

Review Article

What's in a mass?

 Juan J. Calvete¹, Libia Sanz¹, Diana Mora-Obando¹,  Bruno Lomonte², Anita M. Tanaka-Azevedo³, Karen de Moraes-Zani³, Sávio S. Sant'Anna³ and Cleópatra A.S. Caldeira⁴

¹Evolutionary and Translational Venomics Laboratory, Instituto de Biomedicina de Valencia, CSIC, Valencia, Spain; ²Instituto Clodomiro Picado, Facultad de Microbiología, Universidad de Costa Rica, San José, Costa Rica; ³Laboratório de Herpetologia, Instituto Butantan, São Paulo, Brasil; ⁴Centro de Estudos de Biomoléculas Aplicadas a Saúde (CEBio), Fundação Oswaldo Cruz, Fiocruz Rondônia, Porto Velho, RO, Brasil

Correspondence: Juan J. Calvete (jcalvete@ibv.csic.es)

This short essay pretends to make the reader reflect on the concept of biological mass and on the added value that the determination of this molecular property of a protein brings to the interpretation of evolutionary and translational snake venomics research. Starting from the premise that the amino acid sequence is the most distinctive primary molecular characteristics of any protein, the thesis underlying the first part of this essay is that the isotopic distribution of a protein's molecular mass serves to unambiguously differentiate it from any other of an organism's proteome. In the second part of the essay, we discuss examples of collaborative projects among our laboratories, where mass profiling of snake venom PLA₂ across conspecific populations played a key role revealing dispersal routes that determined the current phylogeographic pattern of the species.

Introduction: fundamentals of biological mass spectrometry

The title of this short essay is inspired by Gilbert O'Sullivan's 1968 song 'What's in a Kiss'¹. A kiss is something so everyday that we rarely stop to think about its meaning and the feelings it conveys. A kiss can mean very different things to people of different cultures. One example may suffice to start a reflection in this regard.

In his 1864 book 'Savage Africa', Scottish philosopher, historian, anthropologist and explorer William Winwood Reade (1838–1875) recounts falling in love with Ananga, the daughter of an African king. After several months courting her, he dared to kiss her. Unfortunately, things developed in an unpredictable way. The girl, who had never conceived the possibility of such a thing, 'gave a shriek and bounded from the house like a frightened fawn' [1]. Only later did Reade realise that the young princess had interpreted his kiss as an intention to dine off her!

Something similar happens with the perception of the mass. We are so familiar with the intuitive Newtonian notion of 'mass' as a proxy for quantity of matter that we do not stop to think about all the conceptual complexity of this fundamental quality of everything around us that moves at subluminal speed. The large number of ways and units in which the mass can be measured or operationally defined (i.e. inertial mass, gravitational mass, atomic mass, relativistic mass...) highlights this point [2].

In Natural Sciences, the mass of polyatomic molecules is generally measured in daltons (Da) or unified atomic mass (u) units [3,4]. Prior to the 2019 redefinition of the International System of Units (SI) base unit quantities expressed in Da or u were, by definition, equivalent to g/mol and thus strictly numerically interchangeable. After the 20 May 2019 redefinition of units, this relationship is only nearly equivalent. The definition of the Dalton (1/12 of the mass of an unbound neutral atom of carbon-12 in its nuclear and electronic ground state and at rest, which by definition is exactly 12.000 Da) was not affected by the 2019 redefinition of SI base units. However, the definition of mole

Received: 19 February 2021
 Revised: 27 March 2021
 Accepted: 30 March 2021

Version of Record published:
 30 April 2021

¹*What's in a kiss; Have you ever wondered just what it is; More perhaps than just a moment of bliss; Tell me what's in a kiss* (Gilbert O'Sullivan 1980)

was changed to be the amount of substance consisting of exactly $6.02214076 \times 10^{23}$ entities, and the concept of kilogram (kg) was changed as well by taking for its definition the fixed numerical value of the Planck constant h ($6.62607015 \times 10^{-34}$ J s) in SI units ($\text{kg m}^2 \text{s}^{-1}$), where the meter (m) and the second (s) are defined in terms of c (the speed of light in vacuum, $299\,792\,458 \text{ m s}^{-1}$) and the caesium hyperfine frequency, $\Delta\nu_{\text{Cs}}$ (9 192 631 770 Hz) [3].

Proteins are large size biomolecules that perform a vast array of functions within organisms. Proteins are polymers formed by the linear arrangement of up to 20 building blocks (the amino acids, AAs), which fold into specific 3D structures that determine their activity. Proteins differ from one another primarily in their amino acid sequence, which is dictated by the nucleotide sequence of their genes. The amino acid sequence can thus be considered an important, if not the most distinctive primary molecular characteristic of a protein. The vast majority (99.9%) of the 10 739 protein sequences analysed conformed to a similar basic molecular formula $\text{C}_n\text{H}_{(1.58 \pm 0.015)n}\text{N}_{(0.28 \pm 0.005)n}\text{O}_{(0.30 \pm 0.007)n}\text{S}_{(0.01 \pm 0.002)n}$ [5]. Only molecules with the same structural formula have identical mass, which in the case of proteins means having the same polypeptide length and amino acid composition. Isobaric polypeptides may differ in the order of their building blocks but any change in number or type of residues (other than isobaric isoleucine/leucine replacement) will be reflected in the coefficients of the constituent atoms of the structural formula. The occurrence of absolutely isobaric proteins in an organism's proteome is highly unlikely. The number of isobaric sequences that can theoretically be generated by permutations with repetition of n elements (n = number of residues of the amino acid sequence) of 20 classes ($n_1, n_2, n_3 \dots n_{20}$) of building blocks is given by the formula $n!/(n_1! \times n_2! \times n_3! \times \dots \times n_{20}!)$, which for a 100-residue protein equals 1.5×10^{148} , a number much greater than the estimated number of atoms in the observable Universe (10^{77} – 10^{80})! It is therefore mandatory to conclude that only a small set of all possible theoretical combinations of amino acid sequences of any length occur in the real biological space [6]. Furthermore, biological evolution is intrinsically a creative process of molecular diversity governed by the so-called 'Fourth Law of Thermodynamics', by which when a system can follow various paths that allow it to maximise entropy, it will follow the path(s) that do so most rapidly, given the restrictions to conserve a specific architecture and to retain interactions that mediate functions [7]. Thus, while the structural divergence of the copies of a multiplied gene represents the substrate for molecular evolution, structural convergence of two genes to produce two identical copies has never been documented. In line with this view, except for highly structural and functional constrained short (<50 AAs) proteoforms whose sequences are subject to compensatory mutations of co-evolving residues [8], and proteins encoded by a recently (in the evolutionary scale) duplicated gene, there is no documentation in any organism of two proteins with identical length and amino acid composition (i.e. identical molecular mass) [9]. Additionally, chemical modifications of proteins not only play a central role in encoding and decoding cellular information, but may also modify isobaric polypeptide sequences creating distinct protein species [10,11]. So, can the mass of a given protein serve to unambiguously differentiate it from any other protein molecule? In this mini-review we advocate for a positive response. Consubstantial with this thesis, in what follows we will discuss the critical importance of high-resolution (HR) mass spectrometry (MS) for deriving the exact (to several decimal places) monoisotopic mass of a protein ion from the resolved isotopic distribution of its mass-to-charge ratio (m/z). Access to HRMS is readily available in the field of biomolecular MS. However, the accompanied information on the isotopic distribution in HR spectra is rarely used at its full potential. Here we discuss recent examples from our own research plot that highlight the added value of accurate molecular mass profiling of snake venom toxins in evolutionary and ecological venomics.

