UNIVERSITETET
I OSLO

Anatomical constraints on venom and toxin evolution in centipedes

Sebastian Hverven

Master of Science Project

Ecology and evolution

60 credits

Centre for Ecological and Evolutionary Synthesis

Department of Biosciences

Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

November 2022

Supervisor: Eivind Andreas Baste Undheim

Co-supervisors: Kjetil Lysne Voje

UNIVERSITY
OF OSLO

CEES

Acknowledgement

I would like to thank my supervisors Eivind A.B. Undheim and Kjetil L. Voje for all your support and endless knowledge. I want to thank you for having given me the opportunity to study and further express my great obsession for venomous insects. I want to personally thank Undheim for learning how to better utilize bioinformatics in my project and assisting me through the difficult obstacles. I want to thank Bernd Thiede for performing the LC-MS/MS analysis, and Norwegian Sequencing Centre (NSC) for library preparation and sequences of my samples. Jan Phillip, I want to thank you for performing the sampling of my species during the issues caused by the Covid-19 pandemic, and for the many funny conversations. To Kjetil, I wanna thank you for all your phylogenetic knowledge and Kyrre Grøtan your positive feedback that kept me motivated.

Abstract

Centipedes are among the oldest groups of venomous predatory arthropods. They have had similar morphology and life history for many million years and can be compared with centipedes from over 430 Mya. They offer an opportunity to study a homologous venom system, where the lineages have relatively retained similar body plans through their evolutionary history. Previous study on venom evolution commonly tend to focus on drivers with may facilitate the evolution, and less at what may constrain it. This makes centipedes a great research subject as there is observed discrepancies in venom complexity, which is thought to be caused by morphological limitations of its venom glands. The research question in this project was How have differences in structural venom system morphology influenced the venom complexity, its composition, and the compositional changes along the evolutionary history of different centipede lineages? To explore this, we utilized the first database to cover all five taxonomical orders of centipedes. Moreover, I characterized venom composition of extracted venom from 13 new species, where one of these were *Henia vesuviana*, using the methodology of proteotranscriptomics. Here, I argue that gland complexity is correlated to venom complexity. Strikingly, I discovered 52 new unknown, newly discovered toxin families, where 99 were unique for single orders. My results supported my hypotheses, as all four estimates for the venom complexity had a significant correlation to gland morphometrics. We show consistency with previous studies, and some new advances on the first ever indications of a toxin families identified in all five centipede orders.

Keywords: venom evolution, venom gland morphology, anatomical constraint, centipede, proteomic, transcriptomic, proteotranscriptomics.

Content

1. Introduction

The fascination for venomous organisms as scientific subjects has a long-standing history. This fascination is probably largely due to our instinctive response to venomous organisms as being dangerous, but also perhaps through questions regarding how smaller sized organisms have the ability to inflict such harm on larger predators and prey. Animal venoms are mixtures of various components, collectively referred to as toxins, such as peptides and proteins, salts, small molecules, and organic components such as neurotransmitters and amino acids. Toxins disrupt the regular physiological functions in the envenomated organisms, and while their venom contains various components, the majority of these toxins are proteins and peptides (Jared et al., 2021; Jenner et al., 2019). This biochemical weaponry is a multifunctional adaptation used for various purposes, such as defense, predation, and intraspecific competition (Schendel et al., 2019).

Toxin recruitment and evolution can be facilitated by several mechanisms, but the popular hypothesis on the origin of toxins propose the following: genes encoding for physiological proteins are duplicated, frequently followed by positive selection promoting functional diversity or neofunctionalization, thenceforth additional events of duplications and concerted evolution, which cause increase in effective expression levels in the venom glands. Additional mechanisms may also support the acquisition of venom genes, such as co-option of single genes, horizontal gene transfer, and some post-translational modifications (Hargreaves et al., 2014; Zancolli and Casewell, 2022). Acquiring the recruitment of toxin itself is only a part of the more complex system, as development of toxins is accompanied by various structural innovations (Casewell et al., 2013; Wagner, 2015), where these traits form an integrated venom system (Kazandjian et al., 2022).

Venom systems are among the most frequently evolved adaptative traits in the animal kingdom which have emerged independently over 100 times. To date, there are well over 200,000 identified venomous animal species (Schendel et al., 2019; Zancolli and Casewell, 2022), where they represent remarkable examples of convergent evolutionary novelties. Given the taxonomically diverse occurrences of venom, a large range of behavioral traits and structural innovations have evolved to facilitate its production and utilization. An example of such convergently evolved innovations are mechanical structures (stingers, fang, harpoons, etc.) to deliver venom in a hypodermic needle-like fashion (Casewell et al., 2013; Schendel et al., 2019; Zancolli and Casewell, 2022). These similar structures have originated from various pre-existing structures of different functions, such as the modification of ovipositors in Hymenoptera (Arif and Williams, 2018) and modification of walking legs in centipedes.

Acquiring such an integrated venom system is not a simple process. Multiple interdependent structures, tissues, and physiological proteins need to be altered to serve a different function from its original purpose. Alterations like these, where existing traits evolve to serve multiple functions is referred to as co-option (True and Carroll, 2002), and may represent the first step in the evolution of venom systems. Apart from toxic neofunctionalization of physiological proteins, acquisition of the venom system involves evolution of tissues for toxin production and transportation, and anatomical structures, such as fangs, stingers, etc. for delivering venom. Most venom systems among venomous lineages are usually constructed by a pair of venom glands and their corresponding delivery structures. Morphological structures and properties of these venom systems differ greatly between lineages, despite frequently serving the same general purpose. This structural variation offers information on how morphological, genetic, and ecological limitations may form the complexity of organisms (Schendel et al., 2019; Zancolli and Casewell, 2022).

Toxin evolution is widely regarded as being driven primarily by ecological factors relating to the function of the venom that contain them (predation, defense, etc.) (Schendel et al., 2019). Ecological factors, such as diet of phylogenetic diverse prey, is an example of such factors affecting the complexity of venoms, which is observed in pit vipers (Holding et al., 2021; Undheim et al., 2015b). Venom system morphology is also thought to facilitate the functional evolution and diversification of toxins in venoms that play multiple ecological roles by enabling behavioral control over the composition of secreted venom, thereby reducing evolutionary conflict between potentially conflicting roles, such as predation and defense (Schendel et al., 2019). The latter examples give an insight into the interaction between traits of different levels of biological complexity, and how they affect each other's evolution.

Just as morphological adaptations may facilitate toxin diversification, venom system morphology is also thought to have a negative effect on venom evolution. It has, for example, been hypothesized that venom complexity has an inverse relationship to the animal's ability to physically subdue prey, where the strength of positive selection driving toxin diversification reflected the dependency on venom in a metabolic cost-benefit scenario (Morgenstern and King, 2013). However, this relationship is not always observed, such as in the case of centipedes (Undheim et al., 2015b), which suggests other properties of the venom system may also affect the evolution of venom.

One such property is the number of venom-producing cells in the venom glands. There is emerging evidence that venom production is not even among secretory cells in venom glands, and that each of these cells may instead be specialized to produce a limited number of toxins. Although this

differentiated toxin production was initially thought to be an adaption for modulating venom components according to their function (Dutertre et al., 2014; Walker et al., 2018), recent work on spitting cobras suggest it may instead reflect more fundamental physiological constraints on toxin production (Kazandjian et al., 2022). Such constraints have also been hypothesized to affect the evolution of venom in centipedes (Undheim et al., 2015b), which offer a unique opportunity to test these potential morphological effects on the evolution of molecular traits.

Centipedes are terrestrial arthropods in the class Chilopoda and represent one of the four major lineages of subphylum Myriapoda (Edgecombe and Giribet, 2007; Undheim and King, 2011). There are five living orders of centipedes: Scolopendromorpha, Geophilomorpha, Lithobiomorpha, Craterostigmomorpha, and Scutigeromorpha. These orders contain roughly 3500 extant species, which are located in every continent except Antarctica (Dugon, 2017; Smith and Undheim, 2018; Undheim and King, 2011). Among the morphological traits that is shared among Chilopoda, their perhaps most distinctive shared characteristic are the venomous toxicognaths. This apparatus, also referred to as forcipules, is a modification of the most anterior walking appendages and is used to pierce membrane of prey and deliver venom (Figure 1). The venom is delivered through the opening in the distal end of the forcipules (apical claw) called the meatus that is connected to a venom gland with a cuticular duct. A porous part of the venom duct (calyx) stretches distally into the glands, where each pore is connected to a single secretory unit that is made up of two secretory cells, one canal cell, and one intermediary cell (Figure 1B and 1C). The glands are commonly located in/or near the forcipules, but in some extreme cases, like in *Henia vesuviana* (Dignathodontidae), their abnormally large glands are found between their 12^{th} and 18^{th} body segment (Undheim and King, 2015).

Comparing the forcipules of different orders reveals that they differ substantially between orders (Undheim et al., 2015b) (Figure 1). More striking differences can be seen between some species of Scutigeromorpha and Scolopendromorpha, where Scolopendromorpha are usually larger, stronger, and possess forcipule with greater sclerotized structures to inflict more damage on prey. According to the hypothesis that venom complexity has an inversely relation to the animals' ability to subdue prey, these differences would suggest that scolopendromorphs would contain simpler venom compared to scutigeromorphs. Yet, despite scolopendromorphs having greater physical ability to overpower prey, they have substantially more complex venom than what is observed in scutigeromorphs (Jenner et al., 2019; Undheim et al., 2015b). Consequently, it has been hypothesized that the relatively simple venom gland architecture of centipedes poses morphological

limitations on venom production and hence toxin molecular evolution, thereby explaining the discrepancy of venom complexity in centipedes (Schendel et al., 2019; Undheim et al., 2015b).

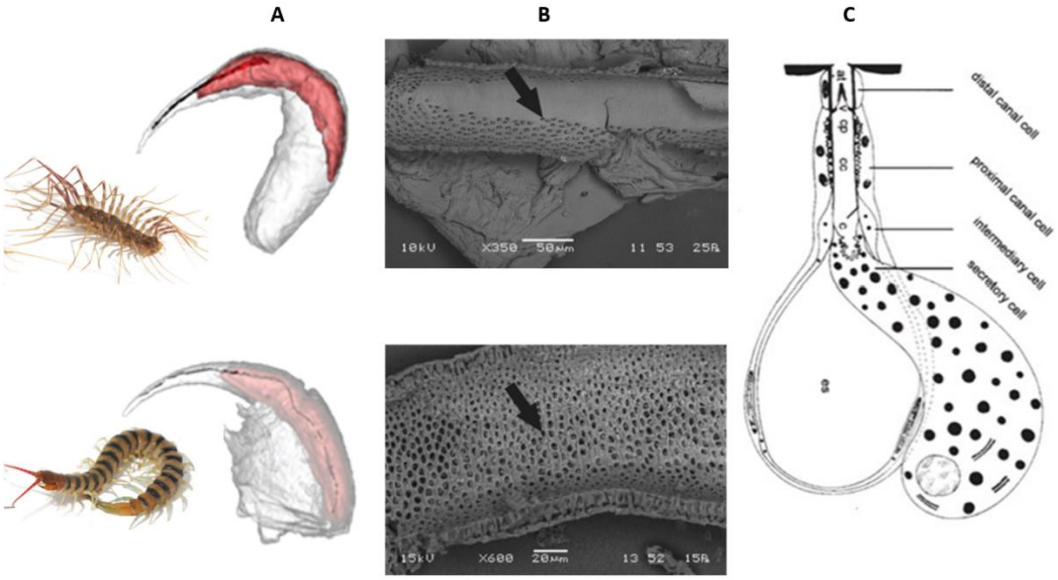

Figure 1 – Comparative structural morphology of the venom system in *Scolopendra morsitans* (lower images) and *Thereupoda longicornis* (upper images). A) Three-dimensional MRI reconstruction illustrating the forcipule (white), venom gland (red), venom duct, and calyx. B) SEM micrographs showing the inner structure of the lumen inside the calyx. Each pore in the calyx is directly connected to a secretory unit. C) Illustration of the structure for the secretory units in the glandular epithelium. The unit is constructed by 4 cells: distal canal cell, proximal canal cell, intermediary cell, and secretory cell. The images are taken from Undheim et al. (2015c) (A-B) and Rosenberg and Hilken (2006) (C).

1.1 Aim and study question

The main goal of this project is to obtain a better understanding of the factors that drive and constrain the evolution of venom by examining how the evolution of venom is affected by the properties of the venom duct and the venom-producing tissues. To achieve this goal, I characterized the secreted venoms from a wide taxonomic range of centipedes using a proteotranscriptomic approach to expand our current knowledge on centipede venom composition. This improved taxonomic sampling then allowed me to investigate the following research question: How have differences in structural venom system morphology influenced the venom complexity, its composition, and the compositional changes along the evolutionary history of different centipede lineages? To answer this question, I test the following hypotheses:

i) Venom gland complexity correlates with venom complexity.

ii) Venom gland complexity correlates with the net number of toxin family functional recruitments.

2. Material and methods

In order to test the hypothesis that venom evolution in centipedes is constrained by venom gland morphology, I analyzed the venom proteomes of a total of 22 species, which includes 9 previously published and 13 new species. Different estimates of venom complexity were correlated to different measures of venom gland morphology and the species body size to assess if there were any significant relations.

2.1 Sampling

Because the published taxonomic sampling of centipede venom proteomes is insufficient for any meaningful testing of the correlation between venom gland morphology and venom complexity, I characterized and annotated the venom proteomes of an additional 13 species using a combined proteotranscriptomic and bioinformatic approach. All additional species, apart from *Henia vesuviana* and *Cryptops iheringi* (De Lucca Caetano et al., 2021), were sampled by Eivind A.B. Undheim (supervisor). Of these species, the venom gland transcriptomes of twelve (all except *Henia vesuviana*) had already been sequenced, trimmed, and assembled, and their venoms analyzed, but not annotated. Three mature *Henia vesuviana*, along with three *Henia* spp. were collected 28th of September 2021 from Remscheid Lüttringhausen, and Oberkassel, Bonn, in eastern Germany by Jan Philip Øyen. Prior to species identification, all specimens were kept in a temperature-controlled room (24 °C) in ventilated containers, where a high humidity was obtained by using damp moss from the same location the specimens were collected. They refused food during their 2-3 weeks in captivity. No mortality among the three specimens occurred prior to venom gland extraction.

2.3 Venom extraction

To obtain a venom proteome of high-quality, it is crucial to that the extracted venom is extruded from the venom system rather than being collected through dissection (Walker et al., 2020). All three specimens of *Henia vesuviana* were anaesthetized using CO_2 before they were restrained on their dorsal side using rubber bands on a curved surface. By using a custom tool for electric stimulation, which was designed and built by my supervisor (Undheim) from a pair of micro forceps, we tried to stimulate muscle contraction to force them to expel venom which could be collected using fine disposable capillary pipette tips. However, *H. vesuviana* showed exceptional defensive behavior, which made them willing to use their venom against the capillary pipette tips and the restrains. Compared to other similar extractions of larger species, where they let the forcipules grasp on to a surface (e.g., scoopula), venom was collected from *H. vesuviana* directly from the forcipules using a micro-pipette tip. The venom samples were kept at -20 °C until

proteomic analysis. This extraction method was also performed on all other species in this project (E. A. B. Undheim, *pers. comm.*).

Table 1- The list of species that is included in this project. The list is divided in "Current study", the species I processed during this project, and "Previous studies", which is species with already existing proteomes from Jenner et al. (2019). The table illustrates taxonomical information, sample location, and additional information about the number of specimens and the sex. References to corresponding studies and suppliers are marked with footnotes.

