the radiation of African spitting cobras (highlighted by arrow). Black-filled circles at each node represent Bayesian posterior probabilities of 1 and grey-filled circles represent values between 0.90 and 0.99. Tips labels are coloured based on the different spitting and non-spitting lineages indicated in the key. The sequence alignment used for tree construction is displayed in Data S5.

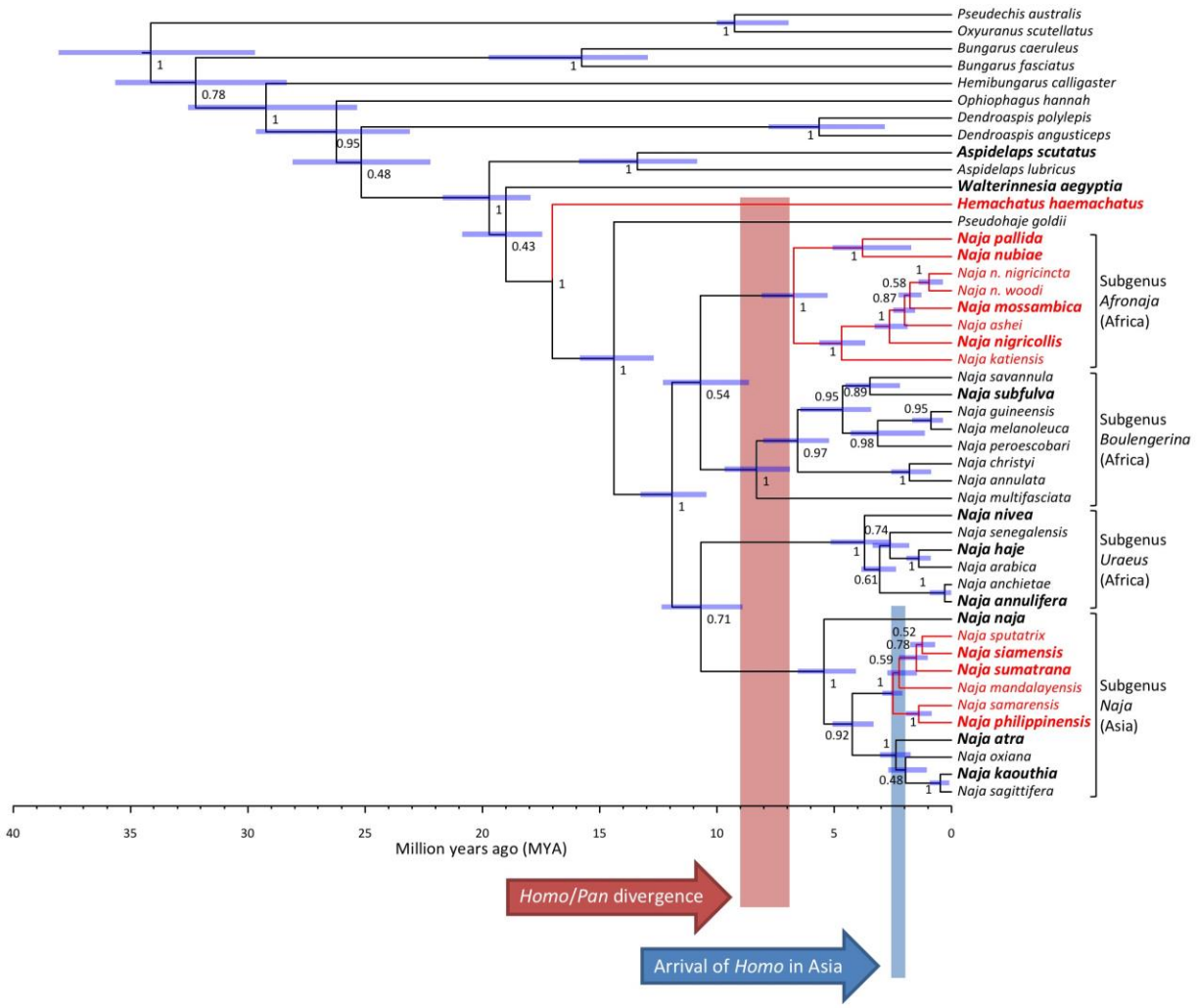


Fig. S14. Time-calibrated coalescent species tree of cobras and related elapids. Node labels denote posterior probabilities and blue bars at each node represent age range estimations. Tip labels in bold and enlarged font indicate species included in venom analyses, red tip labels and branches indicate spitting species.

