A Pseudoscorpion's Promising Pinch: The Venom of *Chelifer cancroides* Contains a Rich Source of Novel Compounds

Jonas Krämer[a], Steve Peigneur[b], Jan Tytgat[b], Ronald Jenner[c], Ronald van Toor[d], Reinhard Predel[a]

[a]*Institute of Zoology, University of Cologne, D-50674, Cologne, Germany*
[b]*Toxicology and Pharmacology, University of Leuven (KU Leuven), B-3000 Leuven, Belgium*
[c]*Department of Life Sciences, Natural History Museum, London, United Kingdom*
[d]*The New Zealand Institute for Plant and Food Research Limited, Canterbury Agriculture & Science Centre, Lincoln 7608, New Zealand*

Abstract

With pedipalps modified for venom injection, some pseudoscorpions possess a unique venom delivery system, which evolved independently from those of other arachnids like scorpions and spiders. Up to now, only a few studies have been focused on pseudoscorpion venom, which either identified a small fraction of venom compounds, or were based on solely transcriptomic approaches. Only one study addressed the bioactivity of pseudoscorpion venom. Here, we expand existing knowledge about pseudoscorpion venom by providing a comprehensive proteomic and transcriptomic analysis of the venom of *Chelifer cancroides*. We identified the first putative genuine toxins in the venom of *C. cancroides* and we showed that a large fraction of the venom comprises novel compounds. In addition, we tested the activity of the venom at specific ion channels for the first time. These tests demonstrate that the venom of *C. cancroides* causes inhibition of a voltage-gated insect potassium channel (Shaker IR) and modulates the inactivation process of voltage-gated sodium channels from *Varroa destructor*. For one of the smallest venomous animals ever studied, today's toolkits enabled a comprehensive venom analysis. This is demonstrated by allocating our identified venom compounds to more than half of the prominent ion signals in MALDI-TOF mass spectra of venom samples. The present study is a starting point for understanding the complex composition and activity of pseudoscorpion venom and provides a potential rich source of bioactive compounds usable for basic research and industrial application.

Keywords

Mass spectrometry, Toxins, Electrophysiology, Activity tests, Venom extraction, Proteomics, Transcriptomics

1. Introduction

Animal venoms are a rich source of bioactive compounds optimized by evolutionary processes for various purposes like subduing prey, defense against predators/microbes, and competition with conspecifics. As venoms contain substances that act on a wide variety of targets in different organisms, venoms represent an enormous reservoir for identifying lead compounds for developing novel pharmaceutics or pesticides (Herzig et al., 2020). In at least eleven cases, the determined structure of venom compounds has already been successfully used for developing commercial drugs (Bordon et al., 2020). The majority of these were developed on the basis of venom compounds from snakes, mainly because of the high venom amount and the research focus on snake venoms in the past (King, 2013). Within metazoans, venoms have evolved at least 100 times (Schendel et al., 2019) and solely for spiders more than 20 million venom compounds are estimated (King and Hardy, 2013). In the light of these numbers, the number of drugs developed to date based on venom compounds does not seem particularly high. However, only a small fraction of substances makes it to the final step of drug development, mainly due to lack of efficacy and side effects discovered in clinical trials during the process (Harvey, 2014). On top of that, venom research was biased towards harmful and relatively large-bodied (due to methodological constraints) taxa. Today, -omics approaches and increasing instrument sensitivity allow identifying venom compounds starting from very little amounts of substance, and modern synthesizers or recombinant manufacturing procedures allow sufficient quantities to be obtained for pharmaceutical or biotechnological approaches (Boldrini-França et al., 2017). Venom sampling across a wider phylogenetic range is very likely to increase the probability of identifying novel compounds with activities not observed before (Lüddecke et al., 2019). Even in well-studied venomous groups like spiders, uncommon venom compositions with novel venom compounds can still be found as demonstrated by recent findings on the venom composition of the wasp spider (Lüddecke et al., 2020). In addition, several studies on previously unstudied venomous animals (Drukewitz et al., 2018; von Reumont et al., 2020,

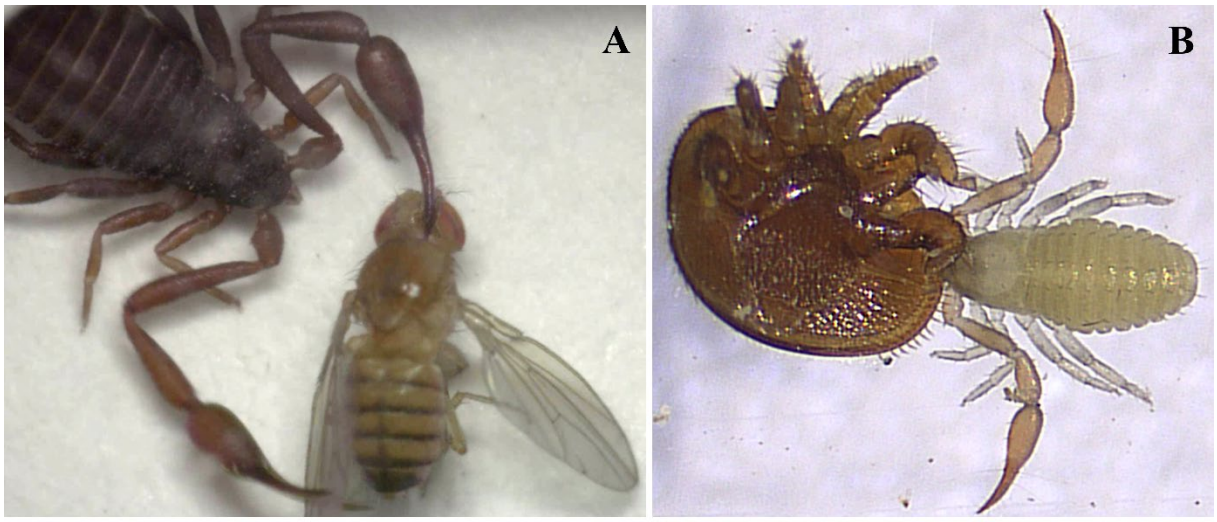

Figure 1: Images of *Chelifer cancroides* subduing its prey. **A**) Adult specimen paralyzing a fruit fly (*Drosophila melanogaster*). **B**) First instar eating a captured *Varroa destructor* (photo: Sam Read).

2014a, 2014b; Walker et al., 2018) also identified entirely novel bioactive compounds. This emphasizes the necessity to include a wide range of venomous animals into venom research in order to gain a more profound understanding of venom evolution and identify novel venom compounds with previously unknown targets or modes of action.

A very promising group for venom research are pseudoscorpions. Comprising more than 3,600 species, pseudoscorpions are more diverse than the 'true' scorpions and offer the promise of a large library of bioactive compounds. Within these small terrestrial arachnids, a unique venom delivery system has evolved independently from those of scorpions or spiders, as pseudoscorpions of the suborder Iocheirata inject venom with the pincers of their pedipalps (Chamberlin, 1924). With this venom delivery system, pseudoscorpions (adults and nymphs) are capable of subduing prey even exceeding their own body size (Fig. 1). Depending on the subgroup of Iocheirata, venom glands can be present in both fingers or be reduced in either the fixed or the movable finger of the chelal hand (Harvey, 1992). The external parts of the venom delivery system are already well described, and comprise the venom tooth with lateral pore and the *lamina defensor*, a seta closely associated with the venom tooth (Chamberlin, 1924; Krämer et al., 2019). The internal parts consist of a narrow venom canal extending proximally to form one or more tubes (depending on the species), which are presumably surrounded by glandular tissue. Studies on chemical aspects of pseudoscorpion venom are still rare. A single study examined the activity of pseudoscorpion venom by testing the effect of crude venom from *Paratemnoides nidificator* on the binding of the neurotransmitter L-Glutamate to its receptor in

rat brains (dos Santos and Coutinho-Netto, 2006). Two studies have investigated the potential venom compositions of the pseudoscorpion species *Synsphyronus apimelus* (Garypidae) and *Wyochernes asiaticus* (Chernetidae) by means of transcriptomic approaches (Lebenzon et al., 2021; Santibáñez-López et al., 2018). However, the study on *W. asiaticus* was based solely on a whole-body transcriptome. A comprehensive analysis of pseudoscorpion venom also comprising proteomics was hampered in the past by the small size of these animals, which mostly do not exceed a body length of 5mm. In a previous study, we addressed this issue by developing a venom extraction procedure for pseudoscorpions followed by a combined transcriptomic and proteomic analysis of the venom of *Chelifer cancroides* (Cheliferidae) (Krämer et al., 2019). This enabled the identification of checacins, the first genuine venom compounds of pseudoscorpions that are potential antimicrobial peptides. However, for that study proteomic data was based solely on a top-down approach, with the identification of venom compounds limited to substances below 3kDa. Therefore, the first aim of our current study was to identify the full set of major venom compounds from *C. cancroides* using a combined proteo-transcriptomic approach. Another objective was to compare the venom composition of *C. cancroides* and *S. apimelus*, which represent different pseudoscorpion families. Finally, another goal was to provide the first activity tests at the cellular level with crude venom.

The oldest known fossils of pseudoscorpions date from 360 million years ago (Harms and Dunlop, 2017). Due to their early divergence from the sister group Scorpiones (Ontano et al., 2021) and their unique venom delivery system, we expect a fairly high number of novel venom compounds in *C. cancroides*.

2. Material and Methods

2.1. Collection and rearing of pseudoscorpions

Specimens of *C. cancroides* used for proteomics were collected in North Rhine Westphalia, Germany. The methodology for collection and rearing is described in (Krämer et al., 2019), which also contains information about transcriptomics using this population. For transcriptome analyses, 31 adult specimens were collected from honeybee hives at Lincoln, Canterbury, New Zealand. These animals were collected from refuges in the hives, in which they lived for two months with access to *Varroa* mites (*Varroa destructor*), psocids, wax moth larvae and other small arthropods. After collection, the specimens were transferred into a micro-tube each and kept at 4°C.

2.2. Venom collection

Venom was collected as described in (Krämer et al., 2019). To increase the yield, venom was extracted from both fingers of one chelal hand during one extraction procedure. Each venom sample was extracted into 1µl of either Milli-Q water or ND96 buffer containing 96mM NaCl, 2mM KCl, 1.8mM CaCl2, 1mM MgCl2, and 5mM HEPES (pH = 7.4).

2.3. Quadrupole Orbitrap mass spectrometry with nanoflow HPLC

Four Quadrupole Orbitrap mass spectrometry (MS) experiments were performed, three of these were bottom-up analyses with digested samples, the fourth experiment was performed without digestion step, but with reduction/alkylation of the sample. For the three bottom-up experiments, we used venom extracted from 24, 44 and 64 specimens, respectively. For the top-down experiment, venom was extracted from 12 specimens and mixed with an equal volume of urea buffer (8M urea/50mM triethylammonium bicarbonate buffer (TEAB)) for denaturation prior to reduction/alkylation. For desalting and removal of urea, poly (styrene divinylbenzene) reverse phase (SDB-RP)-StageTip purification was performed before Orbitrap MS analyses according to the StageTip purification protocol from the CECAD Proteomics Facility, University of Cologne (http://proteomics.cecad-labs.uni-koeln.de/Protocols.955.0.html). As the protein quantity of the venom samples for bottom-up analyses was too low to be measured, the single-pot, solid-phase-enhanced sample preparation (SP3) (Hughes et al., 2019) was utilized for digestion of venom samples, which is especially useful for low concentrated samples. For the SP3 procedure, sodium dodecyl sulfate was added to the venom samples with a final concentration of 5% and reduction/alkylation were performed according to the SP3 protocol (http://proteomics.cecad-labs.uni-koeln.de/Protocols.955.0.html). The SP3 procedure with integrated trypsin/LysC digestion was performed with 0.5ug trypsin and 0.5ug LysC and a beads/protein ratio of 10:1. Afterwards, the venom compounds/tryptic peptides were separated on an EASY nanoLC 1000 UPLC system (Thermo Fisher Scientific, Bremen, Germany). For this purpose, inhouse packed RPC18-columns with a length of 50cm were used (fused silica tube with ID 50µm±3µm, OD 150µm; Reprosil 1.9µm, pore diameter 60A°; Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). The HPLC separation was performed with a binary buffer system (A: 0.1% formic acid (FA), B: 80% acetonitrile, 0.1% FA): linear gradient from 2 to 62% in 110min, 62–75% in 30min, and final washing from 75 to 95% in 6min (flow rate 250nl/min). Re-equilibration was performed with 4% B for 4min. The HPLC was coupled to a Q-Exactive Plus (Thermo Fisher Scientific) mass spectrometer. HCD

fragmentations were performed for the 10 most abundant ion signals from each survey scan in a mass range of m/z 300–3000. The resolution for full MS1 acquisition was set to 70,000 with automatic gain control target (AGC target) at 3e6 and a maximum injection time of 80ms. In order to obtain the HCD spectra, the run was performed at a resolution of 35,000, AGC target at 3e6, a maximum injection time of 240ms, and 28eV normalized collision energy; dynamic exclusion was set to 25s.