	Order	Family	Species	Sample location	No. specimens & sex
Current study	Lithobiomorpha	Henicopidae	*Henicops maculatus*	Tasmania, Australia	2 (one male, one female)
	Geophilomorpha	Dignathodontidae	*Henia vesuviana*	Lüttringhausen, Remscheid, Germany	3 (unknown sex)
	Geophilomorpha	Zelanophilidae	*Tasmanophilus opinatus*	Tasmania, Australia	1 (unknown sex)
	Scolopendromorpha	Scolopocryptopidae	*Scolopocryptops sexspinosus*	Northern Florida, USA	2 (unknown sex)
	Scolopendromorpha	Scolopendridae	*Ethmostigmus trigonopodus*	BugzUK (https://www.bugzuk.com)	4 (two males, two females)
	Scolopendromorpha	Scolopendridae	*Hemiscolopendra marginata*	North Carolina, USA	1 (unknown sex)
	Scolopendromorpha	Scolopendridae	*Rhysida* sp.	Cairns, Queensland, Australia	1 (unknown sex)
	Scolopendromorpha	Scolopendridae	*Scolopendra laeta*	Glenmorgan, QLD, Australia	3 (unknown sex)
	Scolopendromorpha	Scolopendridae	*Scolopendra oraniensis*	Ibiza, Spain	2 (one male, one female)
	Scolopendromorpha	Scolopendridae	*Scolopendra polymorpha*	BugzUK (https://www.bugzuk.com)	2 (unknown sex)
	Scolopendromorpha	Scolopendridae	*Cormocephalus* sp.	Glenmorgan, Queensland, Australia	1 (unknown sex)
	Scolopendromorpha	Cryptopidae	*Cryptops iheringi*[3]	São Paulo, Brazil	7 (unknown sex)
	Scolopendromorpha	Cryptopidae	*Theatops posticus*	Georgia, USA	1 (unknown sex)
Previous studies	Scutigeromorpha	Scutigeridae	*Scutigera coleoptrata*[1]	Ibiza, Spain	20 (unknown sex)
	Scutigeromorpha	Scutigeridae	*Thereuopoda longicornis*[2]	Purchased[5]: Cairns, QLD, Australia	5 (unknown sex)
	Lithobiomorpha	Lithobiidae	*Lithobius forficatus*[1]	Tooting Bec Common and Hither Green, London, England	8 (unknown sex)
	Craterostigmomorpha	Craterostigmidae	*Craterostigmus tasmanianus*[1]	Hellyer Gorge, Tasmania, Australia	5 (unknown sex)
	Geophilomorpha	Linotaeniidae	*Strigamia maritima*[1]	Brora, Scotland	80 males and 57 females
	Scolopendromorpha	Scolopendridae	*Cormocephalus westwoodi*[2]	Launceston region, Tasmania, Australia	5 (unknown sex)
	Scolopendromorpha	Scolopendridae	*Ethmostigmus rubripes*[2]	Purchased[5]: Cairns, QLD, Australia	5 (unknown sex)

	Scolopendromorpha	Scolopendridae	*Scolopendra morsitans*[1]	Glenmorgan, QLD, Australia	1 female
	Scolopendromorpha	Scolopendridae	*Scolopendra subspinipes*[4]	Purchased[6]: Ludwigshafen, Germany	5 (unknown sex)

[1] Jenner et al., 2019; [2] Undheim et al., 2014a; [3] De Lucca Caetano et al., 2021; [4] Smith and Undheim, 2018; [5] Minibeast Wildlife (https://www.minibeastwildlife.com.au); [6] Zoohaus W&S (https://www.zoohaus-ws.de).

2.5 Transcriptomics

To obtain a complete catalogue of genes expressed in the venom gland, including toxins, we sequenced the venom gland transcriptomes of each species.

2.5.1 Dissection and extraction of venom gland

Three days after depletion of venom, we extracted the venom glands for transcriptomics. After being anaesthetized enough to get the specimens in position, they were euthanized by penetrating a needle right beneath the coxosternite and dissected in RNAlater (ThermoFisher, Waltham, MA, USA). Being aware of the glands position (between the 12th and the 18th segment), dissection started from the head, as we wanted to examine the venom duct as well. Using a tampered blade held by a special clamp, we made a central separation on its ventral side until the glands were visible (Figure 2A). From here, the glands were separated from the venom duct and preserved in RNAlater in separate 1.5 mL Eppendorf tubes, and stored at -80 °C. The tubes were labelled as VG (venom gland) 1 to 3 for the different specimens. We also retained the remaining body tissue from each specimen, storing these as per the venom sample.

2.5.2 RNA extraction and Transcriptomic sequencing

Extraction of venom gland total RNA was carried out on every species by using a standardized TRIzol protocol. For H. vesuviana, RNA was extracted from both dissected tissue types (venom gland and whole body) from each specimen. Separate plastic pestles were used to homogenize each sample in 1000 µL TRI Reagent solution (Invitrogen) in 1.5 mL sterile, nuclease-free DNA low-bind Eppendorf tubes. To separate RNA from DNA and protein, 200µL chloroform was added to each tube and incubated for 15 minutes after a 30 sec of gently shaking the reagent. In the first centrifuge of the separation phase, we changed the 12k rcf to 13k rcf for 10 minutes. Technical issues with the temperature regulator in the occurred during this centrifuge, which forced us to switch equipment mid-process caused the samples to go through majority of this process at 20 °C, instead of 4 °C as recommended. After separation, the top, clear, RNA containing solution was transferred to new reagents to go through RNA precipitation, which was achieved by adding 500µL

isopropyl alcohol to each tube, which was incubated for 10 new minutes. From here, the centrifuge process was repeated in pellet RNA. The supernatant was carefully removed, and the remaining pellet washed with 1mL 75% Ethanol and centrifuged at 7,5k rcf for 5 minutes. The washing step was repeated before samples were air-dried at room temperature for no more than 5 minutes in order to remove all the remaining ethanol in the reagents. Finally, the pellet was dissolved in 22 µL nuclease-free water, immediately removing a total of 2 µL for quantitative and qualitative measurements.

The yield and purity of RNA from each sample was estimated by measuring absorbance at 230, 260, and 280 nm using a Nanodrop (Thermo Scientific, USA), using water as blank and adding 1 µL sample. A ratio of 260nm/280nm and 260nm/230nm was used to assess RNA purity. This essentially indicated how pure the samples are from contaminating protein, DNA, and extraction reagent: With higher and lower than 1.8–2.2 in RNA samples (260nm/230nm: 2.00-2.20; 260nm/280nm: 1.80-2.00) will indicate protein/salt contamination and contamination by reagents such as ethanol, respectively (Watt, 2014). RNA intactness was measured using a Bioanalyzer (Agilent) RNA Pico kit, following the manufacturer's protocol. Once the RNA extracted samples were complete, they were submitted to the Norwegian Sequencing Centre (NSC) for library preparation using a TruSeq 3 paired end stranded RNA kit (Illumina, USA), and sequencing on a NovaSeq 500 using a NovaSeq SP (300 cycles; Illumina).

2.5.3 Assembly

In the absence of reference genome, the reads were assembled *de novo*. For *H. vesuviana,* separate assemblies were generated from each tissue type. Prior to this, the quality of the raw reads was examined using FASTQC v0.11.9 (Andrews, 2010) to determine appropriate trimming parameters. FASTQC offer a simple way to visualize your data and their quality scores. The raw reads were trimmed using Trimmomatic v0.39 (Bolger et al., 2014), which removes low quality bases and adapter sequences. In this case, we used Trimmomatic to trim the raw data using an average quality score threshold (WQ) of 30 across a 4 bp length sliding window (WL), and discarding reads less than a minimum read length (ML) of 80 bp. As the software reads the average quality of the 4 WL, it will trim at the point where this score falls below the set WQ 30. If this read is then shorter than the set ML (80), the read is removed. By doing this process, it reduces the amount of possible false positive caused by sequencing error and adapter induced miss-assemblies. Trimmomatic then produces four output files, two containing paired forward and reverse reads and two containing unpaired forward and reverse reads, where one read for each read pair was discarded.

Before the trimmed paired-end reads were assembled in a *de novo* transcriptomic assembly, the reads were checked again with FASTQC v0.11.9. Reads from all three VG samples were assembled together into contigs with Trinity v2.10.0 (Grabherr et al., 2011) using default parameters. The completeness of the resulting *H. vesuviana* VG assembly was checked with BUSCO v5.0.0 (Simão et al., 2015), searching against the Arthropoda single copy orthologs database, while assembly statistics were calculated using Transrate v1.0.3 (Smith-Unna et al., 2016). The whole-body samples were not processed further due to time constraints and are not presented in the results.

We next identified and translated coding DNA sequences (CDS) in the VG transcriptome using Transdecoder v5.5.0 (Haas, 2021) and compared these to proteomic data obtained from the venom. We used the proteomic data to isolate all sequences coding for toxic components, and further filter out non-toxic paralogs and reduce the overestimations of false positives. In Transdecoder, we used a minimum ORF length of 50 amino acids by using the function "Transdecoder.LongORFs". Reducing this from the default length setting of 100 amino acids can significantly affect the false positive rate of the dataset. However, this is done because some short peptide toxin coding regions are shorter than 100 aa, and that matching these to proteomic data is an effective way to eliminate these false positives.

2.6 Venom proteomics

In order to identify contigs encoding venom proteins and peptides we used proteomic analyses of milked venom. All samples were processed as the protocol described for *H vesuviana* below, except that all but *H. vesuviana* were analyzed on a 5600 triple-TOF mass spectrometer (SCIEX, USA).

To analyze the venom composition of *Henia* vesuviana, the milked venom was reduced and alkylated. This process is done in order to break the disulfide bonds in the proteins and peptides and thereby linearize them. 1µL from each venom was mixed with 8µl of 100mM ammonium bicarbonate and 5% acetonitrile (ACN) and 1 µL 50mM dithiothreitol (DTT) added for reduction of disulfide bonds. This solution was incubated for 5 minutes at 65°C to catalyze the reaction. After breaking the disulfide bonds, free sulfides were alkylated by adding 1 µL 100mM iodoacetamide (IAA) and incubating for 30 minutes at 30°C. Because iodoacetamide is unstable in light, the samples were incubated in the dark. 1 µL of 250 ng/µL trypsin was added to the reduced alkylated samples, giving a final concentration of about 20 ng/µL, which was incubated at 37 °C overnight.

To stop the reaction, the solutions' pH was reduced to below by adding 12µL of 2,5% formic acid (FA) in each reagent, diluting it to 1,25% FA in the 24µL reagents. Due to miscommunication, two

of the reagents (Hv1 and Hv2) had 20µL FA added instead. From here the samples were dried in a vacuumed chamber for 30 minutes and reconstituted in 1 % FA.

To desalt the samples prior to LC-MS/MS analysis we used C18 ZipTips (Pierce, Fisher Scientific, USA). Ziptips were activated in 100µL of an 80% ACN and 1% FA solution and equilibrated in 5 % ACN in 1 % FA before samples were loaded by pipetting up and down at least 10 times. Desalting was achieved by washing the samples in 1 % FA, and the bound peptides eluted in 80 % ACN in 1 % FA. The eluted peptides were dried in a vacuum centrifuge before they were redissolved in 5 % ACN in 1 % FA and analyzed.

The samples were analyzed by LC-MS using a timsTOF Pro (Bruker Daltonik, Bremen, Germany) which was coupled online to a nanoElute nanoflow liquid chromatography system (Bruker Daltonik, Bremen, Germany) via a CaptiveSpray nanoelectrospray ion source. The peptides were separated on a reversed phase C18 column (25 cm x 75 µm, 1.5 µm, 100Å (PepSep, Marslev, Denmark) at 50 °C. Mobile phase A contained water with 0.1% formic acid, and acetonitrile with 0.1% formic acid was used as mobile phase B. The peptides were separated by a gradient from 0-35% mobile phase B over 30 min at a flow rate of 300 nl/min at a column temperature of 50 °C. MS acquisition was performed in DDA-PASEF mode. The capillary voltage was set to 1.5 kV with a mass range of 100 to 1700 m/z. The number of PASEF ranges was set to 20 with a total cycle time of 1.16 s, charge up to 5, target intensity of 20,000, intensity threshold of 1,750, and active exclusion with release after 0.4 min. An inversed reduced TIMS mobility (1/k0) of 0.85-1.40 Vs/cm2 was used with a range time of 100 ms, an accumulation time of 100 ms, a duty cycle of 100%, and a ramp rate of 9.51 Hz. Precursors for data-dependent acquisition were fragmented with an ion mobility-dependent collision energy, which was linearly increased from 20 to 59 eV.

The LC/MS data were searched against the transcriptome file (370,636 entries; including sequences from common contaminations such as keratin and trypsin), with PEAKS X+ software version 10.6 (Bioinformatics Solutions, Waterloo, ON, Canada). The following parameters were used: digestion enzyme, none or trypsin; maximum missed cleavage, 2; fragment ion mass error tolerance, 0.03 Da; parent ion error tolerance, 15 ppm. Oxidation of methionine was specified as variable modifications. A false-discovery rate of 1% was applied to the datasets.

2.7 Proteotranscriptomics

Venom components from all species except *H. vesuviana* were identified by searching LC-MS/MS spectra against their respective translated transcriptomes with Protein Pilot v5.0 (SCIEX, USA),

using thorough search method while allowing for amino acid substitutions to account for assembly errors resulting from collapsed isoforms and intrapopulation variation. To minimalize contaminations in the upstream processes, known contaminating sequences were added to the data to remove potential cross-contaminations. False discovery rates (FDR) were estimated using a target decoy search against reversed amino acid sequences. Identified proteins were those ranked higher than that corresponding to a local FDR of 1%, ranked higher than any false positives, and had at least one or two high confidence (> 95%) peptide assignments for peptide and protein precursors, respectively.

As for *H. vesuviana*, the spectral data was searched against their transcriptomes by peptide spectral matching through PEAKS Studio (PEAKS Studio Xpro, Bioinformatics Solutions Inc.), where these samples were processed by Dr Bernd Thiede. For the output files, I used an FDR at 1% and the suggested ascore of 20.

2.8 Venom annotation

Annotation of venom components was done by using the annotated list of proteins previously published by Jenner et al. (2019) and incrementally expanding this database to include the latest annotated species. The annotation pipeline was semi-automated in bash scripts (Appendix 1) and included the following steps.

2.8.1 Extraction and clustering of sequences

Sequences of identified proteins were first extracted from the translated transcriptome using the Protein Pilot protein summary output (Appendix 1 , function 1). This was already done by Peaks Studio for *H. vesuviana;* hence this step was left out. All species sequences were then clustered using CD-HIT v4.8.1 (Li and Godzik, 2001) to cluster sequences of 100% and 70% sequence identity for later comparison (Appendix 1, function 2), which was then later searched against an annotated database of all known centipede toxins using the blastp function in BLAST+ v2.10.1 (Altschul et al., 1990; Altschul et al., 1997) (Appendix 1, function 4). Known sequences were then grouped in toxin families (Appendix 1, function 5–7). Sequences with no significant hits to known centipede toxins where then extracted as unknown sequences (Appendix 1, function 8). Toxins are almost invariably secreted proteins. We therefore used SignalP v5.0b (Almagro Armenteros et al., 2019) to filter the unknow sequences based on the presence of a signal peptide region (Appendix 1, function 10).

To identify unknown sequences containing signal peptides, we incorporated an additional script, created by postdoc Jan Philip Øyen (Appendix 1, function 12), that runs the filtered sequences through Interproscan v5.47 (Jones at al., 2014), which identifies domains that can be used for manually assigning sequences to protein superfamilies and identify putative functions. Sequences with signal peptides but without domains recognized by interproscan were searched against UniProtKB and, in the case of cysteine-rich peptides, classified according to cysteine-pattern manually.

Finally, potential hemolymph contamination was assessed by searching annotation results for hemocyanin in order to minimize inclusion and suppression by non-venom proteins, which would result in inflated and deflated estimates of toxin complexity, respectively.

2.8.2 Evolutionary dynamics

To investigate the evolutionary dynamics of centipede venom evolution, we used a character mapping approach to estimate the number of venom family recruitments and losses of families found in venom proteomes of each species. A time-calibrated phylogenetic species tree was created with TreeGraph v2.15.0-887-beta (Stöver and Müller, 2010) by manually compiling previously published centipede phylogenies (Bonato et al., 2014; Fernández et al., 2016; Siriwut et al., 2016; Vahtera et al., 2012; Vahtera et al., 2013) and personal communication with Dr Gregory D. Edgecombe (*pers. comm.*, 27/04/2022). The resulting species tree was used to reconstruct the evolutionary dynamics of centipede venom composition using a character mapping by maximum parsimony using PAUP* v4.0a builds 168 (Swofford, 2003). I used the parsimony character optimization algorithm accelerated transformation (ACCTRAN) because it is the most conservative model with regards to the number of lineage-specific toxin gene family recruitment compared to gene family losses, and hence degree of parallel evolution (Farris, 1970; Jenner et al., 2019; Swofford and Maddison, 1987).

2.8.3 Toxin diversity

There is no single way to define, nor calculate venom *complexity*, as different definitions are used among different studies (D'Suze and Carlos, 2010; Jenner et al., 2019; Pekár et al., 2018). Here, I defined venom complexity as *1)* the number of unique toxin families found in the venom, *2)* biochemical richness (toxin-encoding ORFs clustered at 100% or 70% amino acid sequence similarity), *3)* toxin diversity using Shannon-Wiener diversity index calculated from the distribution of biochemical richness (at 100 % and 70 % sequence similarity) across toxin families, and *4)*

evolutionary complexity defined as the net number of toxin recruitment events estimated to occur along the organismal phylogenetic branches leading to each species. For toxin family and biochemical richness estimates, these metrics were calculated for each species as counts. Each of these are total estimates of the sheer diversity that secretory cells are able to produce.

Toxin diversity was estimated by using the Shannon-Wiener index formula:

$$H = -\sum_{i=1}^{s} p_i log\, p_i$$

Where *s* is the total number of toxin families in the database, and p_i is the proportion of the total toxin sequences represented by toxin family *i*. The index was used in MS Excel v2209 for calculating the alpha-diversity in the venom of each species, respectively. The index takes richness and evenness into account, and in my case the biochemical diversity will increase with either greater number of unique toxin families, or increased evenness of the components. Additionally, the model sensitive to rare toxin families. This gives me an advantage as venom profiles may contain multiple toxin families of low abundance, as observed in supplementary material S3 in Jenner et al. (2019) where numerous toxin families are only represented by one transcript among the species. In addition to the estimate of total biochemical diversity, of which secretory cells can produce, this estimate also assumes additional costs associated with producing different, unrelated toxin families compared to multiple, related toxins. This may involve greater number of different components for the maturation and secretion of proteins.