Supplementary Tables

Table S1. The relative representation of phospholipases A_2 (PLA_2) and three-finger toxins (3FTX) in elapid venom proteomes described in the literature. The data displayed represents the percentage of each toxin family in the whole venom, and note that the method of quantification applied varies among studies.

Species	3FTX		PLA_2	Citation
	CTX	Other		
Aipysurus laevis	0	25.3	71.2	(*83*)
Aspidelaps lubricus cowlesi	4.9	71.2	4.9	(*40*)
Aspidelaps lubricus lubricus	2.1	75.7	5.7	(*40*)
Aspidelaps scutatus intermedius	33.7	49.0	6.1	(*40*)
Bungarus caeruleus	0	19.0	64.5	(*84*)
Bungarus candidus	0	30.1	25.2	(*85*)
Bungarus fasciatus	0	1.3	66.8	(*86*)
Bungarus fasciatus	0	17.4	44.2	(*85*)
Bungarus multicinctus	0	32.6	66.4	(*87*)
Bungarus multicinctus	0	27.5	15.3	(*86*)
Bungarus sindanus	0	50.3	32.6	(*88*)
Calliophis intestinalis	15.9*	5.1	43.8	(*89*)
Dendroaspis angusticeps	0	69.2	0	(*90*)
Dendroaspis angusticeps	0	64.2	0	(*11*)
Dendroaspis jamesoni jamesoni	0	65.5	2.2	(*11*)
Dendroaspis jamesoni kaimosae	0	74.8	1.8	(*11*)
Dendroaspis polylepis	0	31.0	0	(*91*)
Dendroaspis polylepis	0	40.0	0	(*11*)
Dendroaspis viridis	0	76.7	0.2	(*11*)
Hydrophis curtus	0	26.3	62.0	(*92*)
Hydrophis cyanocinctus	0	81.1	18.9	(*93*)
Hydrophis platurus	0	49.9	32.9	(*94*)
Hydrophis schistosus	0	70.5	27.5	(*95*)
Laticauda colubrina	0.3	66.1	33.3	(*96*)
Micropechis ikaheka	0	9.2	80.0	(*97*)
Micrurus alleni	0	77.3	10.9	(*98*)
Micrurus altirostris	0	79.5	13.7	(*65*)
Micrurus corallinus	34.8		38.5	(*99*)
Micrurus dumerilii	0	28.1	52.0	(*100*)
Micrurus frontalis	0	42.4	49.2	(*101*)
Micrurus lemniscatus carvalhoi	2.3		48.6	(*99*)
Micrurus lemniscatus lemniscatus	34.3		19.4	(*99*)
Micrurus mipartitus	0	61.1	29.0	(*102*)
Micrurus mipartitus	0	83.0	8.2	(*102*)
Micrurus mosquitensis	0	22.5	55.6	(*98*)
Micrurus nigrocinctus	0	38.0	48.0	(*103*)