2.4. MALDI-TOF MS

For MALDI-TOF MS analysis of reduced/alkylated venom samples, 10µl diluted venom extracted from 10 specimens was used. One µl of this venom sample was mixed with an equal amount of ethanol/water/trifluoroacetic acid (TFA; 35/64.95/0.05) for MALDI-TOF MS analysis without prior reduction/alkylation. The remaining 9µl were mixed with 9µl urea buffer to ensure denaturation of the venom sample. Afterwards, the venom sample was reduced, alkylated and the urea was removed utilizing SDB-RP StageTips. For MALDI-TOF MS analysis, 0.3 µl of venom samples were directly spotted onto the sample plate for MALDI-TOF MS and mixed with the same volume of 10 mg/ml 2.5-dihydroxybenzoic acid (Sigma Aldrich, Steinheim, Germany) matrix, dissolved in 50% acetonitrile/0.05% TFA. For an optimal crystallization of the matrix, samples were blow-dried with a hairdryer. An ultrafleXtreme TOF/TOF mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany) was used in reflectron positive mode with overlapping mass ranges of m/z 800–4500 and m/z 3000−10,000. For an optimal signal-to-noise ratio, laser intensity and the number of laser shots were adjusted for each sample. Laser frequency was set to 666 Hz. For external calibration, a mixture containing proctolin ($[M+H]^+$, 649.3), Drm-sNPF-2^{12-19}, ($[M+H]^+$, 974.5), Pea-FMRFa-12 ($[M+H]^+$, 1009.5), Lom-PVK ($[M+H]^+$, 1104.6), Mas-allatotropin ($[M+H]^+$, 1486.7), Drm-IPNa ($[M+H]^+$, 1653.9), Pea-SKN ($[M+H]^+$, 2010.9), and glucagon ($[M+H]^+$, 3481.6) was used for the mass range of m/z 800–4500 and a mixture of bovine insulin ($[M+H]^+$, 5731.5), glucagon and ubiquitin ($[M+H]^+$, 8560.6) was used for the mass range of m/z 3000−10,000. Ion signals were identified by using the peak detection algorithm SNAP from the flexAnalysis 3.4 software package. In addition, each spectrum was manually checked to ensure that the monoisotopic peaks were correctly identified. MSMS experiments were conducted using Bruker LIFT™ technology without CID. Peptide sequences were identified by manual analysis of fragment ions and subsequent comparison of predicted (http://prospector.ucsf.edu) and experimentally obtained fragment patterns.

2.5. RNA extraction, transcriptome sequencing and de novo assembly of nucleotide sequences

Two transcriptomes were generated for the pedipalps, both based on the same 31 individuals of *C. cancroides*. One transcriptome is based on the chelal hands containing the venom glands, while the other (negative control without venom glands) is based on the two proximal segments (patella and femur) of the remaining pedipalps. The specimens were anesthetized by freezing prior to dissection either by keeping them at -18°C for 24h, at -80°C for 5min, or by snap freezing in liquid nitrogen. The pedipalps were severed at the trochanter and placed into a micro-tube containing 1ml fresh RNA*later*™ (Qiagen, Hilden, Germany). Total RNA was extracted using the standard TRIzol protocol (ThermoFisher). Further sample processing and sequencing were performed by the sequencing facility of the Core Research Laboratories at the Natural History Museum in London. RNA was quantified using a Qubit RNA HS Assay Kit (ThermoFisher), and quality was checked with an Agilent TapeStation with RNA ScreenTape. The sequencing library was prepared with the Illumina MiSeq V3 kit following the manufacturer's protocol. Paired end sequencing (2 x 250bp) was performed on an Illumina MiSeq machine. Raw sequences were demultiplexed and adapters were removed using the MiSeq Reporter Software v.2.6 (Illumina). De novo assembly of RNA sequence data was performed with Trinity v2.2.0 (Grabherr et al., 2011) based on default settings. This assembly and the already existing assembly from a German population (Krämer et al., 2019) were used to search for peptide sequences obtained by Quadrupole Orbitrap and MALDI-TOF MSMS experiments. To assess the completeness of transcriptomic data, BUSCO 3 (Waterhouse et al., 2018) was used. The transcriptome data of the chelal hands has been submitted to NCBI (Bioproject: PRJNA752025).

2.6. Identification of venom compounds

Precursors of potential venom compounds were identified by matching the fragment spectra of Quadrupole Orbitrap MS analyses against the chelal hand transcriptome of *C. cancroides*, utilizing the software PEAKS 10 (PEAKS Studio 10; BSI, Toronto, Canada). PEAKS was run with a parent error mass tolerance of 10ppm and 0.05Da for fragment ions. As posttranslational modifications (PTMs), carboxymethyl was set as fixed modification and acetylation (N-terminus, Lys), amidation, carbamidomethylation, carboxylation (Glu), half of a disulfide bridge, oxidation (at Met, His, Trp) and pyroglutamate from Glu were accepted as variable modifications in the analysis. For enzymatically digested samples (bottom-up analyses), enzyme mode was set to 'Trypsin and LysC'. For the sample without digestion (top-down

analyses) 'None' was selected as enzyme mode. To evaluate the significance of the hits, identified precursor sequences were each examined for the presence of a signal peptide with SignalP 5.0 (Almagro Armenteros et al., 2019) and for the presence of a stop codon. In a second step, we searched for the presence of the respective precursors in the negative control (transcriptome of the proximal pedipalp segments without venom glands). For all matches, the expression level in the transcriptomes was assessed with Kallisto (Bray et al., 2016). Another criterion for defining venom compounds was the presence of corresponding ion signals in MALDI-TOF mass spectra of venom samples. For this purpose, theoretical masses calculated for each of the potential bioactive venom peptides were searched against a list of MALDI-TOF ion signals, considering potential PTMs and cleavage sites (e.g., dibasic including quadruplet-cleavage sites; (Kozlov et al., 2005), and cleavage at the 'LEAP'-motif described for *C. cancroides* (Krämer et al., 2019)). A classification based on similarity to compounds from the online database UniProt (The UniProt Consortium, 2021) was made. For this purpose, a local BLAST search of the amino acid precursor sequences was performed against the following UniProt databases. The Metazoa database was searched with the term 'taxonomy:"Metazoa [33208]' and the Tox-Prot database with the term 'taxonomy:"Metazoa [33208]" (keyword:toxin OR annotation:(type:"tissue specificity" venom)). In the case of the Metazoa-database, an E-value of $1e^{-5}$ was used. For the search against the Tox-Prot database, an E-value of 10 was used. The matches were then classified based on the description of the BLAST hits in both databases, in case of the Tox-Prot-database only if the E-value was lower than 0.055. Then, the matches were filtered with respect to the quality of MS data, the coverage between transcriptomic and proteomic data (false discovery rate (-10lgP) > 30, coverage > 7%) and the presence of a signal peptide. Finally, all identified venom precursors were analyzed with InterProScan (Blum et al., 2021) to perform a functional annotation.

2.7. Comparison of the venom compositions of *C. cancroides* and *S. apimelus*

To find orthologous precursors of venom compounds for *C. cancroides* and *S. apimelus,* several BLAST searches were performed. First, the precursors of venom compounds identified for *C. cancroides* were searched against the chelal hand transcriptome of *S. apimelus* with an E-value of 10^{-5}. The second step was to search the proposed precursors of venom compounds described for *S. apimelus* against the chelal hand transcriptome of *C. cancroides*. Finally, a BLAST search was performed to compare the precursor sequences of identified (*C. cancroides*) and proposed (*S. apimelus*) venom compounds with each other. To evaluate the significance of the BLAST

hits, the bitscore was used, which is less dependent on database size. BLAST hits with a bitscore >40 were considered significant (Pearson, 2013).

2.8. Electrophysiological characterization of the crude venom

We followed the protocols described in detail previously (Camargos et al., 2011; Peigneur et al., 2021) For the expression of VdNav1, the auxiliary subunits TipE and the Shaker IR in Xenopus laevis oocytes, the linearized plasmids were transcribed using the T7 mMESSAGE-mMACHINE transcription kit (Ambion). In total, 50nL of cRNA (1 ng/nL) was injected into oocytes, which were incubated in ND96 solution, supplemented with 50mg/L gentamycin sulfate. Recordings were performed using a Geneclamp 500 amplifier (Molecular Devices) controlled by a pClamp data acquisition system (Axon Instruments); bath solution was ND96. Voltage and current electrodes were filled with 3 M KCl. Resistances of both electrodes were kept between 0.7 and 1.5 MΩ. Elicited currents were sampled at 1kHz and filtered at 0.5kHz (for potassium currents) or sampled at 20kHz and filtered at 2kHz (for sodium currents) using a four-pole low-pass Bessel filter. Leak subtraction was performed using a -P/4 protocol. Currents were evoked by a 100ms (Nav) or 500ms (Kv) depolarization to the voltage corresponding to the maximal activation of the channels in control conditions from a holding potential of −90mV. In general, current-voltage relationships were determined by 50-ms step depolarizations between −90 and 70mV, using 5mV increments. Toxin-induced effects on the steady-state inactivation were investigated by using a standard two-step protocol. In this protocol, 100ms conditioning 5mV step prepulses ranging from −90 to 70mV were followed by a 50ms test pulse to 0mV. For current–voltage relationship studies of Kv channels, currents were evoked by 10mV depolarization steps from -90mV to 70mV for 250s from a holding potential of -90mV. All data were obtained in at least six independent experiments ($n \geq 6$). To test the effect of crude pseudoscorpion venom on different ion channels, 2µl venom with concentrations of 2µg/µl or 4µg/µl were applied to measuring chambers containing the oocytes in 80µl ND96 buffer. For determining the concentration of the *C. cancroides* venom samples a ND1000 nanodrop was used. The concentration of the venom samples was adjusted by dilution with ND96-buffer.

3. Results

3.1. Combined transcriptomic and proteomic analysis

Next-generation sequencing of samples from the New Zealand population of *C. cancroides*, after adapter removal, yielded 16,826,640 paired-end reads for the chelal hand (pedipalp)

sample, and 8,983,558 paired-end reads for the proximal pedipalp segments sample (negative control). Sequence assemblies resulted in 80,317 contigs for the chelal hands transcriptome and 69,810 contigs for the proximal pedipalp segments transcriptome. Regarding completeness of transcriptome data, the chelal hand transcriptome contains 45.3% complete and 30.2% fragmented BUSCOs, whereas the proximal pedipalp segments transcriptome comprises 68.1% complete and 20.4% fragmented BUSCOs.

