Venomics survey of six myrmicine ants provides insights into the molecular and structural diversity of their peptide toxins

Valentine Barassé, Nathan Téné, Christophe Klopp, Françoise Paquet, Niklas Tysklind, Valérie Troispoux, Hadrien Lalägue, Jérôme Orivel, Benjamin Lefranc, Jérôme Leprince, et al.

In preparation for *Insect Biochemistry and Molecular Biology*

Venomics survey of six myrmicine ants provides insights into the molecular and structural diversity of their peptide toxins

Valentine Barassé[1], Nathan Téné[1], Christophe Klopp[2], Françoise Paquet[3], Niklas Tysklind[4], Valérie Troispoux[4], Hadrien Lalägue[5], Jérôme Orivel[5], Benjamin Lefranc[6], Jérôme Leprince[6], Martin Kenne[7], Maurice Tindo[7], Michel Treilhou[1], Axel Touchard[1,5#*], Elsa Bonnafé[1#]*

[1] EA-7417, Institut National Universitaire Champollion, Place de Verdun, 81012 Albi, France ; nathan.tene@univ-jfc.fr (N.T.) ; elsa.bonnafe@univ-jfc.fr (E.B.) ; michel.treilhou@univ-jfc.fr (M.T.)

[2] Unité de Mathématique et Informatique Appliquées de Toulouse, UR0875, Genotoul Bioinfo, INRAE Toulouse, 31326 Castanet-Tolosan, France ; christophe.klopp@inrae.fr (C.K.)

[3] Centre de Biophysique Moléculaire – CNRS – UPR 4301, 45071 Orléans, France ; francoise.paquet@cnrs-orleans.fr (F.P.)

[4] INRAE, UMR EcoFoG (Agroparistech, CNRS, Cirad, Université des Antilles, Université de la Guyane), Campus Agronomique, 97310 Kourou, French Guiana ; niklas.tysklind@ecofog.gf (N.T.) ; valerie.troispoux@ecofog.gf (V.T.)

[5] CNRS, UMR EcoFoG (AgroParisTech, CNRS, CIRAD, INRAE, Université des Antilles, Université de Guyane), 97310 Kourou, France; axel.touchard2@gmail.com (A.T) ; hadrien.lalague@ecofog.gf (H.L.) ; jerome.orivel@ecofog.gf (J.O.) ;

[6] Inserm U 1239, Normandie Univ, UNIROUEN, Plate-forme de Recherche en Imagerie Cellulaire Normandie (PRIMACEN), 76000 Rouen, France ; benjamin.lefranc@univ-rouen.fr (B.L.); jerome.leprince@univ-rouen.fr (J.R)

[7] Laboratory of Animal Biology and Physiology, Faculty of Science, University of Douala, Cameroon, P.O.Box. 24157 Douala, Cameroon; tindodouala@yahoo.com (M.T.); medoum68@yahoo.fr (M.K.)

[#] These authors contributed equally to this study.

Corresponding authors:

***** EA-7417, Institut National Universitaire Champollion, Place de Verdun, 81012 Albi, France. Email: valentine.barasse@gmail.com

* CNRS, UMR EcoFoG (AgroParisTech, CNRS, CIRAD, INRAE, Université des Antilles, Université de Guyane), 97310 Kourou, France. Email: axel.touchard2@gmail.com

Declarations of interest: none

Abstract:

Among ants, Myrmicinae represents the most speciose subfamily. The venom composition previously described for these social insects is extremely variable, with alkaloids predominant in some genera while, conversely, proteomics studies have revealed that some myrmicine ant venoms are peptide-rich. Using integrated transcriptomic and proteomic approaches, we characterized the venom peptidomes of six ants belonging to the different tribes of Myrmicinae. We identified a total of 79 myrmicitoxins precursors which can be classified into 38 peptide families according to their mature sequences. Myrmicine ant venom peptidomes showed heterogeneous compositions, with linear and disulfide-bonded monomers as well as dimeric toxins. Several peptide families were exclusive to a single venom whereas some were retrieved in multiple species. A hierarchical clustering analysis of precursor signal sequences led us to divide the myrmicitoxins precursors into eight families, including some that have already been described in other aculeate hymenoptera such as secapin-like peptides and

voltage-gated sodium channel (Na_V) toxins. Evolutionary and structural analyses of two representatives of these families highlighted variation and conserved patterns that might be crucial to explain myrmicine venom peptide functional adaptations to biological targets.

Keywords: toxin precursor; secapins; voltage-gated sodium channel (Na_V) peptide; NMR structure; dimeric peptide

Key Contribution: This study provides novel insights into both peptide diversity and diversification of myrmicine ant venoms.

1. Introduction

Venoms are fine-tuned biochemical arsenals mainly used by animals to defend themselves and/or capture prey (Casewell et al., 2013). Investigations conducted on arthropod venoms (*e.g.*, scorpions, spiders, centipedes, and insects) revealed complex mixtures of toxins with peptides as the dominant reported components. These venoms represent an immense source of structurally diverse toxins with variable amino acid sequences and tridimensional structures (Walker et al., 2018; Wilson and Daly, 2018). However, the peptide toxins of most insect venoms such as those from ants remain largely uncharacterized, despite being among the most abundant and ecologically dominant terrestrial venomous animals. The small size of these insects and consequently the difficulty in recovering large amounts of venom partly explains why ant venom peptides have been so far overlooked (Touchard et al., 2016a). With over 14,000 extant species, ants form a diverse group of social venomous hymenopterans capable to inject or spray secretions from a venom reservoir mainly for defense or prey capture. Ant venoms harbor a clear potential for the discovery of novel peptides with original scaffolds and pharmacology. Recent proteo-transcriptomic investigations provided comprehensive venom peptidomes from several subfamilies including Pseudomyrmecinae, Paraponerinae, Myrmicinae, Ponerinae and Myrmeciinae (Aili et al., 2020; Barassé et al., 2019; Kazuma et al., 2017; Robinson et al., 2018; Touchard et al., 2018, 2020a) and started to reveal the molecular diversity of these toxins. Although these venoms appeared to be less complex in peptide numbers and diversity than those from cone snails or spiders, several interesting peptide structures have been indeed characterized, such as linear helicoidal, dimeric, β-hairpin, EGF-like, ICK-like or Kunitz-like peptides (Barassé et al., 2019; Radis-Baptista et al., 2020;

Touchard et al., 2020b, 2016b) (Barassé et al., 2019; Radis-Baptista et al., 2020; Touchard et al., 2020, 2016b).

An extensive inclusion of ant species from unexplored clades and with broader ecological diversity would thus enhance our ant venom molecular diversification understanding and lead to further discoveries. Among ants, Myrmicinae is the most speciose ant subfamily with 147 valid extant genera (Antweb, consulted on 22/04/22 and AntCat, consulted on 22/04/22), among which 111 are stinging species capable of injecting their venom (Blanchard and Moreau, 2017; and personal observations A.T.). Myrmicines occur in most terrestrial habitats and exhibit wide variation in ecology, behavior, and diet (Blaimer et al., 2018; Ward et al., 2015). Reported observations demonstrated a predominance of alkaloids in some genera such as *Solenopsis* and *Monomorium* (Jones et al., 2003; Morgan, 2008; Touchard et al., 2016a). Also, proteomic analyses have revealed that the venoms of four genera (*i.e.*, *Tetramorium*, *Pogonomyrmex*, *Myrmica*, and *Manica*) are peptide-rich (Heep et al., 2019a, 2019b; Hurka et al., 2022; Schmidt and Blum, 1978; Touchard et al., 2020a, 2018; von Sicard et al., 1989). Further investigations into the venom composition of other myrmicine species is therefore necessary to report the molecular and structural peptide diversity, to understand the heterogeneity and the intrinsic chemical properties of these venoms as well as to gather information on relationships between ant phylogeny and venom composition.

Here, we combined venom gland transcriptomics and venom peptidomics data to decipher the venom composition of six ant species representative of the six Myrmicinae tribes (Ward et al., 2015): *Myrmica ruginodis* (Myrmicini), *Pogonomyrmex californicus* (Pogonomyrmecini), *Stenamma debile* (Stenammini), *Solenopsis saevissima* (Solenopsidini), *Daceton armigerum* (Attini) and *Tetramorium africanum* (Crematogastrini). In light of the peptide composition of myrmicine ant venoms, the tridimensional structure of the first representative member of the most abundant peptide family was determined by NMR spectroscopy. In addition, we further conducted a preliminary evolutionary analysis on the two most ubiquitous peptide families in order to test and determine the site of action of the driving selection forces.

2. Materials and Methods

2.1. Collection and preparation of venom samples

Ant workers were collected from different locations (*i.e.*, France, French Guiana, Cameroon, and USA) and kept alive in the laboratory between 1 and 3 days before dissection.

Between 10 and 52 ant venom reservoirs per species were dissected and pooled in a solution containing 10% acetonitrile (ACN)/ ultrapure water (v/v) (Supplementary Table I). The membranes were then disrupted using ultrasonic waves for 2 min. Finally, the samples were centrifuged for 5 min at 14,400 rpm, and the supernatant was collected and dried using a speed vacuum prior to storage at −20 °C until proteomic analysis.

We tried to follow the practical guide proposed by Walker *et al.* (2020) to study venoms by proteo-transcriptomics (Walker et al., 2020). However, we chose to extract the venoms by dissection which may induce minor contaminations by housekeeping proteins since for some ant species the venom is difficult to obtain by other extraction methods. This approach yielded peptide toxins having a typical pattern for proteins secreted in ant venoms as toxins.

2.2. Mass spectrometry analysis

A preliminary LC–MS analysis of crude venoms was conducted on the LCQ-Ion trap Advantage equipped with an ESI-LC system Accela (ThermoFisher Scientific, Courtabœuf, France) (Supplementary Table I). An Acclaim RSLC C_{18} column (2.2 µm; 2.1 × 150 mm; ThermoFisher Scientific) was used to separate the peptides. A gradient prepared from 0.1% formic acid (FA)/water (v/v) (solvent A) and 0.1% FA/ACN (v/v) (solvent B) constituted the mobile phase. The peptides were eluted using a linear gradient from 0 to 50% of solvent B over 45 min, then from 50 to 100% over 10 min, and finally held for 5 min at a 250 $\mu L.min^{-1}$ flow rate. The electrospray ionization mass spectrometry detection was done in positive mode with the following optimized parameters: the capillary temperature was set at 300 °C, the spray voltage was 4.5 kV, and the sheath gas and auxiliary gas were set at 50 and 10 psi, respectively. The acquisition range was from 100 to 2000 *m/z*. The peak ion extraction function in Xcalibur software (version 4.0, ThermoFisher Scientific) was used to manually integrate the area value of each peak corresponding to a peptide. This relative area provides a measure of relative abundance, as it indicates the contribution of each peptide to all the peptides identified in the venom.

Although peptides structured by disulfide bonds are less abundant within ant venoms than in marine snail, spider, or scorpion venoms, some studies revealed the presence of venom peptides with one, two, and three disulfide bonds (Aili et al., 2014; Pan and Hink, 2000; Touchard et al., 2020a, 2018, 2015). To detect them, a reduction/alkylation was done on several venoms depending on the quantity of crude venom harvested (*i.e.*, *Daceton armigerum*, *Pogonomyrmex californicus* and *Myrmica ruginodis*). Each lyophilized sample of crude venom

was solubilized with 30 µL of ultrapure water. Then, reduction of disulfide bonds was achieved by mixing these aqueous solutions of crude venoms (*i.e.*, 24 venom reservoirs of *D. armigerum*, 25 of *My. ruginodis*, and 10 of *P. californicus*) with 30 µL of 100 mM ammonium bicarbonate buffer (pH 8) containing 10 mM dithiothreitol (DTT) followed by an incubation for 30 min at 56°C. Then, the reduced venoms were alkylated by adding 10 µL of 50 mM iodoacetamide (IA) for 15 min at room temperature in the dark. These samples were then analyzed through LC-MS according to the protocol described above. Since chemical reduction/alkylation results in a mass increase of 57 Da for each cysteine, the examination of mass shifts in the mass spectra of reduced/alkylated samples allowed us to determine the presence and the number of disulfide bonds in the corresponding peptides.

