



HAL
open science

Etude de peptides de venin de fourmis : diversité moléculaire et lien avec la fonction immunitaire

Valentine Barassé

► **To cite this version:**

Valentine Barassé. Etude de peptides de venin de fourmis : diversité moléculaire et lien avec la fonction immunitaire. Sciences agricoles. Institut National Polytechnique de Toulouse - INPT, 2020. Français. NNT : 2020INPT0111 . tel-04193624

HAL Id: tel-04193624

<https://theses.hal.science/tel-04193624>

Submitted on 1 Sep 2023

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.



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Pathologie, Toxicologie, Génétique et Nutrition

Présentée et soutenue par :

Mme VALENTINE BARASSE

le mardi 3 novembre 2020

Titre :

Etude de peptides de venin de fourmis : diversité moléculaire et lien avec
la fonction immunitaire

Ecole doctorale :

Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Unité de recherche :

Biochimie et Toxicologie des Substances Bioactives (BTSB)

Directeur(s) de Thèse :

M. MICHEL TREILHOU

MME ELSA BONNAFE

Rapporteurs :

M. DENIS SERVENT, CNRS

Mme MARYLÈNE POIRIE, UNIVERSITE COTE D'AZUR

Membre(s) du jury :

M. JEAN-LUC GATTI, INRA PACA, Président

Mme ELSA BONNAFE, INSTITUT NATIONAL UNIV. CHAMPOLLION, Membre

M. MICHEL TREILHOU, ENSFEA DE TOULOUSE, Membre

Remerciements

Les premiers remerciements reviennent à mes directeurs, Michel Treilhou et Elsa Bonnafé. Merci à vous pour votre confiance sur ce projet, de mon stage de M2 jusqu'aux termes de ce doctorat. Merci de m'avoir accordée ma chance et de m'avoir accompagnée de par votre expertise, votre écoute et vos recommandations au cours de ces trois ans et demi. Malgré vos occupations, vous trouvez toujours un petit moment pour moi si le besoin s'en ressent. J'ai réellement apprécié ce travail à vos côtés et suis très heureuse de continuer.

Merci aux membres du comité de pilotage de ce travail de thèse pour leurs conseils avisés : Jérôme Orivel, Jérôme Leprince, Céline Brochier-Armanet. Je tiens à remercier personnellement Elisabeth Huguet, mon ancienne professeure et directrice de master qui encore aujourd'hui s'applique à me conseiller lorsque j'en ressens le besoin. Elisabeth, merci pour tes recommandations et pour cette dévotion dont tu fais preuve envers tes étudiants (anciens et actuels) que j'admire particulièrement.

Merci à Denis Servent et Marylène Poirié d'avoir évalué ce manuscrit malgré les courtes limites de temps, ainsi qu'à Jean-Luc Gatti d'avoir accepté le rôle d'examineur.

Nathan, un grand merci pour tes travaux d'analyse des venins et de relecture de ce manuscrit. Jean-Michel, merci de prendre autant soin de nos fourmis. A vous deux, un immense merci pour votre bonne humeur quotidienne et pour l'intérêt que vous avez porté à mes travaux. Les manip' auraient été nettement moins chouettes sans vous pour regarder par-dessus mon épaule et me demander « à quel point j'allais coûter cher » ou si « j'allais encore en zigouiller beaucoup ».

A Pierre (fournisseur officiel de rhum arrangé et distributeur de rire), Alice, Caroline, Florence, Jean-Luc, merci pour votre bonne humeur. Merci également à Arnaud, notre dernier venu, pour ses conseils avisés. Je vous souhaite à tous de vous épanouir au sein de vos champs de recherche et d'enseignement respectifs.

Une pensée particulière pour Jean, envahisseur voleur de piment et de café. Merci à toi pour ton soutien lors de la rédaction de ce manuscrit. Merci infiniment pour ton écoute, tes conseils et pour tes mauvaises blagues qui ne manquent jamais de me dérider.

Je tiens également à remercier Delphine pour son accueil et son écoute depuis mon arrivée à Champollion. Delphine, tu as été d'un soutien sans faille, dans tous les aspects et pour ceci tu as toute ma gratitude. Je remercie également Jenny Moro pour son accueil et sa sympathie.

Un merci spécial pour Karen Chevalier et Valérie Fernandez. Merci à vous pour votre gentillesse et pour m'avoir guidée au travers du parcours du combattant qu'est l'administration. C'est toujours un plaisir de venir vous voir.

A Raphaël et Emma, merci d'avoir été mes petites mains. Vous encadrer a été un vrai plaisir et je ne doute aucunement de votre futur dans la recherche. Belle route à vous.

Mon arrivée dans le sud de la France a ainsi été marquée de belles rencontres, tant sur le plan professionnel que personnel. Marion, Vincent, vous êtes présents depuis le début de mon aventure albigeoise, d'abord en tant que collègues puis en tant qu'amis. Je vous suis infiniment reconnaissante pour votre soutien sans faille. Malgré les embûches, vous faites preuve d'un courage et d'une détermination à toute épreuve, tout en restant présents pour chacun de vos proches. A nos retrouvailles, nos ragots, nos barbecues cramés, aux « bières pour hommes fragiles » que Vincent aime tant, et à votre petite Zoé.

A Axel, sans qui une bonne partie de ce travail de thèse n'aurait pu s'effectuer. Deux sarthois, de même formation, se retrouvent à Albi. Le monde est décidément bien petit et je n'aurais pu espérer mieux comme encadrant. Merci d'avoir partagé ta passion avec moi, de m'avoir formée et soutenue, et surtout merci de la patience et de l'implication dont tu fais encore preuve aujourd'hui. A très vite « chef », je te souhaite le meilleur dans tous les aspects de ta vie.

Au groupe des doctorants, anciens et nouveaux, un immense merci général pour votre présence et votre bonne humeur. Steven, Ibtissam, je n'ai aucun doute sur vos qualités scientifiques et vous souhaite le meilleur pour la fin de votre doctorat. Aux ergo, Eugénie et Charlotte, les moments passés en votre compagnie sont toujours un plaisir. A très vite !

Je tiens à remercier plus particulièrement Manon, Morgane (même galère les filles !), Ludo et Lucie (deux petits diables sur mon épaule gauche). Merci à vous pour votre écoute et votre soutien sans faille depuis notre rencontre. Ce n'est pas avec vous que je pourrais entamer un régime si l'envie m'en prenait mais le fait est que je n'aurais pu rêver de meilleurs acolytes. A nos séances de psy, nos embuscades en terrasses et nos repas beaucoup trop gras pour que les recettes soient prononcées. Santé !

A ceux qui m'ont toujours soutenue dans mes projets : sarthois (Hélène, Estelle, Camille), bourguignons (Yann, Nicolas, Guillaume) et tourangeaux (Aimie, Joey, Manon, Elisa, Marine et Vincent). Merci à vous pour votre bienveillance et votre présence. En attendant de pouvoir vous rendre visite, sachez que ma porte vous sera toujours ouverte.

A ma famille, merci à vous pour votre amour et votre soutien immuables, pour vos conseils ainsi que pour votre positivité. Merci de me supporter dans mes moments de panique, de vous mettre dans des situations improbables dans le seul but d'amuser la galerie et de me « booster » quand le besoin s'en fait ressentir. J'espère que ce manuscrit vous initiera au monde merveilleux des fourmis et vous permettra de comprendre les raisons de mon absence. J'ai hâte de vous retrouver.

Sommaire

1		
2	LISTE DES TRAVAUX	5
3	LISTE DES FIGURES ET DES TABLEAUX	7
4	INTRODUCTION GENERALE	11
5	CHAPITRE I : ÉTAT DE L'ART	15
6	1. LA FONCTION VENIMEUSE DANS LE REGNE ANIMAL : COMPOSITION, EVOLUTION ET	
7	INTERET SCIENTIFIQUE	15
8	1.1. Définition des venins et répartition dans le règne animal	15
9	1.2. Fonctions des venins dans le règne animal	16
10	1.3. Production des venins et diversité des systèmes vulnérants	20
11	1.4. Diversité moléculaire des venins : focus sur les venins de nature protéique	21
12	1.4.1. Synthèse et maturation des toxines protéiques	21
13	1.4.2. Évolution des toxines protéiques de venins – Mécanismes impliqués	25
14	1.5. Intérêt scientifique de l'étude des venins	28
15	2. METHODES D'ETUDES DE LA COMPOSITION DES VENINS : FOCUS SUR LES VENINS	
16	PEPTIDIQUES	31
17	2.1. Protéomique appliquée à la vénomique	32
18	2.2. Transcriptomique appliquée à la vénomique	33
19	2.3. Génomique appliquée à la vénomique	35
20	2.4. Traitement intégré des données issues des différentes approches « omiques » pour l'étude de la	
21	composition des venins	37

22	2.5. L'approche intégrative « vénomique » : une révolution pour la caractérisation exhaustive des venins	
23	d'espèces négligées	39
24	3. LES PEPTIDES DE VENINS DE FOURMIS	44
25	3.1. Les paraponeratoxines (Paraponerinae)	45
26	3.2. Les myrmeciitoxines (Myrmeciinae)	47
27	3.3. Les poneritoxines (Ponerinae)	52
28	3.4. Les ectatotoxines	54
29	3.5. Les pseudomyrmecitoxines	55
30	3.6. Les myrmecitoxines	57
31	4. L'IMMUNITE DES INSECTES	61
32	4.1. Immunité acquise vs. Immunité innée : définitions et application aux insectes	61
33	4.2. Médiation par la réponse cellulaire	63
34	4.3. Médiation par la réponse humorale	65
35	4.3.1. La voie Toll	66
36	4.3.2. La voie IMD	67
37	4.3.3. Host Defense Peptides de <i>Drosophila melanogaster</i> et peptides de venin de <i>Tetramorium</i>	
38	<i>bicarinatum</i>	69
39	5. OBJECTIFS GENERAUX	70
40	CHAPITRE II : DIVERSITE MOLECULAIRE DE PEPTIDES DE VENINS DE FOURMIS MYRMICINES	73
41	1. ARTICLE 1	79
42	2. ARTICLE 2	93
43	3. ARTICLE 3	105
44	CHAPITRE III : LIEN ENTRE LES TOXINES PEPTIDIQUES ET LA FONCTION IMMUNITAIRE	139

45	1.	ETUDE PRELIMINAIRE DU LIEN ENTRE LES PEPTIDES DU VENIN DE <i>TETRAMORIUM</i>	
46		<i>BICARINATUM</i> ET L'IMMUNITE INNEE DE LA FOURMI	140
47	2.	LOCALISATION DES PEPTIDES EN DEHORS DES APPAREILS VULNERANTS	144
48	2.1.	Localisation des transcrits codant certains peptides de venin dans l'abdomen de <i>T. bicarinatum</i>	145
49	2.2.	Localisation de la Bicarinaline et du P17 dans l'abdomen de <i>T. bicarinatum</i>	146
50	3.	REPONSE DES GENES CODANT LES PEPTIDES A UNE INFECTION BACTERIENNE	152
51	3.1.	Évaluation au niveau transcriptomique par la technique de qPCR	152
52	3.2.	Évaluation au niveau transcriptomique par la technique de la ddPCR	158
53	3.3.	Évaluation de la réponse au niveau protéique	161
54	4.	MATERIEL ET METHODES	166
55	4.1.	<i>Tetramorium bicarinatum</i> : généralités et élevage	166
56	4.2.	Dissections	169
57	4.3.	Localisation des peptides en dehors des appareils vulnérants	170
58	4.3.1.	Localisation des transcrits	170
59	4.3.2.	Localisation de la Bicarinaline et du P17	172
60	4.4.	Etude de la réponse des gènes codant pour les peptides à une infection bactérienne	173
61	4.4.1.	Microbiologie et infection des fourmis par voie orale	173
62	4.4.2.	Etude au niveau transcriptomique	175
63	4.4.3.	Etude au niveau protéique	178
64	4.5.	Protocoles utilisés lors de l'étude préliminaire	179
65	4.5.1.	Première localisation des transcrits codant pour la Bicarinaline et le P17 en dehors de l'abdomen	
66		179	
67	4.5.2.	Première estimation de la réponse des gènes bicarinaline et p17 suite à une infection bactérienne	
68		179	
69	4.5.3.	Extraction, traitement de l'ARN et rétrotranscription	179
70	4.5.4.	Amplification par PCR	179
71	4.5.5.	Protocole de qPCR	180

72	CONCLUSION GENERALE / PERSPECTIVES	181
73	REFERENCES BIBLIOGRAPHIQUES	185
74	SUPPLEMENTARY MATERIALS	207
75		
76		

Liste des travaux

77

78

79 Publications :

80 ➤ **Barassé, V.**; Touchard, A.; Téné, N.; Tindo, M.; Kenne, M.; Klopp, C.; Dejean, A.;
81 Bonnafé, E.; Treilhou, M. The peptide venom composition of the fierce stinging ant
82 *Tetraponera aethiops* (Formicidae: Pseudomyrmecinae). *Toxins* (Basel). **2019**, 11.

83 ➤ Touchard, A.; Aili, S.R.; Téné, N.; **Barassé, V.**; Klopp, C.; Dejean, A.; Kini, R.M.;
84 Mrinalini, M.; Coquet, L.; Jouenne, T.; et al. Venom peptide repertoire of the European
85 myrmicine ant *Manica rubida*: identification of insecticidal toxins. *J. Proteome Res.*
86 **2020**, 19.

87

88 Communications orales :

89 ➤ **Barassé, V.**; Touchard, A.; Téné, N.; Bonnafé, E.; Treilhou, M. Exploring the molecular
90 diversity of ant venoms reveals conserved toxin precursors and distinctive mature
91 peptides features among phylogenetic subfamilies. *19ème Colloque sur la Biologie de*
92 *l’Insecte (CBI)*, Juin 2019, Albi, France.

93 ➤ **Barassé, V.**; Touchard, A.; Téné, N.; Bonnafé, E.; Treilhou, M. Exploring the molecular
94 diversity of ant venoms reveals conserved toxin precursors and distinctive mature
95 peptides features among and within phylogenetic subfamilies. *30ème congrès de*
96 *l’Union Internationale de l’Etude des Insectes Sociaux (UIEIS) – section française.*
97 Août 2019, Avignon, France.

98 ➤ **Barassé, V.** ; Touchard, A. Les fourmis possèdent-elles les antibiotiques de demain ?
99 *Pause Café - Pause Musée - Muséum d’Histoire Naturelle de Gaillac.* Septembre 2019.
100 Gaillac, France.

101 ➤ **Barassé, V.** Explorer la diversité moléculaire des venins de fourmis. *Flash’Conf*
102 *Mystères. Nuit européenne des chercheur-e-s.* Septembre 2019, Albi, France.

103

104 Posters:

105 ➤ **Barassé, V.**; Touchard, A.; Téné, N.; Bonnafé, E.; Treilhou, M. Exploring the molecular
106 diversity of ant venoms reveals conserved toxin precursors and distinctive mature
107 peptides features among phylogenetic subfamilies. *21ème réunion du Groupe Français*
108 *des Peptides et des Protéines (GFPP21)*, Mai 2019, Amboise, France.

Liste des figures et des tableaux

Chapitre I : Etat de l'art

Figure 1 : Diversité taxonomique et fonctions primaires des venins.	15
Figure 2 : Exemples d'appareils vulnérants.	18
Figure 3 : Cycle sécrétoire classique de toxines protéiques.	20
Figure 4 : Cibles pharmacologiques de toxines protéiques isolées des venins.	22
Figure 5 : Exemples de mécanismes impliqués dans l'évolution des gènes codant les toxines.	25
Tableau I : Médicaments dérivés des venins approuvés par la FDA (Food Drug Administration).	27
Figure 6 : Approche de spectrométrie de masse dans le cadre de la détermination de la composition des venins.	31
Figure 7 : Méthodes de recherche des transcrits codant les toxines peptidiques à partir de données de RNA-Seq.	33
Figure 8 : Complémentarité des séquençages short et long reads dans l'assemblage de novo d'un génome de référence.	35
Figure 9 : Méthodologie intégrative « vénomique » et implication des différentes méthodes « -omiques » dans l'étude des toxines rentrant dans la composition des venins.	37
Figure 10 : Alignements des précurseurs de myrmecitoxines du venin de <i>Tetramorium bicarinatum</i> et séparation en superfamilles de précurseurs sur la base des régions prépro.	39
Figure 11 : Cladogrammes des précurseurs de toxines peptidiques du venin de <i>Tetramorium bicarinatum</i> .	41
Figure 1 : Diversité spécifique des sous-familles de Formicidae.	42
Figure 13 : Glandes à venin et réservoir de la fourmi myrmicine <i>Tetramorium bicarinatum</i> .	43
Figure 14 : Paraponeratoxines contenues dans les venins de différents spécimens de <i>Paraponera clavata</i> .	44
Figure 15 : Myrmeciitoxines contenues dans le venin de <i>Myrmecia pilosula</i> .	46
Tableau II : Toxines peptidiques contenues dans les venins des fourmis du genre <i>Myrmecia</i> .	48
Figure 16 : Alignements de précurseurs de toxines peptidiques de venin d'Hyménoptères sur la base des régions signal et propeptides.	49
Figure 17 : Poneritoxines isolées du venin d' <i>Anochetus emarginatus</i> .	51
Figure 18 : Poneritoxines de type Kunitz et ICK issus du venin de <i>Dinoponera quadriceps</i> .	52

Figure 19 : Ectatotoxines retrouvées dans les venins d' <i>Ectatomma tuberculatum</i> et <i>E. brunneum</i> .	53
Figure 20 : Pseudomyrmecitoxines retrouvées chez <i>Pseudomyrmex triplarinus</i> et <i>P. penetrator</i> .	55
Figure 21 : Myrmecitoxines isolées des venins de <i>Myrmica rubra</i> et <i>Manica rubida</i> .	56
Figure 22 : Séquences et représentations 3D des peptides matures Bicarinaline et P17.	57
Figure 23 : Alignement des séquences des myrmecitoxines matures de <i>Tetramorium bicarinatum</i> avec d'autres peptides de venins d'hyménoptères aculéates.	58
Figure 24 : Motifs moléculaires associés aux pathogènes et aux lésions tissulaires déclenchant la réponse immunitaire innée des insectes.	61
Figure 25 : Réponses cellulaires exercées par les hémocytes des insectes lors d'une infection microbienne.	62
Figure 26 : Détails moléculaires de la voie de signalisation intracellulaire Toll dans le cas d'une infection bactérienne.	65
Figure 27 : Détails moléculaires de la voie de signalisation intracellulaire IMD dans le cas d'une infection bactérienne.	66
Figure 28 : Structures des précurseurs de Host Defense Peptides de l'hémolymphe de <i>Drosophila melanogaster</i> et des toxines du venin de <i>Tetramorium bicarinatum</i> .	67
 <u>Chapitre II : Diversité moléculaire de peptides de venins de fourmis myrmecines</u>	
Figure 29 : Modèles d'études, positionnement dans la famille des Formicidae et détails de leur écologie.	71
Tableau III : Composition des peptidomes de venins de fourmis pseudomyrmecines et myrmecines.	74
 <u>Chapitre III : Lien entre les toxines peptidiques et la fonction immunitaire</u>	
Figure 30 : Schéma général de l'anatomie interne d'une fourmi et des organes composant le tube digestif et le système vulnérant.	135
Figure 31 : Expression des gènes <i>bicarinaline</i> et <i>p17</i> dans différents organes contenus dans l'abdomen de la fourmi <i>Tetramorium bicarinatum</i> .	135
Figure 32 : Expression des gènes <i>dorsal</i> , <i>relish</i> , <i>bicarinaline</i> et <i>p17</i> .	136
Figure 33 : Expression des gènes <i>pgrp-le</i> , <i>dorsal</i> , <i>bicarinaline</i> , et <i>p17</i> .	137
Figure 34 : Expression des gènes <i>rpl-18</i> , <i>actine</i> , <i>bicarinaline</i> , <i>p17</i> , <i>u3-Tb1a</i> , <i>u4-Tb1a</i> , <i>u5-Tb1a</i> , <i>u7-Tb1a</i> , <i>u8-Tb1a</i> , <i>u9-Tb1a</i> et <i>u16-Tb1a</i> .	140
Figure 35 : Localisation de la Bicarinaline et du P17 dans les glandes à venin de <i>Tetramorium bicarinatum</i> réalisée par immunohistochimie indirecte.	141

Figure 36 : Localisation de la Bicarinaline et du P17 dans les tubes digestifs de <i>Tetramorium bicarinatum</i> réalisée par immunohistochimie indirecte.	141
Figure 37 : Localisation de la Bicarinaline et du P17 dans les tubes de Malpighi de <i>Tetramorium bicarinatum</i> réalisée par immunohistochimie indirecte.	142
Figure 38 : Localisation de la Bicarinaline et du P17 dans les corps gras de <i>Tetramorium bicarinatum</i> réalisée par immunohistochimie indirecte.	143
Figure 39 : Localisation de la Bicarinaline et du P17 dans les corps gras de <i>Tetramorium bicarinatum</i> réalisée par immunohistochimie indirecte.	144
Figure 40 : Cellules du corps gras de la fourmi <i>Mycetarotes parallelus</i> (Myrmicinae : Attini).	145
Figure 41 : Répartition des C_T moyens et ratio d'expressions ($2^{-\Delta\Delta C_T}$) obtenus par qPCR dans les glandes à venin de <i>Tetramorium bicarinatum</i> .	148
Figure 42 : Répartition des C_T moyens et ratio d'expressions obtenus par qPCR dans les tissus de <i>Tetramorium bicarinatum</i> saines ou infectées par <i>Serratia marcescens</i> .	149
Tableau IV : Valeurs de rangs (W) et probabilités (P) obtenus suite aux tests de Mann-Whitney exécutés entre les valeurs des C_T moyens des gènes <i>rpl-18</i> , <i>dorsal</i> , <i>relish</i> , <i>p17</i> , <i>u3-Tb1a</i> , <i>u8-Tb1a</i> , <i>bicarinaline</i> et <i>u9-Tb1a</i> .	150
Tableau V : Valeurs de rangs (W) et probabilités (P) obtenus suite aux tests de Mann-Whitney exécutés entre les ratios d'expression (i.e. $2^{-\Delta\Delta C_T}$) des gènes de l'immunité (<i>dorsal</i> et <i>relish</i>) et codant pour les toxines du venin de <i>Tetramorium bicarinatum</i> (<i>p17</i> , <i>u3-Tb1a</i> , <i>u8-Tb1a</i> , <i>bicarinaline</i> et <i>u9-Tb1a</i>).	150
Figure 43 : Comparaison des données brutes obtenues en qPCR et en ddPCR pour un échantillon de glandes à venin.	153
Tableau VI : Valeurs de rangs (W) et probabilités (P) obtenus suite aux tests de Mann-Whitney exécutés entre les nombres de copies observées suite à une ddPCR pour les gènes <i>rpl-18</i> , <i>relish</i> , <i>p17</i> et <i>u3-Tb1a</i> chez des <i>Tetramorium bicarinatum</i> saines ou infectées par <i>Serratia marcescens</i> .	153
Figure 44 : Expressions des gènes <i>rpl-18</i> , <i>relish</i> , <i>p17</i> et <i>u3-Tb1a</i> déterminées par ddPCR dans différents organes de <i>Tetramorium bicarinatum</i> saines ou infectées par <i>Serratia marcescens</i> .	154
Figure 45 : Localisation de la Bicarinaline et du P17 dans des glandes à venin de <i>Tetramorium bicarinatum</i> des lots contrôle et infecté, réalisée par immunohistochimie indirecte.	156
Tableau VII : Valeurs de la statistique t, des degrés de liberté et de la probabilité obtenus suite aux tests t de Student, exécutés entre les intensités lumineuses mesurées pour les peptides Bicarinaline et P17 dans les organes de <i>Tetramorium bicarinatum</i> saines ou infectées par <i>Serratia marcescens</i> .	157

Figure 46 : Intensité des signaux observés par immunohistochimie indirecte pour les peptides Bicarinaline et P17 dans des organes de <i>Tetramorium bicarinatum</i> saines ou infectées par <i>Serratia marcescens</i> .	157
Figure 47 : Localisation de la Bicarinaline et du P17 dans des tubes de Malpighi de <i>Tetramorium bicarinatum</i> des lots contrôle et infecté, réalisée par immunohistochimie indirecte.	158
Figure 48 : Localisation de la Bicarinaline et du P17 dans des corps gras de <i>Tetramorium bicarinatum</i> des lots contrôle et infecté, réalisée par immunohistochimie indirecte.	159
Figure 49 : <i>Tetramorium bicarinatum</i> (Hymenoptera : Formicidae).	161
Figure 50 : Carte de répartition de <i>Tetramorium bicarinatum</i> .	162
Figure 51 : Nid artificiel de <i>Tetramorium bicarinatum</i> .	163
Tableau III : Couples d'amorces utilisés pour la PCR.	165
Figure 52 : Microbiologie et infection par voie orale de la fourmi <i>Tetramorium bicarinatum</i> par la bactérie <i>Serratia marcescens</i> .	168
Tableau IX : Amorces utilisées pour la qPCR.	170
Tableau X : Couples d'amorces utilisés pour les PCR et la qPCR.	174
Tableau XI : Couple d'amorces utilisé pour le gène <i>pgrp-le</i> en qPCR.	174
Figure 53 : Caractéristiques structurales des familles de toxines peptidiques matures caractérisées dans les venins de fourmis myrmicines et pseudomyrmecine.	176

Introduction générale

112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143

Avec plus de 220 000 espèces venimeuses décrites, les organismes venimeux représentent environ 15% de la biodiversité animale actuelle (Schendel, Rash, Jenner, & Undheim, 2019). Les venins sont des mélanges complexes constitués d'une multitude de molécules appelées toxines (e.g. sels, polyamines, nucléosides, protéines, peptides), et constituent un trait adaptatif très complexe. Le coût métabolique de la production de venin est tel qu'au cours de l'évolution, seules les toxines les plus efficaces selon les contextes d'utilisations ont été sélectionnées (Morgenstern & King, 2013). Les venins sont donc des bibliothèques naturelles de composés bioactifs hautement optimisés. La plupart d'entre eux sont composés de peptides, petites protéines dont la longueur maximale est comprise entre 50 et 100 acides aminés. L'étude des peptides de venins d'animaux emblématiques tels que les serpents ou les cônes marins ont d'ailleurs donné lieu au développement de médicaments actuellement commercialisés (Pennington, Czerwinski, & Norton, 2018). Des nombreux animaux venimeux n'ont cependant pas été étudiés. La nature et les propriétés de leurs venins demeurent ainsi inconnues et inexploitées. C'est dans ce contexte que l'équipe d'accueil « Biochimie et Toxicologie des Substances Bioactives » (EA-7417-BTSB) étudie les venins de fourmis. Le développement des technologies dans les domaines de la transcriptomique et de la protéomique offre aujourd'hui la possibilité d'explorer ces venins en détail. La caractérisation du venin de la fourmi *Tetramorium bicarinatum* a mené à l'identification de toxines peptidiques qui, bien que présentant des séquences matures divergentes, se classent seulement en 3 grandes familles de précurseurs. Certaines se sont révélées être décrites chez d'autres arthropodes venimeux, posant la question de l'origine évolutive et des mécanismes impliqués dans la diversification des toxines de venins de fourmis. Il est de plus apparu que les gènes codant certaines de ces toxines s'expriment en dehors du système vulnérant, notamment dans des organes impliqués dans l'immunité innée des insectes, interrogeant sur l'existence d'un lien entre la fonction venimeuse et l'immunité innée de la fourmi à titre individuel. Ces résultats nous ont conduits à élaborer ce travail de thèse qui a pour objectifs :

- Apporter un premier aperçu de la diversité peptidique contenue dans les venins de fourmis myrmicines
- Évaluer l'implication des toxines de venin de fourmis dans leur immunité innée à titre individuel.