Matching of the proteomic data from four Orbitrap MS experiments (three samples with trypsin digestion, one sample without digestion) against the chelal hands transcriptome initially yielded 1270 hits. After quality filtering (coverage and P-value), removal of precursors without a signal peptide and redundant matches, 124 precursors contributing to the venom composition of *C. cancroides* were identified (Supplementary material 1). These precursors were first separated into precursors with cysteine-containing peptides and the remaining precursors. Both groups were further classified, based on sequence similarities to annotated sequences from online databases, into precursors of potential peptide toxins, antimicrobial peptides, enzymes, 'other' or, in case no matches were found within the databases, as 'novel' (Fig. 2). Where the InterProScan analysis resulted in functional annotations, this information is added in Supplementary material 1. Precursors of identified venom compounds with presumed orthology

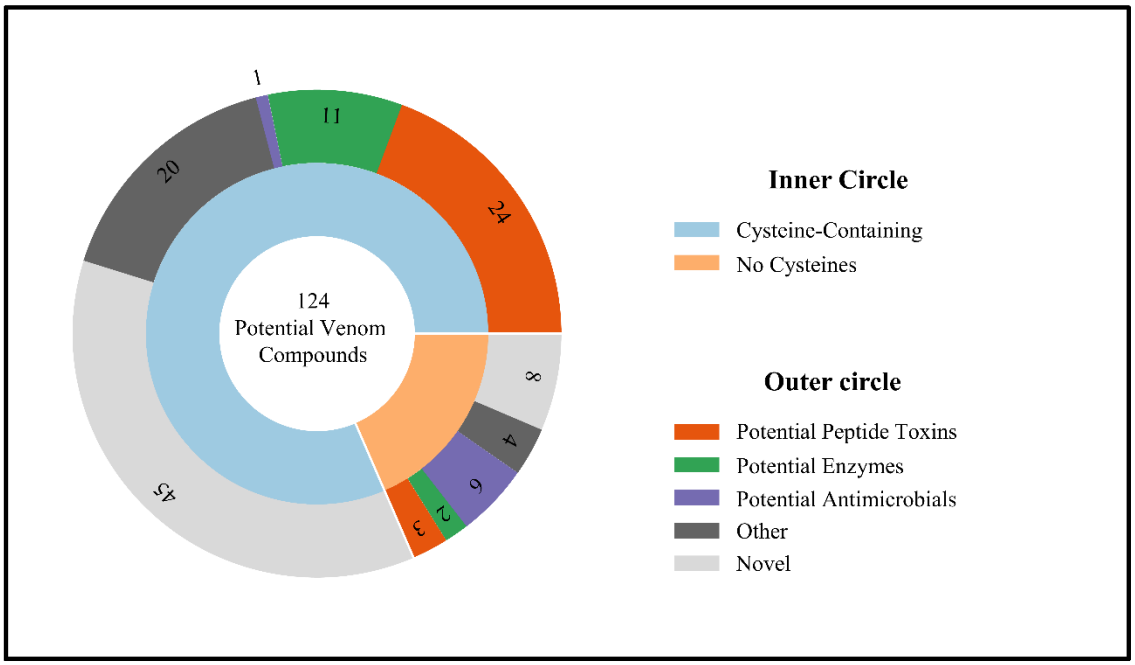

Figure 2. Pie Chart showing a classification of the venom compounds identified for *C. cancroides* by a combined transcriptomic and proteomic approach. The inner circle shows a differentiation of the venom compounds based on cysteine content. In the outer circle, venom compounds are classified based on a presumed orthology to sequences in UniProt (The UniProt Consortium, 2021) entries.

305 to annotated toxin sequences of other taxa were named according to rational nomenclature

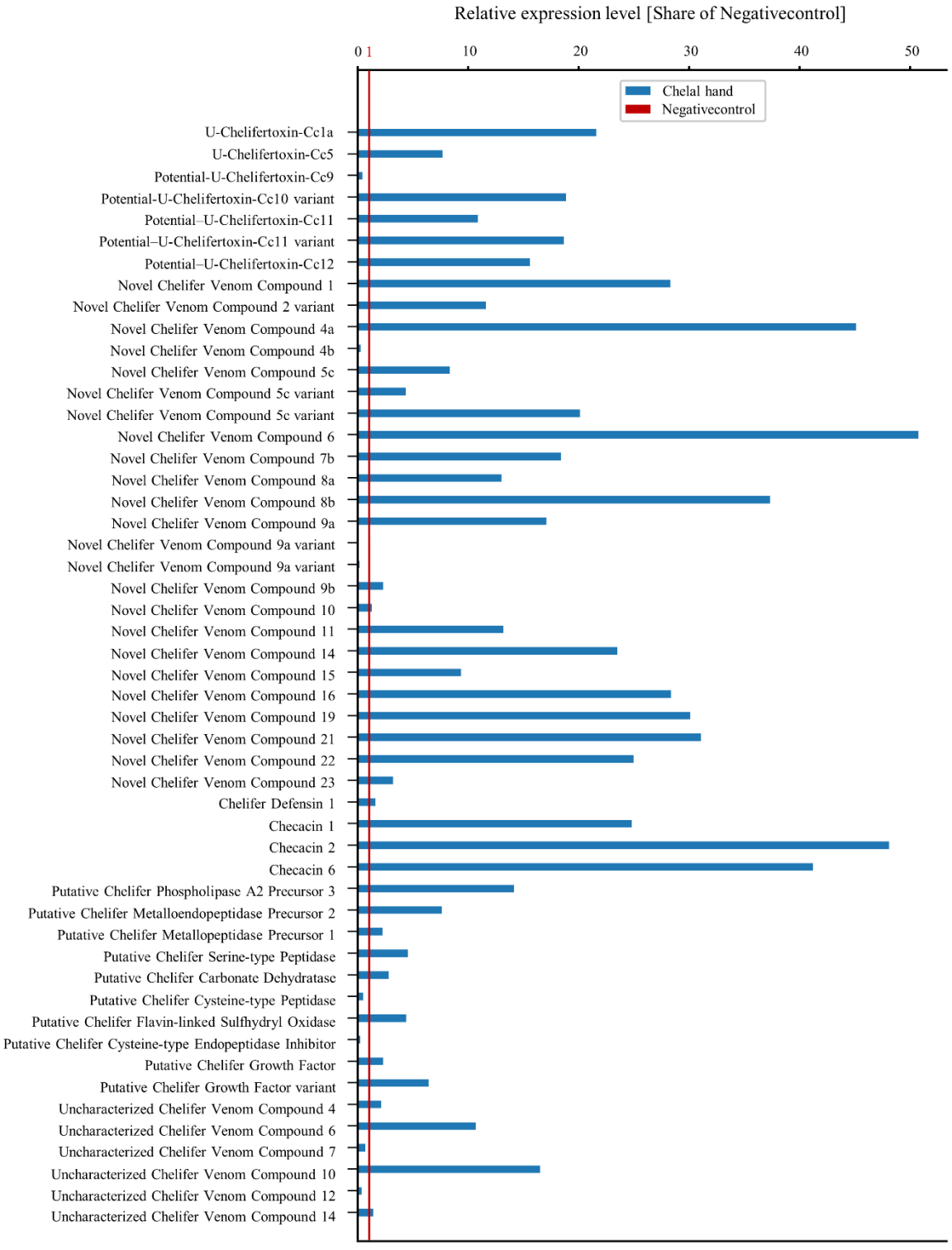

Figure 3. Bar chart showing relative expression levels of venom precursors identified in a chelal hand transcriptome of *Chelifer cancroides*. Only the values of those precursors also present in the negative control are shown. Values are presented as share of the expression levels of respective precursors from the negative control. (Transcriptome of remaining pedipalps). Precursors were identified by a combined proteo-transcriptomic approach.

306 guidelines (King et al., 2008). For the precursors that were also detected in the negative control,
307 a comparison of the expression levels of the corresponding trancripts of chelal hands and
308 remaining pedipalps (negative control) is shown in Fig. 3.
309 Information on 11 precursors, whose products have been detected in the venom samples of *C.*
310 *cancroides* and show sequence similarity to known arthropod toxins, is summarized in Table 1.
311 All of these are cysteine-rich precursors with 3-5 disulfide bonds and the biochemically
312 confirmed or predicted mature peptide toxins are in the mass range below 10 kDa. Precursor
313 genes of CHTX-Cc1a and 1b, CHTX-Cc2a and 2b as well as those of CHTX-Cc8a and 8b likely
314 represent paralogs, respectively. CHTX-Cc1a and 1b exhibit the best BLAST-matches to
315 potential arthropod toxins, the latter classified as potential potassium channel toxins. An
316 alignment of these sequences is shown in Supplementary material 2. All of the listed precursors
317 either show significantly higher expression levels in chelal hands (i.e., in venom glands)
318 compared to the remaining pedipalps or are absent in the negative control. Additional potential
319 toxin precursors can be found in Supplementary material 1.
320 Fifty-two precursors with biochemical confirmation of corresponding peptides in venom
321 samples, but without significant sequence similarity to annotated sequences in the UniProt
322 databases are classified as Novel *Chelifer* Venom Compounds (NCVCs). Of these, a selection
323 of 19 is shown in in Table 2. Among these precursors are products of several paralogous genes
324 (NCVS-4a/b/c, 7a/b, 8a/b, 9a/b), and except for NCVC-4b, all precursors showed increased
325 expression levels in the *Chelifer* hand transcriptome. Three NCVCs (NCVC-4a, 5a, 8b) showed
326 exceptionally high expression levels. Confirmed or predicted mature peptides are mostly
327 cysteine-rich, only products of the precursors of NCVC-11 and 12 represent linear peptides,
328 whereas a single disulfide bridge is present in NCVC-2 (Table 2). Both the linear peptides and
329 the mature peptides of NCVC-1, 2 are C-terminally amidated.
330 Table 3 lists precursors of 8 potentially antimicrobial peptides. In addition to the three checacin
331 precursors already described (Krämer et al., 2019), four new checacin precursors could be
332 identified. The corresponding *checacin* genes always show much higher expression levels in
333 the chelal hands transcriptome than in the proximal pedipalp segments, although the expression
334 level itself is quite different for the various *checacin* genes. All checacins are c-terminally
335 amidated linear peptides. A cysteine-rich putative antimicrobial peptide without sequence
336 similarity to checacins is named *Chelifer* defensin (Table 3). The corresponding precursor
337 shows similarity to Tddefensin, which has been identified in the transcriptome of the scorpion
338 *Tityus discrepans* (D'Suze et al., 2009). Different from the *checacin* precursor genes, the

Table 1. List of potential *C. cancroides* toxin precursors with sequence similarity to annotated toxins of arthropods. Precursors were identified by a combined proteomic and transcriptomic approach. Grey, signal peptide; blue, potential bioactive peptide; green, amidation signal; yellow, cysteine (half of the disulfide bond); red, potential cleavage site. Black underlined, confirmed by MSMS; red underlined, confirmed only by MSMS digested samples; dashed line, mass match in MALDI-TOF MS. The column Orbitrap MS indicates which of the four proteomic analyses provided confirmation of the precursor sequence: D, Bottom-up analysis with the digestion of samples with low (l, venom from 24 specimens), medium (m, venom from 44 specimens) or high (h, venom from 64 specimens) venom amount; ND, Top-down analysis without digestion. MALDI-TOF MS: Confirmation of precursor products in mass spectra of venom samples either by mass match (+) or MSMS (+*). The number of cystines (C-C) was confirmed by mass shifts of the respective ion signals in MALDI-TOF mass spectra after reduction/alkylation. Assumed PTMs include amidation (A), disulfide bridges (C-C) and modification of N-terminal glutamine to pyroglutamic acid (pQ).

