Examining venom variation in a centipede species complex *Scolopendra morsitans* from Peninsular India

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Parikh Richard Pareshkumar

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

Date: March, 2025

Under the guidance of

Supervisor: Dr. Jahnavi Joshi,

Senior Scientist CSIR-CCMB, Hyderabad

From May 2024 to Mar 2025

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

Certificate

This is to certify that this dissertation entitled "**Examining venom variation in a centipede species complex *Scolopendra morsitans* from Peninsular India**" towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by **Parikh Richard Pareshkumar** at Department of Wildlife Conservation and Ecology, Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad under the supervision of **Dr. Jahnavi Joshi**, Senior Scientist CSIR-CCMB, Hyderabad, during the academic year **2024-2025**.

Dr. Jahnavi Joshi

Senior Scientist

CCMB, Hyderabad

Committee:

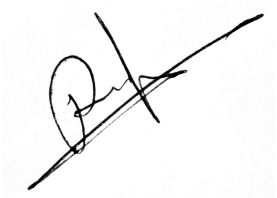
Name of your Guide: **Dr. Jahnavi Joshi**, Senior Scientist CSIR-CCMB, Hyderabad

Name of Your TAC: **Dr. Devapriya Chattopadhyay**, Professor, Department of Earth and Climate Science, IISER Pune

This thesis is dedicated to my parents.

Declaration

I hereby declare that the matter embodied in the report entitled "**Examining venom variation in a centipede species complex *Scolopendra morsitans* from Peninsular India**" are the results of the work carried out by me at the Department of Wildlife Conservation and Ecology, Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, under the supervision of **Dr. Jahnavi Joshi**, and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.



Name of Student: **Parikh Richard Pareshkumar**

Roll No. 20201148

Table of Contents

List of Tables

List of Figures

Abstract

Integrative taxonomic framework incorporates evidence from multiple sources including morphological, molecular, behavioural, and ecological characters to provide a robust species hypothesis. In this context, we employ the use of multiple molecular markers and the ecologically functional trait ‘venom’ to throw light on the evolutionary relationships within the cryptic widely distributed centipede species *Scolopendra morsitans*. We examine the morphologically cryptic diversity through molecular phylogenetic and species delimitation analyses. We also look at the interspecific variation in venom within the *Scolopendra morsitans* species complex distributed across peninsular India. We used mitochondrial and nuclear marker datasets to reconstruct the molecular phylogeny and delimit the putative species within the species complex. Our results show that *S. morsitans* species complex is paraphyletic lineage harboring multiple distinct species. We find a monophyletic clade of *S. morsitans* from peninsular India and Southeast Asia, another distinct clade of *S. morsitans* from Odisha, western India, and Himachal Pradesh, and a distinct clade of African *S. morsitans*. The *S. morsitans* clade from peninsular India and Southeast Asia is named *S. morsitans 1* while the other distinct clade from India is named *S. morsitans 2*. Both clades had three distinct species from India. We further characterized the venom using a proteo-transcriptomic approach and performed various diversity and ordination analysis to study the venom variation between these putative species. There was significant difference in the venom composition between the two major clades, however less so among the putative species within the peninsular Indian clade.

Acknowledgments

I am extremely grateful to my thesis supervisor, Dr. Jahnavi Joshi, for her persistent support, guidance, and patience for my project work throughout the period. I thank her for all her knowledge and inputs, helping me carry out the project, conceptualizing the questions, interpreting the results, and drafting the thesis.

Next, I would like to thank Aditi, who helped me at every stage of the project, taught me proteomics work, and supported me. I want to thank Pragyadeep for addressing many of my doubts and helping me with phylogenetic analyses. Next, I would like to thank Sudhanshu, Dhanush, Karunakar, Madhumithraa, and Sharvani for helping me in different ways with the project. Additionally, I am very thankful to all the members of the Evolutionary Ecology Lab for their fieldwork and for collecting the samples I have used for my project. I want to thank the HLS staff at LaCONES, CCMB, for getting the supplies from CCMB, taking my samples to the Sanger Sequencing Facility at CCMB, and helping with various other tasks. Lastly, I express my thanks to Abhishek, Marwa, Shriya, Devanand, Anupama, Pooja, Shalini, Swati, and Aniruddha, along with all the previously mentioned lab members, for their valuable suggestions and for making this period a fun experience.

I express my gratitude to my thesis expert, Dr. Devapriya Chattopadhyay, for being a TAC member of the thesis and providing me with valuable feedback for the project.

Most importantly, I am forever grateful to my parents for their wholehearted support.

Contributions

Contributor name	Contributor role
JJ, AD, RP	Conceptualization Ideas
RP, AD	Methodology
RP, AD, PR	Software
RP, AD, JJ	Validation
RP, AD, JJ	Formal analysis
JJ, AD	Investigation
JJ	Resources
RP, AD, JJ	Data Curation
RP, AD, JJ	Writing - original draft preparation
RP, JJ, AD	Writing - review and editing
RP, PR, AD, JJ	Visualization
JJ	Supervision
JJ	Project administration
JJ	Funding acquisition

RP: Richard Parikh, IISER Pune

JJ: Jahnavi Joshi, CSIR-CCMB, Hyderabad

AD: Aditi, CSIR-CCMB, Hyderabad

PR: Pragyadeep Roy, CSIR-CCMB, Hyderabad

This contributor syntax is based on the Journal of Cell Science CRediT Taxonomy[1].

[1] https://journals.biologists.com/jcs/pages/author-contributions

1. Introduction

An evolutionary lineage of populations with a common ancestor and gene flow among them is defined as a species (De Queiroz, 2007). The process of splitting such lineages into two or more as a result of some barrier and thus giving rise to new lineages is known as speciation (Queiroz, 1998). Evolutionary lineages diverge over time, accumulating differences in their morphological, genetic, and behavioral traits. During speciation, these traits could be fixed among the lineages, leading to the formation of two distinct species. This can be caused by geographical isolation, natural selection, ecological shifts, genetic drift, and polyploidy (Barraclough and Nee, 2001; Albert and Schluter, 2005; Barraclough, 2019).

In ecological speciation, divergent natural selection acts on specific traits, leading to reproductive isolation among the populations. Ecological factors, including climate, resource availability, competition, and predation, can cause trait divergence (Schluter, 2001). Studying traits that show divergence and contribute to reproductive isolation between species helps understand the mechanism for ecological speciation (Rundle and Nosil, 2005). However, it needs to be noted that trait divergence does not always lead to speciation. Divergence can occur within species, forming multiple populations without any reproductive isolation. Thus, it becomes important to ask which 'trait' might play a role in the reproductive isolation of populations, as this might help infer the evolutionary process that may have shaped the speciation process (Schluter and Rieseberg, 2022).

Understanding the evolution of morphologically cryptic species

Taxonomic identification and classification are commonly carried out using morphological key characteristics (traits). However, there exists an unreliability when species are classified based on morphological traits alone. This is because traits could have evolved through convergent evolution, can show phenotypic plasticity, or cryptic nature. The morphological traits, however visually striking they may be to the human eye, may not always be crucial for the organism's fitness or survival (Dobzhansky, 1956). The challenges in determining taxonomic groups based solely on morphological data has been well reported (Pfenninger *et al.*, 2006). Such an

issue arises when working with cryptic species, lineages which are morphologically cryptic but genetically distinct. Cryptic species have proven difficult to deal with when studying their evolutionary history and biodiversity (Slippers *et al.*, 2014).

However, before studying cryptic species, we must be confident about the cryptic nature of the species we are studying (Shin and Allmon, 2023). The lack of divergence on the morphological axis suggests that the cryptic species could have diverged on some other axis of functional traits. Since focusing solely on the morphological traits would be a biased way to understand the evolutionary history of the species. Dawson *et al.* (2021) shed light on the 'traits' that are used by ecologists to study the evolutionary history of species and highlight the importance of using alternative axes other than the morphological axis, like genetic, behavioural, and geographic axes. For example, a species complex in the fruit fly genus *Anastrepha* known to have morphologically cryptic species, has been reported to have differences in their sex chromosome length, geographic occurrences, ecological habits, wing shape; and these traits differences were able to predict the distinct species in this species complex (Selivon *et al.*, 2022). Cryptic species could also be differentiated based on their geographic distribution; for example, a study on fig wasps has reported an allopatric distribution of morphologically cryptic species and a sympatric co-occurrence of morphologically distinct species (Ct and Jm, 2017). Additionally, it also need to be noted that when studying cryptic species, it is important to focus on species which have maintained their morphological similarity over a long evolutionary period, also known as stasis, as compared to studying species which have recently diverged, as a short evolutionary period might not be enough for the species to accumulate morphological difference (Struck and Cerca, 2019).

Centipedes and their Venom as model systems

Centipedes (Chilopoda), a predatory group of soil arthropods, comprise approximately 3300 extant species. They are fossorial organisms which have found a habitat in leaf litter, soil, rock crevices, and under tree barks (Chitty, 2022). A cosmopolitan occurrence and a fossil history of 420 million years make them one of the most widespread ancient venomous animal lineages. The class Chilopoda comprises the five extant orders, which are Scutigeromorpha, Craterostigmomorpha, Lithobiomorpha, Scolopendromorpha, Geophilomorpha, and one extinct order

Devonobiomorpha (Shear and Bonamo, 1988; Edgecombe and Giribet, 2007). Centipedes are predominantly known to feed on invertebrate prey, but some species have been found to feed on vertebrate prey like bats (Lewis, 1981; Molinari *et al.*, 2005; Edgecombe and Giribet, 2007; Undheim and King, 2011).

A key trait of centipedes is their venom, which enables them to subdue their prey. Venom is an advantageous trait that impedes the prey's homeostasis (Casewell *et al.*, 2020; Arbuckle, 2021). The effectiveness of venom as an adaptive trait is reflected by the number of animal lineages, including snakes, cone snails, sea anemones, centipedes, spiders, scorpions, *etc*., in which it has convergently evolved (Holford *et al.*, 2018). Venom being a toxic substance composed of proteinaceous toxins, which is injected into another organism for predation, defence, or deterring competitors, any changes in the genes coding the venom proteins can directly alter the ecological interactions between predator and prey or between competitors. Thus venom acts as a mediator for interaction between at least two animals, and this fact provides a strong connection between the genotype, phenotype, and ecology of the animal (Arbuckle, 2020).

Despite the long evolutionary history of centipedes as venomous predators, there is limited understanding of their venom as compared to other venomous lineages like snakes, cone snails, *etc*. The centipede venom system involves modifying the first pair of legs into venom-injecting tools known as forcipules. These forcipules harbour an opening at their tip through which they directly inject venom into their prey. The venom cocktail of the centipedes is composed of enzymes like metalloproteases, serine proteases, γ-Glutamyl Transpeptidases (GGTs), Phospholipase A2 (PLA2), CentiPADs (Peptidylarginine deiminase) *etc*. These enzyme components may contribute to the extra-oral digestion of prey. Additionally, the centipede venom also comprises non-enzymatic components like neurotoxins (Scoloptoxins), cytotoxins (Beta pore-forming toxins- βPFTx), cysteine-rich secretory proteins (CRISPs), antigen 5 (Ag5), and pathogenesis-related 1 (Pr-1) proteins (CAP proteins) and various other characterized and uncharacterized proteins whose function remains unknown (Undheim *et al.*, 2015). Many of these components have been recruited once. However, some have been recruited more than once across the evolutionary history of centipedes and their venom (Undheim *et al.*, 2015).

Keeping this in mind, we would like to understand what role (if any) the adaptive trait 'venom' would play in the speciation process of morphologically cryptic centipedes. Studies at higher taxonomic levels have shown strong differences in the venom composition across the various orders of centipedes (Jenner *et al.*, 2019). Here, we would like to examine venom variation at the species level, particularly the *S. morsitans* species complex.

***Scolopendra morsitans* species complex**

Scolopendra morsitans Linnaeus, 1758, is one of the oldest known centipede species and the type species for the genus *Scolopendra*. In India, this species is widely distributed across all states except Kashmir (Shelley *et al.*, 2005). *S. morsitans* has been reported from every continent, including Asia, North America, South America, Africa, Australia, and Europe; however, the European records are considered dubious (Shelley *et al.*, 2005).

In peninsular India, the genus *Scolopendra* started diversifying approximately 85 million years ago during the Late Cretaceous Epoch (Joshi and Karanth, 2011). The diversity of *Scolopendra* species in peninsular India remains poorly understood. Among the many species of *Scolopendra* recognized in India, the two most commonly found are *Scolopendra morsitans* and *Scolopendra hardwickei*, while the rest of the species have been described based on one or two specimens/occurrences (Jangi and Dass, 1984; Lewis, 2010).