144 Les études menées dans le cadre de ce doctorat ont été financées par un contrat doctoral
145 (ex-bourse MRT) ainsi que par le budget propre à l'équipe BTSB. Ce manuscrit se divise en
146 trois chapitres.

147 Le premier chapitre présente un état de l'art axé sur la fonction venimeuse dans le règne
148 animal et sur l'immunité innée des insectes. Dans une première partie, la répartition de ces
149 cocktails biochimiques dans les différents phyla ainsi que leurs différents contextes d'utilisation
150 d'un point de vue écologique seront exposés. La diversité de composition des venins en termes
151 de toxines protéiques sera décrite ainsi que leurs processus de synthèse et certains mécanismes
152 impliqués dans leur évolution. Le dernier point de cette partie montrera l'intérêt scientifique de
153 l'étude de ces toxines, et les applications en ayant résulté. Dans la seconde partie de l'état de
154 l'art, un bilan des méthodes utilisées dans le cadre de l'étude de la composition des venins de
155 nature peptidique sera effectué. A cette occasion, j'expliquerai en quoi le développement des
156 méthodes « omiques », et le traitement intégré des données en résultant, constituent une
157 avancée majeure pour l'étude de la composition de venins d'espèces négligées de par leur taille.
158 Cette stratégie sera notamment étayée par la description de l'approche intégrative vénomique
159 ayant permis la caractérisation du venin de la fourmi *T. bicarinatum*, modèle historique de
160 l'équipe BTSB. La troisième partie de ce chapitre présente une synthèse des toxines peptidiques
161 caractérisées jusqu'à présent dans les venins de fourmis. Dans ce bilan, je présenterai
162 l'ensemble des caractéristiques biochimiques de ces toxines matures, leurs activités biologiques
163 et la structure de leurs précurseurs. Je poursuivrai dans une quatrième partie en décrivant les
164 principaux mécanismes impliqués dans l'immunité innée des insectes. Dans ce contexte, les
165 voies de signalisation immunitaires les plus décrites seront présentées. Je décrirai également les
166 structures des précurseurs et les activités biologiques de certains peptides impliqués dans
167 l'immunité innée de *Drosophila melanogaster*. Ce chapitre se conclura par la présentation des
168 objectifs généraux de ce travail de thèse.

169 Le second chapitre s'intitule « Diversité moléculaire des peptides de venins de fourmis
170 myrmicines » et présente les résultats relatifs à la caractérisation des profils peptidiques des
171 venins de huit espèces de fourmis appartenant à différentes sous-familles et tribus
172 phylogénétiques. J'introduirai ce chapitre en présentant les origines phylogénétiques et les
173 écologies des modèles biologiques. Je dresserai ensuite un aperçu de la diversité moléculaire
174 des toxines peptidiques contenues dans ces venins. Les résultats sont détaillés dans trois articles,
175 dont un en cours de préparation.

176 Le troisième chapitre est intitulé « Lien entre les toxines peptidiques et la fonction
177 immunitaire » et présente une étude exploratoire du lien entre la fonction venimeuse et
178 l'immunité innée de *T. bicarinatum*. Je décrirai dans un premier temps les résultats de l'étude
179 préliminaire des lieux d'expression des gènes codant pour les deux peptides majoritaires de ce
180 venin ainsi que leur réponse à une infection bactérienne. Je détaillerai ensuite ces mêmes
181 paramètres, étendus à d'autres peptides du venin de cette fourmi dans le cadre de ce travail de
182 thèse.

183 Enfin, ce manuscrit s'achèvera par un récapitulatif des principaux résultats et des
184 perspectives de poursuite d'étude.

185

186

188

Chapitre I : État de l'art

189

190 **1. La fonction venimeuse dans le règne animal : composition, évolution et intérêt** 191 **scientifique**

192 1.1. *Définition des venins et répartition dans le règne animal*

193 Les venins sont par définition des mélanges de substances bioactives communément
194 appelées « toxines » telles que des protéines et d'autres composés (Ashwood, Norton,
195 Undheim, Hurwood, & Prentis, 2020; B. G. Fry et al., 2009; Rima et al., 2018; Senji Laxme,
196 Suranse, & Sunagar, 2019; Vassilevski, Kozlov, & Grishin, 2009). Sécrétés par des glandes
197 spécialisées et injectés activement via un système vulnérant, ils constituent de véritables
198 arsenaux biochimiques développés au cours de l'évolution (Casewell, Wüster, Vonk, Harrison,
199 & Fry, 2013).

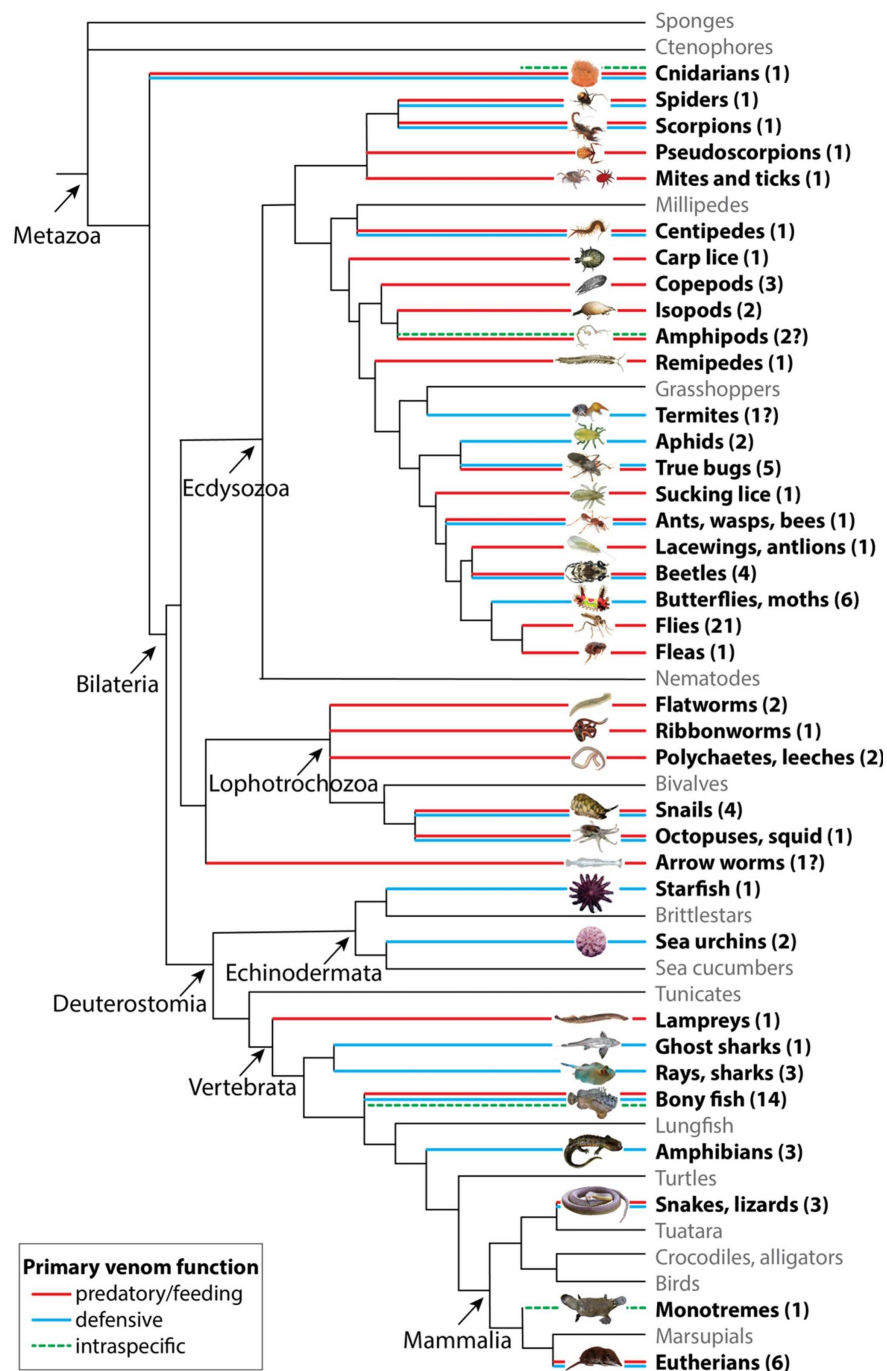
200 Parmi les 9 millions d'espèces animales présentes sur Terre, plus de 200 000 produisent des
201 venins, dont les composants ont été sélectionnés durant 600 millions d'années d'évolution pour
202 une efficacité et une sélectivité optimale (Jimenez, Ikonomopoulou, Lopez, & Miles, 2017;
203 Zancolli & Casewell, 2019). Ces espèces sont réparties au sein de phyla majeurs tels que les
204 Chordés (reptiles, poissons, amphibiens, mammifères), les Échinodermes (étoiles de mer, les
205 oursins), les Mollusques (cônes, pieuvres), les Annélides (sangues), les Némertes, les
206 Arthropodes (araignées, insectes, myriapodes) et les Cnidaires (anémones de mer, méduses,
207 coraux) (Casewell et al., 2013; Ménez, Stöcklin, & Mebs, 2006; Schendel et al., 2019; Zancolli
208 & Casewell, 2019). Les venins sont ainsi apparus de manière indépendante plus de 100 fois
209 dans une large gamme de taxons incluant au moins 8 phyla séparés (Figure 1) (Schendel et al.,
210 2019). Le pourcentage d'espèces venimeuses est très variable dans un même phylum ou une
211 même classe. Il est par exemple faible chez les mammifères où seules quelques espèces sont
212 concernées comme les ornithorynques, les taupes, les musaraignes ou encore les loris du genre
213 *Nycticebus* (Nekaris, Moore, Johanna Rode, & Fry, 2013; Schendel et al., 2019). Il est en
214 revanche égal à 100% chez les scorpions et les cnidaires.

215 Les venins sont donc omniprésents dans le règne animal et constituent des innovations
216 évolutives clés dans une large gamme de lignées phylogénétiques, notamment de par leurs
217 multiples fonctions (Figure 1) (B. G. Fry et al., 2009).

218

219 *1.2. Fonctions des venins dans le règne animal*

220 Les fonctions les plus évidentes et communes des venins sont décrites dans des contextes
221 de prédation et de défense (Schendel et al., 2019). Les serpents, araignées, scorpions et
222 myriapodes utilisent par exemple leurs venins pour immobiliser ou tuer la proie avant de la
223 consommer. *A contrario*, les abeilles et la plupart des poissons venimeux utilisent
224 principalement leurs venins pour se défendre contre les prédateurs (Arbuckle, 2017).



225

226 **Figure 1** : Diversité taxonomique et utilisations majoritaires des venins. Cladogramme illustrant la
 227 fréquence d'apparition des venins dans le règne animal. Les branches colorées indiquent les lignées
 228 présentant des organismes venimeux. Les branches rouges indiquent un rôle offensif des venins
 229 (prédation). Les branches bleues indiquent un rôle défensif. Les branches vertes en pointillés indiquent
 230 un rôle dans la compétition intraspécifique. Les points d'interrogation indiquent les taxons pour

231 lesquels aucune détermination de la nature de leur venin n'a été effectuée. Tiré de (Schendel et al.,
232 2019).

233 Néanmoins, d'autres fonctions ont été observées à travers le règne animal. Dans des
234 contextes intraspécifiques, les ornithorynques mâles utilisent leurs éperons venimeux pour
235 affronter d'autres mâles durant la saison de reproduction. Les scorpions mâles de plusieurs
236 groupes taxonomiques injecteraient également une petite quantité de venin dans le corps de la
237 femelle durant l'accouplement, ce qui produirait des effets sédatifs et aphrodisiaques (Arbuckle,
238 2017).

239 Dans des contextes inter-spécifiques, au-delà de la prédation immédiate, les venins peuvent
240 être utilisés dans la conservation des proies ou encore dans certains cas leur pré-digestion.
241 Ainsi, les taupes et les musaraignes injectent leur venin dans leurs proies afin de les paralyser
242 sans les tuer, les laissant immobilisées dans leurs terriers pour les consommer plus tard (Dufton,
243 1992). De manière similaire, certaines guêpes de la famille des Pompilidae paralysent des
244 tarentules pour pouvoir pondre leurs œufs sur ces dernières (= ectoparasitisme) (Petrunkevitch,
245 1926; F. X. Williams, 1956). Certains endoparasitoïdes utilisent leur venin pour transformer les
246 organismes en un nouvel habitat ou nurserie pour leur descendance lors de l'éclosion. A titre
247 d'exemple, la guêpe *Nasonia vitripennis* utilise son venin pour supprimer certaines fonctions
248 immunitaires de son hôte de manière sélective, pour arrêter son développement et manipuler
249 l'environnement nutritionnel interne (Martinson et al., 2014). Concernant les cas de pré-
250 digestion, certaines triatomes (Reduviidae) tels que *Pristhesancus plagipennis* possèdent des
251 venins paralysants et liquéfiant qui facilitent la prise de repas sanguin (Walker et al., 2017). Il
252 a également été observé une utilisation de venin en tant qu'herbicide chez la fourmi
253 *Myrmelachista schumanni*. Cette fourmi amazonienne injecte son venin composé d'acide
254 formique dans toutes les plantes présentes autour de sa plante hôte (myrmécophyte). Ce
255 comportement original tue les autres plantes et favorise ainsi le développement du
256 myrmécophyte dans une zone de monoculture appelée « jardins du Diable » (Frederickson,
257 Greene, & Gordon, 2005).

258 Outre la fonction de prédation, l'utilisation des venins dans des contextes particuliers de
259 défense, comme la détoxification, a pu être observée. En effet, les fourmis *Nylanderia fulva*
260 appliquent leurs venins composés d'acide formique directement sur leurs cuticules afin de
261 contrer l'effet du venin de la fourmi de feu *Solenopsis invicta* (LeBrun, Jones, & Gilbert, 2014).
262 Certains venins possèderaient également des propriétés antimicrobiennes (Yacoub, Rima,
263 Karam, Sabatier, & Fajloun, 2020). Les guêpes *Polistes dominulus* se recouvrent de venin en

264 guise de protection contre les pathogènes (Turillazzi et al., 2006). Un autre exemple est celui
265 des abeilles *Apis mellifera*, *A. cerana*, *A. dorsata* et *A. andreniformis*. Ces quatre espèces
266 présentent non seulement des traces de venin sur leurs cuticules mais répartissent également
267 leurs venins respectifs dans les cires des nids. Outre l'utilisation classique de défense contre les
268 prédateurs et les pathogènes à titre individuel, le venin semble également être un composant de
269 l'immunité sociale pour ces espèces (Baracchi, Francese, & Turillazzi, 2011). Cette utilisation
270 particulière du venin a également été recensée chez la fourmi *Formica paralugubris* (Brütsch,
271 Jaffuel, Vallat, Turlings, & Chapuisat, 2017).

272 L'utilisation du venin en tant que désinfectant est également notée chez d'autres espèces de
273 Formicidae dans le cas où un individu est contaminé, notamment *via* des cas de toilettages
274 (auto-procuré, interindividuel entre adultes ou au couvain) en vue de limiter la propagation de
275 l'infection et pouvant mener au sacrifice de l'individu infecté (Cremer, 2019; Tragust, 2016).
276 Il a par exemple été montré que le blocage de l'ouverture des glandes à venin des ouvrières
277 *Lasius neglectus* résultait en la survie réduite des adultes et des larves lors d'une infection par
278 *Metharhizium anisopliae* (Allcock et al., 2017). Les ouvrières *L. neglectus* sont en effet
279 capables de repérer les larves infectées par le champignon et de les désinfecter en vaporisant
280 leur venin, tuant d'un seul coup le pathogène et la larve (Baracchi & Tragust, 2020; Pull et al.,
281 2018; Tragust et al., 2013). Chez les fourmis, cette utilisation particulière du venin pour la
282 protection individuelle ou celle des congénères n'a néanmoins été décrite que chez des espèces
283 de Formicinae, dont les venins sont majoritairement composés d'acide formique (Baracchi &
284 Tragust, 2020).

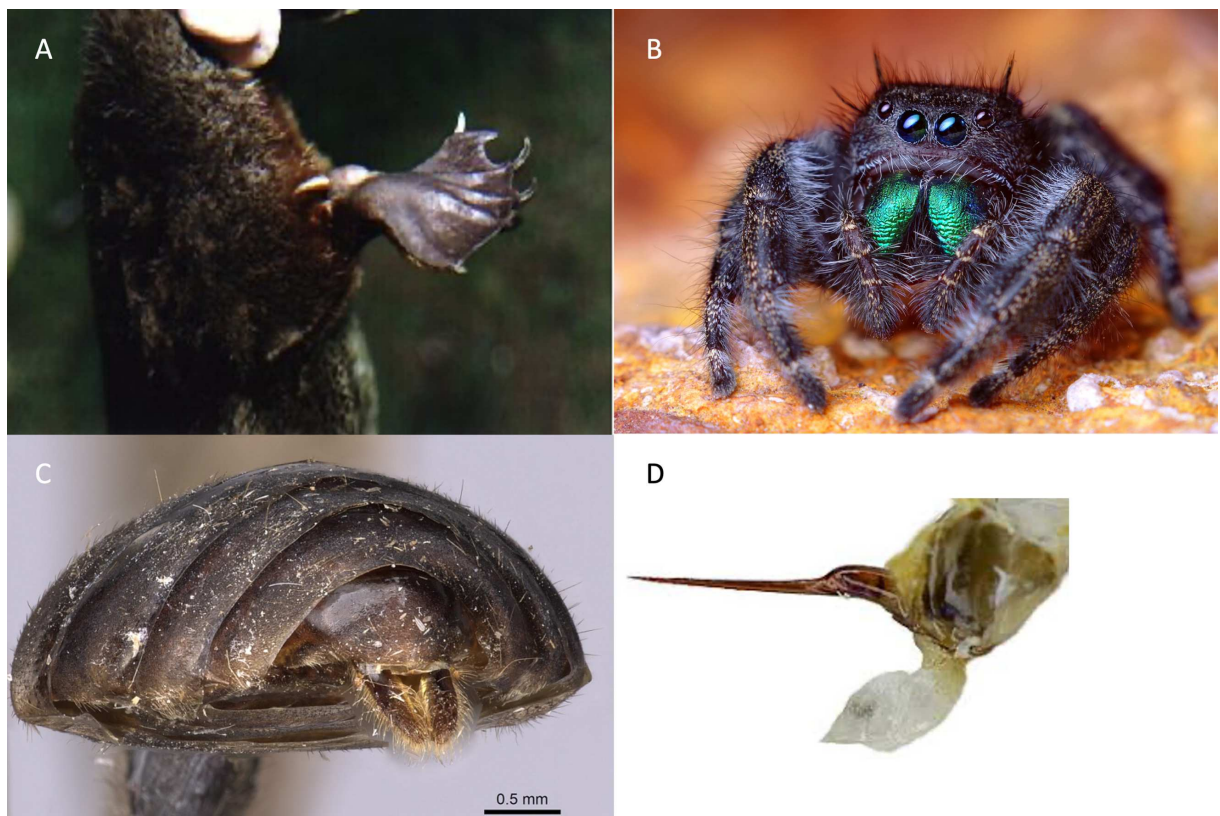
285 Il est essentiel de noter que, si les venins peuvent servir différentes fonctions, ils peuvent
286 également être utilisés à titre multifonctionnel par la majorité des espèces venimeuses
287 (Schendel et al., 2019). En effet, les araignées, serpents, scorpions et myriapodes utilisent non
288 seulement leur venin dans des contextes de prédation mais l'emploient également pour se
289 défendre contre des prédateurs *via* des morsures ou piqûres défensives (Schendel et al., 2019).
290 De plus, chez certains animaux, tels que le cône marin *Conus geographus* ou la punaise
291 *Prithesancus plagipennis*, il existe une production localisée du venin dont la composition est
292 dépendante du contexte de son utilisation. Cette production s'effectue soit de manière localisée
293 à l'intérieur d'un même canal venimeux chez *C. geographus*, soit dans des structures
294 anatomiques distinctes chez *P. plagipennis* (Dutertre et al., 2014; Walker, Mayhew, et al.,
295 2018).

296

297 1.3. Production des venins et diversité des systèmes vulnérants

298 Selon les phyla d'animaux venimeux et l'utilisation de leurs venins, les systèmes vulnérants
299 présentent des morphologies variées. On distingue par exemple les crochets buccaux des
300 araignées, les aiguillons abdominaux des hyménoptères, les rostres piqueurs de certains insectes
301 ou encore les éperons des membres postérieurs des ornithorynques mâles, tous relatifs à une
302 administration par injection (Rollard, Chippaux, & Goyffon, 2015; Whittington & Belov,
303 2007). Les venins peuvent également être « projetés », comme chez le cobra cracheur *Naja*
304 *nigricollis* (Rivel et al., 2016), ou même appliqués directement sur la cible comme pour la
305 fourmi *Myrmicaria opaciventris* qui possède un aiguillon modifié en forme de spatule
306 (Gathalkar & Sen, 2018) (Figure 2). Dans tous les cas, ces appareils vulnérants sont associés à
307 des glandes spécialisées de type exocrine en charge de la production du venin. L'origine
308 embryonnaire des cellules constituant les glandes à venin est, dans la quasi-totalité des cas,
309 ectodermique (Rollard et al., 2015).

310



311 **Figure 2** : Exemples d'appareils vulnérants. (A) Éperon de membre postérieur d'ornithorynque mâle. (B)
312 Chélicères de l'araignée *Phidippus audax*. (C) Aiguillon modifié en forme de spatule de la fourmi
313 *Myrmicaria opaciventris* (Antweb, le 29/07/20). (D) Aiguillon d'*Apis mellifera*.

315

316 *1.4. Diversité moléculaire des venins : focus sur les venins de nature protéique*

317 Bon nombre de scientifiques s'accordent à dire que la nature et la diversité moléculaire des
318 venins reflètent souvent leurs fonctions. Par exemple, les venins dits défensifs, comme ceux
319 des poissons ou encore des abeilles, présenteraient une composition moléculaire simplifiée et
320 hautement conservée, provoquant une douleur immédiate et localisée chez le prédateur
321 (Casewell et al., 2013; J. O. Schmidt, 2019a; Walker, Robinson, et al., 2018; Ziegman &
322 Alewood, 2015). *A contrario*, les venins servant à la prédation seraient souvent plus complexes,
323 et variables dans leurs compositions et leurs effets physiologiques, en relation notamment avec
324 la spécialisation du régime alimentaire de l'animal venimeux (Casewell et al., 2013; Dutertre
325 et al., 2014; Jin et al., 2019; Michalek, Kuhn-Nentwig, & Pekar, 2019; Pla et al., 2017). Cette
326 affirmation reste cependant sujette à controverse, certaines données suggérant que différentes
327 molécules seraient trouvées chez des espèces d'un même genre pour tuer le même type de
328 proies, comme c'est le cas pour les serpents du genre *Bothriechis* (Pla et al., 2017). Dutertre et
329 ses collaborateurs ont également montré que les venins défensifs des cônes marins *C.*
330 *geographus* et *C. marmoreus* sont plus complexes en termes de composition que leurs venins
331 offensifs respectifs (Dutertre et al., 2014). Ainsi, d'autres études reliant la composition des
332 venins et leurs fonctions écologiques chez des modèles biologiques variés seraient nécessaires
333 pour amender ou réfuter cette affirmation (Schendel et al., 2019).

334 Selon les phyla d'animaux venimeux voire selon les espèces dans un même genre, la
335 composition d'un venin est variable. Certains facteurs tels que la provenance géographique, le
336 poids, la taille ou l'âge de l'animal peuvent également entrer en compte (Chippaux, Williams,
337 & White, 1991; Servent, 2016).

338 Parmi les venins les plus étudiés, ceux de serpents sont connus pour être riches en enzymes
339 (phospholipases A2, métalloprotéases, L-aminoacide oxidase) et autres protéines (Rima et al.,
340 2018). Les venins des cônes marins et des arthropodes (araignées, scorpions, myriapodes,
341 insectes) sont connus pour être riches en peptides, de petites protéines dont la longueur
342 maximale est comprise entre 50 et 100 acides aminés (Prashanth et al., 2016; Senji Laxme et
343 al., 2019).

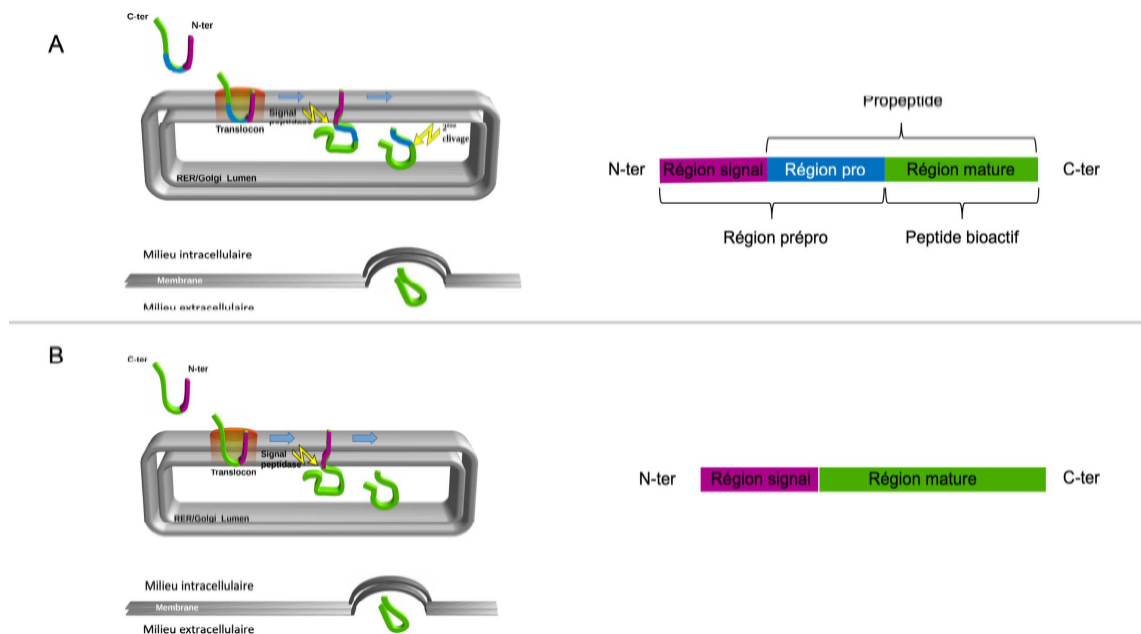
344

345 *1.4.1. Synthèse et maturation des toxines protéiques*

346 La synthèse de toxines protéiques par les glandes spécialisées obéit à un cycle sécrétoire
347 classique : un adressage des précurseurs en cours de synthèse au réticulum endoplasmique (RE)
348 *via* une séquence signal majoritairement N-terminale, puis, après un passage par l'appareil de
349 Golgi, une sécrétion apicale *via* l'exocytose des vésicules de sécrétion. Durant ce trafic

350 intracellulaire, les précurseurs subissent des modifications post-traductionnelles (PTMs)
 351 variées incluant un ou plusieurs clivages protéolytiques qui aboutissent à la formation d'une
 352 toxine mature (Figure 3). Ainsi, l'organisation des précurseurs est majoritairement constante :
 353 une séquence signal hautement conservée à l'extrémité N-terminale, une région pro-peptide
 354 intermédiaire, et une unique copie de la région mature de la toxine à l'extrémité C-terminale. Il
 355 y a ensuite sécrétion de la toxine mature par clivage de la région prépro (Figure 3). Il existe des
 356 modèles de production moins communs où un précurseur donne plusieurs toxines mures.
 357 C'est par exemple le cas des laticrines-6 isolées du venin de l'araignée *Lachesana tarabaevi*
 358 (Kozlov et al., 2006).