Venoms from *Solenopsis* species are known to be rich in alkaloids. However, the presence of peptides has already been suggested as possible venom components (Dos Santos Pinto et al., 2012; Fox et al., 2013). We thus did an hexane extraction on *So. saevissima* crude venom according to the protocol described by Fox *et al.* (2013) to remove most of the alkaloids. Briefly, the hexane extraction was conducted by mixing an aqueous solution of crude venom (*i.e.*, 32 lyophilized venom reservoirs solubilized with 500 µL of ultrapure water) with 500 µL of hexane. This operation was repeated twice. The combined extracts were evaporated under a stream of nitrogen. The residue was redissolved in 50 µL of hexane and 1 µL was injected in splitless mode. GC–MS analysis were carried out on a Trace GC 1300 coupled with a Triple Quadrupole mass spectrometer (TSQ Duo, ThermoScientific) controlled by *Xcalibur* software. The gas chromatograph was equipped with an TG-5SilMS capillary column (30 m×0.25 mm i.d.; 0.25 µm film thickness). The injector temperature was 250 °C. The oven temperature started from 120 °C during 1 min, increased at 15 °C.min^{-1} to 280 °C and was maintained for 15 min. The mass spectrometric detector was operated in electron impact ionization mode with an ionizing energy of 70 eV, scanning from m/z = 50 to 600. The ion source and transfer line temperature were 300 and 280 °C, respectively. The carrier gas was helium (alphagaz 2; Air Liquide, France) at a flow rate of 1.1 mL·min^{-1}. The hexane fraction was then analyzed through GC-MS, whereas the aqueous phase was concentrated prior to LC-MS according to the protocol described above.

2.3. De novo orbitrap mass spectrometry-based sequencing

Each crude venom was re-suspended in water and then desalted using ZipTip® C_{18} (Merck Millipore, Burlington, VT, USA) after adding trifluoroacetic acid (TFA) at a final

concentration of 0.5% (Supplementary Table I). Then, the venom sample was subjected to *de novo* sequencing using a Q-Exactive Plus mass spectrometer coupled to a Nano-LC Proxeon 1000 (ThermoFisher Scientific, Waltham, MA, USA). Peptides were separated through chromatography with the following parameters: Acclaim PepMap100 C_{18} pre-column (2 cm, 75 µm i.d., 3 µm, 100 Å), Pepmap-RSLC Proxeon C_{18} column (50 cm, 75 µm i.d., 2 µm, 100 Å), 300 nL min−1 flow rate, a 98 min gradient from 95% solvent A (water, 0.1% FA) to 35% solvent B (99.9% ACN, 0.1% FA) for a total time of 2 h. Peptides were analyzed in the Orbitrap cell, at a resolution of 120,000, with a mass range of *m/z* 350–1550. Fragments were obtained through high collision-induced dissociation (HCD) activation with a collisional energy of 27%. Data were acquired in the Orbitrap cell in a Top20 mode, at a resolution of 17,500. For the identification step, all MS and MS/MS data were processed with an in-house Peaks software (BSI, version 6.0) to perform *de novo* sequencing. The mass tolerance was set to 10 ppm for precursor ions and 0.02 Da for fragments. The following modifications were allowed: oxidation (Met), C-terminal amidation and pyroglutamic acid. *De novo* peptide sequences with Average Local Confidence (ALC) higher than 60% were used for the peptide identifications.

2.4. Venom gland transcriptomics

2.4.1. RNA extraction and sequencing

The RNA extraction of the Neotropical ant species *Daceton armigerum* venom glands were prepared by dissecting ant venom apparatus in ultrapure water. Venom reservoirs plus venom glands were immediately placed into 1 mL of RNAlater. The samples were stored at −80 °C prior to RNA extraction. The RNAlater was removed with a glass Pasteur pipette and the venom gland tissues were disrupted with a TissueLyser II (Qiagen, Germantown, MD, USA) in RLT buffer containing 10% (v/v) of 2-mercaptoethanol (RNeasy Mini Kit, Qiagen). RNA was first isolated with a phenol-chloroform (5:1) solution followed by washing with a solution of chloroform-isoamyl alcohol (25:1) to remove the phenol. The RNA was then bound to a Qiagen column and washed as per the manufacturer's instructions. DNAse I (Roche Diagnostics GmbH, Mannheim, Germany) was added to remove DNA fragments. The RNA was eluted in 50 µL of RNase free water and a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific) was used to determine 260/280 and 260/230 nm ratios. Finally, RNAstable® LD (Biomatrica, San Diego, CA, USA) was added to the purified RNA and the sample was dried using a Speed Vac (RC1010, Jouan, Saint Herblain, France). *D. armigerum* RNA was dried together with RNAstable® (Biomatrica, SigmaAldrich, Burlington, MA, USA) before

shipment for transcriptomic analysis. The RNA quantification, the mRNA-seq stranded libraries and sequencing on HiSEQ RAPID pair-end 250pb was carried out by IGA Technology Services (Udine, Italy).

Venom apparatus (*i.e.*, venom glands and reservoir) of *Tetramorium africanum*, *Solenopsis saevissima*, *Myrmica ruginodis*, and *Pogonomyrmex californicus* workers were dissected in a Phosphate Buffered Saline (PBS) solution and immediately placed in 500 µL of TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and store at -80°C prior to RNA extraction. Given the small size of *Stenamma debile* workers combined with the difficulty to collect workers in large numbers, the transcriptome database for this species was done using the whole ant tissues (Supplementary Table II). Total RNAs were extracted with the RNeasy Micro Kit (Qiagen) following the manufacturer's instructions. Contaminating genomic DNA was removed using a DNA-free kit (Applied Biosystem) according to the manufacturer's instructions.

RNA quantity and quality were assessed using a Nanodrop and a bioanalyzer (Nanodrop 2000, ThermoFisher Scientific; Agilent 2100 Bioanalyzer System). The extracted RNA was sequenced through RNAseq at the GeT-PlaGe core facility (INRAE Toulouse, France). Briefly, RNA-seq libraries were prepared according to Illumina's protocols using the Illumina TruSeq Stranded mRNA sample prep kit to analyze mRNA. mRNA was selected using poly-T beads. Then, the RNA was fragmented to generate double stranded cDNA and adaptors were ligated to be sequenced. Eleven cycles of PCR were applied to amplify the libraries. Library quality was assessed using a Fragment Analyser, and the libraries were quantified through qPCR using the Kapa Library Quantification Kit. RNA-seq libraries were sequenced on an Illumina HiSeq3000 using a paired-end read length of 2 × 150 pb with the Illumina HiSeq3000 sequencing kits.

2.4.2. Contig quantification

The read pairs were assembled twice with *drap* (version 1.9.1) (Cabau et al., 2017) using the *de Bruijn* graph assemblers called Oases and Trinity (parameters: –dbg oases/trinity). The assembly metrics were produced with the assemblathon_stats.pl scripts. Raw reads were aligned on the contigs with *bwa mem* (version 0.7.12-r1039) (Li and Durbin, 2010) using the default parameters and the alignment files were sorted, compressed, and indexed with *samtools view*, *sort*, and *index* (version: 1.3.1) using the default parameters (Li et al., 2009). The quantification files were generated with *samtools idxstats* (version: 1.3.1), giving us the length of each contig in base pairs along with the number of hits, corresponding to the number of

sequences from RNAseq reads which aligned on a given contig. To calculate the expression rate of each contig, we calculated the transcripts per million values (TPM) by dividing the number of aligned reads for each contig by the contig length, then dividing this value by the ratio of counts to contig length for all contig. This value was then multiplied by 1 million to generate TPM. Only contigs exhibiting more than 300 TPM were considered for annotation.

2.4.3. Precursor identifications and mature sequences

Contigs were translated using a translate program command lines to obtain the potential Open Reading Frames (ORFs) (*emboss* package, command line: transeq). Then, the fragments of sequences obtained during the *de novo* Orbitrap mass spectrometry-based sequencing were aligned against these data by using the command-line NCBI *BLAST* program with adapted parameters for short sequences (*ncbi-blast-2.6.0+* package, command line: blastp, parameter: -matrix PAM30), allowing us to find the complete peptide sequences and the name of the contigs on which they aligned. The masses of mature peptide sequences, obtained from these different approaches, were systematically verified using the peptide mass program from *ExPASy portal* (https://expasy.org) and compared to those obtained through mass spectrometry. Signal sequences and transmembrane domains were predicted with the *Phobius* program available at http://phobius.sbc.su.se/. Sequence similarities were searched for using the *NCBI BLAST* program presented in the Uniprot server with the default parameters. Alignments were achieved using the *Muscle* program in *Seaview* version 4.6.1 (Gouy et al., 2010), and edited using *BOXSHADE* version 3.2 (https://embnet.vital-it.ch/software/BOX_form.html). Sequence identity and similarity percentages were calculated with the software infoalign from the *EMBOSS* suite of bioinformatic tools (Rice et al., 2000). We also performed a Hierarchical Cluster Analysis (HCA) on signal sequences of myrmicitoxins identified in this study and those previously defined in myrmicine ant venom peptidomes (Touchard et al., 2020a, 2018). Briefly, signal sequences were predicted with *SignalP 5.0* (Almagro Armenteros et al., 2019). Multiple alignments were then achieved with the *ClustalW* program and pairwise distances were computed using *MEGAX* version 10.1.7 with default parameters (Kumar et al., 2018). HCA was then performed using Ward's method with the *R* software (R Core Team, 2017).

2.4.4. Annotation of most expressed contigs

Open reading frames (≥100 amino-acids length), found by translating RNAseq data, were extracted from the most abundant contigs (*i.e.*, over 300 TPMs) and then submitted to the

NCBI *BLAST* program against the Uniprot refseq protein database on the computational cluster of the Genotoul bioinformatic facility (INRA Toulouse, France) (*ncbi-blast-2.6.0* + package, command line: blastp, parameter: -matrix BLOSUM62).

2.4.5. Evolutionary analysis of U_{17} and U_3/U_{33} encoding sequences

U_{17} and U_3/U_{33} nucleotide sequences were obtained from our transcriptome assemblies of *T. africanum*, *My. ruginodis*, *P. californicus*, and *So. saevissima* venom glands as well as previously published *T. bicarinatum* and *Ma. rubida* transcriptomes (PRJNA234295 and PRJEB34828, respectively) and blast searches on transcriptomes available for *Solenopsis invicta* and *Myrmica sulcinodis* on NCBI (GFUX00000000.1 and PRJDB4088, respectively). Regarding their similar biochemical features, U_{33} nucleotide sequences were added to the U_3 nucleotide sequences group for the analysis. Alignments of nucleotide sequences encoding mature (*i.e.*, including the C-terminal amidation signal) or prepro-regions were performed on *MEGAX* with *CLUSTALW* (Codons). To estimate the rate of substitution per synonymous sites (dS) or per non-synonymous sites (dN), we computed a pairwise distance using the modified Nei-Gojobori model on *MEGAX* within each group or between all group pairs. We performed a codon Z-test of purifying or positive selection of overall sequence pairs, or over sequence pairs of each group using the Nei-Gojobori model with the pairwise deletion option on *MEGAX*. We tested sequences encoding only mature regions (*i.e.*, with amidation signal) or sequences encoding only prepro-regions. We then conducted a site-by-site analysis to detect pervasive (*i.e.*, continuous changes) or episodic (*i.e.*, single change) positive or purifying selection on each coding site. For this, we used two methods available on datamonkey web site: Mixed Effects Model of Evolution (MEME, pervasive/episodic) and Fixed Effect Likelihood model (FEL, pervasive).