The pillars of protein mass profiling: mass resolving power, accuracy, and the isotope distribution

A mass spectrometer is a device that measures the mass-to-charge ratio of ions. Protein molecules, as any other terrestrial matter, incorporate elemental isotopes in proportion to their natural abundances [12]. Consequently, the molecular mass of a protein is not a single value but a distribution of values, the isotopic distribution, represented by the spectral peaks of the monoisotopic ion (M) and a series of peaks ($M + 1n, M + 2n, M + 3n \dots$) differing from each other in the number (n) of isotopes in their atomic composition (Figure 1).

Accurate mass measurement requires the highest possible mass resolution to ensure that only a single molecular structure populates the protein's monoisotopic mass spectral peak. The accuracy (e.g. the difference between the measured and the experimentally measured mass-to-charge (m/z) of an ion) and resolving power

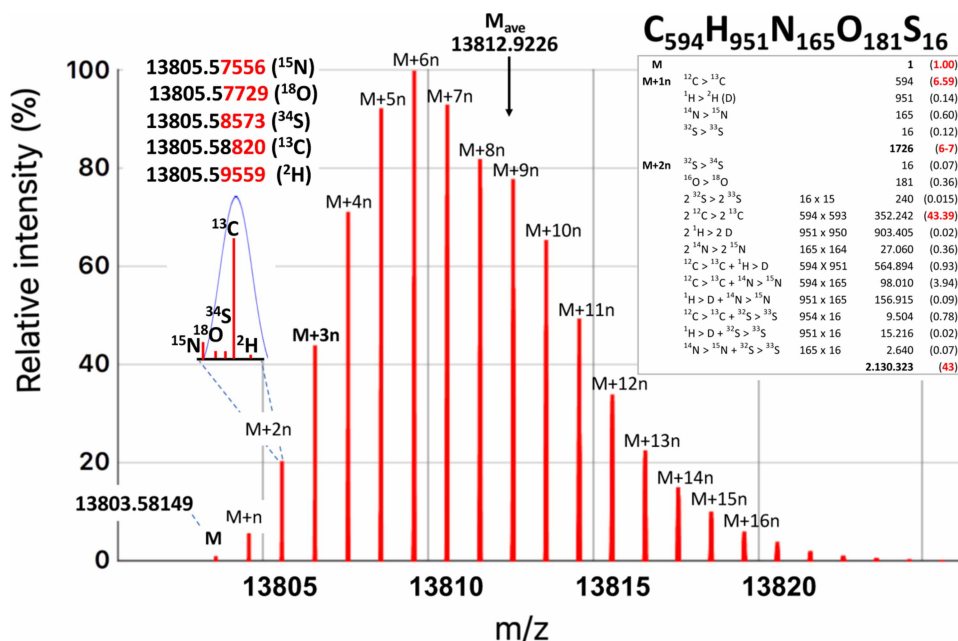


Figure 1. Deconvoluted $z = 1$ cluster of the isotopic distribution of electrospray ionisation mass spectrum of a PLA₂ molecule of structural formula C₅₉₄H₉₅₁N₁₆₅O₁₈₁S₁₆.

The monoisotopic (M) and isotope-averaged (M_{ave}) masses are highlighted. The fine structure subjacent to the M + 2n isotopic peak and the calculated relative contributions and masses for the ²H, ¹³C, ¹⁵N, ¹⁸O and ³⁴S isotopologues is displayed. The inserted right panel shows theoretically number of distinct ensembles of all the quasi-isobaric isotopic topological variants containing an equal number of additional neutrons with respect to the monoisotopic species, that could form part of the envelope of the isotopic peaks M + n and M + 2n. Isotopic combinations that, due to the relative abundance of the isotopes they contain, may contribute at least one structure, are rendered in red.

(defined as $M/\Delta M$, where M is the m/z of an analyte's ion peak and ΔM is the full width at half maximum (FWHM) of that peak) of protein mass measurements by biomolecular mass spectrometry has constantly increased over the last decade. The highest available broadband mass resolving power and mass accuracy is from Fourier transform (FT) orbital trapping (i.e. Orbitrap) and ion cyclotron resonance (ICR)-MS mass analysers. With resolving power of 1 000 000–10 000 000 and mass accuracy ± 0.5 –1 ppm (0.0005–0.001u) at 1000u, these MS configurations are capable of determining the monoisotopic mass of multicharged medium-size (15–25 kDa) protein ions to >3–4 decimal places.

The monoisotopic mass, represented by the 'lightest' peak which is also the least abundant species of the isotopic distribution, is the most reliable mass characteristics of biomolecules. The main reason is that this mass spectral signal has a unique isotopic composition made up by the most abundant, which are also the lightest, isotopes of biomolecular elements (¹²C, ¹H, ¹⁴N, ¹⁶O, and ³²S) and therefore its mass is unaffected by isotopic variabilities. Mass assignments based on any of the other (usually the most abundant) isotopic peak may introduce an error due to the incorrect assignment of the number of heavy isotopes in that peak. Furthermore, due to the different mass defect contributed by the different heavy isotopes of CHNOS, isotopic peaks other than the monoisotope are ensembles of all the isotopic variants containing equal number of additional neutrons with respect to the monoisotopic species (Figure 1). Mass defect differences between the light and heavy isotopes of CHNOS are due to the different relativistic binding energy to accommodate a neutron in the nucleus. Every isotope of every chemical element has a subtly different mass defect, meaning that each of the isotopic distribution mass peaks exhibit isotopic fine structure, which can however be resolved at sufficiently resolving power. For example, at ~ 2 Da above the monoisotopic mass of a protein, the isotopologues having two ¹⁵N, one ³⁴S, one ¹⁸O, two ¹³C, and two ²H are 1.99407 Da, 1.995796 Da, 2.004244 Da, 2.00671 Da, and 2.014102 Da, respectively, above the monoisotopic mass. Resolution of the fine structure of isotopologue peaks of the 16+ charge state of a 15.8 kDa protein is achievable with the resolving power in excess of 8×10^6 of high magnetic field (9.4T) FT-ICR instruments [13].

