



HAL
open science

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.

► To cite this version:

Valentine Barassé, Nathan Téné, Christophe Klopp, Françoise Paquet, Niklas Tysklind, et al.. Venomics survey of six myrmicine ants provides insights into the molecular and structural diversity of their peptide toxins. *Insect Biochemistry and Molecular Biology*, 2022, 151, pp. 103876. 10.1016/j.ibmb.2022.103876 . hal-04446316

HAL Id: hal-04446316

<https://hal.science/hal-04446316v1>

Submitted on 15 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

1 In preparation for *Insect Biochemistry and Molecular Biology*

2

3 **Venomics survey of six myrmicine ants provides insights into the** 4 **molecular and structural diversity of their peptide toxins**

5

6 *Valentine Barassé^{1*}, Nathan Téné¹, Christophe Klopp², Françoise Paquet³, Niklas Tysklind⁴,*
7 *Valérie Troispoux⁴, Hadrien Lalague⁵, Jérôme Orivel⁵, Benjamin Lefranc⁶, Jérôme Leprince⁶,*
8 *Martin Kenne⁷, Maurice Tindo⁷, Michel Treilhou¹, Axel Touchard^{1,5#*}, Elsa Bonnafé^{1#}*

9

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

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

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

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

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

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

27 ⁷ Laboratory of Animal Biology and Physiology, Faculty of Science, University of Douala,
28 Cameroon, P.O.Box. 24157 Douala, Cameroon; tindodouala@yahoo.com (M.T.);
29 medoum68@yahoo.fr (M.K.)
30

31 # These authors contributed equally to this study.
32

33 Corresponding authors:

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

36 * CNRS, UMR EcoFoG (AgroParisTech, CNRS, CIRAD, INRAE, Université des
37 Antilles, Université de Guyane), 97310 Kourou, France. Email:
38 axel.touchard2@gmail.com
39

40 Declarations of interest: none
41

42 **Abstract:**

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

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

58

59 **Keywords:** toxin precursor; secapins; voltage-gated sodium channel (Nav) peptide; NMR
60 structure; dimeric peptide

61

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

64

65 **1. Introduction**

66 Venoms are fine-tuned biochemical arsenals mainly used by animals to defend
67 themselves and/or capture prey (Casewell et al., 2013). Investigations conducted on arthropod
68 venoms (*e.g.*, scorpions, spiders, centipedes, and insects) revealed complex mixtures of toxins
69 with peptides as the dominant reported components. These venoms represent an immense
70 source of structurally diverse toxins with variable amino acid sequences and tridimensional
71 structures (Walker et al., 2018; Wilson and Daly, 2018). However, the peptide toxins of most
72 insect venoms such as those from ants remain largely uncharacterized, despite being among the
73 most abundant and ecologically dominant terrestrial venomous animals. The small size of these
74 insects and consequently the difficulty in recovering large amounts of venom partly explains
75 why ant venom peptides have been so far overlooked (Touchard et al., 2016a). With over 14,000
76 extant species, ants form a diverse group of social venomous hymenopterans capable to inject
77 or spray secretions from a venom reservoir mainly for defense or prey capture. Ant venoms
78 harbor a clear potential for the discovery of novel peptides with original scaffolds and
79 pharmacology. Recent proteo-transcriptomic investigations provided comprehensive venom
80 peptidomes from several subfamilies including Pseudomyrmecinae, Paraponerinae,
81 Myrmicinae, Ponerinae and Myrmeciinae (Aili et al., 2020; Barassé et al., 2019; Kazuma et al.,
82 2017; Robinson et al., 2018; Touchard et al., 2018, 2020a) and started to reveal the molecular
83 diversity of these toxins. Although these venoms appeared to be less complex in peptide
84 numbers and diversity than those from cone snails or spiders, several interesting peptide
85 structures have been indeed characterized, such as linear helicoidal, dimeric, β -hairpin, EGF-
86 like, ICK-like or Kunitz-like peptides (Barassé et al., 2019; Radis-Baptista et al., 2020;

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

89 An extensive inclusion of ant species from unexplored clades and with broader
90 ecological diversity would thus enhance our ant venom molecular diversification understanding
91 and lead to further discoveries. Among ants, Myrmicinae is the most speciose ant subfamily
92 with 147 valid extant genera (Antweb, consulted on 22/04/22 and AntCat, consulted on
93 22/04/22), among which 111 are stinging species capable of injecting their venom (Blanchard
94 and Moreau, 2017; and personal observations A.T.). Myrmicines occur in most terrestrial
95 habitats and exhibit wide variation in ecology, behavior, and diet (Blaimer et al., 2018; Ward
96 et al., 2015). Reported observations demonstrated a predominance of alkaloids in some genera
97 such as *Solenopsis* and *Monomorium* (Jones et al., 2003; Morgan, 2008; Touchard et al., 2016a).
98 Also, proteomic analyses have revealed that the venoms of four genera (*i.e.*, *Tetramorium*,
99 *Pogonomyrmex*, *Myrmica*, and *Manica*) are peptide-rich (Heep et al., 2019a, 2019b; Hurka et
100 al., 2022; Schmidt and Blum, 1978; Touchard et al., 2020a, 2018; von Sicard et al., 1989).
101 Further investigations into the venom composition of other myrmicine species is therefore
102 necessary to report the molecular and structural peptide diversity, to understand the
103 heterogeneity and the intrinsic chemical properties of these venoms as well as to gather
104 information on relationships between ant phylogeny and venom composition.

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

115

116 **2. Materials and Methods**

117 *2.1. Collection and preparation of venom samples*

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

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

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

130

131 2.2. Mass spectrometry analysis

132 A preliminary LC–MS analysis of crude venoms was conducted on the LCQ-Ion trap
133 Advantage equipped with an ESI-LC system Accela (ThermoFisher Scientific, Courtabœuf,
134 France) (Supplementary Table I). An Acclaim RSLC C_{18} column ($2.2\ \mu\text{m}$; $2.1 \times 150\ \text{mm}$;
135 ThermoFisher Scientific) was used to separate the peptides. A gradient prepared from 0.1%
136 formic acid (FA)/water (v/v) (solvent A) and 0.1% FA/ACN (v/v) (solvent B) constituted the
137 mobile phase. The peptides were eluted using a linear gradient from 0 to 50% of solvent B over
138 45 min, then from 50 to 100% over 10 min, and finally held for 5 min at a $250\ \mu\text{L}\cdot\text{min}^{-1}$ flow
139 rate. The electrospray ionization mass spectrometry detection was done in positive mode with
140 the following optimized parameters: the capillary temperature was set at 300°C , the spray
141 voltage was 4.5 kV, and the sheath gas and auxiliary gas were set at 50 and 10 psi, respectively.
142 The acquisition range was from 100 to 2000 m/z . The peak ion extraction function in Xcalibur
143 software (version 4.0, ThermoFisher Scientific) was used to manually integrate the area value
144 of each peak corresponding to a peptide. This relative area provides a measure of relative
145 abundance, as it indicates the contribution of each peptide to all the peptides identified in the
146 venom.

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

153 was solubilized with 30 μ L of ultrapure water. Then, reduction of disulfide bonds was achieved
154 by mixing these aqueous solutions of crude venoms (*i.e.*, 24 venom reservoirs of *D. armigerum*,
155 25 of *My. ruginodis*, and 10 of *P. californicus*) with 30 μ L of 100 mM ammonium bicarbonate
156 buffer (pH 8) containing 10 mM dithiothreitol (DTT) followed by an incubation for 30 min at
157 56°C. Then, the reduced venoms were alkylated by adding 10 μ L of 50 mM iodoacetamide (IA)
158 for 15 min at room temperature in the dark. These samples were then analyzed through LC-MS
159 according to the protocol described above. Since chemical reduction/alkylation results in a mass
160 increase of 57 Da for each cysteine, the examination of mass shifts in the mass spectra of
161 reduced/alkylated samples allowed us to determine the presence and the number of disulfide
162 bonds in the corresponding peptides.

163 Venoms from *Solenopsis* species are known to be rich in alkaloids. However, the presence
164 of peptides has already been suggested as possible venom components (Dos Santos Pinto *et al.*,
165 2012; Fox *et al.*, 2013). We thus did an hexane extraction on *So. saevissima* crude venom
166 according to the protocol described by Fox *et al.* (2013) to remove most of the alkaloids.
167 Briefly, the hexane extraction was conducted by mixing an aqueous solution of crude venom
168 (*i.e.*, 32 lyophilized venom reservoirs solubilized with 500 μ L of ultrapure water) with 500 μ L
169 of hexane. This operation was repeated twice. The combined extracts were evaporated under a
170 stream of nitrogen. The residue was redissolved in 50 μ L of hexane and 1 μ L was injected in
171 splitless mode. GC-MS analysis were carried out on a Trace GC 1300 coupled with a Triple
172 Quadrupole mass spectrometer (TSQ Duo, ThermoScientific) controlled by *Xcalibur* software.
173 The gas chromatograph was equipped with an TG-5SilMS capillary column (30 m \times 0.25 mm
174 i.d.; 0.25 μ m film thickness). The injector temperature was 250 °C. The oven temperature
175 started from 120 °C during 1 min, increased at 15 °C.min⁻¹ to 280 °C and was maintained for
176 15 min. The mass spectrometric detector was operated in electron impact ionization mode with
177 an ionizing energy of 70 eV, scanning from $m/z = 50$ to 600. The ion source and transfer line
178 temperature were 300 and 280 °C, respectively. The carrier gas was helium (alphagaz 2; Air
179 Liquide, France) at a flow rate of 1.1 mL \cdot min⁻¹. The hexane fraction was then analyzed through
180 GC-MS, whereas the aqueous phase was concentrated prior to LC-MS according to the protocol
181 described above.

182

183 2.3. *De novo* orbitrap mass spectrometry-based sequencing

184 Each crude venom was re-suspended in water and then desalted using ZipTip[®] C₁₈
185 (Merck Millipore, Burlington, VT, USA) after adding trifluoroacetic acid (TFA) at a final

186 concentration of 0.5% (Supplementary Table I). Then, the venom sample was subjected to *de*
187 *nov*o sequencing using a Q-Exactive Plus mass spectrometer coupled to a Nano-LC Proxeon
188 1000 (ThermoFisher Scientific, Waltham, MA, USA). Peptides were separated through
189 chromatography with the following parameters: Acclaim PepMap100 C₁₈ pre-column (2 cm,
190 75 μm i.d., 3 μm, 100 Å), Pepmap-RSLC Proxeon C₁₈ column (50 cm, 75 μm i.d., 2 μm, 100
191 Å), 300 nL min⁻¹ flow rate, a 98 min gradient from 95% solvent A (water, 0.1% FA) to 35%
192 solvent B (99.9% ACN, 0.1% FA) for a total time of 2 h. Peptides were analyzed in the Orbitrap
193 cell, at a resolution of 120,000, with a mass range of *m/z* 350–1550. Fragments were obtained
194 through high collision-induced dissociation (HCD) activation with a collisional energy of 27%.
195 Data were acquired in the Orbitrap cell in a Top20 mode, at a resolution of 17,500. For the
196 identification step, all MS and MS/MS data were processed with an in-house Peaks software
197 (BSI, version 6.0) to perform *de novo* sequencing. The mass tolerance was set to 10 ppm for
198 precursor ions and 0.02 Da for fragments. The following modifications were allowed: oxidation
199 (Met), C-terminal amidation and pyroglutamic acid. *De novo* peptide sequences with Average
200 Local Confidence (ALC) higher than 60% were used for the peptide identifications.