For evolutionary complexity, recruitments (added) and losses (subtracted) were summed along branches leading to each species and added to the number of species-specific families (Figure 5). Toxin families predicted to be present in the venom of the last common venomous centipede ancestor were treated as constants and not counted. This estimate provides insight into whether toxin recruitment itself may be constrained by the presence of existing toxin family complexity.

2.9 Venom system morphology, type, and size

The venom system is believed to be an integrated system containing multiple structures such as the venom gland (Kazandjian et al., 2022; Undheim and King, 2011). Although the venom system is a homologous structure with little changes across millions of years, the glands' morphology varies across different orders of centipedes (Undheim et al., 2015b), resulting in a wide range in the

number secretory units. Gland complexity was defined based on calyx properties and the number of secretory units attached to each pore, classifying them as either simple (type 1) or complex (type 2), depending on how far the calyx stretches distally into the glands, which was obtained from my supervisor (E. A. B. Undheim, *pers. comm.*).

In addition to classifying glands as simple or complex, I also calculated venom gland length based on data from published literature. Due to the lack of studies done on this subject, it is tempting to assume that gland length is correlated to body length. Although this assumption is supported by multiple studies that have found gland dimensions to be proportional to body size (Antoniazzi et al., 2009; Cooper et al., 2014), the relationship does not appear to be constant across orders. I therefore calculated gland lengths using different estimations used for the five orders: for scolopendromorphs I assumed that 5.5% of the body size will represent the gland length (Antoniazzi et al., 2009; Cooper et al., 2014); for geophilomorphs except *Henia vesuviana*, I used an estimate of 1 % of body length based on approximate forcipule to full body lengths; for Craterostigmomorpha I used 5% based on approximate forcipule to full body lengths; while I used 5-6% for both lithobiomorphs and scutigeromorphs based on approximate forcipule to full body lengths (E. A. B. Undheim, *pers. comm.*). Body lengths were obtained by obtaining mean full body length estimates for each species (Table 9). For *H. vesuviana*, where the venom gland is not contained in the forcipule, I used measurements from digital pictures taken of the specimens to estimate its gland length.

2.9.1 Morphological Correlation to Composition

To assess any potential correlation between the gland properties and venom compositions, I used Rstudio v4.1.1 (Rstudio Team, 2020) to perform various correlation tests and log-linear regressions. Here, Akaike information criterion (AIC) (Akaike, 1974) was used for model selection. The reason for log-transforming trait values was due to the great numerical differences observed in absolute values. By using log-scaled datasets, we can assess the proportional relation between our response- and predictor variables. Regressions parameters were estimated using Ordinary Least Squares (OLS) in R. Note that this regression technique does not take phylogeny into account, which means we assume species can be considered independent data points. This assumption is likely to be violated, but OLS was preferred compared to more parameter rich models taking evolutionary history into account, since I have a rather low number of species in my dataset. Multiple graphs were created to illustrate the best models describing the relationship between the estimates for toxin complexity and the predictor variables, using both of the clustered datasets (100% and 70% identity clustering).

3. Results

The following sections, until the venom composition, will only include the results from *H. vesuviana*. This is due to the earlier part of the process has already been performed among the other new species, and *H. vesuviana* is the only species I have taken through the entire procedure.

3.1 Venom gland dissection and RNA extraction

The placement for the venom glands match with our gained background knowledge of *H. vesuviana* (Undheim and King, 2015). After opening the ventral side of the specimens, the glands could be found between its 12th and 18th segment (Figure 2). The white glands were laid one in front of the other, which was attached to the transparent smooth venom duct making its way to the forcipules. It was observed muscle fibers going in a diagonal fashion around the venom glands.

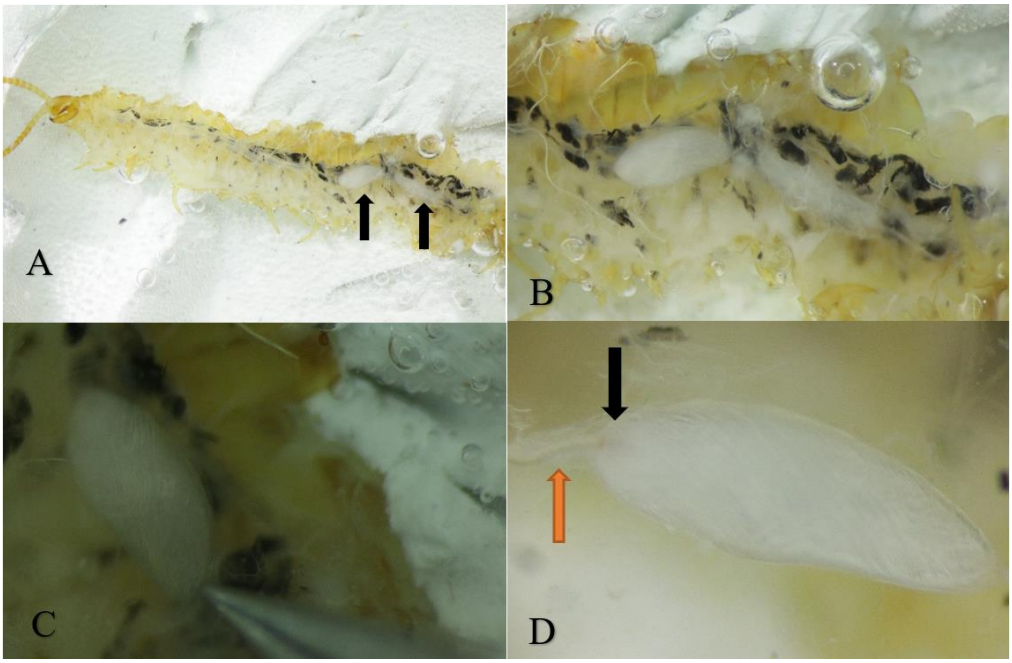

Figure 2 – Pictures taken from venom gland extraction of *Henia vesuviana*: A) Specimen after opening of its ventral side. The venom glands are indicated by two black arrows. B) Closer picture of the two gland between the 12th and 18th body segments. C) The first gland showing lines of muscle fibers going in a diagonal direction. D) The transparent smooth venom duct (orange arrow) stretching from the gland, which makes it way to the forcipule. The calyx (black arrow), also known as the porous venom duct, is shaped like a small globe and does not stretch far into the glandular epithelium.

As the glands were semi-transparent, I could observe the size and shape of the poison calyx. In addition, this gave me an insight in the properties of the porous part of the venom duct. The calyx seemed to have a small globe-like shape, similar to the calyx shape found in another geophilomorph, namely *Strigamia maritima* (Dugon and Arthur 2012a). The size and placement

may suggest that its unable to contain a great number of pores, which indicates that the large glands are constructed of fewer enlarged and elongated secretory units.
The Nanodrop calculated RNA concentration was 9.1 ng/µL in (VG1), 11.9 ng/µL (VG2), and 4.9 ng/µL (VG3) in 20 µL. The minimum recommended concentration for such samples is no less than 10 ng/µL, hence the values were lower than expected. The different values from the 260nm/280nm and 260nm/230nm purity ratios can be found in Table 1. The different absorbance values in the venom gland samples were all lower than the recommended values. Compared to the 260/280 ratio, the 260/230 values were much lower than the recommended 2.00-2.20, indicating that the samples were contaminated by impurities such as salt or other contaminants that absorbs at 230 nm. Despite these suboptimal values, we decided to go ahead with analysis of RNA integrity, library preparation, and sequencing.

Table 1 – RNA concentrations from Nanodrop of the three *Henia vesuviana* samples (VG). The purity measurements from the results included concentrations in nanogram per microliter (ng/µl) and the purity of the two nanometer ratios (260nm/280nm and 260nm/230nm).

Purity / **Samples**	**Conc. Ng/µl**	**260nm/280nm**	**260nm/230nm**
	RECM. >10 ng/µL	RECM. 1.80-2.00	RECM. 2.00-2.20
VG1	9.1	1.53	0.96
VG2	11.9	1.63	0.13
VG3	4.9	1.57	0.17

The RNA integrity analyses suggested some degradation, but that RNA was sufficiently intact for library preparation and Illumina sequencing (Figure 3). Although automatically estimated RNA integrity number (RIN) values were low, these are not reliable when determining RNA intactness for arthropods due to differences in ribosomal RNA used for RIN calculation.

3.2 Transcriptomic results

For assessing expressed toxins in the venom gland tissue among the examined species, polyA-enriched RNA was sequenced using the Illumina NovaSeq platform. The transcriptomic sequencing of the three *H. vesuviana* venom glands resulted in a total of 40 627 948 (VG1), 26 260 528 (VG2), and 33 585 582 (VG3) paired reads prior to trimming (Appendix 3). Raw reads were trimmed with a minimum read length 80 bp and a minimum average Phred quality score of 30, leaving 15 330 763 (VG1), 11 076 243 (VG2), and 13 868 200 (VG3) paired reads.

After *de novo* assembly, the contiguity and completeness of the assembly was assessed using Transrate and BUSCO. The venom gland assembly yielded a total of 138 146 contigs containing 59

473 open reading frames (ORF). The assembly had a contiguity (N50) of 2 051 bp, a mean length of 1147.78 bp among the contigs (Table 3), while the assembly's completeness (BUSCO score) as measured using the Arthropoda database was 96.9 % (Table 2).

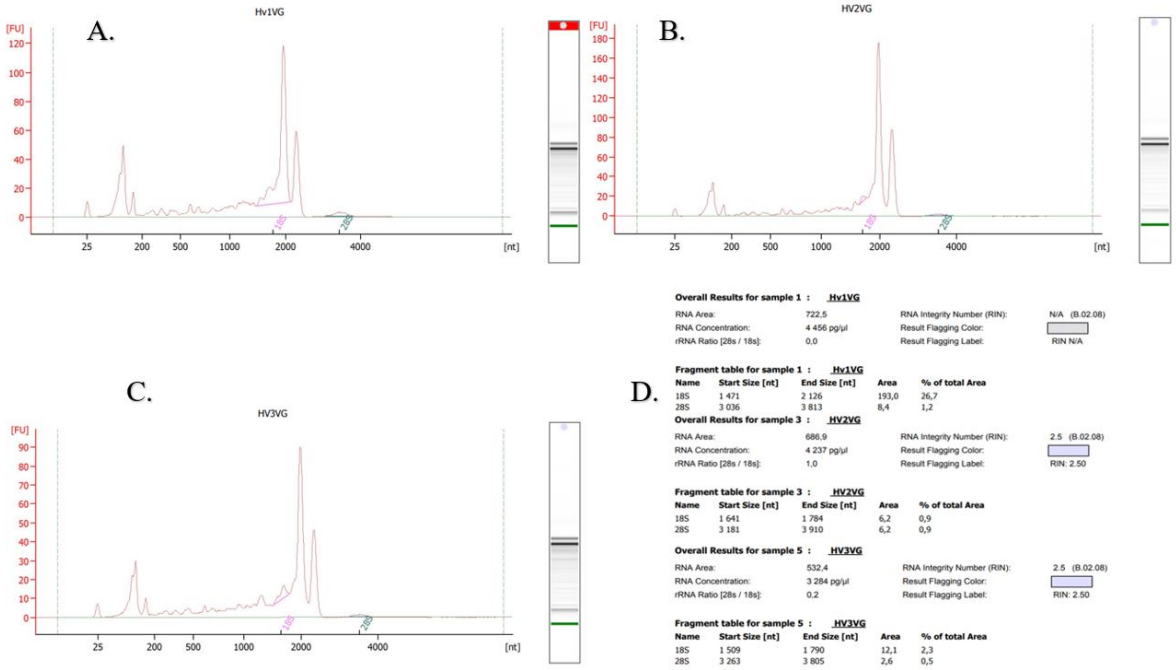

Figure 3 – Intactness measurements from the Bioanalyzer (Agilent) RNA Pico kit. A) Venom gland results from specimen 1 (Hv1VG). B) from specimen 2 (Hv2VG). C) from specimen 3 (Hv3VG). D) List of output values from the different samples.

Table 2 – Statistical information of assembly from the Venom gland tissue samples using Busco. Results from Busco showing I complete Buscos, (S) complete and single-copy, (D) complete and duplicated, (F) fragmented (M) missing, and (n) total Busco groups searched.

Busco scores	C	S	D	F	M	n
Venom gland	979 (96.6%)	293 (28.9%)	686 (67,7%)	24 (2.4%)	10 (1.0%)	1013

Table 3 – Statistical information from the Venom gland tissue samples using Transrate. Transrate software gives general information regarding contig the assemblies, which is useful for determining the assemblies' contiguity.

Transrate scores	n Sequences	Smallest Sequence	Largest Sequence	n bases	mean length	n under 200	n over 1k	n over 10k	n with orf	mean orf percent	n50
Venom gland	138146	191	25900	158561964	1147.77839	5	49043	349	59473	65.97498	2051

3.4 Proteomics and venom composition

We searched the LC-MS/MS spectra against the translated transcriptomes and annotated these using our custom toxin annotation pipeline. The result contained a total of 141 putative toxin families from 2535 unique transcriptomic sequences. Among these families, the venom of the newly annotated species contributed to a total of 52 previously undescribed, putative toxin families in centipede venom. These new families made up almost 37% of the total toxin families, and around 16% of the total sequences in the complete database of venom components across Chilopoda. Throughout the project, the toxin "family" CAP (cysteine-rich secretory proteins, antigen 5 and pathogenesis-related), was clustered in a collection of multiple protein superfamilies. The reason for doing this was insufficient time to perform phylogenetic analyses.

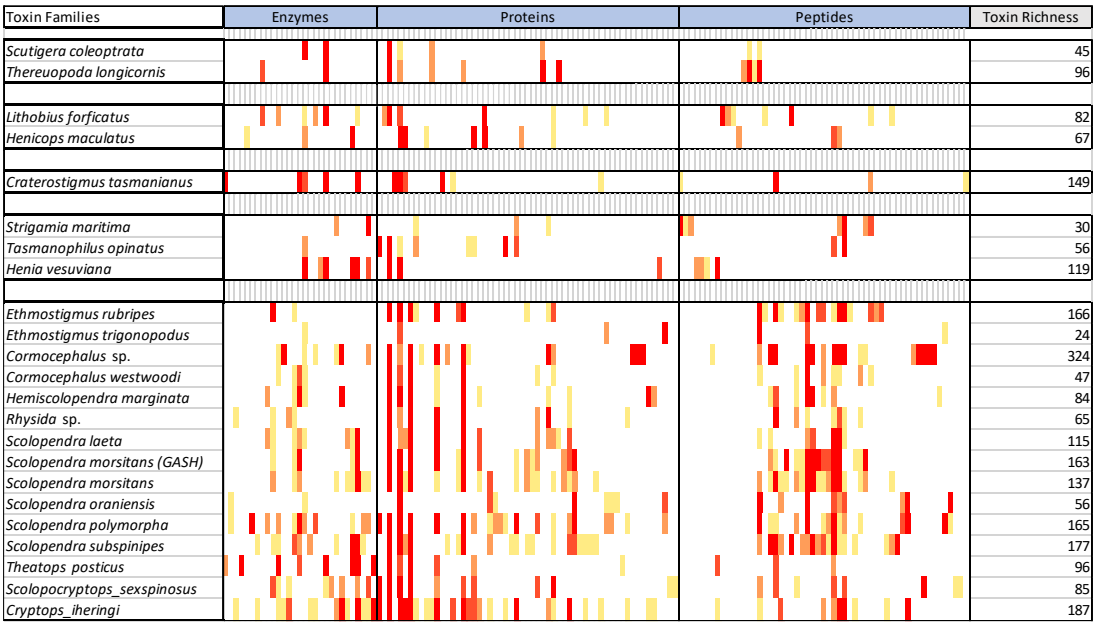

Figure 4 – General composition of centipede venoms. The figure illustrates the distributions of venom encoding sequences that are proteins, peptides, or enzymes. The sequences are shown as a heatmap, and the number of sequences found in the venom (toxin richness) are shown on the right. The species in grouped into their corresponding taxonomical order.

During the annotation, I came across cases of hemocyanin contamination present in *Ethmostigmus trigonopodus*, *Tasmanophilus opinatus*, and *Cryptops iheringi*. note that the values in the previous section have taken this into account as they were not included in the complete database. In *C. iheringi*'s data, taken from the study done by De Lucca Caetano et al. (2021), this contamination also appears in the prominent band group 2 in their SDS-PAGE analysis of the venom (Figure 5 in the paper). The study does not further mention anything about the contaminations, nor about its impact on the data. Therefore, to prevent overestimation of both putative toxin families and non-toxic paralogs, I removed all novel putative toxin families from these species—a total of 13 putative toxin families—which were either unique or exclusive to these three species. Toxin families that

were shared with other, not contaminated species were retained for further analyses assessing toxin evolution. I also considered the potential contribution of contaminants in the correlation analyses between venom complexity and morphological traits in the statistical tests (see below).

Table 4 – Total number of venom-specific sequences found in each species, distributed in order. Contaminated species are marked with "*".