The fine structure underlying the $M + 1$ and $M + 2$ isotopologue peaks can give information about the chemical structure of the monoisotopic ion. Hence, identification of the ^{34}S isotopomer can unequivocally confirm the presence and number of ^{32}S atoms in the monoisotopic species [13]. The increased probability for multiple heavy isotopes as the mass of a molecule increases causes a decrease in the relative abundance of the monoisotopic peak. Fine structure resolved isotopic distribution is particularly useful for proteins with masses larger than 13 kDa, such as the snake venom phospholipase A_2 whose mass spectrum is displayed in Figure 1, where the monoisotopic mass (13803.58149 Da) is only 0.0197% of the most abundant ion, and mass imprecision (which ultimately limits mass accuracy) depends directly on the signal-to-noise ratio [14].

The impressive analytical capability of frequency-based mass analysers does not necessarily represent the theoretical limit of the performance characteristics of FTMS. Leucine and isoleucine are positional isomeric amino acids, and thus have identical mass. Positional isomers have different heat of formation, and thus different relativistic mass, $\Delta E = \Delta mc^2 = \sim 1 \text{ eV}$, which corresponds to a mass difference of $\sim 10^{-9}$ Da. Mass resolving power increases linearly with increasing magnetic field strength, and thus achieving ultrahigh resolving power to resolve unique elemental assignment for positional isomeric compositions will likely go hand in hand with the development and implementation of higher superconducting magnets.

Since its development for routine analysis at the turn of the current century, molecular mass spectrometry along with omics technologies have revolutionised our way of conceiving biological research. The continuous increase in resolving power of mass analysers represents a quantum leap to made inroads into integrative biological research in the gas phase [15–17]. In the previous section we have reflected on the information that can be derived from the simple act of measuring the mass of a polyatomic compound in general, and of a protein in particular. Based on the instrument resolution, different levels of detail can be achieved, i.e. aggregated, fine, and hyper-fine structure [18]. An inherent drawback with cutting-edge technologies capable of revealing the hyperfine structure of the isotopic distribution of molecular mass, is the high cost of both the instrumentation and its maintenance. Fortunately, on the one hand, most today's benchtop mass spectrometry configurations are endowed with analytical capabilities for fine mass measurements, and on the other hand, not every biological experiment requires determining molecular masses with hyper-fine resolving power. Consequently, the trend in biological mass spectrometry is to have in-house instrumentation to carry out routine work and submit samples, whose analysis requires specific technology, to specialised services or technology leading laboratories. The aim of this section is to show examples of how the concepts exposed in the previous section, particularly mass profiling of snake venom PLA_2 molecules, have been useful to elaborate biological stories in the context of ecological-evolutionary venomomics projects. We focused on PLA_2 s because due to its size (13–14 kDa) and absence of post-translational modifications (other than seven absolutely conserved intramolecular disulfide bonds), the monoisotopic molecular mass can be accurately derived both *in silico* (calculated from the amino acid sequence translated from a species-specific transcriptomic database) and experimentally determined (i.e. by electrospray-ionisation mass spectrometry, ESI-MS) using most of the available benchtop mass spectrometry systems [19]. However, it should be emphasized that the concepts and applications of mass profiling in venomomics are not limited to a particular type of toxin. Theoretically, any toxin class whose identity can be unambiguously derived from its mass can serve the same purpose. This is the case of PI snake venom metalloproteinases (PI-SVMP), cysteine-rich secretory proteins (CRISP), and disintegrins (DISI), commonly found in Viperidae venoms, and toxins from multigene families typically expressed in elapid venoms, such as Kunitz-type inhibitors (KUN) and three-finger toxins (3FTxs) [20,21]. This cast of molecules expands the possibilities of applying mass profiling to study important snake species whose venoms lack PLA_2 s, e.g. the African non-spitting cobras *Naja annulifera* [22], *N. senegalensis* [23], *N. haje* [24] and *N. nivea* [24].

Venom toxins evolve by accelerated evolution under strong adaptive selective pressures [25]. Acting through a variety of mechanisms neutral, positive and balancing selection promote structural and functional innovation, whereas purifying (or negative) selection operates conserving structurally and functionally important residues in venom toxins. Conserved cysteines and buried residues appear to be under strong purifying selection due to their role in maintaining the active conformation of the toxin's fold, whereas residues under diversifying selection are generally surface-exposed, often localised in loops extending from the compact central core. Consequently, absolute conservation of toxin structure may occur only within and between recently (in evolutionary terms) genetically isolated populations [26]. The smallest mass difference generated by a single, quasi-isobaric Gln/Lys, amino acid change is ± 0.03638 Da; any other single residue replacement contributes a minimum Δ_{mass} of ± 0.98401 Da. At unit mass resolution provided by quadrupole and ion trap analysers, a map of connectivities between conspecific snake populations that express venom toxins of identical or

quasi-isobaric monoisotopic masses reveals a global dispersal pattern of the species. However, the directionality of the resulting biogeographic branching pattern from the ancestral and subsequent derived populations is - difficult, if not impossible, to establish. Eco-evolutionary feedback is required to unveil fine spatio-phenotypic venomics correlations that inform about the genetic flow underlying individual dispersal steps that ultimately determine the biogeographic structuring of venom variation within wide-ranging species. All these aspects were taken into account for the elaboration of this essay. Research selected here to illustrate the biological value of determining molecular masses of venom toxins in snake venomics projects were recorded initially on a Sciex QTRap2000 MS instrument (resolving power of 6.000 at full-width at half-maximum, FWHM) and, more recently, using a Synapt G2 HR mass spectrometer (R in excess of 40.000 FWHM).

On the added value of measuring the mass of venom toxins in evolutionary and translational snake venomics

Venoms are integrated phenotypes used for predatory and defensive purposes by a wide phylogenetic range of organisms [27–29]. Understanding these complex evolutionary and intrinsically ecological traits demands a detailed picture of the functional significance and evolutionary history of the proteins that make up venoms. Adaptations to ecosystems require evolutionary changes of both morphological and molecular phenotypic traits that maximise the organism's fitness in local environments. Despite being traits of moderate genetic complexity in terms of the number of genes that encode toxin families, within- and between-species venom variability seems to be a common feature at all taxonomic levels. Genomic reorganisations and post-transcriptional regulation of the expression patterns of toxin messengers have been reported to underlie in-space (geographic) and in-time (ontogenetic) venom variability [30]. Understanding how natural selection shapes the genetic variation in populations underlying the adaptations that drive intra-specific venom evolution is of paramount importance in evolutionary biology, and requires that research questions be formulated from an ecologically informed perspective. The geographic pattern of intraspecific venom toxin conservation among disjunct conspecific populations provides clues for tracing dispersal routes that accounted for the current biogeographic distribution of the species [31,32]. The occurrence of intraspecific individual variability resulting through divergence in allopatry highlights the concept that a species should be considered as a group of metapopulations [33]. This, in turn, may help rationalising the often documented compositional and toxinological variability of snake venoms, particularly in wide-ranging species. The examples that follow illustrate this point.