359 D'autres PTMs telles que les glycosylations, l'amidation C-terminale ou encore la
 360 formation de ponts disulfures peuvent également survenir (Buczek et al., 2005).
 361

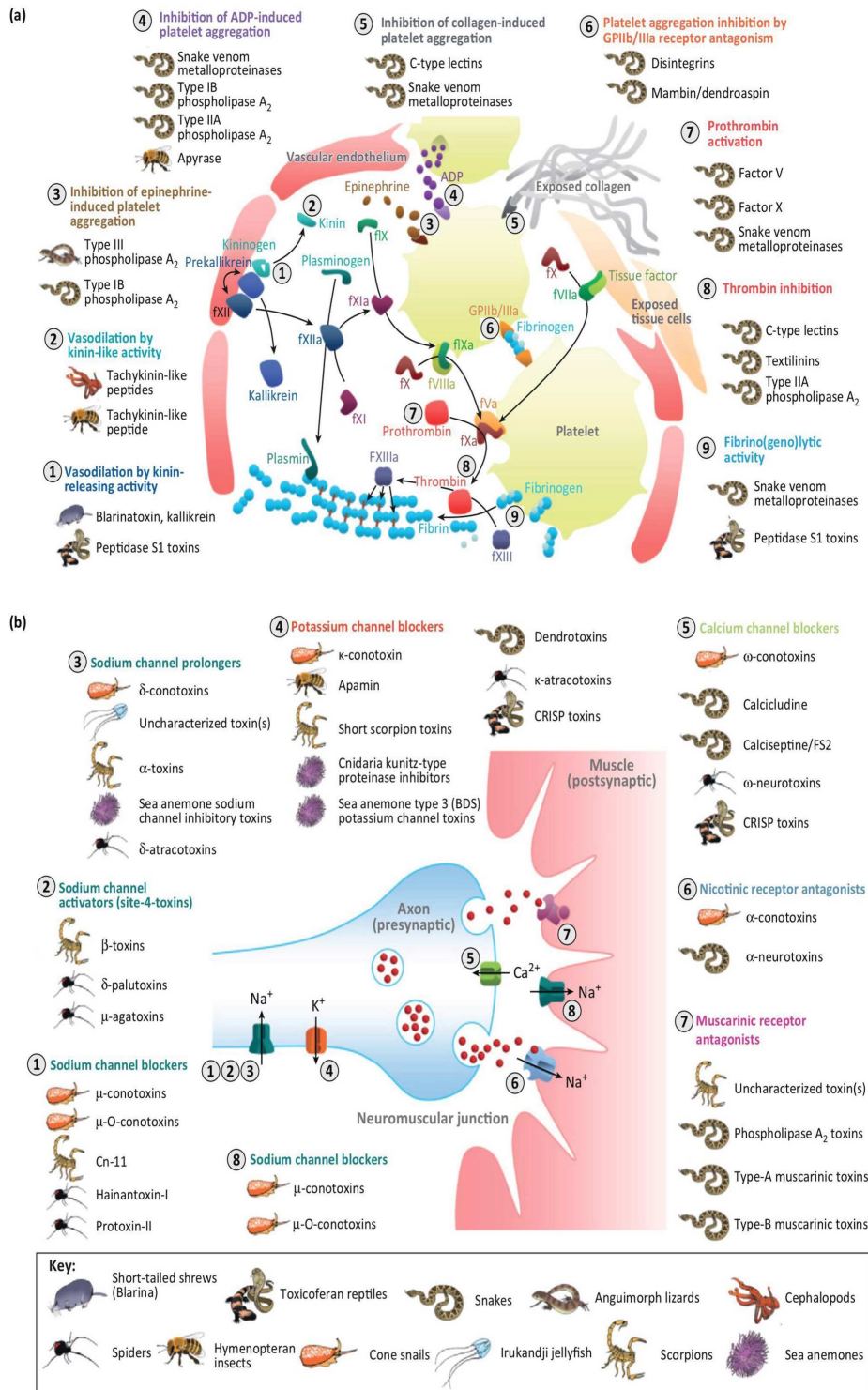


362
 363 **Figure 3 :** Cycle sécrétoire classique de toxines protéiques. Les précurseurs en cours de synthèse sont
 364 adressés au réticulum endoplasmique rugueux (RER) grâce à une séquence signal N-terminale. Le
 365 clivage de la région signal par la Signal peptidase permet la libération du précurseur dans la lumière du
 366 RER et la poursuite de la maturation jusqu'à la sécrétion. (A) Beaucoup de toxines sont synthétisées en
 367 tant que précurseurs inactifs, appelés propeptides. Leur activation est effectuée via le clivage de la région
 368 pro. (B) D'autres précurseurs de toxines ne présentent pas de région pro. La région mature est libérée
 369 dans la lumière du RER et la maturation se poursuit jusqu'à la sécrétion sans second clivage.

370 D'un point de vue pharmacologique, les cibles des venins incluent la plupart des voies
 371 physiologiques et tissus accessibles *via* le système circulatoire (Casewell et al., 2013) (Figure
 372 4). Une grande majorité de toxines peptidiques agissent sur les canaux ioniques en tant que

373 modulateurs. Elles affectent généralement le système nerveux et sont les composants actifs
374 principaux des venins d'araignées, des scorpions, des cônes marins et des serpents Elapidae
375 (Kalia et al., 2015; Utkin, 2015). Les neurotoxines contenues dans les venins d'araignées et de
376 scorpions agissent principalement sur des canaux ioniques voltage-dépendants, tandis que
377 celles contenues dans les venins de serpent ciblent des canaux chimio-dépendants comme les
378 récepteurs nicotiniques de l'acétylcholine (Utkin, 2015). Chez les cônes marins, plusieurs
379 conotoxines modulent les canaux ioniques (calcium, sodium et potassium), les récepteurs
380 nicotiniques de l'acétylcholine, les transporteurs de la noradrénaline, les récepteurs N-méthyl-
381 D-aspartate et les récepteurs à la neurotensine (Mir, Karim, Kamal, Wilson, & Mirza, 2016).

382 Les toxines protéiques d'un même venin vont agir de manière synergique contre leurs cibles
383 biologiques et causer des dysfonctionnements significatifs des systèmes nerveux,
384 cardiovasculaire et musculaire (Utkin, 2015). Ceci a notamment été prouvé pour le venin du
385 scorpion *Mesobuthus eupeus*, dont les neurotoxines, caractérisées comme ciblant les canaux
386 sodiques voltage-dépendants, augmentent l'efficacité contre des bactéries du peptide Meucin-
387 49 contenu dans le même venin (Gao, Dalziel, Tanzi, & Zhu, 2018). King et ses collaborateurs
388 ont également démontré que la toxicité du venin de *Vespula vulgaris* (Hymenoptera : Vespidae)
389 requérait l'action synergique de deux composés venimeux, la phospholipase A₁ et la toxine
390 peptidique Mastoparane (T. P. King, Jim, & Wittkowski, 2003).



391

392 **Figure 4 :** Cibles pharmacologiques de toxines protéiques isolées des venins. (a) Les numéros
 393 représentent les récepteurs, ciblés par différentes toxines peptidiques, impliqués dans le système de
 394 coagulation sanguine. (b) Les numéros représentent les récepteurs du système nerveux ciblés par des
 395 toxines peptidiques. Schéma tiré de (Casewell et al., 2013).

396

397 *1.4.2. Évolution des toxines protéiques de venins – Mécanismes impliqués*

398 D'un point de vue évolutif, les venins sont des outils écologiques multifonctionnels avec
399 de multiples toxines assurant une ou plusieurs fonctions et formant ensemble un trait adaptatif
400 très complexe (Schendel et al., 2019). En effet, de par leurs multiples apparitions au sein du
401 règne animal, les venins procurent un avantage adaptatif à l'organisme qui les produit, quel que
402 soit le contexte de leur utilisation (Casewell et al., 2013; Rollard et al., 2015; Schendel et al.,
403 2019). Ainsi, les toxines composant les venins sont sélectionnées et optimisées au cours de
404 l'évolution.

405 Le modèle classique expliquant l'origine des toxines implique des événements de
406 recrutements de gènes codant des protéines impliquées dans des processus physiologiques clés
407 de l'organisme. Les gènes les codant seraient ainsi sélectivement exprimés dans les glandes à
408 venin (B. Fry, 2005; B. G. Fry et al., 2009; Zancolli & Casewell, 2019).

409 Selon Fry (2009), les protéines recrutées de manière convergente en tant que toxines
410 partagent plusieurs caractéristiques :

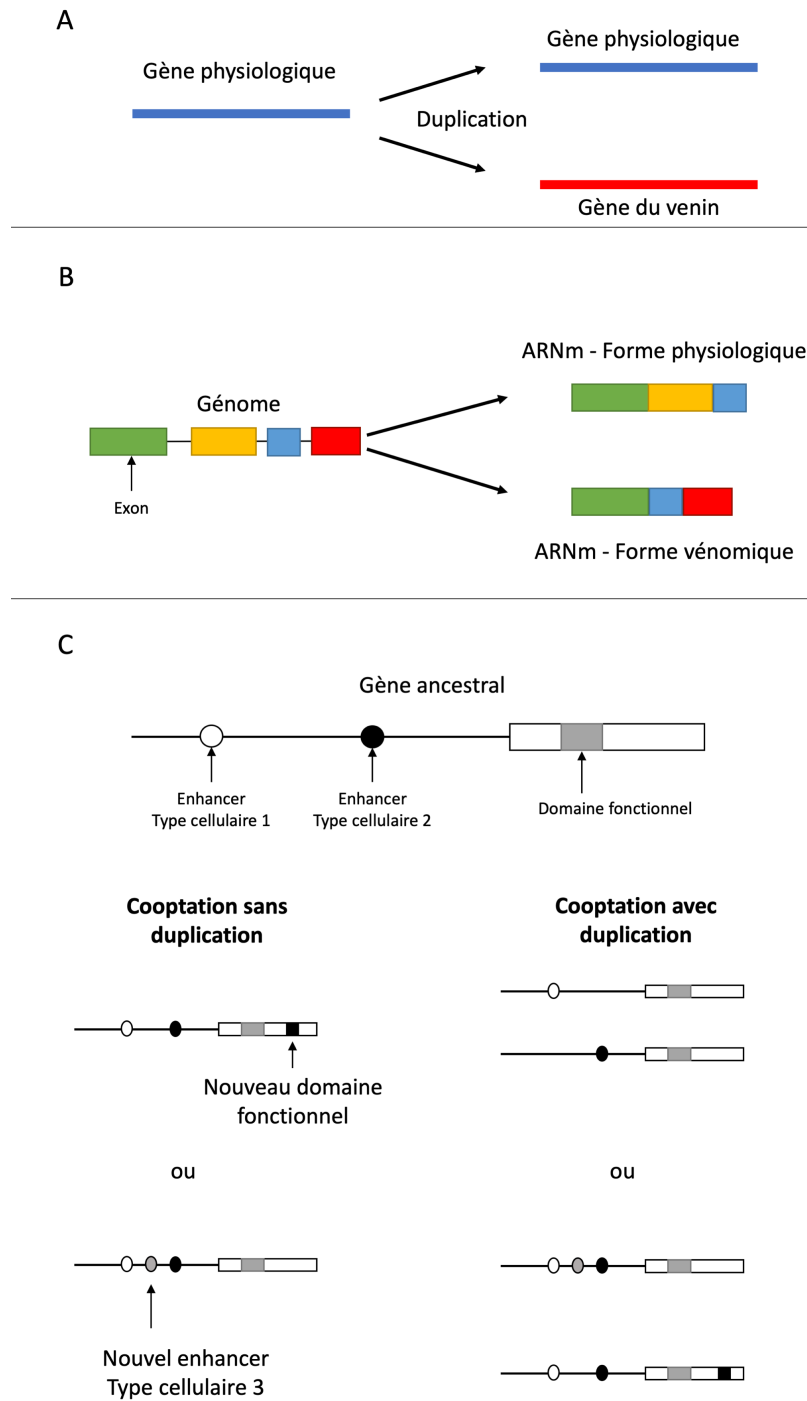
- 411 i. Elles proviennent d'une protéine ancestrale sécrétée.
- 412 ii. Elles appartiennent à des familles de protéines rencontrées dans différentes fonctions
413 métaboliques. Ainsi, les homologues effectuent des réactions biochimiques identiques
414 ou très similaires.
- 415 iii. Leur structure tertiaire est stable, notamment grâce à la présence de plusieurs ponts
416 disulfures.

417 Dans la plupart des cas, les gènes qui les codent ont été dupliqués de multiples fois pour
418 obtenir des familles multigéniques. Ces phénomènes de duplication sont généralement suivis
419 de mutations entraînant des néofonctionnalisations et des délétions de certaines copies, ou
420 rendant ces copies non-fonctionnelles (i.e. pseudogènes). La nouvelle famille multigénique de
421 toxines préserve souvent la structure moléculaire de la protéine ancestrale (y compris la
422 structure tertiaire). En revanche, les résidus fonctionnels en dehors de la structure moléculaire
423 de base sont modifiés, permettant l'acquisition de nouvelles activités (B. G. Fry et al., 2009).
424 Un exemple typique de famille multigénique parmi les plus étudiées est celui des
425 phospholipases A₂ (PLA₂s) retrouvées dans les venins de serpents. Les PLA₂s de type I_B
426 proviennent du pancréas tandis que celles du type II_A proviennent du liquide synovial (B. Fry,
427 2005). L'activité biologique des protéines ancestrales réside dans l'induction de la libération de
428 l'acide arachidonique alors que celle de leurs dérivés consiste en des propriétés neurotoxiques
429 et antiplaquettaires (B. Fry, 2005). La structure des gènes de PLA₂s du groupe II a été
430 particulièrement étudiée chez des espèces de Viperidae, révélant une conservation des régions

431 non-codantes ainsi qu'une diversification des exons codant les protéines matures (Kordis &
432 Gubensek, 2000). Le haut degré de conservation des introns, des régions flanquantes et non-
433 traduites ainsi que des régions « signal » indiquent que ces gènes sont issus des duplications et
434 divergences d'un seul gène ancestral (Kordis & Gubensek, 2000).

435 Les enzymes PLA₂s ont également été recrutées dans le venin de nombreux organismes et
436 présentent une variété de nouvelles activités, de *Apis dorsata* (Hymenoptera : Apidae) à
437 *Adamsia carciniopados* (Actiniaria : Hormathiidae) (B. Fry, 2005). Ce phénomène d'évolution
438 convergente peut être observé d'un point de vue structural mais également fonctionnel. Par
439 exemple, les toxines de type Kunitz issues du venin d'anémone de mer possèdent la même
440 activité inhibitrice sur les canaux potassiques que les peptides du même nom dérivés des venins
441 des serpents du genre *Dendroaspis* (Schweitz et al., 1995). Ces toxines ont été recrutées de
442 manière indépendante dans les venins de cnidaires, de cônes marins, d'insectes, de scorpions,
443 de reptiles et d'araignées et constituent donc un exemple d'évolution convergente en termes de
444 cibles biologiques (Casewell et al., 2013). Un autre exemple est celui des défensines, des
445 peptides riches en cystéines avec des propriétés antimicrobiennes ayant été identifiées chez la
446 plupart des organismes et faisant partie intégrante du système immunitaire (Silva, Gonçalves,
447 & Santos, 2014). Ces molécules ont été classées selon l'arrangement tridimensionnel des ponts
448 disulfures en α -, β -, θ - et α/β -défensines. Des peptides de structure similaire à celle des β -
449 défensines ont été retrouvés dans les venins d'ornithorynque, de scorpions et de reptiles (B. G.
450 Fry et al., 2009). Whittington et ses collaborateurs (2008) ont d'ailleurs montré que les peptides
451 de venin d'ornithorynque de type défensine avaient évolué à partir des β -défensines impliquées
452 dans l'immunité innée. En effet, les gènes codant ces peptides de venin sont situés sur le même
453 chromosome que leurs paralogues impliqués dans l'immunité innée. De plus, leurs promoteurs
454 présentent 89 à 94% d'identité et les arrangements des cystéines des peptides matures sont
455 similaires (Whittington et al., 2008).

456 D'autres mécanismes, tels que le phénomène de cooptation de gènes à copie unique, où une
457 nouvelle utilité émerge pour un gène préexistant ou nouvellement recruté, sans forcément de
458 duplication suivie de néofonctionnalisation (Martinson, Mrinalini, Kelkar, Chang, & Werren,
459 2017), ou l'épissage alternatif contribueraient également à l'ensemble des mécanismes
460 impliqués dans l'acquisition de gènes codant pour des toxines (Zancolli & Casewell, 2019)
461 (Figure 5). Tous ces mécanismes ont généré la grande diversité génétique et fonctionnelle
462 observée aujourd'hui dans les venins.



463

464 **Figure 5** : Exemples de mécanismes impliqués dans l'évolution des gènes codant les toxines. (A) Duplication
 465 d'un gène impliqué dans une fonction physiologique et néofonctionnalisation du gène dupliqué. (B) Épissage
 466 alternatif résultant en des ARN messagers codant respectivement pour une protéine physiologique et une
 467 toxine. (C) Cooptation de gènes avec ou sans duplication. Des gènes peuvent être cooptés pour générer des
 468 nouveautés physiologiques via (i) le changement de leurs motifs de régulation, (ii) le changement de
 469 fonctions des protéines qu'ils codent ou (iii) les deux.

470

471 *1.5. Intérêt scientifique de l'étude des venins*

472 L'étude des venins a tout d'abord été motivée par deux aspects principaux : l'un concernant
473 la conception d'antivenins, l'autre concernant la recherche de nouvelles molécules bioactives
474 d'intérêts en santé humaine (anti-cancéreux, antibactériens, immunothérapie) et en agronomie
475 (bioinsecticides).

476 Dans le cadre de la recherche pour le développement d'antivenins, la majorité des études a
477 été réalisée sur des espèces constituant un problème de santé publique, comme les serpents, les
478 scorpions, les cônes ou encore les araignées. En effet, une meilleure connaissance générale des
479 venins est essentielle au développement de protections contre les envenimations (Chippaux et
480 al., 1991; Ménez et al., 2006). Ce premier aspect de la recherche sur les animaux venimeux a
481 particulièrement concerné les serpents et les scorpions. Encore aujourd'hui, les cas
482 d'envenimations constituent un problème de santé publique d'envergure avec plus de 100 000
483 morsures de serpents au niveau mondial chaque année (Knudsen et al., 2019). Les
484 envenimations par les scorpions représentent aujourd'hui entre 30-200 cas / 100 000 habitants
485 rien que pour le genre *Tityus* en Amérique du Sud, avec des taux de mortalités relativement
486 élevés (Borges et al., 2020). Les araignées ne font pas exception, même si les cas
487 d'envenimations sont plus rares et localisés avec par exemple 3 000 à 5 000 cas recensés chaque
488 année en Australie (Matsumura et al., 2018). La nécessité de production d'antivenins s'est fait
489 ressentir essentiellement de par les actions toxiques (neurotoxiques, nécrolytiques,
490 hémolytiques, inflammatoires, etc.) de ces venins sur les populations locales (Borges et al.,
491 2020; Judge, Henry, Mirtschin, Jelinek, & Wilce, 2006; Knudsen et al., 2019; D. J. Williams,
492 Habib, & Warrell, 2018). Néanmoins, la production d'antivenins suffisamment efficaces n'est
493 pas un processus acquis et reste toujours d'actualité, du fait de la haute variabilité de
494 composition des venins, que ce soit en fonction de la zone géographique, de l'espèce voire de
495 l'individu (Chippaux et al., 1991).

496 Dans le cadre de la recherche de nouvelles molécules bioactives, une attention particulière
497 a été portée aux toxines peptidiques (Pennington et al., 2018). Elles présentent en effet
498 l'avantage d'être peu rémanentes dans l'environnement tout en restant efficaces sur une large
499 gamme de cibles allant des canaux ioniques aux pathogènes. De plus, elles agissent en général
500 à très faible dose, à l'échelle du nano- voire du pico-molaire, concentration active reflétant
501 directement l'affinité et la sélectivité de la toxine pour sa cible pharmacologique (Wulff,
502 Christophersen, Colussi, Chandy, & Yarov-Yarovoy, 2019). Actuellement, sept médicaments
503 dérivés de toxines peptidiques de venins ont été approuvés par la « Food and Drugs
504 Administration » (FDA U.S.) et sont sur le marché (Tableau I). Les deux exemples les plus

505 connus sont le Captopril® et le Prialt® dérivés respectivement de toxines peptidiques du
 506 *Bothrops jararaca* et de *Conus magus*, et agissant contre l'hypertension artérielle et en tant
 507 qu'antidouleur (G. F. King, 2011) (Tableau I). La majorité des médicaments commercialisés
 508 agissent au niveau du système circulatoire avec deux d'entre eux commercialisés en tant
 509 qu'anticoagulants (Aggrastat® et Integrilin®, développés à partir de désintégrines de venins de
 510 serpents) ou en tant qu'agent anti-plaquettaire (Refludan® et Angiomax®, développés à partir
 511 de l'hirudine de la sangsue *Hirudo medicinalis*). Le dernier médicament dérivé de toxines
 512 peptidiques de venins est actuellement utilisé dans le traitement du diabète de type II en tant
 513 qu'hypoglycémiant et a été développé à partir d'un peptide issu du venin d'*Heloderma*
 514 *suspectum* (i.e. monstre de Gila) (Pennington et al., 2018) (Tableau I).

515

516 **Tableau I** : Médicaments dérivés des venins approuvés par la FDA (Food Drug Administration), d'après
 517 Pennington et al. (Pennington et al., 2018). ACE : enzyme de conversion de l'angiotensine ; Ca_v : canaux
 518 calciques voltages-dépendants.

Médicament	Organisme	Espèce	Molécule	Effet
Captopril®	Serpent	<i>Bothrops jararaca</i>	Bradykinin-potentiating Peptide	Inhibiteur de l'ACE
Aggrastat®	Serpent	<i>Echis carinatus</i>	Désintégrine	Anticoagulant
Integrilin®	Serpent	<i>Sistrurus miliarius</i>	Désintégrine	Anticoagulant
Refludan®	Sangsue	<i>Hirudo medicinalis</i>	Hirudine	Inhibiteur de la thrombine
Angiomax®	Sangsue	<i>Hirudo medicinalis</i>	Hirudine	Inhibiteur de la thrombine
Prialt®	Cône marin	<i>Conus magus</i>	ω-conotoxine MVII-A	Bloqueur de canaux Ca _v neuronaux (antidouleur)
Byetta®	Monstre de Gila	<i>Heloderma suspectum</i>	Glucagon-Like Peptide	Hypoglycémiant

519

520 D'autres peptides sont actuellement en phases d'essais cliniques. Les chlorotoxines,
 521 issues du venin du scorpion *Leiurus quinquestriatus hebraeus*, initialement connues pour
 522 inhiber les canaux chlorures, possèdent également une haute affinité pour les cellules tumorales.
 523 Elles sont actuellement en court de développement pour le traitement radiochimique des
 524 cellules malignes ainsi qu'en tant qu'outils de diagnostics *in vivo* des cancers par imagerie
 525 (Pennington et al., 2018). Dans le même ordre d'idée, le peptide SOR-C13 issu du venin de la
 526 musaraigne *Blarina brevicauda* bloque l'influx de calcium activant les récepteurs TRPV6 (pour
 527 Transient Receptor Potential Vanilloid calcium channel subtype 6), particulièrement impliqués
 528 dans la tumorigenèse. Les premiers tests sur des cellules cancéreuses ovariennes ont montré
 529 que la partie C-terminale de ce peptide inhibe ce récepteur avec une IC₅₀ (Inhibitory
 530 Concentration) de 14 nmol.L⁻¹ et entraîne ainsi la mort cellulaire (Pennington et al., 2018). Le

531 peptide ShK issu du venin de l'anémone de mer *Stichodactyla helianthus* inhibe avec une haute
532 affinité (10 pmol.L⁻¹) les canaux K_v1.3, directement impliqués dans l'activation d'un sous-
533 ensemble de lymphocytes T connus comme étant effecteurs de la mémoire cellulaire
534 (Pennington et al., 2018). Il en résulte une diminution de la prolifération cellulaire et une
535 suppression de la production d'interleukines de type 2 (Beeton et al., 2005; Jimenez et al.,
536 2017), effets pouvant faire de ce peptide une molécule permettant de traiter les maladies auto-
537 immunes.

538 Plusieurs autres toxines de venins sont en phase d'essais précliniques, notamment en
539 tant qu'antidouleurs. Ainsi, la protoxine-II, isolée du venin de la tarentule *Thrixopelma pruriens*
540 inhibe les canaux Na_v1.7 avec une IC₅₀ de 300 pM, et les α -conotoxines sont des antagonistes
541 des récepteurs nicotiques $\alpha 9\alpha 10$ nAChRs. Ces peptides sont actuellement développés pour
542 la prévention des douleurs neuropathiques chroniques induites par la chimiothérapie
543 (Pennington et al., 2018).

544 D'autres peptides de venins, qualifiés « d'antimicrobiens », exercent une action
545 cytotoxique directe ou cytostatique sur un panel de microorganismes (Mahlapuu, Håkansson,
546 Ringstad, & Björn, 2016; Zasloff, 2002). A titre d'exemple, Coutinho das Neves *et al.* (2019)
547 ont récemment montré l'action bactériostatique de peptides de venins de guêpes et de scorpions
548 contre *Acinetobacter baumannii*, une bactérie Gram négative résistante aux antibiotiques et
549 responsable d'infections nosocomiales en milieu hospitalier (Coutinho das Neves, Mortari,
550 Ferroni Schwartz, Kipnis, & Junqueira-Kipnis, 2019). Bon nombre d'études démontrent
551 l'action antimicrobienne de toxines de venin. Cependant, aucune n'est encore entrée en phase
552 de développement clinique pour le traitement d'infections bactériennes. En effet, malgré
553 l'efficacité *in vitro* des peptides antimicrobiens, leur usage est limité en raison de leur faible
554 biodisponibilité *in vivo* due à leur instabilité, leur cytotoxicité ainsi que leur coût de production
555 (Moreno & Giralt, 2015).