Micrurus paraensis	12.3		65.9	(*99*)
Micrurus pyrrhocryptus	0	27.0	17.0	(*104*)
Micrurus spixii spixii	9.5		64.0	(*99*)
Micrurus spixii spixii	0	56.5	37.4	(*101*)
Micrurus surinamensis	90.1		6.6	(*99*)
Micrurus surinamensis	36.8		33.0	(*99*)
Micrurus surinamensis	0	95.4	4.2	(*101*)
Micrurus tschudii	0	95.2	4.1	(*105*)
Naja atra	52.9	23.5	16.8	(*106*)
Naja atra	59.4	20.5	14.0	(*106*)
Naja atra	65.3	19.0	12.2	(*87*)
Naja haje	54.0	6.0	4.0	(*107*)
Naja kaouthia	45.7	18.0	23.5	(*108*)
Naja kaouthia	27.6	50.7	12.2	(*108*)
Naja kaouthia	44.9	31.5	17.4	(*108*)
Naja kaouthia	27.9	28.6	26.9	(*109*)
Naja katiensis	62.7	4.4	29.0	(*45*)
Naja melanoleuca*	25.2	31.9	12.9	(*110*)
Naja mossambica	67.7	1.6	27.1	(*45*)
Naja naja	63.8		11.4	(*111*)
Naja naja	69.3	4.8	21.4	(*112*)
Naja naja	71.6	8.9	14.0	(*112*)
Naja nigricollis	72.8	0.5	21.9	(*45*)
Naja nubiae	58.3	12.6	26.4	(*45*)
Naja pallida	64.9	2.8	30.1	(*45*)
Naja philippinensis	21.3	45.3	22.9	(*113*)
Naja sputatrix	48.1	16.1	31.2	(*114*)
Notechis scutatus	0	5.6	74.5	(*115*)
Ophiophagus hannah	9.0	55.2	2.8	(*82*)
Ophiophagus hannah	0.5	42.5	4.0	(*116*)
Oxyuranus scutellatus	0	4.2	68.3	(*117*)
Oxyuranus scutellatus	0	1.5	79.4	(*117*)
Pseudechis papuanus	0	3.1	90.2	(*118*)
Toxicocalamus longissimus	0	92.1	6.5	(*93*)

* BlastP analysis of the proteomic fragments reported in this study reveals that six of the seven proteins annotated as cytotoxins do not match with cytotoxins as their closest match in the NCBI database. The remaining sequence showed only 44% identity with a cytotoxin.

** This study was published before the recent reclassification of *N. melanoleuca* into five distinct species (*119*). Based on the locality provided in this publication, the venom used in this study was likely sourced from either *N. melanoleuca* or *N. subfulva*.

Table S2. A summary of the Phylogenetic Generalized Least Squares (PGLS) tests performed under the scenario of *Naja atra* and *Naja kaouthia* being coded as non-spitting species. All tests were performed using the monophyletic grouping of *Naja* and *Hemachatus*, with the outgroup species of *Walterinnesia aegyptia* and *Aspidelaps scutatus* used to provide additional phylogenetic context. Significant values (P <0.05) are represented by bold red font. Degrees of freedom = 15.

Test Variable	t	P-value
PLA_2 proteomic abundance	**4.24**	**0.0007**
CTX proteomic abundance	-0.83	0.42
'Other 3FTX' proteomic abundance	-1.21	0.25
Venom lethality by murine lethal dose 50 (LD_{50})	0.86	0.40
Potency of neuronal cell activation (EC_{50})	**-4.48**	**0.0004**
PLA_2 enzymatic activity	**2.24**	**0.04**
Venom cytotoxicity via HET-CAM assay	1.08	0.30
Neuronal cell activation (EC_{50}) + PLA_2 activity	**-4.45**	**0.0004**

Table S3. Time-dependent irritation scores described for Luepeke's *in vivo* hen's egg test-chorioallantoic membrane (HET-CAM) assay (*77*).

Effect	Time (min)		
	0.5	2.0	5.0
Hyperaemia	5	3	1
Haemorrhage	7	5	3
Coagulation	9	7	5

Table S4. The previously described relationship between the cumulative irritation score of Luepeke's *in vivo* then's egg test-chorioallantoic membrane (HET-CAM) assay (see Table S3) and 'irritation potential' (*77*).

Cumulative score	**Irritation assessment (based on Luepke** (*77*)**)**	**Irritation potential**
0.0-0.9	Practically none	No irritation
1.0-4.9	Slight	Slight irritation
5.0-8.9	Moderate	Moderate irritation
9.0-21.0	Strong	Severe irritation

Table S5. The murine lethality of the elapid snake venoms used in this study represented by intravenous median lethal dose (LD_{50}) values and corresponding 95% confidence intervals (CI).