Species	Number of sequences
Cormocephalus sp.	324
*Cryptops iheringi**	187
Scolopendra subspinipes	177
Ethmostigmus rubripes	166
Scolopendra polymorpha	165
Scolopendra morsitans (GASH)	163
Craterostigmus tasmanianus	149
Scolopendra morsitans	137
Henia vesuviana	119
Scolopendra laeta	115
Thereuopoda longicornis	96
Theatops posticus	96
Scolopocryptops sexspinosus	85
Hemiscolopendra marginata	84
Lithobius forficatus	82
Henicops maculatus	67
Rhysida sp.	65
*Tasmanophilus opinatus**	56
Scolopendra oraniensis	56
Cormocephalus westwoodi	47
Scutigera coleoptrata	45
Strigamia maritima	30
*Ethmostigmus trigonopodus**	24

The species with the greatest number of sequences in their venom was *Comrocephalus* sp. with a total of 324 sequences (Figure 4 and Table 4 and 9). Surprisingly, the venom of this species contained 137 more contigs, compared to the contaminated *Cryptops iheringi* (187), which had the second greatest number of toxin sequences. The 10 greatest values where mostly dominated by 7

scolopendromorphs, and 3 non-scolopendromorphs (*C. tasmanianus*, *H. vesuviana*, and *T. longicornis*). To further assess the distribution of venom component groups, we investigated if the families were found uniquely at different taxonomical levels. When including all species, the amount of toxin families unique for each order of centipede was: 4 (Scutigeromorpha), 3 (Craterostigmomorpha), 8 (Lithiobiomorpha), 7 (Geophilomorpha), and 77 (Scolopendromorpha).

Table 5 – Number of toxin families unique to different species.

Species	Unique families
Cormocephalus sp.	9
Henia vesuviana	5
Scolopocryptops sexspinosus	5
Lithobius forficatus	4
Scolopendra subspinipes	4
Henicops maculatus	3
Craterostigmus tasmanianus	3
Ethmostigmus rubripes	3
Thereuopoda longicornis	2
Strigamia maritima	2
Scolopendra polymorpha	2
Theatops sp.	2
Hemiscolopendra marginata	1
Scolopendra oraniensis	1

Among the toxin families were 46 of these uniquely found in different species (Table 5). The number of unique families in species ranged between 1–9 toxin families each. Consistent with the results before, *Comrocephalus* sp. contained the greatest number of toxin families only found in this species. The remaining values ranged between 1–5.

3.4.1 Newly discovered Toxins

Among the 52 newly discovered toxin families, we identified 12 enzymes, 22 proteins, and 18 peptides, which represented components of both known and unknown functions (Table 6 and Appendix 2). The following sections provide a brief description of these new putative toxin families.

Table 6 – Enzymes, proteins, and peptides representing the potential newly discovered toxin families from this project.

Enzymes	Proteins	Peptides
Adenylate kinase	Alpha-2-macroglobulin	GEOTX03
Alpha-Mannosidase	Cystine-knot cytokine	GEOTX04
Arylsulfatase	CAA-like	GEOTX05
Bacterial transglycosylase-like	DUF1459	Insulin-like
Carbon-nitrogen hydrolase	Leucine-rich repeat	Kazal
Chitinase	Lipid-transport protein	Neurohypophysial hormone
Concanavalin A-like lectin	Saposin-related	SLPTX42
Cyclase-like protein	Serpine	SLPTX47
Lysosomal thioesterase PPT2	Unchar henicops	SLPTX50
Peptidase M13	Unchar20	SLPTX51
Triacylglycerol lipase family	Unchar23	SLPTX55
M12B	Unchar24	SLPTX57
	Unchar28	SLPTX58
	Unchar29	SLPTX59
	Unchar30	SLPTX63
	Unchar31	SLPTX66
	Unchar32	SLPTX68
	Unchar33	SLPTX69
	Unchar34	
	Unchar35	
	Unchar37	
	Unchar38	

3.4.1.1 Enzymes

We were able to identify 12 new enzymes from the new species. Among these enzymes were concanavalin A-like lectin, peptidase M13, M12B, chitinase, Lysosomal thioesterase PPT2, bacterial transglycosylase-like, alpha-mannosidase, cyclase-like protein, adenylate kinase, carbon-nitrogen hydrolase, arylsulfatase, and triacylglycerol lipase families. These families were found among 11 of the newly added species: *Cormocephalus* sp., *C. iheringi*, *H. vesuviana*, *Henicops maculatus*, *Hemiscolopendra marginata*, *Scolopendra laeta*, *S. polymorpha*, *S. oraniensis*, *Scolopocryptops sexspinosus*, *Rhysida* sp., and *Theatops posticus*. Strikingly, metallopeptidase M12B, which was only identified in *S. sexspinosus*, is also found in the venom of some snakes (Snake venom metalloproteinase: SVMPs), where possess hemorrhagic activities and may potentially obstruct the hemostatic system in prey (Takeda, 2016). This is the second snake-like enzyme found in this species, where the known M13-type enzymes, share similarities to adamlysin-like metalloproteases in snake venom described in Ellsworth et al. (2019). This enzymatic occurrence in their venom of *S. sexspinosus* represents a striking case of molecular convergent evolution.

3.4.1.2 Non-enzymatic proteins

Among the 22 proteins identified in the venom of the new species, these proteins represent seven of known functions and the remaining 15 were proteins with uncharacterized structures or functions. The first 7 proteins were Alpha-2-macroglobulin, cystine-knot cytokine, CAA-like protein, leucine-rich repeat protein, lipid-transport protein, saposin-related protein, and serpine. Furthermore, we identified multiple new uncharacterized proteins with unknown functions. These were, DUF1459 (Protein of unknown function 1459), Uncharacterized protein family 20 (Unchar20), Unchar23-24, Unchar28-35, Unchar37-38, and Unchar_henicops (uncharacterized protein found only in *H. maculatus*).

3.4.1.3 Peptides

Among the 18 newly identified peptides were neurohypophyseal hormone-like peptides with functions that may be similar to those of oxytocin, which have been found in the venom of two marine snails of the genus *Conus* (Acher, 2004), Kazal domain peptides, Insulin-like peptides, three additional families of uncharacterised Geotoxins (GEOTX: GEOTX03-05), and 12 new uncharacterized Scoloptoxins (SLPTX: SLPTX42, 47, 51, 55, 57-59, 63, 66, and 68-69). Among the uncharacterized peptides representing cysteine-rich peptides were: GEOTX03 and GEOTX05, SLPTX47, 51, 55, 57, 58, and 68.

3.4.2 Compositional activities

When examining the results of the compositional activities, there were no clear observed trends which could explain the variation. In order to identify potential trends in the compositional dynamics, a measurement of protein, peptide, and enzymatic activities was calculated for each species, respectively (Table 7). In the initial inspection, the percentage of different activities showed large intraspecific variation between species within the same order. Furthermore, there was no indication of correlation between activities and the morphometric parameters of the different species.

Table 7 – Compositional activities in the centipede species' venom showed in percentages. The components in the venom are assigned into three groups based on their peptide, protein, or enzymatic properties. The dashed lines distinguish the five different centipede orders. Contaminated species are marked with "*".

Species	Enzymes	Proteins	Peptides
Scutigera coleoptrata	37.8 %	57.8 %	4.4 %
Thereuopoda longicornis	8.3 %	75.0 %	16.7 %
Lithobius forficatus	53.7 %	28.0 %	18.3 %
Henicops maculatus	11.9 %	77.6 %	10.4 %
Craterostigmus tasmanianus	21.5 %	23.5 %	55.0 %
Strigamia_maritima	20.0 %	13.3 %	66.7 %
*Tasmanophilus opinatus**	3.6 %	82.1 %	14.3 %
Henia vesuviana	50.4 %	37.0 %	12.6 %
Ethmostigmus rubripes	3.0 %	46.4 %	50.6 %
Ethmostigmus trigonopodus *	4.2 %	37.5 %	58.3 %
Cormocephalus sp.	9.6 %	26.5 %	63.9 %
Cormocephalus westwoodi	12.8 %	57.4 %	29.8 %
Hemiscolopendra_marginata	27.4 %	51.2 %	21.4 %
Rhysida sp.	7.7 %	72.3 %	20.0 %
Scolopendra laeta	13.0 %	69.6 %	17.4 %
Scolopendra morsitans_(GASH)	8.6 %	47.9 %	43.6 %
Scolopendra morsitans	10.9 %	48.9 %	40.1 %
Scolopendra oraniensis	3.6 %	33.9 %	62.5 %
Scolopendra polymorpha	20.0 %	44.8 %	35.2 %
Scolopendra subspinipes	13.0 %	50.3 %	36.7 %
Theatops posticus	71.9 %	19.8 %	8.3 %
Scolopocryptops sexspinosus	17.6 %	61.2 %	21.2 %
Cryptops iheringi *	42.3 %	38.6 %	19.0 %

3.4.3 Venom complexity: Toxin richness

To estimate toxin complexity, we calculated both toxin richness and diversity in the venom for each species. When calculating the toxin richness we used two different parameters: 1) Toxin family richness as the number of identified putative toxin families present in each venom, and 2) toxin richness as the total number of sequences in the venom of each species. The number of unique sequences were counted for both cases in two datasets, where one had a CD-HIT clustering at 100 % sequence identity, and the other at 70 % identity. The species possessing the most complex venoms, in terms the greatest number of identified toxin families, were *S. subspinipes* with 47 identified families, *S. polymorpha* with 46, and *S.morsitans* with 45 (Table 9). On the other end of the scale, the species with the least number of toxin families were *S. coleoptrata*, with only 8 different toxin families. The complete database did experience a single reduction in overall toxin

family richness upon clustering toxin sequences at 70 % (see below). Additionally, there was observed a reduction in the number of toxin families in two species: Both representatives of the protein Unchar35 in *S. polymorpha* (TR35108|c0_g1_i1|CDS3 at 99.20% and TR35108|c0_g1_i3|CDS3 at 99.14%), where clustered with an transcriptome from the Aldo/keto reductase family, which is presumed to be an non-toxin due to its type of enzymatic activity and lack of sufficient evidence among the non-contaminated venoms. The SLPTX69 transcript in *S. sexpinosus* (TRINITY_ DN20266_c0_g2_i6|CDS1 at 90.91%) was clustered with SLPTX55. The SLPTX69 peptide is the only loss in richness which the complete database experienced.

Of the 2535 total unique, putative toxin sequences in the 100% clustered data, the species covering the 10 greatest numbers of sequences are dominated by 8 scolopendromorphs, 1 geophilomorph (*H. vesuviana*), and 1 createrostigmomorph (*C. tasmanianus*) (Table 9). *Cormocephalus* sp. was the species with the greatest toxin richness, with 324 unique sequences, which was 137 more sequences compared to the species with the second richest venom. The second greatest toxin richness belonged to *C. iheringi* (187), followed by *S. subspinipes* (177), *E. rubripes* (166), *S. polymorpha* (165), *S. morsitans* GASH (163), *C. tasmanianus* (149), *S. morsitans* (137), *H. vesuviana* (119), and lastly *S. laeta* (115). Of these species, it should be mentioned that the toxin counts from *C. iheringi* are potentially inflated due to substantial hemolymph contamination.
Compared to the 100% identity clustered database, clustering at 70% sequence identity reduced the observed overall toxin richness from 2535 to 1317 sequences (~52% reduction) (Appendix 4). On species-level, the greatest toxin richness was again observed in *Cormocephalus* sp. (155), *Cryptops iheringi* (127), and S. subspinipes (113), having similar order as the 100% identity clustering. From here, the following greatest toxin family richness was observed in *S. morsitans* (112), *S. polymorpha* (101), and *S. laeta* (75).

3.4.4 Venom complexity: Toxin diversity

The third parameter we have used to estimate toxin complexity is the diversity of toxins. As richness alone does not take evenness into account and is not a good general measurement for diversity, I calculated the alpha-diversity of the transcript data for the 100% identity clustering for each species, respectively.

Table 5 – Shannon-Weiner diversity among all the species. The different taxonomical orders are colorized in green (Scolopendromorpha), blue (Geophilomorpha), yellow (Lithobiomorpha), orange (Craterostigmomorpha), and grey (Scutigeromorpha). Contaminated species are marked with "*".

	Shannon-Weiner index
Scolopendra polymorpha	3.4465
Scolopendra morsitans	3.2872
Cryptops_iheringi*	3.2755
Scolopendra subspinipes	3.2520
Scolopendra morsitans (GASH)	3.0944
Scolopocryptops_sexspinosus	3.0208
Cormocephalus sp.	2.9958
Hemiscolopendra marginata	2.7507
Scolopendra oraniensis	2.7298
Ethmostigmus rubripes	2.7252
Scolopendra laeta	2.6971
Cormocephalus westwoodi	2.5820
Rhysida sp.	2.5575
Theatops sp.	2.4703
Strigamia maritima	2.3323
Lithobius forficatus	2.1965
Henicops maculatus	2.0502
Henia vesuviana	1.9962
Craterostigmus tasmanianus	1.7812
Tasmanophilus opinatus*	1.7279
Ethmostigmus trigonopodus*	1.6552
Thereuopoda longicornis	1.6396
Scutigera coleoptrata	1.4828

For assessing the diversity of each species, I used the Shannon-Weiner diversity index (Shannon, 1948), which in my case is sensitive to "rare" toxin families. In our results using the Shannon-Weiner diversity index, the diversity values ranged between 1.483–3.447 (Table 8). Moreover, there seems to be some indication of clustering of the five orders and the toxin diversity values, apart from Geophilomorpha. All but *E. trigonopodus* (1.655) from the order Scolopendromorpha, contained greater diversity in their venom, ranging from 2.470-3.447, compared any other orders of centipedes. Moreover, the species with the greatest diversity was *S. polymorpha*. The geophilomorphs (1.728-2.332) were dispersed and did not have any specific placement which could represent the order. The remaining orders were Lithiobiomorpha (2.050-2.196), Craterostigmomorpha (1.781) with its single representative, and lastly Scutigeromorpha (1.489-1.640) with the lowest diversity range among the orders.

3.5 Evolutionary complexity

Evolutionary complexity was defined as the net number of toxin recruitments. By using our phylogenetic tree as the reference in PAUP*, we used ACCTRAN as the parsimony optimization algorithm to calculate the compositional changes according to my tree (Figure 5). Components that were present in the last common venomous ancestor were not included in the calculations, but their potential losses along the branches were taken into account. Contaminated species (where hemocyanin was detected in the venom) were not included due to the uncertainties of their effect on the data. The net recruitment of the included species varied between 4–43 recruitments (Table 9), where species with more than 10 net recruitments included all scolopendromorphs and two non-scolopendromorphs, namely *L. forficatus* (16) and *C. tasmanianus* (11). The 4 species with the greatest net evolutionary events were *S. subspinipes* (43), *S. morsitans* (42), *S. polymorpha* (41), and *Cormocephalus* sp. (33). Moreover, the species with the fewest events were *S. coleoptrata* with only 4 net recruitments.

According to our set phylogenetic tree, the majority of compositional evolutions occurred within the orders, which is similar to what was described in Jenner at al. (2019). Little compositional evolution is evident in the branches predating the separate orders, while 94.5% of the compositional evolutions occurred within the orders. Moreover, this supports previous assumptions stating that venom complexity has evolved in parallel along the species lineages (Jenner et al., 2019).