201

202 2.4. *Venom gland transcriptomics*

203 2.4.1. *RNA extraction and sequencing*

204 The RNA extraction of the Neotropical ant species *Daceton armigerum* venom glands were
205 prepared by dissecting ant venom apparatus in ultrapure water. Venom reservoirs plus venom
206 glands were immediately placed into 1 mL of RNAlater. The samples were stored at -80 °C
207 prior to RNA extraction. The RNAlater was removed with a glass Pasteur pipette and the venom
208 gland tissues were disrupted with a TissueLyser II (Qiagen, Germantown, MD, USA) in RLT
209 buffer containing 10% (v/v) of 2-mercaptoethanol (RNeasy Mini Kit, Qiagen). RNA was first
210 isolated with a phenol-chloroform (5:1) solution followed by washing with a solution of
211 chloroform-isoamyl alcohol (25:1) to remove the phenol. The RNA was then bound to a Qiagen
212 column and washed as per the manufacturer's instructions. DNase I (Roche Diagnostics
213 GmbH, Mannheim, Germany) was added to remove DNA fragments. The RNA was eluted in
214 50 μL of RNase free water and a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher
215 Scientific) was used to determine 260/280 and 260/230 nm ratios. Finally, RNastable® LD
216 (Biomatrica, San Diego, CA, USA) was added to the purified RNA and the sample was dried
217 using a Speed Vac (RC1010, Jouan, Saint Herblain, France). *D. armigerum* RNA was dried
218 together with RNastable® (Biomatrica, SigmaAldrich, Burlington, MA, USA) before

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

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

231 RNA quantity and quality were assessed using a Nanodrop and a bioanalyzer (Nanodrop
232 2000, ThermoFisher Scientific; Agilent 2100 Bioanalyzer System). The extracted RNA was
233 sequenced through RNAseq at the GeT-PlaGe core facility (INRAE Toulouse, France). Briefly,
234 RNA-seq libraries were prepared according to Illumina's protocols using the Illumina TruSeq
235 Stranded mRNA sample prep kit to analyze mRNA. mRNA was selected using poly-T beads.
236 Then, the RNA was fragmented to generate double stranded cDNA and adaptors were ligated
237 to be sequenced. Eleven cycles of PCR were applied to amplify the libraries. Library quality
238 was assessed using a Fragment Analyser, and the libraries were quantified through qPCR using
239 the Kapa Library Quantification Kit. RNA-seq libraries were sequenced on an Illumina
240 HiSeq3000 using a paired-end read length of 2×150 pb with the Illumina HiSeq3000
241 sequencing kits.

242

243 2.4.2. Contig quantification

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

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

257

258 2.4.3. Precursor identifications and mature sequences

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

281

282 2.4.4. Annotation of most expressed contigs

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

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

288

289 2.4.5. Evolutionary analysis of *U₁₇* and *U₃/U₃₃* encoding sequences

290 *U₁₇* and *U₃/U₃₃* nucleotide sequences were obtained from our transcriptome assemblies
291 of *T. africanum*, *My. ruginodis*, *P. californicus*, and *So. saevissima* venom glands as well as
292 previously published *T. bicarinatum* and *Ma. rubida* transcriptomes (PRJNA234295
293 and PRJEB34828, respectively) and blast searches on transcriptomes available for *Solenopsis*
294 *invicta* and *Myrmica sulcinodis* on NCBI (GFUX00000000.1 and PRJDB4088, respectively).
295 Regarding their similar biochemical features, *U₃₃* nucleotide sequences were added to the *U₃*
296 nucleotide sequences group for the analysis. Alignments of nucleotide sequences encoding
297 mature (*i.e.*, including the C-terminal amidation signal) or prepro-regions were performed on
298 *MEGAX* with *CLUSTALW* (Codons). To estimate the rate of substitution per synonymous sites
299 (dS) or per non-synonymous sites (dN), we computed a pairwise distance using the modified
300 Nei-Gojobori model on *MEGAX* within each group or between all group pairs. We performed
301 a codon Z-test of purifying or positive selection of overall sequence pairs, or over sequence
302 pairs of each group using the Nei-Gojobori model with the pairwise deletion option on *MEGAX*.
303 We tested sequences encoding only mature regions (*i.e.*, with amidation signal) or sequences
304 encoding only prepro-regions. We then conducted a site-by-site analysis to detect pervasive
305 (*i.e.*, continuous changes) or episodic (*i.e.*, single change) positive or purifying selection on
306 each coding site. For this, we used two methods available on datamonkey web site: Mixed
307 Effects Model of Evolution (MEME, pervasive/episodic) and Fixed Effect Likelihood model
308 (FEL, pervasive).

309

310 2.5. Fmoc solid phase synthesis of *U₁₇-MYRTX-Tb1a* and *U₃-MYRTX-Tb1a*

311 All Fmoc amino acid residues, *O*-benzotriazol-1-yl-*N,N,N',N'*-tetramethyluronium
312 hexafluorophosphate (HBTU) and Rink amide 4-methylbenzhydrylamine (MBHA) resin were
313 purchased from Christof Senn Laboratories (Dielsdorf, Switzerland) or IRIS Biotech
314 (Marktredwitz, Germany). *N,N*-Diisopropylethylamine (DIEA), piperidine, trifluoroacetic acid
315 (TFA), triisopropylsilane (TIS), *tert*-butylmethylether (TBME), MeOH, NaOMe, DMSO and
316 Fmoc-Thr[GalNAc(Ac)₃- α -D]-OH were supplied from Sigma-Aldrich (Saint-Quentin-
317 Fallavier, France). *N*-methylpyrrolidone (NMP), dimethylformamide (DMF), dichloromethane

318 (DCM) and acetonitrile were from Fisher Scientific (Illkirch, France). Peptides U₃-MYRTX-
319 Tb1a (VLPALPLLAGLMSLPFLQHKLTN-NH₂) and [Thr(GalNAc- α -D)¹]U₁₇-MYRTX-
320 Tb1a (XIINAPNRCPPGHVVVKGRICRIA-NH₂, X= Thr(GalNAc- α -D)) were synthesized by
321 Fmoc solid phase methodology on a Liberty microwave assisted automated peptide synthesizer
322 (CEM, Saclay, France) using the standard manufacturer's procedures at 0.1 mmol scale. All
323 Fmoc-amino acids (0.5 mmol, 5 eq.) were coupled (25 W, 75°C, 300 sec) except for histidine
324 and cysteine (0 W, 50°C, 120 sec and 25 W, 50°C, 240 sec) on Rink amide resin, by *in situ*
325 activation with HBTU (0.5 mmol, 5 eq.) and DIEA (1 mmol, 10 eq.) before Fmoc removal with
326 a 20% piperidine in DMF (35 W, 75°C, 30 sec and 35 W, 75°C, 180 sec). Reactive side chains
327 were protected as follow: Asn, Gln, trityl (Trt) amide; His, trityl (Trt) amine; Ser, Thr, *tert*-
328 butyl (*t*Bu) ether; Arg, pentamethyldihydrobenzofuran (Pbf) sulfonylamide; Lys, *tert*-
329 butyloxycarbonyl (Boc) carbamate; Cys, acetamidomethyl (Acm) thioether. After completion
330 of the chain assembly, peptides were deprotected and cleaved from the resin by adding 10 mL
331 of an ice-cold mixture of TFA/TIS/H₂O (9.5:0.25:0.25, v/v/v) and agitating 3 hours at room
332 temperature. Peptides were then obtained by precipitation in TBME followed by centrifugation
333 (4500 rpm, 15 min). The synthetic peptides U₃-MYRTX-Tb1a and [Thr(GalNAc(Ac)₃- α -
334 D)¹]U₁₇-MYRTX-Tb1a were purified by reversed-phase HPLC on a 21.2 x 250 mm Jupiter C₁₈
335 (5 μ m, 300 Å) column (Phenomenex, Le Pecq, France) using a linear gradient (20-70% and 10-
336 60% over 45 min, respectively) of acetonitrile/TFA (99.9:0.1) at a flow rate of 10 mL/min.
337 Deacetylation of the Thr(GalNAc(Ac)₃- α -D) residue was performed with sodium methoxide.
338 The lyophilized glycopeptide was dissolved in methanol (5 mg/mL) and a solution of sodium
339 methoxide (0.3 M) was added dropwise to a final concentration of 30 mM. After complete
340 removal of O-acetyl protection groups in about 6 h, the reaction mixture was frozen after adding
341 water and freeze-dried. The formation of the disulfide bond was then carried out by oxidation
342 of the deprotected peptide in a mixture of H₂O/CH₃COOH/DMSO (75:5:20) (0.5 mg/mL) at
343 pH 5.2. The reaction was stopped after 19 h, freeze-dried and the resulting cyclic glycopeptide
344 was purified in the same conditions as described above. The purified peptides, U₃-MYRTX-
345 Tb1a and [Thr(GalNAc- α -D)¹]U₁₇-MYRTX-Tb1a, were then characterized by MALDI-TOF
346 mass spectrometry on a ultrafleXtreme (Bruker, Strasbourg, France) in the reflector mode using
347 α -cyano-4-hydroxycinnamic acid as a matrix. Analytical RP-HPLC, implemented on a 4.6 x
348 250 mm Jupiter C₁₈ (5 μ m, 300 Å) column, indicated that the purity of the peptides was >99.9%.
349 Net peptide content was accurately determined on a Unicube elemental analyser (Elementar,
350 Lyon, France) using manufacturer's protocols.

351

352 *2.6. NMR Experiments*

353 As no tridimensional structure of secapin-like peptide and Nav toxin from myrmicine ant
354 venoms has been published so far, we completed our study by doing a NMR analysis of the
355 peptides U₁₇-MYRTX-Tb1a (glycosylated, U₁₇-Tb1a) and U₃-MYRTX-Tb1a (U₃-Tb1a) from
356 *T. bicarinatum* venom, the first representatives of these kinds of peptides in myrmicine ants
357 venoms (Touchard et al., 2018). Prior to NMR analysis each synthesized peptide was dissolved
358 in the appropriate solvent. U₁₇-Tb1a was dissolved in H₂O:D₂O (9:1 v/v) at a concentration of
359 0.94 mM with pH adjusted to 5.2, and U₃-Tb1a was dissolved in 2,2,2-trifluoroethanol-d₂
360 (TFE) 50% pH 4.6 at a concentration of 0.8 mM. Then, 2D ¹H-NOESY (200 and 150 ms,
361 respectively), 2D ¹H-TOCSY (70 and 80 ms, respectively), a sofast-HMQC (Schanda et al.,
362 2005) (¹⁵N natural abundance) and a ¹³C-HSQC (¹³C natural abundance) spectra were
363 performed at 298K on an Avance III HD BRUKER 700 MHz spectrometer equipped with a
364 cryoprobe. ¹H chemical shifts were referenced to the water signal (4.77 ppm at 298K). NMR
365 data were processed using the Topspin software version 3.6.2TM (Bruker, Billerica, MA, USA)
366 and analyzed with CCPNMR version 2.2.2 (Vranken et al., 2005).