Species	Group	Lineage	LD_{50} (µg/mouse)	95% CI
H. haemachatus	Spitter	*H. haemachatus*	57.96	47.28-80.00
N. mossambica	Spitter	African spitter	24.48	19.68-29.51
N. nubiae	Spitter	African spitter	23.65	19.17-27.37
N. pallida	Spitter	African spitter	9.29	3.76-13.23
N. nigricollis	Spitter	African spitter	27.49	22.55-38.21
N. annulifera	Non-spitter	African non-spitter	33.75	29.72-37.18
N. subfulva	Non-spitter	African non-spitter	5.84	3.27-7.47
N. nivea	Non-spitter	African non-spitter	15.05	10.20-18.70
N. haje	Non-spitter	African non-spitter	8.15	6.55-9.89
N. philippinensis	Spitter	Asian spitter	2.36	0.64-4.20
N. siamensis	Spitter	Asian spitter	18.43	14.02-22.17
N. sumatrana	Spitter	Asian spitter	17.43	13.29-20.97
N. atra	Non-spitter	Asian non-spitter	17.86	14.16-27.89
N. kaouthia	Non-spitter	Asian non-spitter	6.69	3.89-8.54
N. naja	Non-spitter	Asian non-spitter	11.1	7.90-13.10
A. scutatus	Non-spitter	Outgroup	5.75 *	5.29-6.26
W. aegyptia	Non-spitter	Outgroup	15.08	11.78-20.68

* Previously determined by Whiteley et al. (*40*).

Table S6. A summary of the Phylogenetic Generalized Least Squares (PGLS) tests performed under the scenario of *Naja atra* and *Naja kaouthia* being coded as spitting species. All tests were performed using the monophyletic grouping of *Naja* and *Hemachatus*, with the outgroup species of *Walterinnesia aegyptia* and *Aspidelaps scutatus* used to provide additional phylogenetic context. Significant values ($P < 0.05$) are represented by bold red font. Degrees of freedom = 15.

Test Variable	t	P-value
PLA_2 proteomic abundance	**2.86**	**0.01**
CTX proteomic abundance	-0.57	0.58
'Other 3FTX' proteomic abundance	-0.76	0.46
Venom lethality by murine lethal dose 50 (LD_{50})	1.25	0.23
Potency of neuronal cell activation (EC_{50})	**-4.59**	**0.0004**
PLA_2 enzymatic activity	1.80	0.09
Venom cytotoxicity via HET-CAM assay	**3.29**	**0.004**
Neuronal cell activation (EC_{50}) + PLA_2 activity	**-4.57**	**0.0004**

Table S7. NCBI GenBank accession numbers of the nucleotide sequences used in the construction of the species tree.