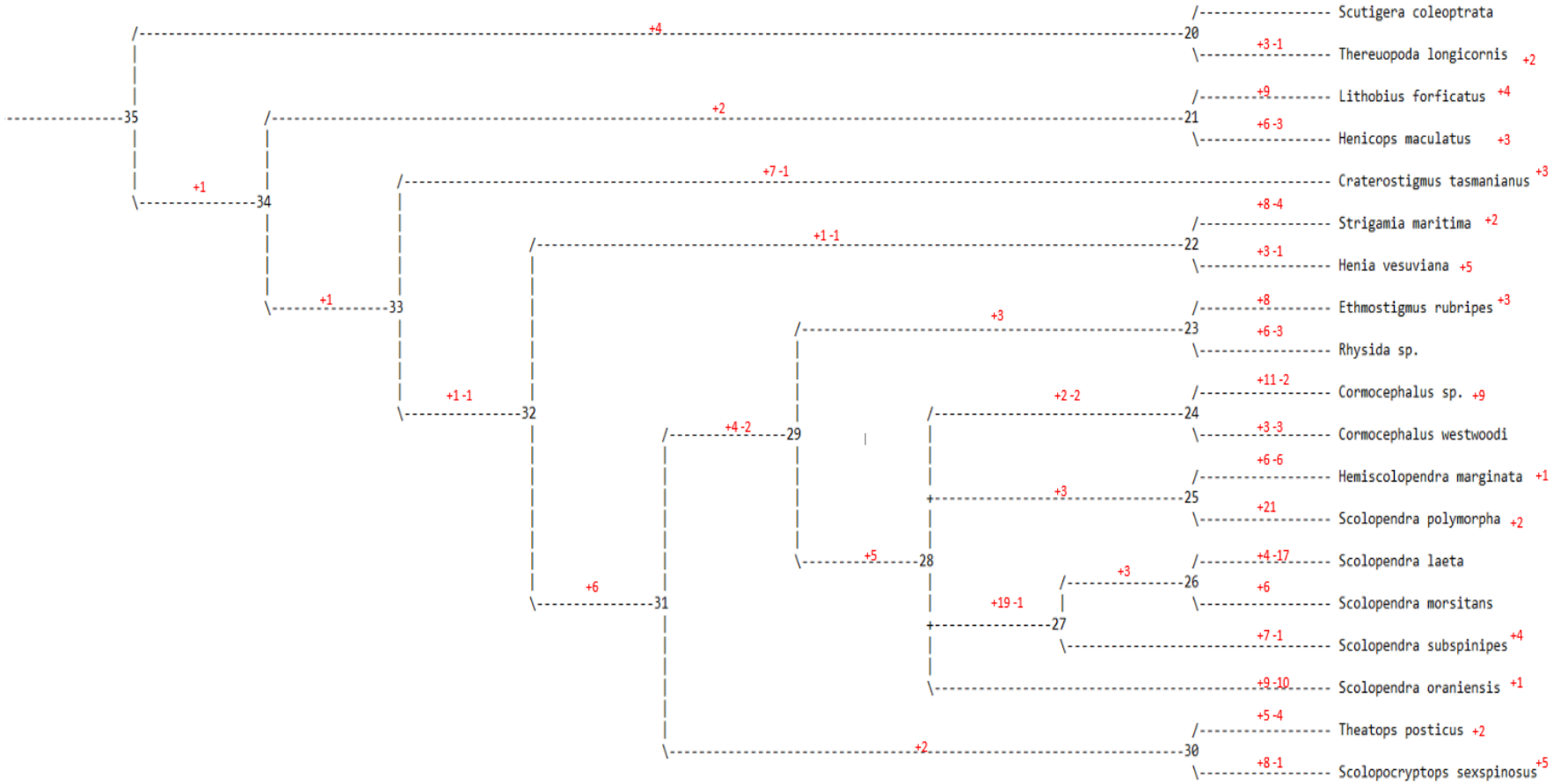

Figure 5 – The graph displays ACCTRAN estimations of functional recruitment (+) and losses (-) of putative toxin families in the different lineages of centipedes. The recruitments presented after the species names represents the putative toxin families unique to the associated species. Number in each node are merely identity-tags, which is default for this output in PAUP*. Notice that contaminated species has been removed from this model due to the risk of disturbance. The length of the tree was 252, with a Consistency index (CI) of 0.3770 (CI excluding uninformative characters = 0.3319), homoplasy index (HI) of 0.6230 (HI excluding uninformative characters = 0.6681), retention index (RI) of 0.3819, and rescaled consistency index (RC) of 0.1440.

3.6 Morphology

Information regarding the different venom gland dimensions was scarce, forcing me to perform multiple estimations based on the assumption on gland dimensions from the studies by Antoniazzi et al. (2009), Cooper et al. (2014), and personal comments from my supervisor (E. A. B. Undheim, *pers. comm.*) (Table 9). In the different centipede orders, gland lengths varied between the ranges 0.83-3.3mm (Scutigeromorpha), 1.18-1.85mm (Lithobiomorpha), 1.5mm (Craterostigmomorpha), 0.35-2.0mm (Geophilomorpha), and 1.5-8.14mm (Scolopendromorpha).

Table 9 – Morphologic and proteomic information described for the included species. The table includes variables, such as body length (BL) (mm), gland length (GL) (mm) gland type, toxin family richness (TD), toxin richness (TR), toxin diversity (TD), and Net recruitments (NR) as evolutionary complexity. The dashed lines distinguish the different orders, and contaminated species are marked with "*". The different abbreviations are: (BL) body length, (GL) gland length, (GT) gland type, (FR) family richness at 100% and 70% clustering, (TR) Toxin richness at 100% and 70% clustering, (TD) toxin diversity at 100% and 70% clustering, and (NR) net recruitments of venom components (evolutionary events) for each species and average for each orders.

Species	BL	GL	GT	FR 100%	FR 70%	TR (100 %)	TR (70 %)	TD 100%	TD 70%	NR	
Scutigera coleoptrata	15	0.83	1	8	8	45	34	1.483	1.457	4	6
Thereuopoda longicornis	60	3.3	1	12	12	96	31	1.640	1.603	8	
Lithobius forficatus	21.5	1.18	1	20	20	82	47	2.196	2.422	16	12.5
Henicops maculatus	30	1.65	1	13	13	67	39	2.050	2.126	9	
Craterostigmus tasmanianus	30	1.5	1	15	15	149	33	1.781	2.414	11	11
Strigamia maritima	35	0.35	1	12	12	30	16	2.332	2.361	8	8.5
Tasmanophilus opinatus*	80	0.8	1	11	11	56	24	1.728	1.884	-	
Henia vesuviana	50	2	1	13	13	119	33	1.996	2.005	9	
Ethmostigmus rubripes	101	5.56	2	28	28	166	51	2.725	2.903	24	27.8
Ethmostigmus trigonopodus*	120	6.6	2	7	7	24	14	1.655	1.451	-	
***Cormocephalus* sp.**	90	4.95	2	37	37	324	155	2.996	3.080	33	
Cormocephalus westwoodi	64	3.52	2	19	19	47	32	2.582	2.707	15	
Hemiscolopendra marginata	35	1.91	2	23	23	84	47	2.751	2.895	19	
***Rhysida* sp.**	60	3.3	2	20	20	65	47	2.557	2.493	16	
Scolopendra laeta	35	1.5	2	27	27	115	75	2.697	2.677	23	
Scolopendra morsitans GASH	125	6.88	2	34	32	149	73	3.122	2.994	42	
Scolopendra morsitans	125	6.88	2	45	45	137	112	3.287	3.226	42	
Scolopendra oraniensis	60	3.3	2	19	18	56	30	2.730	2.717	15	
Scolopendra polymorpha	100	5.5	2	46	45	165	102	3.447	3.473	41	
Scolopendra subspinipes	148	8.14	2	47	47	177	113	3.252	3.209	43	
Theatops posticus	40	2.2	2	17	17	96	35	2.470	2.614	13	
Scolopocryptops sexspinosus	40	2.2	2	27	26	79	48	3.021	2.720	22	
Cryptops iheringi*	80	4.4	1	49	40	187	127	3.276	3.142	-	

As for the different gland type complexities, all scolopendromorphs but *Cryptops iheringi* have type 2 (long calyx, i.e., complex) venom gland, and the remaining orders possessed type 1 (short calyx, i.e., simple) glands.

3.5.1 Morphological Effect on Composition

I performed multiple statistical tests to assess if morphological properties of the venom gland predict the venom composition and its evolution in centipedes. Given that the three species with hemolymph contamination in the venom may influence results of statistical analysis, all statistical tests were performed with and without data from these species. However, I only illustrated the results excluding the contaminants, just to be sure. The datasets were analyzed in Rstudio, where log-linear relationships were investigated using morphological variables as independent variables and the estimates of complexity as dependent variables, namely toxin family richness, toxin richness, toxin diversity, and evolutionary complexity. Models were fitted using both the 100% and 70% identity clustered datasets (except evolutionary complexity). The morphological variables were log-transformed prior to analysis, and I used AIC to evaluate relative model fit (Table 10).

Table 10 –Akaike Information Criterion (AIC) score for the different linear models investigating for a relationship between venom composition and gland type and length. The table consists of values from the models that included (values in parentheses) and excluded contaminated species based on both the 100% and 70% clustered datasets. The AIC values marked with thick letters were the models used when creating graphs.

	Gland type	Gland length	Gland type + length[1]	Gland type X length[2]
100%				
Family Richness	16.5 (38.1)	17.4 (34.9)	**11.0** (37.8)	12.5 (36.1)
Toxin Richness	36.5 (48.3)	**26.8** (42.7)	30.0 (44.6)	28.1 (43.8)
Toxin Diversity	-21.9 (-4.4)	-5.2 (0.1)	**-24.4** (-1.5)	-20.2 (-3.3)
Net Recruitments	27.56	27.93	**22.01**	23.14
70%				
Family Richness	17.1 (36.3)	17.5 (33.7)	**11.9** (36.4)	13.3 (34.6)
Toxin Richness	31.0 (47.0)	**25.0** (41.7)	27.5 (45.1)	26.2 (44.0)
Toxin Diversity	**-19.3** (-1.3)	-8.8 (1.3)	-19.2 (1.8)	-18.1 (0.03)

[1] Assumes two intercepts and a common slope; [2] assumes two intercepts and two slopes

The models that best described the dependent variables did change across the clustered datasets in toxin diversity, which changed from a multiple linear model with the 100% data to a single variable model for the 70% data. The models used for toxin family richness, toxin richness, and evolutionary complexity did not change across the identity clustered datasets. Toxin richness from both clustered datasets was best described by gland length, giving a significant p-value of 0.001046 in 100%

(Figure 6A), and 0.0004624 in the 70% dataset (Figure 6B). The model illustrated in Figure 6A has a rate of increase (regression slope) around 0.48386 (std. error: 0.1240), where 45.8% (Multiple R-squared) of the variation in toxin richness is explained by the gland length. In Figure 6B the rate of increase is 0.5054 (std. error: 0.1184), and here gland length explained 50.3% of the variation in toxin richness. The coefficients indicates that a 100% increase in the gland length variable will lead to a proportional increase in toxin richness by 48.4% and 50.1% in the different models. Compared to the 70% dataset, where type 2 glands had greater percentage increase in toxin richness with greater gland length, the 100% plot indicated type 1 having the greatest slope. Here, *H. vesuviana* (marked as green) with an estimated gland length of ~2.0mm (4% of its body length) had a much greater value of toxin richness compared to *S. maritima* (marked as orange) from the same order, which possess glands around 4 times shorter than *H. vesuviana* (1% of its body length).

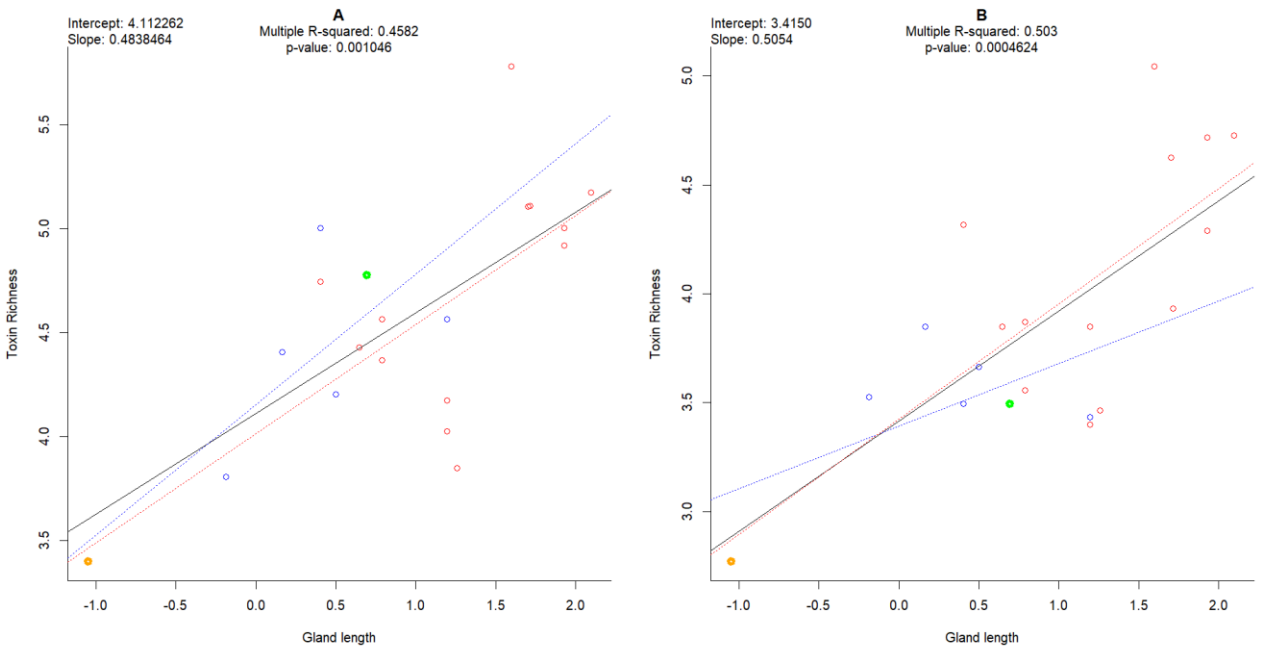

Figure 6 – Scatterplot of the log-scaled relationship between toxin richness and gland length with type 1 (blue, orange: S. maritima, and green: H. vesuviana) and type 2 (red) venom glands. The graphs contain regression lines for all datapoints (solid black line), type 1 glands (dashed blue line) and type 2 glands (dashed red line). (A) Values from the 100% identity clustering. (B) 70% identity clustering. The graphs indicated contradicting results where A indicates type 1 glands had the greatest slope, while in plot B the type 2 glands had the greater slope.

Compared to the toxin richness, the toxin family richness was best described by gland type and gland length as covariates in both clustered datasets (Figure 7). Here the 100% model had a p-value of 5.553e-05 (Figure 7A), and 70% had a p-value of 8.396e-05 (Figure 7B). These two models had two separate intercept and slope estimates i.e., a significant interaction between gland length and gland complexity. In Figure 7A (100%), a 100% increase of gland length would lead to a 7.4% (std. error: 0.15859) increase in family richness among species with type 1 glands, while species with

type 2 glands would have an increase of 44.8% (std. error: 0.21551). The corresponding multiple R-squared implies that 74.3% of the family richness variation is explained by the gland length and gland type. In Figure 7B, type 1 regression had identical 44.8% (std. error: 0.16210) increase as type 1 in the previous model, since only 2 scolopendromorphs experienced a reduction in family richness across the clustered datasets. The type 2 regression would have an increase in family richness with 45.2% (std. error: 0.22028). Here, 72.9% of the variation in family richness can be explained by the gland length and gland type.

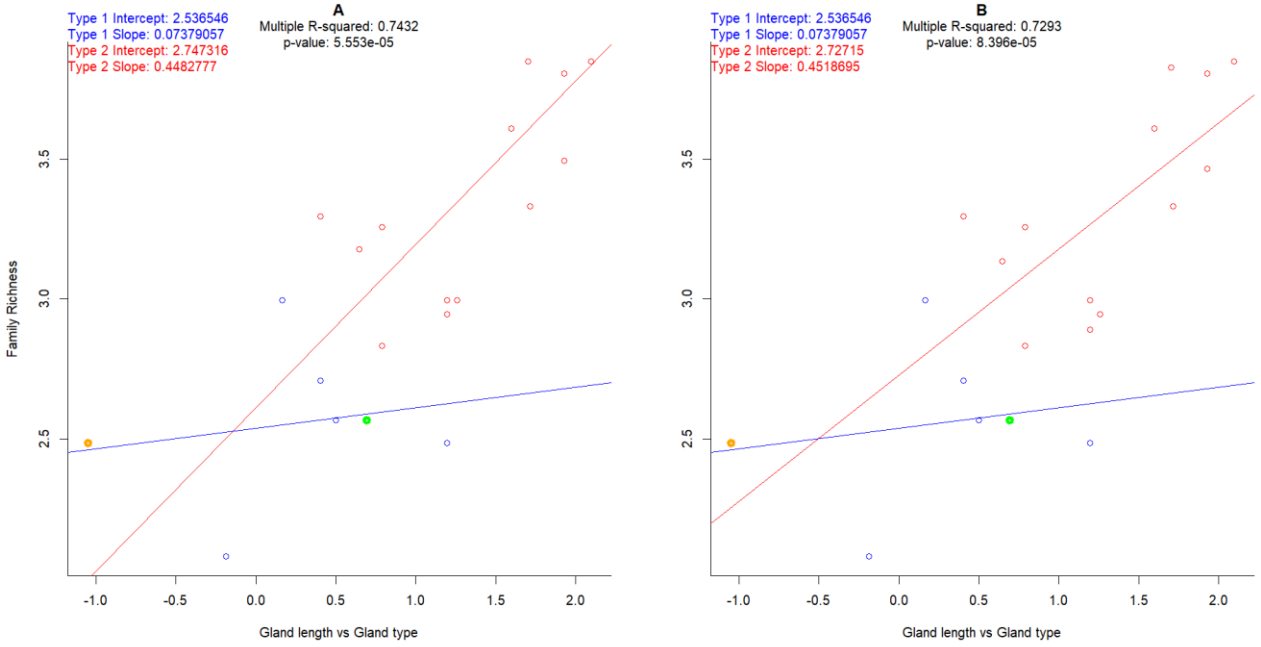

Figure 7 – Scatterplot of log-scaled relationship between toxin family richness and the covariables gland type and gland length. The graphs' models all exclude contaminated species and distinguish between type 1 (blue, orange: *Strigamia maritima*, and green: *H. vesuviana*) and type 2 (red) venom glands. The regression lines for the two different gland types had different intercept and slope in both plots, indicating that type 2 glands possess a greater rate of increase in family richness with increasing gland length. (A) Plot using values from the 100% identity clustering dataset. (B) Values from the 70% identity clustering.

The outputs of these models differ minimally, and both regression models implied that type 2 glands had a greater percentage increase of family richness, compared to type 1 when increasing gland length. The only differences are observed in *S. polymorpha* and *S. sexpinosus* as earlier mentioned. These results indicate that species with type 2 glands experiences significant greater response in family richness, per increased gland length, compared to type 1 species. This can also be observed between *H. vesuviana* and *S. maritima*, which differ greatly in gland length but experience similar percentage increase in toxin family richness.