We focus on the study group *Scolopendra morsitans*, which could be a species complex rather than a single taxonomic identity. Variations within *S. morsitans* across different geographic regions, particularly the high molecular divergence, suggest the possibility of multiple cryptic species that have yet to be formally described. The debate surrounding *S. morsitans* is further complicated by its relationship with *Scolopendra amazonica*, another species reported from India. Both species are morphologically similar and have 20th leg tarsal spurs. This has led researchers to question whether these two taxa are truly distinct or just sibling species (Lewis, 2010; Joshi and Karanth, 2011). Siriwut (2016) found the monophyly of the Southeast Asian *S. morsitans*. However, monophyly of the Indian, African, and Southeast Asian *S. morsitans* needs to be examined (Joshi and Karanth, 2011).

Given the likelihood of the cryptic diversity within *S. morsitans*, further molecular and morphological studies are necessary to clarify its taxonomic status.

In this study, we ask the following questions:

1. Is *Scolopendra morsitans* a species complex or a single widely distributed species?
2. Are the Indian *Scolopendra morsitans* monophyletic?
3. Do the distinct clade/putative species within the species complex show variation in their venom phenotype, and what might drive such variation?

Given that the species of the genus *Scolopendra* show a high level of morphological similarity despite their diversification since the Late Cretaceous Epoch, they shall be considered truly cryptic species which have exhibited morphological stasis over a long evolutionary time period (Joshi and Karanth, 2011; Struck and Cerca, 2019). Previous studies using an integrative approach of morphological, molecular, and climate data on the centipede genus *Digitipes* in peninsular India have shown a presence of high cryptic diversity within this genus (Joshi and Karanth, 2012). Thus, we might find similar patterns of high cryptic diversity for the species belonging to the genus *Scolopendra*.

We are dealing with cryptic centipedes known to have low dispersal ability, so they might have an allopatric distribution (Dool *et al.*, 2022; Bharti *et al.*, 2023). An allopatric distribution means that the centipede groups are geographically isolated. Geographic isolation means they might have differing ecological niches in terms of abiotic (environment) and biotic (diet) factors. Thus the ecological traits of the centipedes would diverge and adapt to the different environments, leading to reproductive isolation (Schluter, 2001). In our case, we are looking at venom as an adaptive ecological trait. Studies in snakes have shown that environmental differences have caused venom variation (Gren *et al.*, 2017). Studies in cone snails have shown morphologically cryptic groups to have distinct venom proteomes (Himaya *et al.*, 2022). Similarly, we expect the different putative species of the morphologically cryptic *S. morsitans* species complex to have distinct venom phenotype, given they also show allopatric distribution.

2. Materials and Methods

2.1 Taxon Sampling

From 2021 to 2024, centipedes were sampled across peninsular India, *i.e.*, the Western and Eastern Ghats, from the states of Maharashtra, Karnataka, Kerala, Tamil Nadu, Telangana, Odisha, and Andhra Pradesh. The specimens were actively sampled using forceps to catch centipedes by turning rocks, wooden logs, barks, leaf litter, *etc*. The live specimens were collected and brought back in appropriately sized containers. *Scolopendra morsitans* individuals were collected from unique locations, ranging from wet forests to savannah to lateritic plateaus.

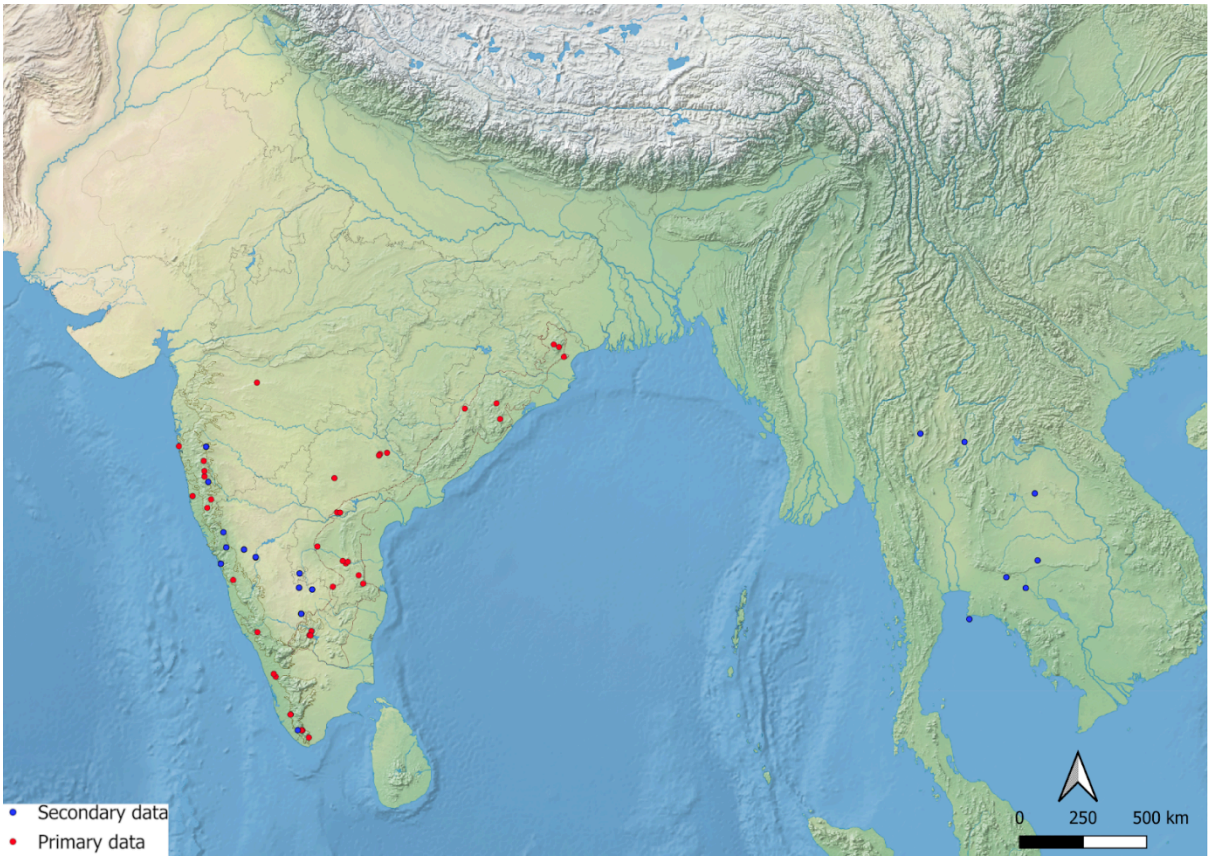


Fig 1: Sampling locations from peninsular India and Southeast Asia for DNA and protein dataset. Red dots represent the combined primary DNA (n=42) and protein (n=33) dataset, while blue dots represent the secondary DNA dataset (n=50) for *S. morsitans.*

2.2. Morphological Identification

Preliminary identification of the *Scolopendra morsitans* samples was done along with other species based on morphological keys described by Jangi and Dass (1984). *S. morsitans* is a large-sized species, and the individuals are usually easy to identify even in the field, provided their characteristic morphological patterns and the first tergite being anteriorly overlaid by a cephalic plate. The members of the Evol-EcoLab, Aditi, a graduate student, and Karunakar Majhi, a research fellow in the Evol-EcoLab, performed the identification using the Olympus SZX2-MDCU stereo zoom microscope.

2.3. DNA work

2.3.1. DNA Extraction

DNA was extracted from 58 samples of *Scolopendra* individuals and 34 samples of *Cormocephalus* individuals (sister group) preserved in 70% ethanol using the *QIAGEN DNeasy Blood & Tissue Kit*. First, the leg tissues of the centipedes were dissected using sterilised forceps and blades to perform these extractions. Depending on the individual's size, the number of legs used for DNA extraction varied between 1 and 4. The leg tissues were cut using a surgical blade and transferred to a 1.5 mL microcentrifuge tube containing a lysis buffer from the kit. The tissue was then crushed using a micro-pestle to ensure that the chitin layer on the outer part of the leg tissue didn't prevent the lysis buffer from coming in contact with the inner part of the leg tissue. Proteinase K was added to degrade proteins in the lysate, and the mixture was incubated overnight at 56°C 800 RPM. The next day, the *QIAGEN DNeasy Blood & Tissue Kit* protocol was followed to perform the downstream processing of the samples. The DNA was eluted in 50uL of Elution buffer provided in the kit. The concentration and quality of DNA were measured using a NanoDrop Spectrophotometer, and later, the DNA extracts were stored at -20 °C until further use.

2.3.2. Polymerase Chain Reaction (PCR)

The laboratory technique of PCR was used to amplify the sequence of the three markers COI, 16s rDNA, and 28s rDNA for all the samples (Mullis *et al.*, 1986). The total volume of the reaction mixture used in PCR was 25 uL. The reaction mixture

contains dNTPs, DNA template, Taq polymerase, MilliQ Water, Buffer, and Forward and Reverse primers. The amplification was performed for 35 cycles. The annealing temperatures for COI, 16s rDNA and 28s rDNA were 46°C, 49°C and 52°C, respectively. The PCR protocol from Joshi and Karanth (2011) was closely followed, and appropriate changes were made if the standard protocol did not amplify the amplicon sequence.

Step	**Temp (°C)**	**Time**
initital_den	94	1.5 min
den	95	40 sec
ann	46/49/52	45 sec
ext	72	40 sec
final_ext	72	6 min
cycles	35	
hold	4	20 min

Table 1: PCR program protocol for the three markers.

2.3.3. DNA sequence trimming and editing

The software Chromas (version 2.6.6) was used to visualize the chromatogram of both the forward and the reverse strands of each marker. The sequences were trimmed appropriately based on the Phred quality score of the sequence. Then pairwise alignment was performed on the forward and the reverse sequence in the software MEGA (version 11 and 12) using the MUSCLE algorithm with default settings (Tamura *et al.*, 2021; Kumar *et al.*, 2024). The sequences were trimmed and edited to get a consensus sequence. The consensus sequence was then exported as a FASTA file.

2.3.4. Multiple sequence alignment (MSA)

The cleaned consensus DNA sequences for each marker (COI, 16S, 28S) were used to obtain a multiple sequence alignment file. For COI, alignments were checked for stop codons to detect any pseudogenes in the data. The MSA were performed using the Invertebrate Mitochondrial codon table in MEGA (version 11 and 12). After

performing MSA, any erroneous sequences causing complications in the MSA were removed after carefully looking through the original chromatograms.

2.3.5. Maximum Likelihood Phylogeny

The command line version of IQTree2 (Minh *et al.*, 2020), as well as the IQTree web server (Trifinopoulos *et al.*, 2016), were used to construct phylogenetic trees under the Maximum Likelihood Framework for the three markers. An ultrafast bootstrap implemented in IQTree2 (Hoang *et al.*, 2018) was used to obtain branch support. ModelFinder implemented in IQTree2 (Kalyaanamoorthy *et al.*, 2017) was used to decide on a substitution model for the coding and non-coding genes. The data used for the construction of the phylogenetic trees consisted of the DNA sequence data generated in this study along with an already existing dataset of different species belonging to the genera *Scolopendra*, *Asanada*, and *Cormocephalus* (sub-family *Scolopendrinae*). The closely related *Otostigminae* subfamily was chosen as the outgroup to root the tree. The secondary data was obtained from two sources, *i.e.*, DNA sequences generated by the lab and DNA sequences available on the NCBI public repository. The trees were made using the three markers, COI, 16s rDNA, 28s rDNA, and a concatenated dataset of 16s-COI markers (n=295), and 16s-28s-COI markers (n=295). The reason for constructing a sub-family level tree is to get a better picture of the phylogenetic relationships between all the genera. The constructed trees were later visualised using FigTree (v1.4.4) and were rooted accordingly.

2.3.6. Bayesian Inference Phylogeny

The MPI version of MrBayes 3.2.7a was used to reconstruct the phylogenetic tree under the Bayesian Framework for the concatenated dataset of three markers (Huelsenbeck et al., 2001). The data was partitioned into five blocks, one for 16s marker, one for 28s marker, and three for the three codon positions of COI marker. The substitution models for each partition were obtained using the ModelFinder implemented in IQTree2 after restricting the models to the available models in MrBayes. Two runs for a total of 300 million generations were performed, and the runs were stopped once a standard deviation of split frequencies was less than or equal to 0.008. The default value for burnin of pruning the first 25% trees was used,

and a consensus tree was obtained from the remaining trees. The rooted tree was visualised in FigTree (v1.4.4).