Name	BLAST hit	Expression level [tpm]	Expression level negative control [tpm]	PTM	Predicted Mass [M+ H^+]	Orbitrap MS	Confirmed disulfide-bridges (MALDI)	MALDI-TOF MS
U-Chelifertoxin-Cc1a*	Potassium channel toxin alpha-KTx Tx308 (*Buthus occitanus israelis*), 30%, Acc: B8XH30 MKCYLFVILLVVCAIGMDSVQGQKWACENGGAECDKMCRSIGKMGACSPGGPGVLLCRCI	2659	123	pQ,C-C	3905.75	D (h, m), ND	3	+*
U-Chelifertoxin-Cc1b*	Potassium channel toxin alpha-KTx 1.16 (*Mesobuthus eupeus*), 28%, Acc: C0HJQ8 MKCYLFIILLVVCAIGMDSVQGQKWRCDNGGEECYKMCRRIGKVGECSPGGPGVPLCRCI	716	-	pQ,C-C	4161.91	D (m), ND	3	+*
U-Chelifertoxin-Cc2a*	Toxin CSTX-17 (*Cupiennius salei*), 45%, Acc: B3EWT2 MSRLILFLCFSVLVMVSLAMAEDTPGEESEHISKRACVPDYGKCKQNGIKKNNCCNKVSCYCNLTFTNCYCKPPLFGK	2133	-	A,C-C	4616.02	D (l, m), ND	4	+
U-Chelifertoxin-Cc2b*	Toxin CSTX-17 (*Cupiennius salei*), 41%, Acc: B3EWT2 MSRLILFLCFSALVMVSLAMAEDTPGEEPEQISKRACVPDYGKCKWTSKGKKNNCCNDVSCYCNLSLTDCYCNPPIFG	54	-	A,C-C	4689.02	D (l, m)	4	+

U-Chelifertoxin-Cc3	Putative neurotoxin LTDF S-06 (*Dolomedes fimbriatus*), 34%, Acc: A0A0K1D8H2 MSKLIFALLFSGLVLASLVMAEEEEEETLEISKRSCIKEYGTCQWKGLGAKSQCCDNRNCVCNIALNNCKCKPSPSQLLAKVFG	679	-	A,C-C	5357.46	D (h, l, m), ND	4	+
U-Chelifertoxin-Cc4*	U28-Sparatoxin-Hju1n (*Heteropoda jugulans*), 41%, Acc: A0A4Q8KD95 MKVAFFVFLVVLSAAALAKAIEDGQEENMEISKRDTCLAVGDNCQGNTGKCCDGAKCVCRKDFILGFSGSHIITRCNCKK	13	-	C-C	4874.14	D (h, l, m), ND	4	-
U-Chelifertoxin-Cc5	Putative neurotoxin-H (*Lychasmucronatus*), 30%, Acc: D9U2B4 MAAVEMGRASWILAVLVLTAVFWTCEADALCDKGAETCNLSCYRKSYQLVGYCDRNRDGKTHCRCMKKSDASLIGR	84	10	A,C-C	5311.4	D (h, l, m)	3	-
U-Chelifertoxin-Cc6	U20-Liphistoxin-Lsp1a (*Liphistius* sp.), 43%, Acc: A0A4Q8K5N5 MWRCWWTVLLLWLVAEARYATWADFEAAHGRRPPQARALAACARAGPARDLCERCAKVTRSEVVFPFCCDDTRDVRAWCERFLDFGLQNL	23	-	C-C	8294.96	D (h, l, m), ND	3	-
U-Chelifertoxin-Cc7	U68-Liphistoxin-Lsp1a (*Liphistius* sp.), 30%, Acc: A0A4Q8K539 LTIVLALVILAVVAEAERKCFIHRRDCSKDECCAGVGIVGVCKKLAQAGEKCRIIDSFDCPCAKGLECLPFGIIRGICFKKKDETPAQDLAE	7	-	C-C	8351.00	D (h, m)	5	-
U-Chelifertoxin-Cc8a	Kappa-Theraphotoxin-Ct1a_1 (*Coremiocnemis tropix*), 37%, Acc: A0A482Z9G0 MYKFSVIFLLAAAVILVAAEYDDEDGRRYLATEKRSCSISKCNIQECCPGYVCRKGARHSSGSVCVNSG	149	-	C-C, A	3455.54	D (l, h), ND	3	-
U-Chelifertoxin-Cc8b*	Kappa-Theraphotoxin-Ct1a_1 (*Coremiocnemis tropix*), 37%, Acc: A0A482Z9G0	219	-	C-C	3930.85	D(l)	3	-

	MYKFSVIFLLAAAVILVAAEYDDEDGRRYLATEKRSCSISKCNIQECCPGYVCRKGALRSSGSVCVDSGLTIL

Table 2. List of selected precursors of novel venom compounds identified in the venom of *C. cancroides*. Listed precursors either show a corresponding MALDI-signal or were considered most evident after manual inspection of PEAKs-results (matches between proteomic and transcriptomic data). Additional precursors of novel venom compounds can be found in Supplementary 1. For further explanations see Table 1.

Name	Expression level [tpm]	Expression level negativecontrol [tpm]	PTM	Predicted Mass [M+ H^+]	OrbitrapMS	Confirmed disulfide-bridges (MALDI)	MALDI-TOF MS
Novel*Chelifer* Venom Compound 1	680	24	A,C-C	5124.19	D (h, l, m), ND	3	+
	MKTFCLALLLVGVLAGVMETEAVVAGCPDESKCHAWCLSQFPKYQAVTTGFCVNSNRCACHVDTNEDPTGK						
Novel*Chelifer* Venom Compound 2*	2150	-	A,C-C	2672.40	D (h, m), ND	1	+*
	MKTFVVLFFGAVLLAFAAADIENEAALESEMLDLESDLAELLEAPSPIGILQCLGRKDTTWKECLNKNNKGK						
Novel*Chelifer* Venom Compound 3	376	-	C-C	2659.15	ND	2	+
	MSRLLVVLVVAAVVLTAVVSVEAETESEVMDESTVEESPECVCNPPESTCCFAKGQVYNKNKT						
Novel*Chelifer* Venom Compound 4a	26718	592	C-C	7924.93	D (h, l, m), ND	3	+
	MKYVALSLALVLCLAVLARAEDQGVQDGDVCIIDRVLGEIKCIGKGINKIYKSIFKSYQKCKEFCKQYEAQGYKCKQKGISDYKCTNKK						
Novel*Chelifer* Venom Compound 4b	8	28	C-C	7678.77	D (h, l, m), ND	3	+
	MKTWFYLAAVAAMLTLATRAEEDPPEGGKCIIDAVLDEIKCIGKAINKVYKNKFTSYQKCVKFCKDYEAKGFKCKSKGPLSDYKCTDK						
Novel*Chelifer* Venom Compound 4c	3775	-	C-C	7940.01	D (h, l, m), ND	2	+
	MNSCALFLLVVLSLCALSWAEEEKKKTVLDKVGEELKKVGQGMKDIYNNIYKSYNKCKDFCKQYESKGYTCQKKLLSVSDYKCAPKKP						

Novel*Chelifer* Venom Compound 5a*	11672	-	C-C	9409.56	D (h, l, m), ND	3	+
	MKVAVSLLCLLLAAVLAAVSCTADQHVQDEQELESPDGIVDWLKKELGDRADSIYKGTMRNPITKVYGKYQKCQEECKNKPDKRCKCQMMRFLKKMEDKQEHKCMCISLLDHSMD						
Novel *Chelifer* Venom Compound 5b*	145	-	C-C	9392.6	D (h, l, m)	3	-
	MKVAVSLLCLLLAAVSCTADQLVQDEQELESPDVVLDWFKKEVGSRAESIYKGTMRNPVTKVYGKYQKCQEECKNKPDKRCKCQLSKFFRSMENKQEHKCMCISLLDRSMD						
Novel *Chelifer* Venom Compound 5c*	1038	124	C-C	9276.71	D (h, l, m)	3	-
	MKVAVSLLCLLLAAVSCTADQLVQDEQELESPDGVLDWLKKEIGDRAEAIYKGVLRNPVTKVYNKYLKCQDECKGKPDKRCKCQLSRFLKPYSEKQEHKCMCVDLLDKSFD						
Novel *Chelifer* Venom Compound 6	3069	60	C-C	6244.33	D (h, l, m)	3	+
	MKYLQIVCLLLALTVFASAFQQEEEELETELDELDTPGWGKLFGVIKKGARFVLKRGQKLMRNRKKCRAQCKDPAFHCKCDPISTKCKCVAN						
Novel *Chelifer*Venom Compound 7a	288	-	A, C-C	7938.98	D (h, l, m), ND	2	-
	MNMKILKILIIGLTITLNLLCSSNAADLQEDEGNTENEALPSFESYPVYDLSKGKPEKCPEGMGFYNGKCHKLHCAIPGYVLKDKKCVRKPKTLLK*G*						
Novel *Chelifer* Venom Compound 7b	611	33	A, C-C	8388.23	D (h, l, m)	2	-
	MNMKILKILIISLIITLNLVCSSNAAELQEEEGNTEYEALPSFAILLDMNPEHGKSVGKYEKCPEGTGRFNGECRILNCGIPGYVLKGDKCVPKRRRILK*G*						
Novel *Chelifer*Venom Compound 8a	297	22	C-C	9508.53	D (h, l, m)	3	-
	MRAAIVLGLLLAVALETTAASYLEAEDGGMLAWMRKELGDKSARLYGMFADPIKKVYGKYVKCQEECRGQADKRCRCQLLKALYSKENQQAHKCMCVSMLSE						
Novel *Chelifer*Venom Compound 8b	33315	892	C-C	10581.28	D (h, l, m), ND	3	-
	MKSQLLVLCLCLAVAAAELQDLEKPDELEAGESEATYMQWITGEVGAKWAKAVYMAIKGQVVKVYQKYKTCQTTCTAPDKRCKCQLLRFLKPLSKKQEHKCMCKGMLEE						
	4420	258	C-C	10181.11	D (h, l, m)	3	-

Novel *Chelifer* Venom Compound 9a*	MNGLQWTVTALCLALFSASVQAAHVQDDQELEVGVTEAALLAWVTYEVGPKFAKAVYNALTGQIKKVYDKYLACQATCTAPDMRCNCQL LRFLYPMEKKQEKKCMCLNIKKN						
Novel *Chelifer* Venom Compound 9b	70	30	C-C	10140.9	D (h, l, m)	4	-
	MNGLQWTVTALCLALFSASVQAAHVQDDQELEVGVTEGALIAWITAEVCPKVAKAVYEALSGQITKVYDKYKACQATCTGANMKCHCQL LRFLKPMDKKQEHKCMCKNSNETSC						
Novel *Chelifer* Venom Compound 10	349	258	C-C	13085.92	D (h, l, m)	4	-
	MLLLVCALLVVAATGTSAQSCEVEKEWDIQQVFCNMKDGEKQFATCEEMMPDEAKEMIKNCNTAAQGENPSQTAFQYSCADCASLQKLK QCLSDNSMDQKIKNMNKSDQEKMAKSVVCVMSLYKKETGKDMEL						
Novel *Chelifer* Venom Compound 11	1000	75	A	3870.16	D (h, l, n); ND	-	+
	MKTLALVVCGLAVLLVASAEQEDSELGSYASDSLQSPLDEMMNQYANEDEESLSLESDLAYQWEQMELESPFWKKMKSFFKDKVIPKVQ QAYSLYNKLQHKLG						
Novel *Chelifer* Venom Compound 12	559	79	A	959.52	-	-	+
	MKNLLLCLFIFGLVLSNGVAFEDGNSVDELLESWAEESWVQEEKMPLESPGKPFQPMRG						

352

353 **Table 3.** List of precursors of potential antimicrobial peptides identified in the venom of *C. cancroides*. For further explanations see Table 1.