556 En dehors du domaine thérapeutique, les toxines de venins peuvent également servir
557 pour des applications dans le domaine agronomique. En effet, les peptides dérivés de venins de
558 prédateurs d'insectes entomophages (e.g. scorpions, araignées) font l'objet d'un intérêt de plus
559 en plus grand pour la recherche d'alternatives naturelles et sélectives aux pesticides. Le peptide
560 ω -HXTX-Hv2a, isolé à partir du venin de *Hyadronyche versuta*, possède par exemple une
561 activité inhibitrice élevée de canaux calciques voltage-dépendant d'invertébrés. A partir de ce
562 peptide, la compagnie Vestaron Corporation a développé un analogue peptidique inhibant les
563 canaux calciques voltage-dépendant et potassiques calcium-dépendant des insectes, nommé
564 ω/κ -HXTX-Hv1a. Il présente ainsi une forte activité insecticide contre une large gamme

565 d'insectes ravageurs de cultures, tout en ayant une toxicité moindre pour les ouvrières *Apis*
566 *mellifera* par ingestion ($DL_{50} > 100\mu\text{g}/\text{abeille}$) (G. F. King, 2019; Nakasu et al., 2014; Powell
567 et al., 2020). Depuis 2017, cette toxine modifiée est commercialisée comme agent de
568 biocontrôle (Spear-T[®]) (Corporation, n.d.).

569 Certaines toxines constituent également des outils pour l'étude fondamentale de systèmes
570 physiologiques complexes (nerveux, cardiovasculaire, hormonal, immunitaire) (Ménez et al.,
571 2006). En effet, de par leurs actions, les toxines peuvent jouer un rôle crucial dans la
572 compréhension de ces processus physiologiques. Par exemple, la tétrodotoxine est couramment
573 utilisée pour l'étude des canaux sodiques voltage-dépendant (Clark et al., 2019). D'autres
574 toxines, comme celles formant des pores membranaires, ont également aidé à élucider la
575 structure et la fonction des membranes cellulaires, des protéines membranaires et des voies de
576 signalisations transmembranaires (Clark et al., 2019). La melittine issues du venin d'*Apis*
577 *mellifera* et les actinoporines des venins d'anémones de mer ont notamment permis de
578 déterminer des mécanismes de rupture des membranes cellulaires (Akbari et al., 2018; Rojko,
579 Dalla Serra, Maček, & Anderluh, 2016). Il est également essentiel de noter que les organismes
580 venimeux sont des modèles d'intérêt pour l'étude de points clés relatifs à la biologie évolutive
581 et moléculaire. Les toxines de venin constituent en effet des modèles de choix dans l'étude des
582 processus évolutifs impliquant l'acquisition de nouvelles fonctions génétiques ou les
583 dynamiques de recrutements convergents de protéines (Casewell et al., 2013; Zancolli &
584 Casewell, 2019).

585 **2. Méthodes d'étude de la composition des venins : focus sur les venins peptidiques**

586 Un des plus grands défis des chercheurs travaillant sur les venins reste la caractérisation en
587 termes de séquences et de structures des différentes toxines les composant. D'un point de vue
588 historique, les premières études s'effectuaient majoritairement sur des toxines isolées *via* des
589 méthodes de biochimie, telle que l'électrophorèse et la chromatographie liquide haute
590 performance (en anglais High Performance Liquid Chromatography – HPLC). Les structures
591 primaires, secondaires et tertiaires des toxines ainsi isolées et purifiées étaient ensuite
592 caractérisées par différentes méthodes telles que la spectrométrie de masse, la dégradation
593 d'Edman, la Résonance Magnétique Nucléaire (RMN) et la cristallographie aux rayons X
594 (Shin, Lee, & Lee, 2008; Wilson & Daly, 2018; Yee et al., 2005). L'inconvénient de ces trois
595 dernières méthodes est qu'elles requièrent une quantité importante de toxine purifiée et donc
596 de venin brut de départ (Himaya & Lewis, 2018). Elles sont peu adaptées pour l'étude du venin

597 d'espèces de petites tailles, pour lesquelles il est difficile de récolter une grande quantité de
598 venin.

599 L'étude de la composition des venins de ce type d'organismes est désormais un domaine
600 de recherche en pleine expansion grâce à une méthodologie intégrative appelée « vénomique ».
601 Elle consiste en l'étude de la composition globale des venins, par l'utilisation à large échelle de
602 différentes méthodes « -omiques » (e.g. génomique, transcriptomique, protéomique) et le
603 croisement des données obtenues *via* des outils bioinformatiques (Wilson & Daly, 2018). La
604 combinaison de ces méthodes, couplées aux grandes avancées technologiques dans chacune
605 d'entre elles permet d'obtenir rapidement une vision complète de la composition des venins,
606 même à partir de faibles quantités.

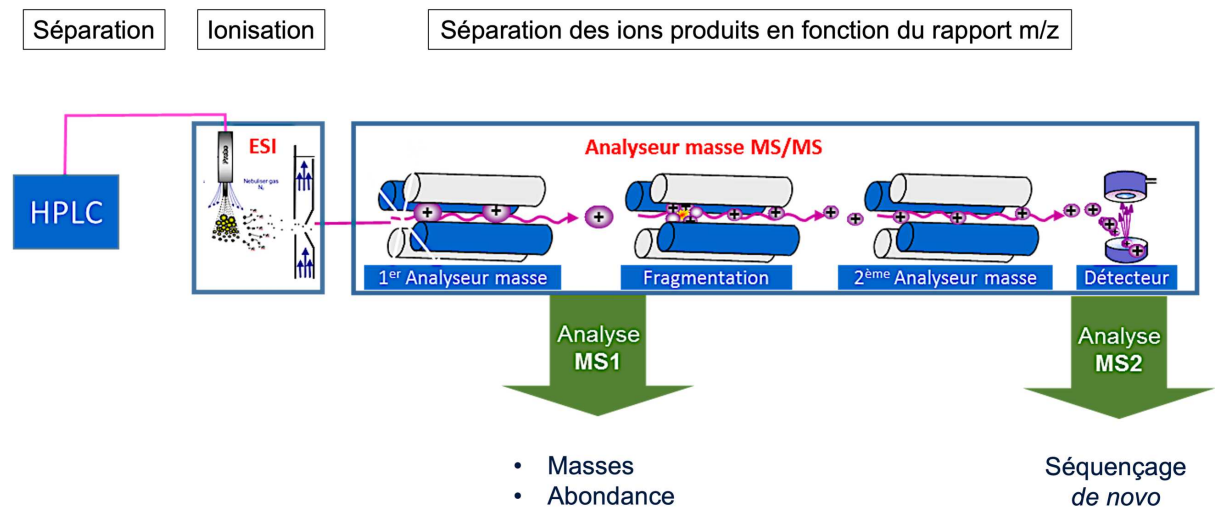
607

608 *2.1. Protéomique appliquée à la vénomique*

609 La fraction protéique est généralement la fraction dominante des venins d'arthropodes
610 (Daly & Wilson, 2018; dos Santos-Pinto, Perez-Riverol, Musacchio-Lasa, & Palma, 2018;
611 Senji Laxme et al., 2019; Walker, Robinson, et al., 2018). Les approches protéomiques sont
612 donc des outils indispensables à leur étude. Elles permettent en effet d'analyser les toxines
613 protéiques sécrétées dans le venin et constituent donc une approche complémentaire aux
614 approches transcriptomiques et génomiques, qui apportent une vision globale des produits des
615 gènes, qu'ils soient sécrétés ou non.

616 Les approches majoritairement utilisées sont les méthodes associées à la spectrométrie de
617 masse (MS) de par leur rapidité d'analyse, leurs hautes sensibilité et résolution. Ces approches
618 nécessitent désormais une moindre quantité de venin (e.g. quelques nanogrammes) et
619 permettent surtout une caractérisation rapide et complète d'un peptidome pour un coût réduit
620 (e.g. 150€ par échantillon) (Wilson & Daly, 2018). La MS est une technique d'analyse
621 permettant de détecter et d'identifier des molécules par mesure de leur masse, ainsi que de
622 caractériser leur structure chimique. Son principe réside dans la séparation en phase gazeuse de
623 molécules chargées (ions) en fonction de leur rapport masse/charge (m/z). Les venins étant par
624 définition des mélanges de toxines, il est nécessaire de séparer ces toxines par une
625 chromatographie liquide à haute performance (HPLC) avant le passage dans le spectromètre de
626 masse (ESI-LC-MS). Ce premier passage permettra d'obtenir l'ensemble des masses présentes
627 dans l'échantillon injecté ainsi que leur abondance. Par la suite, il est également possible de
628 déterminer la structure primaire des toxines peptidiques (i.e. séquences d'acides aminés) *via*
629 des méthodes de fragmentation suivies d'une seconde analyse MS, appelée séquençage *de novo*
630 ou spectrométrie de masse en tandem (couramment nommée MS/MS) (Figure 6). D'autres

631 stratégies liées à la MS peuvent aussi être employées à large échelle pour déterminer la structure
 632 primaire des toxines, telles que la digestion enzymatique d'un venin suivie d'une LC-MS/MS.
 633 Ces méthodes nécessitent néanmoins l'utilisation de données transcriptomiques pour confirmer
 634 les séquences des toxines, en levant les incertitudes liées aux amino acides isobares, à certaines
 635 modifications post-transcriptionnelles, voire à quelques modifications artéfactuelles.



636
 637 **Figure 6 :** Approche de spectrométrie de masse dans le cadre de la détermination de la composition des
 638 venins.

639 La MS permet également d'identifier les modifications post-traductionnelles (PTMs pour
 640 Post-Translational Modifications), telles que les glycosylations *via* le choix de différentes
 641 techniques de fragmentation (Mechref, 2012), ou encore en procédant à une
 642 réduction/alkylation du venin brut ou à une digestion enzymatique pour identifier la présence
 643 de ponts disulfures et leurs connectivités (Walker, Robinson, Hamilton, Undheim, & King,
 644 2020). D'autres techniques nécessitent un travail sur les toxines isolées et purifiées. La
 645 dégradation d'Edman permet par exemple de déterminer la structure primaire d'une toxine. La
 646 RMN fournit des informations sur la connectivité des ponts disulfures et plus généralement sur
 647 la structure tridimensionnelle d'une toxine (Himaya & Lewis, 2018; Shin et al., 2008; Yee et
 648 al., 2005).

649
 650 *2.2. Transcriptomique appliquée à la vénomique*

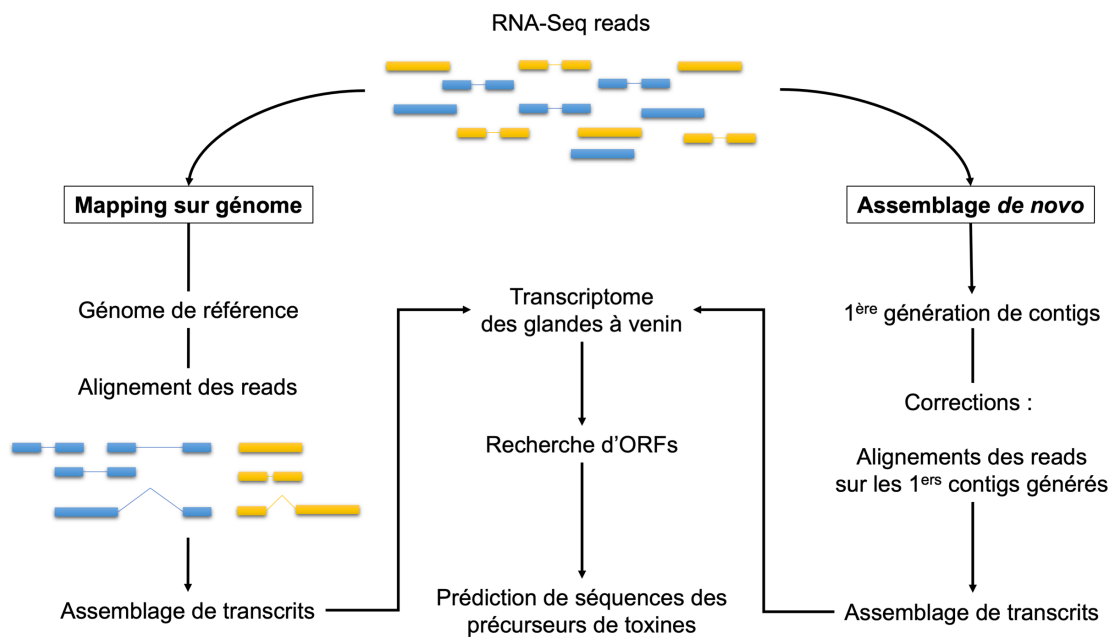
651 L'étude du transcriptome permet d'accéder à la séquence des transcrits matures, au patron
 652 d'expression spatial et temporel des gènes ainsi qu'à leur niveau d'expression. Les objectifs

653 principaux des méthodes transcriptomiques sont (i) de construire un catalogue de tous les
654 transcrits pour un organisme ou un tissu donné (i.e. transcriptome de référence), (ii) d'établir la
655 structure et l'organisation de ces transcrits (e.g. extrémités 5'-3' ; prédiction de leurs six cadres
656 de lecture (ORFs pour Open Reading Frames) ; détermination des motifs d'épissage si génome
657 de référence disponible), ainsi que (iii) de quantifier tout changement de niveau d'expression
658 des transcrits sous différentes conditions (i.e. RNA-seq en différentiel, si génome ou
659 éventuellement transcriptome de références disponibles) (Wilson & Daly, 2018).

660 Actuellement, la méthode transcriptomique la plus courante appliquée à la vénomique fait
661 tout d'abord intervenir une méthode de séquençage qui implique les plateformes de séquençage
662 NGS « short-reads » (Cf. partie 2.3) et appelée RNAseq. Néanmoins, le séquençage ne
663 s'effectue pas directement sur le brin source d'ARN. En effet, de l'ADN complémentaire
664 (ADNc) est généré par transcription inverse. L'ADNc est ensuite préparé en librairie via une
665 amplification par réaction de polymérisation en chaîne (PCR). Ainsi, les limites du RNAseq
666 sont celles rencontrées pour les séquençages NGS, auxquelles s'ajoute le taux d'erreur de la
667 transcription inverse. De plus, les réactions sont directement gouvernées par la processivité de
668 la Taq Polymérase impliquée dans la PCR pour la préparation de la librairie, c'est-à-dire la
669 capacité de l'enzyme à catalyser des réactions successives sur une même molécule sans la
670 relâcher (Wilson & Daly, 2018). Il existe également d'autres biais dus aux méthodes
671 d'assemblage des données brutes (i.e. reads) en séquences nucléotidiques continues et
672 ordonnées (i.e. contigs), pour lesquelles on distingue différentes approches : *de novo* ou *via* un
673 alignement sur un génome de référence (i.e. mapping). L'assemblage *via* mapping sur génome
674 de référence s'effectue classiquement pour des espèces modèles et attribue les reads
675 directement aux gènes référencés. *A contrario*, l'assemblage *de novo* d'un transcriptome ne
676 nécessite pas obligatoirement d'information « externe », bien qu'il y ait possibilité d'utiliser
677 des bases de données de protéines référencées de l'organisme modèle ou d'un organisme
678 proche. Ce type d'assemblage est en conséquence utilisé pour les espèces pour lesquelles il
679 n'existe pas de génome de référence (Figure 7), et représente d'ailleurs le point de départ dans
680 la construction d'un transcriptome de référence. Le RNAseq est actuellement le plus utilisé, du
681 fait de son coût (environ 350 € / échantillon) et de son accessibilité pour l'étude d'espèces
682 venimeuses non modèles. L'assemblage *de novo* s'effectue en plusieurs étapes impliquant une
683 normalisation des reads suivie de leur fusion en contigs. De nombreuses étapes de corrections
684 se succèdent, incluant notamment l'alignement des reads sur les contigs précédemment générés
685 afin de corriger les séquences. Ce type d'assemblage présente donc des biais dus à la variabilité

686 des données brutes (Cabau et al., 2017; Korpelainen, Tuimala, Somervuo, Huss, & Wong,
687 2015).

688 Dans le contexte de la vénomique, l'annotation des transcriptomes aide à identifier les
689 protéines exprimées dans les glandes à venin. Les études transcriptomiques des glandes à venins
690 ne donnent néanmoins accès qu'aux séquences des transcrits et à leurs ORFs potentiels. La
691 recherche d'ORFs sur les transcrits permettra ensuite de prédire les séquences des précurseurs
692 de toxines peptidiques (Figure 7). La qualité de cette prédiction dépend directement de celle de
693 l'assemblage, et peut diminuer notamment dans le cas des séquences répétées, et des contigs
694 trop courts. Ces derniers peuvent en effet contenir des ORFs incomplets. Les méthodes
695 transcriptomiques seules ne permettent pas de prédire toutes les PTMs que subissent les
696 précurseurs. Il est ainsi nécessaire de combiner les données transcriptomiques avec les données
697 protéomiques.



698
699 **Figure 7** : Méthodes de recherche des transcrits codant les toxines peptidiques à partir de données de RNA-
700 Seq. Les données brutes peuvent être assemblées selon deux méthodes : un alignement sur un génome de
701 référence ou *de novo*. La recherche d'Open Reading Frames (ORFs) sur les transcrits permet de prédire les
702 séquences des précurseurs de toxines.

703 704 2.3. Génomique appliquée à la vénomique

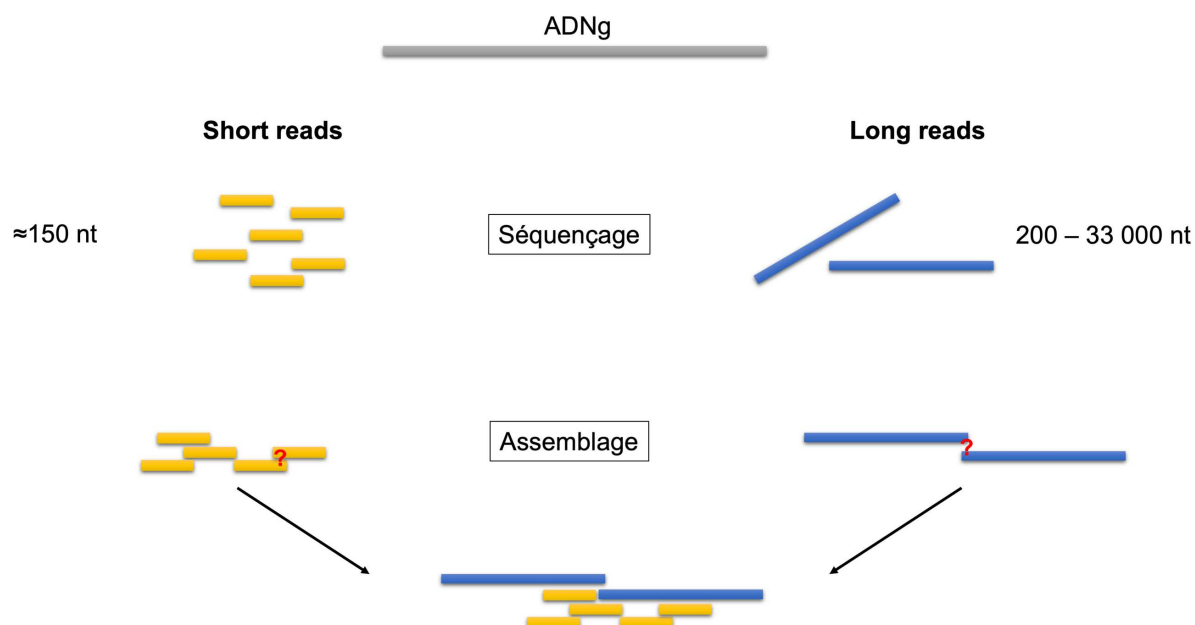
705 Le domaine de la génomique n'a fait son entrée que très récemment dans la vénomique,
706 notamment en raison du coût de séquençage (Himaya & Lewis, 2018) (e.g. environ 7 000 €
707 pour un organisme eucaryote). La connaissance du génome peut en effet aider à l'analyse de la

708 composition des venins peptidiques puisque ces composés sont des produits directs des gènes
709 (Babb et al., 2017). Le génome d'une espèce venimeuse contient les informations codantes et
710 non-codantes, pour chaque toxine protéique et peptidique qu'elle soit exprimée ou non. Son
711 séquençage permet donc de confirmer les séquences des toxines, de détecter des variants dus à
712 des épissages alternatifs et peut également aider à l'assemblage des données de RNAseq. Les
713 données génomiques donnent de plus des informations concernant les gènes codant les toxines
714 telles que leur structure, leur régulation, leur positionnement sur les chromosomes, permettant
715 de déterminer les mécanismes impliqués dans la diversification des toxines de venins.

716 Actuellement, les séquençages sont réalisés majoritairement *via* l'utilisation des
717 technologies de « nouvelle-génération » (NGS, pour Next-Generation Sequencing). Ces
718 technologies sont subdivisées en deux catégories : les séquençages « short-reads » et « long-
719 reads » (Wilson & Daly, 2018) (Figure 8).

720 Le séquençage « short-reads » fait intervenir les plateformes Illumina (Hi-Seq, Mi-Seq,
721 etc.) et présente un taux d'erreur bas (< 5 %) en plus de fournir un génotypage précis dans les
722 régions contenant des séquences non répétées. Il ne permet cependant pas l'assemblage *de novo*
723 de contigs mitoyens, limitant ainsi la possibilité de reconstructions de séquences répétées ainsi
724 que la détection de variations structurales complexes (Illumina Inc., 2011; Jain et al., 2018).

725 Le séquençage « long-reads » fait intervenir les plateformes Pacific Biosciences (PacBio)
726 ou Oxford (Nanopore) avec des séquenceurs à molécules unique. Ces technologies fournissent
727 des lectures plus longues et peuvent aider à surmonter les limitations associées aux séquences
728 répétées, aux génomes de grande taille, et améliorent le séquençage ainsi que l'assemblage *de*
729 *novo* d'un génome. Elles présente cependant un taux d'erreur élevé, 10-15% pour Pacific
730 Biosciences et 5-20% pour Oxford Nanopore, et nécessitent des données de séquençage « short-
731 reads » complémentaires pour procéder aux assemblages *de novo* de génomes de référence (Jain
732 et al., 2018; Sedlazeck et al., 2018) (Figure 8).



733
 734 **Figure 8** : Complémentarité des séquençages short et long reads dans l'assemblage *de novo* d'un génome de
 735 référence. La combinaison des deux techniques de séquençage aide à lever les ambiguïtés, notamment celles
 736 liées aux séquences répétées.

737 Les données génomiques ne sont cependant pas suffisantes pour déterminer les spécificités
 738 et les dynamiques d'expression au niveau cellulaire et tissulaire ainsi que pour définir
 739 précisément les bornes introns/exons. Pour ceci, les approches transcriptomiques sont requises
 740 en complément (Sunagar, Morgenstern, Reitzel, & Moran, 2016).

741
 742 *2.4. Traitement intégré des données issues des différentes approches « omiques » pour*
 743 *l'étude de la composition des venins*

744 Chacune des différentes approches présentées permet d'apporter des informations qui
 745 restent fragmentaires si elles sont appréhendées de manières isolées. La combinaison des
 746 données issues des différentes approches « omiques » permet alors de préciser ces données et
 747 d'obtenir une vision exhaustive de la composition des venins, du gène au peptide mature.

748 Les outils de bioinformatique appliqués à la vénomique permettent de croiser l'ensemble
 749 des données obtenues en génomique, transcriptomique et protéomique afin de fournir
 750 rapidement une image complète du venin étudié en termes de composition, de dynamique
 751 d'expression ainsi que d'évolution des gènes codant pour les toxines (Figure 9).

752 Dans le cas d'un « workflow » classique d'étude de la composition d'un venin à large
 753 échelle, l'analyse en LC-MS/MS d'un venin brut, ou ayant subi un traitement tel qu'une

754 digestion enzymatique ou encore une réduction/alkylation, donne une liste de masses présentes
755 dans l'échantillon, leurs abondances ainsi que des séquences prédictives des toxines matures.

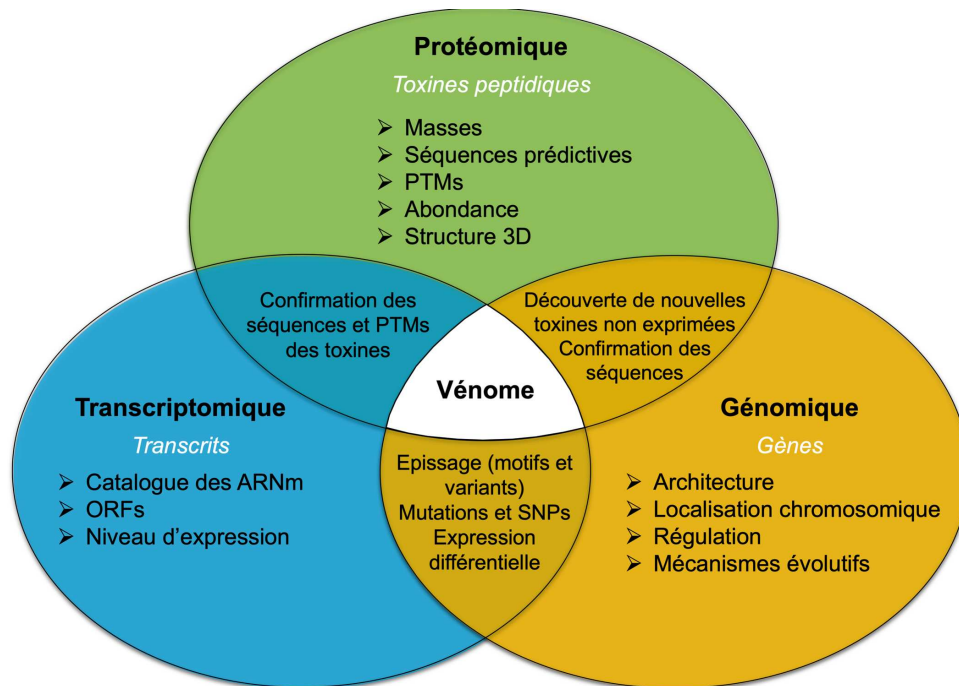
756 Une fois les données de RNAseq assemblées, les contigs sont traduits *in silico* en séquences
757 d'acides aminé au moyen d'outils bioinformatiques. La recherche d'ORFs peut ensuite
758 s'effectuer au moyen d'alignements (Camacho et al., 2009), soit avec des séquences de
759 précurseurs déjà connus, soit directement avec les séquences prédictives obtenues des données
760 de LC-MS/MS, manuellement ou de manière automatique à partir des données de RNA-Seq
761 traduites. A noter qu'il existe également des bases de données de référence pour les espèces
762 venimeuses modèles, par exemple ConoServer pour les cônes marins ou Arachnoserver pour
763 les araignées (Kaas, Yu, Jin, Dutertre, & Craik, 2012; Pineda et al., 2018).

764 Une fois les séquences prédictives des précurseurs de toxines obtenues, la prédiction de
765 séquence signal aide à définir le site de coupure pour la libération de la toxine mature. Cette
766 prédiction peut être effectuée grâce à différents serveurs tels que Phobius
767 (<http://phobius.sbc.su.se>) ou Signal IP (Almagro Armenteros et al., 2019). Il est ensuite
768 nécessaire de relier ces séquences prédictives de toxines matures aux masses et aux séquences
769 prédictives acquises en LC-MS/MS pour les confirmer (Figure 9).