2.3.7. Species delimitation

mPTP (multi-rate Poisson Tree Processes), a tree-based approach to species delimitation, was used for molecular delimitation within the *S. morsitans* species complex (Kapli *et al.*, 2017). Since mPTP uses single-locus data, we performed species delimitation on the separate mitochondrial marker datasets and a concatenated dataset of mitochondrial markers, which could be a single-locus dataset (Budowle *et al.*, 1999). Upon the construction of the sub-family level phylogenetic tree, the sub-clade belonging to the *Scolopendra morsitans* species complex was extracted from the larger tree; correspondingly, we used the MSA of only *Scolopendra morsitans* species complex. We provided the ML tree as an input to mPTP along with the MSA. mPTP works on a maximum likelihood and Markov Chain Monte Carlo framework (MCMC) and can provide posterior support values based on millions of MCMC generations. mPTP considers the different evolutionary rates of different groups in a phylogeny. It delimits species based on that, as compared to the other methods where all the species or genus in the phylogeny have the same evolutionary rate. Because of these, mPTP can provide a conservative estimate of species number similar to the actual taxonomy as compared to other distance and tree-based methods for species delimitation, which mostly show overestimation of species number (Kapli *et al.*, 2017). mPTP was performed using the command line of Ubuntu Linux. The final results were later mapped onto the maximum likelihood tree obtained from the concatenated dataset of three markers (16s,28s, and COI)

2.4. Venom Proteomics work

2.4.1. Venom Extraction

Upon arrival of the live animals from the field, they were housed in a cool, dark environment with regular ventilation to ensure optimal conditions. Venom extraction was carried out using a commercial TENS (Transcutaneous Electrical Nerve Stimulation) unit *via* electrostimulation. Before extraction, the animals were anesthetised by exposure to gaseous CO_2 for varying durations, depending on

factors such as body size and container sizes. Care was taken to prevent excessive pressure and prolonged CO_2 exposure within the chamber. Once anesthetised, the animal was secured upside down on a custom Velcro-fitted bench using clean forceps. A sterile metal spatula was inserted between its forcipules, followed by electrostimulation with the TENS unit (40 seconds to 3 minutes). The intensity and duration of stimulation were adjusted according to the individual's body size. Anesthesia parameters were flexibly modified as needed to accommodate situational variations. Following venom extraction, protein concentration was measured using a NanoDrop spectrophotometer. The venom was then aliquoted and stored at -80°C for future use.

2.4.2. Venom protein digestion and mass spectrometry

To identify and characterize the protein components present in the venom of the centipedes, the proteomic techniques liquid chromatography (LC) and Tandem mass spectrometry (MS/MS) are performed together, this is known as LC-MS/MS. This approach is essential for analysing large protein sets within intricate biological samples. Proteomic studies are performed in two different ways. One approach is the top-down approach, where the proteins are analysed in their intact forms, while the other approach, the bottom-up approach, relies on the breakdown of the proteins before performing LC-MS/MS (Roberts *et al.*, 2024). This approach applies the use of a proteolytic enzyme, Trypsin, which cleaves the protein chain at their carboxyl end in Lysine and Arginine residues.

To understand the venom proteome of *S. morsitans* species complex in detail and the inter-specific venom variation, we performed a bottom-up proteomic study of the individual samples by facilitating the in-solution digestion of the proteins. In this method of in-sol digestion, the protein is directly broken down into peptides with the addition of a digestion buffer in the solution, unlike the method of in-gel digestion where the gel bands are excised first and then digested into peptides. The in-sol method is used for the following reasons: reduced sample loss when working with smaller sample volumes, faster processing times, and broader protein coverage. This method of protein digestion is also helpful when the protein amount is in lesser

concentrations to begin with (such as centipede venom in this case). The protein sample preparation for LC-MS/MS analysis was performed using the *Thermo Fisher Scientific In-Solution Tryptic Digestion and Guanidination Kit.* After the sample preparation, the samples were subjected to nano-LC followed by Electrospray Ionization MS/MS in the Q Exactive Orbitrap Mass Spectrometer at the central Proteomics Facility, CCMB.

2.4.3. Proteomics data processing

Following mass spectrometry, the Total Ion Chromatogram (TIC) for each sample was retrieved and analysed using Thermo Xcalibur software. A proteomic scan was conducted in Proteome Discoverer (PD) version 2.2.0.388 at the Central Proteomics Facility, CCMB Hyderabad, to identify proteins corresponding to the ion chromatograms.

Uniprot Centipede all venom (reviewed and non-reviewed) and the custom-built venom gland transcriptomes of *Scolopendra*, assembled and annotated by Aditi and Pragyadeep Roy from the Evolutionary Ecology Lab, CCMB, served as the reference databases for the proteomic scans. These transcriptomes were translated *in-silico* using a minimum cutoff of 50 amino acids to generate the reference proteomes. After the PD searches were completed, the FASTA and Excel result files were retrieved for each sample and used for matrix generation and data analysis.

2.5. Data analysis

2.5.1 Minimum Convex Polygons (MCPs)

A Minimum Convex Polygon (MCP) is used in ecology to estimate the spatial range of species based on their occurrence points. When comparing two species, drawing MCPs for their geographical distributions helps assess their range overlap, habitat use, and potential interactions. We drew polygons covering the outermost geographically distributed samples from each clade to see if any geographical overlap existed and to infer whether the phylogenetic clusters formed with respect to *S. morsitans* species complex were congruent with their geographical distribution. The R package sf was used for the same.

2.5.2. Diversity Comparisons

We calculated the estimates of Hill numbers (q=0 and q=1) to compare the venom diversity across the clusters of *S. morsitans* species complex (Hill, 1973). The q=0 and q=1 estimates are diversity indices derived from Hill numbers, commonly used to assess biodiversity. When applied to venom composition, they provide insights into the richness and evenness of venom protein diversity across different groups. The q=0 (Species Richness / Protein Richness) metric represents the total number of unique venom proteins detected, treating all proteins equally regardless of their abundance. A higher q=0 value suggests greater venom diversity in terms of the number of distinct proteins present, without considering their relative abundances, *i.e* it gives an idea about the venom protein diversity with respect to richness. On the other hand, the q=1 estimate, also known as the Shannon diversity index in the context of Hill numbers, measures abundance-weighted diversity. q=0 counts the number of unique venom proteins (richness), while q=1 considers both the presence and the abundance of each protein. Thus, q=1 provides a better understanding of the proteome by also considering the evenness of each unique protein. Higher q=1 values indicate that venom proteins are more evenly distributed, meaning there isn't a strong dominance of a few proteins. Lower q=1 values suggest that venom composition is skewed, with a few proteins being extremely abundant while many others are rare.

2.5.3. Principal Component Analysis (PCA)

PCA or Principal Component Analysis reduces high-dimensional data while preserving key variance, making visualization and interpretation easier. It helps identify venom composition patterns, clusters, and outliers, revealing ecological and evolutionary influences. Additionally, PCA highlights the main features driving venom variation across different populations or clusters. The ordination method of PCA was used to visualise the venom data in a venom trait space. PCA considers the n-dimensional *i.e*. multivariate data, reducing it to a lower dimensional plot (usually a 2-D plot). Here, the two best axes of the 2-D plot account for explaining the maximum variation in the dataset and extracting the important information of which variables in particular are causing this variation.

We performed PCA on the PCN (protein copy number) matrix, which was obtained post-data processing, containing the abundance data of the 113 venom proteins

identified from the venoms of *S. morsitans* species complex. PCA also helped us to see if the clusters obtained after phylogenetic analysis separate in the venom trait space. Along with the visualisation, we also extracted information about which venom proteins are the most important for explaining the observed variation. The python library ‘scikit-learn’ was used to perform the analysis.

2.5.4. Non-Metric Multidimensional Scaling (nMDS)

Since the data obtained in the PCN matrix may not be normally distributed and the relation between variables might not be linear, we employed an ordination method like nMDS. nMDS takes a distance matrix as input and then reduces the dataset to fewer dimensions from multidimensional data while preserving the rank-order relationship among the target variables based on the input variables. The PCN matrix was used to obtain a Bray-Curtis dissimilarity matrix, which was then used as an input for the nMDS analysis. Unlike PCA, which considers only abundance values that might be sensitive to the outlier values, nMDS is more robust when dealing with such data. It clusters based on their protein composition (species composition) as it takes into account both the presence/absence and abundance of different proteins in different clusters. Thus, points in nMDS plot are closer if they have similar protein composition in the original matrix while farther points have dissimilar protein composition. The python libraries ‘scikit-learn’ and ‘scipy’ performed the nMDS and distance matrix analyses.

3. Results

3.1. Data generation

3.1.1. DNA and Proteomics work

DNA extractions were carried out for *Scolopendra morsitans* (n=52) and *Scolopendra hardwickei* (n=6). These samples were carefully chosen based on their distinct geographic areas and morphological characteristics to ensure a complete taxon sampling across peninsular India. The primary DNA dataset generated in this study consisted of 52 samples belonging to the species complex *S. morsitans* from 39 unique locations across peninsular India. Along with this, a secondary dataset of samples from India (n=42) from (Joshi and Karanth, 2011; Joshi and Edgecombe, 2018; Joshi *et al.*, 2020), South-East Asia (n=7) from (Siriwut *et al.*, 2015), and Africa (n=5) from ((Vahtera *et al.*, 2012); and the molecular data generated in the Evol-EcoLab for the specimens preserved at the American Museum of Natural History, New York, USA) was used for a complete global sampling and representation of *Scolopendra morsitans*.

The venom protein dataset included 33 samples from the *S. morsitans* species complex from peninsular India. Similarly to the DNA dataset, the samples were carefully chosen from unique geographic locations across peninsular India for a complete taxon sampling. The generation of the protein dataset was a combined effort by me and Aditi.

The following table shows the data for this study from the DNA and venom datasets for the different clades of *Scolopendra morsitans*.

Genus	**Species**	**No.of samples**	**DNA sequence**			**Venom LC-MS**
			COI	16s	28s	
Primary data						
Scolopendra	*morsitans 1*	46	31	38	38	28
Scolopendra	*morsitans 2*	6	6	6	6	5
Secondary data						
Scolopendra	*morsitans 1*	32	20	30	24	0
Scolopendra	*morsitans 2*	10	10	8	9	0
Scolopendra	*morsitans 1* (SE Asia)	7	7	7	0	0
Scolopendra	*morsitans* (Africa)	5	3	5	4	0

Table 2: Primary and secondary dataset for *S. morsitans*.

3.1.2. Venom Extraction

From June 2024 to December 2024, I actively contributed to the venom extractions of around 320 centipede samples belonging to the genus *Scolopendra* and other scolopendrid genera.

Genus	No. of Individuals (Venom Extracted)	Average protein concentration (mg/ml)
Scolopendra	32	5.93 mg/ml
Asanada	4	1.04 mg/ml
Cormocephalus	38	1.49 mg/ml
Rhysida	138	3.49 mg/ml

Ethmostigmus	13	3.90 mg/ml
Digitipes	57	0.79 mg/ml
Miscellaneous (unidentified)	40	-

Table 3: Venom extraction data

3.2. Phylogenetic analyses

3.2.1 Maximum Likelihood and Bayesian Inference Phylogeny

The maximum likelihood and the bayesian inference phylogeny from the 16s rDNA dataset, COI dataset, 16s-COI concatenated dataset, and 16s-28s-COI concatenated dataset did not recover monophyly of the *S. morsitans* as the Australian *Scolopendra laeta* (Hasse, 1887) and *Scolopendra leki* (Waldock and Edgecombe, 2012) were nested within Indian *S. morsitans*. However, we observed a high bootstrap support of 99% and a posterior probability of 0.6 for monophyly among the peninsular Indian and Southeast Asian samples. Interestingly, samples from Eastern Ghats (Odisha), western India (Gujarat, Maharashtra), and Himachal Pradesh formed a distinct cluster from the rest of the *S. morsitans* samples from peninsular India with strong support (bootstrap support of 86% and posterior probability of 1). Within the *Scolopendra morsitans*, the African *S. morsitans* formed one cluster (bootstrap support of 100% and posterior probability of 1), and were sister to the rest of *S. morsitans*, *S. laeta*, and *S. leki*. The subfamily level ML and BI tree of the combined dataset (16s-28s-COI) is available from in the supplementary materials.