Similar to the family richness, toxin diversity from the 100% data was best described by gland length and gland type as covariates. This was not the case in the 70%, where the best model to explain the venom diversity only utilized the gland type parameter. The 100% clustered data had a p-value of 6.401e-06 (see Figure 8A), while the 70% had a value of 2.425e-05 (see Figure 8B). In Figure 8A, the slope estimates for the gland types indicates a decrease in biochemical diversity of 9.1% (std. error: 0.06545) for type 1 species, and an increase of 12.3% (std. error: 0.08894) for type 2 when the gland length increase with 100%. The variation in diversity is 80.5% explained by the dependent covariates. In Figure 8B, the diversity variation in the model is 63.8% explained by the gland types.

Compared to previous models, *S. maritima* which had both lower family- and toxin richness, had a greater toxin diversity with shorter glands than *H. vesuviana*. In both graphs in Figure 8, there is a clear distinction in the diversity scores between type 1 and 2, whereas the previous toxin complexity estimates had some degree of overlapping scores.

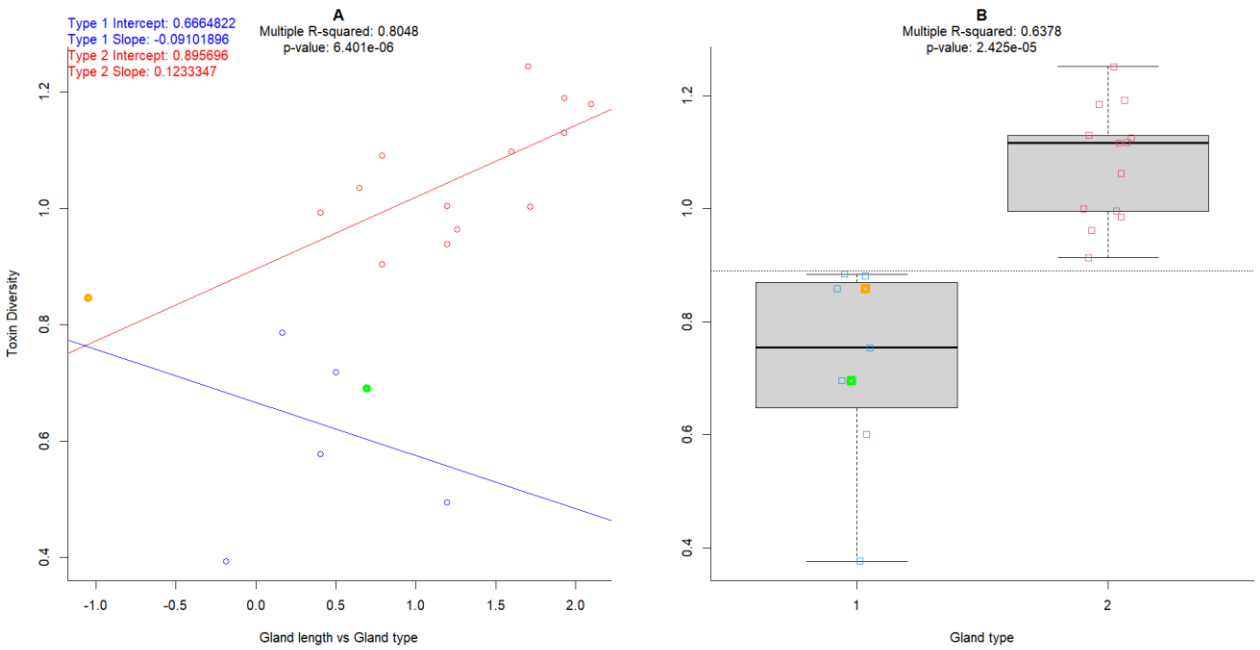

Figure 8 – Relationship between log toxin diversity and (A) log of the gland length where gland types have different slopes and intercepts using the 100% clustered dataset, and (B) the gland types using the 70% dataset. The graphs' models all exclude contaminated species and distinguish between type 1 (blue, orange: Strigamia maritima, and green: H. vesuviana) and type 2 (red) venom glands. The dashed line in boxplot B visualizes the non-overlapping values between the two different gland types.

The evolutionary complexity correlated with the gland length and gland type (using the 100% dataset). Similar to diversity and toxin family richness, there is a clear difference between the two gland types, where both types experienced an increase of 28.4% (type 1; std. error: 0.2844) and 75.2% (type 2; std. error: 0.4672) when gland length is increased by 100% (Figure 9). The model

had a p-value of 5.418e-05, where 74.4% of the variation in evolutionary complexity was explained be the covariates.

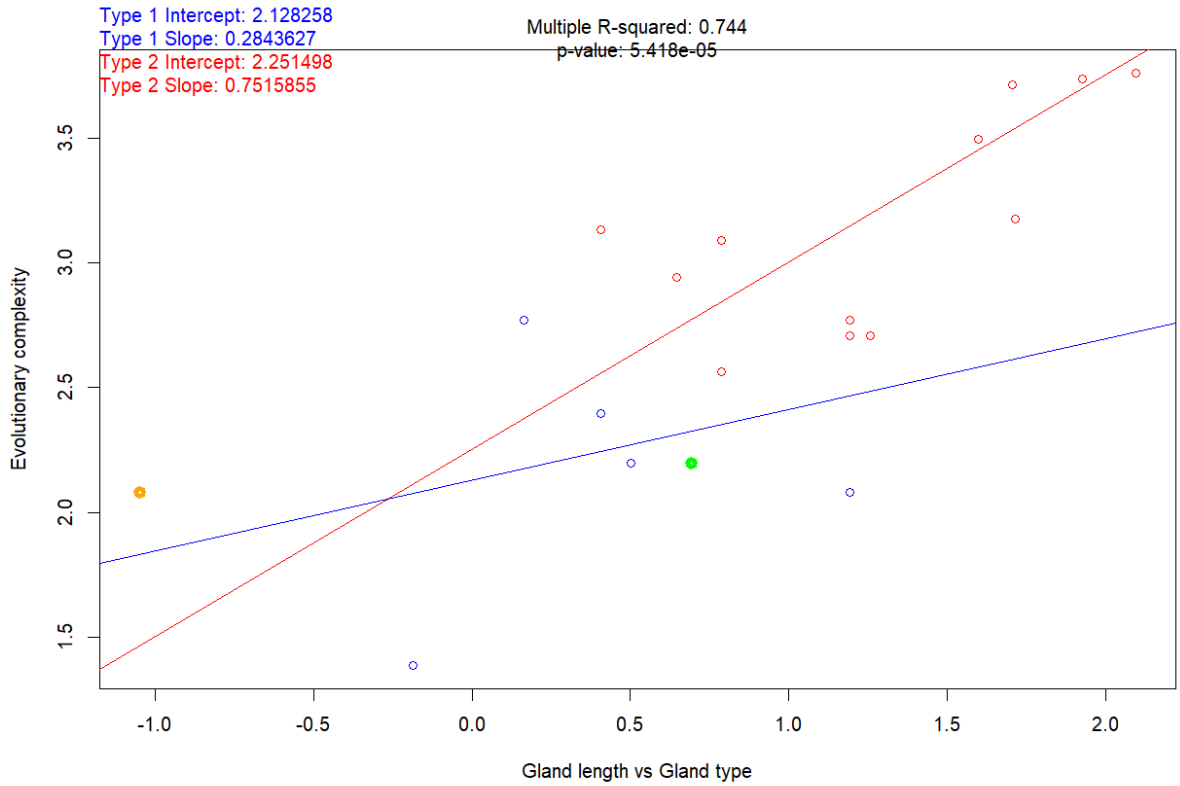

Figure 9 – Relationship between log net recruitments (evolutionary complexity) and covariables gland type and gland length. The graphs' models all exclude contaminated species and distinguish between type 1 (blue, orange: *Strigamia maritima*, and green: *H. vesuviana*) and type 2 (red) venom glands. The regression lines for the two different gland types had different intercept and slope in both plots, indicating that type 2 glands possess a greater rate of increase in net recruitments with increasing gland length.

4. Discussion

The aim of this study was to gain knowledge regarding the discrepancy of venom complexity in centipede venom, and how morphological limitations of the venom system may have affected the venom molecular evolution and composition. This was done by building on the venom specific database published by Jenner et al. (2019), to characterize the venom composition of 13 new species through proteotranscriptomics. This method offers fewer false positive toxin assignments and more accurate venom proteomes, which is caused by the reduction in the assignment of non-toxin paralogs as toxins, compared to when using a transcriptome-only approach (Smith and Undheim, 2018). Next, morphometric measurement of the venom system were collected and correlated with estimates of molecular and evolutionary venom complexity.

Despite their potentially greater ability to physically overpower prey, the members of the family Scolopendridae possess greater venom complexity. Based on my results, the discrepancy in venom complexity may be explained by the difference in properties of the species' venom systems where venom complexity is highly influenced by the complexity of their venom glands. Consequently, species with simpler glands (smaller or less porous calyx) are unable to obtain a greater venom complexity, despite them possibly being more dependent on their venom to subdue prey. Consistent with previous observations, the dynamic of compositional evolutions (functional recruitments and losses) was more frequent beyond shared stem lineages between orders. Furthermore, the four estimates for venom complexity (diversity, family richness, toxin richness, and evolutionary complexity), all supported my hypotheses that gland complexity significantly correlates with venom complexity. In the following sections, I will discuss these interpretations starting with the venom composition, its evolution, and phylogenetic distribution of toxin families. I then discuss the morphological impact on venom complexity and composition, and lastly, the study's limitations and further implications.

4.1 Composition evolution and distribution

Compared to other studies, our contribution has led to the construction of one of the largest, if not the largest, annotated list of potential centipede toxins known today. In Jenner et al. (2019), the study revealed 93 phylogenetically distinct putative toxin families, where 67 families were unique to individual orders. The complete database from this project is created from the previous existing venom profiles (proteotranscriptomic), and the added annotated data from 13 additional species (De Lucca Caetano et al., 2021; Jenner et al., 2019; Smith and Undheim, 2018; Undheim et al., 2014a). The newly 13 annotated species, composed of 1443 sequences of known and unknown families of the total 2535 from our complete list, making up 57% of the identified venom components. From these new sequences we discovered 52 new putative toxin families that did not match components found in Jenner et al. (2019) or other venom profiles of which we had covered (De Lucca Caetano et al., 2021; Smith and Undheim, 2018; Undheim et al., 2014a). Including the new species, our results produced a total of 141 distinct putative toxin families, of which 99 were unique for single orders and 46 unique to single species. This indicates high levels of evolutionary dynamics in venom compositional evolution, supporting the previous findings by Jenner et al. (2019).

It is thought that the evolution of centipede venoms is highly dynamic, and that venom complexity have evolved in parallel within the five different centipede orders (Jenner et al., 2019). Functional recruitments and losses are considered rare events, nevertheless, they have appeared frequently along the evolutionary history of centipede venoms, according to the results of previous and the current study (Jenner et al., 2019). Like previous observations, compositional evolutions (functional recruitments and losses) were not concentrated in their early evolutionary history of shared stem lineages, but occurred frequently within orders. As seen in Figure 5, the compositional evolutions increase significantly past the shared stem lineages, with 94.5% (242 of 256) of the compositional changes occurred within the orders. This is consistent with previous findings by Jenner et al. (2019).

Although our results still indicate low levels of taxonomical overlap in venom compositions, the overlap are greater than what was observed previously (Figure 4). In Jenner et al. (2019), there were no putative toxin families which could be identified in the venom of all five orders of centipedes. For the first time, through denser taxonomic sampling, we have identified 2 toxin families present in all extant orders of centipedes. The components, which are now present in the venom of all orders, are two enzymes M12A (astacin-like zinc metalloprotease, MEROPS family M12A, common in centipede toxin) and GH18 (chitinase-like glycoside hydrolase, CAZY family GH18 (Undheim et al., 2014a; Smith and Undheim, 2018).

The M12A is an endopeptidase previously only discovered in the venoms of some Scutigeromorpha, Lithobiomorpha, and Craterostigmomorpha. Although, this protease was thought to be lost within the epimorphan lineages, our results show that M12A is a more broadly distributed venom component than previously thought. This toxin family appeared in 30.4% (7/23: this ratio assumes *S. morsitans* GASH as an own species) of the total species. The new species representing M12A in the remaining orders were the Geophilomorph, *Henia vesuviana*, and the scolopendromorphs *Theatops posticus* and *Scolopcryptops sexspinosus*. The GH18, which was previously not identified in Geophilomorpha, is now observed in both newly added geophilomorphs *Henia vesuviana* and *Tasmanophilus opinatus*. The GH18 were present in 56.5% (13/23) of the total species.

Another noteworthy taxonomical expansion of a toxin family is the β-PFTxs. Jenner et al. (2019) found that β-pore-forming toxins (β-PFTxs), which is thought to have been a component in the

earliest centipede venoms, had been lost from the order Geophilomorpha. However, our deeper taxonomic sampling of Geophilomorpha show that the loss of β-PFTx did not occur in an early ancestor of the order. This lack of specimens from the same order may arguably be the case for Craterostigmomorpha, which also is thought to have experienced a functional loss of β-PFTx. On the other hand, as there are only 2 identified members of Craterostigmomorpha (Jenner et al., 2019), a single member may arguably be sufficient to represent the whole order. Including the remaining species from this order may be necessary to further address this.

4.2 Morphological impact on venom complexity

Venom complexity had a significant correlation with gland type complexity, which supported my hypotheses. All estimates of venom complexity correlated with morphological complexity of the venom glands. The four different measurements were used for estimating toxin complexity, toxin family richness (number of toxin families), toxin richness (number of sequences), toxin diversity (using Shannon-Wiener diversity index), and evolutionary complexity (net functional toxin recruitments). My results indicate that gland length and gland type correlate with family richness, toxin diversity, and evolutionary complexity, where the effect of gland types led to different rate of increases between the two types (Figure 7–9). In all three latter cases, species with type 2 glands experienced greater increase in venom complexity, when increasing gland length, compared to type 1.

Interestingly, the toxin diversity slope of type 1 glands (100% dataset) did not increase but decreased with greater gland lengths. This may be caused by the extreme score from *S. maritima* having the greatest type 1 toxin diversity, which is caused by the greater evenness of toxins in its venom. Moreover, although toxin diversity in the 70% clustered dataset was a function of only gland type, the values from type 2 glands rest well above the type 1 scores. The non-overlapping ranges further indicates the differences in percentage between the gland types. Based on the current quantity of venom, it can be hypothesized that this non-overlapped area could represent a “border” between morphological properties and greatest potential toxin diversity in simple (type 1) glands.

Toxin richness supported my hypotheses, as there was a significant correlation between the gland length and the number of toxin sequences. The log-linear regression model that best described this richness only used gland length as the predictor variable (Figure 6). Compared to the latter venom

complexity estimates, this model did not express as great differences in slopes between the gland types, and the scores appeared more overlapped. Unexpectedly, while type 2 had a slightly greater slope in the 70% clustered dataset, the 100% clustering expressed type 1 having the greater slope increase in richness.

Although there are some differences in the regressions of each gland type, these are surprisingly smaller compared to the other three estimates. If we were to rely on the 100% clustered data (Figure 6A), we would see *H. vesuviana* having type 1 glands and yet its venom would have contained greater toxin richness than 7 of the 13 included scolopendromorphs (Table 9). With a toxin richness of 119 (*H. vesuviana*), where its gland lengths were only greater than two of these 7 scolopendromorphs (*S. laeta* 1.5mm; *H. marginata* 1.9mm), this would have contradicted our hypothesis on toxin richness related calyx length and its number of secretory units. As each secretory unit can only produce limited various components in high abundance (Undheim et al., 2015b), increased calyx, thus increased number of pores and units, may facilitate greater toxin richness. *H. vesuviana*, as we know, possess a much smaller calyx proportional to its glands, similar to *S. maritima*. Although the calyx contains pores covering its whole circumference, which is not observed in other orders of what I have covered (Dugon and Arthur, 2012), it should not facilitate greater number of pores compared to others' calyx. This might suggest that there may be some facilitating properties having larger secretory units. Nonetheless, this, as well as the 70% dataset, plays a role in venom gland complexity, of which support my hypothesis that gland complexity correlates to venom complexity.

We can only speculate why there is a significant difference in toxin richness between *H. vesuviana* and *S. maritima*. As we were not able to collect more detailed information regarding the calyx in *H. vesuviana*, we can not overlook the fact that there is a large difference in calyx properties. The only visible main differences we can point out is the gland sizes and the venom duct lengths. As their venoms possessed similar amount of toxin families, but greatly differed in toxin richness, it may be that larger secretory units are able to produce greater toxin richness, or that produced venom is stored in *H. vesuviana*'s unusually long venom duct. To make any assumptions on the matter, it would acquire more morphometric information and comparative tests, which is a possible subject for future studies.