Mass profiling of PLA₂ molecules provides a venomics ground to trace Russell's viper landscape of structural and toxinological variability across the Indian subcontinent

The well-documented variable clinical manifestations of envenomings caused by Russell's viper (*Daboia russelii*) in different areas of the Indian subcontinent indicate geographical variations in venom composition. However, only venom obtained from a single source, the Irula tribal snake venom extraction cooperative situated in the south Indian state of Tamil Nadu, is being used for manufacturing antivenoms to treat Russell's viper bites in India and surrounding countries. Consequently, Indian anti-Russell's viper antivenoms are sub-optimal, and in some cases even inadequate, to fully neutralise all the geographic-specific clinical complications across its wide range (consult [34] and therein cited references). PLA₂s represent the dominant toxin family in the venom proteomes of *D. russelii* from Bangladesh, Sri Lanka, Pakistan, and South India. Figure 2 displays the distribution of *D. russelii* venom PLA₂s across the Indian subcontinent. Connecting populations that share identical PLA₂ molecules depicts a rough picture of the dispersal pattern of Russell's viper ancestors to its current biogeographic distribution. The data led us to propose two alternative routes of radiation for Russell's viper, towards the northeast of India and Bangladesh and towards south India and Sri Lanka, with a common origin in Pakistan. The mass profiling informed phylogenomics landscape provided a ground for rationalising the geographic variability in venom composition and the documented North-South divide in Russell's viper envenoming in India. The dispersal route towards the south explains the paraspecific neutralisation activity exhibited by antivenoms raised against *D. russelii* venom from Tamil Nadu or Sri Lanka towards the toxic effects, including lethality, of *D. russelii* venom from Pakistan, and their poor outcome against venoms from

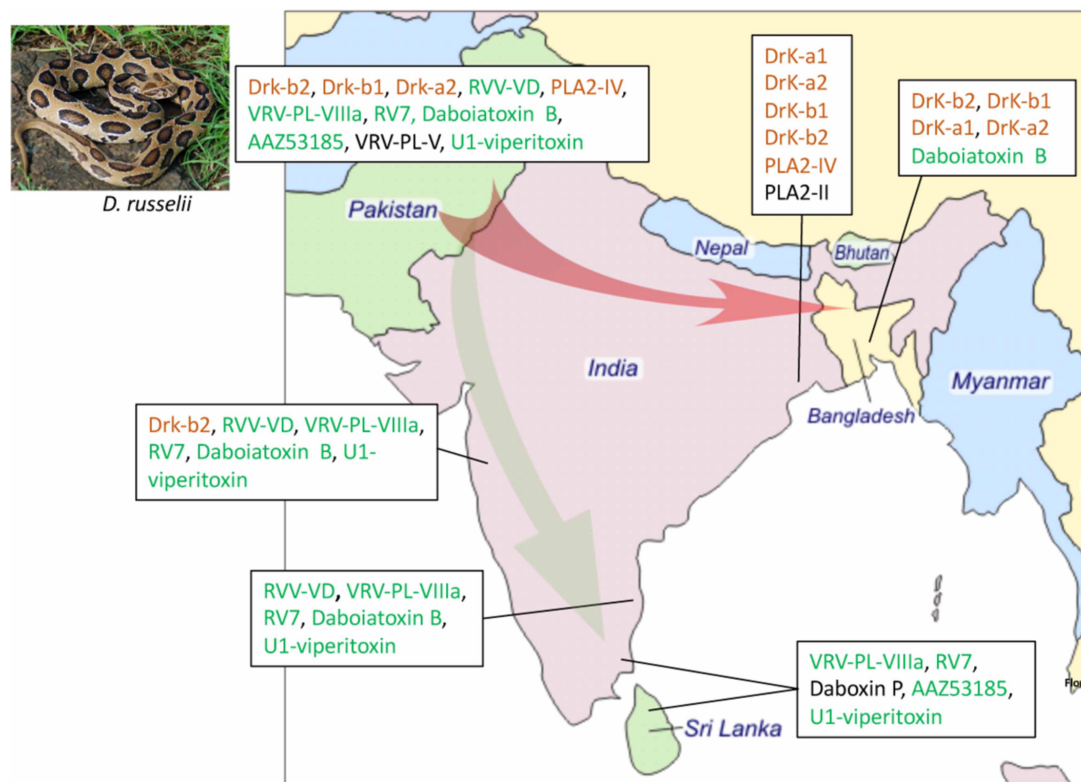


Figure 2. Distribution of venom PLA₂ molecules across Russell's viper range.

The pale green and brown arrows highlight, respectively, the proposed Northwest-South and Northwest-East dispersal routes that emerge by connecting geographic variations of Russell's viper venoms sharing PLA₂ molecules. Shared PLA₂s distributed along the northwest-northeast and northwest-south dispersal routes are color-coded in brown and green, respectively (for primary data and bibliography, please consult [34]).

northern India and Bangladesh [34]. On the other hand, this knowledge offers the possibility of improving current antivenoms in a scientifically informed way.

Tracing dispersal patterns of the wide-ranging species *Bothrops atrox* and *B. asper*

The genus *Bothrops* includes at least 50 species of venomous pitvipers (Viperidae: Crotalinae) which are widely distributed throughout the Americas, from México to southern Argentina. *Bothrops* species inhabit diverse habitats, from tropical and subtropical forests to arid and semiarid regions, and from sea level to altitudes of more than 3000 m [35]. The common ancestor of the extant species within genus *Bothrops* was the first viperid to colonise South America, sometime during the late Miocene. Rapid dispersal and diversification of *Bothrops* across South America may have been favoured through adaptive radiations into habitats devoid of viperid competitors. A single species, the ancestor of *B. asper*, reinvaded Central America, where it remains the only widespread species of *Bothrops*. Biogeographic studies support *B. asper* as the first species complex to split from the *B. atrox* clade in the Pliocene, ~3.02–2.32 Mya, as the result of geological events, including the uplift of the Guanacaste, Central, and Talamanca mountain ranges in Mesoamerica, the late Miocene through the Pliocene; the closure of the Isthmus of Panamá; and the uplift of the Andes mountain range in South America. Habitat fragmentation promoted new ecological niche opportunities and led to allopatric speciation in these groups [36,37]. We have applied PLA₂ mass profiling to elucidate biogeographical patterns of the highly adaptable and widely distributed *B. atrox* and *B. asper*, which are considered the most dangerous snakes throughout much of their distribution range in the Amazon region (*B. atrox*) [31] and throughout Central America and west of the Andes in northern South America (*B. asper*) [32].