3.2.2 Species Delimitation

We performed species delimitation analyses, mPTP, on the *S. morsitans* to delineate the distinct units. The results from mPTP on the 16s dataset indicated the presence of 6 distinct clades within the peninsular Indian *S. morsitans*, 1 distinct clade of *S. morsitans* from Odisha (Eastern Ghats), 1 distinct clade from individuals from

western India (Gujarat & Maharashtra), and 1 distinct clade of *S. morsitans* from Himachal Pradesh. However, for the COI dataset, mPTP results show a presence of 13 distinct clades within the peninsular Indian *S. morsitans,* and for the rest of the regions in India, the results were similar. The results from mPTP of combined 16s-COI dataset were similar to results from 16s dataset. We used posterior support of 1 to get a more conservative estimate for the distinct units within the *S. morsitans* species complex. We found 3 distinct units from peninsular India, referred to as *S. morsitans 1a*, *S. morsitans 1b*, and *S. morsitans 1c*; 2 distinct units from SE Asia referred to as *S. morsitans 1d*, and *S. morsitans 1e*. The clades from peninsular India and southeast Asia are together referred to as S. *morsitans 1*. The distinct clade from Odisha is referred to as *S. morsitans 2a*, the clade from western India as *S. morsitans 2b*, and the clade from Himachal Pradesh as *S. morsitans 2c*; these clades together are referred as S. morsitans 2.

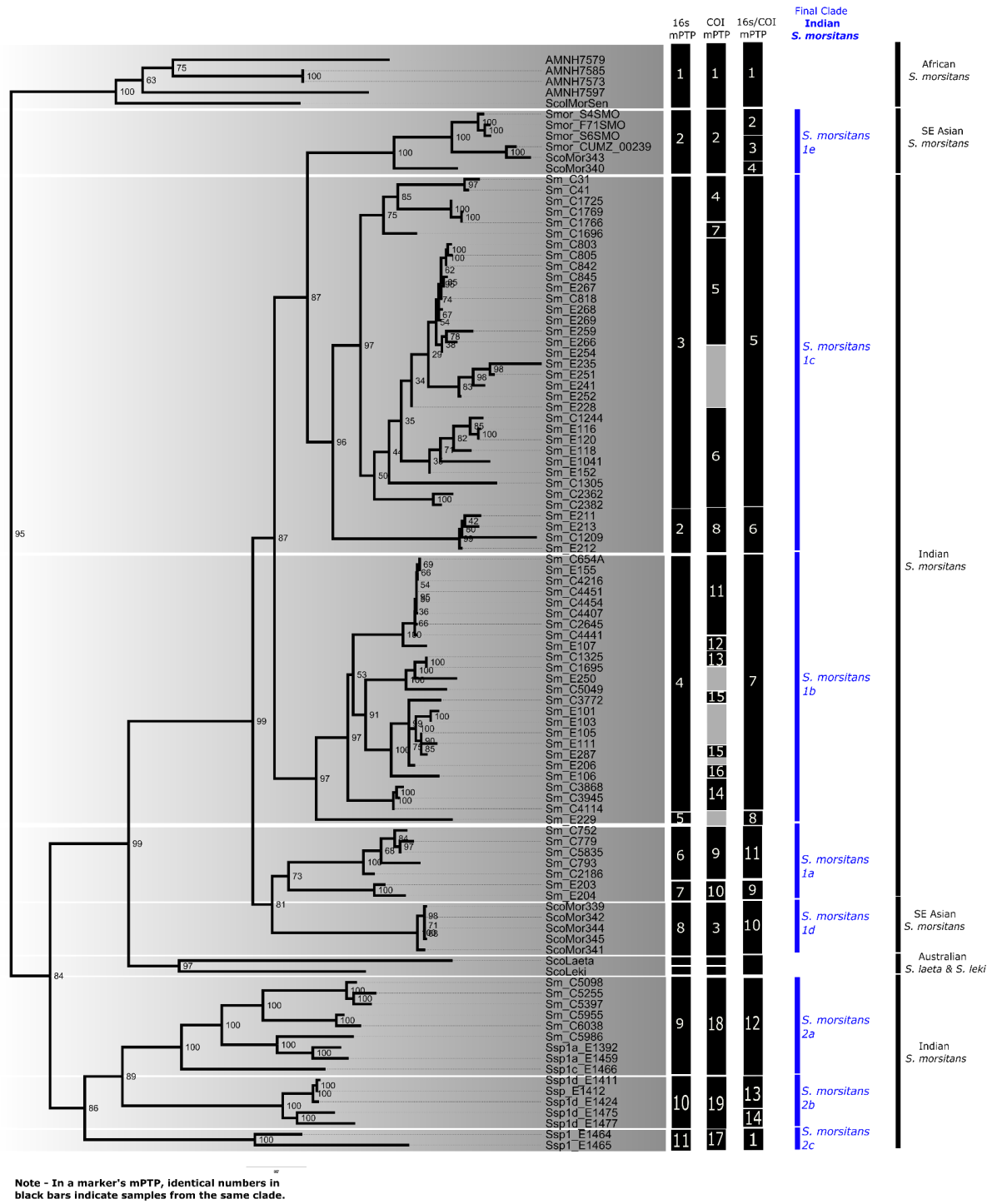


Fig 2: Maximum Likelihood tree of the global *S. morsitans* with mPTP results mapped onto the ML tree.

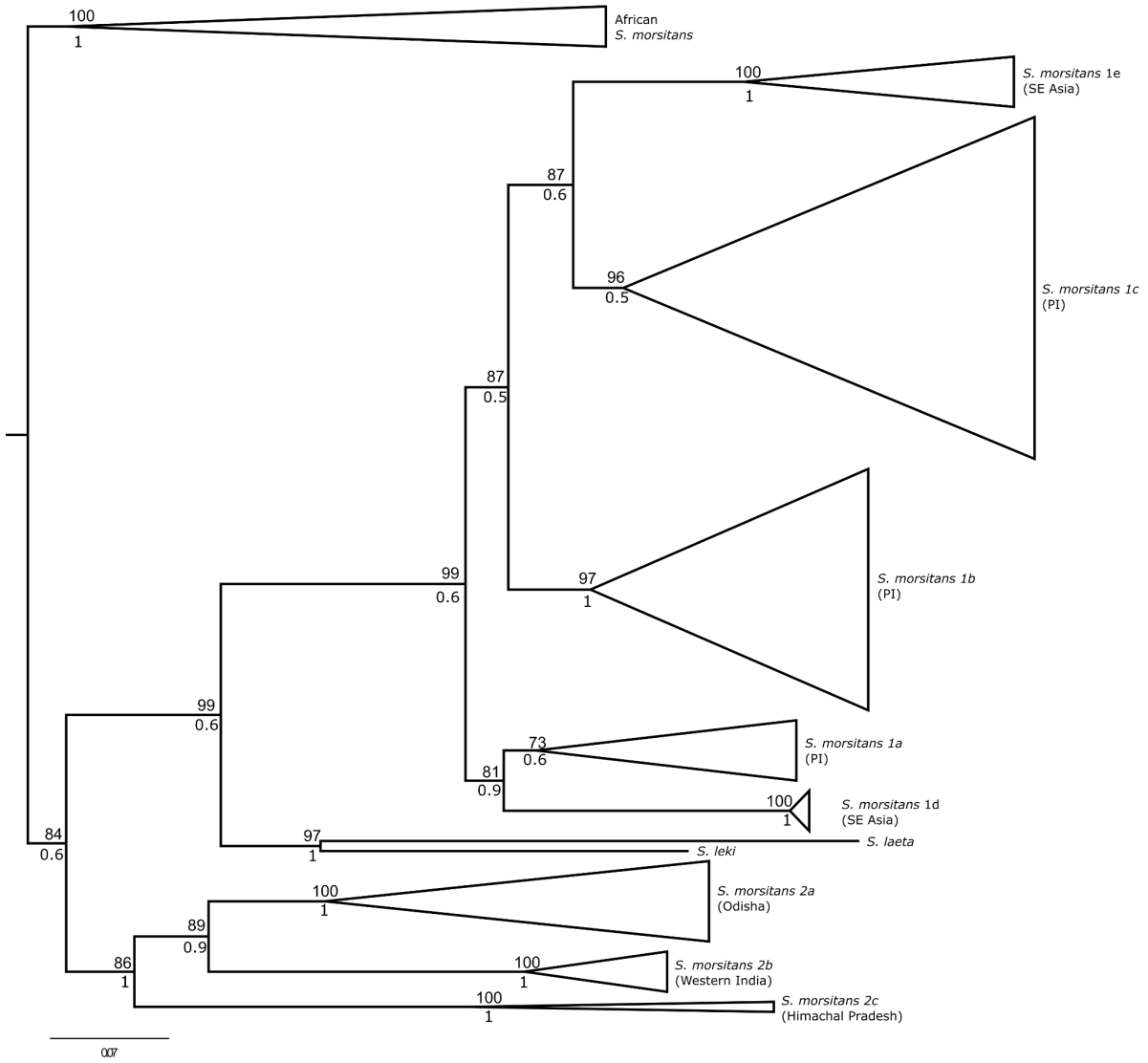


Fig 3: Phylogenetic tree of the distinct clades formed by global *S.morsitans.* The values above and below the nodes indicate the bootstrap support and posterior probability from the Maximum Likelihood and Bayesian Inference Tree respectively.

Further analysis were performed considering the *S. morsitans 1a*, *S. morsitans 1b*, and *S. morsitans 1c* from the peninsular Indian clade, and the *S. morsitans 2a* from the Odisha, as this clades had both venom and genetic dataset available. For easier interpretation *S. morsitans 2a* from Odisha is referred to as *S. morsitans 2*, while the *S. morsitans 1a*, *S. morsitans 1b*, and *S. morsitans* 1c are collectively referred to as *S. morsitans 1*.

To understand the molecular divergence between these clades, we have also calculated the pairwise genetic distance (COI marker) among them using the Tamura-Nei (1993) substitution model.

	S. morsitans (Africa)	*S. morsitans 1c*	*S. morsitans 1a*	*S. morsitans 1b*	*S. morsitans 2a*	*S. morsitans 2b*	*S. morsitans 2c*
S. morsitans (Africa)	0.141						
S. morsitans 1c	0.226	0.124					
S. morsitans 1a	0.238	0.174	0.104				
S. morsitans 1b	0.229	0.176	0.169	0.095			
S. morsitans 2a	0.228	0.246	0.246	0.241	0.143		
S. morsitans 2b	0.211	0.219	0.237	0.224	0.214	0.013	
S. morsitans 2c	0.229	0.256	0.259	0.276	0.247	0.211	0.079

Table 4: The average pairwise genetic distance (COI marker) between the different *S. morsitans* clades calculated using the Tamura-Nei (1993) substitution model. The diagonal values represent the within group average genetic distances.

3.3. Minimum Convex Polygons

The MCPs for *S. morsitans 1* and *S. morsitans 2* did not overlap, indicating allopatric distributions (i.e., the two groups are geographically separated).

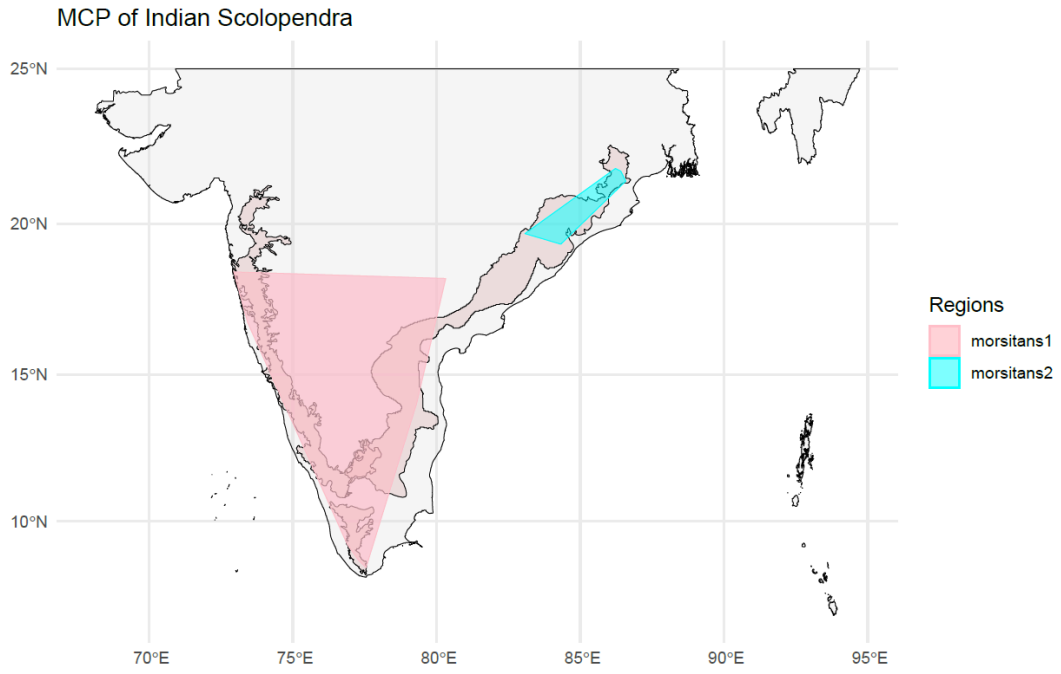


Fig 4: MCP of *S. morsitans 1* and *S. morsitans 2*

The three distinct units (*S. morsitans 1a*, *S. morsitans 1b*, and *S. morsitans 1c*) within the peninsular Indian *S. morsitans* clade also had non-overlapping MCPs.