	ND4	cytb	PRLR	NT3	UBN1	c-mos	RAG1
Aspidelaps lubricus	MT346783-9	MT346603-9	MT347202-12	MT347034-41	MT347442-53	MT346956-7	MT347362-3
Aspidelaps scutatus	AY058969, MT346790-8	AY188007	MT347214-21	KX694994, MT347042-8	MT347454-8	AY058923, AY187968, KX694796, MT346958-60	MT347364-6
Bungarus caeruleus	AJ830220, MT346799-800	AJ749305	MT347222-	MT347049-50	MT347459	MT346961-2	MT347367
Bungarus fasciatus	EU547037, MT346801	EU547086, MT346610	MT347224-	FJ434090, KX694998, MT347051-2	MT347460	AF544732, AY058924, EU366447, MT346963-4	EU366438, JF357954, MT347368
Dendroaspis angusticeps	MT346803-16	MT346611-24	MT347226-38	MT347053-63	Mt347461-71	AF544735, MT346965-6	MT347369
Dendroaspis polylepis	MT346817-25	MT346625-33	MT347239-42	MT347064-7	MT347472-5	AY058928, FJ387197	MT347370-1
Hemachatus haemachatus	MT346826-38	MT346634-46	MT347244-56	MT347068-78	MT347476-88	MT346968-9	MT347373-4
Hemibungarus calligaster	KX130764	EF137411	MT347243	MT347192	MT347489	EF137419, MT346967	MT347372
Naja (Afronaja) ashei	GQ359575, MT346839-42	GQ359493, MT346647-50	MT347257-8	MT347079-80	MT347497-8	MT346970-1	MT347375-6
Naja (Afronaja) katiensis	GQ359576, MT346843-5	GQ359494, MT346651-3	MT347259-60	MT347081-2	MT347499-500	MT346972-3	MT347377
Naja (Afronaja) mossambica	GQ359577, MT346846-53	AF399747, GQ359495, MT346654-60	MT347261-3	MT347083-5	MT347501-3	MT346974-5	MT347378-9
Naja (Afronaja) nigricincta nigricincta	MT346854-5	MT346661-2	MT347264-5	MT347086-7	MT347504-5	MT346976-7	MT347380
Naja (Afronaja) nigricincta woodi	MT346889-92	MT346693-6	MT347290-1	MT347105-6	MT347525-6	MT346984-5	MT347385
Naja (Afronaja) nigricollis	AY713377, MT346856-79	AF399746, GQ359505, MT346663-85	MT347266-84	MT347088-99	MT347506-19	MT346978-9	MT347381
Naja (Afronaja) nubiae	GQ359579, MT346880-81	AF399751, GQ359497, MT346686	MT347285-7	MT347100-2	MT347520-2	MT346980-1	MT347382
Naja (Afronaja) pallida	GQ359578, MT346882-8	AF399748, GQ359496, MT346687-92	MT347288-9	MT347103-4	MT347523-4	MT346982-3	MT347383-4
Naja (Boulengerina) annulata	MT346893-6	MT346697-700	MT347292-8	MT347107-9	MT347490-2	AY058925, AY187971, MT346986	MT347386-7
Naja (Boulengerina) christyi	MT346897	MT346701	MT347299	MT347110	MT347493		
Naja (Boulengerina) guineensis	MH337376-88	MH337570-82	MH337474-5, MH337484-95	MT347111-5	MH337511-24	MT346987-8	MT347388
Naja (Boulengerina) melanoleuca	MH337389-402	MH337583-96	MH337472-83, MH337496-8, MH337500, MT347301-	MT347116-21	MH347505-10, MH337525-30, MH337567-8, MT347494-6	AY611904, MT346989-90	JF357955, MT347389-90
Naja (Boulengerina) multifasciata	AY058985, MT346898	AF217837, MT346702	MT347303	MT347122	MT347527	AY058941	MT347391