367 Structures were calculated using the Crystallography and NMR System (CNS) (Brünger,
368 2007; Brünger et al., 1998) through the automatic assignment software ARIA2 version 2.3
369 (Rieping et al., 2007) with NOE derived distances, hydrogen bonds (in accordance with the
370 observation of typical distance NOE cross peaks network for β-sheets in the case of U₁₇-Tb1a
371 and for α-helix in the case of U₃-Tb1a), backbone dihedral angle restraints determined with the
372 DANGLE program (Cheung et al., 2010) and the imposed disulfide bridge between Cys⁹ and
373 Cys²⁰ only present in U₁₇-Tb1a. Strong Hα(X)–Hδ(Pro) NOEs were diagnostic of trans
374 conformations about each X-Pro bond. The ARIA2 protocol, with default parameters, used
375 simulated annealing with torsion angle and Cartesian space dynamics. The iterative process was
376 repeated until the assignment of the NOE cross peaks was completed. The last run for each
377 peptide was done with 1,000 initial structures and 250 structures were refined in water. Fifteen
378 structures were selected on the basis of total energies and restraint violation statistics, to
379 represent the tridimensional structure of U₁₇-Tb1a and U₃-Tb1a in solution. The quality of final
380 structures was evaluated using PROCHECK-NMR (Laskowski et al., 1996) and PROMOTIF
381 (Hutchinson and Thornton, 1996).

382 The structure calculations were done without taking into account the C-terminal amidation
383 especially as no NOE could be observed between the amido C-terminal protons (A²³-NH₂ for

384 U₁₇-Tb1a and Asn²³-NH₂ for U₃-Tb1a) and all other protons of each peptide. However, the
385 Coulombic electrostatic potential (ESP) and Molecular lipophilicity potential (MLP) were
386 determined at the Connolly surface of peptide after C-terminal amidation using ChimeraX 1.3
387 (Pettersen et al., 2021). The figures were prepared with PYMOL (DeLano, 2002) and
388 ChimeraX.

389

390 2.7. Toxin nomenclature and data accessibility

391 Novel myrmecitoxin sequences were named following the standard nomenclature for
392 animal venom peptides (King et al., 2008), and we used myrmecitoxin (MYRTX) to define the
393 venom peptides from the subfamily Myrmicinae and genus/species descriptors as follow;
394 *Myrmica ruginodis* (Mru), *Pogonomyrmex californicus* (Pc), *Stenamma debile* (Sd), *Daceton*
395 *armigerum* (Da), and *Tetramorium africanum* (Ta) (Touchard et al., 2016a). In cases of
396 sequence similarities, we also named the novel myrmecitoxins according to the same subscripts
397 used to denote the ‘unknown’ activity descriptor prefixes previously defined for myrmecine ant
398 venom peptides (U₁-U₂₀), and forming mature myrmecitoxin families (Touchard et al., 2020a,
399 2018).

400 The transcriptomes raw data are available through the European Nucleotide Archive
401 (ENA) under the project number PRJEB51841. The mass spectrometry data were deposited in
402 the ProteomeXchange Consortium *via* the PRIDE (Perez-Riverol et al., 2022) partner repository
403 with the dataset identifier PXD033547. NMR data and tridimensional structure coordinates
404 have been deposited with SMSDep as Biological Magnetic Resonance Bank (BMRB) accession
405 codes: 21098 (U₁₇-MYRTX-Tb1a) and 21099 (U₃-MYRTX-Tb1a).

406

407 3. Results

408 3.1. Proteotranscriptomics data

409 3.1.1. Proteomic data

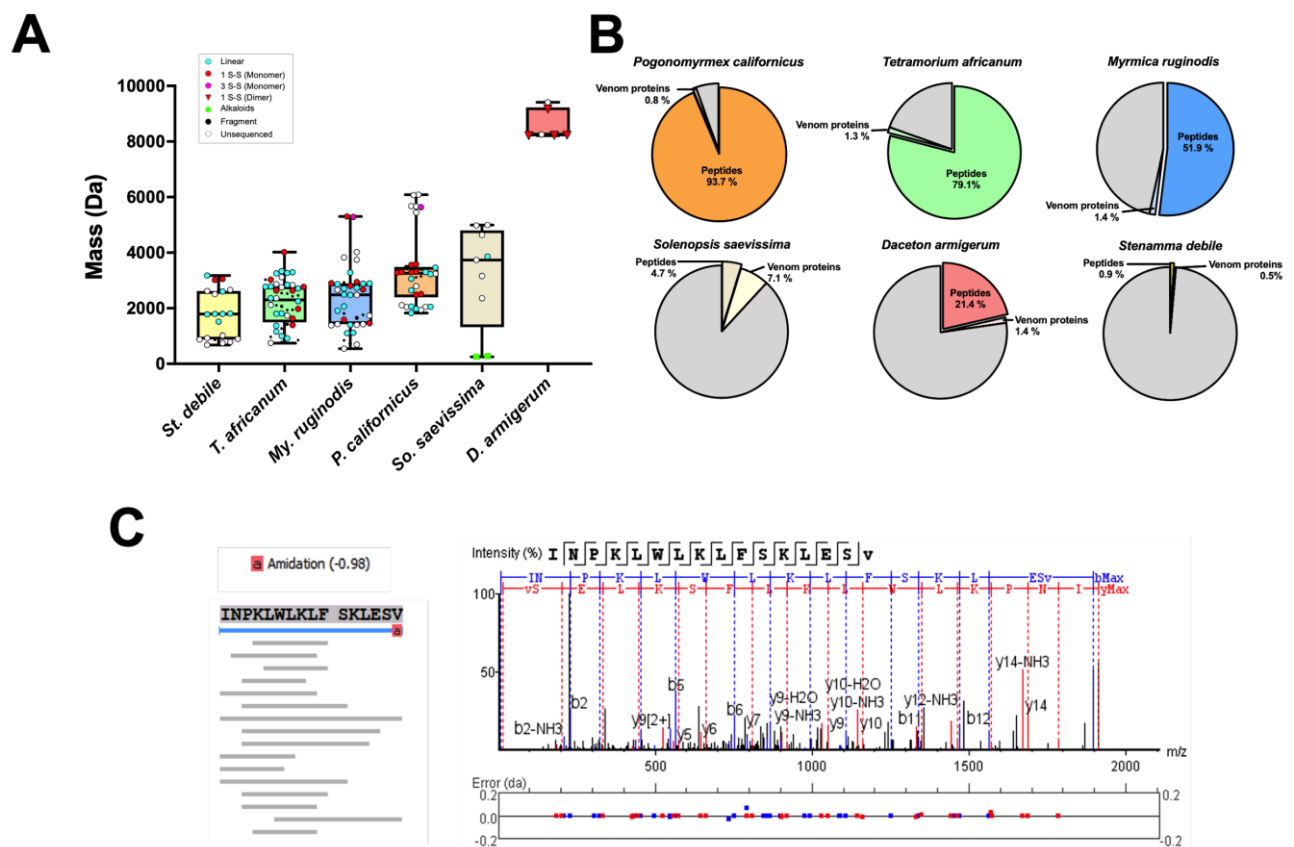
410 LC-MS analysis revealed heterogeneous profiles, with great variation in the peptide
411 molecular weight and retention time (*i.e.*, hydrophobicity) among species, overall ranging from
412 682 to 9,416 Da and from 1.84 to 51.62 min (Figures 1 and 2 – A). A total of 155 peptides were
413 detected with great variations among species: from only 6 peptides detected in the venom of
414 *Daceton armigerum* to up to 49 peptides in *Tetramorium africanum* venom (Figure 2 – A;
415 Supplementary tables III-XIII). Venom peptides from *T. africanum* were distributed relatively
416 widely between 1.84 and 46.24 min of elution, ranging from 744 and 3,338 Da (Supplementary

417 Table III – Figure 2 A). This profile is consistent with previous observations of venom from
418 the genus *Tetramorium* (Rifflet et al., 2012; Touchard et al., 2018). Overall, the venoms of the
419 species studied here contained mostly small peptides (*i.e.*, molecular weight < 5,000 Da) except
420 for *D. armigerum* venom in which all peptides were detected in the narrow mass range 8,208-
421 9,416 Da.

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

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

441 Finally, a LC-MS/MS analysis of the six crude venoms was done using a Q-exactive
 442 orbitrap to achieve *de novo* sequencing and so acquire predictive peptide sequence tags. This
 443 resulted in a total of 6,314 sequence tags with an ALC score higher or equal to 60%, ranging
 444 from 145 to 2,898 sequence tags for *St. debile* and *T. africanum* venoms, respectively. These
 445 sequence tags were then blasted on the transcriptomic data to determine the complete sequences
 446 of venom peptides and to identify precursors.
 447
 448



449
 450 **Figure 2:** Proteotranscriptomic data of the venoms of *Stenamma debile*, *Tetramorium africanum*, *Myrmica*
 451 *ruginodis*, *Pogonomyrmex californicus*, *Solenopsis saevissima*, and *Daceton armigerum*. **(A)** Repartition of
 452 masses detected by LC-MS analysis in the six myrmicine ant venoms. **(B)** Proportion of addressed contigs
 453 expressing venom peptides and venom proteins in the six myrmicine ant venom gland transcriptomes. **(C)**
 454 An example of MS/MS spectrum for U₁₄-Mru1a leading to amino acid sequence confirmation by Peaks
 455 software.

456 3.1.2. Transcriptomic data

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

460 all venoms, with a mean contig size of 1,880 and 2,174 bp, respectively. Both assembly
461 methods gave various numbers of contigs for each venom gland transcriptome, ranging from
462 13,701 for *D. armigerum* (Oases) to 48,353 for *St. debile* (Trinity) (Supplementary table XV).
463 For each venom, a combined database from both assemblies was used to search for peptide
464 sequences generated from LC-MS/MS analyses. Additionally, the signal sequences of the
465 previously published myrmecitoxins (*i.e.*, from *Ma. rubida* and *T. bicarinatum*) were searched
466 against these databases to find peptides sharing the same signal sequences. This permitted us to
467 infer that most of the venom peptide encoding genes had high transcription levels as they
468 accounted from 21% to 94% of the addressed contigs by the venom glands of *D. armigerum*
469 and *P. californicus*, respectively (Figure 2 – B; Supplementary Information).

470 Apart from the other myrmecine ant transcriptomes presented here, only 7% of the
471 addressed contigs from *So. saevissima* venom glands encoded putative venom peptides (Figure
472 2 – B, Supplementary Information), which is consistent with the low relative abundances of
473 venom peptides noted in the proteomic data. Also, the transcripts encoding venom peptides of
474 *St. debile* represented 1% of the addressed contigs (Figure 2 – B, Supplementary Information).
475 As these transcriptomic data were not issued only from venom glands, our strategy does not
476 allow us to evaluate the expression level in these organs but rather permits us to identify putative
477 venom peptide precursors. However, the combination of proteomic and transcriptomic data
478 allowed us to assign only one of the encoding sequences to one of the measured masses.