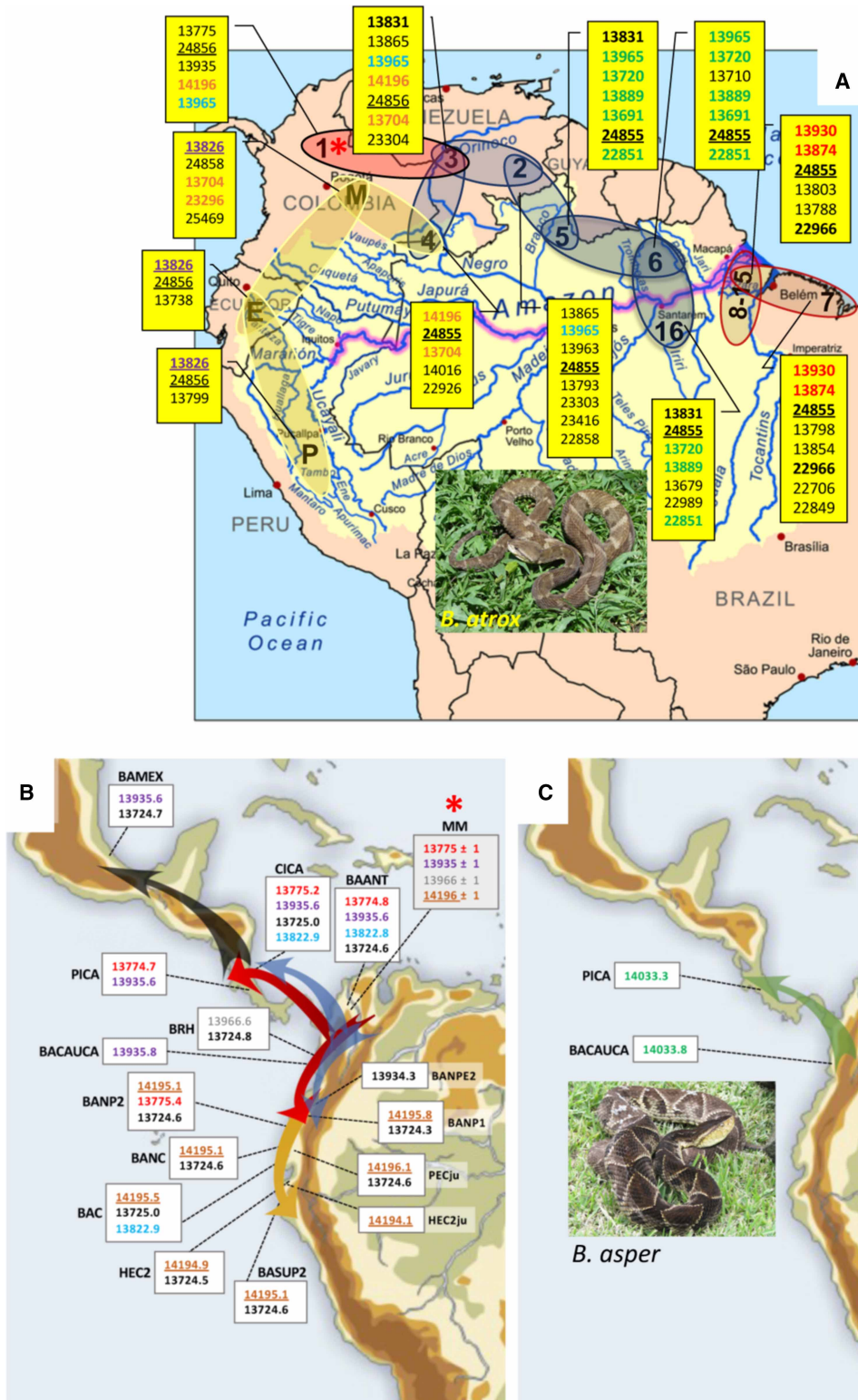


Figure 3. Part 1 of 2
 (A) Physical map of northern South America highlighting the Amazon River basin and the sampling localities for the *B. atrox* venoms investigated in Colombia: Magdalena Medio (Antioquia Department) (1); Venezuela: El Paují (2) (Orinoquía) and Puerto Ayacucho (3) (Amazonía); and Brazil: São Gabriel de Cachoeira (4) and Presidente Figueiredo (5) (Amazonas), São Bento (7) (Maranhão), and Monte Alegre (6), Ananindeua (8), Santa Isabel (9), Tucuruí (10), Icoaraci (11), Barcarena (12), Acara (13), Belém

Figure 3.

Part 2 of 2

(14), Ilha de Mosqueiro (15), and Santarém (16) (Pará). Populations of *B. atrox* whose venoms contain common molecules (PLA₂s, 15 kDa; PI-SVMPs, 23 kDa) are linked revealing a connectivity map. The Andean corridor connecting the Colombian and Venezuelan llanos with the southern block of savannas through the eastern Andean slopes (M-E-P), which is hypothesised to have served in the trans-Amazonian southward expansion of *B. atrox*, is highlighted in gold. **(B)** Displays the distribution of conserved PLA₂ molecules between *B. atrox* from Venezuela and *B. asper* from the Colombian Magdalena Medio Valley (MM), used as molecular markers for tracing the initial dispersal events of ancestral *B. asper* from northwestern Colombia southwards along the Pacific slopes of the Colombian and Ecuadorian Andes ('West Andes route') and northwestwards entering the Isthmus of Panamá towards Costa Rica and México ('Panamá-México route') [32]. **(C)** Highlights a proposed more recent 'Colombia-Costa Rica' dispersal route from the south-western Colombian population BACAUCA to the Pacific versant of Costa Rica, an event that may have occurred after the separation of the Caribbean and the Pacific *B. asper* lineages, as a consequence of the uplift of the Central Mountain Chain during the late Pliocene/early Pleistocene, dated at 3.20 (CI_{95%} = 2.32–4.31) Mya [37]. Panel **A** was adapted from [31] and panels **B** and **C** are reproduced from [32]. These figures are used here under a CC-BY 4.0 license. The *B. asper* picture in panel **C** was kindly provided by Mahmood Sasa[®].