2.5. Fmoc solid phase synthesis of U_{17}-MYRTX-Tb1a and U_3-MYRTX-Tb1a

All Fmoc amino acid residues, *O*-benzotriazol-1-yl-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HBTU) and Rink amide 4-methylbenzhydrylamine (MBHA) resin were purchased from Christof Senn Laboratories (Dielsdorf, Switzerland) or IRIS Biotech (Marktredwitz, Germany). *N,N*-Diisopropylethylamine (DIEA), piperidine, trifluoroacetic acid (TFA), triisopropylsilane (TIS), *tert*-butylmethylether (TBME), MeOH, NaOMe, DMSO and Fmoc-Thr[GalNAc(Ac)$_3$-α-*D*]-OH were supplied from Sigma-Aldrich (Saint-Quentin-Fallavier, France). *N*-methylpyrrolidone (NMP), dimethylformamide (DMF), dichloromethane

(DCM) and acetonitrile were from Fisher Scientific (Illkirch, France). Peptides U_3-MYRTX-Tb1a (VLPALPLLAGLMSLPFLQHKLTN-NH_2) and [Thr(GalNAc-α-*D*)^1]U_{17}-MYRTX-Tb1a (XIINAPNRCPPGHVVVKGRCRIA-NH_2, X= Thr(GalNAc-α-*D*)) were synthesized by Fmoc solid phase methodology on a Liberty microwave assisted automated peptide synthesizer (CEM, Saclay, France) using the standard manufacturer's procedures at 0.1 mmol scale. All Fmoc-amino acids (0.5 mmol, 5 eq.) were coupled (25 W, 75°C, 300 sec) except for histidine and cysteine (0 W, 50°C, 120 sec and 25 W, 50°C, 240 sec) on Rink amide resin, by *in situ* activation with HBTU (0.5 mmol, 5 eq.) and DIEA (1 mmol, 10 eq.) before Fmoc removal with a 20% piperidine in DMF (35 W, 75°C, 30 sec and 35 W, 75°C, 180 sec). Reactive side chains were protected as follow: Asn, Gln, trityl (Trt) amide; His, trityl (Trt) amine; Ser, Thr, *tert*-butyl (*t*Bu) ether; Arg, pentamethyldihydrobenzofuran (Pbf) sulfonylamide; Lys, *tert*-butyloxycarbonyl (Boc) carbamate; Cys, acetamidomethyl (Acm) thioether. After completion of the chain assembly, peptides were deprotected and cleaved from the resin by adding 10 mL of an ice-cold mixture of TFA/TIS/H_2O (9.5:0.25:0.25, v/v/v) and agitating 3 hours at room temperature. Peptides were then obtained by precipitation in TBME followed by centrifugation (4500 rpm, 15 min). The synthetic peptides U_3-MYRTX-Tb1a and [Thr(GalNAc$(Ac)_3$-α-*D*)^1]U_{17}-MYRTX-Tb1a were purified by reversed-phase HPLC on a 21.2 x 250 mm Jupiter C_{18} (5 µm, 300 Å) column (Phenomenex, Le Pecq, France) using a linear gradient (20-70% and 10-60% over 45 min, respectively) of acetonitrile/TFA (99.9:0.1) at a flow rate of 10 mL/min. Deacetylation of the Thr(GalNAc$(Ac)_3$-α-*D*) residue was performed with sodium methoxide. The lyophilized glycopeptide was dissolved in methanol (5 mg/mL) and a solution of sodium methoxide (0.3 M) was added dropwise to a final concentration of 30 mM. After complete removal of O-acetyl protection groups in about 6 h, the reaction mixture was frozen after adding water and freeze-dried. The formation of the disulfide bond was then carried out by oxidation of the deprotected peptide in a mixture of H_2O/CH_3COOH/DMSO (75:5:20) (0.5 mg/mL) at pH 5.2. The reaction was stopped after 19 h, freeze-dried and the resulting cyclic glycopeptide was purified in the same conditions as described above. The purified peptides, U_3-MYRTX-Tb1a and [Thr(GalNAc-α-*D*)^1]U_{17}-MYRTX-Tb1a, were then characterized by MALDI-TOF mass spectrometry on a ultrafleXtreme (Bruker, Strasbourg, France) in the reflector mode using α-cyano-4-hydroxycinnamic acid as a matrix. Analytical RP-HPLC, implemented on a 4.6 x 250 mm Jupiter C_{18} (5 µm, 300 Å) column, indicated that the purity of the peptides was >99.9%. Net peptide content was accurately determined on a Unicube elemental analyser (Elementar, Lyon, France) using manufacturer's protocols.

2.6. NMR Experiments

As no tridimensional structure of secapin-like peptide and Na_V toxin from myrmicine ant venoms has been published so far, we completed our study by doing a NMR analysis of the peptides U_{17}-MYRTX-Tb1a (glycosylated, U_{17}-Tb1a) and U_3-MYRTX-Tb1a (U_3-Tb1a) from *T. bicarinatum* venom, the first representatives of these kinds of peptides in myrmicine ants venoms (Touchard et al., 2018). Prior to NMR analysis each synthesized peptide was dissolved in the appropriate solvent. U_{17}-Tb1a was dissolved in H_2O:D_2O (9:1 v/v) at a concentration of 0.94 mM with pH adjusted to 5.2, and U_3-Tb1a was dissolved in 2,2,2-trifluoroethanol-d2 (TFE) 50% pH 4.6 at a concentration of 0.8 mM. Then, 2D ^{1}H-NOESY (200 and 150 ms, respectively), 2D ^{1}H-TOCSY (70 and 80 ms, respectively), a sofast-HMQC (Schanda et al., 2005) (^{15}N natural abundance) and a ^{13}C-HSQC (^{13}C natural abundance) spectra were performed at 298K on an Avance III HD BRUKER 700 MHz spectrometer equipped with a cryoprobe. ^{1}H chemical shifts were referenced to the water signal (4.77 ppm at 298K). NMR data were processed using the Topspin software version 3.6.2^{TM} (Bruker, Billerica, MA, USA) and analyzed with CCPNMR version 2.2.2 (Vranken et al., 2005).

Structures were calculated using the Cristallography and NMR System (CNS) (Brünger, 2007; Brünger et al., 1998) through the automatic assignment software ARIA2 version 2.3 (Rieping et al., 2007) with NOE derived distances, hydrogen bonds (in accordance with the observation of typical distance NOE cross peaks network for β-sheets in the case of U_{17}-Tb1a and for α-helix in the case of U_3-Tb1a), backbone dihedral angle restraints determined with the DANGLE program (Cheung et al., 2010) and the imposed disulfide bridge between Cys^9 and Cys^{20} only present in U_{17}-Tb1a. Strong Hα(X)–Hδ(Pro) NOEs were diagnostic of trans conformations about each X-Pro bond. The ARIA2 protocol, with default parameters, used simulated annealing with torsion angle and Cartesian space dynamics. The iterative process was repeated until the assignment of the NOE cross peaks was completed. The last run for each peptide was done with 1,000 initial structures and 250 structures were refined in water. Fifteen structures were selected on the basis of total energies and restraint violation statistics, to represent the tridimensional structure of U_{17}-Tb1a and U_3-Tb1a in solution. The quality of final structures was evaluated using PROCHECK-NMR (Laskowski et al., 1996) and PROMOTIF (Hutchinson and Thornton, 1996).

The structure calculations were done without taking into account the C-terminal amidation especially as no NOE could be observed between the amido C-terminal protons (A^{23}-NH_2 for

U_{17}-Tb1a and Asn^{23}-NH_2 for U_3-Tb1a) and all other protons of each peptide. However, the Coulombic electrostatic potential (ESP) and Molecular lipophilicity potential (MLP) were determined at the Connoly surface of peptide after C-terminal amidation using ChimeraX 1.3 (Pettersen et al., 2021). The figures were prepared with PYMOL (DeLano, 2002) and ChimeraX.

2.7. Toxin nomenclature and data accessibility

Novel myrmicitoxin sequences were named following the standard nomenclature for animal venom peptides (King et al., 2008), and we used myrmicitoxin (MYRTX) to define the venom peptides from the subfamily Myrmicinae and genus/species descriptors as follow; *Myrmica ruginodis* (Mru), *Pogonomyrmex californicus* (Pc), *Stenamma debile* (Sd), *Daceton armigerum* (Da), and *Tetramorium africanum* (Ta) (Touchard et al., 2016a). In cases of sequence similarities, we also named the novel myrmicitoxins according to the same subscripts used to denote the 'unknown' activity descriptor prefixes previously defined for myrmicine ant venom peptides (U_1-U_{20}), and forming mature myrmicitoxin families (Touchard et al., 2020a, 2018).

The transcriptomes raw data are available through the European Nucleotide Archive (ENA) under the project number PRJEB51841. The mass spectrometry data were deposited in the ProteomeXchange Consortium *via* the PRIDE (Perez-Riverol et al., 2022) partner repository with the dataset identifier PXD033547. NMR data and tridimensional structure coordinates have been deposited with SMSDep as Biological Magnetic Resonance Bank (BMRB) accession codes: 21098 (U_{17}-MYRTX-Tb1a) and 21099 (U_3-MYRTX-Tb1a).

3. Results

3.1. Proteotranscriptomics data

3.1.1. Proteomic data

LC-MS analysis revealed heterogeneous profiles, with great variation in the peptide molecular weight and retention time (*i.e.*, hydrophobicity) among species, overall ranging from 682 to 9,416 Da and from 1.84 to 51.62 min (Figures 1 and 2 – A). A total of 155 peptides were detected with great variations among species: from only 6 peptides detected in the venom of *Daceton armigerum* to up to 49 peptides in *Tetramorium africanum* venom (Figure 2 – A; Supplementary tables III-XIII). Venom peptides from *T. africanum* were distributed relatively widely between 1.84 and 46.24 min of elution, ranging from 744 and 3,338 Da (Supplementary

Table III – Figure 2 A). This profile is consistent with previous observations of venom from the genus *Tetramorium* (Rifflet et al., 2012; Touchard et al., 2018). Overall, the venoms of the species studied here contained mostly small peptides (*i.e.*, molecular weight < 5,000 Da) except for *D. armigerum* venom in which all peptides were detected in the narrow mass range 8,208-9,416 Da.

As *Solenopsis* is known to possess alkaloid-rich venoms, an hexane extraction was chosen and the aqueous fraction was submitted to LC-MS analysis (Figure 1). Nine masses having a molecular weight in the range 2,361 – 4,979 Da and eluting from 21.91 to 41.98 min were detected in the aqueous phase of *So. saevissima.* However, they were present in very low amount, as 90% of the aqueous phase was constituted by two masses (253.00 and 279.22 Da), which were identified as remaining piperidinic alkaloids (*i.e.*, 2-methyl-6-undecylpiperidine and 2-methyl-6-tridecyl-6-piperideine, respectively) thanks to the parallel GC-MS analysis of the hexane phase (Supplementary Table XIII – Figures 1 and 2-A).

LC-MS analysis of reduced/alkylated venoms revealed 21 disulfide-bonded venom peptides. Most of them had a single disulfide bond while two venom peptides were structured by three disulfide bonds in *My. ruginodis* and *P. californicus* venoms (Figure 2 – A). Furthermore, the HPLC profile of the venom of *D. armigerum* appeared completely different following the reduction/alkylation. The peaks were shifted for several minutes and the masses ranged from 4,162.3 to 4,765.7 Da (Figure 1), highlighting the presence of dimeric peptides.