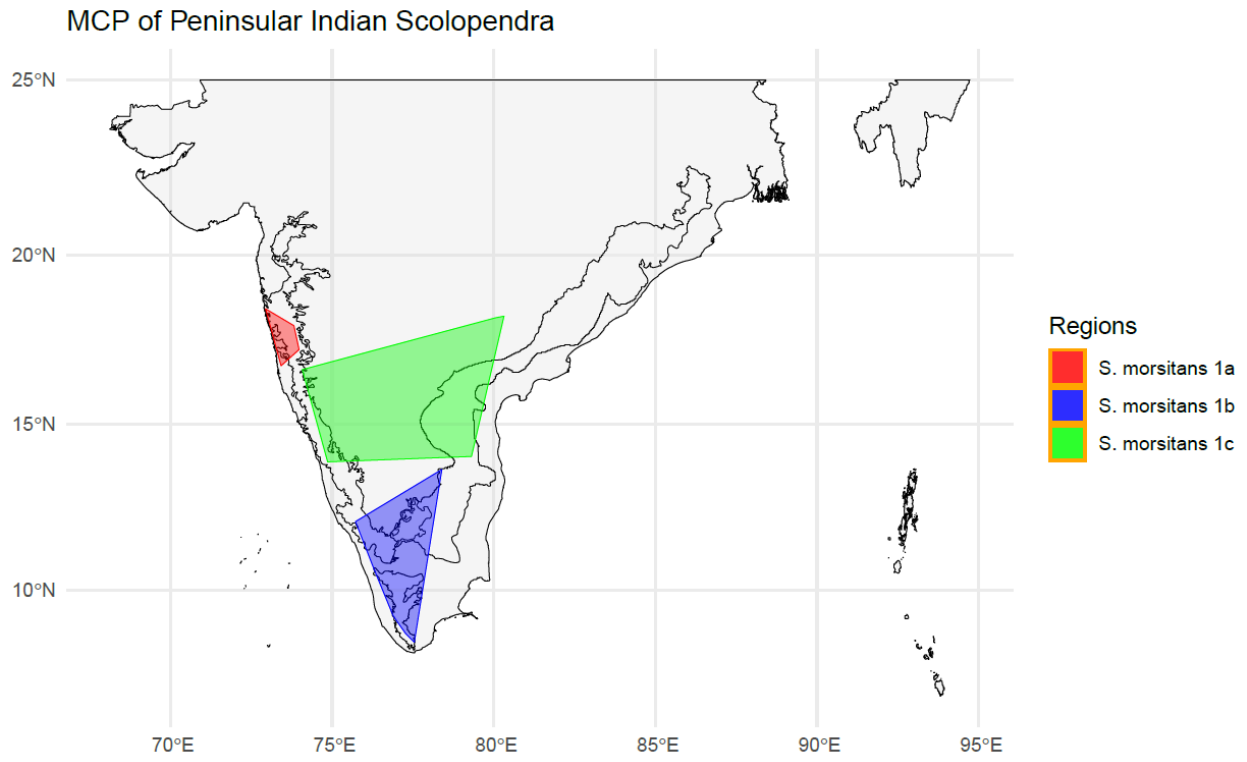


Fig 5: MCP of *S. morsitans 1a*, *S. morsitans 1b*, and *S. morsitans 1c*

3.4. Venom analyses

3.4.1 General Venom Composition

To gain insight into the overall venom composition of the peninsular Indian *Scolopendra morsitans* species complex, we grouped the 113 identified proteins into major groups. These included Enzymes, Scoloptoxins, Beta pore-forming toxins, Unchar (uncharacterized) proteins, LDLA domain-containing proteins, CAP proteins, proteins with domains of unknown function (DUF), as well as other toxin and non-toxin proteins. Using the two broad phylogenetic and distinct clusters obtained from the DNA data, we then characterised the venom profiles of these clusters by analysing the cumulative copy numbers (*i.e.*, abundance values) of the

corresponding proteins. The analysis was performed for the two clusters named *S. morsitans 1* and *S. morsitans 2*.

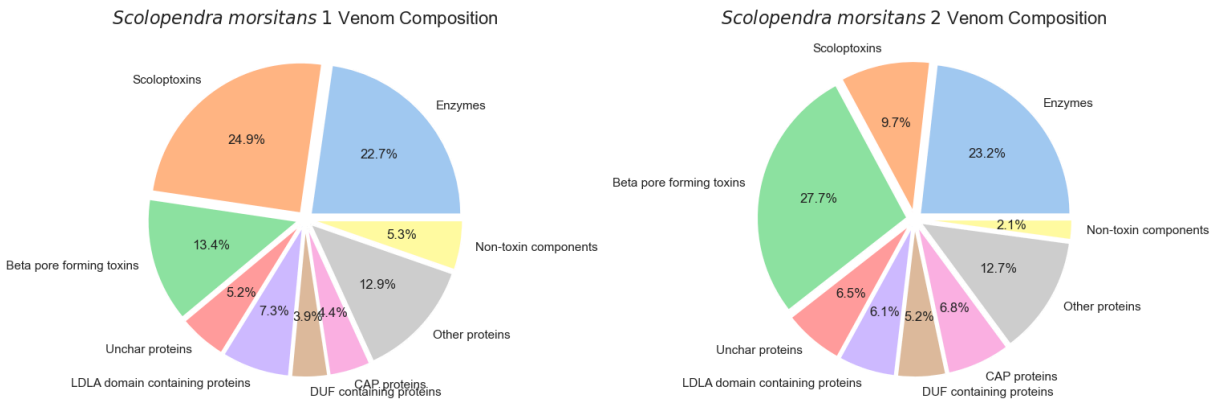


Fig 6: Venom Composition of *S. morsitans* 1 (n=28) and *S. morsitans* 2 (n=5)

Both clusters had a venom composition dominated by groups Enzymes, Scoloptoxins and Beta pore-forming toxins, although to variable degrees of dominance. The venom of *S. morsitans 1* was dominated by Scoloptoxins (24.9%), Enzymes (22.7%) and Beta pore-forming toxins (13.4%). In comparison, the composition of *S. morsitans 2* was dominated by Beta pore-forming toxins (27.7%), enzymes (23.2%), and Scoloptoxins (9.7%). Across all the 113 identified proteins, 40 toxin proteins and 16 non-toxins were unique to *S. morsitans* 1. In comparison, only 3 toxin proteins were unique to *S. morsitans* 2, while the other 54 proteins were found in both the clades.

Proteins	**Presence in *S. morsitans* 1 (n=28)**	**Presence in *S. morsitans* 2 (n=5)**
Acid phosphatase (Enzyme)	1	Absent
Apolipoprotein D (Others)	3	Absent
ARYLSULFATASE (Enzyme)	5	Absent
CAA-like (Others)	1	Absent
Calycin-Lipocalin (Others)	7	Absent
Carboxypeptidase (Enzyme)	Absent	1
Cyclase-like protein (Enzyme)	2	Absent

DUF1397	2	Absent
Fibronectin or immunoglobulin-like (Others)	4	Absent
GH18 (Enzyme)	5	Absent
Haem-binding protein (Others)	10	Absent
Hyaluronidase (Enzyme)	Absent	1
Icarapin-like (Others)	13	Absent
Ige-ESP-like (Others)	7	Absent
Kazal domain (Others)	1	Absent
LysozymeC (Enzyme)	27	Absent
M12A (Enzyme)	1	Absent
Melanotransferrin 2 (Others)	Absent	1
Purple acid phosphatase (Enzyme)	1	Absent
Saposin-related (Others)	6	Absent
SCTX01 (Others)	1	Absent
SCTX02 (Others)	3	Absent
Serpine (Others)	16	Absent
SLPTX02 (Scoloptoxin)	5	Absent
SLPTX04 (Scoloptoxin)	24	Absent
SLPTX05 (Scoloptoxin)	3	Absent
SLPTX06 (Scoloptoxin)	8	Absent
SLPTX07 (Scoloptoxin)	14	Absent
SLPTX08 (Scoloptoxin)	3	Absent
SLPTX09 (Scoloptoxin)	7	Absent
SLPTX13 (Scoloptoxin)	11	Absent
SLPTX19 (Scoloptoxin)	23	Absent
SLPTX21 (Scoloptoxin)	14	Absent
SLPTX29 (Scoloptoxin)	2	Absent
SLPTX30 (Scoloptoxin)	13	Absent
SLPTX47 (Scoloptoxin)	1	Absent

TGFbeta (Others)	14	Absent
Unchar30 (Uncharacterized)	7	Absent
Unchar31 (Uncharacterized)	3	Absent
Unchar32 (Uncharacterized)	5	Absent
Unchar33 (Uncharacterized)	3	Absent
Vigilin (Others)	1	Absent
WAP (Others)	5	Absent

Table 5: List of unique toxin proteins in each of the two clusters. The word inside the brackets refer to the broad toxin category used to characterise the venom proteome.

3.4.2 Venom Diversity comparisons

In the case of the diversity estimates, the *S. morsitans 1* (n=28) had a higher average protein richness(q=0) value of 49.5 than the members of *S. morsitans* 2 (n=5), which had a protein richness(q=0) of 30.8, indicating greater venom protein diversity in the former. This pattern was also seen to be consistent with the Shannon's index (q=1) estimate. The average Shannon's index (q=1) value for *S. morsitans 1* was 29.26, greater than that of *S. morsitans* 2, which was 15.76, indicating higher evenness and functional diversity of venom proteins in the former group than the latter. The average protein abundance in *S. morsitans 2,* which was 300, was greater than that of *S. morsitans 1,* which was 183.36, indicating that the venom composition of the former group was influenced by a fewer number of proteins that are not very diverse, albeit harbouring high copy numbers.

Species	Average Protein Richness (q=0)	Average Shannon's Index (q=1)
S. morsitans 1 (n=28)	49.5	29.26
S. morsitans 2 (n=5)	30.8	15.76

Table 6: Average values of protein richness and Shannon's index of the two putative species

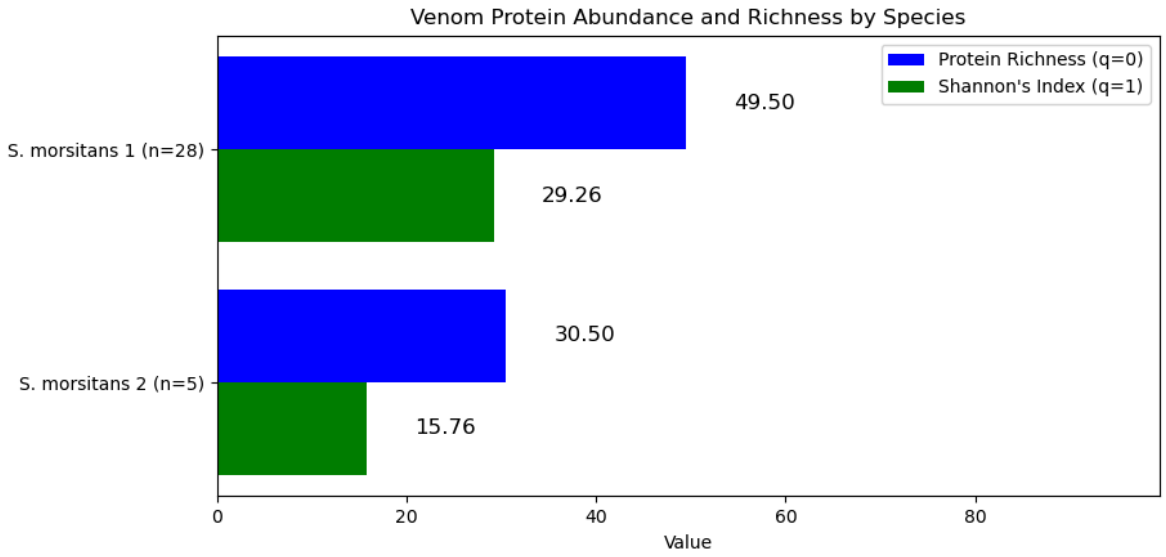


Fig 7: Bar plot depicting the average protein richness (q=0) and Shannon's index (q=1) of the two putative species.