Name	BLAST hit	Expression level [tpm]	Expression level negative control [tpm]	PTM	Predicted Mass [M+ H^+]	Orbitrap MS	MALDI-TOF MS
Checacin 1	Megicin-18 (*Mesobuthus gibbosus*); 47.2%; Acc: A0A059U8Y9	3316	133	A	2937.76	D (h, l, n); ND	+*
	MKYLQIVCLVLSLAVLTSAFPMEEQLSESELKELEAPFFGAIAKLAMKFLPAIYKQIQKKRKGRSLEAQ						
Checacin 2	-	4563	94	A	2757.72	D (h, l, n); ND	+
	MKYIQVVCLVLSMAVFTSAFEVEDLTESELQELEAPFVGLLAKLAAYVIPQIVKRFQKKKGKRSLEWEDDDA						
	-	1563	55	A	2652.63	-	+

Checacin 3	MKYIQVVCLVLSMAVFTSA**FQVEELTESELQELEAPFIGIMATLASLVIPKLIEKIKQARGRRSLEEDELFF**						
Checacin 4	-	341	-	A	2755.77	D (h, l, n)	+
	MKYIQVVCLVLSMAVFTSA**FQVEDLTESELQELEAPFVGLLARLAAFVIPQIVKRFQKKNGKRSLEWEEE**						
Checacin 5	-	2459	-	A	2967.75	D (h, l, n); ND	+
	MKYLQFVCLLLSLAVFTSA**FQVEEELSESELKELEAPFFGVIAKMAMKFLPAIFKQIQKKRKGRSLEDQ**						
Checacin 6	-	146	3	A	5459.89	D (l, n); ND	-
	MKYLQIVCLVISLAVLASS**FPLEEQLTESDLNELESLWRGTTHFVHQYKIMPFRKLVFKRRKNGRRG**						
Checacin 7	-	717	-	A	2630.59	ND	+
	MKYIQVVCLVLSMAVFTSA**FQVEELTESELQELEAPFFGAFAAIASLVIPKLIEKIKQARGRRSLEDEEFVF**						
Chelifer Defensin 1	Tddefensin (*Tityusdiscrepans*); 57%; Acc: P0CF77	34	21	A, C-C	4570.86	D (h, l, n); ND	-
	MKLLGVVCLSALLLCLGFHMAEA**ISGANGCPMNEGRCEDHCMRRGRPGGHCGGSMRRSCICDSNLP**						

354

transcription level of the precursor gene for *Chelifer defensin* is not much higher in the chelal hand transcriptome compared to the transcriptome of the proximal pedipalp segments. Thirteen precursors biochemically identified in the venom of *C. cancroides* show similarity to enzymes from the UniProt database (Supplementary material 1). Eleven of these could be functionally annotated based on interProScan results and were named accordingly. Most of these precursors are either absent from the proximal pedipalp segments or show higher expression levels in the venom gland transcriptome. An exception is the putative *Chelifer* cysteine-type peptidase. Two of the precursors (*Chelifer* phospholipase A2 precursor 3 and *Chelifer* metalloendopeptidase precursor 1) show particularly high expression levels in the chelal hand transcriptome compared to the remaining enzyme precursors. For *Chelifer* metalloendopeptidase precursor 1, an alignment with sequences from the closest BLAST hits is shown in Supplementary material 2. Supplementary material 1 also includes 21 precursors classified as 'Other'. Three of these could be functionally annotated as cysteine-type endopeptidase inhibitors and growth factors based on interProScan results, and the precursors were named accordingly. The remaining precursors were classified as Uncharacterized *Chelifer* Venom Compounds. None of these precursors exhibit high expression levels in the chelal hand transcriptome.

3.2. Allocation of venom compounds to signals in MALDI-TOF MS

MALDI-TOF mass spectra allow rapid screening of venom compounds released from individual venom glands and provide sufficient information on (1) the relative abundance of venom compounds, (2) mature (main) products of the various precursors that contribute to the

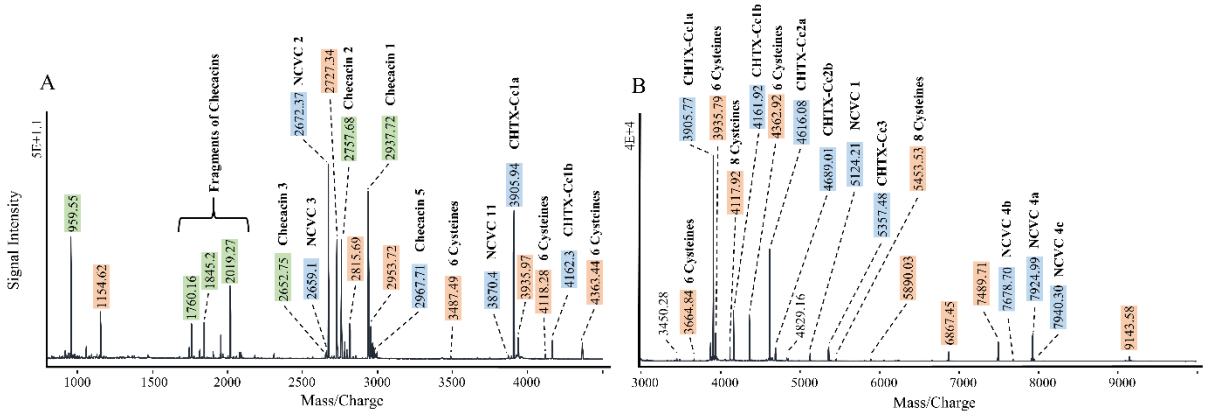

Figure 4. MALDI-TOF mass fingerprinting of venom samples of *C. cancroides* highlighting venom compounds that can be correlated with our transcriptomic and Orbitrap MS data. Ion signals highlighted in green represent venom compounds identified in Krämer et al. (2019). Ion signals highlighted in blue are venom compounds identified with their precursors in the current study. Remaining signals are marked in red. **A**) Lower mass range (m/z 800-4500). **B**) Higher mass range (m/z 3000-10,000).

venom composition, (3) changes in the venom composition over time, and (4) the completeness of precursors already described, i.e., what is the percentage of ion signals in mass spectra that can be assigned to the known precursors. As shown in Fig. 4, most of the prominent ion signals in the mass spectra (m/z 900 – 10,000) are products of precursors with particularly high expression levels (see Tables 1 – 3). Of the 11 precursors with assumed orthology to known arthropod toxins (Table 1), a number of ion signals were identified in MALDI-TOF mass spectra of venom samples that were mass-identical to products of precursors for U-chelifertoxin-Cc1a, Cc1b, Cc2a, Cc2b, and Cc3 (Fig. 4), and subsequent analyses of reduced/alkylated venom samples confirmed the expected number of disulfide bonds for these substances (Fig. 5). The mature U-chelifertoxins Cc-1a and Cc1b each comprise the complete precursor sequence without a signal peptide and are N-terminally blocked by pyroglutamate. The sequence of both peptides could be confirmed by MALDI-TOF MSMS fragmentations (Supplementary Material S3), in addition to the Orbitrap MSMS analyses. The precursors of U-chelifertoxin-Cc2a, 2b, and 3 each contain an internal Arg-Lys cleavage signal which is efficiently used to cleave an N-terminal precursor peptide from the mature C-terminally amidated U-chelifertoxins.

For three of the 19 precursors representing potential novel venom precursors (including 7 paralogs; Table 2), the predicted masses of the mature peptides are above the analyzed mass range. Mass matches were found for products of eight of the remaining 13 precursors (Fig. 4). Due to the loss of long-chain peptides during Stage-Tip purification of reduced/alkylated samples, the number of disulfide bonds could not be confirmed for most of these relatively large peptides. Confirmed disulfide bonds were, however, obtained for NCVC-1 and 2 (Fig. 5). The latter peptides are both amidated. While NCVC-1 comprises the complete precursor sequence (without signal peptide) downstream to the C-terminal Gly-Lys motif, mature NCVC-2 represents only the C-terminal sequence of the corresponding precursor (Table 2). The NCVC-2 precursor contains the LEAP cleavage motif described for *C. cancroides* checacin precursors (Krämer et al., 2019), but mature NCVC-2 is cleaved two amino acids C-terminally from that motif, i.e., C-terminally from Ser-Pro (Table 2).

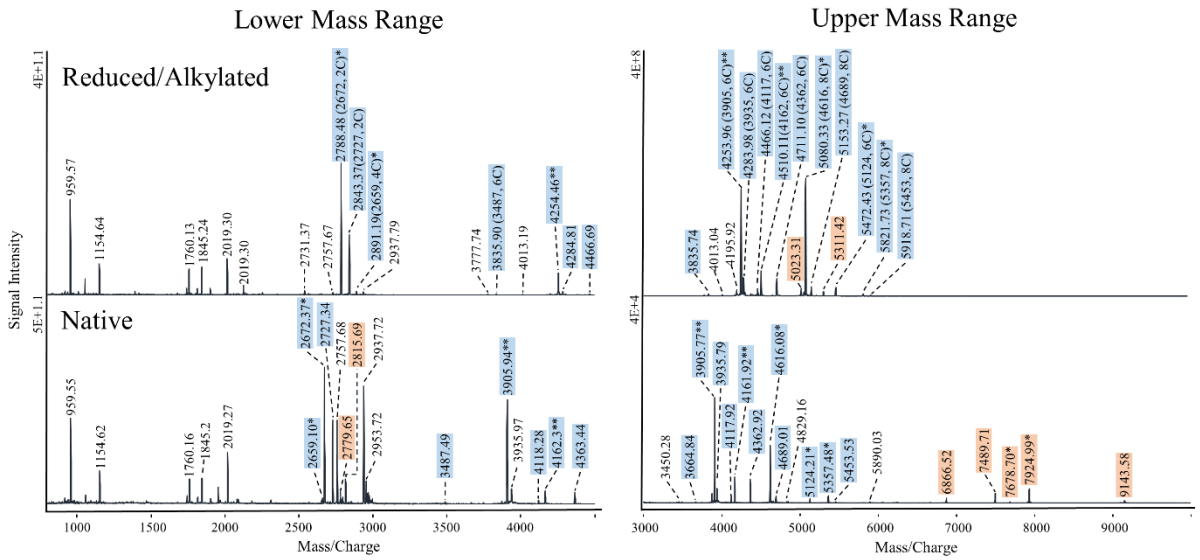

Figure 5. Comparison of MALDI-TOF mass spectra of untreated and reduced/alkylated venom samples from *C. cancroides*. Ion signals suggesting a mass shift due to reduction/alkylation are highlighted in blue. For these ion signals, the original mass and the proposed number of cysteines is added in brackets. Ion signals without corresponding signal in untreated samples are highlighted in beige. All ion signals mass-identical to venom compounds identified by our combined transcriptomic and proteomic data are marked with '*'. In case the sequences of these peptides could be confirmed in the same samples by MALDI-TOF MSMS, they are labeled with '**'.

The expression level of the paralogous *checacin* genes is highly different (Table 3), and this is also reflected in the ion signal intensity of the mature checacins (Fig. 5). Overall, ion signals mass-identical with checacins of six checacin precursors, always amidated at the C-terminus, were detected in the MALDI-TOF mass spectra (Fig. 4). These checacins are all N-terminally cleaved at the LEAP motif within the precursor sequence, and start with Phe as the N-terminal amino acid (Table 3). Truncated checacins cleaved predominantly at internal Lys or Arg-Lys were occasionally detected in the mass spectra, but with much lower signal intensity than those of full-length checacins. The lowest expression level was found for the *checacin* 6 gene, and the ion signals predicted for checacin 6 were not detectable at all. The precursor of checacin 6 does not contain an internal LEAP motif and the predicted mature checacin 6 therefore potentially contains a much longer N-terminus (Table 3).

All experimental data considered, the MALDI-TOF mass spectra suggest that a majority of the more enriched venom compounds of *C. cancroides* are identified for the mass range examined (Fig. 4).

3.3. Comparing the venom compositions of *C. cancroides* and *S. apimelus*

A recent study described potential venom precursors of the pseudoscorpion *S. apimelus* based on a chelal hand transcriptome and a BLAST search within this dataset using known toxins and other venom compounds from arthropods (Santibáñez-López et al., 2018). Are *C. cancroides* orthologs of these proposed venom precursor genes from *S. apimelus* responsible for the composition of the venom in *C. cancroides*? Supplementary material 4 lists BLAST results of searching the potential venom precursors described for *S. apimelus* in the chelal hand transcriptome of *C. cancroides* on the one hand, and in the venom compounds biochemically identified for *C. cancroides* on the other hand. Most of the potential *S. apimelus* venom precursors have a corresponding BLAST hit in the chelal hand transcriptome of *C. cancroides*. However, a majority of the predicted *S. apimelus* venom precursors did not show significant BLAST hits with the precursors that contribute substantially to the venom compounds biochemically identified here for *C. cancroides*. For example, many of the putative U8-agatoxin-like peptides described for *S. apimelus* yielded significant BLAST hits in the chelal hand transcriptome of *C. cancroides* (in some cases >90% sequence identity), though products of the corresponding *C. cancroides* genes were not found in our MS datasets. This suggests that these genes are not specifically expressed in the venom glands.