Panel A of **Figure 3** shows a map of connectivities revealing a phylogeographic pattern consistent with a gradual expansion of *B. atrox* range. Analysed in the frame of current molecular evidence, which consistently places the northern Venezuelan haplotype as the sister group to all other South American mainland *Bothrops* haplotypes [38], the proteomic data suggest the existence of an Andean corridor linking northern and southern Amazonian *B. atrox* populations. The data also showed a 'PLA₂ connection' between the *B. atrox* population of Puerto Ayacucho (Venezuelan Amazonia, 3) and the Colombian Magdalena Medio Valley *B. asper* population (1) (**Figure 3B**). This finding is consistent with a model of dispersal of the ancestors of *B. asper* along a north-east to northwest path, presumably through corridors between the Venezuelan Cordillera de Mérida and the Colombian Cordillera Oriental and Sierra de Perijá and Sierra Nevada de Santa Marta (asterisk in **Figure 3A**), reaching the Pacific coast of Isthmian Central America. The pattern of PLA₂ conservation among Ecuadorian, Colombian, Costa Rican, and Mexican *B. asper* lineage venoms displayed in panels B and C of **Figure 3** can be explained by at least two dispersal events, a Y-shaped path with one branch of the fork running southwards along the narrow strip of lowlands of the Pacific slopes of the Colombian and Ecuadorian Andes ('West Andes route'), with the other branch entering the Isthmus of Panamá towards Costa Rica and México ('Panamá-México route'), and another radiation from southern Ecuador northward into Mesoamerica ('Colombia-Costa Rica route') characterised by the introgression of PLA₂s [14 033 Da] in the Pacific population but not in the *B. asper* stock of the Atlantic side of Costa Rica. This scenario is compatible with Savage's two-step dispersal-pulse hypothesis [39], by which the first episode took place some 3.4 Mya when the sea level fell and South American taxa were able to invade Mesoamerica before the closure of the Panamanian land bridge, and the second dispersal episode one million years later, after the final closure of the Isthmian bridge and after the isolation of the Caribbean and Pacific Costa Rican *B. asper* phylogroups.

A case of hybridisation between *Bothrops bilineatus* subspecies revealed by PLA₂ mass profiling

The arboreal South American palm pitviper (*Bothrops bilineatus*) lineage emerged within the Pliocene *Bothrops taeniatus* clade c. 4.3 Mya [35], during a period in which the rapid uplift of the eastern Andes enhanced environmental opportunities for ecological divergence and speciation. Two broad-ranging subspecies are currently recognised, the Amazonian palm pitviper *B. bilineatus bilineatus* (Wied-Neuwied, 1821) in the northern and western Amazon region, and the two-striped forest pitviper *B. bilineatus smaragdinus* (Hoge, 1966) with Atlantic Forest, Guiana Shield, and central Amazonian clades (**Figure 4A**). A recent comparative proteomics analysis of the venoms of a *B. b. bilineatus* specimen collected from an adult snake in Porto Velho (State of Rondônia, Brazil) and two *B. b. smaragdinus* geographic variants (from the Peruvian Department of Junin and from two snakes born in captivity in the Laboratory of Herpetology of Instituto Butantan, São Paulo, Brazil) [40] revealed that the two Brazilian taxa share an identical PLA₂ molecule between themselves but not with Peruvian *B. b. smaragdinus* (**Figure 4B**). This seemingly unusual result only found a logical explanation when the origin of the poisons donor lineages was investigated. Venom from *B. b. bilineatus* was collected from an

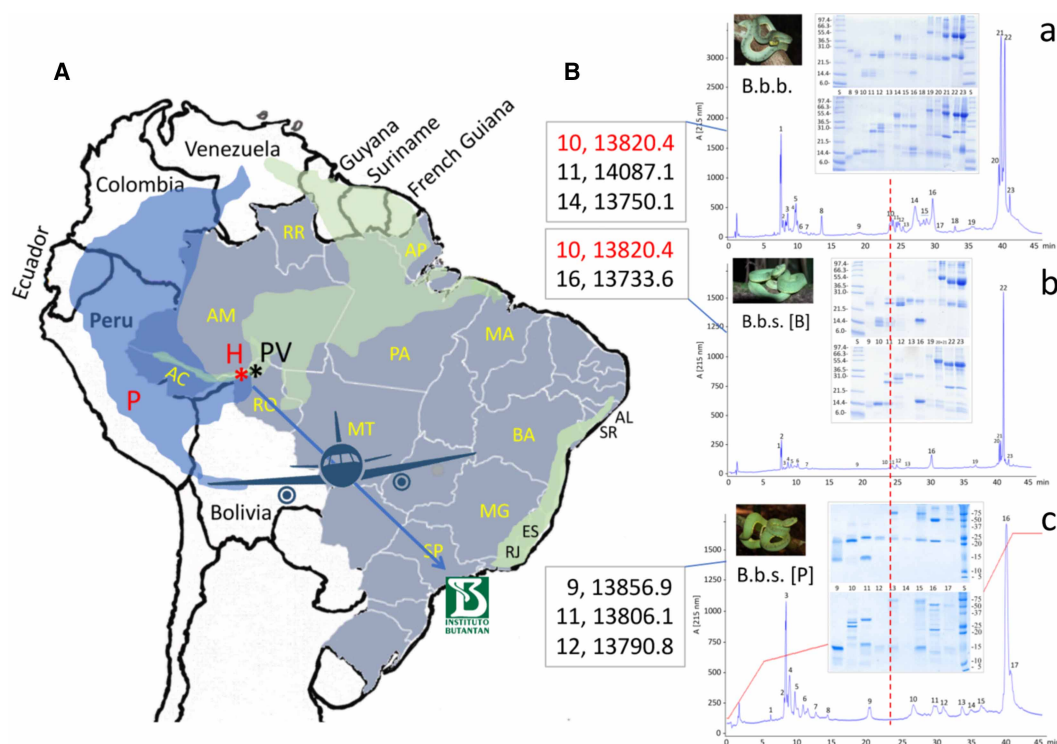


Figure 4. (A) Displays the approximate distribution ranges of South American palm pitvipers, *B. b. bilineatus* in light green and *B. b. smaragdinus* in light blue. The geographic origin of Brazilian *B. b. bilineatus* and *B. b. smaragdinus* (Porto Velho, Rondônia) and the *B. b. smaragdinus* from Chanchamayo Province, Department of Junin (Perú) are labelled PV, H, and P, respectively. The arrow and the logo of an airplane leaving Rondônia (RO) and arriving in São Paulo (SP) refers to the transportation to Instituto Butantan in 2011 of a pregnant *B. b. smaragdinus* female rescued in September 2011 from the Santo Antônio Hydroelectric Plant (Porto Velho -Rondônia) and sent to the Laboratory of Herpetology of Instituto Butantan (São Paulo), where it gave birth to a litter of which a female and a male specimens kept since then in captivity [41] contributed the venom analysed by Sanz and co-workers [40]. (B) Shows reversed-phase chromatographic separations of the venom proteins of *B. b. bilineatus* (*Bbb*, Porto Velho, Rondoônia, Br) (a); *B. b. smaragdinus* (*Bbs* [B] from Instituto Butantan (b); and *B. b. smaragdinus* (*Bbs* [P] from the Chanchamayo Province, Department of Junin (Perú) (c). This figure is adapted from Figures 1 and 2 of ref. [40], and used here under CC-BY 4.0 license.

adult snake in Porto Velho in 2008 and was part of reference venom collection of the Centro de Estudos de Biomoléculas Aplicadas a Saúde (CEBio), Fiocruz Rondônia. Venom provided by Instituto Butantan was pooled from an adult female and an adult male specimens born October 21, 2011 to a *B. b. smaragdinus* female rescued in September 2011 from the Santo Antônio Hydroelectric Plant on the Madeira River and sent to the Laboratory of Herpetology of Instituto Butantan [41]. The Santo Antônio hydroelectric dam on the Madeira River is 6 km southwest of Porto Velho, suggesting that both *B. bilineatus* subspecies may live in sympatry in Rondônia, and may eventually hybridise in areas of secondary contact resulting in introgression of venom genes.