479

480 3.1.3. Global description of myrmecine ant venom peptidomes and molecular features 481 of mature myrmecitoxins

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

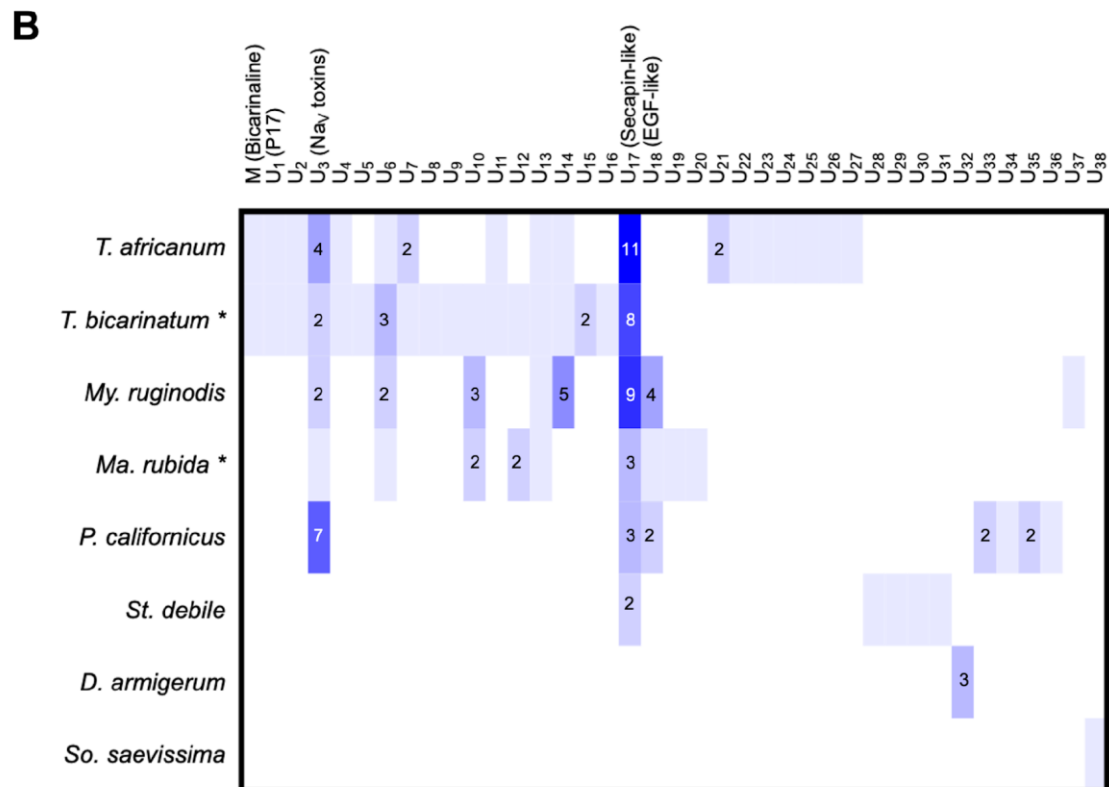
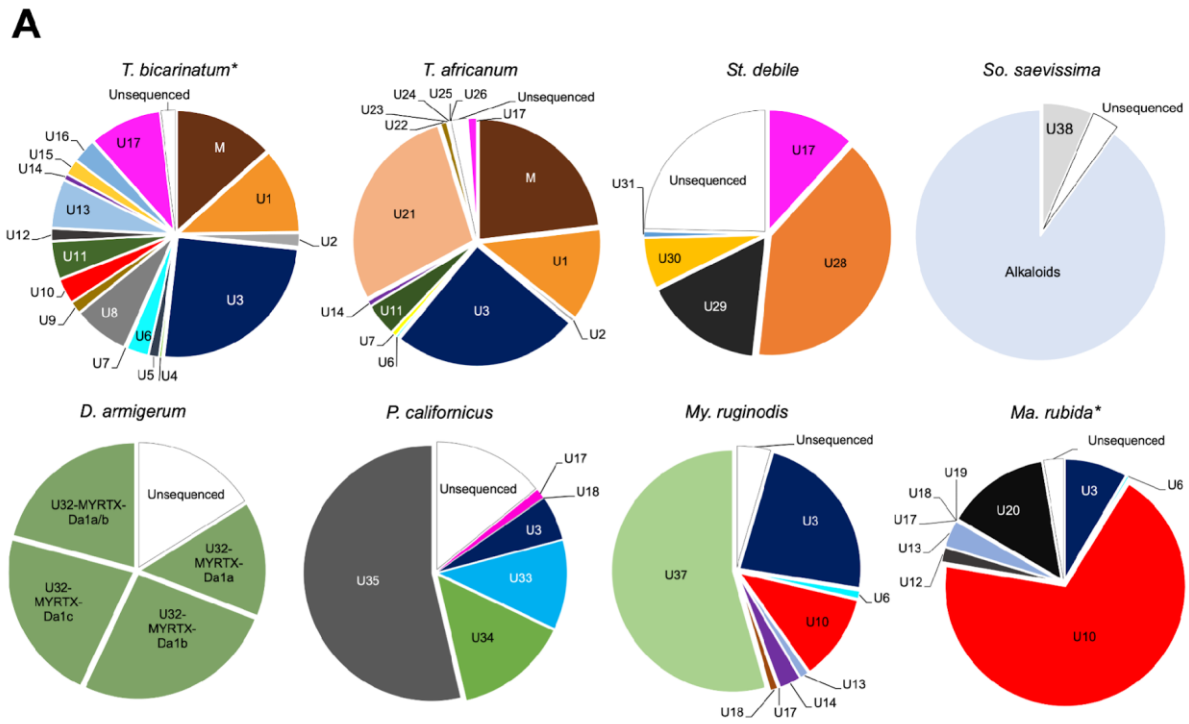
489 Most of the ant venoms studied here had a high content of linear polycationic peptides,
490 containing several arginine and lysine residues (*i.e.*, families M, U₁, U₃, U₁₀, U₁₃, U₁₄, U₂₄, U₂₅,
491 U₂₆, U₂₈, U₃₀, U₃₁, U₃₃ and U₃₄) (Figure 3 – A – Supplementary Figure 1). Those short
492 myrmecitoxins (*i.e.*, 10-31 amino acids long) constituted 62% of *T. africanum* peptidome,

493 whereas they composed 49% of *St. debile* venom (Figure 3 – A). However, the venom
494 peptidomes of *P. californicus* and *My. ruginodis* were both dominated by two linear venom
495 peptides uniquely found in these two species and exhibiting a negative net charge, due to the
496 high content of aspartic and glutamic acids (*i.e.*, U₃₅ and U₃₇, respectively). These two peptides
497 represented 54% of their venom peptidome compositions. Besides, three families gathering
498 anionic linear venom peptides were found in *T. africanum*, *St. debile* and *So. saevissima*
499 venoms, representing 29% (*i.e.*, U₂₁), 17% (*i.e.*, U₂₉) and 7% (*i.e.*, U₃₈) of their compositions,
500 respectively (Figure 3 – A, Supplementary Figure 1).

501 Apart from linear peptides, disulfide-bonded venom peptides were also found within
502 myrmicine ant venom peptidomes. From 13 to 34 amino acids long, ten of them contained a
503 single intrachain disulfide bond and were classified in nine families of mature toxins (*i.e.*, U₄,
504 U₆, U₇, U₁₁, U₁₇, U₁₈, U₂₂, U₂₃ and U₂₇) (Figure 3 – A). Most of them were also polycationic
505 due to the presence of several lysine, arginine, and histidine residues (Supplementary Figure
506 2). They were, however, less abundant than their linear counterparts as their abundances ranged
507 from 2% to 12% in *P. californicus* and *St. debile*, respectively (Figure 3 – A). Furthermore, six
508 myrmicotoxins containing three disulfide bonds were identified in *P. californicus* and *My.*
509 *ruginodis* venoms, accounting for less than 1% of each venom peptidome (*i.e.*, U₁₈). They were
510 also predicted by ScanProsite to have an Epidermal Growth Factor-like domain (EGF), which
511 is consistent with the EGF-like peptides recently reported from several ant venoms (Hurka et
512 al., 2022; Robinson et al., 2018; Touchard et al., 2020). Also, the toxin U₁₈-MYRTX-Mru1b
513 exhibited a missing 146 Da mass unit in its calculated mass, corresponding to an additional
514 fucose glycan (Supplementary Tables XI and XII) in accordance with the presence of the
515 consensus O-glycosylation site observed in its primary sequence -C₂XXXXS/TC₃-
516 (Supplementary Figure 2).

517 Apart from other myrmicine ant peptidomes, the venom of *D. armigerum* was found to
518 be exclusively composed of dimeric venom peptides. The transcriptomic data of *D. armigerum*
519 venom glands led to the identification of three transcripts with calculated masses ranging from
520 4,104 to 4,581 Da and exhibiting a single cysteine. Together with the results from the LC-MS
521 analysis of the reduced/alkylated venom, it enabled the identification of three monomers (-
522 Da1a, -Da1b and -Da1c). Four theoretical masses for dimeric peptides matched with masses
523 measured in the total ion chromatogram of the crude venom (*i.e.*, 8,231 Da, 9,161 Da, 8,209 Da
524 and 8,220 Da), corresponding to the homodimer-Da1b, Da1c, Da1a and the heterodimer-
525 Da1a/b, respectively) (Figure 1). These amphiphilic and polycationic dimers accounted for 84%

526 of the venom peptidome (Figure 3 – A). The chains composing them were very similar with a
527 mean of 78% identity (Supplementary Figure 2). As they displayed no similarity with
528 previously reported ant venom peptides, they were therefore gathered in a single family (*i.e.*,
529 U₃₂).