The results of a BLAST search in the chelal hand transcriptome of *S. apimelus* with those *C. cancroides* precursors whose products are specifically enriched in venom samples of *C. cancroides* are included in Supplementary material 1. It is noteworthy that some of the prominent *Chelifer* venom compounds with presumed orthology to arthropod toxins (e.g., CHTX-Cc1 and 2) did not yield significant matches in the *S. apimelus* transcriptome or resulted in BLAST hits with only moderate sequence similarity (high E-values). Only four of these compounds (see Table 1) exhibit significant similarity to venom precursors described for *S. apimelus* (CHTX-Cc 2a, 2b, 3,7). In addition, most of the NCVCs identified in this study (see Table 2) did not yield significant BLAST-hits in the *S. apimelus* transcriptome, exceptions being e.g., NCVC 7a, 7b and 11. For the checacins (see Table 3), which belong to the most abundant venom compounds of *C. cancroides*, and are provisionally grouped among antimicrobial peptides due to a moderate similarity to scorpion megicin (Diego-García et al., 2014), no BLAST hits were identified in the *S. apimelus* transcriptome. The BLAST-hits with the highest similarity in both species were found for metalloprotease and phospholipase precursors (Supplementary material 4).

3.4. Electrophysiological characterization of the crude venom

In control conditions, the conductance over the membrane of the oocytes remains minimal over a large voltage range (Fig. 6A). At a concentration of 0.1µg/µl, an outwardly rectifying venom-dependent conductance was observed in non-injected oocytes, which can be interpreted as evidence for a pore-forming activity or cytolysis as induced by the *C. cancroides* venom (i.e. induction of 'leaky cells'). As the cytolytic damage of the oocytes at high concentrations makes it impossible to investigate potential ion channel modulation, we tested the venom at a concentration of 0.05µg/µl on insect voltage-gated sodium (Nav) channels and voltage-gated potassium (Kv) channels. At this concentration, crude venom modulates the Nav channels from the mite *V. destructor* (VdNav). In the presence of 0.05µg/µl venom an increase of the sodium peak current and a slowing down of the inactivation could be observed, resulting in sustained currents (Fig. 6B). A small but significant shift of the midpoint of activation was noted with $V_{1/2}$ values of -33.9 ± 0.1mV and -36.1 ± 0.2mV in control and venom conditions, respectively. The $V_{1/2}$ of inactivation shifted from -55.9 ± 0.17mV in control to -58.6 ± 0.2mV in the presence of the venom (Figure 6C). In the presence of 0.05µg/µl *C. cancroides* venom an inhibition of Kv channels from *Drosophila melanogaster* (*Shaker* IR) occurred with 42.3 ± 2.5% (Fig. 6D). At this concentration, no modulation of the activation was observed since the $V_{1/2}$ values yielded 17.9 ± 2.0mV in control and 16.1 ± 1.6mV after application of venom (Fig. 6E).

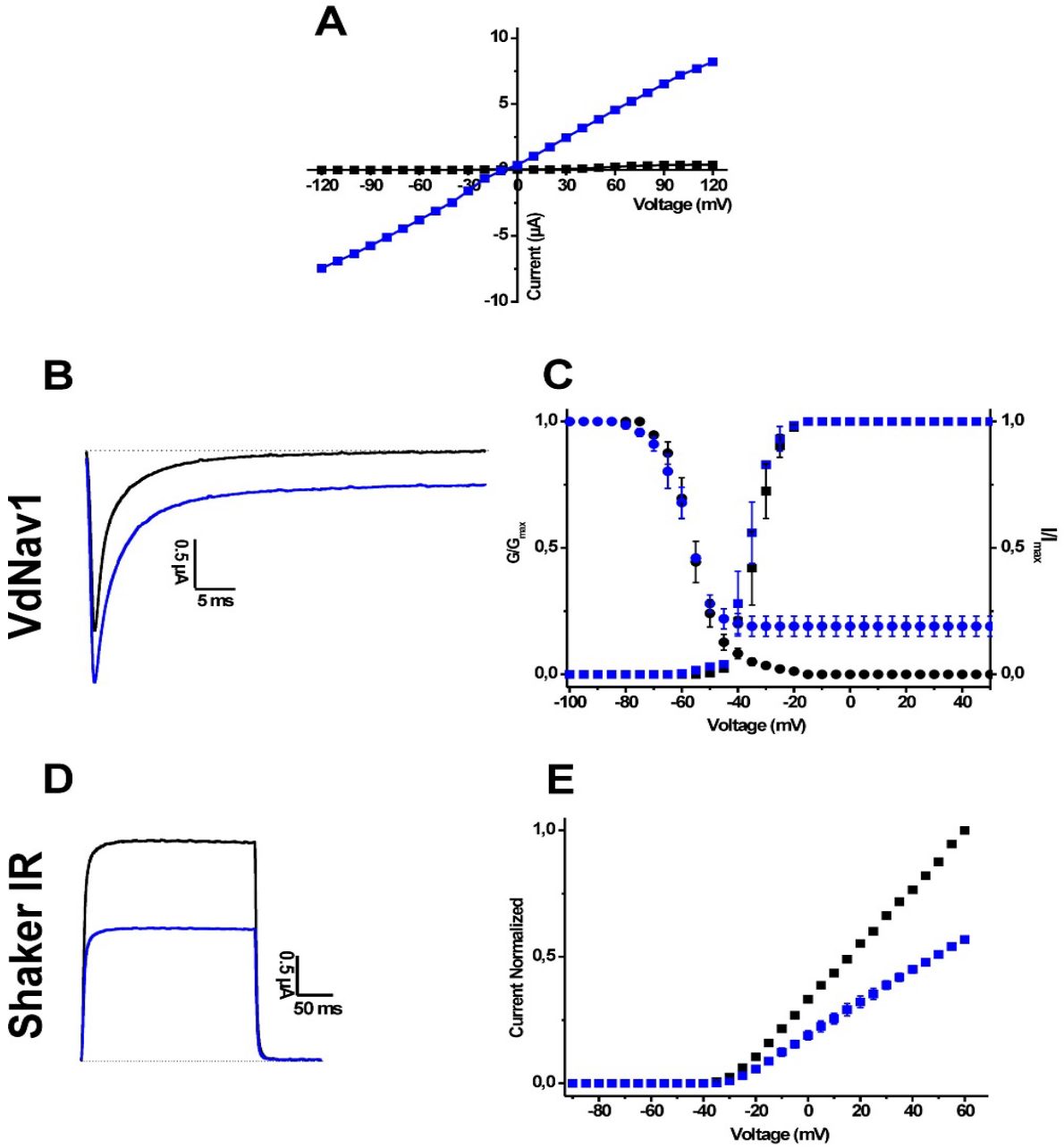

Figure 6. Electrophysiological profiling of *Chelifer cancroides* total venom. **A**) In the absence of the venom, the conductance over the membrane of the oocytes remains minimal over a large voltage range, which is indicative for a healthy cell (control, black symbols). In the presence of higher concentrations of venom (> 0.05µg/µl) an outwardly rectifying venom-dependent conductance in non-injected oocytes (blue symbols) was observed. **B**) Whole-cell current traces were recorded from *Xenopus laevis* oocytes expressing cloned VdNav1 in control or 0.05µg/µl venom. The dotted line indicates the zero current level. Blue traces were recorded after the application of venom. **C**) steady-state activation (square symbols) and inactivation (circle symbols) curves in control conditions (black) and in the presence of venom (blue). **D**) representative whole-cell current through *Shaker* IR channels in control (black) and in the presence of 0.05µg/µl venom (blue). **E**) Current-voltage dependencies of *Shaker* IR. Black symbols, control; blue symbols, after application of venom.

473

4. Discussion

The present study provides the first comprehensive analysis on the composition of a pseudoscorpion's venom based on a combined transcriptomic and proteomic approach. To obtain information on real venom compounds, we performed proteomics analyses of venom samples. Peptides were considered to be venom-specific if they could be identified by MS in venom samples and the corresponding precursors showed higher expression levels in the chelal hand transcriptome compared to the transcriptome of the proximal pedipalp segments. In the case of the rather tiny pseudoscorpions, venom analysis is complicated by two factors in particular. First, milking the crude venom was hampered in the past mainly because handling these small animals is challenging. This problem has been solved in previous experiments by developing a sophisticated extraction methodology for pseudoscorpions (Krämer et al. 2019). Second, the very low volume of venom released per milking (estimated to be 5 nl; see Krämer et al. 2019) requires multiple venom extractions to obtain a sufficient amount for biochemical analyses. We found that venom from approximately 50 specimens of *C. cancroides* was sufficient (two venom samples per specimen) to perform bottom-up analyses including reduction/alkylation/digestion, and subsequent Orbitrap MS analyses. The MALDI-TOF MS required only single venom samples. As verified in parallel experiments, analyzing the venom of, for example, 64 instead of 44 individuals did not significantly increase the number of identified peptides anymore. Therefore, we may have identified a large proportion of the more prominent venom compounds. This is supported by the high number of identified venom compounds represented by their ion signals in the MALDI-TOF mass spectra of venom samples.

In total, peptides from more than 124 precursors were identified in the venom of *C. cancroides*. One hundred and seventeen of the corresponding genes were found with higher expression levels in the transcriptome of chelal hands compared to the transcriptome of the proximal pedipalp segments. Mature peptides derived from these precursors show few PTMs, among them disulfide-bonds in most of the venom peptides, except for the checacins and C-terminal amidations. C-terminal amidation delays proteolytic degradation by exopeptidases; N-terminal pyroglutamate formation, which has been demonstrated for CHTX-Cc1, has a similar effect. The low number of PTMs is consistent with findings on spider and scorpion venom for which only disulfide bonds and C-terminal amidation are more frequent (Delgado-Prudencio et al., 2019; King and Hardy, 2013). The precursors derived from these genes were provisionally

grouped into putative orthologs of known arthropod toxins, precursors of antimicrobial peptides, enzyme precursors, and novel precursors without significant similarity to known venom precursors of arthropods. Many of the precursors listed as orthologs of known arthropods toxins exhibit only moderate sequence similarity to their respective arthropod precursors, and there is a smooth transition to precursors listed as novel, i.e., those precursors without significant orthology to known precursors of arthropods. Nevertheless, all orthologs of arthropod toxins derived from these *C. cancroides* precursors are stabilized by disulfide bridges, and some of these cysteine-rich peptides (e.g., CHTX-Cc 2a, 2b, 3 and 4) display toxin-specific cysteine patterns (Extra Structural Motif (ESM) and Principal Structural Motif (PSM)) typical for ion channel toxins from spider venom (Kozlov et al., 2005). Such cysteine-rich toxins often exhibit ICK motifs as known from peptide toxins in, e.g., spiders (Langenegger et al., 2019) and scorpions. The corresponding peptides are named knottins and some knottins are known to block e.g. potassium channels with high specificity (e.g., Kuzmenkov et al., 2018). It is probable that at least some of the here-identified venom compounds also exhibit the ICK-motif.