Naja (Boulengerina) peroescobari	MH337440	MH337634	MH337499	MT347123-4	MH337563	MT346991-2	MT347392-3
Naja (Boulengerina) savannula	MH337403-8	MH337597-602	MH337501-4	MT347125-8	MH337532-5	MT346993-4	MT397394-5
Naja (Boulengerina) subfulva	MH337409-21, MH337426-39	MH337603-33	MH337441-71	MT347129-45	MH337531, MH337537-66	MT346995-6	MT347396-7
Naja (Naja) atra	MT346899-900	MT346703-4	MT347304-5	MT347146-7	MT347528-9	KX694797, MT346997-8	MT347398-9
Naja (Naja) kaouthia	EU624209, MT346901-4	GQ359507, MT346705-8	MT347306-8	MT347148-50	MT347530-2	AY058938, MT346999	MT347400
Naja (Naja) mandalayensis	MT346905-6	MT346709-10	MT347309-10	MT347151-2	MT347533-4	MT347000	MT347401-2
Naja (Naja) naja	AY713378, MH337375, MT346907-8	GQ359506, MH337569, MT346711-2	MT347311-2	MT347155-6	MT347535-6	EU366445, MT347001-2	EU366432, MT347403
Naja (Naja) oxiana	MT346909-10	MT346713-4	MT347313-4	MT347157	MT347537	MT347003	MT347404
Naja (Naja) philippinensis	MT346911-5	MT346911-5	MT347315-6	MT347158-9	MT347538-9	MT347004-5	MT347405
Naja (Naja) sagittifera	MT346916-7	MT346916-7	MT347317	MT347160	MT347540	MT347006-7	MT347406
Naja (Naja) samarensis	MT346918-20	MT346918-20	MT347318-9	MT347161-2	MT347541-2	MT347008	MT347407-8
Naja (Naja) siamensis	MT346923-6	MT346923-6	MT347320-1	MT347163-5	MT347543-5	MT347009-11	MT347409-10
Naja (Naja) sputatrix	MT346921-2	MT346921-2	MT347322	MT347166	MT347546		
Naja (Naja) sumatrana	MT346927-34	MT346927-34	MT347323-6	MT347153-4, MT347167-8	MT347547-51	MT347012-6	MT347411-5
Naja (Uraeus) anchietae	GQ387085-8	GQ387085-8	MT347327-8	MT347169-70	MT347552-3	MT347017	MT347416
Naja (Uraeus) annulifera	GQ359586, GQ387089	GQ359586, GQ387089	MT347329-31	MT347171-2	MT347554-6	MT347018	MT347417-8
Naja (Uraeus) arabica	GQ359382, GQ387074-9	GQ359500, MT346744-8	MT347332-3	MT347173-4	MT347557-8	MT347019-20	MT347419-20
Naja (Uraeus) haje	GQ359580-1, GQ359583, GQ387062, GQ387064, GQ387066, GQ387068-73	GQ359498, GQ359501, MT346749-58	MT347334-9	MT347175-8	MT347559-64	AY611903, MT347021-2	MT347421-2
Naja (Uraeus) nivea	AY058983, MT346935-6	AF217827, MT346759-60	MT347340-1	MT347179-80	MT347564-6	AY058939, MT347023	MT347423-4
Naja (Uraeus) senegalensis	GQ359584-5, GQ387080, GQ387083-4	GQ359502-3, MT346761-3	MT347342-3	MT347181-2	MT347567-8	MT347024-5	MT347425
Ophiophagus hannah	AY058984, EU921899, MT346937-46	AF217842, EU921899, MT346764-73	MT347344-53	MT347183-91	MT347569-74	AY058940, KX694798, MT347026	MT347426-34
Oxyuranus scutellatus	EU547006, MT346947	EU547051, MT346774	MT347354-5	MT347193-4	MT347575-6	EU546916, MT347027-8	EU546877, MT347435-6
Pseudechis australis	AJ83071, MT346948-50	EU547046, MT346775-7	KX981746-7, MT347356-7	MT347195-7	MT347577-80	EU546912, MT347029	EU546873, MT347437
Pseudohaje goldii	MT346951-2	MT346778-9	MT347358	MT347198	MT347581	MT347030	MT347438
Walterinnesia aegyptia	AY058988, MT346953-5	AF217838, MT346780-2	MT347359-61	MT347199-201	MT347582-4	AY058943, MT347031-3	MT347439-41

Table S8. The snake species, localities of origin, and number of specimens that contributed to the venom pools used in venom proteomic analyses and murine lethality tests.

Species	Locality	Number of individual snakes contributing to the venom pool
Hemachatus haemachatus	Captive bred	2
Naja annulifera	Captive bred	2
Naja atra	Captive bred	2
Naja haje	Uganda	6
Naja kaouthia	Captive bred	3
Naja mossambica	Tanzania	4
Naja naja	Captive bred	2
Naja nigricollis	Nigeria	3
Naja nivea	South Africa	3
Naja nubiae	Captive bred	3
Naja pallida	Tanzania	2
Naja philippinensis	Captive bred	2
Naja siamensis	Captive bred	2
Naja subfulva	Cameroon	2
Naja sumatrana	Captive bred	1
Walterinnesia aegyptia	Captive bred	2

Table S9. PCR primers and typical thermocycling conditions for the genes sequenced for the phylogenetic analysis. The denaturing step involved 30-45 s at 94°C and the extension step 1 min at 72°C, with a final extension step of 5 min at 72°C followed by cooling to 4°C for 15 min.