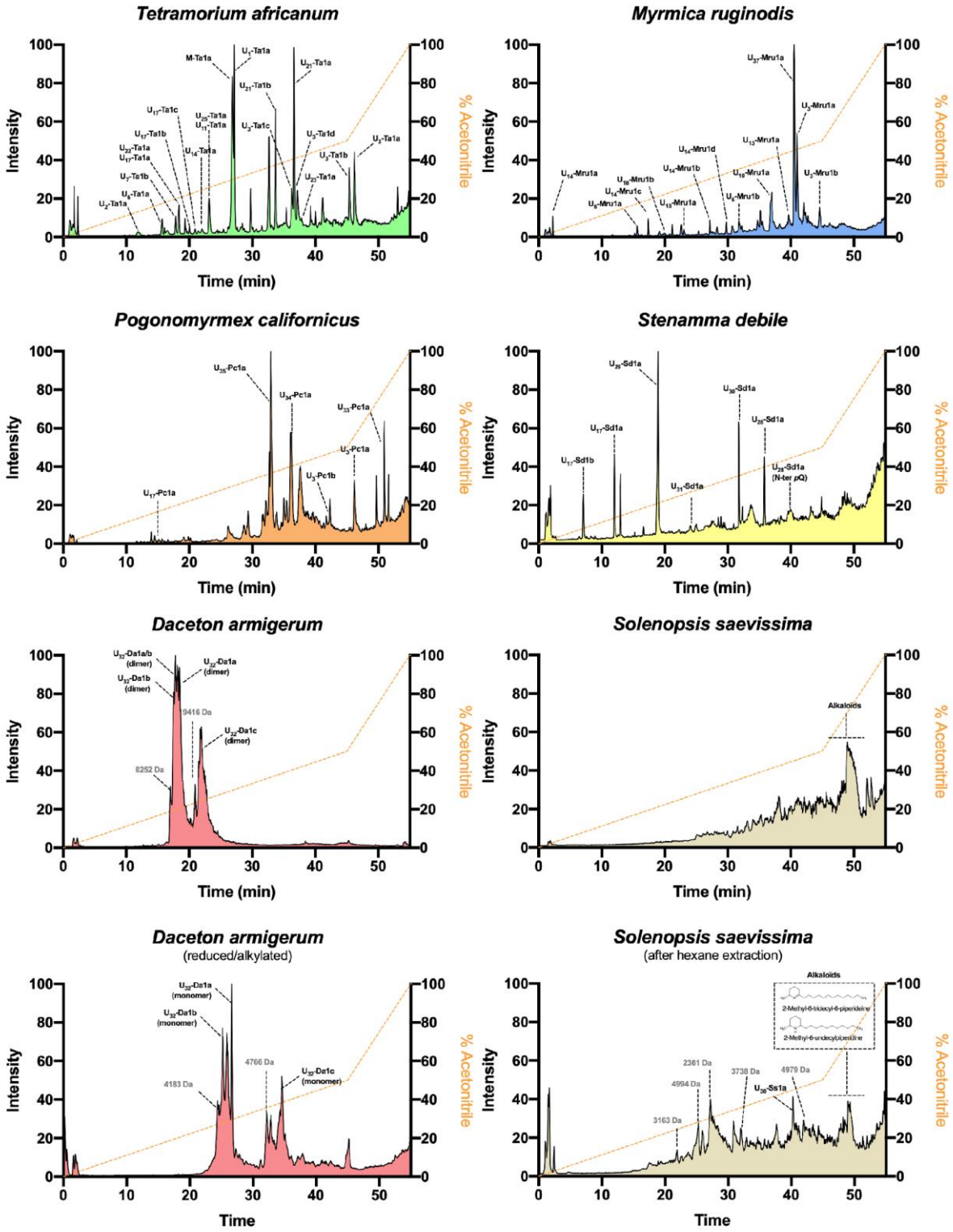

Figure 1: Total ion chromatograms (TIC) of crude venom of *Tetramorium africanum*, *Myrmica ruginodis*, *Pogonomyrmex californicus*, *Stenamma debile*, *Daceton armigerum*, and *Solenopsis saevissima*. Peptides were eluted through RP-HPLC on a C18 column using a linear H_2O/ACN gradient at a flow rate of 250 µL.min^{-1}.

Finally, a LC-MS/MS analysis of the six crude venoms was done using a Q-exactive orbitrap to achieve *de novo* sequencing and so acquire predictive peptide sequence tags. This resulted in a total of 6,314 sequence tags with an ALC score higher or equal to 60%, ranging from 145 to 2,898 sequence tags for *St. debile* and *T. africanum* venoms, respectively. These sequence tags were then blasted on the transcriptomic data to determine the complete sequences of venom peptides and to identify precursors.

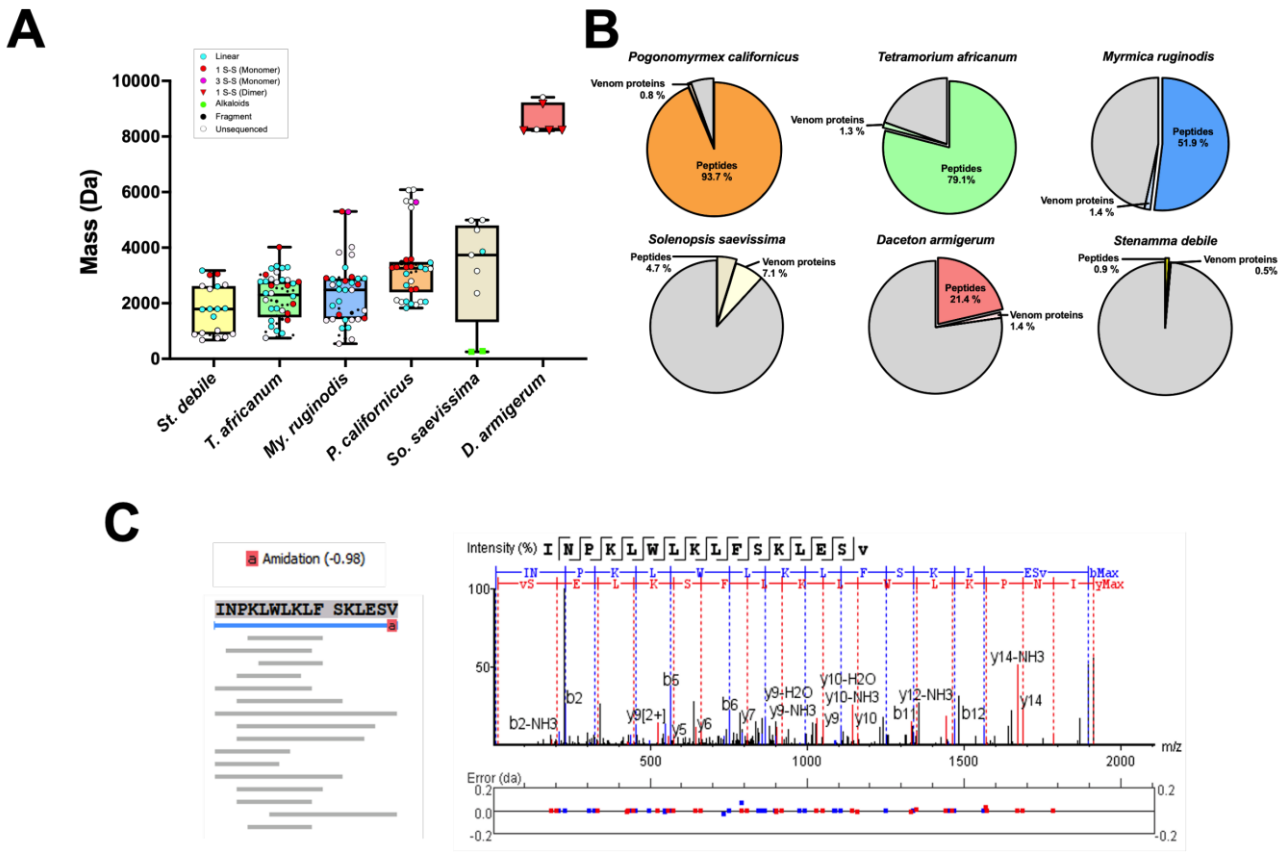

Figure 2: Proteotranscriptomic data of the venoms of *Stenamma debile*, *Tetramorium africanum*, *Myrmica ruginodis*, *Pogonomyrmex californicus*, *Solenopsis saevissima*, and *Daceton armigerum*. **(A)** Repartition of masses detected by LC-MS analysis in the six myrmicine ant venoms. **(B)** Proportion of addressed contigs expressing venom peptides and venom proteins in the six myrmicine ant venom gland transcriptomes. **(C)** An example of MS/MS spectrum for U_{14}-Mru1a leading to amino acid sequence confirmation by Peaks software.

3.1.2. Transcriptomic data

The sequencing of the total venom gland mRNAs resulted in a total of 319,437,422 raw reads, ranging from 19,389,866 to 82,055,072 for *D. armigerum* and *T. africanum*, respectively. *De novo* assembly with Trinity and Oases yielded a total of 181,237 and 102,277 contigs for

all venoms, with a mean contig size of 1,880 and 2,174 bp, respectively. Both assembly methods gave various numbers of contigs for each venom gland transcriptome, ranging from 13,701 for *D. armigerum* (Oases) to 48,353 for *St. debile* (Trinity) (Supplementary table XV). For each venom, a combined database from both assemblies was used to search for peptide sequences generated from LC-MS/MS analyses. Additionally, the signal sequences of the previously published myrmicitoxins (*i.e.*, from *Ma. rubida* and *T. bicarinatum*) were searched against these databases to find peptides sharing the same signal sequences. This permitted us to infer that most of the venom peptide encoding genes had high transcription levels as they accounted from 21% to 94% of the addressed contigs by the venom glands of *D. armigerum* and *P. californicus*, respectively (Figure 2 – B; Supplementary Information).

Apart from the other myrmicine ant transcriptomes presented here, only 7% of the addressed contigs from *So. saevissima* venom glands encoded putative venom peptides (Figure 2 – B, Supplementary Information), which is consistent with the low relative abundances of venom peptides noted in the proteomic data. Also, the transcripts encoding venom peptides of *St. debile* represented 1% of the addressed contigs (Figure 2 – B, Supplementary Information). As these transcriptomic data were not issued only from venom glands, our strategy does not allow us to evaluate the expression level in these organs but rather permits us to identify putative venom peptide precursors. However, the combination of proteomic and transcriptomic data allowed us to assign only one of the encoding sequences to one of the measured masses.

3.1.3. Global description of myrmicine ant venom peptidomes and molecular features of mature myrmicitoxins

The proteotranscriptomics approach led to the identification of 75 to 98% (*i.e.*, *St. debile* and *T. africanum*, respectively) of the peptide venom content of the studied myrmicine ants (Figure 3 – A – Supplementary Tables III-XIV). The amino acid sequences yielded from transcriptomic data were verified manually using mass spectrometry data by matching theoretical masses with those measured. This approach allowed us to find 79 putative precursors and to confirm 60 sequences of mature peptides with some having several post-translational modifications (PTMs) such as C-terminal amidation or disulfide bonds.

Most of the ant venoms studied here had a high content of linear polycationic peptides, containing several arginine and lysine residues (*i.e.*, families M, U_1, U_3, U_{10}, U_{13}, U_{14}, U_{24}, U_{25}, U_{26}, U_{28}, U_{30}, U_{31}, U_{33} and U_{34}) (Figure 3 – A – Supplementary Figure 1). Those short myrmicitoxins (*i.e.*, 10-31 amino acids long) constituted 62% of *T. africanum* peptidome,

whereas they composed 49% of *St. debile* venom (Figure 3 – A). However, the venom peptidomes of *P. californicus* and *My. ruginodis* were both dominated by two linear venom peptides uniquely found in these two species and exhibiting a negative net charge, due to the high content of aspartic and glutamic acids (*i.e.*, U_{35} and U_{37}, respectively). These two peptides represented 54% of their venom peptidome compositions. Besides, three families gathering anionic linear venom peptides were found in *T. africanum*, *St. debile* and *So. saevissima* venoms, representing 29% (*i.e.*, U_{21}), 17% (*i.e.*, U_{29}) and 7% (*i.e.*, U_{38}) of their compositions, respectively (Figure 3 – A, Supplementary Figure 1).

Apart from linear peptides, disulfide-bonded venom peptides were also found within myrmicine ant venom peptidomes. From 13 to 34 amino acids long, ten of them contained a single intrachain disulfide bond and were classified in nine families of mature toxins (*i.e.*, U_4, U_6, U_7, U_{11}, U_{17}, U_{18}, U_{22}, U_{23} and U_{27}) (Figure 3 – A). Most of them were also polycationic due to the presence of several lysine, arginine, and histidine residues (Supplementary Figure 2). They were, however, less abundant than their linear counterparts as their abundances ranged from 2% to 12% in *P. californicus* and *St. debile*, respectively (Figure 3 – A). Furthermore, six myrmicitoxins containing three disulfide bonds were identified in *P. californicus* and *My. ruginodis* venoms, accounting for less than 1% of each venom peptidome (*i.e.*, U_{18}). They were also predicted by ScanProsite to have an Epidermal Growth Factor-like domain (EGF), which is consistent with the EGF-like peptides recently reported from several ant venoms (Hurka et al., 2022; Robinson et al., 2018; Touchard et al., 2020). Also, the toxin U_{18}-MYRTX-Mru1b exhibited a missing 146 Da mass unit in its calculated mass, corresponding to an additional fucose glycan (Supplementary Tables XI and XII) in accordance with the presence of the consensus O-glycosylation site observed in its primary sequence -C_2XXXXS/TC_3- (Supplementary Figure 2).