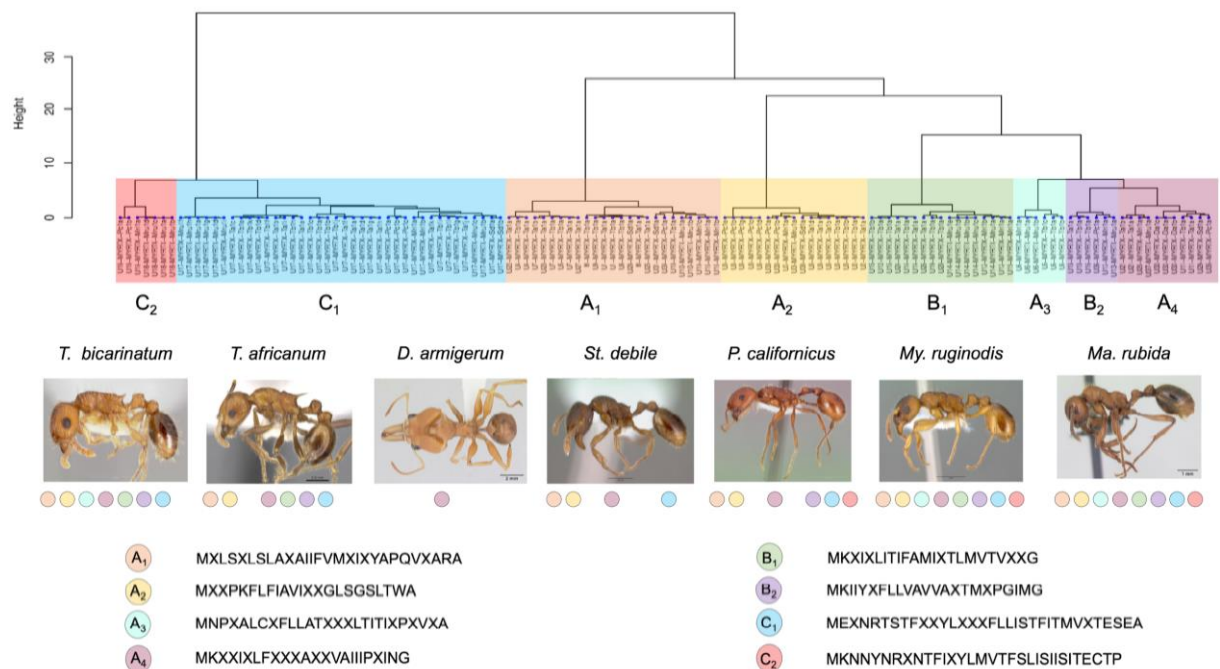


530
 531 **Figure 3:** Global repartition of mature venom peptide families within myrmicine ant venom peptidomes.
 532 “*” denotes previously described venom peptidomes, which were characterized with a similar methodology
 533 from (Touchard et al., 2018). (A) Relative abundances of mature myrmecitoxin families. (B) Number of
 534 peptide variants found by mature myrmecitoxin family.

535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549

3.2. Superfamilies of myrmecitoxins 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.*, A₁, A₂, B₁, B₂ and C) (Touchard et al., 2020a, 2018), and incremented this nomenclature to the new clusters (*i.e.*, A₃, A₄, C₁ and C₂) (Figure 4). However, this nomenclature is tentative and needs to be refined by using genomic data.



550
 551
 552
 553
 554
 555
 556
 557

Figure 4: Clustering analysis of myrmecitoxin signal sequences. We included all described venom peptide precursors from myrmecine 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 myrmecitoxin precursors found. Every ant photography is extracted from AntWeb

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

563 In Touchard *et al.* (2018), precursors from the superfamilies A and B have been defined as
564 “pilosulin-like” peptides and a new superfamily of ant venom precursors, respectively.
565 Precursors from these superfamilies are related to aculeatoxins and exhibit an organization in
566 three regions: an N-terminal signal sequence, an alanine/glutamic acid rich prosequence with
567 variable length and the mature region (Touchard *et al.*, 2018). The prosequences of precursors
568 from superfamily B are shorter than those from superfamily A, but the cleavage maturation site
569 releasing the mature toxin is located after an alanine or a proline residue for both, as previously
570 described for other ant venom precursors (Touchard *et al.*, 2020a, 2018). While mature regions
571 of A₁, A₄, B₁ and B₂ clusters are highly diverse in their sequences and biochemical properties,
572 the mature sequences which are gathered into A₂ and A₃ clusters belonged mainly to the same
573 peptide families, or shared the same physico-chemical properties (*i.e.*, U₃/U₃₃ and U₆ peptides
574 respectively) (Figure 4). The superfamily C was previously defined in *T. bicarinatum* and *Ma.*
575 *rubida* venom peptidomes as gathering secapin-like (U₁₇) and EGF-like peptides (U₁₈)
576 (Touchard *et al.*, 2020a, 2018). Members of the superfamily C differed from precursors of the
577 superfamilies A and B in their organization and maturation profiles. Indeed, they are predicted
578 as Type II single pass transmembrane proteins with a short cytosolic N-terminal domain of five
579 amino acids and an internal signal sequence directly followed by the mature sequence (Eagles
580 *et al.*, 2022; Touchard *et al.*, 2018). Here, the clustering analysis split it into two separate
581 clusters gathering U₁₇ and U₁₈ precursors (*i.e.*, C₁ and C₂, respectively) (Figure 4). The cleavage
582 maturation sites of precursors from the C₁ family were more often located after a serine residue
583 than an alanine or proline residue, whereas the cleavage sites of precursors from the C₂ family
584 were located after a proline but also after a threonine or cysteine residue (Figure 4).

585

586 3.3. NMR structures of the secapin-like peptide U₁₇-MYRTX-Tb1a and the Nav toxin U₃- 587 MYRTX-Tb1a

588 Despite being found in the venom of numerous hymenoptera (Lee *et al.*, 2016), no secapin
589 and Nav toxin (U₃) structure has been characterized yet. The 3D structure of U₁₇-Tb1a
590 (glycosylated) and U₃-Tb1a previously characterized in the venom of the myrmicine ant *T.*
591 *bicarinatum* (Touchard *et al.*, 2018), was thus determined here using NMR spectroscopy.

592 The main U₁₇-Tb1a structure feature is a double-stranded antiparallel β -sheet connected by
593 a β -turn, otherwise known as a β -hairpin motif (Figure 5 – B, Supplementary Figure 6). The
594 single disulfide bridge between Cys⁹ and Cys²⁰ forming a 12-membered macrocycle was likely
595 to maintain the high degree of structure. The N-terminal 6 residues part of the peptide (1-6)
596 appeared as an unstructured and highly flexible tail. Residues Pro¹⁰, Pro¹¹, Gly¹² and His¹³
597 formed a type II β -turn that led to the first β -strand (His¹³, Val¹⁴, Val¹⁵, Val¹⁶) and, Arg¹⁹, Cys²⁰,
598 Arg²¹ and Ile²² made up the other strand. These two β -strands were connected by a β -turn around
599 Gly¹⁸ defined with PROMOTIF as type IV (VVKG) and type I' (VKGR). Some NOEs were
600 observed between the side chains of residues Asn⁷ and Val¹⁵-Val¹⁶ (β 1-strand) and Arg¹⁹-Cys²⁰
601 (β 2-strand), but due to an overlapping of their HN and H α protons, the NOEs characteristics of
602 an antiparallel β -sheet were possible but unobservable. In this case, a third β -strand (Asn⁷-Arg⁸)
603 could be formed and the structure of U₁₇-Tb1a should be a triple-stranded antiparallel β -sheet.
604 AlphaFold predicts this tendency for the complete sequence of U₁₇-MYRTX-Tb1a (57 AA
605 including the signal sequence) as it shown in UniProtKB for the representation of Entry
606 A0A6M6RE84 (SECP_TETBN). U₁₇-Tb1a was a highly charged molecule (Figure 5 – B,
607 Supplementary Figure 6), counting 3 arginine and 1 lysine residues whose side chains were
608 exposed to the solvent and spread on the same side of the molecular surface.

609 The U₃-Tb1a adopted a helical structure from Leu5 to Thr22 (Figure 6 – B; Supplementary
610 Figure 7) defined with PROMOTIF as two α -helices (5-8 and 10-22) with a ω angle of (-45.5
611 ± 2.7) ° for 53% structures or a bent α -helix (5-22) with a deviation of (34.4 ± 2) ° for 47%
612 structures. U₃-Tb1a was a highly hydrophobic peptide (Supplementary Figure 7 – D) as it
613 contained 8 leucines, 3 prolines, 2 alanines, 2 glycines and 1 valine. Only the C-terminal part
614 encompassed polar and charged residues (Gln¹⁸ and Lys²⁰). This would result in different
615 interactions of U₃-Tb1a with cell membranes. It is conceivable that the N-terminal apolar helix
616 favors transmembrane localization while the C-terminal amphiphilic helix is either solvent
617 exposed or interacting with the membrane surface.

618

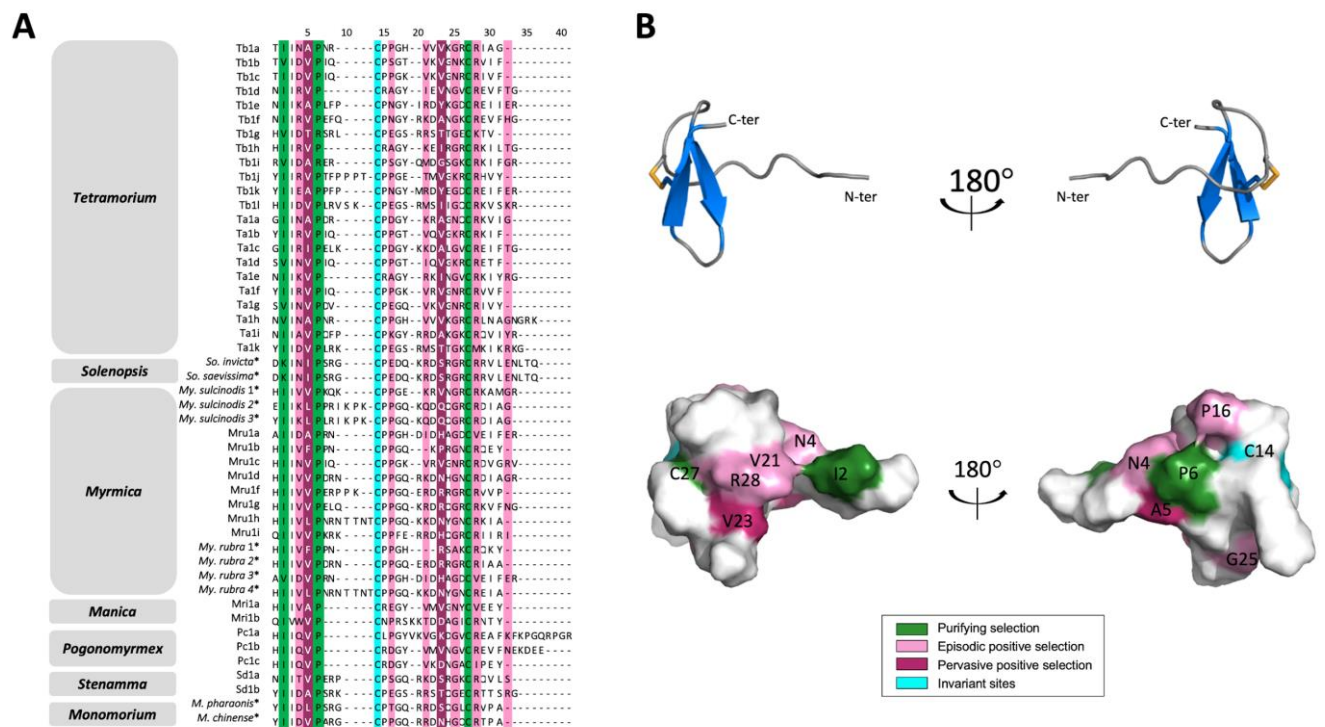
619 3.4. Identified selection on families U₃/U₃₃ and U₁₇

620 The families U₃ (Nav toxins) and U₁₇ (secapin peptides) are mainly ubiquitous among
621 myrmicine ant venoms with several variants among species. The functions described for
622 secapins from *Apis* spp. venoms suggest their defensive roles as they induce pain and
623 inflammation in mammals (Lee et al., 2016; Mourelle et al., 2014), and U₃ peptides were
624 recently characterized as defensive toxins acting on mammalian Nav channels (Robinson *et al.*

625 2022, in review for *Nat. Com.*). The sequence diversity highlighted here provides an ideal
626 playground for analyzing the evolutionary mechanisms leading to these sequence variations.
627 Regarding their similar biochemical features, U₃₃ nucleotide sequences were added to the U₃
628 nucleotide sequences group for the analysis. For this, pairwise comparisons of nucleotide
629 sequences of the mature peptides and the prepro regions of U₃/U₃₃ and U₁₇ precursors were then
630 aligned separately on the software MEGAX (Supplementary Tables XIX to XXII).