Perspectives

- Tracing the natural history of snakes demands a detailed picture of structure-function correlations of the toxins that make up their venoms. Toxin mass profiling across evolutionary clades complements molecular approaches to integrate and rationalise *omics* data in the context of an evolutionary and ecological hypothesis.

- Venoms act as integrated adaptive phenotypes aiding snakes to manage their hunting success while minimising risks posed by predators and dangerous prey, and inherent physical threats in its habitat. Rationalisation in the context of the natural history of the snake lineage of the landscape of venom variability within and between extant conspecific populations, is of paramount importance to understand the molecular basis of local adaptations.
- The depth to which a proteome can be analysed is directly related to the completeness of the database against which the proteomic data are mapped. The emerging field of comparative snake genomics promises to uncover a treasure trove of biological information and give venomomics a big boost in the reconstruction of the evolutionary scenario of venom evolvability-aided local adaptations and speciation of snakes.

Author contributions

Conceptualization: Juan J. Calvete and Bruno Lomonte. Investigation: all authors. Writing — original draft: Juan J. Calvete. Writing — review and editing: all authors.

Funding

This study was partly supported by grant BFU2017-89103-P from the Ministerio de Ciencia, Innovación y Universidades, Madrid (Spain).

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Abbreviations

FT, Fourier transform; FWHM, full width at half maximum; HR, high-resolution; ICR, ion cyclotron resonance; MS, mass spectrometry; SI, International System of Units.

References

- 1 Winwood Reade, W. (1864) *Savage Africa: Being the Narrative of A Tour in Equatorial, Southwestern and Northwestern Africa*, Harper & Brothers Publishers, New York, p. 452
- 2 Hecht, E. (2006) There is no really good definition of mass. *Phys. Teacher* **44**, 40–45 <https://doi.org/10.1119/1.2150758>
- 3 International Bureau of Weights and Measures (2006) *The International System of Units (SI)*, 8th edition, p. 126, ISBN 92-822-2213-6. (<https://www.bipm.org/en/measurement-units/base-units.html>)
- 4 Mohr, P.J., Taylor, B. N. and Newell, D.B. (2011). CODATA Recommended Values of the Fundamental Physical Constants: 2010, (<https://physics.nist.gov/cuu/Constants/index.html>)
- 5 Torabizadeh, H. (2011) All proteins have a basic molecular formula. *World Acad. Sci. Eng. Technol.* **5**, 501–505
- 6 Otaki, J.M., Ienaka, S., Gotoh, T. and Yamamoto, H. (2005) Availability of short amino acid sequences in proteins. *Protein Sci.* **14**, 617–625 <https://doi.org/10.1110/ps.041092605>
- 7 Martínez-Castilla, L.P. and Martínez-Kahn, M. (2010) Darwin y el desarrollo de otra ley de la termodinámica. *Educ. Quím.* **21**, 230–237
- 8 Kowarsch, A., Fuchs, A., Frishman, D. and Pagel, P. (2010) Correlated mutations: a hallmark of phenotypic amino acid substitutions. *PLoS Comput. Biol.* **6**, e1000923 <https://doi.org/10.1371/journal.pcbi.1000923>
- 9 Tiessen, A., Pérez-Rodríguez, P. and Delaie-Arredondo, L.J. (2012) Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes. *BMC Res. Notes* **5**, 85 <https://doi.org/10.1186/1756-0500-5-85>
- 10 Schlüter, H., Apweiler, R., Holzhütter, H.-G. and Jungblut, P.R. (2009) Finding one's way in proteomics: a protein species nomenclature. *Chem. Cent. J.* **3**, 11 <https://doi.org/10.1186/1752-153X-3-11>
- 11 Compton, P.D., Kelleher, N.L. and Gunawardena, J. (2018) Estimating the distribution of protein post-translational modification states by mass spectrometry. *J. Proteome Res.* **17**, 2727–2734 <https://doi.org/10.1021/acs.jproteome.8b00150>
- 12 Atomic Weights and Isotopic Compositions for All Elements (https://physics.nist.gov/cgi-bin/Compositions/stand_alone.pl?ele=&all=all&ascii=html&isotype=some)
- 13 Shi, S.D., Hendrickson, C.L. and Marshall, A.G. (1998) Counting individual sulfur atoms in a protein by ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometry: experimental resolution of isotopic fine structure in proteins. *Proc. Natl Acad. Sci. U.S.A.* **95**, 11532–11537 <https://doi.org/10.1073/pnas.95.20.11532>
- 14 Marshall, A.G., Blakney, G.T., Chen, T., Kaiser, K.N., McKenna, M.A., Rodgers, P.R. et al. (2013) Mass resolution and mass accuracy: how much is enough? *Mass Spectrom.* **2**, S0009 <https://doi.org/10.5702/massspectrometry.S0009>