All tests from this project demonstrate that there is a significant correlation between gland morphometrics and venom complexity, which support my hypotheses. As mentioned, the centipede venom glands and their cellular structure is greatly influenced by the calyx morphology. As the

secretory units are directly connected to singular pores on the calyx, the number of units are therefore dependent on the number of pores (Undheim et al., 2015b). It is identified that in the glands of species, such as the *S. morsitans* (Figure 10), their complex glands possess more defined regional areas with diversified, heterogenous toxin gene expression across multiple secretory units. This evidently suggests that the venom glands in centipedes is a composite gland constructed by different "types" of units, or semiautonomous sub-glands. (Rosenberg and Hilken, 2006; Undheim et al., 2015b). This aligns with my hypotheses, and further support the thought that increased gland complexity leads to greater number of secretory units, which may facilitate increased toxin diversification.

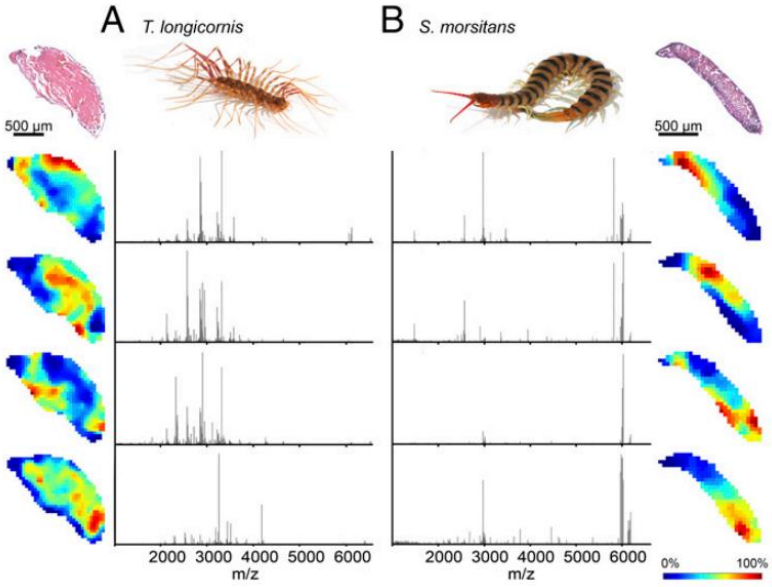

Figure 10 – The distribution of different types of secretory units in the venom glands of centipedes. The image is taken from Figure 6 in Undheim et al. (2015c), where regional groups of units with representative MS spectra is identified by calculations from AIC/pLSA on UMS data that was obtained by linear positive analysis. (A) *T. longicornis*, and (B) *S. morsitans*.

As this project has strictly assessed the morphological properties which may affect the molecular evolution, composition, and complexity in centipede venom, we can not exclude the possibility of any external factors. The discrepancy in toxin complexity may be further explained by differences in foraging strategies, diet (prey taxonomical range), prey capture rate, sex-based composition, and synergistic effect of toxic components. An ecological factor which may facilitate differences in centipede venom glands is foraging strategies. Different centipede lineages have various strategies, where species like scolopendrids actively hunt a larger phylogenomic diversity in prey (generalist), and scutigeromorphs are more opportunistic ambush predators hunting on a narrow phylogenetic prey diversity (specialist) (Undheim et al., 2015b). Among spiders, there is identified that species

that are generalists possess a larger gland compared to specialists, which is thought to be cause by the frequency of prey capture (Pekár et al., 2018). Recent evidence shows that large scolopendrids have the ability to forage arboreally and prey on small vertebrates as well (Hodges and Goodyear, 2021). Increased prey phylogenetic diversity has been observed to correlate with the venom complexity in pit vipers (Holding et al., 2021). Therefore, while the venom dependency in larger centipede species (like scolopendrids) might be lower, in means of overpowering similar prey as fewer complex species, its dependency may increase when predating on more phylogenetically diverse prey (Hodges and Goodyear, 2021; Undheim et al., 2015b). This further validates how the trait interaction of different levels of biological complexity affect each other's evolution. However, to determine the degree of influence from such factors as morphometrics versus ecology, more extensive research is acquired, which offers new subject to future studies.

5. Conclusion

Centipedes are one of the oldest lineages of venomous arthropods where there are around 3500 identified extant species in five different orders (Dugon, 2017; Smith and Undheim, 2018; Undheim and King, 2011). These arthropods offered the opportunity to study a homologous venom systems, where the lineages have relatively retained similar body plans (Jenner et al., 2019). They have received increasing attention for various reasons, where this project has focused on the discrepancy of venom complexity, which thought to be caused by morphological limitations of its venom system (Schendel et al., 2019; Undheim et al., 2015b). Nevertheless, the availability of genomic data has been scarce and venom profiles available only covers a narrow taxonomical range. A previous study was the first to create venom profiles which covered all five centipede orders and created a venom-specific database for centipedes (Jenner et al., 2019).

I have in this study build upon this database to characterize venom composition, using proteotranscriptomics, to further test our hypothesis that gland complexity mirrors venom complexity. According to my results, there is a significant correlation between venom gland complexity and venom complexity. Species possessing complex glands would gain greater venom complexity when their gland dimensions were increased. This trend can be observed in all our log-linear regression models, apart from the toxin richness using the 100% clustered data. Morphological limitations of the complexity of venom glands may therefore affect the molecular evolution, and how it has evolved through evolutionary history. Consistent with previous assumptions, the venom complexity evolved in parallel within the orders, which is supported by the phylogenetic distribution of venom components. Moreover, the assumption on evolutionary

dynamics is also relevant our results, where the number of compositional evolutions substantially increases past the orders shared stem lineages (Jenner et al., 2019). This project highlights the how venom system morphology has played a role in the compositional evolutionary history of centipede venom.

References

Acher, R. (2004). Neurohypophysial Hormone Regulatory Systems.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716–723.

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., Heijne, G.V., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature biotechnology, 37(4), 420-423.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Antoniazzi, M. M., Pedroso, C. M., Knysak, I., Martins, R., Guizze, S. P., Jared, C., & Barbaro, K. C. (2009). Comparative morphological study of the venom glands of the centipede Cryptops iheringi, Otostigmus pradoi and Scolopendra viridicornis. Toxicon, 53(3), 367-374.

Arif, F., & Williams, M. (2018). Hymenoptera Stings (Bee, Vespids and Ants).

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120.

Bonato, L., Drago, L., & Murienne, J. (2014). Phylogeny of G eophilomorpha (C hilopoda) inferred from new morphological and molecular evidence. Cladistics, 30(5), 485-507.

Casewell, N. R., Wüster, W., Vonk, F. J., Harrison, R. A., & Fry, B. G. (2013). Complex cocktails: the evolutionary novelty of venoms. Trends in ecology & evolution, 28(4), 219-229.

Cooper, A. M., Fox, G. A., Nelsen, D. R., & Hayes, W. K. (2014). Variation in venom yield and protein concentration of the centipedes Scolopendra polymorpha and Scolopendra subspinipes. Toxicon, 82, 30-51.

D'Suze, G., & Sevcik, C. (2010). Scorpion venom complexity fractal analysis. Its relevance for comparing venoms. Journal of theoretical biology, 267(3), 405-416.

De Lucca Caetano, L. H., Nishiyama-Jr, M. Y., de Carvalho Lins Fernandes Távora, B., de Oliveira, U. C., de Loiola Meirelles Junqueira-de-Azevedo, I., Faquim-Mauro, E. L., & Magalhães, G. S. (2021). Recombinant production and characterization of a new toxin from cryptops iheringi centipede venom revealed by proteome and transcriptome analysis. Toxins, 13(12), 858.

Dutertre S, Jin A-H, Vetter I, Hamilton B, Sunagar K, Lavergne V, et al. (2014). Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. Nat Commun. 5, 3521. https://doi.org/10.1038/ncomms4521

Dugon, M. M. (2017). Evolution, morphology, and development of the centipede venom system. Evolution of venomous animals and their toxins, 261-278.

Dugon, M. M., & Arthur, W. (2012). Comparative studies on the structure and development of the venom-delivery system of centipedes, and a hypothesis on the origin of this evolutionary novelty. Evolution & Development, 14(1), 128-137.

Edgecombe, G. D., & Giribet, G. (2007). Evolutionary biology of centipedes (Myriapoda: Chilopoda). Annual review of entomology, 52(1), 151-170.

Ellsworth, S. A., Nystrom, G. S., Ward, M. J., de Sousa, L. A. F., Hogan, M. P., & Rokyta, D. R. (2019). Convergent recruitment of adamalysin-like metalloproteases in the venom of the red bark centipede (Scolopocryptops sexspinosus). Toxicon, 168, 1-15.

Farris, J. S. (1970). Methods for computing Wagner trees. Systematic Biology, 19(1), 83-92.

Fernández, R., Edgecombe, G. D., & Giribet, G. (2016). Exploring phylogenetic relationships within Myriapoda and the effects of matrix composition and occupancy on phylogenomic reconstruction. Systematic Biology, 65(5), 871-889.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology, 29(7), 644-652.

Hargreaves, A. D., Swain, M. T., Hegarty, M. J., Logan, D. W., & Mulley, J. F. (2014). Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins. Genome biology and evolution, 6(8), 2088-2095.

Haas, B. (2021). TransDecoder (Find Coding Regions Within Transcripts). Github. https://github.com/TransDecoder/TransDecoder/wiki

Hodges, C. W., & Goodyear, J. (2021). Novel foraging behaviors of Scolopendra dehaani (Chilopoda: Scolopendridae) in Nakhon Ratchasima, Thailand. International Journal of Tropical Insect Science, 41(4), 3257-3262.

Holding, M. L., Strickland, J. L., Rautsaw, R. M., Hofmann, E. P., Mason, A. J., Hogan, M. P., ... & Parkinson, C. L. (2021). Phylogenetically diverse diets favor more complex venoms in North American pitvipers. Proceedings of the National Academy of Sciences, 118(17), e2015579118.

Jared, C., Luiz Mailho-Fontana, P., & Maria Antoniazzi, M. (2021). Differences between poison and venom: An attempt at an integrative biological approach. Acta Zoologica, 102(4), 337-350. https://doi.org/10.1111/azo.12375

Jenner, R. A., von Reumont, B. M., Campbell, L. I., & Undheim, E. A. (2019). Parallel evolution of complex centipede venoms revealed by comparative proteotranscriptomic analyses. Molecular biology and evolution, 36(12), 2748-2763. https://doi.org/10.1093/molbev/msz181

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics, 30(9), 1236-1240.

Kazandjian, T. D., Hamilton, B. R., Robinson, S. D., Hall, S. R., Bartlett, K. E., Rowley, P., ... & Undheim, E. A. (2022). Physiological constraints dictate toxin spatial heterogeneity in snake venom glands. BMC biology, 20(1), 1-14.

Li, W., Jaroszewski, L., & Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics, 17(3), 282-283.

Morgenstern D, King GF (2013) The venom optimization hypothesis revisited. Toxicon 63:120–128.

Pekár, S., Bočánek, O., Michálek, O., Petráková, L., Haddad, C. R., Šedo, O., & Zdráhal, Z. (2018). Venom gland size and venom complexity—essential trophic adaptations of venomous predators: a case study using spiders. Molecular Ecology, 27(21), 4257-4269.

Rosenberg, J., & Hilken, G. (2006). Fine structural organization of the poison gland of Lithobius forficatus (Chilopoda, Lithobiomorpha). Norwegian Journal of Entomology, 53(2), 119.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Schendel, V., Rash, L. D., Jenner, R. A., & Undheim, E. A. (2019). The diversity of venom: the importance of behavior and venom system morphology in understanding its ecology and evolution. Toxins, 11(11), 666.

Shannon, C. E. (1948). A mathematical theory of communication. The Bell Systems Technical Journal, 27: July 379–423.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31(19), 3210-3212.

Siriwut, W., Edgecombe, G. D., Sutcharit, C., Tongkerd, P., & 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, (590), 1. https://doi.org/10.3897/zookeys.590.7950

Smith, J. J., & Undheim, E. A. (2018). True lies: Using proteomics to assess the accuracy of transcriptome-based venomics in centipedes uncovers false positives and reveals startling intraspecific variation in Scolopendra subspinipes. Toxins, 10(3), 96.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., & Kelly, S. (2016). TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome research, 26(8), 1134-1144.

Stöver, B.C., Müller, K.F. (2010). TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11, 7. https://doi.org/10.1186/1471-2105-1

Swofford, D. L., & Maddison, W. P. (1987). Reconstructing ancestral character states under Wagner parsimony. Mathematical biosciences, 87(2), 199-229.

Swofford, D.L. (2003) PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland.

Takeda, S. (2016). ADAM and ADAMTS family proteins and snake venom metalloproteinases: A structural overview. Toxins, 8(5), 155.

True, J. R., & Carroll, S. B. (2002). Gene co-option in physiological and morphological evolution. Annual review of cell and developmental biology, 18(1), 53-80.

Undheim, E. A., & King, G. F. (2011). On the venom system of centipedes (Chilopoda), a neglected group of venomous animals. Toxicon, 57(4), 512-524. https://doi.org/10.1016/j.toxicon.2011.01.004.

Undheim, E. A., Fry, B. G., & King, G. F. (2015). Centipede venom: recent discoveries and current state of knowledge. Toxins, 7(3), 679-704.

Undheim, E. A., Hamilton, B. R., Kurniawan, N. D., Bowlay, G., Cribb, B. W., Merritt, D. J., ... & Venter, D. J. (2015). Production and packaging of a biological arsenal: evolution of centipede venoms under morphological constraint. Proceedings of the National Academy of Sciences, 112(13), 4026-4031.

Undheim, E. A., Jones, A., Clauser, K. R., Holland, J. W., Pineda, S. S., King, G. F., & Fry, B. G. (2014). Clawing through evolution: toxin diversification and convergence in the ancient lineage Chilopoda (Centipedes). Molecular biology and evolution, 31(8), 2124-2148.

Vahtera, V., Edgecombe, G. D., & Giribet, G. (2012). Evolution of blindness in scolopendromorph centipedes (Chilopoda: Scolopendromorpha): insight from an expanded sampling of molecular data. Cladistics, 28(1), 4-20.

Vahtera, V., Edgecombe, G. D., & Giribet, G. (2013). Phylogenetics of scolopendromorph centipedes: can denser taxon sampling improve an artificial classification?. Invertebrate Systematics, 27(5), 578-602.

Wagner, G. P. (2015). Evolutionary innovations and novelties: Let us get down to business!. Zoologischer Anzeiger-A Journal of Comparative Zoology, 256, 75-81.

Walker, A. A., Mayhew, M. L., Jin, J., Herzig, V., Undheim, E. A., Sombke, A., ... & King, G. F. (2018). The assassin bug Pristhesancus plagipennis produces two distinct venoms in separate gland lumens. Nature communications, 9(1), 1-10.

Walker, A. A., Robinson, S. D., Hamilton, B. F., Undheim, E. A., & King, G. F. (2020). Deadly proteomes: a practical guide to proteotranscriptomics of animal venoms. Proteomics, 20(17-18), 1900324.

Zancolli, G., & Casewell, N. R. (2020). Venom systems as models for studying the origin and regulation of evolutionary novelties. Molecular Biology and Evolution, 37(10), 2777-2790.