Apart from other myrmicine ant peptidomes, the venom of *D. armigerum* was found to be exclusively composed of dimeric venom peptides. The transcriptomic data of *D. armigerum* venom glands led to the identification of three transcripts with calculated masses ranging from 4,104 to 4,581 Da and exhibiting a single cysteine. Together with the results from the LC-MS analysis of the reduced/alkylated venom, it enabled the identification of three monomers (-Da1a, -Da1b and -Da1c). Four theoretical masses for dimeric peptides matched with masses measured in the total ion chromatogram of the crude venom (*i.e.*, 8,231 Da, 9,161 Da, 8,209 Da and 8,220 Da), corresponding to the homodimer-Da1b, Da1c, Da1a and the heterodimer-Da1a/b, respectively) (Figure 1). These amphiphilic and polycationic dimers accounted for 84%

of the venom peptidome (Figure 3 – A). The chains composing them were very similar with a mean of 78% identity (Supplementary Figure 2). As they displayed no similarity with previously reported ant venom peptides, they were therefore gathered in a single family (*i.e.*, U_{32}).

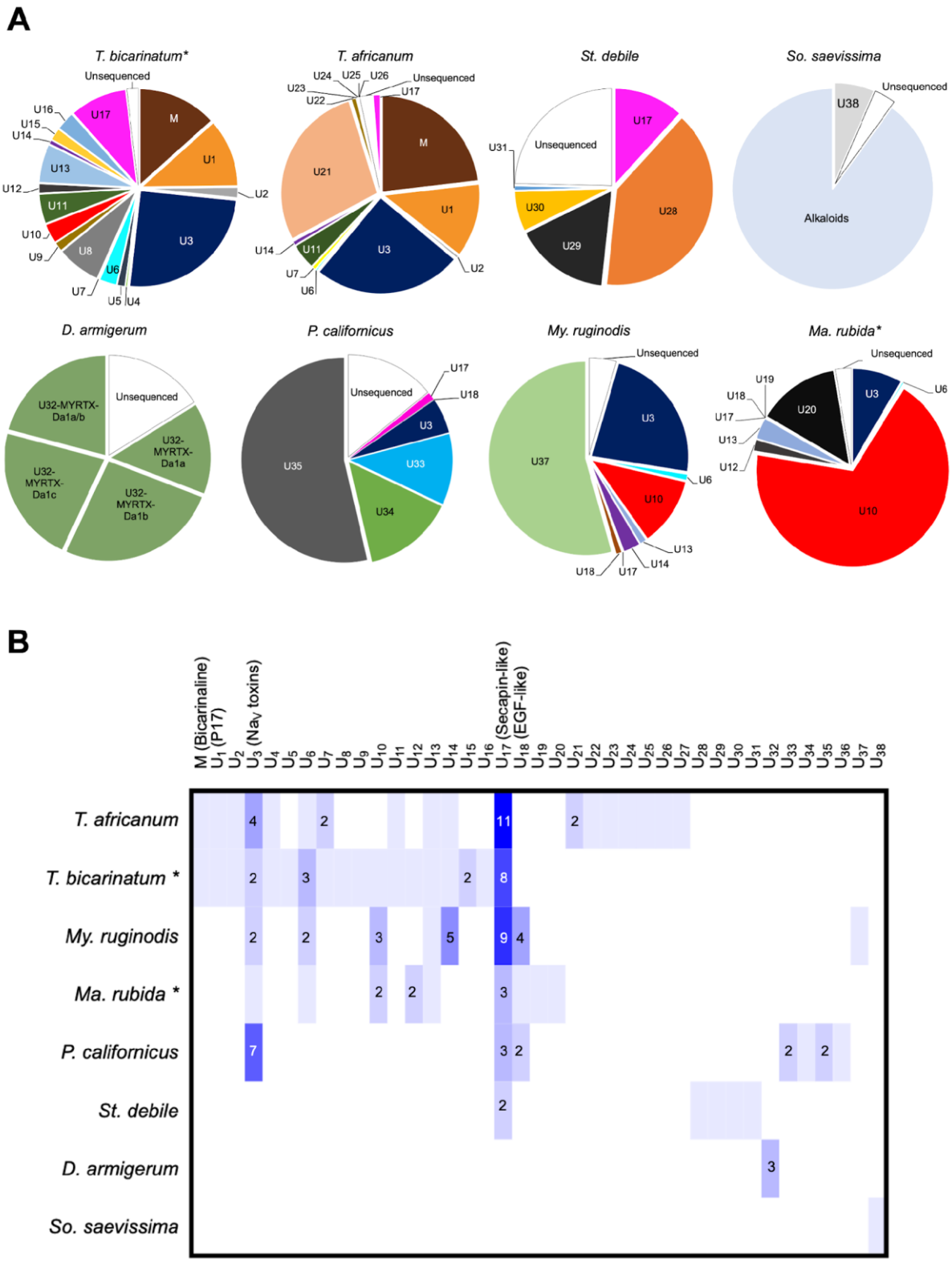

Figure 3: Global repartition of mature venom peptide families within myrmicine ant venom peptidomes. " * " denotes previously described venom peptidomes, which were characterized with a similar methodology from (Touchard et al., 2018). (**A)** Relative abundances of mature myrmicitoxin families. (B) Number of peptide variants found by mature myrmicitoxin family.

3.2. Superfamilies of myrmicitoxins precursors

The venom peptide precursors evolved at different rates, with signal sequences being more conserved than mature sequences. To define venom gene superfamilies, a focus on signal sequences appears thus more accurate (Kaas et al., 2010). To classify the venom peptide precursors, we performed a Hierarchical Cluster Analysis (HCA) based on a matrix distance of 118 aligned signal sequences. This HCA analysis revealed eight main clusters gathering precursors which shared at least 50% sequence identity with the defined consensus sequence (Figure 4 – Supplementary Table XXI). Since we included *T. bicarinatum* and *Ma. rubida* signal sequences in the analysis (Touchard et al., 2020a, 2018), we initially named the clusters according to the precursor superfamilies previously defined in these ant venoms (*i.e.*, A_1, A_2, B_1, B_2 and C) (Touchard et al., 2020a, 2018), and incremented this nomenclature to the new clusters (*i.e.*, A_3, A_4, C_1 and C_2) (Figure 4). However, this nomenclature is tentative and needs to be refined by using genomic data.

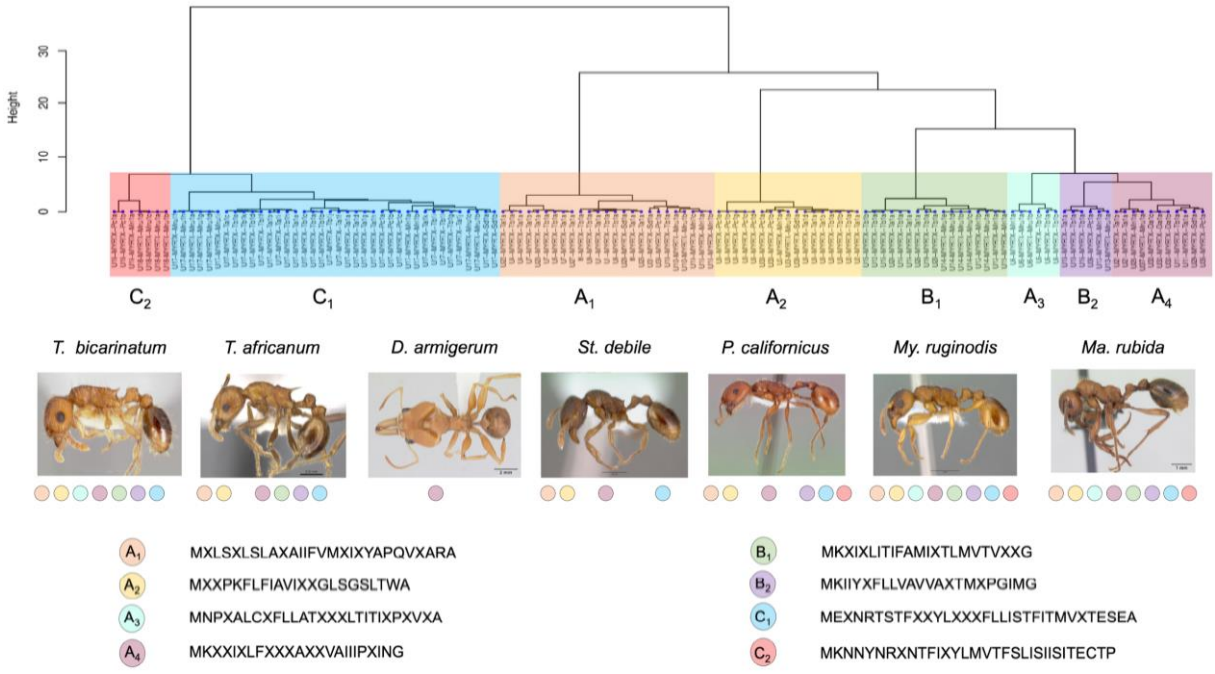

Figure 4: Clustering analysis of myrmicitoxin signal sequences. We included all described venom peptide precursors from myrmicine ant venoms, including those defined in *Tetramorium bicarinatum* and *Manica rubida* venom peptidomes (Touchard et al., 2020, 2018). The families were defined using a hierarchical cluster analysis based on signal sequences. Multiple alignments were achieved with *ClustalW* and pairwise distances were computed with *MEGAX* software (Kumar et al., 2018). Hierarchical cluster analysis was performed with the *R* software (R Core Team, 2017). The colored circles and their position under the pictures refer to superfamilies of myrmicitoxin precursors found. Every ant photography is extracted from AntWeb

(https://www.antweb.org) and can be found under the following specimen codes: *T. bicarinatum* (CASENT0005826), *Tetramorium africanum* (CASENT0280920), *Daceton armigerum* (CASENT0178489), *Stenamma debile* (CASENT0010691), *Pogonomyrmex californicus* (CASENT0005710), *Myrmica ruginodis* (CASENT0008642), *Ma. rubida* (CASENT0173135) (photographers: Aprile Nobile & Estella Ortega).

In Touchard *et al.* (2018), precursors from the superfamilies A and B have been defined as "pilosulin-like" peptides and a new superfamily of ant venom precursors, respectively. Precursors from these superfamilies are related to aculeatoxins and exhibit an organization in three regions: an N-terminal signal sequence, an alanine/glutamic acid rich prosequence with variable length and the mature region (Touchard et al., 2018). The prosequences of precursors from superfamily B are shorter than those from superfamily A, but the cleavage maturation site releasing the mature toxin is located after an alanine or a proline residue for both, as previously described for other ant venom precursors (Touchard et al., 2020a, 2018). While mature regions of A_1, A_4, B_1 and B_2 clusters are highly diverse in their sequences and biochemical properties, the mature sequences which are gathered into A_2 and A_3 clusters belonged mainly to the same peptide families, or shared the same physico-chemical properties (*i.e.*, U_3/U_{33} and U_6 peptides respectively) (Figure 4). The superfamily C was previously defined in *T. bicarinatum* and *Ma. rubida* venom peptidomes as gathering secapin-like (U_{17}) and EGF-like peptides (U_{18}) (Touchard et al., 2020a, 2018). Members of the superfamily C differed from precursors of the superfamilies A and B in their organization and maturation profiles. Indeed, they are predicted as Type II single pass transmembrane proteins with a short cytosolic N-terminal domain of five amino acids and an internal signal sequence directly followed by the mature sequence (Eagles et al., 2022; Touchard et al., 2018). Here, the clustering analysis split it into two separate clusters gathering U_{17} and U_{18} precursors (*i.e.*, C_1 and C_2, respectively) (Figure 4). The cleavage maturation sites of precursors from the C_1 family were more often located after a serine residue than an alanine or proline residue, whereas the cleavage sites of precursors from the C_2 family were located after a proline but also after a threonine or cysteine residue (Figure 4).

3.3. NMR structures of the secapin-like peptide U_{17}-MYRTX-Tb1a and the Na_V toxin U_3-MYRTX-Tb1a

Despite being found in the venom of numerous hymenoptera (Lee et al., 2016), no secapin and Na_V toxin (U_3) structure has been characterized yet. The 3D structure of U_{17}-Tb1a (glycosylated) and U_3-Tb1a previously characterized in the venom of the myrmicine ant *T. bicarinatum* (Touchard et al., 2018), was thus determined here using NMR spectroscopy.