- 15 Marcoux, J. and Robinson, C.V. (2013) Twenty years of gas phase structural biology. *Structure* **21**, 1541–1550 <https://doi.org/10.1016/j.str.2013.08.002>
- 16 Toby, T.K., Fornelli, L. and Kelleher, N.L. (2016) Progress in top-down proteomics and the analysis of proteoforms. *Annu. Rev. Anal. Chem.* **9**, 499–519 <https://doi.org/10.1146/annurev-anchem-071015-041550>
- 17 Wörner, T.P., Shamorkina, T.M., Snijder, J. and Heck, A.J.R. (2021) Mass spectrometry-Based structural virology. *Anal. Chem.* **93**, 620–640 <https://doi.org/10.1021/acs.analchem.0c04339>
- 18 Valkenborg, D., Mertens, I., Lemiere, F., Witters, E. and Burzykowski, T. (2012) The isotopic distribution conundrum. *Mass Spectrom. Rev.* **31**, 96–109 <https://doi.org/10.1002/mas.20339>
- 19 Calvete, J.J. (2014). The expanding universe of mass analyzer configurations for biological analysis. In *Plant Proteomics: Methods and Protocols, Second Edition, Methods in Molecular Biology* (Jorrín-Novo, J.V., Komatsu, S., Weckwerth, W. and Wienkoop, S., eds), vol. **1072**, pp. 61–81, Humana Press, New York
- 20 Calvete, J.J., Juárez, P. and Sanz, L. (2007) Snake venomomics. Strategy and applications. *J. Mass Spectrom.* **42**, 1405–1414 <https://doi.org/10.1002/jms.1242>
- 21 Ainsworth, S., Petras, D., Engmark, M., Süßmuth, R.D., Whiteley, G., Albuлесcu, L.O. et al. (2018) The medical threat of mamba envenoming in sub-Saharan Africa revealed by genus-wide analysis of venom composition, toxicity and antivenomics profiling of available antivenoms. *J. Proteomics* **172**, 173–189 <https://doi.org/10.1016/j.jprot.2017.08.016>
- 22 Tan, K.Y., Wong, K.Y., Tan, N.H. and Tan, C.H. (2021) Quantitative proteomics of *Naja annulifera* (sub-Saharan snouted cobra) venom and neutralization activities of two antivenoms in Africa. *Int. J. Biol. Macromol.* <https://doi.org/10.1016/j.jbiomac.2020.04.173>
- 23 Wong, K.Y., Tan, K.Y., Tan, N.H. and Tan, C.H. (2021) A neurotoxic snake venom without phospholipase A(2): proteomics and cross-neutralization of the venom from Senegalese cobra, *Naja senegalensis* (Subgenus: Uraeus). *Toxins* **13**, 60 <https://doi.org/10.3390/toxins13010060>
- 24 Kazandjian, T.D., Petras, D., Robinson, S.D., van Thiel, J., Greene, H.W., Arbuckle, K. et al. (2021) Convergent evolution of pain-inducing defensive venom components in spitting cobras. *Science* **371**, 386–390 <https://doi.org/10.1126/science.abb9303>
- 25 Jackson, T.N.W. and Koludarov, I. (2020) How the toxin got its toxicity. *Front. Pharmacol.* **11**, 574925 <https://doi.org/10.3389/fphar.2020.574925>
- 26 Malhotra, A. (2017) Mutation, Duplication, and More in the Evolution of Venomous Animals and Their Toxins. In *Evolution of Venomous Animals and Their Toxins. Toxinology* (Gopalakrishnakone, P. and Malhotra, A. eds), Springer, Dordrecht
- 27 Calvete, J.J. (2017) Venomomics: integrative venom proteomics and beyond. *Biochem. J.* **474**, 611–634 <https://doi.org/10.1042/BCJ20160577>
- 28 Undheim, E. and Jenner, R. (2017) *Venom: The Secrets of Nature's Deadliest Weapon*, Natural History Museum, London
- 29 Jackson, T.N.W., Jouanne, H. and Vidal, N. (2019) Snake venom in context: neglected clades and concepts. *Front. Ecol. Evol.* **7**, 332 <https://doi.org/10.3389/fevo.2019.00332>
- 30 Casewell, N.R., Jackson, T.N.W., Laustsen, A.H. and Sunagar, K. (2020) Causes and consequences of snake venom variation. *Trends Pharmacol. Sci.* **41**, 570–581 <https://doi.org/10.1016/j.tips.2020.05.006>
- 31 Calvete, J.J., Sanz, L., Pérez, A., Borges, A., Vargas, A.M., Lomonte, B. et al. (2011) Snake population venomomics and antivenomics of *Bothrops atrox*: Paedomorphism along its transamazonian dispersal and implications of geographic venom variability on snakebite management. *J. Proteomics* **74**, 510–527 <https://doi.org/10.1016/j.jprot.2011.01.003>
- 32 Mora-Obando, D., Salazar-Valenzuela, D., Pla, D., Lomonte, B., Guerrero-Vargas, J.A., Ayerbe, S. et al. (2020) Venom variation in *Bothrops asper* lineages from North-Western South America. *J. Proteomics* **229**, 103945 <https://doi.org/10.1016/j.jprot.2020.103945>
- 33 Calvete, J.J. (2013) Snake venomomics: from the inventory of toxins to biology. *Toxicon* **75**, 44–62 <https://doi.org/10.1016/j.toxicon.2013.03.020>
- 34 Pla, D., Sanz, L., Quesada-Bernat, S., Villalta, M., Baal, J., Chowdhury, M.A.W. et al. (2019) Phylogenomics of *Daboia russelii* across the Indian subcontinent. Bioactivities and comparative *in vivo* neutralization and *in vitro* third-generation antivenomics of antivenoms against venoms from India, Bangladesh and Sri Lanka. *J. Proteomics* **207**, 103443 <https://doi.org/10.1016/j.jprot.2019.103443>
- 35 Hamdan, B., Guedes, T.B., Carrasco, P.A. and Melville, J. (2020) A complex biogeographic history of diversification in Neotropical lancehead pitvipers (Serpentes, Viperidae). *Zool. Scr.* **49**, 145–158 <https://doi.org/10.1111/zsc.12398>
- 36 Salazar-Valenzuela, D., Kuch, U., Torres-Carvajal, O., Valencia, J.H. and Gibbs, H.L. (2019) Divergence of tropical pitvipers promoted by independent colonization events of dry montane Andean habitats. *J. Biogeogr.* **46**, 1826–1840 <https://doi.org/10.1111/jbi.13661>
- 37 Saldarriaga-Córdoba, M., Parkinson, C.L., Daza, J.M., Wüster, W. and Sasa, M. (2017) Phylogeography of the Central American lancehead *Bothrops asper* (SERPENTES: VIPERIDAE). *PLoS ONE* **12**, e0187969 <https://doi.org/10.1371/journal.pone.0187969>
- 38 Wüster, W., Salomão, M.G., Quijada-Mascareñas, J.A. and Thorpe, R.S. (2002). Origins and evolution of the South American pitviper fauna: evidence from mitochondrial DNA sequence analysis. In *Biology of the Vipers* (Schuett, G.W., Höggren, M., Douglas, M.E. and Greene, H.W., eds), pp. 111–128, Eagle Mountain Publishing, LC
- 39 Savage, J.M. (2002) *The Amphibians and Reptiles of Costa Rica: A Herpetofauna Between two Continents, Between two Seas*, University of Chicago Press, Chicago and London
- 40 Sanz, L., Quesada-Bernat, S., Pérez, A., De Morais-Zani, K., Sant'Anna, S.S., Hatakeyama, D.M. et al. (2020) Danger in the canopy. comparative proteomics and bioactivities of the venoms of the South American palm Pit viper *bothrops bilineatus* subspecies *bilineatus* and *smaragdinus* and antivenomics of *B. b. bilineatus* (Rondônia) venom against the Brazilian pentabothropic antivenom. *J. Proteome Res.* **19**, 3518–3532 <https://doi.org/10.1021/acs.jproteome.0c00337>
- 41 Grego, K.F., Fernandes, W., Croce, A.P., Vasconcellos, D.R., Sant'Anna, S.S. and Coragem, J.T. (2012) *Bothrops bilineata smaragdinus* (Green Jaraca) reproduction. *Herpetol. Rev.* **43**, 492