Another group of peptides recently described from *C. cancroides* is the checacins (Krämer et al. 2019). These linear peptides have been classified as antimicrobial peptides based on their similarity to megicin, an antimicrobial peptide from the venom of *Mesobuthus gibbosus* (Diego-García et al., 2014). Checacins are characterized by relatively high net charges and nonpolar amino acids at the N-terminus. This led to the assumption of membrane disruption through pore formation as a potential mode of action of these peptides (Langenegger et al., 2019). Depending on their charge, such peptides can act not only on bacterial membranes/cell walls, but also on those of potential prey. As typical for non-selective toxins, checacins were highly abundant in the venom samples of *C. cancroides*. Another putative antimicrobial peptide identified in our study shows similarity to the defensin family. Defensins are disulfide-rich cationic peptides that are already known to be present in venoms (e.g., scorpion venoms; (Zhu and Tytgat, 2004)), and possibly protect the venom compounds against a wide range of bacteria (Shafee et al., 2017). However, compared to checacins, *Chelifer* defensin is not particularly enriched in venom. The transcriptome data also suggest a low expression level of *Chelifer* defensin in the chelal hands, which is only slightly higher compared to the proximal pedipalp segments. In the present study, we identified four additional checacin precursors, bringing the total number to seven. The expression level of the *checacin* genes is very variable in *C. cancroides*, but mass matches were found in MALDI-TOF mass spectra for six of the seven predicted mature checacins. All checacins are C-terminally amidated, and the six checacins detectable in

MALDI-TOF mass spectra are cleaved from the N-terminal propeptide downstream of a highly conserved LEAP motif. An identical cleavage motif was also observed in the precursor of NCVC-2. However, the mature NCVC-2 peptide is cleaved further downstream, C-terminally from LEAPSP. The Ser-Pro motif appears to function as a cleavage signal even in the absence of the preceding LEAP motif, at least our data on NCVC-3 suggest this. Precursors of NCVC-5 and 6 contain modified LEAP motifs (LESP, LDTP) that appear to efficiently separate the N-terminal propeptides from the cysteine-rich peptides. Other observed cleavage signals that can be attributed to regular intracellular proprotein convertases (Benjannet et al., 1991) include dibasic (mostly Arg-Lys) and quadruplet motifs (Kozlov et al., 2005), which are commonly known from neuropeptide precursors. These cleavage signals either separate the N-terminal propeptide from the potential toxin (CHTX-Cc 2, 3, 4, 8) or, in the case of some checacins (checacin 2, 4), result in cleavage after the C-terminal Gly, which provides the amide group for the preceding amino acid. In addition, dibasic Arg-Arg (Checacin 3, 6, 7) or monobasic Arg (checacin 1, 5) also enable effective C-terminal cleavage of the checacins. The remaining venom peptides of *C. cancroides* either consist of the complete precursor (without signal peptide), or the mature peptides are not yet verified by MS analysis. A number of identified precursors for venom peptides show orthology to enzyme precursors. It is likely that at least the predicted *C. cancroides* phospholipases and metalloproteases are actively involved in envenomation. Both phospholipases and metalloproteases have been identified previously in many venoms, (Carmo et al., 2014; Casewell et al., 2013; Ramos and Selistre-de-Araujo, 2006) and are often described as spreading factors that facilitate the dispersion of toxins by destroying either cell membranes or proteins. However, at least for phospholipases, the range of effects seems to be more complex, as phospholipase homologues cause a variety of pharmacological effects in the case of snake venoms, and also act as neurotoxins themselves (Manjunatha Kini, 2003).

A major achievement of our study is the documentation of the first specific effects of pseudoscorpion crude venom on insect and arachnid ion channels. Interestingly, our activity test with crude venom confirmed inhibition of insect voltage-gated potassium channels. This activity might be caused by CHTX-Cc1a and 1b. However, similarities only provide a first indication about the potential activity of a bioactive peptide/protein and are not sufficient to draw valid conclusions (Stevens et al., 2011). In the case of neurotoxins, small sequence differences can alter the target-sensitivity/-binding (e.g. Peigneur et al., 2012). In addition, even though CHTX-Cc1a and 1b exhibit similarity to scorpion alpha-KTX, both lack the described 'scorpion K_V channel toxin signature' which is important but not obligatory for toxin interaction

with K_V channels (Zhu et al., 2014). Consequently, further studies are needed to test the specific activities of CHTX-Cc1a and 1b on molecular and cellular level.

The modulation of the inactivation process of VdNav1 channel from *V. destructor*, discovered in our activity tests, provides evidence for the presence of sodium channel toxins in the venom. Similar activities were previously described for spider, sea anemone and scorpion toxins binding at the neurotoxin binding site 3 of Nav channels (Stevens et al., 2011). It is conspicuous that no compounds with sequence similarity to these were found in the venom of *C. cancroides*. Of the compounds identified, CHTX-Cc8a and 8b can be speculated to cause the observed effect, as both showed similarity to Kappa-theraphotoxins. These usually bind to potassium channels, but most peptides of this family also act on sodium channels by modulating the inactivation in a similar fashion as observed for the crude venom of *C. cancroides* (e.g., Xiao et al., 2004). Otherwise, it might be that the pseudoscorpion toxins causing this effect belong to a novel structural family of Nav/Kv modulators or even act in a different way, e.g., on another/novel binding site. The first evidence for pore-forming cytolytic effects of the venom can be drawn based on the application of higher venom amounts to *Xenopus*-oocytes, resulting in an outwardly rectifying venom-dependent conductance.

The electrophysiological data suggest that *C. cancroides* represents an interesting new source of pesticidal compounds. Especially the modulation of VdNav1 channels from *V. destructor*, an important pest of honeybee hives, may provide new insights on how *C. cancroides* efficiently control *Varroa* mites. *C. cancroides* has previously been considered to protect honey bees from mite infestations (van Toor et al., 2015).

The potency of the venom also becomes evident from predation rates of first-instar *C. cancroides* given a choice of the similar sized Psocids (*Liposcelis entomophila*) and the much larger *Varroa* mites (Fig. 1B) In an arena containing 15 each of healthy Psocids and *Varroa*, 5 first-instar larvae that had been removed from culture and starved for two days killed on average 52% of Psocids and 25% of *Varroa* within 4 hours over 23 repeats (van Toor, unpublished).

A previous study on *S. apimelus* used transcriptome information obtained from an extract of the chelal hand to discuss the hypothetical venom composition of pseudoscorpions (Santibáñez-López et al., 2018). This allows us to make a detailed comparison of the potential venom precursors proposed for *S. apimelus* with the peptides biochemically identified in *C. cancroides*. Notably, few predicted venom precursors of *S. apimelus* match precursors whose products were biochemically confirmed in the venom of *C. cancroides*. Although orthologs of many of these *S. apimelus* genes that are dominated by enzyme-coding genes were also found in the chelal hand transcriptome of *C. cancroides*, we identified only very few mature compounds from these

precursors in the venom of *C. cancroides*. Vice versa, only a few significant BLAST-hits corresponding to precursors of confirmed venom compounds of *C. cancroides* were found in the chelal hand transcriptome of *S. apimelus*. There are several possible explanations for these findings: 1) The venom of these species is indeed highly different; not only because the peptide sequences derived from orthologous genes are quite different, but also because completely different genes are involved in venom production. For example, it has been postulated for centipedes (Jenner et al., 2019) that the venom composition differs significantly between higher-level taxa, such as the different orders of centipedes. *C. cancroides* and *S. apimelus* belong to the same suborder (Iocheirata) within the Pseudoscorpiones but represent different families. The lineages to which they belong have been separated for more than 200 million years (Benavides et al., 2019). 2) The venom compounds identified here for *C. cancroides* are still incomplete and further studies will show better agreement between the proposed venom precursors of *S. apimelus* and those of *C. cancroides*. 3) Information on venom precursors of *S. apimelus*, currently based on a solely transcriptomic approach, is still incomplete. 4) The actual composition of the venom of *S. apimelus* differs to a greater degree from that described in Sharma et al. (2019). It was shown that solely transcriptomic venom profiles can overestimate venom complexity substantially (e. g., Smith and Undheim, 2018). The correct answer which of the explanations fits best is probably somewhere in the middle, but it would certainly be interesting to verify which compounds actually appear in the venom of *S. apimelus* by proteomics analysis of released venom. Only then will it be possible to assess whether the venom composition of these pseudoscorpions has evolved mainly after the separation of their lineages or is more similar than it currently appears. However, what is already clear from our study on *C. cancroides* is the presence of a strikingly large number of novel venom compounds whose specific cellular targets still await functional deorphanization.

5. Acknowledgements

We thank the CECAD Proteomics Facility, University of Cologne and Lapo Ragionieri (Institute for Zoology, Cologne) for excellent support with Quadrupole Orbitrap analyses. J. K. received funding from the Studienstiftung des deutschen Volkes. This study was supported by grants G0E7120N, GOC2319N, and GOA4919N from the F.W.O. Vlaanderen awarded to J.T., and S.P. was supported by KU Leuven funding (PDM/19/164) and F.W.O. Vlaanderen grant 12W7822N.

6. Literature

Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37, 420–423. https://doi.org/10.1038/s41587-019-0036-z

Benavides, L.R., Cosgrove, J.G., Harvey, M.S., Giribet, G., 2019. Phylogenomic interrogation resolves the backbone of the Pseudoscorpiones tree of life. Mol. Phylogenet. Evol. 139, 106509. https://doi.org/10.1016/j.ympev.2019.05.023

Benjannet, S., Rondeau, N., Day, R., Chretien, M., Seidah, N.G., 1991. PC1 and PC2 are proprotein convertases capable of cleaving proopiomelanocortin at distinct pairs of basic residues. Proc. Natl. Acad. Sci. 88, 3564–3568. https://doi.org/10.1073/pnas.88.9.3564

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A., Finn, R.D., 2021. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49, D344–D354. https://doi.org/10.1093/nar/gkaa977

Boldrini-França, J., Cologna, C.T., Pucca, M.B., Bordon, K. de C.F., Amorim, F.G., Anjolette, F.A.P., Cordeiro, F.A., Wiezel, G.A., Cerni, F.A., Pinheiro-Junior, E.L., Shibao, P.Y.T., Ferreira, I.G., de Oliveira, I.S., Cardoso, I.A., Arantes, E.C., 2017. Minor snake venom proteins: Structure, function and potential applications. Biochim. Biophys. Acta BBA - Gen. Subj. 1861, 824–838. https://doi.org/10.1016/j.bbagen.2016.12.022

Bordon, K. de C.F., Cologna, C.T., Fornari-Baldo, E.C., Pinheiro-Júnior, E.L., Cerni, F.A., Amorim, F.G., Anjolette, F.A.P., Cordeiro, F.A., Wiezel, G.A., Cardoso, I.A., Ferreira, I.G., Oliveira, I.S. de, Boldrini-França, J., Pucca, M.B., Baldo, M.A., Arantes, E.C., 2020. From Animal Poisons and Venoms to Medicines: Achievements, Challenges and Perspectives in Drug Discovery. Front. Pharmacol. 11, 1132. https://doi.org/10.3389/fphar.2020.01132

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527. https://doi.org/10.1038/nbt.3519

Camargos, T.S., Restano-Cassulini, R., Possani, L.D., Peigneur, S., Tytgat, J., Schwartz, C.A., Alves, E.M.C., de Freitas, S.M., Schwartz, E.F., 2011. The new kappa-KTx 2.5 from the scorpion *Opisthacanthus cayaporum*. Peptides 32, 1509–1517. https://doi.org/10.1016/j.peptides.2011.05.017

Carmo, A.O., Oliveira-Mendes, B.B.R., Horta, C.C.R., Magalhães, B.F., Dantas, A.E., Chaves, L.M., Chávez-Olórtegui, C., Kalapothakis, E., 2014. Molecular and functional characterization of metalloserrulases, new metalloproteases from the *Tityus serrulatus* venom gland. Toxicon 90, 45–55. https://doi.org/10.1016/j.toxicon.2014.07.014

Casewell, N.R., Wüster, W., Vonk, F.J., Harrison, R.A., Fry, B.G., 2013. Complex cocktails: the evolutionary novelty of venoms. Trends Ecol. Evol. 28, 219–229. https://doi.org/10.1016/j.tree.2012.10.020

Chamberlin, J.C., 1924. Preliminary note upon the pseudoscorpions as a venomous order of the Arachnida. Entomol News 35, 205–209.