The main U_{17}-Tb1a structure feature is a double-stranded antiparallel β-sheet connected by a β-turn, otherwise known as a β-hairpin motif (Figure 5 – B, Supplementary Figure 6). The single disulfide bridge between Cys^{9} and Cys^{20} forming a 12-membered macrocycle was likely to maintain the high degree of structure. The N-terminal 6 residues part of the peptide (1-6) appeared as an unstructured and highly flexible tail. Residues Pro^{10}, Pro^{11}, Gly^{12} and His^{13} formed a type II β-turn that led to the first β-strand (His^{13}, Val^{14}, Val^{15}, Val^{16}) and, Arg^{19}, Cys^{20}, Arg^{21} and Ile^{22} made up the other strand. These two β-strands were connected by a β-turn around Gly^{18} defined with PROMOTIF as type IV (VVKG) and type I' (VKGR). Some NOEs were observed between the side chains of residues Asn^{7} and Val^{15}-Val^{16} (β1-strand) and Arg^{19}-Cys^{20} (β2-strand), but due to an overlapping of their HN and Hα protons, the NOEs characteristics of an antiparallel β-sheet were possible but unobservable. In this case, a third β-strand (Asn^{7}-Arg^{8}) could be formed and the structure of U_{17}-Tb1a should be a triple-stranded antiparallel β-sheet. AlphaFold predicts this tendency for the complete sequence of U_{17}-MYRTX-Tb1a (57 AA including the signal sequence) as it shown in UniProtKB for the representation of Entry A0A6M6RE84 (SECP_TETBN). U_{17}-Tb1a was a highly charged molecule (Figure 5 – B, Supplementary Figure 6), counting 3 arginine and 1 lysine residues whose side chains were exposed to the solvent and spread on the same side of the molecular surface.

The U_3-Tb1a adopted a helical structure from Leu5 to Thr22 (Figure 6 – B; Supplementary Figure 7) defined with PROMOTIF as two α-helices (5-8 and 10-22) with a ω angle of (-45.5 ± 2.7) ° for 53% structures or a bent α-helix (5-22) with a deviation of (34.4 ± 2) ° for 47% structures. U_3-Tb1a was a highly hydrophobic peptide (Supplementary Figure 7 – D) as it contained 8 leucines, 3 prolines, 2 alanines, 2 glycines and 1 valine. Only the C-terminal part encompassed polar and charged residues (Gln^{18} and Lys^{20}). This would result in different interactions of U_3-Tb1a with cell membranes. It is conceivable that the N-terminal apolar helix favors transmembrane localization while the C-terminal amphiphilic helix is either solvent exposed or interacting with the membrane surface.

3.4. Identified selection on families U_3/U_{33} and U_{17}

The families U_3 (Na_V toxins) and U_{17} (secapin peptides) are mainly ubiquitous among myrmicine ant venoms with several variants among species. The functions described for secapins from *Apis* spp. venoms suggest their defensive roles as they induce pain and inflammation in mammals (Lee et al., 2016; Mourelle et al., 2014), and U_3 peptides were recently characterized as defensive toxins acting on mammalian Na_V channels (Robinson *et al.*

2022, in review for *Nat. Com.*). The sequence diversity highlighted here provides an ideal playground for analyzing the evolutionary mechanisms leading to these sequence variations. Regarding their similar biochemical features, U_{33} nucleotide sequences were added to the U_3 nucleotide sequences group for the analysis. For this, pairwise comparisons of nucleotide sequences of the mature peptides and the prepro regions of U_3/U_{33} and U_{17} precursors were then aligned separately on the software MEGAX (Supplementary Tables XIX to XXII).

Among secapin peptides (U_{17} family), a higher global dN/dS rate was noted within the *Manica* group in comparison to the other groups, except *Pogonomyrmex*. More generally, high global intragroup dN/dS rates were observed for *Manica*, *Stenamma*, and *Myrmica*, suggesting the action of a prospective positive selection on some sequences (Supplementary Table XIX). Consequently, a codon based-Z-test of positive and purifying selection was done, highlighting a significant positive selection within the group *Manica*, as well as in the other groups, except *Pogonomyrmex*. The same significant tendency was also noted for some sequence pairs of *Tetramorium vs. Myrmica; Tetramorium vs. Manica;* and *Myrmica vs. Manica* (Supplementary Table XXI). Although *Manica* and *Myrmica* groups contained only two sequences, these tests showed that sequence pairs encoding secapins were under positive selection (i.e., 50% of the pairs for *Myrmica* and 100% of them for *Manica*). A positive selection was also noted for 20% of the sequence pairs within the *Tetramorium* group (Supplementary Table XXI). MEME and FEL algorithms, used for the site-by-site analysis, predicted positions 5 and 23 as being under persistent positive selection. For position 23, notable modifications concerning the net charge according to whole sequences might have consequences on mature toxin structures and functions. Furthermore, the FEL algorithm predicted that the positions 2 and 6, as well as the second cysteine (*i.e.*, position 27) are under purifying selection. Besides, the positions 4, 16, 21, 23, 25, 28, and 32 were noted to be under episodic positive selection by the algorithm MEME. The first cysteine is thus the only invariant site when the positions presenting a majority of GAP were removed (Figure 5 – A).

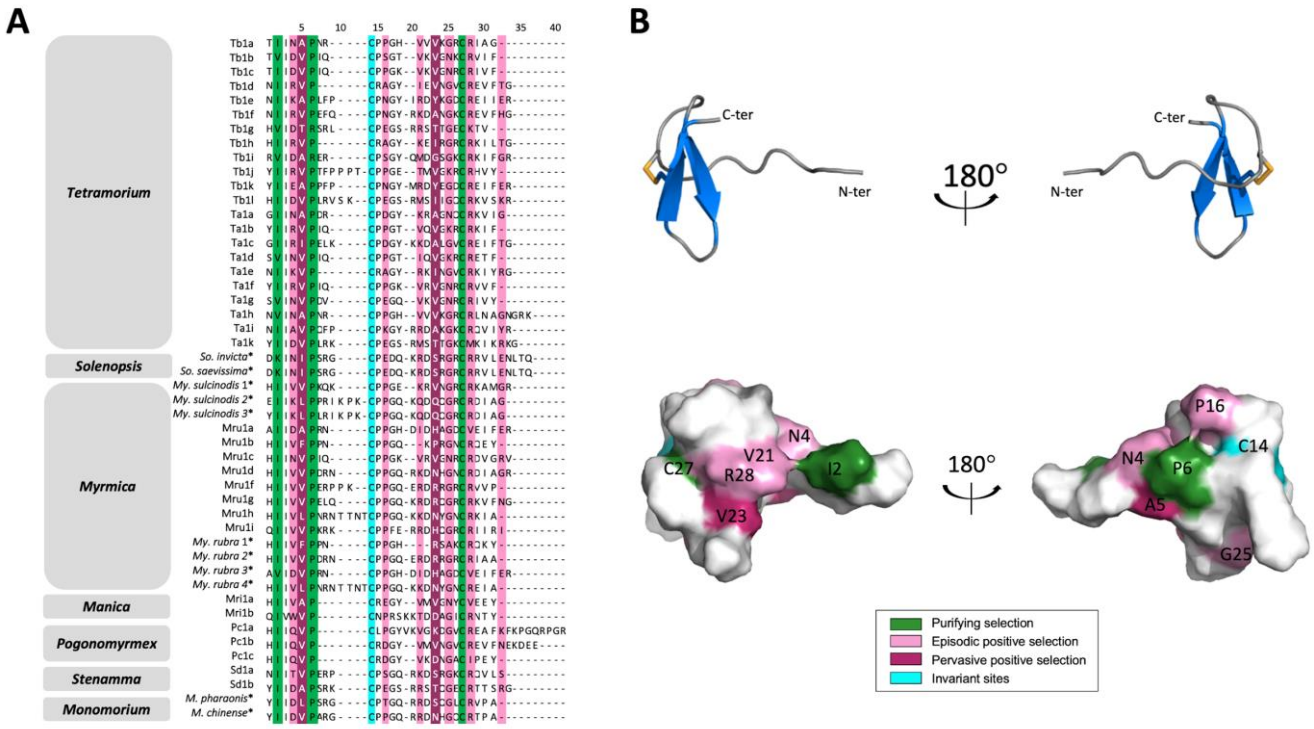

Figure 5: Evolutionary study of secapin-like peptides from myrmicine ant venoms and structural analysis of U_{17}-MYRTX-Tb1a. (**A**) Alignments of mature amino acid sequences of the peptides from U_{17} family and results of the site-by-site analysis. To estimate the evolutionary forces and the positions they were applied on, the alignments and analysis were performed under the software *MEGA X* and with the algorithms MEME and FEL, respectively (Kumar et al., 2018). The sequences followed by "*" were retrieved from transcriptomes which were available online on the ant genomics database - Fourmidable (https://antgenomes.org). (**B**) Primary structure, cartoon representation of the lowest-energy solution structure and molecular surface of glycosylated U_{17}-MYRTX-Tb1a. The disulfide bond is colored in yellow. The numbering of amino acids was defined according to the alignment of all U_{17} peptides.

Conversely, U_3 family dN/dS rates were mostly < 1 between groups or ≃ 1 within groups. Accordingly, codon-Z-based tests revealed only few significant pairs among groups, as well as within them, for both types of forces driving selection (*i.e.*, positive and purifying). The only significant pairs were indeed found inside *Tetramorium* and *Pogonomyrmex* groups (*i.e.*, 2 and 6 pairs, respectively), as well as between these two groups (*i.e.*, 13 pairs) and between *Tetramorium* and *Myrmica* (*i.e.*, 7 pairs) (Supplementary Tables XX and XXII). The site-by-site analysis revealed two common positions under pervasive positive selection (*i.e.*, 10 and 11) for both algorithms (Figure 6 – A). However, the FEL algorithms also predicted ten more positions under purifying selection (Supplementary Table XXII). Moreover, the encoding sequences of the prepro regions showed a low dN/dS ratio within groups, except for *Manica* (*i.e.*, dS = 0) and *Myrmica*. The codon based-Z-test revealed indeed that *Pogonomyrmex* and

Tetramorium prepro sequences were under purifying selection whereas *Manica*'s was under positive selection, as well as some prepro sequences from *Myrmica*. However, this concerned exclusively the prepro regions of the U_3/U_{33} pairs (Supplementary Table XXII).

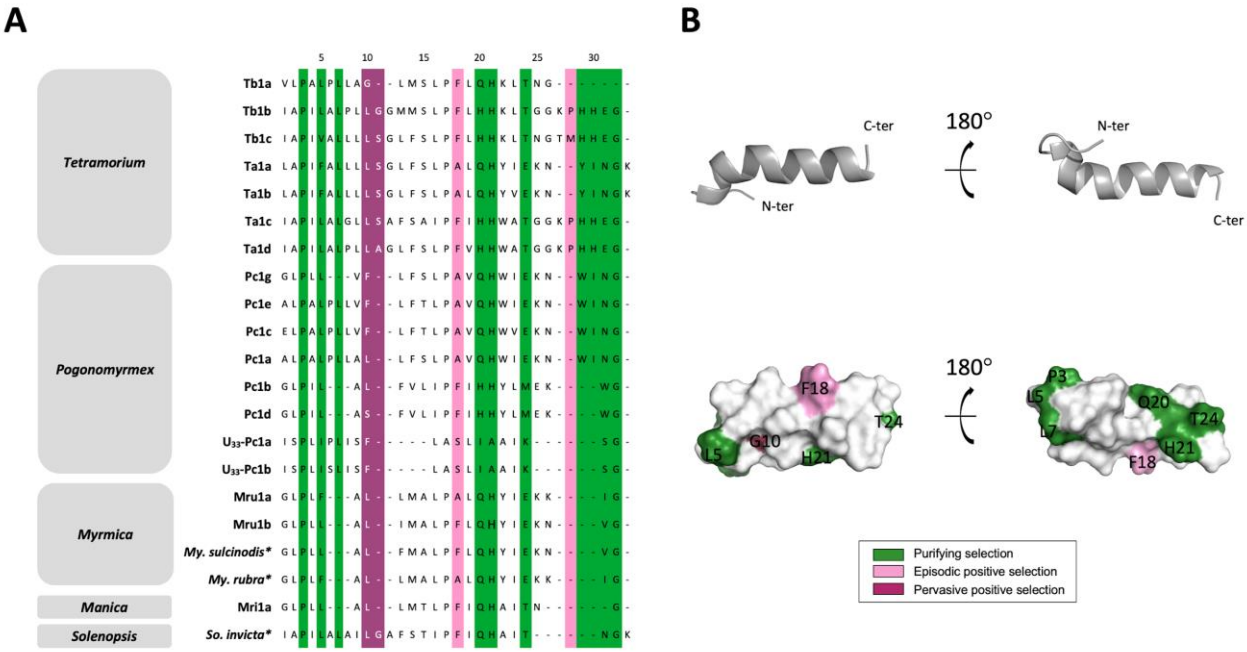

Figure 6: Evolutionary study of Na_V toxins from myrmicine ant venoms and structural analysis of U_3-MYRTX-Tb1a. (**A**) Alignments of mature amino acid sequences of the peptides from the U_3 family and results of the site-by-site analysis. To estimate the evolutionary forces and the positions they were applied on, the alignments and analysis were performed under the software *MEGA X* and with the algorithms MEME and FEL, respectively (Kumar et al., 2018). The sequences followed by "*" were retrieved from transcriptomes which were available online on the ant genomics database - Fourmidable (https://antgenomes.org). (**B**) Primary structure, cartoon representation of the lowest-energy solution structure and molecular surface of U_3-MYRTX-Tb1a. The numbering of amino acids was defined according to the alignment of all U_3 peptides.