770 Le croisement des données protéomiques et transcriptomiques est efficace dans la mesure
771 où les informations de chaque jeu de données sont utilisées pour corriger, confirmer et améliorer
772 les autres. A titre d'exemple, les données de RNA-Seq reflètent tous les transcrits présents dans
773 les glandes à venin au moment de l'extraction de départ et permettent donc (i) d'identifier des
774 toxines présentes en moindre quantité dans le venin et qui ont pu passer inaperçues dans les
775 premières analyses protéomiques, et (ii) d'estimer l'expression des gènes les codant. *A*
776 *contrario*, la majorité des transcrits dans ces données codent des protéines de la machinerie
777 cellulaire et ne portent pas d'ORF correspondant aux toxines sécrétées dans le venin. Ce sont
778 donc les données protéomiques qui permettent de confirmer la présence et les séquences des
779 toxines matures. Les PTMs, dont les clivages protéolytiques, ne sont également détectables
780 qu'à partir des données protéomiques, notamment *via* la comparaison de la masse observée en
781 MS et de la masse moléculaire calculée à partir de l'ORF d'un peptide présumé.

782 Une approche dite « protéotranscriptomique » accélère donc le décryptage d'un peptidome
783 de venin de manière complète et précise. Elle ne permet cependant pas l'étude de la régulation
784 des gènes codant pour les toxines, ni la confirmation des mécanismes évolutifs menant à leur
785 diversification. Pour ceci, les données génomiques sont nécessaires.

786



787

788 **Figure 9 :** Méthodologie intégrative « vénomique » et implication des différentes méthodes « -omiques »
 789 dans l'étude des toxines rentrant dans la composition des venins. PTMs : modifications post-traductionnelles.

790

791 *2.5. L'approche intégrative « vénomique » : une révolution pour la caractérisation*
 792 *exhaustive des venins d'espèces négligées*

793 La recherche en vénomique a été révolutionnée par l'utilisation des techniques
 794 transcriptomiques et protéomiques à haut-débit, ainsi que par la disponibilité croissante de
 795 données génomiques (Sunagar et al., 2016; Bjoern Marcus von Reumont, Campbell, & Jenner,
 796 2014; Wilson & Daly, 2018). La combinaison des méthodes décrites précédemment permet
 797 d'avoir rapidement une vision complète des venins, en termes de composition, de production
 798 et d'évolution. De plus, grâce à une sensibilité de plus en plus élevée des appareils, les grandes
 799 avancées technologiques dans la spectrométrie de masse et le séquençage NGS ont permis (i)
 800 d'identifier les séquences des toxines présentes en moindre quantité dans les venins, et passées
 801 inaperçues dans les précédentes analyses, (ii) l'émergence de nouvelles questions plus
 802 fondamentales quant à la diversification des venins au fil de l'évolution, et (iii) d'étudier les
 803 venins d'organismes jusque-là négligés de par leur taille et la faible quantité de venins
 804 récoltable, tels que les rémipèdes, les polychètes ou encore certains insectes (Rádis-Baptista &
 805 Konno, 2020; Björn M. von Reumont et al., 2014; Björn M. von Reumont, Undheim, Jauss, &
 806 Jenner, 2017; Björn Marcus Von Reumont, 2018; Walker, Dobson, et al., 2018; Walker,
 807 Hernández-vargas, et al., 2018). Drukewitz et ses collaborateurs (2018) ont par exemple
 808 appliqué une approche protéotranscriptomique pour étudier la composition des venins de deux

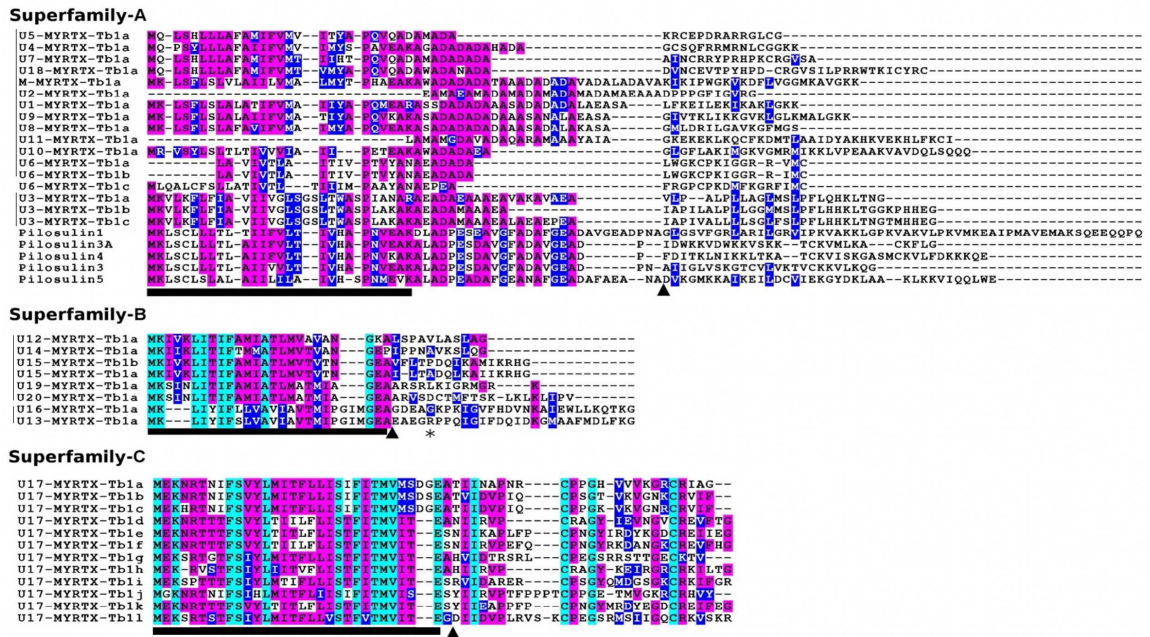
809 espèces européennes d'asilidés, *Eutolmus rufibarbis* et *Machimus arthriticus*, des diptères
810 prédateurs d'autres insectes. Leur étude a révélé que ces venins étaient majoritairement
811 composés de peptides et de protéines non enzymatiques (Drukewitz et al., 2018). Ces auteurs
812 ont ensuite appliqué la même approche protéotranscriptomique pour l'étude du venin d'une
813 autre Asilidae, *Dasyopogon diadema*, avant de procéder à une approche génomique comparative
814 avec d'autres groupes d'insectes (Drukewitz, Bokelmann, Undheim, & von Reumont, 2019).
815 Leurs résultats ont ainsi montré que le venin de *D. diadema* avait probablement évolué de
816 manière multimodale incluant (i) une néofonctionnalisation après une duplication de gènes, (ii)
817 une cooptation expression-dépendante de gènes, et (iii) des gènes spécifiques de la famille des
818 Asilidae avec des origines non déterminées actuellement (Drukewitz et al., 2019).

819 Les insectes constituent un taxon hyperdiversifié regroupant actuellement plus de 5
820 millions d'espèces. Depuis leur apparition, ils ont présenté un rayonnement adaptatif
821 spectaculaire et occupent aujourd'hui des niches écologiques variées. Étant donné leur diversité
822 écologique et spécifique, il n'est pas étonnant que les venins aient évolué de manière
823 indépendante à multiples reprises dans ce taxon (Walker, Robinson, et al., 2018). Peu d'insectes
824 venimeux ont néanmoins été étudiés en détails, excepté celui des hyménoptères aculéates
825 (abeilles, guêpes, fourmis) (Baumann, Dashevsky, Sunagar, & Fry, 2018; dos Santos-Pinto,
826 Perez-Riverol, Musacchio-Lasa, et al., 2018; Senji Laxme et al., 2019; Walker, Robinson, et
827 al., 2018). La majorité de ces études se sont cependant focalisées sur les composés qualifiés
828 d'allergènes telles que les phospholipases, les phosphatases acides, les serine protéases et les
829 antigènes 5 (Baumann et al., 2018), ou sur des peptides isolés (dos Santos-Pinto, Perez-Riverol,
830 Lasa, & Palma, 2018). Peu d'études complètes des peptidomes de venins d'hyménoptères
831 aculéates ont donc été réalisées via des approches protéotranscriptomiques.

832 Ce n'est que récemment que la caractérisation du peptidome complet du venin d'une fourmi
833 myrmicine a été effectuée (Touchard et al., 2018). Grâce à la méthode vénomique, trente-sept
834 précurseurs de toxines peptidiques ont en effet été identifiés dans le venin de la fourmi
835 *Tetramorium bicarinatum*, la plupart montrant une organisation classique de protéine sécrétée
836 avec une région signal suivie d'une région propeptide, puis d'une région dite mature,
837 constituant la toxine bioactive. La combinaison des données protéomiques et transcriptomiques
838 a également permis de mettre en évidence deux types de PTMs sur les parties matures des
839 toxines : des O-glycosylations par des N-acétyl-hexosamine sur des résidus thréonine N-
840 terminaux, et des amidations C-terminales induites par les motifs « GKK » ou « G ». Les
841 caractéristiques biochimiques et les effets biologiques des myrmicitoxines de la fourmi *T.*
842 *bicarinatum* sont détaillées dans la partie 3.6 de ce chapitre. L'analyse des régions signal et

843 propeptide a permis de répartir les myrmecitoxines en trois superfamilles de précurseurs (i.e. A,
844 B et C), et ce malgré la diversité des séquences présentées par les toxines matures des
845 superfamilles A et B (Touchard et al., 2018) (Figure 10). Parmi les superfamilles A et B,
846 quelques précurseurs ont présenté des identités de séquence moins élevées avec les séquences
847 consensus définies. Ces deux superfamilles ont donc été subdivisées (i.e. A₁, A₂, A₃, B₁ et B₂).
848 Les précurseurs regroupés dans la superfamille A possèdent des régions prépro de séquences et
849 structures similaires à celles présentées par d'autres peptides de venin de fourmis (i.e.
850 myrmeciitoxines (Cf. partie 3.2)) (Figure 10). Les superfamilles B et C constituent en revanche
851 de nouvelles familles de précurseurs dans la mesure où les régions prépro n'ont présenté aucune
852 similarité de séquence avec d'autres précurseurs de peptides de venin de fourmis déjà décrits.
853 Curieusement, les parties matures des précurseurs de la superfamille C ont montré des
854 similarités de séquence avec les sécapines, des peptides multifonctionnels trouvés dans les
855 venins d'hyménoptères aculéates (Cf. partie 3.6). Il a également été montré par une analyse
856 selon la méthode « Minimum Evolution », que les précurseurs de peptides issus d'organismes
857 non venimeux, tels que des « Host-Defense Peptides » impliqués dans l'immunité innée de
858 *Drosophila melanogaster*, se plaçaient à proximité de certains précurseurs de toxines du venin
859 de la fourmi *T. bicarinatum* (Figure 11).

860 Ces résultats soutiennent d'une part l'hypothèse qu'une grande diversité de toxines de venin
861 d'Hyménoptères dériveraient d'un nombre réduit de précurseurs, et d'autre part que certaines
862 toxines de venin seraient en lien avec le système immunitaire inné.



863

864 **Figure 10** : Alignements des précurseurs de myrmicitoxines du venin de *Tetramorium bicarinatum* et

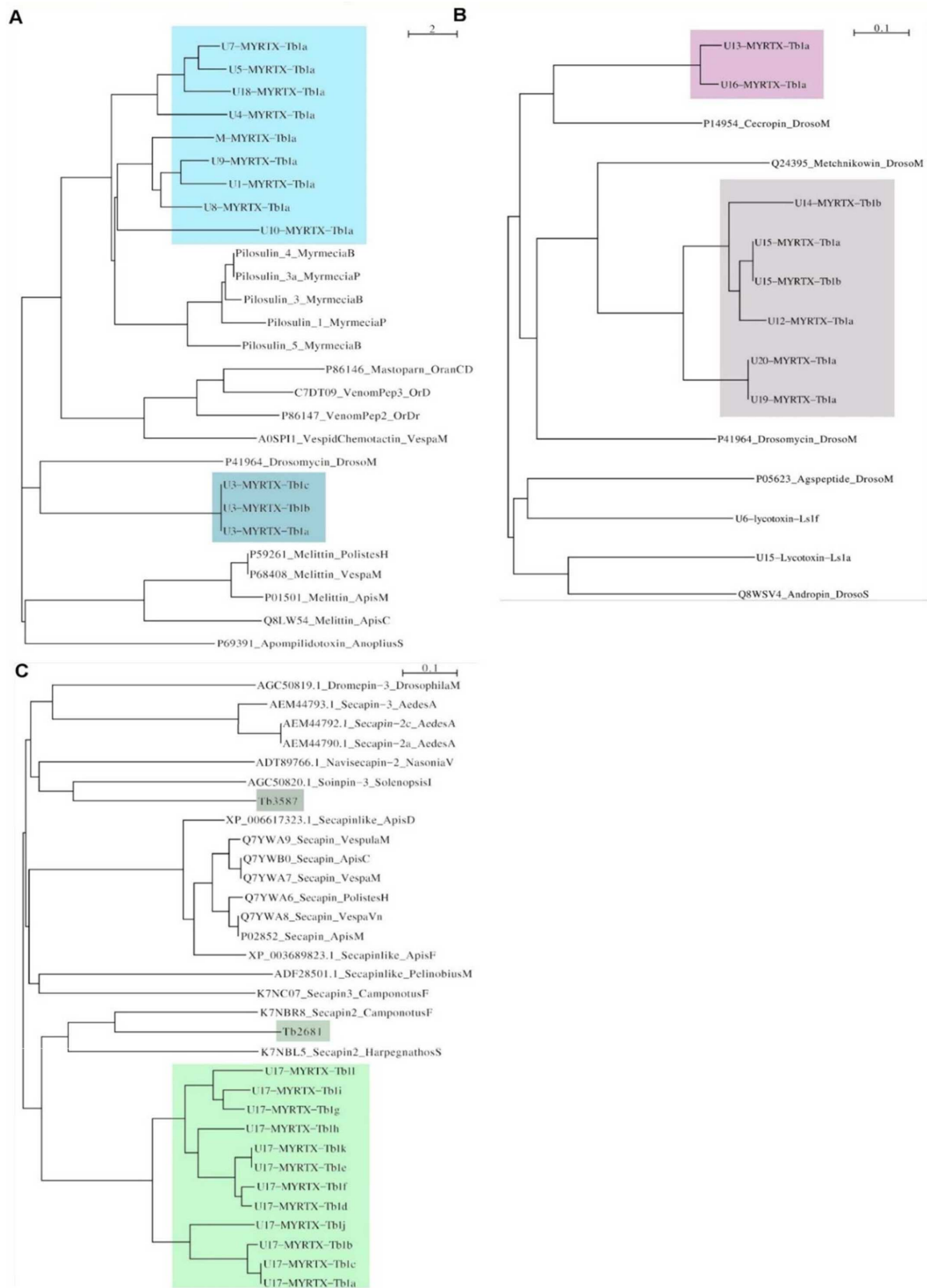
865 séparation en superfamilles de précurseurs (A, B et C) sur la base des région prépro. Les résidus identiques

866 sont surlignés en magenta. Les résidus similaires sont surlignés en bleu tandis que les résidus conservés sont

867 surlignés en cyan. Les triangles noirs indiquent les sites de clivage libérant les toxines matures. * Pour le

868 précurseur U₁₃-MYRTX-Tb1a, le site de clivage libérant la toxine mature est situé après le motif

869 « GEAEAEAG ». Les séquences signal prédictives sont surlignées. Modifié d'après (Touchard et al., 2018).



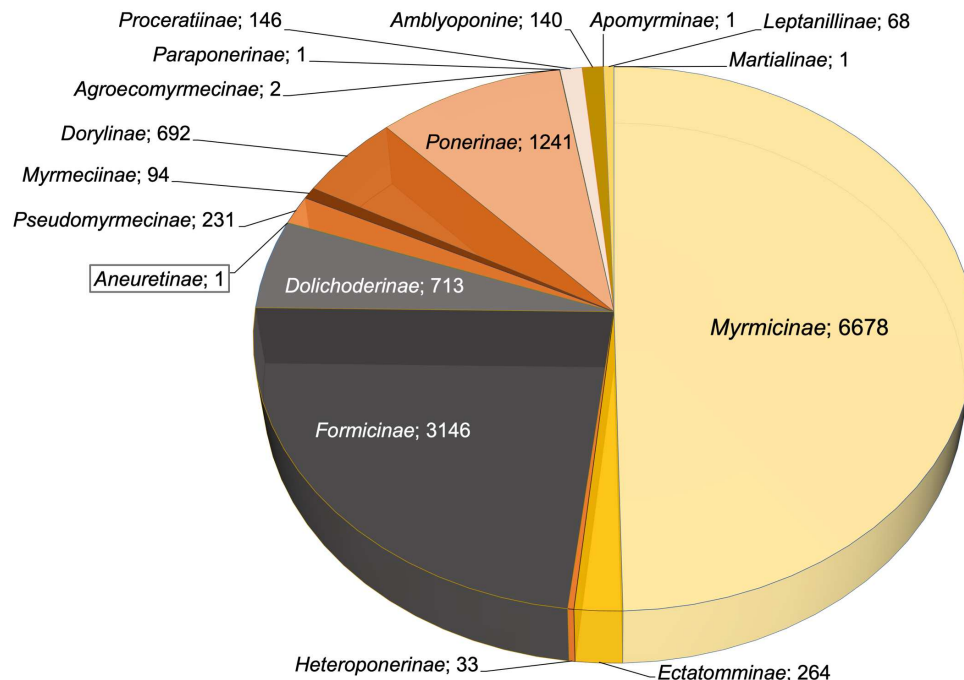
870
 871 **Figure 11** : Cladogrammes des précurseurs de toxines peptidiques du venin de *Tetramorium bicarinatum*.
 872 Les précurseurs de toxines peptidiques du venin de *T. bicarinatum* sont dans les cases colorées. A :
 873 superfamille A, pilosulines et autres précurseurs de peptides de venins et de « Host-Defense Peptides »
 874 (HDPs). B : Superfamille B et précurseurs de HDPs. C : superfamille C et précurseurs de sécapines. Tiré de
 875 (Touchard et al., 2018).

876 3. Les peptides de venins de fourmis

877 Il existe aujourd'hui plus de 16 000 espèces et sous-espèces de fourmis décrites (AntWeb,
878 le 08/03/20). Ces espèces se sont diversifiées et adaptées à de nombreuses niches écologiques
879 (Aili et al., 2014). Cette diversité spécifique et écologique s'illustre également par un vaste
880 champ d'utilisation des venins dans différents contextes de prédation, de défense (i.e.
881 prédateurs et pathogènes), de communication ou encore pour le développement d'une plante-
882 hôte (Frederickson et al., 2005; Orivel et al., 2001; J. O. Schmidt, 1987). En conséquence, la
883 composition de ces venins et la nature chimique des toxines présentent une grande variabilité.

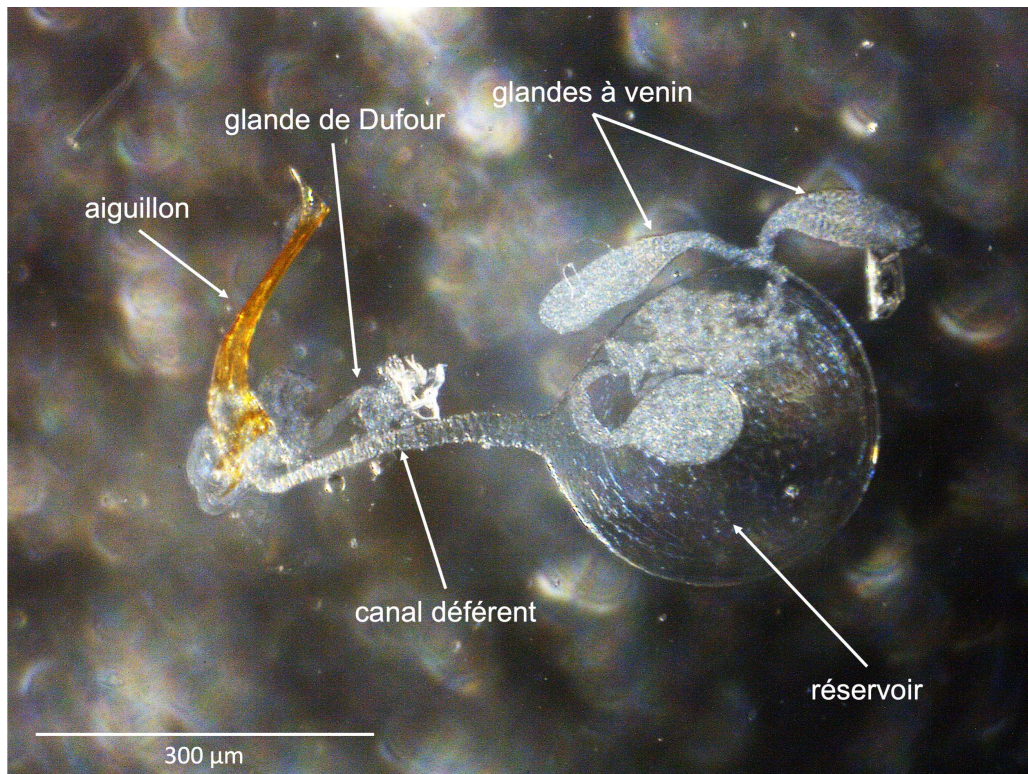
884 La majorité des fourmis (i.e. 70% des espèce) injecte leur venin au moyen d'un aiguillon
885 (Figure 12), venin dont les toxines, à l'exception de rares taxons (i.e. tribu des Solenopsidini)
886 sont majoritairement de nature protéique (Touchard, Aili, et al., 2016). Leur appareil vulnérant
887 est, comme pour la plupart des autres hyménoptères, dérivé d'un système reproducteur ancestral
888 (Robertson, 1968). Il se compose de deux glandes à venin, connectées à un réservoir au niveau
889 d'une partie convolutive, et d'un canal déférent qui relie ce réservoir à l'aiguillon. Ce dernier
890 est également associé à la glande de Dufour (Mitra, 2013), qui peut être atrophiée chez certaines
891 espèces (Figure 13).

892



893

894 **Figure 12 :** Diversité spécifique des sous-familles de Formicidae. Les sous-familles regroupant des espèces
895 de fourmis injectant leur venin au moyen d'un aiguillon sont indiquées dans les parts colorées. Les sous-
896 familles matérialisées en gris regroupent des espèces n'injectant pas leurs venins.



897

898 **Figure 13** : Appareil vulnérant de la fourmi myrmicine *Tetramorium bicarinatum*.

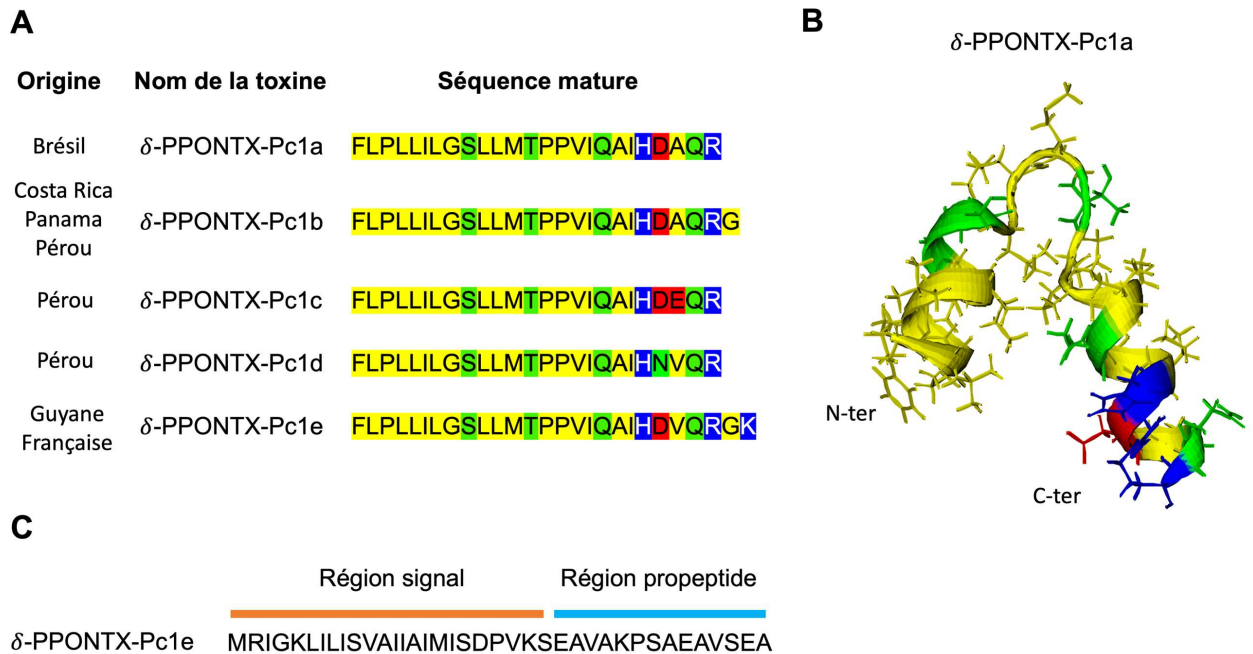
899 Aujourd'hui, la diversité moléculaire des toxines peptidiques des venins de fourmis se
 900 révèle grâce aux approches protéotranscriptomiques. La publication de nouvelles séquences de
 901 peptides croît de manière exponentielle, d'où la nécessité d'adopter une nomenclature
 902 rationnelle et stable. Les toxines peptidiques publiées ont donc été nommées selon la
 903 nomenclature standard définie par King et ses collaborateurs (G. F. King, Gentz, Escoubas, &
 904 Nicholson, 2008). Les fourmis constituent un groupe très diversifié dont les espèces sont
 905 classées en 17 sous-familles (Figure 12), une modification a ainsi été apportée à cette
 906 nomenclature afin d'adapter les noms génériques des toxines de venins de fourmis en fonction
 907 des noms de sous-familles (Touchard, Aili, et al., 2016).

908

909 3.1. Les paraponeratoxines (*Paraponerinae*)

910 La sous-famille *Paraponerinae* ne regroupe qu'une seule espèce, *Paraponera clavata* dite
 911 la fourmi « balle de fusil ». Cette fourmi mesure plus de 2 cm de longueur et est connue en
 912 raison de la douleur infligée par sa piqûre. La première étude portant sur ce venin a isolé, par
 913 des méthodes biochimiques, le composant le plus abondant, une neurotoxine peptidique linéaire
 914 (i.e. sans pont disulfure) longue de 25 acides aminés et initialement nommée poneratoxine
 915 (actuellement, δ -PPONTX-Pc1a) (Piek, Hue, Mantel, Terumi, & Schmidt, 1991) (Figure 14 –
 916 A). Une analyse RMN dans un environnement imitant une bicouche lipidique a démontré que

917 cette toxine était composée de deux hélices α connectées par une boucle au niveau d'un doublet
 918 de prolines (Figure 14 - B) (Szolajska et al., 2004). Elle est capable de moduler les influx
 919 nerveux des vertébrés et des invertébrés, notamment en prolongeant les potentiels d'actions via
 920 l'inhibition de l'inactivation des canaux sodiques voltage-dépendant (i.e. Nav 1.7) (Piek, Duval,
 921 et al., 1991; Piek, Hue, et al., 1991).