Delgado-Prudencio, G., Possani, L.D., Becerril, B., Ortiz, E., 2019. The Dual α-Amidation System in Scorpion Venom Glands. Toxins 11, 425. https://doi.org/10.3390/toxins11070425

Diego-García, E., Caliskan, F., Tytgat, J., 2014. The Mediterranean scorpion *Mesobuthus gibbosus* (Scorpiones, Buthidae): transcriptome analysis and organization of the genome encoding chlorotoxin-like peptides. BMC Genomics 15, 295. https://doi.org/10.1186/1471-2164-15-295

dos Santos, W.F., Coutinho-Netto, J., 2006. Effects of the Paratemnus elongatus pseudoscorpion venom in the uptake and binding of the L-glutamate and GABA from rat cerebral cortex. J. Biochem. Mol. Toxicol. 20, 27–34. https://doi.org/10.1002/jbt.20113

Drukewitz, S., Fuhrmann, N., Undheim, E., Blanke, A., Giribaldi, J., Mary, R., Laconde, G., Dutertre, S., von Reumont, B., 2018. A Dipteran's Novel Sucker Punch: Evolution of Arthropod Atypical Venom with a Neurotoxic Component in Robber Flies (Asilidae, Diptera). Toxins 10, 29. https://doi.org/10.3390/toxins10010029

D'Suze, G., Schwartz, E.F., García-Gómez, B.I., Sevcik, C., Possani, L.D., 2009. Molecular cloning and nucleotide sequence analysis of genes from a cDNA library of the scorpion *Tityus discrepans*. Biochimie 91, 1010–1019. https://doi.org/10.1016/j.biochi.2009.05.005

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconsis, X., Fran, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.

Harms, D., Dunlop, J.A., 2017. The fossil history of pseudoscorpions (Arachnida: Pseudoscorpiones). Foss. Rec. 20, 215–238. https://doi.org/10.5194/fr-20-215-2017

Harvey, A.L., 2014. Toxins and drug discovery. Toxicon 92, 193–200. https://doi.org/10.1016/j.toxicon.2014.10.020

Harvey, M.S., 1992. The phylogeny and classification of the Pseudoscorpionida (Chelicerata: Arachnida). Invert Taxon 6, 1373–1435.

Herzig, V., Cristofori-Armstrong, B., Israel, M.R., Nixon, S.A., Vetter, I., King, G.F., 2020. Animal toxins — Nature's evolutionary-refined toolkit for basic research and drug discovery. Biochem. Pharmacol. 181, 114096. https://doi.org/10.1016/j.bcp.2020.114096

Hughes, C.S., Moggridge, S., Müller, T., Sorensen, P.H., Morin, G.B., Krijgsveld, J., 2019. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat. Protoc. 14, 68–85. https://doi.org/10.1038/s41596-018-0082-x

Jenner, R.A., von Reumont, B.M., Campbell, L.I., Undheim, E.A.B., 2019. Parallel Evolution of Complex Centipede Venoms Revealed by Comparative Proteotranscriptomic Analyses. Mol. Biol. Evol. 36, 2748–2763. https://doi.org/10.1093/molbev/msz181

King, G., 2013. Venoms to Drugs: Translating Venom Peptides into Therapeutics 44, 4.

King, G.F., Gentz, M.C., Escoubas, P., Nicholson, G.M., 2008. A rational nomenclature for naming peptide toxins from spiders and other venomous animals. Toxicon 52, 264–276. https://doi.org/10.1016/j.toxicon.2008.05.020

King, G.F., Hardy, M.C., 2013. Spider-Venom Peptides: Structure, Pharmacology, and Potential for Control of Insect Pests. Annu. Rev. Entomol. 58, 475–496. https://doi.org/10.1146/annurev-ento-120811-153650

Kozlov, S., Malyavka, A., McCutchen, B., Lu, A., Schepers, E., Herrmann, R., Grishin, E., 2005. A novel strategy for the identification of toxinlike structures in spider venom. Proteins Struct. Funct. Bioinforma. 59, 131–140. https://doi.org/10.1002/prot.20390

Krämer, J., Pohl, H., Predel, R., 2019. Venom collection and analysis in the pseudoscorpion *Chelifer cancroides* (Pseudoscorpiones: Cheliferidae). Toxicon 162, 15–23. https://doi.org/10.1016/j.toxicon.2019.02.009

Kuzmenkov, A.I., Nekrasova, O.V., Peigneur, S., Tabakmakher, V.M., Gigolaev, A.M., Fradkov, A.F., Kudryashova, K.S., Chugunov, A.O., Efremov, R.G., Tytgat, J., Feofanov, A.V., Vassilevski, A.A., 2018. KV1.2 channel-specific blocker from *Mesobuthus eupeus* scorpion venom: Structural basis of selectivity. Neuropharmacology 143, 228–238. https://doi.org/10.1016/j.neuropharm.2018.09.030

Langenegger, N., Nentwig, W., Kuhn-Nentwig, L., 2019. Spider Venom: Components, Modes of Action, and Novel Strategies in Transcriptomic and Proteomic Analyses. Toxins 11, 611. https://doi.org/10.3390/toxins11100611

Lebenzon, J.E., Toxopeus, J., Anthony, S.E., Sinclair, B.J., 2021. De novo assembly and characterisation of the transcriptome of the Beringian pseudoscorpion. Can. Entomol. 153, 301–313. https://doi.org/10.4039/tce.2021.2

Lüddecke, T., Vilcinskas, A., Lemke, S., 2019. Phylogeny-Guided Selection of Priority Groups for Venom Bioprospecting: Harvesting Toxin Sequences in Tarantulas as a Case Study 10.

Lüddecke, T., von Reumont, B.M., Förster, F., Billion, A., Timm, T., Lochnit, G., Vilcinskas, A., Lemke, S., 2020. An Economic Dilemma between Molecular Weapon Systems May Explain an Arachno-Atypical Venom in Wasp Spiders (*Argiope bruennichi*). Biomolecules 10, 978. https://doi.org/10.3390/biom10070978

Manjunatha Kini, R., 2003. Excitement ahead: structure, function and mechanism of snake venom phospholipase A2 enzymes. Toxicon 42, 827–840. https://doi.org/10.1016/j.toxicon.2003.11.002

753 Ontano, A.Z., Gainett, G., Aharon, S., Ballesteros, J.A., Benavides, L.R., Corbett, K.F., Gavish-
754 Regev, E., Harvey, M.S., Monsma, S., Santibáñez-López, C.E., Setton, E.V.W., Zehms, J.T.,
755 Zeh, J.A., Zeh, D.W., Sharma, P.P., 2021. Taxonomic Sampling and Rare Genomic Changes
756 Overcome Long-Branch Attraction in the Phylogenetic Placement of Pseudoscorpions. Mol.
757 Biol. Evol. https://doi.org/10.1093/molbev/msab038
758 Pearson, W.R., 2013. An Introduction to Sequence Similarity ("Homology") Searching. Curr. Protoc.
759 Bioinforma. 42. https://doi.org/10.1002/0471250953.bi0301s42
760 Peigneur, S., Béress, L., Möller, C., Marí, F., Forssmann, W., Tytgat, J., 2012. A natural point
761 mutation changes both target selectivity and mechanism of action of sea anemone toxins.
762 FASEB J. 26, 5141–5151. https://doi.org/10.1096/fj.12-218479
763 Peigneur, S., da Costa Oliveira, C., de Sousa Fonseca, F.C., McMahon, K.L., Mueller, A., Cheneval,
764 O., Cristina Nogueira Freitas, A., Starobova, H., Dimitri Gama Duarte, I., Craik, D.J., Vetter,
765 I., de Lima, M.E., Schroeder, C.I., Tytgat, J., 2021. Small cyclic sodium channel inhibitors.
766 Biochem. Pharmacol. 183, 114291. https://doi.org/10.1016/j.bcp.2020.114291
767 Ramos, O.H.P., Selistre-de-Araujo, H.S., 2006. Snake venom metalloproteases — structure and
768 function of catalytic and disintegrin domains. Comp. Biochem. Physiol. Part C Toxicol.
769 Pharmacol. 142, 328–346. https://doi.org/10.1016/j.cbpc.2005.11.005
770 Santibáñez-López, C., Ontano, A., Harvey, M., Sharma, P., 2018. Transcriptomic Analysis of
771 Pseudoscorpion Venom Reveals a Unique Cocktail Dominated by Enzymes and Protease
772 Inhibitors. Toxins 10, 207. https://doi.org/10.3390/toxins10050207
773 Schendel, Rash, Jenner, Undheim, 2019. The Diversity of Venom: The Importance of Behavior and
774 Venom System Morphology in Understanding Its Ecology and Evolution. Toxins 11, 666.
775 https://doi.org/10.3390/toxins11110666
776 Shafee, T.M.A., Lay, F.T., Phan, T.K., Anderson, M.A., Hulett, M.D., 2017. Convergent evolution of
777 defensin sequence, structure and function. Cell. Mol. Life Sci. 74, 663–682.
778 https://doi.org/10.1007/s00018-016-2344-5
779 Smith, J., Undheim, E., 2018. True Lies: Using Proteomics to Assess the Accuracy of Transcriptome-
780 Based Venomics in Centipedes Uncovers False Positives and Reveals Startling Intraspecific
781 Variation in Scolopendra subspinipes. Toxins 10, 96. https://doi.org/10.3390/toxins10030096
782 Stevens, M., Peigneur, S., Tytgat, J., 2011. Neurotoxins and Their Binding Areas on Voltage-Gated
783 Sodium Channels. Front. Pharmacol. 2. https://doi.org/10.3389/fphar.2011.00071
784 The UniProt Consortium, 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids
785 Res. 49, D480–D489. https://doi.org/10.1093/nar/gkaa1100
786 van Toor, R.F., Thompson, S.E., Gibson, D.M., Smith, G.R., 2015. Ingestion of *Varroa destructor* by
787 pseudoscorpions in honey bee hives confirmed by PCR analysis. J. Apic. Res. 54, 555–562.
788 https://doi.org/10.1080/00218839.2016.1184845

von Reumont, B.M., Blanke, A., Richter, S., Alvarez, F., Bleidorn, C., Jenner, R.A., 2014a. The First Venomous Crustacean Revealed by Transcriptomics and Functional Morphology: Remipede Venom Glands Express a Unique Toxin Cocktail Dominated by Enzymes and a Neurotoxin. Mol. Biol. Evol. 31, 48–58. https://doi.org/10.1093/molbev/mst199

von Reumont, B.M., Campell, L.I., Hering, L., Sykes, D., Hetmank, J., Jenner, R.A., Bleidorn, C., 2014b. A Polychaete's Powerful Punch: Venom Gland Transcriptomics of Glycera Reveals a Complex Cocktail of Toxin Homologs. Genome Biol. Evol. 6, 2406–2423. https://doi.org/10.1093/gbe/evu190

von Reumont, B.M., Lüddecke, T., Timm, T., Lochnit, G., Vilcinskas, A., von Döhren, J., Nilsson, M.A., 2020. Proteo-Transcriptomic Analysis Identifies Potential Novel Toxins Secreted by the Predatory, Prey-Piercing Ribbon Worm Amphiporus lactifloreus. Mar. Drugs 18, 407. https://doi.org/10.3390/md18080407

Walker, A.A., Dobson, J., Jin, J., Robinson, S.D., Herzig, V., Vetter, I., King, G.F., Fry, B.G., 2018. Buzz Kill: Function and Proteomic Composition of Venom from the Giant Assassin Fly *Dolopus genitalis* (Diptera: Asilidae) 17.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., Zdobnov, E.M., 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol. Biol. Evol. 35, 543–548. https://doi.org/10.1093/molbev/msx319

Xiao, Y., Tang, J., Yang, Y., Wang, M., Hu, W., Xie, J., Zeng, X., Liang, S., 2004. Jingzhaotoxin-III, a Novel Spider Toxin Inhibiting Activation of Voltage-gated Sodium Channel in Rat Cardiac Myocytes. J. Biol. Chem. 279, 26220–26226. https://doi.org/10.1074/jbc.M401387200

Zhu, S., Peigneur, S., Gao, B., Umetsu, Y., Ohki, S., Tytgat, J., 2014. Experimental Conversion of a Defensin into a Neurotoxin: Implications for Origin of Toxic Function. Mol. Biol. Evol. 31, 546–559. https://doi.org/10.1093/molbev/msu038

Zhu, S., Tytgat, J., 2004. The scorpine family of defensins: gene structure, alternative polyadenylation and fold recognition. Cell. Mol. Life Sci. 61. https://doi.org/10.1007/s00018-004-4149-1