4. Discussion

4.1. Heterogenous compositions of myrmicine ant venoms

Combined, our results and previous studies revealed that myrmicine ants have heterogeneous venom peptide compositions. The complexity of peptidomes varies in terms of venom profile and diversity of peptide families represented as a suggestive consequence to the wide variations of ecology, diet, and venom use of these ants. Some peptide families are exclusive to a single investigated species whereas others are represented in most venoms of our studied species. The venoms of Myrmicines also appear very different from the venoms of ants

belonging to other subfamilies previously studied. With the exception of EGF-like peptides which are also present in *Myrmecia gulosa* venom (Eagles et al., 2022; Robinson et al., 2018), we did not find any common peptide families with ant toxins from other investigated subfamilies such as Ponerinae and Myrmeciinae (Kazuma et al., 2017; Orivel et al., 2001; Radis-Baptista et al., 2020; Robinson et al., 2018; Touchard et al., 2016b). This study also allowed us to appreciate the structural diversity of the peptides. Along with linear peptides, several peptides with a disulfide bond were characterized, some of them even dimeric. In particular, we determined the β-hairpin structure of a secapin-like peptide, a toxin family which is also present in bee venoms but whose structure had never been elucidated before our study. We therefore provide a glimpse of the molecular and structural diversity of the peptides contained in the venoms of the large Myrmicine subfamily of ants. The venoms of the 5,000 stinging myrmicine ant species estimated remain underexplored for the search of new bioactive peptides. Some of these peptides could be promising for the discovery of new ligands to human receptors as exemplified by the peptide P17 (U_1-Tb1a) (Duraisamy et al., 2021).

The defensive function seems to be a major player in the evolution of the myrmicine ant venom peptidome compositions, which are associated with an aggressive collective defensive behavior. Our results, along with previous reported investigations, emphasize that myrmicine venoms contain numerous and diverse algesic peptides. A recent comprehensive pharmacological characterization of U_3 peptides from *T. africanum* and *Ma. rubida* venoms demonstrated that these peptides are vertebrate-selective pain-causing Na_v toxins (Robinson et al. 2022, in review for *Nat. Com*.). U_3 peptides are predominant in the majority of myrmicine ant venoms studied so far (Touchard et al., 2020a, 2018), including here, except for *D. armigerum* and *St. debile*. Na_V toxins (U_3) likely contribute to the painful sting of ants from the genera *Myrmica*, *Manica*, *Pogonomyrmex,* and *Tetramorium* which are defined as aggressive species while defending their nests against predators and intruders. The non-aggressive species *St. debile* and *D. armigerum* have evolved distinctive strategies to defend their colonies (*e.g.*, feign death and powerful bite, respectively). Their venoms are exclusively used for prey capture, which would have relaxed the selective pressures toward these Nav toxins (U_3) since exhibiting very weak lethal and paralytic effects against insects. In the same line, U_{17} peptides constitute another highly represented family in the analyzed venoms here. These toxins were previously shown to share significant sequence similarity with secapins (Touchard et al., 2018), which are toxins secreted in the venom of numerous hymenoptera, and for some of which inducing pain and inflammation in mammals (Lee et al., 2016; Mourelle et al., 2014). Further

pharmacological experiments are necessary to clearly understand the function and the mechanism of secapin-like peptides from ant venoms. The NMR data deposited here will help to this purpose. In addition, the toxin arsenal of *P. californicus, My. ruginodis,* and *Ma. rubida* comprises EGF-like peptides (U_{18}) (Touchard et al., 2020a) with 6 variants described here. If the biological activity of myrmicine ants EGF-like toxins remains to be experimentally demonstrated, a similar peptide from *Myrmecia gulosa* venom (Myrmeciinae) was recently characterized as an agonist of the mammalian receptor ErbB1, involved in pain signaling (Eagles et al., 2022). A similarity of U_{18} peptides from *Ma. rubida*, *My. ruginodis* and *My. rubra* venoms with the insect hormone Spitz were also noted, suggesting it might target the insect ErbB receptor of predators/prey (Eagles et al., 2022; Hurka et al., 2022). Besides, the venoms of both *Tetramorium* species contain U_1 peptides, for one of which is involved in the mammalian nocifensive responses through immunomodulatory effects on vertebrates. Peptide U_1-Tb1a from *T. bicarinatum* is indeed a pro-inflammatory peptide agonist of the MRGPRX2 receptor (Mas-related G protein–coupled receptor X2) (Duraisamy et al., 2021). The venoms described here also comprise several families of polycationic peptides with amphipathic structure which are presumably membrane-acting toxins and may contribute to the pain of these ant stings.

4.2. Molecular diversification of myrmicine ant venom peptides

Conserved signal regions of venom peptide precursors, coupled with divergent sequences of mature toxins, have already been observed in ant venoms with peptide precursors that were assigned to the family "pilosulin-like" (Barassé et al., 2019; Ceolin Mariano et al., 2019; Kazuma et al., 2017; Robinson et al., 2018; Touchard et al., 2018). It was recently suggested that aculeate Hymenoptera venom peptides having similar signal sequences might share a gene family and have evolved through gene duplication and neofunctionalization (Robinson et al., 2018; Touchard et al., 2018). Across myrmicine ant venoms, the 8 clusters of precursors defined here led to the maturation of 38 families of mature peptides. Even if further genomic studies are still needed, our results fit with these evolutive mechanisms. Among the described peptide families, secapin-like peptides and Na_V toxins were represented in almost all myrmicine ant venom peptidomes therefore including multiple variants (*i.e.*, U_3 and U_{17}). Considering the species in which U_3 (*i.e.*, Na_V toxins), and U_{17} (*i.e.*, secapin-like) toxins were retrieved, their venom uses and the biological activities reported, it seems thus likely that these toxins might have been retained and diversified predominantly as defensive weapons against vertebrates.

Further bioinformatic analyses were consequently carried out on the nucleotidic encoding sequences to estimate their evolutionary divergence and the driving forces of selection in action.

We underlined that positive selection affected secapin sequences of myrmicine species. The codon located in the 23rd position of the U_{17} alignment is significantly under pervasive positive selection. This position corresponds to Val_{16} of U_{17}-MYRTX-Tb1a embedded in the first β-strand or in the β-turn connecting the two β-strands. The 23rd position is of particular interest as the amino acid residue variations observed here in charge and polarity could affect the conformation and the function of the venom peptide, and confer functional adaptations to new targets. Position 23 is moreover close to other positions on positive or purifying selection on the 3D representation of U_{17}-Tb1a., constituting a patch on the surface of one side of the peptide. The same phenomenon is observed on the opposite side. These two areas could thus be crucial for the interaction of U_{17}-Tb1a and its molecular target. Variation or conservation patterns in these areas might affect affinity of the peptide with its receptor and confer functional adaptation to new biological targets.

Concerning the Na_V toxins of myrmicine ant venoms, the weak positive selection and the multiple codons under purifying selection suggest a structural and/or functional conservation of U_3 peptides. The positions 10 and 11 are under pervasive positive selection. For U_3-Tb1a, the position 10 corresponds to a glycine which is localized within the helix tilt. Positions under purifying selection are located along the helix, suggesting a global conservation of the helicoidal structure of the U_3 family. The weak global positive selection acting on U_3 sequences could be associated with a conservation of molecular targets and function for U_3 peptides from myrmicine ant venoms. Therefore, the positive selection on positions 10 and 11 could also take part in interactions with new molecular targets. To identify them, it will thus be interesting to test the action of U_3 peptides issued from different myrmicine ant venoms on mammalian Na_V, as well as to extend this study on other vertebrate Na_V.

In venomous animals, there is accumulated evidence suggesting that natural selection for diet is a potent driver of venom evolution being responsible for generating variation in venom compositions. This is potentiated by the dual factors of evolving resistance to venom in some prey and the metabolic expense of producing venom (Casewell et al., 2013). The evolution of venom toxin families *via* the 'birth and death' model is often accompanied by robust evidences of accelerated evolution and positive selection (Kordiš and Gubenšeka, 2000). Positive selection is near-universal among studied trophically venomous taxa, including snakes (Fry et al., 2003), scorpions (Zhu et al., 2004), spiders (Binford et al., 2009), and cone snails

(Chang and Duda, 2012; Dutertre et al., 2014). Gene duplication, positive selection, and protein neofunctionalization therefore appear to work in unison to provide the evolutionary novelty that allows adaptation of venoms to different prey (Duda and Palumbi, 1999). However, defensive toxins do not seem to follow this classic pattern of functional evolution and there is little evidence for defense-related selective pressures on venom compositions, except in cone snails (Dutertre et al., 2014). The understanding of these pressures as well as their role in venom evolution remains poor (Casewell et al., 2013). Thus, defensive toxins evolved slowly and appear with strong sequence conservation (Herzig et al., 2020), which is consistent with our data for Na_V toxins (U_3 peptides). Surprisingly, secapin-like peptides (U_{17}) showed the same pattern as trophically toxins being mostly affected by positive selection. These results raise further evolutive questions and highlight the need for more functional studies. Overall, in the absence of doing the required genomic investigations, further transcriptomic analysis on myrmicine ant venoms will also be needed to access more sequences and confirm these evolutionary mechanisms of diversification.

Acknowledgements: V.B. was the recipient of a PhD fellowship from the French Ministry of Scientific Research. This study was conducted in collaboration with the GeT core facility, Toulouse, France (http://get.genotoul.fr). Financial supports were provided by the France Génomique National institutional infrastructure partly funded "Investissement d'avenir" program managed by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09), by another Investissement d'Avenir grant of the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01) and by the PO-FEDER 2014-2020, Région Guyane (FORMIC, GY0013708). We would like to thank Robert A. Johnson from Arizona State University, for assistance and delivery of *P. californicus* workers. Ant samples were collected under the authorizations of both the Cameroon Ministry of Scientific Research and Innovation and the French Ministry of Ecological and Solidarity Transition, in accordance with Article 17, paragraph 2, of the Nagoya Protocol on Access and Benefit-sharing (Reference number of the permits: TREL1820249A/54 and TREL1916196S/214). The authors thank the NMR division of the MO2VING facility (Orléans, France).