631 Among secapin peptides (U₁₇ family), a higher global dN/dS rate was noted within the
632 *Manica* group in comparison to the other groups, except *Pogonomyrmex*. More generally, high
633 global intragroup dN/dS rates were observed for *Manica*, *Stenamma*, and *Myrmica*, suggesting
634 the action of a prospective positive selection on some sequences (Supplementary Table XIX).
635 Consequently, a codon based-Z-test of positive and purifying selection was done, highlighting
636 a significant positive selection within the group *Manica*, as well as in the other groups, except
637 *Pogonomyrmex*. The same significant tendency was also noted for some sequence pairs of
638 *Tetramorium* vs. *Myrmica*; *Tetramorium* vs. *Manica*; and *Myrmica* vs. *Manica* (Supplementary
639 Table XXI). Although *Manica* and *Myrmica* groups contained only two sequences, these tests
640 showed that sequence pairs encoding secapins were under positive selection (i.e., 50% of the
641 pairs for *Myrmica* and 100% of them for *Manica*). A positive selection was also noted for 20%
642 of the sequence pairs within the *Tetramorium* group (Supplementary Table XXI). MEME and
643 FEL algorithms, used for the site-by-site analysis, predicted positions 5 and 23 as being under
644 persistent positive selection. For position 23, notable modifications concerning the net charge
645 according to whole sequences might have consequences on mature toxin structures and
646 functions. Furthermore, the FEL algorithm predicted that the positions 2 and 6, as well as the
647 second cysteine (i.e., position 27) are under purifying selection. Besides, the positions 4, 16,
648 21, 23, 25, 28, and 32 were noted to be under episodic positive selection by the algorithm
649 MEME. The first cysteine is thus the only invariant site when the positions presenting a
650 majority of GAP were removed (Figure 5 – A).

651



652

653

654

655

656

657

658

659

660

661

Figure 5: Evolutionary study of scapin-like peptides from myrmicine ant venoms and structural analysis of U₁₇-MYRTX-Tb1a. (A) Alignments of mature amino acid sequences of the peptides from U₁₇ 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 U₁₇-MYRTX-Tb1a. The disulfide bond is colored in yellow. The numbering of amino acids was defined according to the alignment of all U₁₇ peptides.

662

663

664

665

666

667

668

669

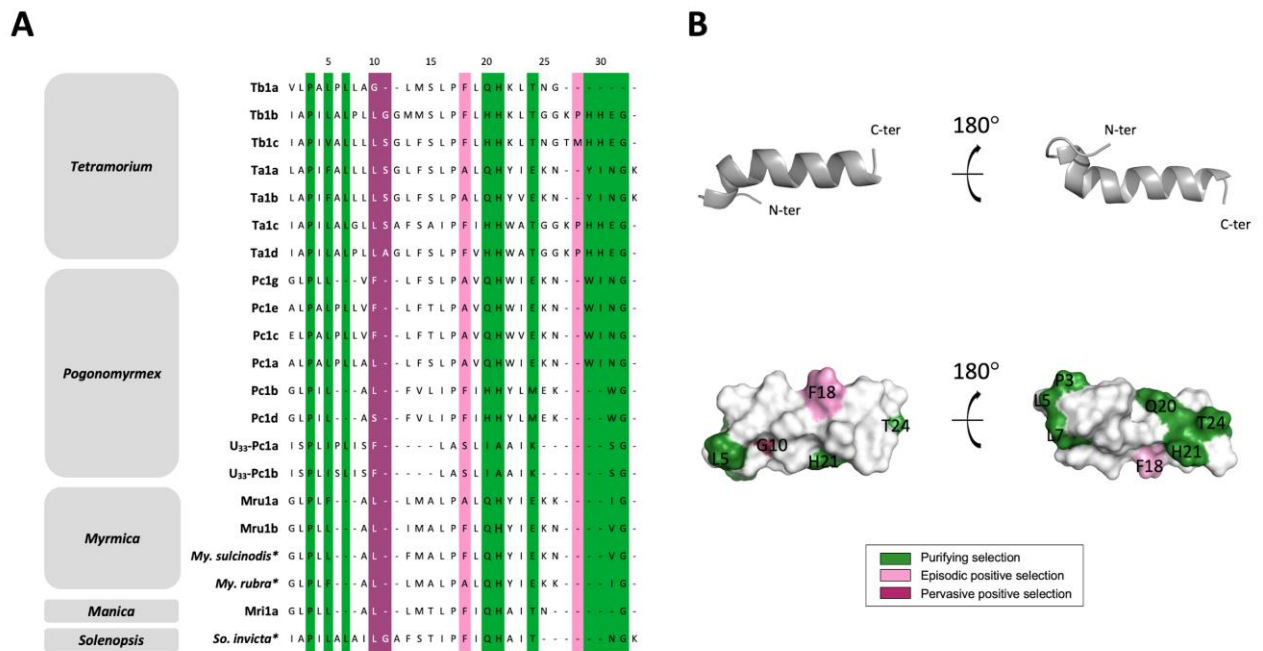
670

671

672

Conversely, U₃ 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

673 *Tetramorium* prepro sequences were under purifying selection whereas *Manica*'s was under
 674 positive selection, as well as some prepro sequences from *Myrmica*. However, this concerned
 675 exclusively the prepro regions of the U₃/U₃₃ pairs (Supplementary Table XXII).



676
 677 **Figure 6:** Evolutionary study of Nav toxins from myrmicine ant venoms and structural analysis of U₃-
 678 MYRTX-Tb1a. (A) Alignments of mature amino acid sequences of the peptides from the U₃ family and
 679 results of the site-by-site analysis. To estimate the evolutionary forces and the positions they were applied
 680 on, the alignments and analysis were performed under the software *MEGA X* and with the algorithms MEME
 681 and FEL, respectively (Kumar et al., 2018). The sequences followed by "*" were retrieved from
 682 transcriptomes which were available online on the ant genomics database - Fourmidable
 683 (<https://antgenomes.org>). (B) Primary structure, cartoon representation of the lowest-energy solution
 684 structure and molecular surface of U₃-MYRTX-Tb1a. The numbering of amino acids was defined according
 685 to the alignment of all U₃ peptides.

686
 687 **4. Discussion**

688 *4.1. Heterogenous compositions of myrmicine ant venoms*

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

695 belonging to other subfamilies previously studied. With the exception of EGF-like peptides
696 which are also present in *Myrmecia gulosa* venom (Eagles et al., 2022; Robinson et al., 2018),
697 we did not find any common peptide families with ant toxins from other investigated
698 subfamilies such as Ponerinae and Myrmeciinae (Kazuma et al., 2017; Orivel et al., 2001;
699 Radis-Baptista et al., 2020; Robinson et al., 2018; Touchard et al., 2016b). This study also
700 allowed us to appreciate the structural diversity of the peptides. Along with linear peptides,
701 several peptides with a disulfide bond were characterized, some of them even dimeric. In
702 particular, we determined the β -hairpin structure of a secapin-like peptide, a toxin family which
703 is also present in bee venoms but whose structure had never been elucidated before our study.
704 We therefore provide a glimpse of the molecular and structural diversity of the peptides
705 contained in the venoms of the large Myrmicine subfamily of ants. The venoms of the 5,000
706 stinging myrmicine ant species estimated remain underexplored for the search of new bioactive
707 peptides. Some of these peptides could be promising for the discovery of new ligands to human
708 receptors as exemplified by the peptide P17 (U₁-Tb1a) (Duraismy et al., 2021).

709 The defensive function seems to be a major player in the evolution of the myrmicine ant
710 venom peptidome compositions, which are associated with an aggressive collective defensive
711 behavior. Our results, along with previous reported investigations, emphasize that myrmicine
712 venoms contain numerous and diverse algescic peptides. A recent comprehensive
713 pharmacological characterization of U₃ peptides from *T. africanum* and *Ma. rubida* venoms
714 demonstrated that these peptides are vertebrate-selective pain-causing Nav_v toxins (Robinson et
715 al. 2022, in review for *Nat. Com.*). U₃ peptides are predominant in the majority of myrmicine
716 ant venoms studied so far (Touchard et al., 2020a, 2018), including here, except for *D.*
717 *armigerum* and *St. debile*. Nav toxins (U₃) likely contribute to the painful sting of ants from the
718 genera *Myrmica*, *Manica*, *Pogonomyrmex*, and *Tetramorium* which are defined as aggressive
719 species while defending their nests against predators and intruders. The non-aggressive species
720 *St. debile* and *D. armigerum* have evolved distinctive strategies to defend their colonies (*e.g.*,
721 feign death and powerful bite, respectively). Their venoms are exclusively used for prey
722 capture, which would have relaxed the selective pressures toward these Nav toxins (U₃) since
723 exhibiting very weak lethal and paralytic effects against insects. In the same line, U₁₇ peptides
724 constitute another highly represented family in the analyzed venoms here. These toxins were
725 previously shown to share significant sequence similarity with secapins (Touchard et al., 2018),
726 which are toxins secreted in the venom of numerous hymenoptera, and for some of which
727 inducing pain and inflammation in mammals (Lee et al., 2016; Mourelle et al., 2014). Further

728 pharmacological experiments are necessary to clearly understand the function and the
729 mechanism of secapin-like peptides from ant venoms. The NMR data deposited here will help
730 to this purpose. In addition, the toxin arsenal of *P. californicus*, *My. ruginodis*, and *Ma. rubida*
731 comprises EGF-like peptides (U₁₈) (Touchard et al., 2020a) with 6 variants described here. If
732 the biological activity of myrmicine ants EGF-like toxins remains to be experimentally
733 demonstrated, a similar peptide from *Myrmecia gulosa* venom (Myrmeciinae) was recently
734 characterized as an agonist of the mammalian receptor ErbB1, involved in pain signaling
735 (Eagles et al., 2022). A similarity of U₁₈ peptides from *Ma. rubida*, *My. ruginodis* and *My.*
736 *rubra* venoms with the insect hormone Spitz were also noted, suggesting it might target the
737 insect ErbB receptor of predators/prey (Eagles et al., 2022; Hurka et al., 2022). Besides, the
738 venoms of both *Tetramorium* species contain U₁ peptides, for one of which is involved in the
739 mammalian nocifensive responses through immunomodulatory effects on vertebrates. Peptide
740 U₁-Tb1a from *T. bicarinatum* is indeed a pro-inflammatory peptide agonist of the MRGPRX2
741 receptor (Mas-related G protein-coupled receptor X2) (Duraisamy et al., 2021). The venoms
742 described here also comprise several families of polycationic peptides with amphipathic
743 structure which are presumably membrane-acting toxins and may contribute to the pain of these
744 ant stings.