922
 923 **Figure 14** : Paraponeratoxines contenues dans les venins de différents spécimens de *Paraponera clavata*.
 924 (A) Détail des séquences matures des cinq paraponeratoxines. (B) Structure 3D de δ -PPONTX-Pc1a. Les
 925 acides aminés neutres apolaires sont colorés en jaune. Les résidus chargés positivement sont colorés en bleu
 926 tandis que ceux chargés négativement sont colorés en rouge. Les acides aminés neutres polaires sont colorés
 927 en vert. (C) Séquence de la région prépro de la toxine δ -PPONTX-Pc1e.

928 Quatre peptides analogues ont été découverts chez des spécimens appartenant à différentes
 929 populations de *P. clavata* d'Amérique du Sud (Costa Rica, Panama, Pérou et Guyane
 930 Française). Ces analogues, δ -PPONTX-Pc1b, -Pc1c, -Pc1d et -Pc1e, diffèrent d'un acide aminé
 931 dans leur région mature C-terminale. Toutes les paraponeratoxines possèdent une région N-
 932 terminale constituée majoritairement d'acides aminés hydrophobes (notamment en leucines)
 933 ainsi qu'une partie C-terminale chargée (Figure 14 – A). D'un point de vue pharmacologique,
 934 les peptides analogues sont tous capables de moduler les Nav 1.7 avec des efficacités différentes
 935 (Aili et al., 2020; Johnson, Rikli, Schmidt, & Evans, 2017). Seul le précurseur complet de δ -
 936 PPONTX-Pc1e a été identifié. La séquence signal n'a été déterminée que récemment et contient
 937 24 acides aminés (i.e. MRIGKLILISVAIIAIMISDPVKS). Elle ne présente pas de similarité

938 comparativement aux autres séquences signal décrites dans les venins de fourmis, et est suivie
939 d'un segment de 14 acides aminés constituant la région propeptide (EAVAKPSAEAVSEA)
940 (Figure 14 – C) (Aili et al., 2020).

941

942 3.2. *Les myrmeciitoxines (Myrmeciinae)*

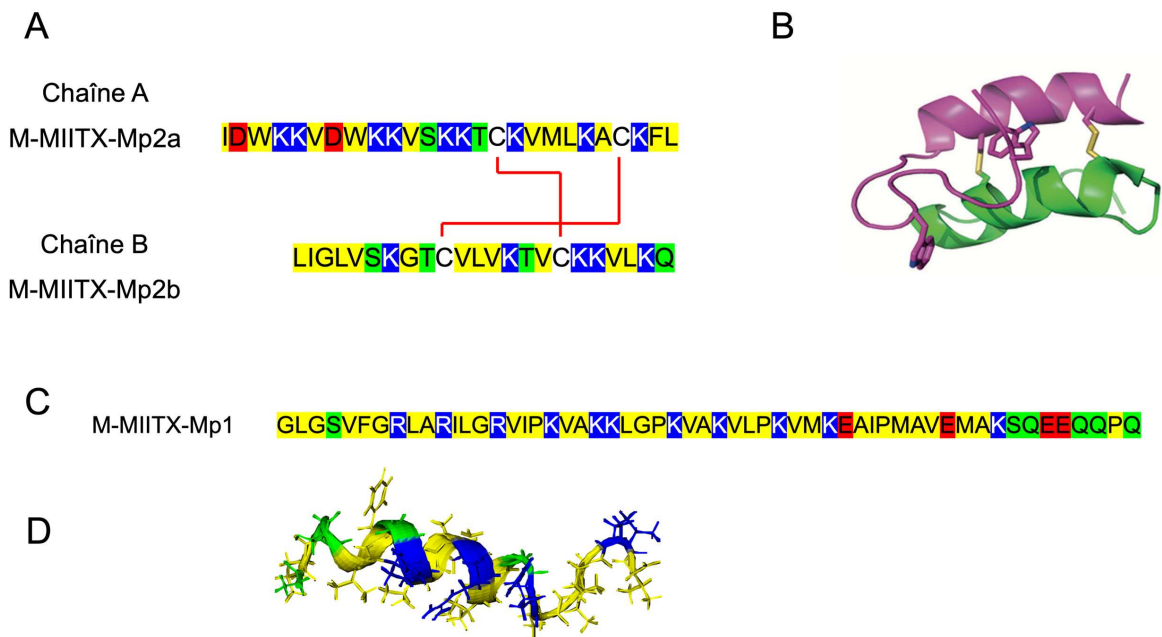
943 Les *Myrmeciinae* regroupent aujourd'hui 94 espèces endémiques d'Australie. Un certain
944 nombre d'espèces du genre *Myrmecia* cause fréquemment de réactions allergiques chez
945 l'Homme. Ces fourmis utilisent leur venin à la fois dans des contextes de prédation, notamment
946 afin de nourrir les larves carnivores, et de défense. Leurs piqûres sont d'ailleurs connues pour
947 être particulièrement douloureuses. Parmi les *Myrmeciinae*, les compositions en toxines
948 peptidiques des venins de deux espèces du groupe taxonomique *Myrmecia pilosula* (*M. pilosula*
949 et *M. banksi*), et celle du venin de *M. gulosa* ont été étudiées en détails (Wanandy, Gueven,
950 Davies, Brown, & Wiese, 2015).

951 Les compositions des venins de *M. pilosula* et *M. banksi* ont montré une majorité de toxines
952 dimériques initialement appelées pilosulines. La présence de peptides dimériques a également
953 été montré chez *M. gulosa*, bien que ce peptidome soit majoritairement composé de peptides
954 linéaires, dont certains possèdent des groupements N-acétylhexosamines (MIITX₁-Mg7a et
955 MIITX₁-Mg7b) (Tableau II) (Robinson et al., 2018). La majorité des myrmeciitoxines décrites
956 jusqu'à maintenant présentent des caractéristiques biochimiques communes dans les séquences
957 en acides aminés de leur partie mature, indépendamment de la présence ou non de cystéines.
958 Elles sont en effet toutes riches en lysine, et la plupart possèdent donc une charge positive nette
959 (Tableau II). De plus, les structures 3D étudiées ont révélé une prédominance d'hélices α
960 amphiphiles (Figures 15).

961 A titre d'exemple, le composant le plus abondant du venin de *M. pilosula* est un
962 hétérodimère composé de deux chaînes cationiques (i.e. M-MIITX-Mp2a et M-MIITX-Mp2b)
963 formant des hélices α reliées de manière antiparallèle par deux ponts-disulfures (Figure 15, A
964 et B). Cet hétérodimère agit par perturbation membranaire. Il exerce une activité
965 bactériostatique contre une large gamme de bactéries, une activité nociceptive sur cellules de
966 mammifères ainsi qu'une activité insecticide. Cette toxine induit en effet une douleur spontanée
967 *in vivo* chez les souris (e.g. comportements caractéristiques de léchage et de tressaillement de
968 la patte), et provoque *in vitro* une augmentation rapide de la concentration en Ca²⁺
969 intracellulaire suivie d'une lyse, indépendamment du type de cellule testé (i.e. cellules
970 neuronales et non-neuronales) (Nixon et al., 2020; Zelezetsky, Pag, Antcheva, Sahl, & Tossi,
971 2005). Les tests d'activité insecticide ont montré qu'elle induisait une paralysie irréversible

972 chez *Drosophila melanogaster* avec une dose létale de l'ordre du picomole/g (Nixon et al.,
 973 2020). Il est également apparu que la dimérisation, couplée à une connectivité antiparallèle des
 974 ponts disulfures, est essentielle à l'activité de cette toxine. Les analogues synthétisés,
 975 l'hétérodimère parallèle, les homodimères potentiels (parallèle et antiparallèle) ou les
 976 monomères, étaient en effet tous moins efficaces que l'hétérodimère antiparallèle natif, voire
 977 n'avaient aucun effet (Nixon et al., 2020).

978 Des activités nociceptive et insecticide ont également été démontrées pour une toxine
 979 peptidique linéaire du venin de *M. gulosa* (i.e. MIITX₁-Mg1a). Contrairement à l'hétérodimère
 980 antiparallèle de *M. pilosula*, l'augmentation en Ca²⁺ intracellulaire provoquée par ce peptide
 981 n'est pas corrélée à une lyse cellulaire sur des ganglions spinaux de souris. Cette toxine induit
 982 également une paralysie dose-dépendante (de l'ordre du µg/g) et réversible chez des criquets
 983 (Robinson et al., 2018).



984
 985 **Figure 15** : Myrmeciotoxines contenues dans le venin de *Myrmecia pilosula*. (A) Séquences des chaînes M-
 986 MIITX-Mp2a et M-MIITX-Mp2b, et connectivité des ponts disulfures de l'hétérodimère antiparallèle. (B)
 987 Représentation 3D de l'hétérodimère antiparallèle, tiré de (Dekan et al., 2017). La chaîne A est colorée en
 988 rose et la chaîne B est colorée en vert. (C) Séquence de M-MIITX-Mp1. (D) Structure 3D de M-MIITX-
 989 Mp1. Les acides aminés neutres apolaires sont colorés en jaune. Les résidus chargés positivement sont
 990 colorés en bleu tandis que ceux chargés négativement sont colorés en rouge. Les acides aminés neutres
 991 polaires sont colorés en vert.

992 Si les myrmeciotoxines décrites montrent une diversité dans les séquences de leurs parties
 993 matures (Tableau II), elles possèdent des régions signal et propeptide très similaires. Les

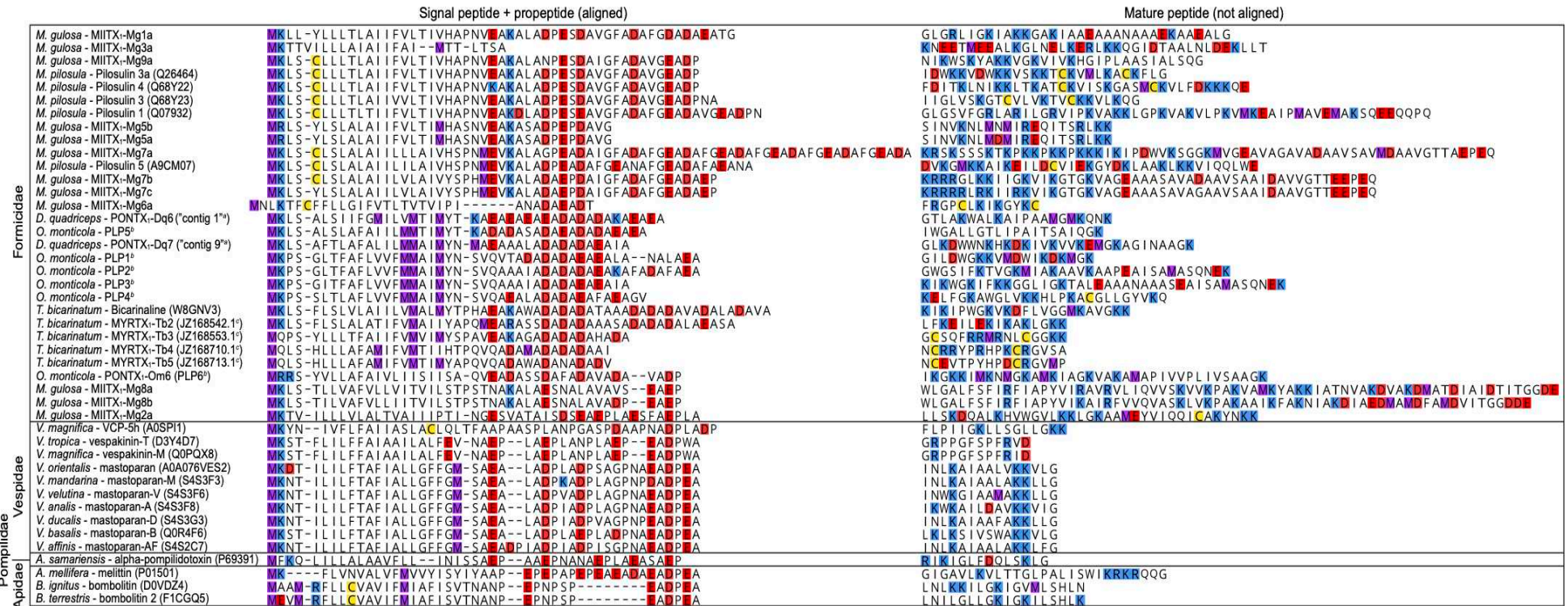
994 régions signal correspondent aux 25 aminés N-terminaux et suivent la séquence consensus
995 suivante : MKLSCLLLTLAIFVLTIVHAPNVEA. Ces régions signal sont suivies de régions
996 propeptides d'une trentaine d'acides aminés (consensus :
997 KALADPESDAVGFAXAFGEADAVGEADPNA) (Figure 16). Ces similarités de séquences
998 prépro sont également observées chez d'autres précurseurs décrits dans les venins de fourmis
999 ponerines et myrmecines (i.e. *Dinoponera quadriceps*, *Odontomachus monticola*, *Tetramorium*
1000 *bicarinatum*) (Kazuma, Masuko, Konno, & Inagaki, 2017; Mariano et al., 2019; Torres et al.,
1001 2014; Touchard et al., 2018), ainsi que chez d'autres hyménoptères aculéates (Figure 16). Les
1002 séquences des toxines matures présentent en revanche une grande diversité. Ces observations
1003 suggèrent que ces toxines peptidiques de venin, nommées « Aculéatoxines », appartiennent à
1004 la même famille de gènes et auraient donc dérivés d'un nombre réduit de précurseurs (Robinson
1005 et al., 2018).
1006

1007
1008

Tableau II : Toxines peptidiques contenues dans les venins des fourmis du genre *Myrmecia*. PTMs, post-translational modifications ; * Amidation C-terminale. Les cystéines sont surlignées en jaune. Les acides aminés chargés positivement sont surlignés en bleu tandis que les résidus chargés négativement sont surlignés en rouge.

Espèces	Nom des toxines	Séquences	PTMs
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg1a	GLGR L IG K IA KK GA K IAA E AAANAAA E KAA E AL*	
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg2a	LLS K D QAL K HVWGV L KK LG K AAM E YVIQ Q I C A K Y N KK	Homodimère
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg3a	K N E E TM E EAL K GLN E L K ER L KK Q G DTAALN L D E K LLT	
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg4a	SLVG C PR P D FLPSWN R C K S C V C K NN K L C P K IL K GSLL K TAA	Homodimère
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg4b	SLVG C PR P PNFLPSWN R C K C I C K NN K PM G R K LPNLL K TTA	Homodimère ?
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg5a	SIN V K NLM D MI R EQITS R L KK	
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg5b	SIN V K NLMNMI R EQITS R L KK	
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg6a	F R G P C L K I K G Y K C	
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg7a	K R S K SS S K T K P K P K K P K K K I K IP D W V K SG G K MV G E AVAGAVADAAVSAVM D AAVGT T A E P E Q	3 HexNAc
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg7b	K R R R GL K K I G K V I K GT G K V A G E AAASAV A D AAVSA A I D AVVGT T E E P E Q	3 HexNAc
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg7c	K R R R R LR K I R K V I K GT G K V A G E AAASAVAGAAVSA A I D AAVGT T E E P E Q	3 HexNAc ?
<i>Myrmecia gulosa</i>	MIITX ₁ -Mg9a	NI K W S K Y A K K V G K V V K H GIPLAASIALSQ*	
<i>Myrmecia gulosa</i>	MIITX ₂ -Mg1a	D IS D Y G D PC S D D L K D Y C I H G D C FF L K ELNQPAC R C Y T G Y Y G S R C E H I D H N	
<i>Myrmecia pilosula</i>	M-MIITX-Mp1	GLGSV F G R LARILGRVIP K V A K K LG P K V A K VLP K V M K E AIPMAV E MA K S Q E E QQ P Q	
<i>Myrmecia pilosula</i>	M-MIITX-Mp2a	I D W K K V D W K K V S K K T G K V M L K A C K F L *	Heterodimère
<i>Myrmecia pilosula</i>	M-MIITX-Mp2b	LIGLVS K GT C V L V K TV C K K V L K Q	
<i>Myrmecia banksi</i>	M-MIITX-Mb1a	IIGLVS K GT C V L V K TV C K K V L K Q *	1 S-S
<i>Myrmecia banksi</i>	M-MIITX-Mb2a	F D IT L N I K K L T K AT C K V I S K GASM C K V L F D K K K Q E	Homodimère
<i>Myrmecia banksi</i>	M-MIITX-Mb3a	D V K G M K K A I K E ILDCV E K G Y D K L A A K L K K V I Q L W E	Homodimère

1009



1010

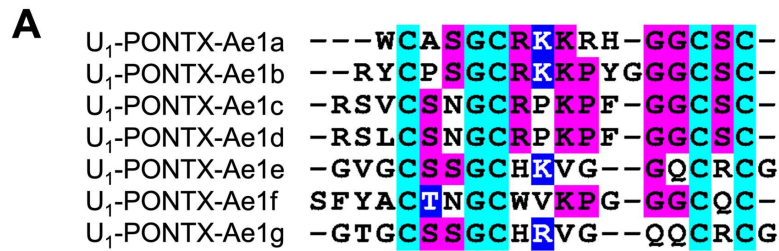
1011 **Figure 16 :** Alignements de précurseurs de toxines peptidiques de venin d'Hyménoptères sur la base des régions signal et propeptide. Les régions matures présentent
 1012 une grande diversité de séquences et n'ont pas été alignées. Les méthionines, lysines, arginines, aspartate/glutamate et cystéines sont respectivement surlignés en violet,
 1013 bleu, rouge et jaune. Les modifications post-traductionnelles n'ont pas été matérialisées. ^a (Torres et al., 2014), ^b (Kazuma et al., 2017) et ^c (Bouid et al., 2014). Tiré
 1014 de (Robinson et al., 2018)

1015 3.3. *Les poneritoxines (Ponerinae)*

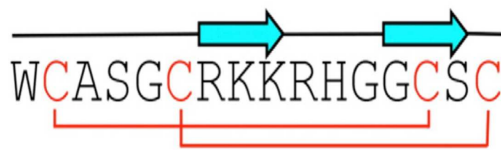
1016 Les *Ponerinae* constituent la troisième sous-famille phylogénétique en termes de nombre
1017 d'espèces, et sont principalement rencontrées dans les régions tropicales. Les ponerines sont
1018 généralement de grande taille et utilisent leur venin à la fois pour la prédation et la défense.
1019 Certaines espèces présentent également de puissantes mandibules, pouvant s'ouvrir jusqu'à
1020 180°, qu'elles utilisent dans la capture de proies. (Borowiec, Moreau, & Rabeling, 2020). Parmi
1021 cette sous-famille, les études des venins de 7 espèces de fourmis appartenant aux genres
1022 *Neoponera* (anciennement *Pachycondyla*), *Dinoponera*, *Odontomachus* et *Anochetus* ont mené
1023 à la caractérisation des séquences de 62 toxines peptidiques.

1024 La plupart des poneritoxines décrites sont des peptides linéaires, polycationiques et forment
1025 des hélices amphiphiles (Cologna et al., 2013; Johnson, Copello, Evans, & Suarez, 2010;
1026 Kazuma et al., 2017; Mariano et al., 2019; Orivel, 2000; Orivel et al., 2001; Torres et al., 2014).
1027 D'un point de vue pharmacologique, les poneritoxines linéaires testées montrent des propriétés
1028 multifonctionnelles (i.e. hémolytiques, cytostatiques contre une large gamme de
1029 microorganismes et insecticides) (Dodou Lima, de Paula Calvacante, & Radis-Baptista, 2020;
1030 Orivel et al., 2001; Radis-Baptista et al., 2020).

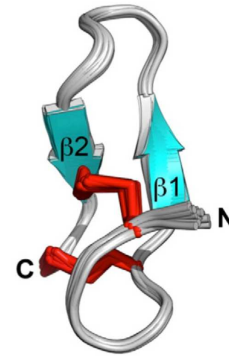
1031 Certaines poneritoxines présentent néanmoins des séquences et des structures originales.
1032 L'étude du venin d'*Anochetus emarginatus* a par exemple mené à l'identification des séquences
1033 de sept peptides possédant des séquences similaires et structurées par deux ponts disulfures
1034 (Touchard, Brust, et al., 2016) (Figure 17 - A). L'analyse RMN de la toxine U₁-PONTX-Ae1a
1035 a de plus révélé une structure tridimensionnelle n'ayant jamais été observée dans un venin
1036 animal : une β -hairpin antiparallèle C-terminale connectée à la région N-terminale par deux
1037 ponts-disulfures (Figure 17 - B et C). Cette toxine induit une paralysie réversible sur des
1038 mouches (PD₅₀ = 8,9 ± 3,1 nmol/g) et inhibe sélectivement les canaux calciques voltage-
1039 dépendant humains de type L (IC₅₀ = 4,6 μ M) (Touchard, Brust, et al., 2016).



B



C

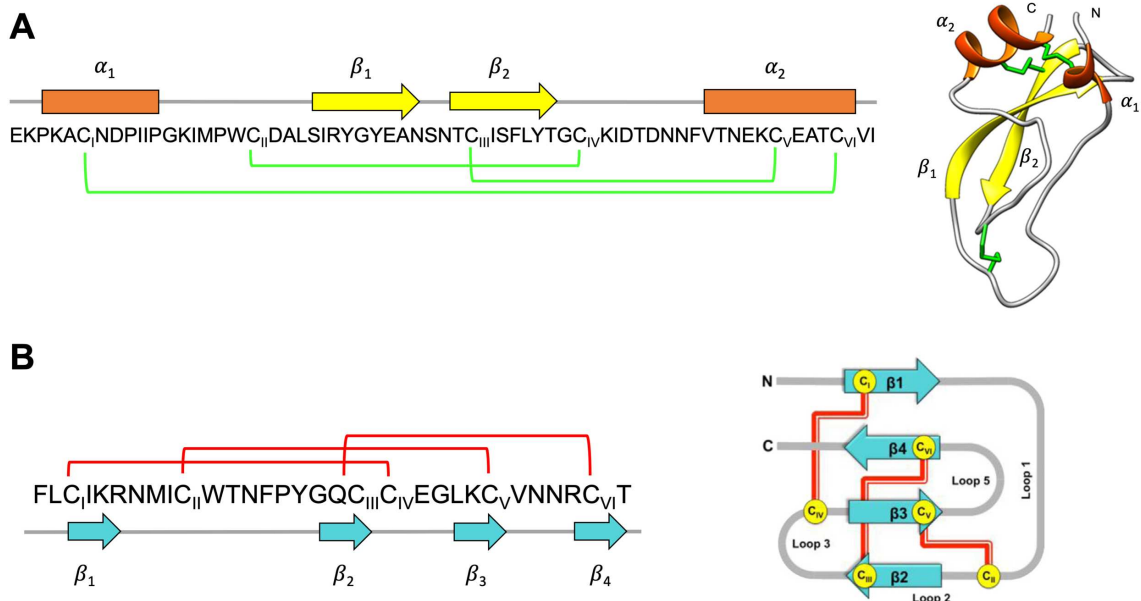


1040

1041 **Figure 17 :** Poneritoxines isolées du venin d'*Anochetus emarginatus*. (A) Les résidus identiques sont
 1042 surlignés en magenta. Les résidus similaires sont surlignés en bleu tandis que les résidus conservés sont
 1043 surlignés en cyan. (B) Structure de U₁-PONTX-Ae1a déterminée à partir d'une analyse RMN, et (C)
 1044 Représentation 3D. Les brins β sont colorés en cyan et les ponts disulfures sont matérialisés en rouge. Tirés
 1045 de (Touchard, Brust, et al., 2016).

1046 Des toxines présentant des structures de types Kunitz et ICK (Inhibitor Cystin Knot) ont
 1047 également été identifiées pour la première fois chez les fourmis dans le venin de *Dinoponera*
 1048 *quadriceps* (Figure 18) (Mariano et al., 2019). Ces structures, conférant stabilité et résistance
 1049 aux protéases grâce à l'agencement de trois ponts disulfures, ont déjà été observées dans de
 1050 nombreux venins (e.g. serpents, araignées). Le motif Kunitz consiste en deux hélices α et deux
 1051 feuillets β interconnectés par trois ponts disulfures (Figure 18 - A). Le motif ICK contient
 1052 quatre feuillets β antiparallèles formant un « pseudo-nœud » : un anneau est formé par deux
 1053 ponts disulfures (C_I-C_{IV} et C_{II}-V) dans lequel passe un troisième pont disulfure (C_{III}-C_{VI}) (Figure
 1054 18 - B).

1055 Seuls les précurseurs des peptides de venins de *D. quadriceps* et *Odontomachus monticola*
 1056 ont été caractérisés (Kazuma et al., 2017; Mariano et al., 2019). Les séquences signal sont
 1057 similaires à celles décrites pour les myrmecitoxines (Cf. partie 3.2) et suivent la séquence
 1058 consensus MKXSXLXLAFFXXVXMMAIMYNXVQATA. Elles sont suivies d'une région
 1059 propeptide longue de 14 à 22 acides aminés, suivant la séquence consensus :
 1060 AAAAXADADAXAEAXAXAXAEA.



1061
 1062 **Figure 18** : Poneritoxines de type Kunitz (**A**) et ICK (**B**) issus du venin de *Dinoponera quadriceps*. La
 1063 structure 3D de la toxine de type Kunitz est tirée de (Mariano et al., 2019) tandis que la représentation
 1064 schématique de la structure ICK est tirée de (Aili et al., 2014).