Gene	Sense	Name	Sequence	Citation	Annealing temp.	Cycles
CytB	Forward	Gludg	5'– TGACTTGAARAACCAYCGTTG – 3'	(*120*)	47°C	39
	Reverse	ATRCB3	5'– TGAGAAGTTTTCYGGGTCRTT – 3'	(*121*)		
		H16064	5'– CTTTGGTTTACAAGAACAATGCTTTA – 3'	(*122*)		
ND4	Forward	NADH4	5'– CACCTATGACTACCAAAAGCTCATGTAG AAGC – 3'	(*123*)	57°C	39
	Reverse	H12763V	5'– TTCTATCACTTGGATTTGCACCA – 3'	(*123*)		
C-MOS	Forward	AV CMOSF	5'– AAGCACATCAAGGATTCGTCG – 3'	(*124*)	60.5°C	44
	Reverse	AV CMOSR	5'– TCTGCCTTGGGTGTGATTTTCT – 3'			
	Forward	G303	5'– ATTATGCCATCMCCTMTTCC – 3'	(*125*)	57°C	35
	Reverse	G708	5'– GCTACATCAGCTCTCCARCA – 3'			
NT3	Forward	NTF3 F1	5'– ATGTCCAATCTTGTTTTATGTGATATTT – 3'	(*126*)	42°C	39
	Reverse	NTF3 R1	5'– ACRAGTTTRTTGTTYTCTGAAGTC – 3'			
PRLR	Forward	PRLR F1	5'– GACARYGARGACCAGCAACTRATGCC – 3'	(*126*)	48°C	39
	Reverse	PRLR F3	5'– GACYTTGTGRACTTCYACRTAATCCAT – 3'			
RAG1	Forward	AV RAG1F	5'– AAATGTGACAGGGTCTCT – 3'	(*124*)	59°C	44
	Reverse	AV RAG1R	5'– GGGCATCTCAAAACCAAATTGT – 3'			
UBN1	Forward	BaUBN F	5'– CCTCTGGTTACTCAGCAGCA – 3'	(*127*)	40°C	39
	Reverse	BaUBN R	5'– ATTGGCCACTCCTTGTGTTC – 3'			

Table S10. The snake species, localities of origin, and number of specimens that contributed to the venom pools used in the hen's egg test-chorioallantoic membrane (HET-CAM) assay.

Species	Locality	Number of individual snakes contributing to the venom pool	Same venom sample as that used for proteomic analysis?
Naja kaouthia	Captive bred	1	No
Naja mossambica	Captive bred	1	No
Naja nivea	Captive bred	1	No
Naja atra	Captive bred	1	No
Naja sumatrana	Captive bred	1	No
Naja siamensis	Captive bred	1	No
Naja pallida	Captive bred	1	No
Naja annulifera	Captive bred	1	No
Naja haje	Captive bred	1	No
Naja subfulva	Captive bred	1	No
Naja naja	Captive bred	1	No
Naja nubiae	Captive bred	1	No
Walterinnesia aegyptia	Captive bred	2	Yes
Hemachatus haemachatus	Captive bred	2	Yes
Naja philippinensis	Captive bred	2	Yes
Naja nigricollis	Nigeria	3	Yes
Aspidelaps lubricus cowlesi	Captive bred	1	N/A (proteomics previously performed)

Supplementary Data Files

Data S1.

Coalescent species trees resulting from the phylogenetic analyses and jackknife analyses. Tree 1: uncalibrated coalescent species tree. Tree 2: time-calibrated coalescent species tree. Tree 3: jackknife analysis: uncalibrated coalescent species tree calculated under exclusion of C-MOS. Tree 4: jackknife analysis: uncalibrated coalescent species tree calculated under exclusion of NT3. Tree 5: jackknife analysis: uncalibrated coalescent species tree calculated under exclusion of PRLR. Tree 6: jackknife analysis: uncalibrated coalescent species tree calculated under exclusion of RAG1. Tree 7: jackknife analysis: uncalibrated coalescent species tree calculated under exclusion of UBN1.

Data S2.

Top-down MS analyses of the venom proteomes described in this study: including African and Asian *Naja* species, the desert black snake (*Walterinnesia aegyptia*), and the rinkhals (*Hemachatus haemachatus*), as summarized in Fig. 1A and Fig. S2.

Data S3.

Nucleotide sequence alignment for the three-finger toxin (3FTX) family, extracted from venom gland transcriptomes from *Naja* spp., the desert black snake (*Walterinnesia aegyptia*) and the rinkhals (*Hemachatus haemachatus*) described in this study, along with data from related elapid snake species (*Aspidelaps scutatus intermedius* (*40*), *Bungarus flaviceps* (*53*), *Bungarus multicinctus* (52), *Dendroaspis* spp. (*11*), *Micrurus fulvius* (*55*) and *Ophiophagus hannah* (53, 54)) sourced from previous, aligned against a sequence from the outgroup species (*Python regius*; NCBI accession number GBIC00000000, from (*58*)).