Appendix

Appendix 1 – Bash script for annotations

```
#! /bin/bash -ue
## usage: sbatch -J <job_name> --
export=PROTEINSUMMARY=<proteinsummary_with_contig_names_in_col7.txt>,TRANSCRIPTOME=<
translated_transcriptome.aa.fasta>,ANNOTATION_REF=<annotated_reference_fasta_file.fasta> centipede-
annotation

#SBATCH --account=<insert account>
#SBATCH --time=01:00:00 --ntasks=1 --cpus-per-task=4 --mem-per-cpu=1Gb
## set to notify by email if batch fails or completes successfully
#SBATCH --mail-type=END
#SBATCH --mail-type=FAIL
#SBATCH --mail-user=<insert email>

## exit on errors
set -o errexit
```

Function 1

```
##extract sequences for proteinpilot id's
awk 'FNR==NR {a[$7] ; next} /^>/{contig=$0 ; sub ( /^>/,"",contig ) ; hit=contig in a?1:0} hit'
$PROTEINSUMMARY $TRANSCRIPTOME > ${PROTEINSUMMARY}.fasta
```

Function 2

```
##cluster identical sequences
module load CD-HIT/4.8.1-intel-2018b
cd-hit -i ${PROTEINSUMMARY}.fasta -o ${PROTEINSUMMARY}.nr.fasta -c 1 -n 5

##load tools
module purge
module use /cluster/projects/nn9825k/software/modules/all/
module load SignalP/5.0b-Linux
module load BLAST+/2.10.1-gompi-2020a
module load InterProScan/5.47-82.0-GCCcore-9.3.0

##make folder for blast database
mkdir ${ANNOTATION_REF}.blast-db
```

Function 3

```
## go to blast database folder, make up-to date blast database, return to submission directory
cd ${ANNOTATION_REF}.blast-db
makeblastdb -in ../$ANNOTATION_REF -title $ANNOTATION_REF -dbtype prot -out
${ANNOTATION_REF}.blast-db
```

```
cd ../
```

Function 4

```
##blastp against reference blastdb
blastp -db ${ANNOTATION_REF}.blast-db/${ANNOTATION_REF}.blast-db -query
${PROTEINSUMMARY}.nr.fasta -max_target_seqs 1 -evalue 0.001 -outfmt 6 >
${PROTEINSUMMARY}_BLASTp_${ANNOTATION_REF}.txt
```

Function 5

```
##extract best hit family for each contig
awk '{ print $1"|"$2"|"$3 }' ${PROTEINSUMMARY}_BLASTp_${ANNOTATION_REF}.txt  | awk -F "|"
'{print $1"|"$2"\t"$5}' > ${PROTEINSUMMARY}_families.txt
```

Function 6

```
##convert proteinsummary fasta file to tab
awk -v RS=">" -v ORS="\n" -v OFS="" '{$1=$1"\t"}1' ${PROTEINSUMMARY}.nr.fasta >
${PROTEINSUMMARY}.fasta.nr.tab
```

Function 7

```
##combine with blast hit families to make annotated fasta file
awk 'NR==FNR{contig[$1]=$2;next}{$3=contig[$1];if($3!="")printf $1"\t"$3"\t"$2"\n"}'
${PROTEINSUMMARY}_families.txt ${PROTEINSUMMARY}.fasta.nr.tab | awk '{print
">"$1"|"$2"\n"$3}' > ${PROTEINSUMMARY}.nr.annotated.fasta
```

Function 8

```
##extract unknowns
awk 'NR==FNR{contig[$1]=$2;next}{$3=contig[$1];if($3=="")printf $1"\t"$3"\t"$2"\n"}'
${PROTEINSUMMARY}_families.txt ${PROTEINSUMMARY}.fasta.nr.tab  | awk 'NR>1 {print
">"$1"\n"$2}' > ${PROTEINSUMMARY}.nr.unknowns.fasta
```

Function 9

```
##split knowns into separate fasta files
mkdir ${PROTEINSUMMARY}.Toxin-families
cd ${PROTEINSUMMARY}.Toxin-families
awk 'NR==FNR{contig[$1]=$2;next}{$3=contig[$1];if($3!="")printf
$1"\t"$3"\t"$2"\n"}' ../${PROTEINSUMMARY}_families.txt ../${PROTEINSUMMARY}.fasta.nr.tab |
awk '{printf ">"$1"|"$2"\n"$3"\n" >> $2; close($2)}'
grep -c ">" * > ${PROTEINSUMMARY}.family-counts.txt
cd ../
```

Function 10

```
##run signalp
signalp -fasta ${PROTEINSUMMARY}.nr.unknowns.fasta
```

Function 11

```
##run interpro
```

interproscan.sh -i ${PROTEINSUMMARY}.nr.unknowns.fasta -cpu 4

Function 12

##extract hits using JP's script
perl /cluster/projects/nn9825k/scripts/sum_ipscan_v2_SignalP5.pl --ips
${PROTEINSUMMARY}.nr.unknowns.fasta.tsv --signalp
${PROTEINSUMMARY}.nr.unknowns_summary.signalp5 --out
${PROTEINSUMMARY}.nr.unknowns.ipr-signalp.summary.tsv

2 – Trimmomatic results

Appendix 2 – Paired reads count from the transcriptomic results of the H. vesuviana venom gland samples using Trimmomatic. The table show values prior and post trimming.

	Vg1		**Vg2**		**Vg3**	
	Raw	Trimmed	Raw	Trimmed	Raw	Trimmed
No. Reads	40,627,948	15,330,763	26,260,528	11,076,243	33,585,582	13,868,200

3 – Toxin table of all toxins identified in the venom

	Scutigera coleoptrata	Thereuopoda longicornis	Lithobius forficatus	Henicops maculatus	Craterostigmus tasmanianus	Strigamia maritima	Tasmanophilus opinatus	Henia vesuviana	Ethmostigmus rubripes	Ethmostigmus trigonopodus	Cormocephalus sp.	Cormocephalus westwoodi	Hemiscolopendra marginata	Rhysida sp.	Scolopendra laeta	Scolopendra morsitans (GASH)	Scolopendra morsitans	Scolopendra oraniensis	Scolopendra polymorpha	Scolopendra subspinipes	Theatops sp.	Scolopocryptops_sexspinosus	Cryptops_iheringi
Unchar23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Unchar20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0
Unchar17	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	2	0	0	0	0
Unchar16	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Unchar15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
Unchar14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Unchar13	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
Unchar12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Unchar11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	2	4	6	3	0	0	0
Unchar10	0	0	0	0	0	0	0	0	0	0	0	0	1	1	3	3	1	0	2	3	0	0	4
Unchar09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0
Unchar08	0	5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Unchar06	0	0	1	1	0	0	0	0	3	0	2	1	0	0	2	0	1	0	1	1	0	5	1
Unchar05	0	0	0	0	0	1	0	0	1	0	4	0	1	5	2	0	0	0	0	0	0	0	2
Unchar04	2	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 3 – The following figures are taken from the complete toxin table. This is a more detailed version of what was illustrated in figure 4. The dashed lines separate the five different orders.

	Scutigera coleoptrata	Thereuopoda longicornis	Lithobius forficatus	Henicops maculatus	Craterostigmus tasmanianus	Strigamia maritima	Tasmanophilus opinatus	Henia vesuviana	Ethmostigmus rubripes	Ethmostigmus trigonopodus	Cormocephalus sp.	Cormocephalus westwoodi	Hemiscolopendra marginata	Rhysida sp.	Scolopendra laeta	Scolopendra morsitans (GASH)	Scolopendra morsitans	Scolopendra oraniensis	Scolopendra polymorpha	Scolopendra subspinipes	Theatops sp.	Scolopocryptops_sexspinosus	Cryptops_iheringi
Unchar03	0	0	0	0	0	0	0	0	0	0	0	1	0	2	2	2	1	0	3	0	0	3	0
Unchar02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0
Unchar01	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	2	0	0	1	0	0	0
Unchar_henicops	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Transferrin	0	0	0	0	0	2	3	0	0	0	0	0	0	0	0	1	1	0	4	1	0	0	4
TGFbeta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Serpine	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Saposin-related	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
Protein_of_unknown_function_DUF1459	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0
PEBP-like	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	1	2	0	0	1
PCPDP	0	0	10	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LysozymeC	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	3	1	0	0	0	0	0	2
Lipid-transport_protein	0	0	0	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	2	3	3
Leucine-rich_repeat	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3
LDLA	0	2	0	0	0	0	0	0	24	0	13	7	6	8	18	15	12	0	6	20	0	3	5

	Scutigera coleoptrata	Thereuopoda longicornis	Lithobius forficatus	Henicops maculatus	Craterostigmus tasmanianus	Strigamia maritima	Tasmanophilus opinatus	Henia vesuviana	Ethmostigmus rubripes	Ethmostigmus trigonopodus	Cormocephalus sp.	Cormocephalus westwoodi	Hemiscolopendra marginata	Rhysida sp.	Scolopendra laeta	Scolopendra morsitans (GASH)	Scolopendra morsitans	Scolopendra oraniensis	Scolopendra polymorpha	Scolopendra subspinipes	Theatops sp.	Scolopocryptops_sexspinosus	Cryptops_iheringi
IL17	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1	1	0	0	1	0	0	0
Ige-ESP-like	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Icarapin-like	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0
DUF4773	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DUF3472	0	0	0	0	0	0	0	0	8	0	7	1	1	5	4	4	1	0	5	0	3	2	4
DUF1397	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CAA-like	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Cystine-knot_cytokine	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Cystatin	0	0	0	0	0	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CUB_domain	0	0	0	0	0	0	0	0	10	0	9	4	4	7	15	14	12	0	6	22	4	6	6
C-type_Lectin	0	0	0	16	3	0	0	0	0	0	0	0	0	0	2	0	1	0	0	2	0	0	4
CAP	1	2	3	7	10	0	1	12	4	3	2	3	12	2	2	8	7	5	12	7	3	9	6
Calycin-Lipocalin	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BPFTx	21	55	5	0	0	0	29	29	22	0	16	10	6	16	26	18	17	0	7	19	6	7	6
AMG-like	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	Scutigera coleoptrata	Thereuopoda longicornis	Lithobius forficatus	Henicops maculatus	Craterostigmus tasmanianus	Strigamia maritima	Tasmanophilus opinatus	Henia vesuviana	Ethmostigmus rubripes	Ethmostigmus trigonopodus	Cormocephalus sp.	Cormocephalus westwoodi	Hemiscolopendra marginata	Rhysida sp.	Scolopendra laeta	Scolopendra morsitans (GASH)	Scolopendra morsitans	Scolopendra oraniensis	Scolopendra polymorpha	Scolopendra subspinipes	Theatops sp.	Scolopocryptops_sexspinosus	Cryptops_iheringi
Alpha-2-macroglobulin	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	7	0	0	12	9
Triacylglycerol_lipase_family	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	23
S8	0	0	0	0	0	4	0	3	0	0	2	0	0	0	0	0	1	0	2	0	0	3	1
S10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	1	0	0	1
S1	0	0	1	0	4	0	0	6	0	0	0	0	0	0	5	6	5	0	0	7	11	2	29
PLA2	0	0	0	5	0	0	0	4	0	0	0	0	0	0	1	1	1	0	1	4	25	0	0
Phosphodiesterase2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	1
Peptidase_M13	0	0	0	0	0	0	0	0	0	0	12	0	7	0	0	0	0	0	0	0	0	2	13
PAM-like	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	2
M12B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
M12A	12	5	35	0	4	0	0	39	0	0	0	0	0	0	0	0	0	0	0	0	4	1	0
Lysosomal_thioesterase_PPT2	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lysosomal_lipase	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0	1
LDLA-chitinase	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1
GH18	5	0	1	2	3	0	2	6	0	1	1	1	1	0	1	0	0	1	2	0	0	0	0

	Scutigera coleoptrata	Thereuopoda longicornis	Lithobius forficatus	Henicops maculatus	Craterostigmus tasmanianus	Strigamia maritima	Tasmanophilus opinatus	Henia vesuviana	Ethmostigmus rubripes	Ethmostigmus trigonopodus	Cormocephalus sp.	Cormocephalus westwoodi	Hemiscolopendra marginata	Rhysida sp.	Scolopendra laeta	Scolopendra morsitans (GASH)	Scolopendra morsitans	Scolopendra oraniensis	Scolopendra polymorpha	Scolopendra subspinipes	Theatops sp.	Scolopocryptops_sexspinosus	Cryptops_iheringi
GGT	0	0	0	0	14	0	0	0	0	0	0	3	12	0	2	5	2	0	4	2	3	0	0
GDH	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	1	3	0	0	0
Cyclase-like_protein	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	3
Concanavalin_A-like_lectin	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	1
COEsteraseB	0	0	2	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	2	1	4	1	1
Chondroitinase	0	0	0	0	0	0	0	0	4	0	0	0	0	1	1	1	2	0	0	1	0	3	0
Chitinase	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	2	0	0	0	0
CentiPAD	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cathepsin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
Carbon-nitrogen_hydrolase	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0
Bacterial_transglycosylase-like	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ARYLSULFATASE_	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0
ALPHA-MANNOSIDASE	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Adenylate_kinase	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
Acid-phosphatase	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0

	Scutigera coleoptrata	Thereuopoda longicornis	Lithobius forficatus	Henicops maculatus	Craterostigmus tasmanianus	Strigamia maritima	Tasmanophilus opinatus	Henia vesuviana	Ethmostigmus rubripes	Ethmostigmus trigonopodus	Cormocephalus sp.	Cormocephalus westwoodi	Hemiscolopendra marginata	Rhysida sp.	Scolopendra laeta	Scolopendra morsitans (GASH)	Scolopendra morsitans	Scolopendra oraniensis	Scolopendra polymorpha	Scolopendra subspinipes	Theatops sp.	Scolopocryptops_sexspinosus	Cryptops_iheringi
WAP	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SLPTX69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
SLPTX68	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
SLPTX66	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	1	0	0	0	0
SLPTX63	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	4	0	0	0	0
SLPTX59	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

	Scutigera coleoptrata	Thereuopoda longicornis	Lithobius forficatus	Henicops maculatus	Craterostigmus tasmanianus	Strigamia maritima	Tasmanophilus opinatus	Henia vesuviana	Ethmostigmus rubripes	Ethmostigmus trigonopodus	Cormocephalus sp.	Cormocephalus westwoodi	Hemiscolopendra marginata	Rhysida sp.	Scolopendra laeta	Scolopendra morsitans (GASH)	Scolopendra morsitans	Scolopendra oraniensis	Scolopendra polymorpha	Scolopendra subspinipes	Theatops sp.	Scolopocryptops_sexspinosus	Cryptops_iheringi
SLPTX58	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0
SLPTX57	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
SLPTX55	0	0	0	0	0	0	0	0	0	0	38	0	0	0	0	0	0	0	0	0	0	4	0
SLPTX51	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0
SLPTX50	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
SLPTX47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	5	0	0	0	4
SLPTX42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0	0	0	0
SLPTX31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
SLPTX30	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0
SLPTX29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
SLPTX25	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SLPTX24	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SLPTX22	0	0	1	0	2	3	0	0	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0
SLPTX21	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	4	2	0	0	0	0	0	0
SLPTX20	0	0	0	0	0	0	0	0	0	0	1	2	0	1	0	1	1	0	1	0	0	0	0

	Scutigera coleoptrata	Thereuopoda longicornis	Lithobius forficatus	Henicops maculatus	Craterostigmus tasmanianus	Strigamia maritima	Tasmanophilus opinatus	Henia vesuviana	Ethmostigmus rubripes	Ethmostigmus trigonopodus	Cormocephalus sp.	Cormocephalus westwoodi	Hemiscolopendra marginata	Rhysida sp.	Scolopendra laeta	Scolopendra morsitans (GASH)	Scolopendra morsitans	Scolopendra oraniensis	Scolopendra polymorpha	Scolopendra subspinipes	Theatops sp.	Scolopocryptops_sexspinosus	Cryptops_iheringi
SLPTX19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1
SLPTX18	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SLPTX17	0	0	0	0	0	6	5	0	5	0	10	0	0	1	1	0	1	3	2	1	0	3	15
SLPTX16	0	0	0	2	0	2	0	0	27	0	10	1	0	3	4	8	6	2	1	1	0	0	6
SLPTX15	0	0	0	3	0	0	3	0	1	0	5	1	2	1	8	12	8	3	5	15	2	3	2
SLPTX14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	0	2	1	0	0	0
SLPTX13	0	0	0	0	0	0	0	0	3	0	2	2	1	0	0	3	1	0	1	3	0	0	2
SLPTX12	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	6	1	0	0	2	0	0	0
SLPTX11	0	0	0	0	0	0	0	0	0	0	61	0	5	0	3	13	11	0	0	6	0	0	0
SLPTX10	0	0	0	0	0	0	0	0	22	3	4	5	4	1	2	9	5	6	9	2	0	0	0
SLPTX09	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	1	0	0	2	0	0	0
SLPTX08	0	0	0	0	0	0	0	0	1	0	0	1	1	2	0	1	2	0	2	0	0	1	0
SLPTX07	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
SLPTX06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1	0	0	0	0	0	0
SLPTX05	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	2	0	2	0	0	3

	Scutigera coleoptrata	Thereuopoda longicornis	Lithobius forficatus	Henicops maculatus	Craterostigmus tasmanianus	Strigamia maritima	Tasmanophilus opinatus	Henia vesuviana	Ethmostigmus rubripes	Ethmostigmus trigonopodus	Cormocephalus sp.	Cormocephalus westwoodi	Hemiscolopendra marginata	Rhysida sp.	Scolopendra laeta	Scolopendra morsitans (GASH)	Scolopendra morsitans	Scolopendra oraniensis	Scolopendra polymorpha	Scolopendra subspinipes	Theatops sp.	Scolopocryptops_sexspinosus	Cryptops_iheringi
SLPTX04	0	0	0	0	78	0	0	0	4	0	9	0	3	4	0	1	8	0	1	11	3	3	0
SLPTX03	0	0	0	0	0	0	0	0	0	0	14	0	1	0	1	2	1	0	1	5	0	0	0
SLPTX02	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SLPTX01	1	8	0	0	0	0	0	0	5	10	2	1	0	0	1	0	2	9	20	2	0	2	1
SCTX03	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SCTX02	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SCTX01	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Neurhyp_horm	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LTHTX03	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LTHTX02	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LTHTX01	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kazal	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1
Insulin-like	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
GEOTX05	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GEOTX04	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	Scutigera coleoptrata	Thereuopoda longicornis	Lithobius forficatus	Henicops maculatus	Craterostigmus tasmanianus	Strigamia maritima	Tasmanophilus opinatus	Henia vesuviana	Ethmostigmus rubripes	Ethmostigmus trigonopodus	Cormocephalus sp.	Cormocephalus westwoodi	Hemiscolopendra marginata	Rhysida sp.	Scolopendra laeta	Scolopendra morsitans (GASH)	Scolopendra morsitans	Scolopendra oraniensis	Scolopendra polymorpha	Scolopendra subspinipes	Theatops sp.	Scolopocryptops_sexspinosus	Cryptops_iheringi
GEOTX03	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GEOTX02	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GEOTX01	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CHILOTX01	0	0	0	0	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unchar38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Unchar37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Unchar35	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	3	2	0	0	0	0
Unchar34	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unchar33	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1
Unchar32	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	1
Unchar31	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0
Unchar30	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0
Unchar29	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0
Unchar28	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1
Unchar24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0