3.4.3 Principal Component Analysis (PCA)

We performed the PCA analysis on two different datasets: one on the venom dataset of all the *S. morsitans* centipedes found across the Indian geography, *i.e.*, *S. morsitans 1* and *S. morsitans 2*. The other PCA was done on the venom dataset of the *S. morsitans* found within peninsular India, *i.e.*, *S. morsitans 1a*, *S. morsitans 1b*, and *S. morsitans 1c*.

In the first dataset, PC1 explained 71.84% of the variation present, with the variations explained by components β-PFTx, CAP, GGT, and DUF3472. PC2 explained 7.15% of the variation in the dataset, the major contributors being CUB-domain_protein metalloproteases, SLPTX15, SLPTX11, and LDLA. The venom trait space for the *S. morsitans 1* and *S. morsitans 2* showed a clear separation.

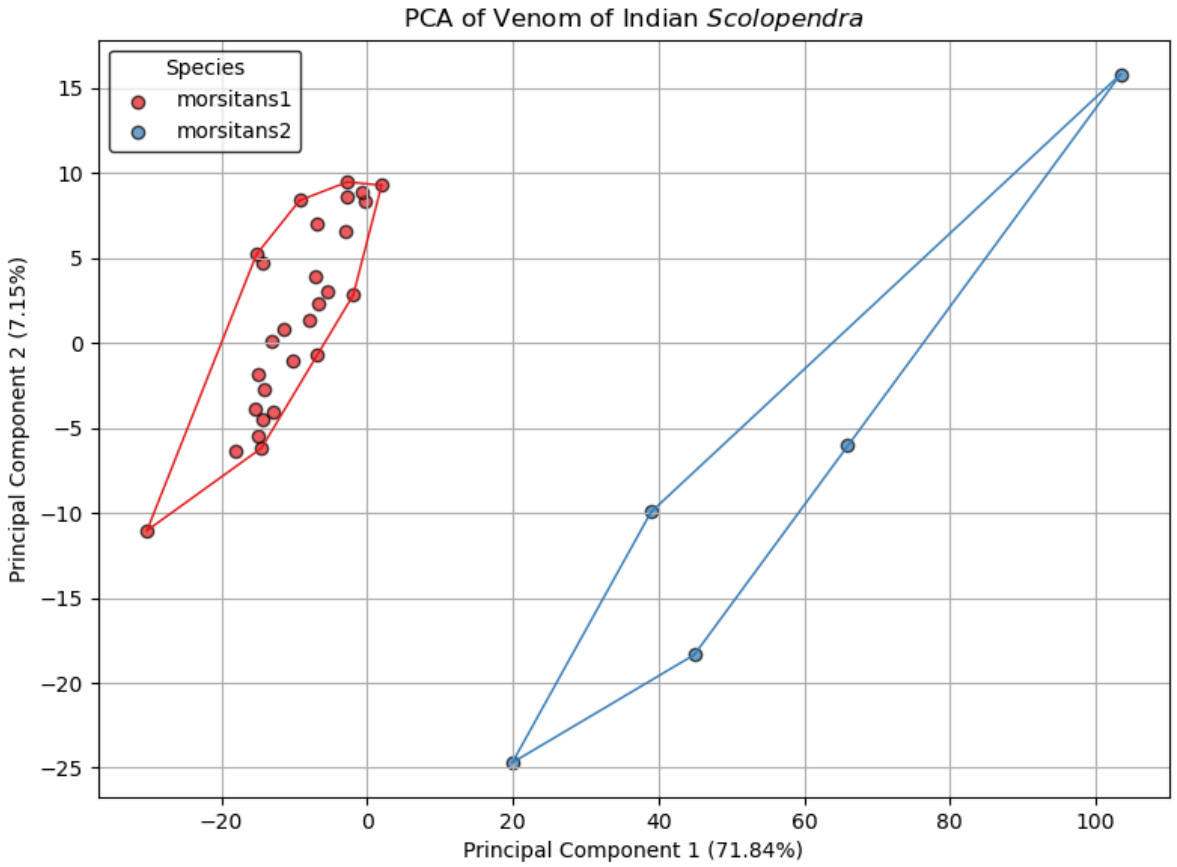


Fig 8: PCA of venom components of *S.morsitans 1* and *S. morsitans 2*

The second dataset consisted of *S. morsitans* 1a, *S. morsitans* 1b, and *S. morsitans 1c*, corresponding to the three phylogenetic clusters/units seen in peninsular India. The clusters formed by the three peninsular Indian clades showed an overlap in the venom trait space. PC1 explained 38.10% of the variation, with the major contributors being β-PFTx, CUB-domain containing metalloproteases, LDLA domain containing protein, SLPTX15 and SLPTX11. PC2 explained 15.39% of total variation, with the highest contributions from SLPTX11, Alpha-2-macroglobulin and β-PFTx.

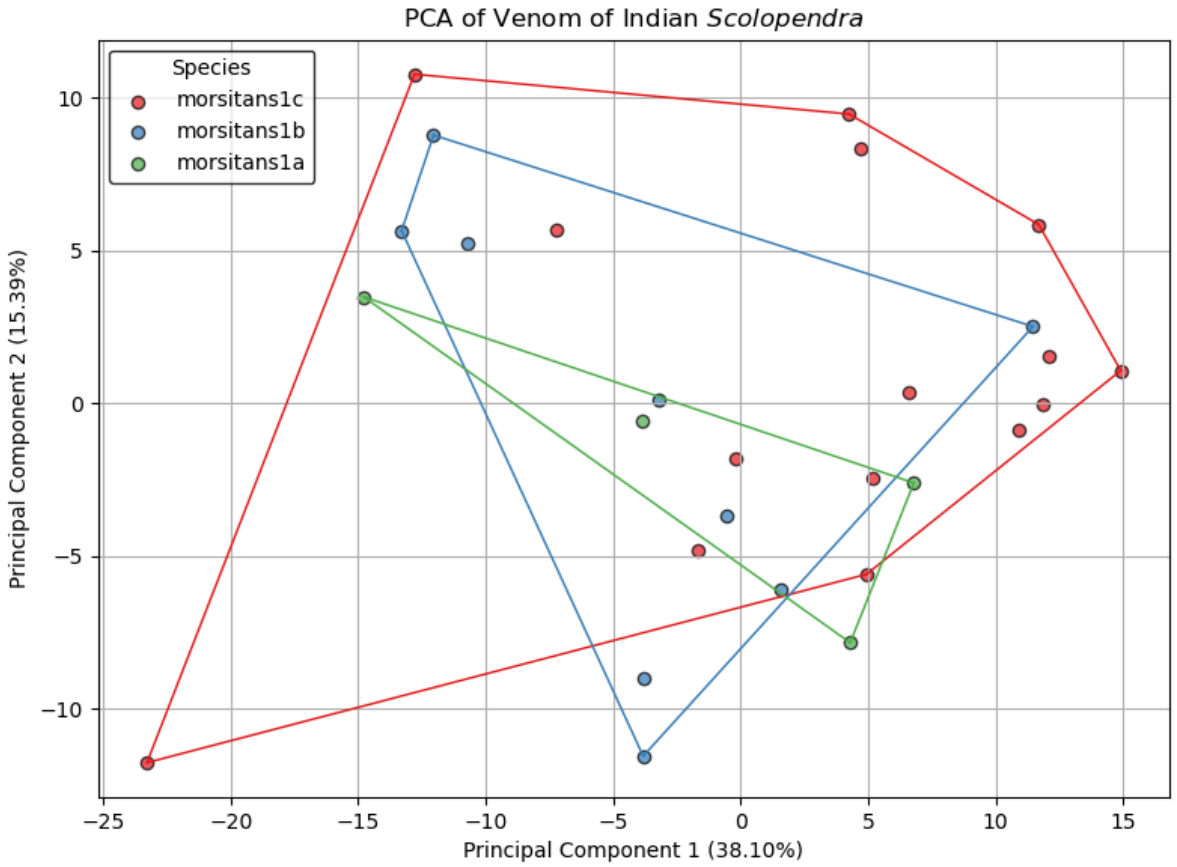


Fig 9: PCA of venom components of *S. morsitans 1a*, *S. morsitans 1b*, and *S. morsitans 1c*

3.4.4 Non-Metric Multidimensional Scaling (nMDS) Analysis

Similar to the PCA analyses, nMDS was performed on the Protein Copy Number matrices for both the datasets. To perform nMDS, a dissimilarity matrix was calculated based on the Bray-Curtis dissimilarity index and then used as the input data. Contrary to PCA, where distances between the targets based on their variables matter, nMDS only maintains the ranks based on the distances between them. Thus, nMDS gives information that is different from PCA. Unlike PCA, nMDS considers both the richness and abundance of the proteins in the venom and clusters the targets based on that data.

In both cases, the analysis of similarities (ANOSIM) test was performed to assess the significance of the clusters' separation. The test directly uses the ranked dissimilarity matrix obtained after calculating the Bray Curtis dissimilarity.

The analysis revealed a high degree of dissimilarity between the two groups in the first dataset comparing *S. morsitans 1* and *S. morsitans 2*.

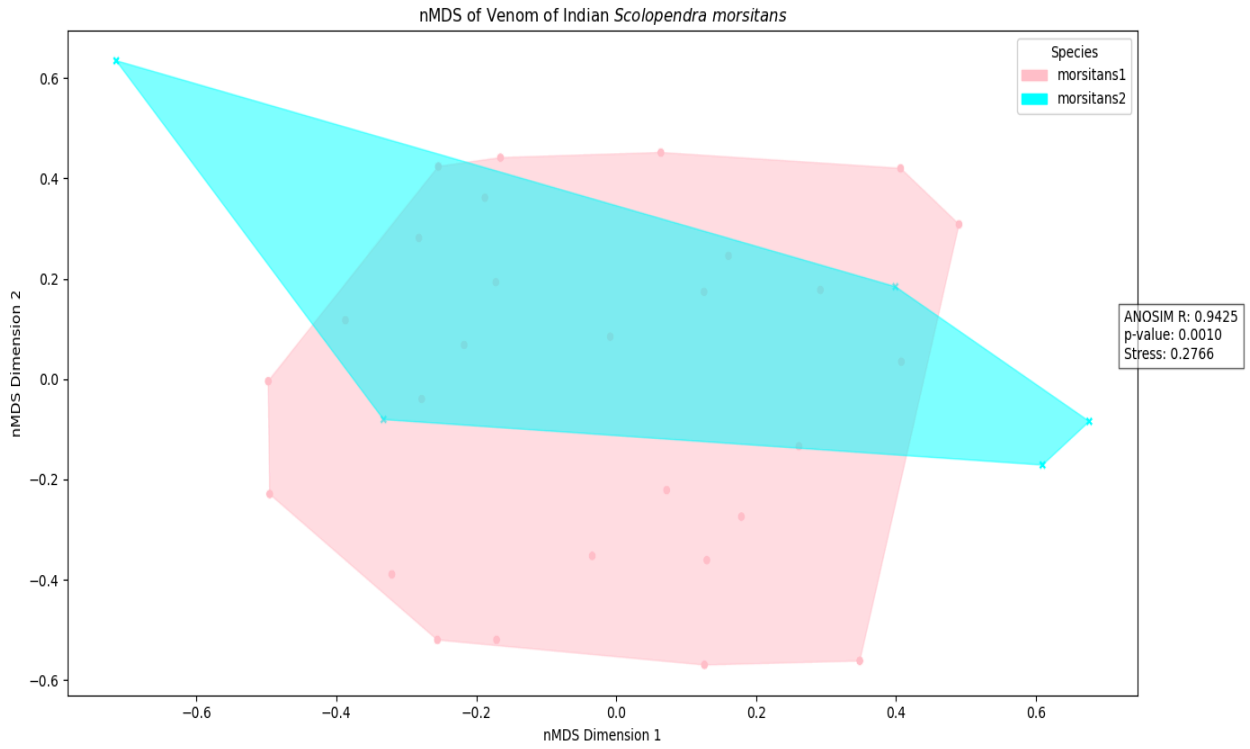


Fig 10: nMDS of venom components of *S. morsitans 1* and *S. morsitans 2*

While there was some overlap observed in the ordination space, the distinction between the groups was supported by the results of the ANOSIM test. The ANOSIM-R value was 0.9425, which indicates a high level of dissimilarity between the two groups relative to within-group variation, and the associated p-value was 0.0010. This meant we could reject the null hypothesis that there was no difference between *S. morsitans 1* and *S. morsitans 2*. This suggested that the venom profiles of these two groups were significantly different, despite some degree of overlap observed in the visual space.