Data S4.

Nucleotide sequence alignment for the cytotoxin (CTX) subset of the three-finger toxin (3FTX) family, extracted from venom gland transcriptomes from *Naja* spp., the rinkhals (*Hemachatus haemachatus*), the shield-nosed cobra (*Aspidelaps scutatus intermedius*, from (*40*)) and the king cobra (*Ophiophagus hannah*, from (*56*)), aligned against a sequence from the outgroup elapid, the red-headed krait (*Bungarus flaviceps*, NCBI accession number GU190795, from (*53*)).

Data S5.

Nucleotide sequence alignment for the phospholipase A2 (PLA_2) toxin family, extracted from venom gland transcriptomes from *Naja* spp., the desert black snake (*Walterinnesia aegyptia*) and the rinkhals (*Hemachatus haemachatus*) described in this study, along with data from related elapid snake species (*Aspidelaps scutatus intermedius* (*40*), *Bungarus caeruleus* (*59*, *60*), *Bungarus candidus* (*61*), *Bungarus fasciatus* (*62*), *Bungarus flaviceps* (*53*, *61*), *Bungarus* multicinctus (*63*, *64*), *Micrurus altirostris* (*65*), *Micrurus corallinus* (*66*), *Micrurus fulvius* (*55*), *Micrurus laticollaris* (*67*), *Micrurus tener* (*68*) and *Ophiophagus hannah* (*57*, *69*, *70*)) aligned against non-toxin PLA_2 sequences from *Python bivittatus* (*71*) and *Anolis carolinensis* (*72*).

Supplementary Acknowledgments

The authors wish to thank the following individuals who very kindly helped with sampling for the phylogenetic dataset:

Bioken (Watamu, Kenya: Sanda Ashe, Anthony Childs, Royjan Taylor), Susen Baier, Patrick Barrière, Joe Beraducci, Phil Berry, Bill Branch, M. Brand, Donald G. Broadley, Ashok Captain, Tom Charlton, Laurent Chirio, F.W.P. Cotteril, Merel J. Cox, Liverpool School of Tropical Medicine (Edouard Crittenden, Paul D. Rowley, R. David G. Theakston), Jenny Daltry, Anslem de Silva, Latoxan (Valence, France: Yvon Doljansky, Franck Principaud), Craig Doria, California Academy of Sciences (Bob Drewes, Joe Slowinski), Damien Egan, Thomas Eimermacher, Johannes Els, Bryan G. Fry, Eli Greenbaum, Michael Griffin, Michel Guillot, Chris Hay, Hans-Werner Herrmann, Ivan Ineich, Andrew Jackson, Kate Jackson, Moshe Kahn, Trinh Xuan Kiem, Warren Klein, Lukas Kratochvil, Adam Leaché, Jonathan Leakey & Dena Crain, Thea Litschka-Koen, Mario Lutz, Anita Malhotra, Youssouph Mané, Bryan Maritz, Mark Marshall, Richard Mastenbroek, Peter Mirtschin, Zoltan Nagy, Arno Naude, Deon Naude, Mark O'Shea, Mike Perry, Heidi Pfeifer, Tony Phelps, Jens B. Rasmussen, Paul D. Rowley, Radko Samek, Kate Sanders, Vishal Santra, Donald Schultz, Sterrin Smallbrugge, Martin Smit, Jeremie Tai-A-Pin, Zoltan Takacs, Bernard Thorens, Roger S. Thorpe, Jean-François Trape, Tanith Tyrr, Romilly van den Bergh, Craig van Rensburg, Mark Vandewalle, David A. Warrell, Zoological Society of London (Heather Hall, Terry March, Esther Wenman), Romulus Whitaker, Chris Wild, David Williams, Marcel Witberg, Addison Wynn, Luke Yeomans, Matthew Yuyek, London Zoo, Chai Koh Shin.