745

746 4.2. Molecular diversification of myrmicine ant venom peptides

747 Conserved signal regions of venom peptide precursors, coupled with divergent sequences
748 of mature toxins, have already been observed in ant venoms with peptide precursors that were
749 assigned to the family “pilosulin-like” (Barassé et al., 2019; Ceolin Mariano et al., 2019;
750 Kazuma et al., 2017; Robinson et al., 2018; Touchard et al., 2018). It was recently suggested
751 that aculeate Hymenoptera venom peptides having similar signal sequences might share a gene
752 family and have evolved through gene duplication and neofunctionalization (Robinson et al.,
753 2018; Touchard et al., 2018). Across myrmicine ant venoms, the 8 clusters of precursors defined
754 here led to the maturation of 38 families of mature peptides. Even if further genomic studies
755 are still needed, our results fit with these evolutive mechanisms. Among the described peptide
756 families, secapin-like peptides and Nav toxins were represented in almost all myrmicine ant
757 venom peptidomes therefore including multiple variants (*i.e.*, U₃ and U₁₇). Considering the
758 species in which U₃ (*i.e.*, Nav toxins), and U₁₇ (*i.e.*, secapin-like) toxins were retrieved, their
759 venom uses and the biological activities reported, it seems thus likely that these toxins might
760 have been retained and diversified predominantly as defensive weapons against vertebrates.

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

763 We underlined that positive selection affected secapin sequences of myrmicine species. The
764 codon located in the 23rd position of the U₁₇ alignment is significantly under pervasive positive
765 selection. This position corresponds to Val₁₆ of U₁₇-MYRTX-Tb1a embedded in the first β -
766 strand or in the β -turn connecting the two β -strands. The 23rd position is of particular interest
767 as the amino acid residue variations observed here in charge and polarity could affect the
768 conformation and the function of the venom peptide, and confer functional adaptations to new
769 targets. Position 23 is moreover close to other positions on positive or purifying selection on
770 the 3D representation of U₁₇-Tb1a., constituting a patch on the surface of one side of the
771 peptide. The same phenomenon is observed on the opposite side. These two areas could thus
772 be crucial for the interaction of U₁₇-Tb1a and its molecular target. Variation or conservation
773 patterns in these areas might affect affinity of the peptide with its receptor and confer functional
774 adaptation to new biological targets.

775 Concerning the Nav_v toxins of myrmicine ant venoms, the weak positive selection and the
776 multiple codons under purifying selection suggest a structural and/or functional conservation
777 of U₃ peptides. The positions 10 and 11 are under pervasive positive selection. For U₃-Tb1a,
778 the position 10 corresponds to a glycine which is localized within the helix tilt. Positions under
779 purifying selection are located along the helix, suggesting a global conservation of the
780 helicoidal structure of the U₃ family. The weak global positive selection acting on U₃ sequences
781 could be associated with a conservation of molecular targets and function for U₃ peptides from
782 myrmicine ant venoms. Therefore, the positive selection on positions 10 and 11 could also take
783 part in interactions with new molecular targets. To identify them, it will thus be interesting to
784 test the action of U₃ peptides issued from different myrmicine ant venoms on mammalian Nav,
785 as well as to extend this study on other vertebrate Nav.

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

794 (Chang and Duda, 2012; Dutertre et al., 2014). Gene duplication, positive selection, and protein
795 neofunctionalization therefore appear to work in unison to provide the evolutionary novelty that
796 allows adaptation of venoms to different prey (Duda and Palumbi, 1999). However, defensive
797 toxins do not seem to follow this classic pattern of functional evolution and there is little
798 evidence for defense-related selective pressures on venom compositions, except in cone snails
799 (Dutertre et al., 2014). The understanding of these pressures as well as their role in venom
800 evolution remains poor (Casewell et al., 2013). Thus, defensive toxins evolved slowly and
801 appear with strong sequence conservation (Herzig et al., 2020), which is consistent with our
802 data for Nav toxins (U₃ peptides). Surprisingly, secapin-like peptides (U₁₇) showed the same
803 pattern as trophically toxins being mostly affected by positive selection. These results raise
804 further evolutive questions and highlight the need for more functional studies. Overall, in the
805 absence of doing the required genomic investigations, further transcriptomic analysis on
806 myrmicine ant venoms will also be needed to access more sequences and confirm these
807 evolutionary mechanisms of diversification.

808

809 **Acknowledgements:** V.B. was the recipient of a PhD fellowship from the French
810 Ministry of Scientific Research. This study was conducted in collaboration with the GeT core
811 facility, Toulouse, France (<http://get.genotoul.fr>). Financial supports were provided by the
812 France Génomique National institutional infrastructure partly funded “Investissement d’avenir”
813 program managed by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09), by
814 another Investissement d’Avenir grant of the Agence Nationale de la Recherche (CEBA: ANR-
815 10-LABX-25-01) and by the PO-FEDER 2014-2020, Région Guyane (FORMIC, GY0013708).
816 We would like to thank Robert A. Johnson from Arizona State University, for assistance and
817 delivery of *P. californicus* workers. Ant samples were collected under the authorizations of
818 both the Cameroon Ministry of Scientific Research and Innovation and the French Ministry of
819 Ecological and Solidarity Transition, in accordance with Article 17, paragraph 2, of the Nagoya
820 Protocol on Access and Benefit-sharing (Reference number of the permits: TREL1820249A/54
821 and TREL1916196S/214). The authors thank the NMR division of the MO2VING facility
822 (Orléans, France).

823

824 **References:**

825 Aili, S.R., Touchard, A., Escoubas, P., Padula, M.P., Orivel, J., Dejean, A., Nicholson, G.M.,
826 2014. Diversity of peptide toxins from stinging ant venoms. *Toxicon* 92, 166–178.

827 <https://doi.org/10.1016/j.toxicon.2014.10.021>

828 Aili, S.R., Touchard, A., Hayward, R., Robinson, S.D., Pineda, S.S., Lalagüe, H., Vetter, I.,
829 Undheim, E.A.B., Kini, R.M., Escoubas, P., Padula, M.P., Myers, G.S.A., Nicholson,
830 G.M., 2020. An integrated proteomic and transcriptomic analysis reveals the venom
831 complexity of the bullet ant *Paraponera clavata*. *Toxins (Basel)*. 12.
832 <https://doi.org/10.3390/toxins12050324>

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

837 Barassé, V., Touchard, A., Téné, N., Tindo, M., Kenne, M., Klopp, C., Dejean, A., Bonnafé,
838 E., Treilhou, M., 2019. The peptide venom composition of the fierce stinging ant
839 *Tetraponera aethiops* (Formicidae: Pseudomyrmecinae). *Toxins (Basel)*. 11.
840 <https://doi.org/10.3390/toxins11120732>

841 Blaimer, B.B., Ward, P.S., Schultz, T.R., Fisher, B.L., Brady, S.G., 2018. Paleotropical
842 diversification dominates the evolution of the hyperdiverse ant tribe Crematogastrini
843 (Hymenoptera: Formicidae). *Insect Syst. Divers.* 2. <https://doi.org/10.1093/isd/ixy013>

844 Brünger, A.T., 2007. Version 1.2 of the crystallography and nmr system. *Nat. Protoc.* 2, 2728–
845 2733. <https://doi.org/10.1038/nprot.2007.406>

846 Brünger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grossekunstleve, R.W.,
847 Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T.,
848 Warren, G.L., 1998. Crystallography & NMR system: a new software suite for
849 macromolecular structure determination. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 54,
850 905–921. <https://doi.org/10.1107/S09074444998003254>

851 Cabau, C., Escudié, F., Djari, A., Guiguen, Y., Bobe, J., Klopp, C., 2017. Compacting and
852 correcting Trinity and Oases RNA-Seq de novo assemblies. *PeerJ* 5, e2988.
853 <https://doi.org/10.7717/peerj.2988>

854 Casewell, N.R., Wüster, W., Vonk, F.J., Harrison, R.A., Fry, B.G., 2013. Complex cocktails:
855 The evolutionary novelty of venoms. *Trends Ecol. Evol.*
856 <https://doi.org/10.1016/j.tree.2012.10.020>

857 Cheung, M.S., Maguire, M.L., Stevens, T.J., Broadhurst, R.W., 2010. DANGLE: a Bayesian
858 inferential method for predicting protein backbone dihedral angles and secondary
859 structure. *J. Magn. Reson.* 202, 223–233. <https://doi.org/10.1016/j.jmr.2009.11.008>

860 Daly, N.L., Wilson, D., 2018. Structural diversity of arthropod venom toxins. *Toxicon* 152, 46–
861 56. <https://doi.org/10.1016/j.toxicon.2018.07.018>

862 DeLano, W.L., 2002. PyMOL: an open-source molecular graphic tool. *Ccp4.Ac.Uk*.

863 Dos Santos Pinto, J.R.A., Fox, E.G.P., Saidemberg, D.M., Santos, L.D., Da Silva Menegasso,
864 A.R., Costa-Manso, E., MacHado, E.A., Bueno, O.C., Palma, M.S., 2012. Proteomic view
865 of the venom from the fire ant *Solenopsis invicta* buren. *J. Proteome Res.* 11, 4643–4653.
866 <https://doi.org/10.1021/pr300451g>

867 Duraisamy, K., Singh, K., Kumar, M., Lefranc, B., Bonnafé, E., Treilhou, M., Leprince, J.,
868 Chow, B.K.C., 2021. P17 induces chemotaxis and differentiation of monocytes via
869 MRGPRX2-mediated mast cell-line activation. *J. Allergy Clin. Immunol.*
870 <https://doi.org/10.1016/j.jaci.2021.04.040>

871 Eagles, D.A., Saez, N.J., Krishnarjuna, B., Bradford, J.J., Chin, Y.K., 2022. A peptide toxin in
872 ant venom mimics vertebrate EGF-like hormones to cause long-lasting hypersensitivity in
873 mammals. *Proc. Natl. Acad. Sci.* 119, e2112630119.
874 <https://doi.org/10.1073/pnas.2112630119/-/DCSupplemental.Published>

875 Fox, E.G.P., Russ Solis, D., Delazari dos Santos, L., Aparecido dos Santos Pinto, J.R., Ribeiro
876 da Silva Menegasso, A., Cardoso Maciel Costa Silva, R., Sergio Palma, M., Correa Bueno,

877 O., de Alcântara Machado, E., 2013. A simple, rapid method for the extraction of whole
878 fire ant venom (Insecta: Formicidae: Solenopsis). *Toxicon* 65, 5–8.
879 <https://doi.org/10.1016/j.toxicon.2012.12.009>

880 Heep, J., Klaus, A., Kessel, T., Seip, M., Vilcinskas, A., Skaljac, M., 2019a. Proteomic analysis
881 of the venom from the ruby ant *Myrmica rubra* and the isolation of a novel insecticidal
882 decapeptide. *Insects* 10. <https://doi.org/10.3390/insects10020042>

883 Heep, J., Skaljac, M., Grotmann, J., Kessel, T., Seip, M., Schmidtberg, H., Vilcinskas, A.,
884 2019b. Identification and functional characterization of a novel insecticidal decapeptide
885 from the myrmicine ant *Manica rubida*. *Toxins (Basel)*. 11, 1–17.
886 <https://doi.org/10.3390/toxins11100562>

887 Hurka, S., Brinkrolf, K., Ozbek, R., Forster, F., Billion, A., Heep, J., Timm, T., Lochnit, G.,
888 Vilcinskas, A., Luddecke, T., 2022. Venomics of the central european myrmicine ants
889 *Myrmica rubra* and *Myrmica ruginodis*. *Toxins (Basel)*. 14.
890 <https://doi.org/10.3390/toxins14050358>

891 Hutchinson, E.G., Thornton, J.M., 1996. PROMOTIF: a program to identify and analyze
892 structural motifs in proteins. *Protein Sci.* 5, 212–220.