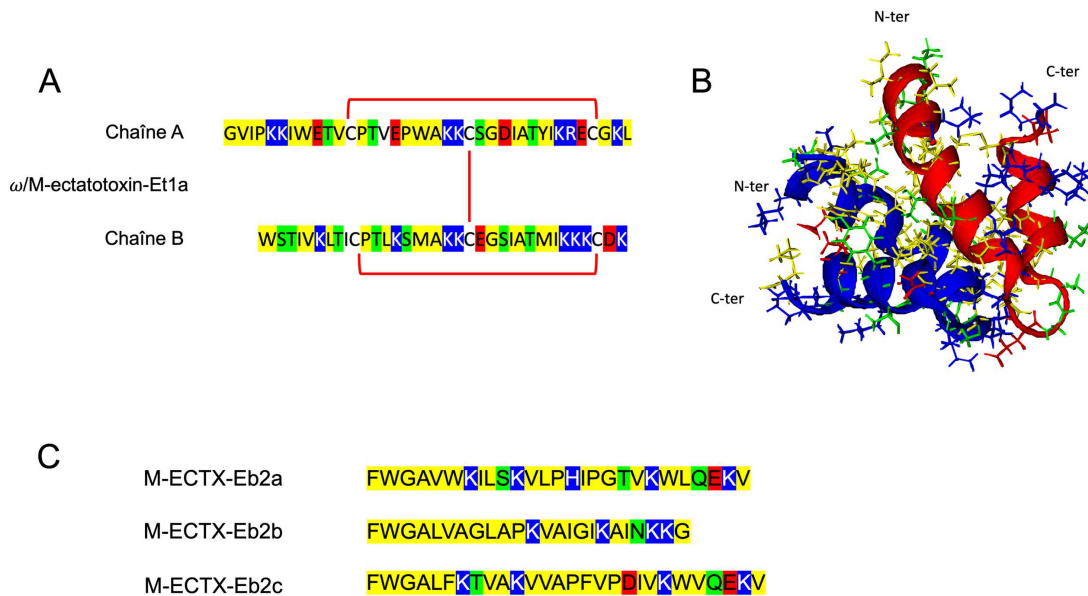
1065
 1066 *3.4. Les ectatotoxines*

1067 Quatre ectatotoxines ont été identifiées dans les venins d'*Ectatomma tuberculatum* et *E.*
 1068 *brunneum* (anciennement *E. quadricens*). La première ectatotoxine, caractérisée dans les
 1069 années 1990 chez *E. tuberculatum*, est un peptide dimérique constitué de deux chaînes
 1070 cationiques, ω /M-ectatotoxin-Et1a (sous-unité A – code Uniprot : P49343) et ω /M-ectatotoxin-
 1071 Et1a (sous unité B – code Uniprot : P49344) (K. A. Pluzhnikov et al., 1994). Les analyses en
 1072 RMN ont démontré que chaque chaîne d'acides aminés se compose de deux hélices α reliées
 1073 par une boucle de quatre résidus, formant une structure en épingle à cheveux stabilisée par un
 1074 pont disulfure intracaténaire. Les deux chaînes sont reliées entre elles par un troisième pont
 1075 disulfure. Ce peptide forme donc un ensemble de quatre domaines α (Nolde, Sobol, Pluzhnikov,
 1076 Grishin, & Arseniev, 1995; K. A. Pluzhnikov et al., 1994) (Figure 19, A et B). Les tests
 1077 d'activité biologiques sur des myocytes isolés de rats ont en effet permis d'identifier cette
 1078 toxine comme étant un inhibiteur de canaux calciques voltage-dépendant, mais elle exerce
 1079 également une action cytolytique sur des cellules de vertébrés et invertébrés (K. Pluzhnikov et
 1080 al., 1999).

1081 Les trois autres ectatotoxines ont été caractérisées chez *E. brunneum* (M-ectatotoxin-Eb2a,
 1082 -Eb2b, -Eb2c), et sont toutes des peptides cationiques linéaires possédant des effets

1083 bactériolytiques contre des bactéries Gram négatif (Kirill A. Pluzhnikov et al., 2014) (Figure
 1084 19, C).

1085 Cependant, aucune étude transcriptomique n'a été effectuée sur les glandes à venin de ces
 1086 fourmis. Les séquences entières des précurseurs des ectatotoxines n'ont donc pas été
 1087 caractérisées.



1088 **Figure 19** : Ectatotoxines retrouvées dans les venins d'*Ectatomma tuberculatum* et *E. brunneum*. (A)
 1089 Séquences des chaînes composant l'hétérodimère ω/M-ectatotoxin-Et1a de *E. tuberculatum*, et connectivité
 1090 des ponts disulfures. (B) Représentation 3D de l'hétérodimère ω/M-ectatotoxin-Et1a, tiré de (K. A.
 1091 Pluzhnikov et al., 1994). La chaîne A est colorée en bleu et la chaîne B est colorée en rouge. (C) Séquences
 1092 des ectatotoxines linéaires trouvées dans le venin de *E. brunneum*. Les résidus chargés positivement sont
 1093 colorés en bleu tandis que les résidus chargés négativement sont colorés en rouge. Les résidus neutre polaires
 1094 sont colorés en vert tandis que les résidus neutres apolaires sont colorés en jaune.

1096

1097 3.5. Les pseudomyrmecitoxines

1098 La sous-famille des *Pseudomyrmecinae* comprend des fourmis majoritairement arboricoles
 1099 des genres *Pseudomyrmex*, *Myrcidris* et *Tetraponera* (Borowiec et al., 2020). Parmi les
 1100 pseudomyrmecines, les venins de deux espèces de fourmis du genre *Pseudomyrmex* ont été
 1101 étudiés (i.e. *P. penetrator* et *P. triplarinus*). Ces fourmis possèdent des venins qui paralysent
 1102 rapidement les arthropodes, et dissuadent efficacement les herbivores vertébrés en induisant
 1103 une forte douleur, ce qui suggère qu'elles contiennent à la fois des composés neurotoxiques et
 1104 cytotoxiques. Bien que les venins de *Pseudomyrmex* possèdent des peptides aux structures
 1105 variées, avec notamment des peptides monomériques linéaires et structurés par des ponts

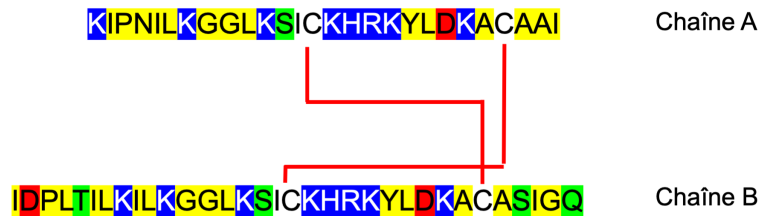
1106 disulfures, toutes les pseudomyrmecitoxines décrites à ce jour sont des toxines dimériques. Les
1107 premières toxines isolées du venin de *P. triplarinus* sont des hétérodimères constitués de
1108 chaînes cationiques amphiphiles de séquences similaires, reliées entre elles par deux ponts
1109 disulfures (Figure 20 – A et C). D’un point de vue fonctionnel, les six hétérodimères produisent
1110 une inhibition dose-dépendante significative d’œdèmes induits par la carraghénine chez des
1111 rats, et pourraient donc exercer une action anti-inflammatoire (Pan & Hink, 2000).

1112 Plus récemment, un hétérodimère antiparallèle a été identifié chez *P. penetrator* (i.e. Δ -
1113 PSDTX-Pp1a). Les deux chaînes composant cet hétérodimère sont cationiques, amphiphiles et
1114 reliées entre elles de manière antiparallèle grâce à deux ponts disulfures (Figure 20 – B). Elles
1115 possèdent également des séquences similaires, avec 85% d’identité de séquence (Figure 20 –
1116 C). La structure antiparallèle de cette toxine lui confère une grande stabilité, avec une demi-vie
1117 de 13 heures en présence de protéinase K. Les tests *in vitro* et *in vivo* ont également montré
1118 qu’elle exerce un effet cytotoxique sur les cellules d’*Aedes albopictus* ($IC_{50} = 1;04 \mu M$), ainsi
1119 qu’une paralysie irréversible sur des mouches menant à la mort en moins de 24h ($LD_{50} =$
1120 3 nmol/g) (Touchard, Mendel, Boulogne, Herzig, & Emidio, 2020).

1121 Aucune étude transcriptomique des glandes à venin de ces espèces n’a été effectuée. Les
1122 séquences signal des pseudomyrmecitoxines restent encore non caractérisées à ce jour.

A

Pseudomyrmecitoxines	Masse (Da)	Sous-unités	
		Petites	Grandes
I	7069	SS1	LS2
II	6998	SS1	LS1
III	7090	SS2	LS2
IV	7143	SS2	LS2
V	7017	SS3	LS1
VI	7069	SS3	LS1

B**C**

	%ID	%S	Espèces
Δ-PSDXT-Pp1a - A	-	-	<i>P. penetrator</i>
Δ-PSDXT-Pp1a - B	85	89	<i>P. penetrator</i>
U ₁ -PSDXT-Pt1-SS1	34	59	<i>P. triplarinus</i>
U ₁ -PSDXT-Pt1-LS2	36	58	<i>P. triplarinus</i>
U ₁ -PSDXT-Pt1-SS2	31	59	<i>P. triplarinus</i>
U ₁ -PSDXT-Pt1-LS1	36	55	<i>P. triplarinus</i>
U ₁ -PSDXT-Pt1-SS3	31	59	<i>P. triplarinus</i>

1123

1124 **Figure 20 :** Pseudomyrmecitoxines retrouvées chez *Pseudomyrmex triplarinus* et *P. penetrator*. (A)1125 Composition des hétérodimères du venin de *P. triplarinus*. (B) Connectivité des ponts-disulfures de

1126 l'hétérodimère Δ-PSDXT-Pp1a. Les acides aminés neutres apolaires sont colorés en jaune. Les résidus

1127 chargés positivement sont colorés en bleu tandis que ceux chargés négativement sont colorés en rouge. Les

1128 acides aminés neutres polaires sont colorés en vert. (C) Alignements des séquences de

1129 pseudomyrmecitoxines décrites à ce jour. Les résidus identiques sont surlignés en magenta tandis que les

1130 résidus similaires sont surlignés en bleu.

1131

3.6. Les myrmecitoxines

1132 Les *Myrmicinae* regroupent actuellement le plus grand nombre d'espèces décrites (i.e. 6678

1133 espèces) (Borowiec et al., 2020). Seuls les venins de trois espèces de fourmis myrmecines ont

1134 cependant été étudiés. Deux myrmecitoxines linéaires ont en effet été isolées des venins de

1135 *Myrmica rubra* et *Manica rubida* (Heep, Klaus, et al., 2019; Heep, Skaljac, et al., 2019). Ces

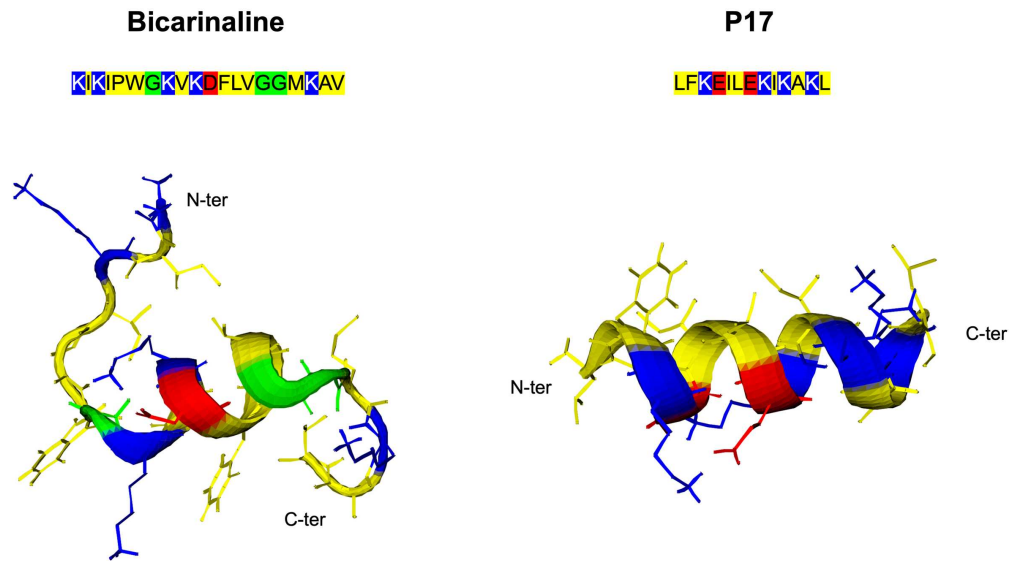
1136 toxines, appelées respectivement U-MYRTX-MRArub1 et U-MYRTX-MANr1, sont
 1137 constituées de 10 acides aminés et montrent 80% d'identité de séquence (Figure 21). Elles
 1138 possèdent toutes deux une charge nette négative, due à la présence d'un acide aspartique et d'un
 1139 acide glutamique. U-MYRTX-MANr1 cause une paralysie réversible chez des pucerons tandis
 1140 que U-MYRTX-MRArub1 augmente leur sensibilité aux insecticides (e.g. imidaclopride,
 1141 méthomyle) (Heep, Klaus, et al., 2019; Heep, Skaljic, et al., 2019).

1142 Les régions signal et propeptides de ces deux toxines n'ont cependant pas été décrites.

		%ID	%S	Espèces
U-MYRTX-MRArub1	IDPKLES ⁺ LA	-	-	<i>Myrmica rubra</i>
U-MYRTX-Man1	IDPKVLES ⁺ SLV	80	90	<i>Manica rubida</i>

1143
 1144 **Figure 21** : Myrmicitoxines isolées des venins de *Myrmica rubra* (i.e. MRArub1) et *Manica rubida* (i.e.
 1145 Man1). Les résidus chargés négativement sont surlignés en rouge tandis que les résidus chargés positivement
 1146 sont surlignés en bleu.

1147 En revanche, la caractérisation du peptidome du venin de *T. bicarinatum* (Cf. partie 2.5 de
 1148 ce chapitre) a mené à l'identification de 37 précurseurs de toxines peptidiques. Ces peptides ont
 1149 été classés en vingt familles selon les similarités de séquence de leur partie mature (i.e. U₁ à
 1150 U₂₀), et consistent majoritairement en des toxines linéaires et cationiques. Trois familles
 1151 regroupent cependant des peptides possédant un pont disulfure (i.e. U₆, U₁₁ et U₁₇). Les deux
 1152 toxines les plus abondantes, M-MYRTX-Tb1a (i.e. Bicarinaline) et U₁-MYRTX-Tb1a (i.e.
 1153 P17), sont linéaires, cationiques et amphiphiles (Figure 22). La caractérisation des séquences
 1154 des précurseurs de ces deux peptides a également révélé que les gènes les codant s'expriment
 1155 en dehors du système vulnérant, dans l'abdomen de la fourmi *T. bicarinatum*. D'un point de
 1156 vue fonctionnel, la Bicarinaline exerce une activité bactériolytique contre une large gamme de
 1157 bactéries Gram positives et négatives tandis que le P17 possède une activité
 1158 immunomodulatrice sur des macrophages humain *via* l'activation d'un RCPG (i.e. MRGPRX2)
 1159 (Benmoussa et al., 2017; Duraisamy et al., 2020; Rifflet et al., 2012; Téné et al., 2014).

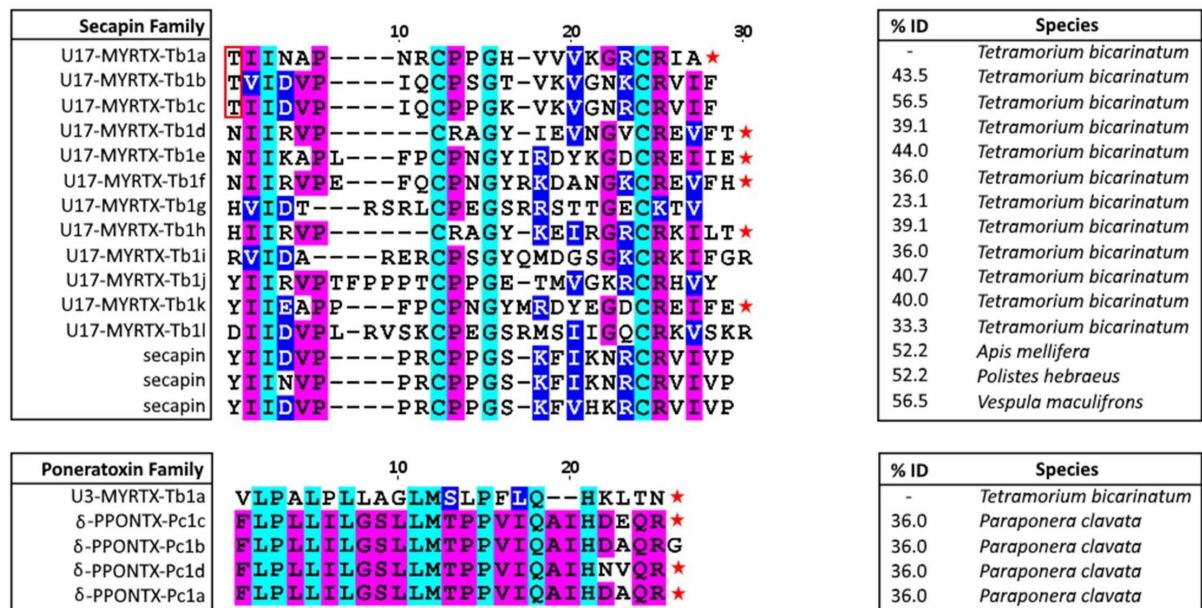


1160

1161 **Figure 22** : Séquences et représentations 3D des peptides matures Bicarinaline et P17. Les résidus chargés
 1162 positivement sont colorés en bleu. Les résidus chargés négativement sont colorés en rouge. Les résidus
 1163 neutres polaires sont colorés en vert. Les résidus neutres apolaires sont colorés en jaune.

1164

1165 Certaines myrmecitoxines matures présentent des similarités de séquences avec des peptides
 1166 caractérisés dans les venins d'autres hyménoptères aculéates. Les peptides U₁₇ possèdent par
 1167 exemple 23 à 56% d'identité de séquence avec des sécapines identifiées dans des venins d'*Apis*
 1168 *mellifera* et de guêpes. Bien que le peptide U₃ et la paraponeratoxine isolée du venin de *P.*
 1169 *clavata* présentent seulement 36% d'identité de séquence, ces peptides montrent des
 1170 caractéristiques biochimiques proches, notamment une majorité d'acides aminés hydrophobes
 1171 concentrés dans la région N-terminale ainsi que des acides aminés chargés en position C-
 1172 terminale (Figure 23) (Touchard et al., 2018).



1173
1174 **Figure 23** : Aligment des séquences des myrmicotoxines matures de *Tetramorium bicarinatum* avec
1175 d'autres peptides de venins d'hyménoptères aculéates. Les résidus identiques sont surlignés en magenta. Les
1176 résidus similaires sont surlignés en bleu tandis que les résidus conservés sont surlignés en cyan. Les
1177 pourcentages d'identité (%ID) sont relatifs au premier peptide de chaque famille. Les étoiles rouges
1178 représentent des amidations C-terminales et les résidus de thréonine encadrés en rouge sont O-glycosylés.
1179 Tiré de (Touchard et al., 2018).

1180
1181 En conclusion, un total de 136 toxines peptidiques de venins a été identifié dans six sous-
1182 familles phylogénétiques des Formicidae. La majorité de ces peptides sont linéaires,
1183 cationiques et amphiphiles, et induisent des effets nociceptifs, cytolytiques et/ou insecticides.
1184 Des structures plus complexes ont été également caractérisées, impliquant notamment la
1185 formation de ponts disulfures (e.g. Kunitz-like, ICK-like, dimères antiparallèles), avec des
1186 activités biologiques tout aussi remarquables. Certaines études ont révélé que les précurseurs
1187 de certaines toxines montrent des séquences signal et propeptide similaires à celles d'autres
1188 peptides identifiés dans les venins d'espèces de sous-familles phylogénétiques voire de familles
1189 différentes. Ceci suggère que ces toxines, bien que présentant une grande diversité de séquences
1190 matures, seraient codées par une même famille de gènes. Néanmoins, seulement 69 séquences
1191 de précurseurs de toxines peptidiques de venins de fourmis ont été identifiées, soit un peu plus
1192 de la moitié du nombre de peptides caractérisés.

1193

1194 4. L'immunité des insectes

1195 Récemment, il a été montré que les régions prépro de « Host Defense Peptides » (HDPs) de
1196 l'hémolymphe de la drosophile se plaçaient dans les mêmes clusters que certaines toxines du
1197 venin de *T. bicarinatum* (Touchard et al., 2018). Ces résultats, combinés aux activités
1198 bactériolytiques et immunomodulatrices de la Bicarinaline et du P17 et l'expression des gènes
1199 les codant en dehors du système vulnérant, soulèvent la question du lien pouvant exister entre
1200 le système immunitaire inné des insectes venimeux et les toxines contenues dans leur venin.

1201 L'immunité des organismes pluricellulaires implique des organes et mécanismes de défense
1202 qui les protègent contre des substances étrangères et des pathogènes via différents systèmes de
1203 reconnaissance et de réponse. Les insectes possèdent un système immunitaire complexe et
1204 efficace, et ce malgré leur courte espérance de vie (Larsen, Reynaldi, & Guzmán-Novoa, 2019).
1205 De par son cycle de vie court et la connaissance de son génome, la mouche *Drosophila*
1206 *melanogaster* est l'insecte qui a été le plus étudié dans cette thématique (Charroux & Royet,
1207 2010). Ce modèle biologique a d'ailleurs permis une meilleure compréhension de l'immunité
1208 chez d'autres organismes. Les systèmes immunitaires des moustiques et des abeilles mellifères
1209 sont par exemple très similaires à celui des drosophiles. De plus, la plupart des voies de
1210 signalisation cellulaire découvertes chez les insectes sont conservées chez les mammifères
1211 (Bang, 2019; Lemaitre & Hoffmann, 2007; Tafesh-Edwards & Eleftherianos, 2020).

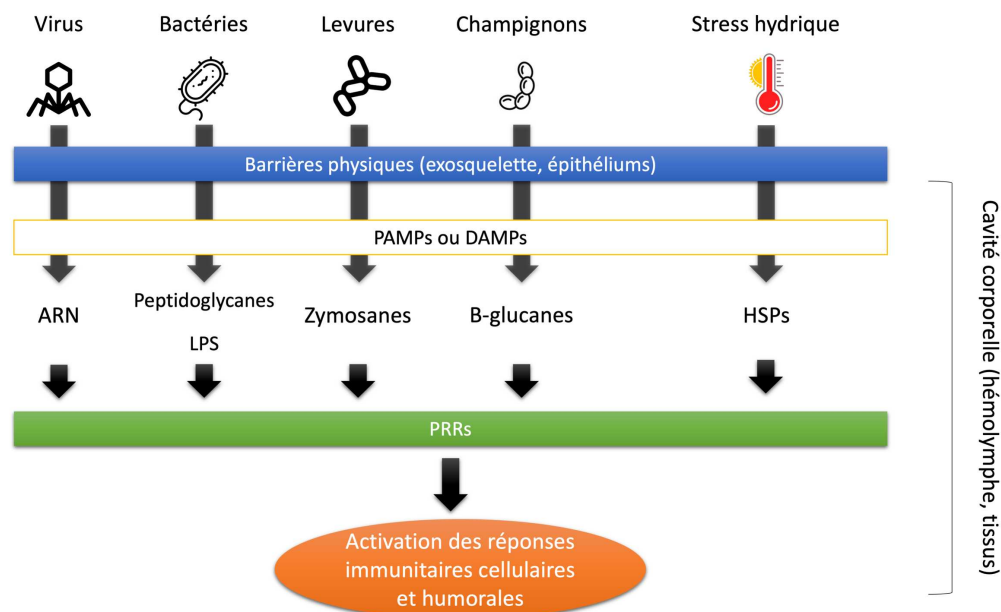
1212 4.1. Immunité acquise vs. Immunité innée : définitions et application aux insectes

1213 Deux types d'immunité sont classiquement distingués : l'immunité innée et l'immunité
1214 acquise. L'immunité acquise, appelée également « spécifique » ou « adaptative », décrit une
1215 protection améliorée contre des réinfections avec un type de pathogène (Larsen et al., 2019).
1216 Cette protection implique donc une mémoire à long terme et des molécules agissant dans la
1217 reconnaissance du soi et du non-soi, telles que les molécules du complexe majeur
1218 d'histocompatibilité, les récepteurs T et les anticorps. Elles agissent de manière spécifique
1219 contre un pathogène donné quand l'organisme y est exposé plus d'une fois (Larsen et al., 2019;
1220 Helge Schlüns & Crozier, 2009). Cette immunité médiée par les anticorps est retrouvée
1221 exclusivement chez les vertébrés (Kimbrell & Beutler, 2001).

1222 L'immunité innée comprend une protection avec une faible capacité de mémoire
1223 immunologique. Cette protection intervient rapidement suite à l'exposition aux pathogènes ou

1224 aux substances toxiques et est considérée comme la première ligne de défense des organismes
1225 pluricellulaires. Elle est en conséquence très conservée d'un point de vue phylogénétique. (Gao,
1226 Tian, & Zhu, 2007; Imler & Hoffmann, 2001). De manière générale, les invertébrés, et plus
1227 particulièrement les insectes, sont dits dépourvus de système immunitaire adaptatif car ils ne
1228 possèdent pas de lymphocyte (Zaidman-Rémy et al., 2006; Zasloff, 2002). La défense de
1229 l'organisme repose ainsi essentiellement sur les réactions du système immunitaire inné
1230 (Satyavathi et al. 2014). Cette protection comprend des barrières physiques (exosquelette,
1231 épithéliums), des molécules de reconnaissance (les PRRs pour Pattern Recognition Receptors)
1232 détectant les motifs moléculaires associés aux pathogènes (les PAMPs pour Pathogen
1233 Associated Molecular Patterns) ou à des lésions tissulaires (les DAMPs pour Damage
1234 Associated Molecular Patterns), des voies de signalisation biochimiques ainsi que des
1235 molécules effectrices, agissant directement sur le pathogène (Larsen et al., 2019; Helge Schlüns
1236 & Crozier, 2009).

1237 Les PAMPs sont des composants essentiels et conservés d'un groupe défini de
1238 pathogènes. On distingue par exemple les lipopolysaccharides et les peptidoglycanes de type
1239 Lys des membranes externes des bactéries à Gram négatif, les peptidoglycanes de type DAP
1240 des bactéries Gram positives, les zymosanes des levures, les β -glucanes des champignons ou
1241 encore les ARN doubles-brins de certains virus (Larsen et al., 2019; Sheehan, Farrell, &
1242 Kavanagh, 2020). Quant aux DAMPs, il s'agit de molécules exprimées dans les cellules ayant
1243 subi des dommages infectieux ou non, telles que les protéines de choc thermique (en anglais,
1244 HSPs pour Heat Shock Proteins) (Figure 24). Ces structures agissent en tant que ligands sur les
1245 PRRs, au niveau des hémocytes circulant dans l'hémolymphe ou à la surface de différents tissus
1246 tels, que les épithéliums et le corps gras (Larsen et al., 2019; Uvell & Engström, 2007). Ces
1247 récepteurs sont liés à des voies de signalisation biochimiques qui enclenchent deux types de
1248 réponses : les réponses cellulaire et humorale (Satyavathi et al. 2014) (Figure 24). Ces réponses
1249 ne sont cependant pas si tranchées dans la mesure où des facteurs humoraux régulent certaines
1250 réponses cellulaires, et que des réponses cellulaires sont sources de molécules de défense
1251 humorale tels que les radicaux libres et la phénoloxydase.



1252

1253 **Figure 24 :** Motifs moléculaires associés aux pathogènes (PAMPs pour Pathogen Associated Molecular
 1254 Patterns) ou aux lésions tissulaires (DAMPs pour Damage Associated Molecular Patterns) déclenchant la
 1255 réponse immunitaire innée des insectes. S'ils réussissent à entrer dans la cavité corporelle, ces motifs seront
 1256 détectés par les PRRs (Pathogen Recognition Receptors), déclenchant les différentes réponses immunitaires
 1257 cellulaires et humorales.