The second dataset compares the within-group differences between the peninsular Indian clade of *S. morsitans 1a*, *S. morsitans 1b*, and *S. morsitans 1c*. There was a substantial degree of overlap between all three groups.

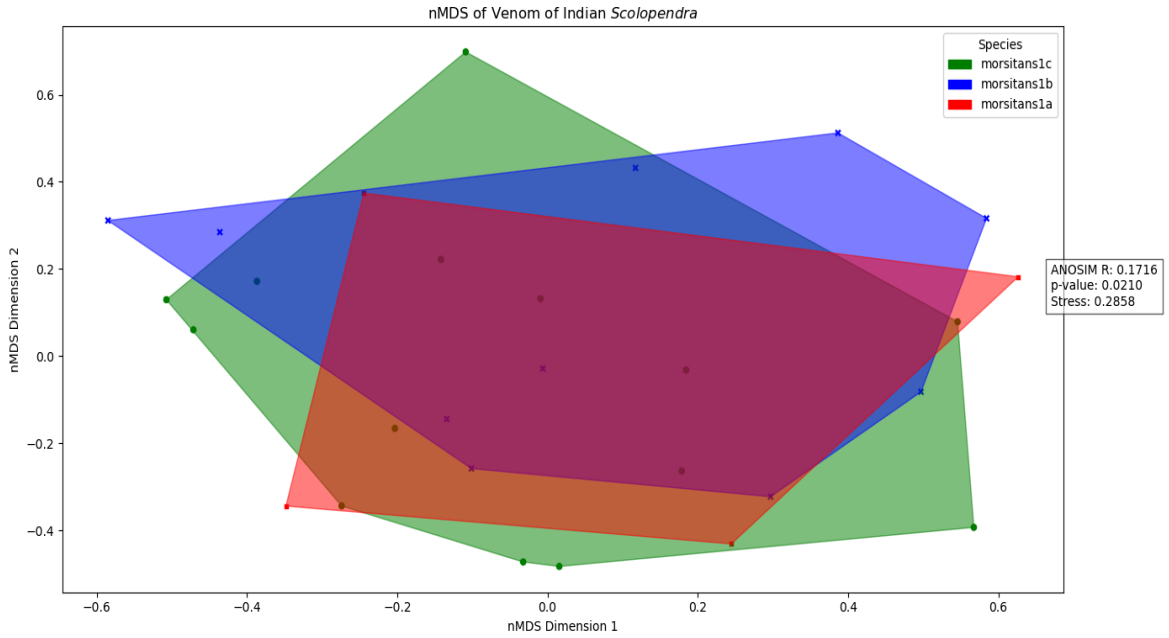


Fig 11: nMDS of venom components of *S.morsitans 1a*, *S. morsitans 1b*, and *S. morsitans 1c*

For the second dataset of *S. morsitans* 1a, *S. morsitans* 1b, and *S. morsitans* 1c, the R-value was 0.1716, with a significance value of 0.0210. In this case, a low R-value suggests that there was a low level of dissimilarities in terms of venom composition between the clusters.

4. Discussion

We have employed multiple lines of evidence using an integrative taxonomic framework for building a robust species hypothesis for the *S. morsitans* species complex.

The results from the molecular phylogenetic approach using the mitochondrial and nuclear dataset showed *Scolopendra morsitans* as a species complex, species with limited variations among their morphological traits. The species complex showed varying levels of molecular divergence as seen from the average pairwise genetic distance between the different clades. The phylogenetic trees reconstructed using the methods of maximum likelihood and bayesian inference show congruence with each other. The African *S. morsitans* were the most divergent from the rest of the *S. morsitans* in the study. We found that the Southeast Asian and peninsular Indian *S. morsitans* clustered together, however, not all the Indian samples identified as *S. morsitans* fall within this cluster, as the *S. morsitans* from Odisha (Eastern Ghats), western India (Gujarat and Maharashtra), and Himachal Pradesh formed a separate cluster. These results agree with previous research that *S. moristans* could be a species complex with multiple cryptic/sibling species under the same name (Lewis, 2010; Joshi and Karanth, 2011; Siriwut *et al.*, 2016).

The *S. morsitans* was not monophyletic as the Australian centipedes *S. laeta* and *S. leki* were nested within the *S. morsitans*, making the global *S. morsitans* to be paraphyletic. Both *S. laeta* and *S. leki* were represented by only one sample, and additional samples from Australia could confirm these results.

Thus, based on our molecular phylogenetic analyses, we divided *Scolopendra morsitans* into two unique clades, one clade belonging to the peninsular India and southeast Asia named *S. morsitans 1* while the other clade belonging to the *Scolopendra morsitans* from Odisha, western India, and Himachal Pradesh named *S. morsitans* 2. Within the peninsular Indian clade, the species delimitation results have shown a varying number of putative species, however to prevent any overestimation because of oversplitting, we have chosen a conservative number *i.e.*, three species among the peninsular Indian clade, referred as *S. morsitans 1a*, *S.*

morsitans 1b, and *S. morsitans 1c*. All the further analyses were conducted with these clades in consideration.

Giving more confidence to results, the MCPs also show a similar trend. Between the peninsular Indian and Odisha clades of *S. morsitans*, we see a clear separation of the geographical distribution of these two clades, showing allopatric distribution. These results seem to provide more evidence that geography might act as an important factor in determining the genetic structure of these dispersal-limited ancient organisms. This might suggest that *S. morsitans 1* and *S. morsitans 2* may be ecologically or evolutionarily distinct, potentially undergoing divergent adaptation to their respective environments. In the peninsular Indian clade, we see that the clade *S. morsitans 1a* (samples from Maharashtra), which consistently showed up as a separate group when performing phylogenetic analysis, also seems to be geographically isolated from the other two clades. And the other two clades, *S. morsitans 1b* (samples from Kerala, Tamil Nadu, and Telangana) and *S. morsitans 1c* (samples from Karnataka, Telangana, and Andhra Pradesh) also seem to show a clear geographical separation. Thus, even within peninsular India, we see a congruence of geography with the phylogeny. This could be further studied using Species Distribution Models (SDMs), as they might be much more suitable to determine the realised niche of the species, taking into account both the geography and climatic conditions (Elith and Leathwick, 2009). Thus, if the separation of our clusters results from different geographical boundaries having differing environments, SDMs would reflect as such and would be much more suitable for making such inferences.

The comparisons of venom between the two major clades has shown some stark distinctions between them, while *S. morsitans 1* shows a higher richness of protein diversity with an even distribution of proteins, *S. morsitans 2* shows less protein diversity but higher abundances for the few proteins present in the venom, this is consistent with the numbers of unique toxin proteins seen in *S. morsitans* 1 which is 40. In contrast, only 3 unique toxin proteins are found in *S. morsitans 2*. The results from PCA also showed that there was a difference in the venom composition at the abundance level of the cytotoxin (βPFTx) and the neurotoxins (SLPTX11 and SLPTX15), as seen from the loading of PC1 and PC2, respectively. The results from

ANOSIM-R of nMDS have a value of 0.9425 with a p-value of 0.0010. Within the peninsular India, the clades of *S. morsitans 1a*, *S. morsitans 1b*, and *S. morsitans 1c*, we see mostly an overlap of their venom composition from the results of PCA as well as nMDS (R-value of 0.1716 and p-value of 0.0210).

The greater Shannon's index (q=1) value of 29.26 in *S. morsitans 1* shows that the venom composition is more balanced, which means multiple proteins contribute significantly rather than just a few dominating ones. This aligns with the higher protein richness (q=0) value of 49.5, suggesting that *S. morsitans 1* has more unique venom proteins and maintains a relatively even distribution across them. On the other hand, the lower Shannon's index (q=1) value of 15.76 in *S. morsitans* 2 suggests that a few proteins dominate while others are present in much lower abundances. This implies a more specialised venom profile, where only certain proteins are functionally significant in predation or defense. Combined with its lower protein richness (q=0) value of 30.8, *S. morsitans 2* likely has a narrower and less even venom composition. This pattern could also be understood by the average abundances of total venom proteins across the two clusters.

This suggests that *S. morsitans 1* might have a more balanced venom composition indicating a generalist approach towards their prey. While *S. morsitans 2*'s venom might indicate specialization, where certain key toxins dominate envenomation. Further functional assays and characterisation studies of these important toxins might help in shedding more light on this. Since *S. morsitans 1* is found across a wider geographic range as shown from the MCP, the distinct units within *S. morsitans 1* might have to adapt to the varying environments across the wider peninsular Indian geography, however, *S. morsitans 2* being restricted to a small geographic area (Odisha), might have adapted to a narrow environment (Gunnell, 1997). Following this, it would be interesting to compare the venom profiles *Scolopendra morsitans* from a different geographic area which inhabits a similar environment to *S. morsitans 2* to test if environment plays a role, something which is seen in snake venom (Smith *et al.*, 2023). For this, one can use the bioclimatic variables and the venom phenotype data for an environmental niche modelling, where the model shall predict the presence of a particular venom phenotype for an environment across a geography, and this can be compared with the actual

centipede venom phenotypes found in those geographies. This might give a clearer idea of the importance of any bioclimatic variable in contributing to the differences in the venom phenotype (Smith *et al.*, 2023). Similarly, the role of diet should also be assessed as a factor for variation in venom. The more complex venom of *S. morsitans 1* might suggest it having a phylogenetically diverse diet as shown by studies, while the less complex venom of *S. morsitans 2* suggests a more specialized diet (Phuong *et al.*, 2016; Davies and Arbuckle, 2019).

In future, we would like to perform species delimitation on the resulting tree of the multi-locus dataset using BPP, a validation method (Yang, 2015). Following this, to understand the effect of many drivers known to cause venom variation in venomous animals like snakes, cone snails, *etc*. (Chang *et al.*, 2015; Casewell *et al.*, 2020; Tasima *et al.*, 2022), we plan to perform regression analysis, to determine the effect of various factors like temperature, precipitation, geographic distance, geographic location, elevation on the changes in the venom phenotype across the distinct species.

Our study has shed light on the different axes of species divergence, *i.e*., phylogenetic and ecological (venom) axes. By showing that the genetically distinct centipedes, which have not diverged on their morphological axes, have rather diverged on their ecological axis, we have shown that venom can act as an important trait for major evolutionary change like speciation in animal taxa. Accompanying this, we have also gained knowledge about the composition of centipede venom at the species level of *Scolopendra morsitans* and brought to light the putative species that exist within such a species complex.

Conclusion

In this project, we have delimited the putative species within the *Scolopendra morsitans* species complex. Firstly, we found that the *Scolopendra morsitans* species complex is paraphyletic. Then we found that there are different numbers of putative species across its geography which includes one species from Africa, two species from Southeast Asia, and six species from India. The species from India were divided into two clades referred to as *S. morsitans* 1 and *S. morsitans* 2. To provide more evidence to the Indian putative species, we used geographical location data which showed a mostly allopatric distribution for these species across India. Along with this, we assessed the venoms of centipedes as an axis to understand differences between species to provide more evidence to the species hypothesis. We found a significant difference in the venom composition between the clades *S. morsitans* 1 and *S. morsitans* 2. Within the clade *S. morsitans* 1 which contains phylogenetically and geographically distinct the Indian *S. morsitans 1a* , *S. morsitans 1b*, and *S. morsitans 1c* had similar venom profiles. Thus, incorporating evidence from more axes could further help in differentiating these putative species.