893 Jones, T.H., Zottig, V.E., Robertson, H.G., Snelling, R.R., 2003. The venom alkaloids from
894 some african *Monomorium* species. *J. Chem. Ecol.* 29, 2721–2727.

895 Kaas, Q., Westermann, J.C., Craik, D.J., 2010. Conopeptide characterization and
896 classifications: An analysis using ConoServer. *Toxicon* 55, 1491–1509.
897 <https://doi.org/10.1016/j.toxicon.2010.03.002>

898 Kazuma, K., Masuko, K., Konno, K., Inagaki, H., 2017. Combined venom gland transcriptomic
899 and venom peptidomic analysis of the predatory ant *Odontomachus monticola*. *Toxins*
900 (Basel). 9, 1–15. <https://doi.org/10.3390/toxins9100323>

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

904 Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary
905 genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549.
906 <https://doi.org/10.1093/molbev/msy096>

907 Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996.
908 AQUA and PROCHECK-NMR: programs for checking the quality of protein structures
909 solved by NMR. *J. Biomol. NMR* 8, 477–486.

910 Lee, K.S., Kim, B.Y., Yoon, H.J., Choi, Y.S., Jin, B.R., 2016. Secapin, a bee venom peptide,
911 exhibits anti-fibrinolytic, anti-elastolytic, and anti-microbial activities. *Dev. Comp.*
912 *Immunol.* 63, 27–35. <https://doi.org/10.1016/j.dci.2016.05.011>

913 Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler
914 transform. *Bioinformatics* 26, 589–595. <https://doi.org/10.1093/bioinformatics/btp698>

915 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
916 Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25,
917 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>

918 Mariano, D.O.C., de Oliveira, Ú.C., Zaharenko, A.J., Pimenta, D.C., Rádis-Baptista, G., Prieto-
919 Da-Silva, Á.R. de B., 2019. Bottom-up proteomic analysis of polypeptide venom
920 components of the giant ant *Dinoponera quadriceps*. *Toxins (Basel)*. 11.
921 <https://doi.org/10.3390/toxins11080448>

922 Morgan, E.D., 2008. Chemical sorcery for sociality: exocrine secretions of ants (Hymenoptera:
923 Formicidae). *Myrmecol. News* 11, 79–90.

924 Mourelle, D., Brigatte, P., Bringanti, L.D.B., De Souza, B.M., Arcuri, H.A., Gomes, P.C.,
925 Baptista-Saidemberg, N.B., Ruggiero Neto, J., Palma, M.S., 2014. Hyperalgesic and
926 edematogenic effects of Secapin-2, a peptide isolated from Africanized honeybee (*Apis*

927 mellifera) venom. *Peptides* 59, 42–52. <https://doi.org/10.1016/j.peptides.2014.07.004>

928 Orivel, J., Redeker, V., Le Caer, J.P., Krier, F., Revol-Junelles, A.M., Longeon, A., Chaffotte,
929 A., Dejean, A., Rossier, J., 2001. Ponericins, new antibacterial and insecticidal peptides
930 from the venom of the ant *Pachycondyla goeldii*. *J. Biol. Chem.* 276, 17823–17829.
931 <https://doi.org/10.1074/jbc.M100216200>

932 Pan, J., Hink, W.F., 2000. Isolation and characterization of myrmexins, six isoforms of venom
933 proteins with anti-inflammatory activity from the tropical ant, *Pseudomyrmex triplarinus*.
934 *Toxicon* 38, 1403–1413. [https://doi.org/10.1016/S0041-0101\(99\)00233-0](https://doi.org/10.1016/S0041-0101(99)00233-0)

935 Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S.,
936 Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M.,
937 Walzer, M., Wang, S., Brazma, A., Vizcaíno, J.A., 2022. The PRIDE database resources
938 in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.*
939 50, D543–D552. <https://doi.org/10.1093/nar/gkab1038>

940 Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H.,
941 Ferrin, T.E., 2021. UCSF ChimeraX: structure visualization for researchers, educators,
942 and developers. *Protein Sci.* 30, 70–82. <https://doi.org/10.1002/pro.3943>

943 R Core Team, 2017. R: a language and environment for statistical computing. *R. Found. Stat.*
944 *Comput.*

945 Radis-Baptista, G., Dodou, H. V., Prieto-da-Silva, A.R.B., Zaharenko, A.J., Nihei, K., Inagaki,
946 H., Mori-Yasumoto, K., Konno, K., 2020. Comprehensive analysis of peptides and low
947 molecular weight components of the giant ant *Dinoponera quadriceps* venom. *Biol. Chem.*
948 <https://doi.org/10.1515/hsz-2019-397ja-01>

949 Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open
950 Software Suite. *Trends Genet.* 16, 276–277. [https://doi.org/10.1016/S0168-](https://doi.org/10.1016/S0168-9525(00)02024-2)
951 [9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2)

952 Rieping, W., Bardiaux, B., Bernard, A., Malliavin, T.E., Nilges, M., 2007. ARIA2: automated
953 NOE assignment and data integration in NMR structure calculation. *Bioinformatics* 23,
954 381–382. <https://doi.org/10.1093/bioinformatics/btl589>

955 Rifflet, A., Gavalda, S., Téné, N., Orivel, J., Leprince, J., Guilhaudis, L., Génin, E., Vétillard,
956 A., Treilhou, M., 2012. Identification and characterization of a novel antimicrobial peptide
957 from the venom of the ant *Tetramorium bicarinatum*. *Peptides* 38, 363–370.
958 <https://doi.org/10.1016/j.peptides.2012.08.018>

959 Robinson, S.D., Mueller, A., Clayton, D., Starobova, H., Hamilton, B.R., Payne, R.J., Vetter,
960 I., King, G.F., Undheim, E.A.B., 2018. A comprehensive portrait of the venom of the giant
961 red bull ant, *Myrmecia gulosa*, reveals a hyperdiverse hymenopteran toxin gene family.
962 *Sci. Adv.* 4, eaau4640. <https://doi.org/10.1126/sciadv.aau4640>

963 Schmidt, J.O., Blum, M.S., 1978. The biochemical constituents of the venom of the harvester
964 ant, *Pogonomyrmex badius*. *Comp. Biochem. Physiol. Part C, Comp.* 61, 239–247.
965 [https://doi.org/10.1016/0306-4492\(78\)90137-5](https://doi.org/10.1016/0306-4492(78)90137-5)

966 Touchard, A., Aili, S.R., Fox, E.G.P., Escoubas, P., Orivel, J., Nicholson, G.M., Dejean, A.,
967 2016a. The biochemical toxin arsenal from ant venoms. *Toxins (Basel)*. 8, 1–28.
968 <https://doi.org/10.3390/toxins8010030>

969 Touchard, A., Aili, S.R., Téné, N., Barassé, V., Klopp, C., Dejean, A., Kini, R.M., Mrinalini,
970 M., Coquet, L., Jouenne, T., Lefranc, B., Leprince, J., Escoubas, P., Nicholson, G.M.,
971 Treilhou, M., Bonnafé, E., 2020. Venom peptide repertoire of the european myrmicine ant
972 *Manica rubida*: identification of insecticidal toxins. *J. Proteome Res.* 19, 1800–1811.
973 <https://doi.org/10.1021/acs.jproteome.0c00048>

974 Touchard, A., Brust, A., Cardoso, F.C., Chin, Y.K.Y., Herzig, V., Jin, A.H., Dejean, A.,
975 Alewood, P.F., King, G.F., Orivel, J., Escoubas, P., 2016b. Isolation and characterization
976 of a structurally unique β -hairpin venom peptide from the predatory ant *Anochetus*

977 emarginatus. *Biochim. Biophys. Acta - Gen. Subj.* 1860, 2553–2562.
978 <https://doi.org/10.1016/j.bbagen.2016.07.027>

979 Touchard, A., Koh, J.M.S., Aili, S.R., Dejean, A., Nicholson, G.M., Orivel, J., Escoubas, P.,
980 2015. The complexity and structural diversity of ant venom peptidomes is revealed by
981 mass spectrometry profiling. *Rapid Commun. Mass Spectrom.* 29, 385–396.
982 <https://doi.org/10.1002/rcm.7116>

983 Touchard, A., Téné, N., Chan Tchi Song, P., Lefranc, B., Leprince, J., Treilhou, M., Bonnafé,
984 E., 2018. Deciphering the molecular diversity of an ant venom peptidome through a
985 venomics approach. *J. Proteome Res.* 17, 3503–3516.
986 <https://doi.org/10.1021/acs.jproteome.8b00452>

987 von Sicard, N.A.E., Candy, D.J., Anderson, M., 1989. The biochemical composition of venom
988 from the pavement ant (*Tetramorium caespitum* L.). *Toxicon* 27, 1127–1133.
989 [https://doi.org/10.1016/0041-0101\(89\)90006-8](https://doi.org/10.1016/0041-0101(89)90006-8)

990 Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L.,
991 Markley, J.L., Ionides, J., Laue, E.D., 2005. The CCPN data model for NMR spectroscopy:
992 Development of a software pipeline. *Proteins Struct. Funct. Genet.* 59, 687–696.
993 <https://doi.org/10.1002/prot.20449>

994 Walker, A.A., Robinson, S.D., Yeates, D.K., Jin, J., Baumann, K., Dobson, J., Fry, B.G., King,
995 G.F., 2018. Entomo-venomics: the evolution, biology and biochemistry of insect venoms.
996 *Toxicon* 154, 15–27. <https://doi.org/10.1016/j.toxicon.2018.09.004>

997 Ward, P.S., Brady, S.G., Fisher, B.L., Schultz, T.R., 2015. The evolution of myrmicine ants:
998 phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae).
999 *Syst. Entomol.* 40, 61–81. <https://doi.org/10.1111/syen.12090>

1000