References:

Aili, S.R., Touchard, A., Escoubas, P., Padula, M.P., Orivel, J., Dejean, A., Nicholson, G.M., 2014. Diversity of peptide toxins from stinging ant venoms. Toxicon 92, 166–178.

https://doi.org/10.1016/j.toxicon.2014.10.021
Aili, S.R., Touchard, A., Hayward, R., Robinson, S.D., Pineda, S.S., Lalagüe, H., Vetter, I., Undheim, E.A.B., Kini, R.M., Escoubas, P., Padula, M.P., Myers, G.S.A., Nicholson, G.M., 2020. An integrated proteomic and transcriptomic analysis reveals the venom complexity of the bullet ant Paraponera clavata. Toxins (Basel). 12. https://doi.org/10.3390/toxins12050324
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
Barassé, V., Touchard, A., Téné, N., Tindo, M., Kenne, M., Klopp, C., Dejean, A., Bonnafé, E., Treilhou, M., 2019. The peptide venom composition of the fierce stinging ant Tetraponera aethiops (Formicidae: Pseudomyrmecinae). Toxins (Basel). 11. https://doi.org/10.3390/toxins11120732
Blaimer, B.B., Ward, P.S., Schultz, T.R., Fisher, B.L., Brady, S.G., 2018. Paleotropical diversification dominates the evolution of the hyperdiverse ant tribe Crematogastrini (Hymenoptera: Formicidae). Insect Syst. Divers. 2. https://doi.org/10.1093/isd/ixy013
Brünger, A.T., 2007. Version 1.2 of the crystallography and nmr system. Nat. Protoc. 2, 2728–2733. https://doi.org/10.1038/nprot.2007.406
Brünger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grossekunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L., 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921. https://doi.org/10.1107/S0907444998003254
Cabau, C., Escudié, F., Djari, A., Guiguen, Y., Bobe, J., Klopp, C., 2017. Compacting and correcting Trinity and Oases RNA-Seq de novo assemblies. PeerJ 5, e2988. https://doi.org/10.7717/peerj.2988
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. https://doi.org/10.1016/j.tree.2012.10.020
Cheung, M.S., Maguire, M.L., Stevens, T.J., Broadhurst, R.W., 2010. DANGLE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. J. Magn. Reson. 202, 223–233. https://doi.org/10.1016/j.jmr.2009.11.008
Daly, N.L., Wilson, D., 2018. Structural diversity of arthropod venom toxins. Toxicon 152, 46–56. https://doi.org/10.1016/j.toxicon.2018.07.018
DeLano, W.L., 2002. PyMOL: an open-source molecular graphic tool. Ccp4.Ac.Uk.
Dos Santos Pinto, J.R.A., Fox, E.G.P., Saidemberg, D.M., Santos, L.D., Da Silva Menegasso, A.R., Costa-Manso, E., MacHado, E.A., Bueno, O.C., Palma, M.S., 2012. Proteomic view of the venom from the fire ant Solenopsis invicta buren. J. Proteome Res. 11, 4643–4653. https://doi.org/10.1021/pr300451g
Duraisamy, K., Singh, K., Kumar, M., Lefranc, B., Bonnafé, E., Treilhou, M., Leprince, J., Chow, B.K.C., 2021. P17 induces chemotaxis and differentiation of monocytes via MRGPRX2-mediated mast cell-line activation. J. Allergy Clin. Immunol. https://doi.org/10.1016/j.jaci.2021.04.040
Eagles, D.A., Saez, N.J., Krishnarjuna, B., Bradford, J.J., Chin, Y.K., 2022. A peptide toxin in ant venom mimics vertebrate EGF-like hormones to cause long-lasting hypersensitivity in mammals. Proc. Natl. Acad. Sci. 119, e2112630119. https://doi.org/10.1073/pnas.2112630119/-/DCSupplemental.Published
Fox, E.G.P., Russ Solis, D., Delazari dos Santos, L., Aparecido dos Santos Pinto, J.R., Ribeiro da Silva Menegasso, A., Cardoso Maciel Costa Silva, R., Sergio Palma, M., Correa Bueno,

O., de Alcântara Machado, E., 2013. A simple, rapid method for the extraction of whole fire ant venom (Insecta: Formicidae: Solenopsis). Toxicon 65, 5–8. https://doi.org/10.1016/j.toxicon.2012.12.009
Heep, J., Klaus, A., Kessel, T., Seip, M., Vilcinskas, A., Skaljac, M., 2019a. Proteomic analysis of the venom from the ruby ant Myrmica rubra and the isolation of a novel insecticidal decapeptide. Insects 10. https://doi.org/10.3390/insects10020042
Heep, J., Skaljac, M., Grotmann, J., Kessel, T., Seip, M., Schmidtberg, H., Vilcinskas, A., 2019b. Identification and functional characterization of a novel insecticidal decapeptide from the myrmicine ant Manica rubida. Toxins (Basel). 11, 1–17. https://doi.org/10.3390/toxins11100562
Hurka, S., Brinkrolf, K., Ozbek, R., Forster, F., Billion, A., Heep, J., Timm, T., Lochnit, G., Vilcinskas, A., Luddecke, T., 2022. Venomics of the central european myrmicine ants Myrmica rubra and Myrmica ruginodis. Toxins (Basel). 14. https://doi.org/10.3390/toxins14050358
Hutchinson, E.G., Thornton, J.M., 1996. PROMOTIF: a program to identify and analyze structural motifs in proteins. Protein Sci. 5, 212–220.
Jones, T.H., Zottig, V.E., Robertson, H.G., Snelling, R.R., 2003. The venom alkaloids from some african Monomorium species. J. Chem. Ecol. 29, 2721–2727.
Kaas, Q., Westermann, J.C., Craik, D.J., 2010. Conopeptide characterization and classifications: An analysis using ConoServer. Toxicon 55, 1491–1509. https://doi.org/10.1016/j.toxicon.2010.03.002
Kazuma, K., Masuko, K., Konno, K., Inagaki, H., 2017. Combined venom gland transcriptomic and venom peptidomic analysis of the predatory ant Odontomachus monticola. Toxins (Basel). 9, 1–15. https://doi.org/10.3390/toxins9100323
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
Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096
Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8, 477–486.
Lee, K.S., Kim, B.Y., Yoon, H.J., Choi, Y.S., Jin, B.R., 2016. Secapin, a bee venom peptide, exhibits anti-fibrinolytic, anti-elastolytic, and anti-microbial activities. Dev. Comp. Immunol. 63, 27–35. https://doi.org/10.1016/j.dci.2016.05.011
Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595. https://doi.org/10.1093/bioinformatics/btp698
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Mariano, D.O.C., de Oliveira, Ú.C., Zaharenko, A.J., Pimenta, D.C., Rádis-Baptista, G., Prieto-Da-Silva, Á.R. de B., 2019. Bottom-up proteomic analysis of polypeptide venom components of the giant ant Dinoponera quadriceps. Toxins (Basel). 11. https://doi.org/10.3390/toxins11080448
Morgan, E.D., 2008. Chemical sorcery for sociality: exocrine secretions of ants (Hymenoptera: Formicidae). Myrmecol. News 11, 79–90.
Mourelle, D., Brigatte, P., Bringanti, L.D.B., De Souza, B.M., Arcuri, H.A., Gomes, P.C., Baptista-Saidemberg, N.B., Ruggiero Neto, J., Palma, M.S., 2014. Hyperalgesic and edematogenic effects of Secapin-2, a peptide isolated from Africanized honeybee (Apis

mellifera) venom. Peptides 59, 42–52. https://doi.org/10.1016/j.peptides.2014.07.004
Orivel, J., Redeker, V., Le Caer, J.P., Krier, F., Revol-Junelles, A.M., Longeon, A., Chaffotte, A., Dejean, A., Rossier, J., 2001. Ponericins, new antibacterial and insecticidal peptides from the venom of the ant Pachycondyla goeldii. J. Biol. Chem. 276, 17823–17829. https://doi.org/10.1074/jbc.M100216200
Pan, J., Hink, W.F., 2000. Isolation and characterization of myrmexins, six isoforms of venom proteins with anti-inflammatory activity from the tropical ant, Pseudomyrmex triplarinus. Toxicon 38, 1403–1413. https://doi.org/10.1016/S0041-0101(99)00233-0
Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., Walzer, M., Wang, S., Brazma, A., Vizcaíno, J.A., 2022. The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 50, D543–D552. https://doi.org/10.1093/nar/gkab1038
Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., Ferrin, T.E., 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82. https://doi.org/10.1002/pro.3943
R Core Team, 2017. R: a language and environment for statistical computing. R. Found. Stat. Comput.
Radis-Baptista, G., Dodou, H. V., Prieto-da-Silva, A.R.B., Zaharenko, A.J., Nihei, K., Inagaki, H., Mori-Yasumoto, K., Konno, K., 2020. Comprehensive analysis of peptides and low molecular weight components of the giant ant Dinoponera quadriceps venom. Biol. Chem. https://doi.org/10.1515/hsz-2019-397ja-01
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2
Rieping, W., Bardiaux, B., Bernard, A., Malliavin, T.E., Nilges, M., 2007. ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics 23, 381–382. https://doi.org/10.1093/bioinformatics/btl589
Rifflet, A., Gavalda, S., Téné, N., Orivel, J., Leprince, J., Guilhaudis, L., Génin, E., Vétillard, A., Treilhou, M., 2012. Identification and characterization of a novel antimicrobial peptide from the venom of the ant Tetramorium bicarinatum. Peptides 38, 363–370. https://doi.org/10.1016/j.peptides.2012.08.018
Robinson, S.D., Mueller, A., Clayton, D., Starobova, H., Hamilton, B.R., Payne, R.J., Vetter, I., King, G.F., Undheim, E.A.B., 2018. A comprehensive portrait of the venom of the giant red bull ant, Myrmecia gulosa, reveals a hyperdiverse hymenopteran toxin gene family. Sci. Adv. 4, eaau4640. https://doi.org/10.1126/sciadv.aau4640
Schmidt, J.O., Blum, M.S., 1978. The biochemical constituents of the venom of the harvester ant, Pogonomyrmex badius. Comp. Biochem. Physiol. Part C, Comp. 61, 239–247. https://doi.org/10.1016/0306-4492(78)90137-5
Touchard, A., Aili, S.R., Fox, E.G.P., Escoubas, P., Orivel, J., Nicholson, G.M., Dejean, A., 2016a. The biochemical toxin arsenal from ant venoms. Toxins (Basel). 8, 1–28. https://doi.org/10.3390/toxins8010030
Touchard, A., Aili, S.R., Téné, N., Barassé, V., Klopp, C., Dejean, A., Kini, R.M., Mrinalini, M., Coquet, L., Jouenne, T., Lefranc, B., Leprince, J., Escoubas, P., Nicholson, G.M., Treilhou, M., Bonnafé, E., 2020. Venom peptide repertoire of the european myrmicine ant Manica rubida: identification of insecticidal toxins. J. Proteome Res. 19, 1800–1811. https://doi.org/10.1021/acs.jproteome.0c00048
Touchard, A., Brust, A., Cardoso, F.C., Chin, Y.K.Y., Herzig, V., Jin, A.H., Dejean, A., Alewood, P.F., King, G.F., Orivel, J., Escoubas, P., 2016b. Isolation and characterization of a structurally unique β-hairpin venom peptide from the predatory ant Anochetus

emarginatus. Biochim. Biophys. Acta - Gen. Subj. 1860, 2553–2562. https://doi.org/10.1016/j.bbagen.2016.07.027
Touchard, A., Koh, J.M.S., Aili, S.R., Dejean, A., Nicholson, G.M., Orivel, J., Escoubas, P., 2015. The complexity and structural diversity of ant venom peptidomes is revealed by mass spectrometry profiling. Rapid Commun. Mass Spectrom. 29, 385–396. https://doi.org/10.1002/rcm.7116
Touchard, A., Téné, N., Chan Tchi Song, P., Lefranc, B., Leprince, J., Treilhou, M., Bonnafé, E., 2018. Deciphering the molecular diversity of an ant venom peptidome through a venomics approach. J. Proteome Res. 17, 3503–3516. https://doi.org/10.1021/acs.jproteome.8b00452
von Sicard, N.A.E., Candy, D.J., Anderson, M., 1989. The biochemical composition of venom from the pavement ant (Tetramorium caespitum L.). Toxicon 27, 1127–1133. https://doi.org/10.1016/0041-0101(89)90006-8
Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., Laue, E.D., 2005. The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins Struct. Funct. Genet. 59, 687–696. https://doi.org/10.1002/prot.20449
Walker, A.A., Robinson, S.D., Yeates, D.K., Jin, J., Baumann, K., Dobson, J., Fry, B.G., King, G.F., 2018. Entomo-venomics: the evolution, biology and biochemistry of insect venoms. Toxicon 154, 15–27. https://doi.org/10.1016/j.toxicon.2018.09.004
Ward, P.S., Brady, S.G., Fisher, B.L., Schultz, T.R., 2015. The evolution of myrmicine ants: phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). Syst. Entomol. 40, 61–81. https://doi.org/10.1111/syen.12090