Supplementary materials

- Thesis_Data_Supplementary.xlsx
- Thesis_Phylogenetic trees_Supplementary.pdf

References

1. Albert, AYK, and Schluter, D (2005). Selection and the origin of species. Curr Biol 15, R283-288.
2. Arbuckle, K (2020). From molecules to macroevolution: Venom as a model system for evolutionary biology across levels of life. Toxicon: X 6, 100034.
3. Arbuckle, K (2021). Special Issue: Evolutionary Ecology of Venom. Toxins (Basel) 13, 310.
4. Barraclough, TG (2019). What causes speciation? In: The Evolutionary Biology of Species, ed. TG Barraclough, Oxford University Press, 0.
5. Barraclough, TG, and Nee, S (2001). Phylogenetics and speciation. Trends in Ecology & Evolution 16, 391–399.
6. Bharti, DK, Pawar, PY, Edgecombe, GD, and Joshi, J (2023). Genetic diversity varies with species traits and latitude in predatory soil arthropods (Myriapoda: Chilopoda). Glob Ecol Biogeogr 32, 1508–1521.
7. Budowle, B, Wilson, MR, DiZinno, JA, Stauffer, C, Fasano, MA, Holland, MM, and Monson, KL (1999). Mitochondrial DNA regions HVI and HVII population data. Forensic Science International 103, 23–35.
8. Casewell, NR, Jackson, TNW, Laustsen, AH, and Sunagar, K (2020). Causes and Consequences of Snake Venom Variation. Trends Pharmacol Sci 41, 570–581.
9. Chang, D, Olenzek, AM, and Duda, TF (2015). Effects of geographical heterogeneity in species interactions on the evolution of venom genes. Proceedings of the Royal Society B: Biological Sciences 282, 20141984.
10. Chitty, JR (2022). Myriapods (centipedes and Millipedes). In: Invertebrate Medicine, John Wiley & Sons, Ltd, 399–412.
11. Ct, D, and Jm, C (2017). Cryptic diversity in a fig wasp community-morphologically differentiated species are sympatric but cryptic species exhibit competitive exclusion. Molecular Ecology 26.
12. Davies, E-L, and Arbuckle, K (2019). Coevolution of Snake Venom Toxic Activities and Diet: Evidence that Ecological Generalism Favours Toxicological Diversity. Toxins (Basel) 11, 711.
13. Dawson, SK, Carmona, CP, González-Suárez, M, Jönsson, M, Chichorro, F, Mallen-Cooper, M, Melero, Y, Moor, H, Simaika, JP, and Duthie, AB (2021).

The traits of “trait ecologists”: An analysis of the use of trait and functional trait terminology. Ecology and Evolution 11, 16434–16445.

14. De Queiroz, K (2007). Species concepts and species delimitation. Syst Biol 56, 879–886.
15. Dobzhansky, T (1956). What is an Adaptive Trait? The American Naturalist 90, 337–347.
16. Dool, SE, Picker, MD, and Eberhard, MJB (2022). Limited dispersal and local adaptation promote allopatric speciation in a biodiversity hotspot. Molecular Ecology 31, 279–295.
17. Edgecombe, GD, and Giribet, G (2007). Evolutionary Biology of Centipedes (Myriapoda: Chilopoda). Annual Review of Entomology 52, 151–170.
18. Elith, J, and Leathwick, JR (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics 40, 677–697.
19. Gren, EC, Kelln, W, Person, C, McCabe, JG, Kornhauser, R, Hart, AJ, Erbas-White, K, Pompe, LR, and Hayes, WK (2017). Geographic variation of venom composition and neurotoxicity in the rattlesnakes Crotalus oreganus and C. helleri: assessing the potential roles of selection and neutral evolutionary processes in shaping venom variation. The Biology of Rattlesnakes II, 228–252.
20. Gunnell, Y (1997). Relief and climate in South Asia: the influence of the western ghats on the current climate pattern of peninsular India. International Journal of Climatology 17, 1169–1182.
21. Hill, MO (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. Ecology 54, 427–432.
22. Himaya, SWA, Arkhipov, A, Yum, WY, and Lewis, RJ (2022). Comparative Venomics of C. flavidus and C. frigidus and Closely Related Vermivorous Cone Snails. Marine Drugs 20, 209.
23. Hoang, DT, Chernomor, O, von Haeseler, A, Minh, BQ, and Vinh, LS (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35, 518–522.
24. Holford, M, Daly, M, King, GF, and Norton, RS (2018). Venoms to the rescue. Science 361, 842–844.
25. Huelsenbeck, JP, Ronquist, F, Nielsen, R, and Bollback, JP (2001). Bayesian

inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

26. Jangi, BS, and Dass, CMS (1984). Scolopendridae of the Deccan. Journal of Scientific and Industrial Research 43, 27–54.
27. Jenner, RA, von Reumont, BM, Campbell, LI, and Undheim, EAB (2019). Parallel Evolution of Complex Centipede Venoms Revealed by Comparative Proteotranscriptomic Analyses. Molecular Biology and Evolution 36, 2748–2763.
28. Joshi, J, and Edgecombe, GD (2018). Molecular phylogeny and systematics of the centipede genus Ethmostigmus Pocock (Chilopoda : Scolopendromorpha) from peninsular India. Invert Systematics 32, 1316.
29. Joshi, J, and Karanth, KP (2011). Cretaceous–Tertiary diversification among select Scolopendrid centipedes of South India. Molecular Phylogenetics and Evolution 60, 287–294.
30. Joshi, J, and Karanth, KP (2012). Coalescent Method in Conjunction with Niche Modeling Reveals Cryptic Diversity among Centipedes in the Western Ghats of South India. PLOS ONE 7, e42225.
31. Joshi, J, Karanth, PK, and Edgecombe, GD (2020). The out-of-India hypothesis: evidence from an ancient centipede genus, Rhysida (Chilopoda: Scolopendromorpha) from the Oriental Region, and systematics of Indian species. Zoological Journal of the Linnean Society 189, 828–861.
32. Kalyaanamoorthy, S, Minh, BQ, Wong, TKF, von Haeseler, A, and Jermiin, LS (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14, 587–589.
33. Kapli, P, Lutteropp, S, Zhang, J, Kobert, K, Pavlidis, P, Stamatakis, A, and Flouri, T (2017). Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. Bioinformatics 33, 1630–1638.
34. Kumar, S, Stecher, G, Suleski, M, Sanderford, M, Sharma, S, and Tamura, K (2024). MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for Adaptive and Green Computing. Molecular Biology and Evolution 41, msae263.
35. Lewis, JGE (2010). A key and annotated list of the <I>Scolopendra</I> species of the Old World with a reappraisal of <I>Arthrorhabdus</I>

(Chilopoda: Scolopendromorpha: Scolopendridae). Int J Myriapodology 3, 83–122.

36. Lewis, JGE (John GE (1981). The biology of centipedes, Cambridge [Eng.] ; New York : Cambridge University Press.
37. Minh, BQ, Schmidt, HA, Chernomor, O, Schrempf, D, Woodhams, MD, von Haeseler, A, and Lanfear, R (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution 37, 1530–1534.
38. Molinari, J, Eliecer, É, Gutiérrez, Antonio, A, Ascencao, D, Jafet, M, Nas, *et al.* (2005). Predation by Giant Centipedes , Scolopendra gigantea , on Three Species of Bats in a Venezuelan Cave.
39. Mullis, K, Faloona, F, Scharf, S, Saiki, R, Horn, G, and Erlich, H (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol 51 Pt 1, 263–273.
40. Pfenninger, M, Cordellier, M, and Streit, B (2006). Comparing the efficacy of morphologic and DNA-based taxonomy in the freshwater gastropod genus Radix (Basommatophora, Pulmonata). BMC Evolutionary Biology 6, 100.
41. Phuong, MA, Mahardika, GN, and Alfaro, ME (2016). Dietary breadth is positively correlated with venom complexity in cone snails. BMC Genomics 17, 401.
42. Queiroz, K de (1998). The General Lineage Concept of Species, Species Criteria, and the Process of Speciation: A Conceptual Unification and Terminological Recommendations. In: Endless Forms: Species and Speciation, ed. DJ Howard, and SH Berlocher, Oxford University Press, 0.
43. Roberts, DS, Loo, JA, Tsybin, YO, Liu, X, Wu, S, Chamot-Rooke, J, Agar, JN, Paša-Tolić, L, Smith, LM, and Ge, Y (2024). Top-down proteomics. Nat Rev Methods Primers 4, 38.
44. Rundle, HD, and Nosil, P (2005). Ecological speciation. Ecology Letters 8, 336–352.
45. Schluter, D (2001). Ecology and the origin of species. Trends Ecol Evol 16, 372–380.
46. Schluter, D, and Rieseberg, LH (2022). Three problems in the genetics of speciation by selection. Proc Natl Acad Sci U S A 119, e2122153119.
47. Selivon, D, Perondini, ALP, Hernández-Ortiz, V, doVal, FC, Camacho, A,

Gomes, FR, and Prezotto, LF (2022). Genetical, Morphological, Behavioral, and Ecological Traits Support the Existence of Three Brazilian Species of the Anastrepha fraterculus Complex of Cryptic Species. Front Ecol Evol 10.

48. Shear, WA, and Bonamo, PM (Patricia M) (1988). Devonobiomorpha, a new order of centipeds (Chilopoda) from the Middle Devonian of Gilboa, New York State, USA, and the phylogeny of centiped orders. American Museum novitates ; no. 2927.
49. Shelley, R, Edwards, GB, and Chagas, A (2005). Introduction of the centipede Scolopendra morsitans L., 1758, into northeastern Florida, the first authentic North American record, and a review of its global occurrences (Scolopendromorpha: Sscolopendridae: Scolopendrinae). Entomological News.
50. Shin, CP, and Allmon, WD (2023). How we study cryptic species and their biological implications: A case study from marine shelled gastropods. Ecol Evol 13, e10360.
51. Siriwut, W, Edgecombe, GD, Sutcharit, C, and Panha, S (2015). The Centipede Genus Scolopendra in Mainland Southeast Asia: Molecular Phylogenetics, Geometric Morphometrics and External Morphology as Tools for Species Delimitation. PLoS ONE 10, e0135355.
52. Siriwut, W, Edgecombe, GD, Sutcharit, C, Tongkerd, P, and Panha, S (2016). A taxonomic review of the centipede genus Scolopendra Linnaeus, 1758 (Scolopendromorpha, Scolopendridae) in mainland Southeast Asia, with description of a new species from Laos. Zookeys, 1–124.
53. Slippers, B, Roux, J, Wingfield, MJ, van der Walt, FJJ, Jami, F, Mehl, JWM, and Marais, GJ (2014). Confronting the constraints of morphological taxonomy in the Botryosphaeriales. Persoonia 33, 155–168.
54. Smith, CF, Nikolakis, ZL, Ivey, K, Perry, BW, Schield, DR, Balchan, NR, Parker, J, Hansen, KC, Saviola, AJ, Castoe, TA, *et al.* (2023). Snakes on a plain: biotic and abiotic factors determine venom compositional variation in a wide-ranging generalist rattlesnake. BMC Biology 21, 136.
55. Struck, TH, and Cerca, J (2019). Cryptic Species and Their Evolutionary Significance. In: eLS, John Wiley & Sons, Ltd, 1–9.
56. Tamura, K, Stecher, G, and Kumar, S (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution

38, 3022–3027.
57. Tasima, LJ, Hatakeyama, DM, Aguiar, W da S, Lima, EOV de, Miyamoto, JG, Tashima, AK, Sant'Anna, SS, Grego, KF, Morais-Zani, K de, and Tanaka-Azevedo, AM (2022). Analyzing the influence of age and sex in *Bothrops pauloensis* snake venom. Toxicon 214, 78–90.
58. Trifinopoulos, J, Nguyen, L-T, von Haeseler, A, and Minh, BQ (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research 44, W232–W235.
59. Undheim, EAB, Fry, BG, and King, GF (2015). Centipede Venom: Recent Discoveries and Current State of Knowledge. Toxins (Basel) 7, 679–704.
60. Undheim, EAB, and King, GF (2011). On the venom system of centipedes (Chilopoda), a neglected group of venomous animals. Toxicon 57, 512–524.
61. Vahtera, V, Edgecombe, GD, and Giribet, G (2012). Evolution of blindness in scolopendromorph centipedes (Chilopoda: Scolopendromorpha): insight from an expanded sampling of molecular data. Cladistics 28, 4–20.
62. Yang, Z (2015). The BPP program for species tree estimation and species delimitation. Current Zoology 61, 854–865.