1258

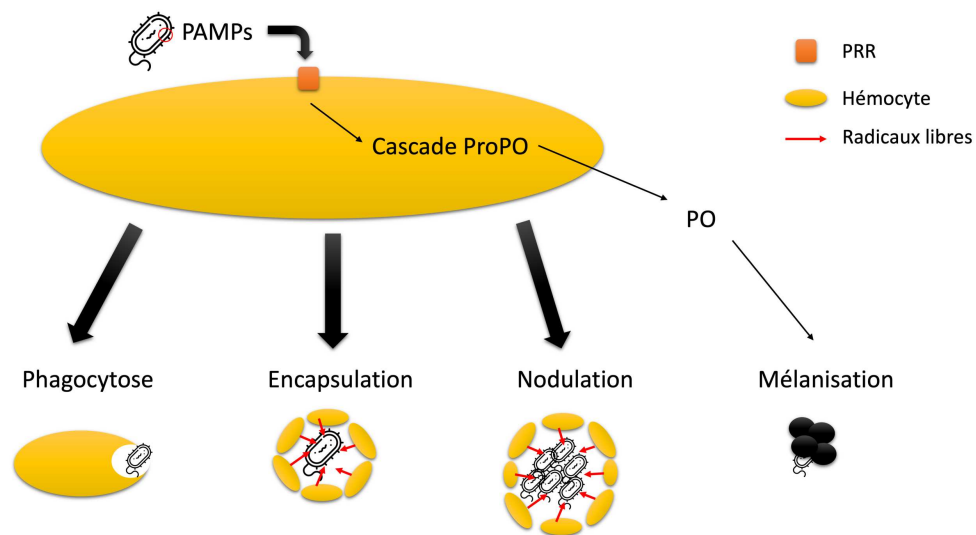
1259 4.2. Médiation par la réponse cellulaire

1260 La réponse cellulaire fait référence aux actions exercées par les hémocytes, à savoir la
 1261 phagocytose, l'encapsulation, la nodulation, la mélanisation et la cascade de la phénoloxydase
 1262 (Ali Mohammadie Kojour, Han, & Jo, 2020; Larsen et al., 2019; Satyavathi, Minz, & Nagaraju,
 1263 2014) (Figure 25). Il est nécessaire de noter que, chez les insectes, les hémocytes, en association
 1264 avec d'autres organes, synthétisent et stockent également des peptides antimicrobiens, appelés
 1265 ici Host-Defense Peptides (HDPs) (Larsen et al., 2019).

1266 Les mécanismes immunitaires cellulaires contribuent à l'élimination des pathogènes en les
 1267 phagocytant, en les lysant, ou encore en les cernant pour les neutraliser (Strand, 2008). Les
 1268 agents pathogènes de petite taille sont phagocytés. Les agents de plus grande taille déclenchent
 1269 l'encapsulation tandis que la nodulation fait référence aux agrégats d'hémocytes emprisonnant
 1270 un grand nombre de bactéries. Ces deux derniers mécanismes impliquent tous deux une action
 1271 coopérative des hémocytes (Larsen et al., 2019; Satyavathi et al., 2014). Il y a en effet mise en
 1272 place d'agrégats d'hémocytes à la surface de l'agent pathogène, suivie d'une sécrétion de

1273 radicaux libres de l'oxygène et d'azote pour sa destruction (Figure 25). Une sécrétion
1274 simultanée d'agents antioxydants est également observée afin de minimiser les dommages à
1275 l'encontre des hémocytes (Larsen et al., 2019; Satyavathi et al., 2014).

1276 La mélanisation est une combinaison de processus cellulaires et humoraux qui s'effectue
1277 lors de l'encapsulation, la nodulation et plus généralement la cicatrisation. Ce mécanisme
1278 permet de traiter les blessures, qu'elles soient liées à des agents pathogènes ou non (Larsen et
1279 al., 2019). Elle est déclenchée de manière locale en réponse à une rupture de la barrière
1280 cuticulaire, ou de manière systémique lors de l'invasion microbienne de l'hémolymphe
1281 (Nakhleh, El Moussawi, & Osta, 2017). Son activation est médiée par la cascade de la
1282 prophénoloxydase, elle-même déclenchée par l'activation de PRRs déployés par les hémocytes
1283 (Larsen et al., 2019) (Figure 25).



1284

1285 **Figure 25 :** Réponses cellulaires exercées par les hémocytes des insectes lors d'une infection microbienne.
1286 Suite à la reconnaissance des PAMPs effectuée par les PRRs à la surface des hémocytes, quatre types de
1287 réponses cellulaires peuvent être observées. Les pathogènes de petite taille seront majoritairement
1288 phagocytés. Les agents de plus grande taille déclenchent l'encapsulation tandis que la nodulation fait
1289 référence aux agrégats d'hémocytes emprisonnant un grand nombre d'agents pathogènes. Ces deux derniers
1290 mécanismes impliquent tous deux une action coopérative des hémocytes. Suite à une agrégation à la surface
1291 du pathogène, il y aura sécrétion de radicaux libres pour sa destruction. La mélanisation est une combinaison
1292 de processus cellulaires et humoraux et s'effectue durant l'encapsulation, la nodulation et plus généralement
1293 la cicatrisation. Lors d'une invasion microbienne de l'hémolymphe, la cascade de la prophénoloxydase
1294 (ProPO) sera déclenchée par les PRRs, qui déclenche une sécrétion de phénoloxydase (PO) et enfin un dépôt
1295 de mélanine autour du pathogène.

1296 4.3. Médiation par la réponse humorale

1297 La réponse humorale correspond majoritairement à la synthèse d'HDPs et autres protéines
1298 impliquées dans la défense de l'organisme via l'activation de voies de signalisation
1299 intracellulaires (Larsen et al., 2019; Sackton et al., 2007; Satyavathi et al., 2014). Elle constitue
1300 le système de défense le plus important chez les insectes, notamment de par sa vitesse d'action.
1301 A titre d'exemples, la réponse humorale est détectée dans les 24 à 48 heures post-infection orale
1302 chez *Bombus pascuorum*, et dès 2 heures chez *Anopheles stephensis* (Larsen et al., 2019). Les
1303 deux voies immunitaires les plus décrites sont des voies conservées appelées « Toll » et « IMD
1304 » (Immune deficiency), principalement activées par des infections fongiques et bactériennes
1305 (Bier & Guichard, 2012; Cooper, Chamberlain, & Lowenberger, 2009; Imler & Hoffmann,
1306 2001; Larsen et al., 2019; Lemaitre & Hoffmann, 2007; Sackton et al., 2007; Viljakainen,
1307 2015). D'autres voies de signalisation existent également, telles que les voies c-Jun N-terminal
1308 Kinase (JNK) et JAK/STAT, impliquées majoritairement dans des réponses antivirales et
1309 antifongiques (Agaisse & Perrimon, 2004; Bang, 2019; Geng et al., 2016; Larsen et al., 2019;
1310 Tafesh-Edwards & Eleftherianos, 2020).

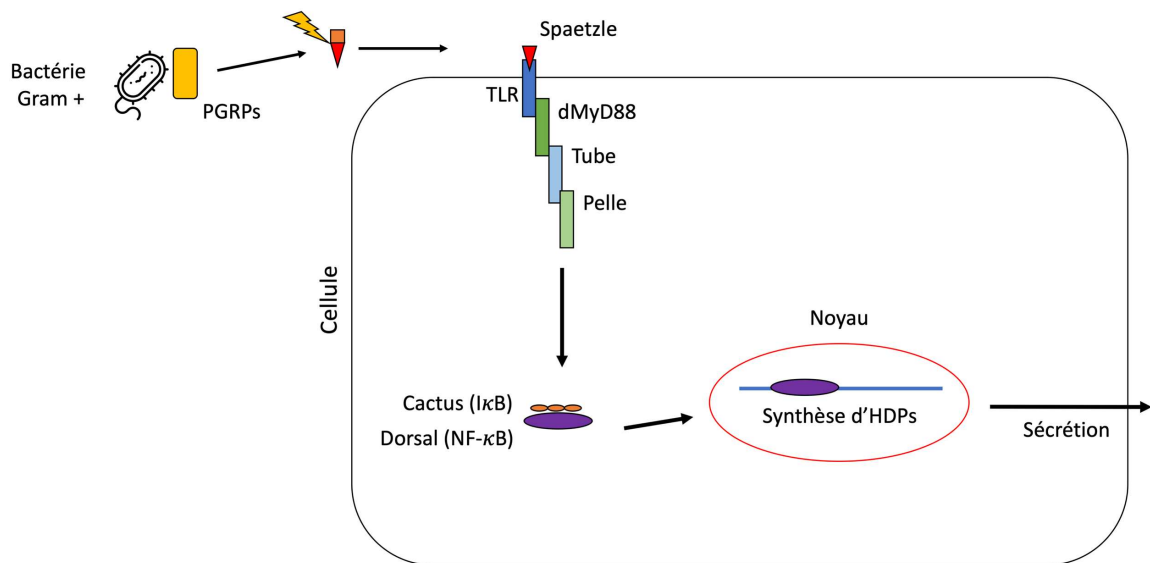
1311 Les HDPs produits sont généralement très conservés et sont sécrétés dans l'hémolymphe.
1312 Cette production, consécutive à l'activation des gènes, se retrouve dans différents organes tels
1313 que le corps gras (analogue du foie des mammifères) ainsi que les organes en contact avec le
1314 milieu extérieur (systèmes digestif, respiratoire et reproducteur) (Kimbrell & Beutler, 2001;
1315 Satyavathi et al., 2014; Tapadia & Verma, 2012; Tzou et al., 2000; K. Wu et al., 2016). Ainsi,
1316 il a été démontré chez *D. melanogaster* que des gènes codant pour différents HDPs
1317 s'exprimaient dans les mésentérons (équivalent de l'iléon chez les mammifères), dans les
1318 trachéoles, les glandes salivaires et labellaires, les conduits reproducteurs ainsi que dans les
1319 tubes de Malpighi (équivalent des reins chez les mammifères) suite à une infection par injection
1320 (Tzou et al., 2000). Les HDPs peuvent également être stockés dans les cellules (Kuhn-Nentwig,
1321 2003). L'activation des gènes est alors constitutive et c'est la sécrétion qui est contrôlée. Ceci
1322 a été démontré dans les hémocytes et les glandes salivaires de *Pseudacanthotermes spiniger*,
1323 où la libération dans l'hémolymphe de la termicine et la spinigérine s'effectue après un
1324 challenge immunitaire (Lamberty et al., 2001).

1325 4.3.1. La voie Toll

1326 Les récepteurs activant la voie Toll, appelés TLR pour Toll-Like Receptors ont été
1327 découverts initialement chez *D. melanogaster*. Cette voie de signalisation a été définie dans un
1328 premier temps comme étant impliquée dans l'établissement de la polarité dorso-ventrale. Elle
1329 est également impliquée dans le système immunitaire via la synthèse d'HDPs, notamment dans
1330 la réponse aux infections fongiques et de bactéries à Gram positif (Ali Mohammadie Kojour et
1331 al., 2020; Bier & Guichard, 2012).

1332 Dans le cas d'une infection bactérienne, il y aura tout d'abord reconnaissance des
1333 composants des peptidoglycanes de type Lys des parois bactériennes à Gram positif, par des
1334 récepteurs nommés PGRPs (Peptidoglycan Recognition Proteins). Ceci initiera des cascades
1335 protéolytiques extracellulaires qui provoqueront le clivage du facteur pro-Spaetzle. Ce ligand
1336 Spaetzle ainsi clivé se liera au TLR et activera alors la voie de signalisation intracellulaire Toll
1337 (Figure 26).

1338 Le TLR activé recrute en effet un complexe de protéines « DEATH-domain » (dMyD88,
1339 Tube et Pelle), ce qui résulte en la dissociation de la protéine inhibitrice de type IκB Cactus du
1340 facteur de transcription de type κB Dorsal (NF- κB) chez *Apis mellifera* (Larsen et al., 2019),
1341 ou Dif/Dorsal chez *D. melanogaster* (Bier & Guichard, 2012). Ce facteur de transcription est
1342 ensuite transloqué dans le noyau et active la transcription de gènes codant pour des HDPs, tels
1343 que la drosomycine et la metchnikowine chez *D. melanogaster*, ou encore la défensine-1 chez
1344 *A. mellifera* (Bier & Guichard, 2012; Kimbrell & Beutler, 2001; Larsen et al., 2019; Lourenço,
1345 Florecki, Simões, & Evans, 2018) (Figure 26).



1346

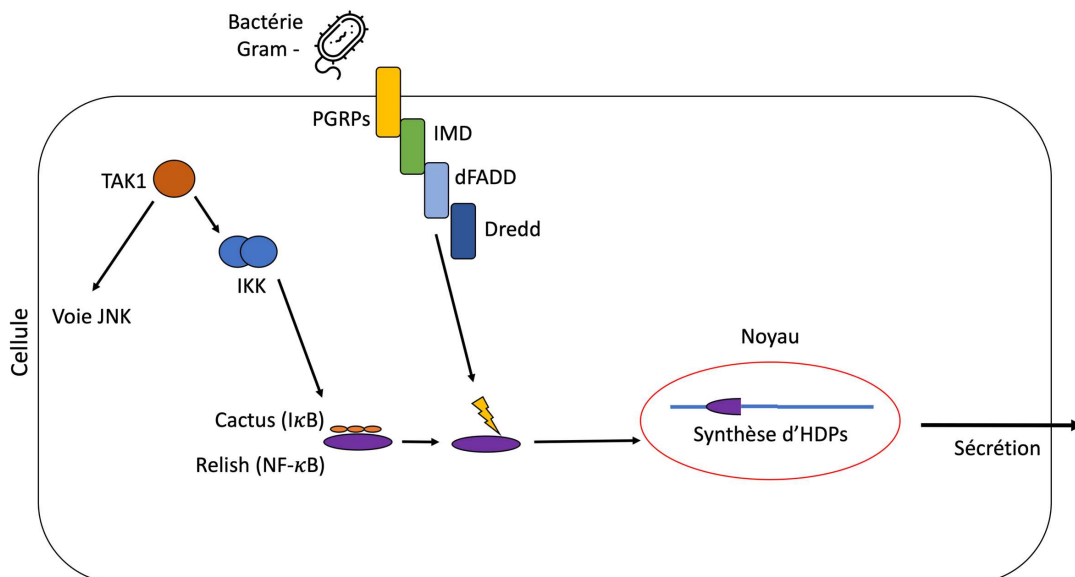
1347 **Figure 26** : Détails moléculaires de la voie de signalisation intracellulaire Toll chez *Apis mellifera* dans le
 1348 cas d'une infection bactérienne. La voie Toll est connue pour répondre majoritairement à des infections
 1349 causées par des bactéries à Gram positif. Il y a dans un premier temps reconnaissance du peptidoglycane de
 1350 type Lys des bactéries à Gram positif par des PGRPs, menant au clivage du facteur pro-Spaetzle *via* des
 1351 cascades protéolytiques extracellulaires. Le ligand Spaetzle se lie ensuite au Toll-Like Receptor (TLR). Ce
 1352 dernier recrute un complexe de protéines « DEATH-domain » (dMyD88, Tube et Pelle), menant à la
 1353 dissociation de la protéine inhibitrice de type IκB Cactus du facteur de transcription de type κB Dorsal (NF-
 1354 κB). Ce facteur de transcription est ensuite transloqué dans le noyau et active la transcription de gènes codant
 1355 pour des HDPs. Une fois synthétisés, les HDPs sont sécrétés dans l'hémolymphe.

1356

1357 4.3.2. La voie IMD

1358 Cette voie de signalisation contrôle la majorité de l'expression de HDPs chez *D.*
 1359 *melanogaster*, et constitue donc un élément central dans la réponse immunitaire contre les
 1360 microorganismes (Ali Mohammadie Kojour et al., 2020). Les PRRs de la voie IMD sont des
 1361 PGRPs et reconnaissent le peptidoglycane de type DAP, constitutif de la paroi des bactéries à
 1362 Gram négatif et de quelques bactéries à Gram positif telles que *Bacillus spp.* et *Listeria spp.* (Ali
 1363 Mohammadie Kojour et al., 2020; Hedengren-Olcott et al., 2004). Le nombre de PRRs et leur
 1364 appellation peut varier selon les espèces d'insectes étudiés. Chez la drosophile, on distingue
 1365 PGRP-LC et PGRP-LE tandis que seul PGRP-LC est retrouvé chez *A. mellifera* (Ali
 1366 Mohammadie Kojour et al., 2020; Evans et al., 2006; H. Schlüns & Crozier, 2007).

1367 A titre d'exemple, la reconnaissance des PAMPs par les PGRPs entraîne l'activation de
 1368 la protéine adaptatrice IMD. Il existe ensuite deux embranchements distincts menant à
 1369 l'activation du NF- κ B Relish. Un embranchement mène à la dissociation du NF- κ B de la
 1370 protéine I κ B Cactus via le complexe IKK, lui-même précédemment activé par la protéine
 1371 TAK1. Le second embranchement mène au clivage du NF- κ B par la protéine FADD,
 1372 précédemment activée par la caspase Dredd. Relish sera ensuite transloqué dans le noyau et
 1373 activera la transcription de gènes codant pour des HDPs tels que l'abaécine et l'hyménoptaécine
 1374 chez *A. mellifera*, ou la diptéricine chez *D. melanogaster* (Bier & Guichard, 2012; Cooper et
 1375 al., 2009; Costa, Jan, Sarnow, & Schneider, 2009; Kimbrell & Beutler, 2001; Satyavathi et al.,
 1376 2014; H. Schlüns & Crozier, 2007; Tzou et al., 2000) (Figure 27).



1377

1378 **Figure 27** : Détails moléculaires de la voie de signalisation intracellulaire IMD dans le cas d'une infection
 1379 bactérienne. La voie IMD est connue pour répondre majoritairement aux infections par des bactéries à Gram
 1380 négatif. La reconnaissance du peptidoglycane de type DAP des bactéries à Gram négatif par les PGRPs
 1381 entraîne l'activation de la protéine adaptatrice IMD. Il existe ensuite deux embranchements distincts menant
 1382 à l'activation du NF- κ B Relish. Un embranchement mène à la dissociation du NF- κ B de la protéine I κ B
 1383 Cactus via le complexe IKK, lui-même précédemment activé par la protéine TAK1. Cette dernière est
 1384 également connue pour être impliquée dans la voie de signalisation cellulaire JNK. Le second embranchement
 1385 mène au clivage du NF- κ B par la protéine FADD, précédemment activée par la caspase Dredd. Relish est
 1386 ensuite transloqué dans le noyau et active la transcription de gènes codant pour des HDPs. Suite à leur
 1387 synthèse, les HDPs sont sécrétés dans l'hémolymphe.

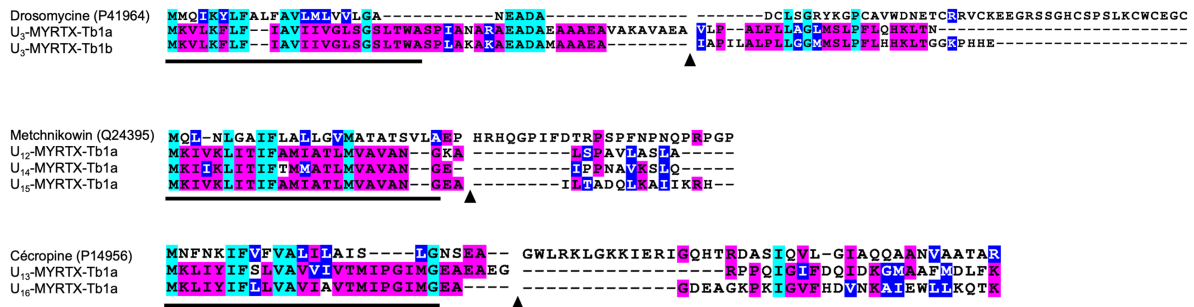
1388 4.3.3. Host Defense Peptides de *Drosophila melanogaster* et peptides de venin de
1389 *Tetramorium bicarinatum*

1390 Les alignements effectués lors de l'étude des précurseurs de *T. bicarinatum* ont montré que
1391 la drosomycine, la metchnikowin et la cécropine B issus de l'hémolymph de *D. melanogaster*
1392 se positionnent dans les mêmes clusters que certains précurseurs de toxines peptidiques des
1393 superfamilles A (i.e. U₃) et B (i.e. U₁₂, U₁₃, U₁₄, U₁₅ et U₁₆) (Touchard et al., 2018). La
1394 metchnikowin et la cécropine B sont des peptides linéaires et cationiques à action cytolytique
1395 dont les synthèses sont respectivement activés par les voies Toll et IMD suite à une infection
1396 fongique ou bactérienne (Gram positive et négative) (Brady, Grapputo, Romoli, & Sandrelli,
1397 2019; Ekengren & Hultmark, 1999; Tonk et al., 2019; Q. Wu, Patočka, & Kuča, 2018). La
1398 drosomycine est également un peptide cationique dont la synthèse est activée à la fois par la
1399 voie Toll et la voie IMD suite à une infection fongique (Ferrandon et al., 1998; Hanson &
1400 Lemaitre, 2020), et présente une structure plus complexe en raison de la présence de 4 ponts
1401 disulfures. Ces derniers lui confèrent une conformation CS $\alpha\beta$ similaire à celle adoptée par les
1402 défensines (Q. Wu et al., 2018; Zhang & Zhu, 2009).

1403 Les précurseurs de ces HDPs sont composés d'une séquence signal longue d'une vingtaine
1404 d'acides aminés, suivie d'une région propeptide courte (i.e. 2 à 5 acides aminés) puis de la
1405 partie mature sécrétée (Figure 28). Bien que montrant de faibles identités de séquences (i.e. 14
1406 à 20%) avec les précurseurs de toxines du venin de *T. bicarinatum* (Touchard et al., 2018), les
1407 précurseurs de HDPs montrent la même structure. Le caractère linéaire et cationique des
1408 séquences matures est également retrouvé chez de nombreuses toxines du venin de cette fourmi.

1409 Les phénomènes de recrutement de gènes de l'immunité afin de produire de nouvelles
1410 toxines de venin ont déjà été démontrés chez l'ornithorynque pour les toxines de type défensine
1411 (Cf. partie 1.4.2) (Whittington et al., 2008). Il a également été montré chez *Apis cerana* que le
1412 gène codant le peptide du venin, AcSecapin-1, s'exprime en dehors du système vulnérant, et
1413 est surexprimé en cas de challenge immunitaire, montrant que cette toxine est un effecteur de
1414 l'immunité innée (Lee, Kim, Yoon, Choi, & Jin, 2016). Ainsi, compte tenu des activités
1415 bactériolytiques et immunomodulatrices présentées respectivement par les peptides
1416 Bicarinaline et P17 du venin de *T. bicarinatum* (Benmoussa et al., 2017; Rifflet et al., 2012), et
1417 de l'expression des gènes les codant en dehors du système vulnérant (Cf. partie 2.5 de ce
1418 chapitre), il semble probable que certains peptides du venin de cette fourmi soient en lien avec
1419 le système immunitaire inné.

1420



1421
 1422 **Figure 28** : Structure et séquences des précurseurs des Host Defense Peptides de l'hémolymphe de
 1423 *Drosophila melanogaster* et des toxines du venin de *Tetramorium bicarinatum*. Les résidus conservés sont
 1424 surlignés en cyan tandis que les résidus identiques sont surlignés en magenta. Les résidus similaires sont
 1425 surlignés en bleu. Les régions signal sont soulignées en noir tandis que le triangle noir symbolise le clivage
 1426 entre la région propeptide et le peptide mature. Modifié à partir de (Touchard et al., 2018).

1427 5. Objectifs généraux

1428 Les venins de Formicidae constituent une source importante de peptides bioactifs. Seuls
 1429 cinq peptidomes complets ont été caractérisés grâce la méthode intégrative vénomique chez des
 1430 fourmis appartenant aux sous-familles phylogénétiques des *Paraponerinae*, *Ponerinae*,
 1431 *Myrmicinae* et *Myrmeciinae* (Aili et al., 2020; Kazuma et al., 2017; Mariano et al., 2019; Radis-
 1432 Baptista et al., 2020; Robinson et al., 2018; Touchard et al., 2018). Ces études ont révélé que
 1433 les toxines peptidiques décrites dans ces venins possédaient des séquences signal similaires à
 1434 celles trouvées dans les venins d'autres Hyménoptères Aculéates. Ceci suggère que les toxines
 1435 peptidiques de ces organismes appartiendraient à une même superfamille de gènes, les
 1436 Aculeatoxines, et auraient donc dérivé d'un nombre réduit de précurseurs (Robinson et al.,
 1437 2018). L'étude du venin de *Tetramorium bicarinatum* a également montré l'existence de deux
 1438 autres superfamilles de précurseurs (Touchard et al., 2018). De plus, les gènes codant pour
 1439 certaines myrmecitoxines du venin de cette fourmi s'expriment en dehors du système vulnérant,
 1440 suggérant l'existence d'un lien entre certaines toxines du venin et d'autres fonctions
 1441 physiologiques telles que l'immunité innée.

1442 Ces résultats posent les questions des mécanismes impliqués dans la diversification des
 1443 toxines peptidiques de venins de fourmis, ainsi que de leur potentiel rôle en dehors de la
 1444 fonction venimeuse. Pour répondre à ces problématiques, le premier objectif de ce travail de
 1445 thèse est d'apporter un aperçu de la diversité peptidique contenue dans les venins de fourmis
 1446 myrmicines. Pour ceci, les peptidomes des venins de sept fourmis appartenant à la sous-famille
 1447 des *Myrmicinae* et d'une fourmi appartenant à la sous-famille des *Pseudomyrmecinae* ont été

1448 caractérisés. Ce travail a mené à l'identification de 100 toxines matures et de leurs précurseurs.
1449 Ces résultats font l'objet d'un chapitre intitulé « Diversité moléculaire de peptides de venins de
1450 fourmis myrmicines » et ont mené à l'écriture de trois articles dont un en préparation :

1451 ➤ **Article 1 : Barassé, V.**; Touchard, A.; Téné, N.; Tindo, M.; Kenne, M.; Klopp, C.;
1452 Dejean, A.; Bonnafé, E.; Treilhou, M. The peptide venom composition of the fierce
1453 stinging ant *Tetraponera aethiops* (Formicidae: Pseudomyrmecinae). *Toxins (Basel)*.
1454 **2019, 11.**

1455 ➤ **Article 2 :** Touchard, A.; Aili, S.R.; Téné, N.; Barassé, V.; Klopp, C.; Dejean, A.; Kini,
1456 R.M.; Mrinalini, M.; Coquet, L.; Jouenne, T.; et al. Venom peptide repertoire of the
1457 European myrmicine ant *Manica rubida*: identification of insecticidal toxins. *J.*
1458 *Proteome Res.* **2020**, acs.jproteome.0c00048.

1459 ➤ **Article 3 : Barassé, V.** ; Téné, N. ; Klopp, C. ; Tysklind, N. ; Lalägue, H. ; Orivel, J. ;
1460 Troispoux, V. ; Petitclerc, F. ; Tindo, M. ; Kenne, M. ; Treilhou, M. ; Bonnafé, E. ;
1461 Touchard, A. Myrmicine ant venoms: first insights into their peptidic diversity. (*in*
1462 *prep.*)

1463 Le second objectif de ce travail de thèse consiste à explorer le lien potentiel entre les toxines
1464 de venin et le système immunitaire inné individuel de la fourmi *T. bicarinatum*. Cette étude a
1465 mis en évidence la présence de transcrits dans des organes impliqués dans l'immunité innée des
1466 insectes, ainsi qu'une production de deux peptides dans le corps gras suite à un challenge
1467 immunitaire. Ces résultats ont mené à l'écriture du troisième chapitre de ce manuscrit : « Lien
1468 entre les toxines peptidiques et la fonction immunitaire ».

1469

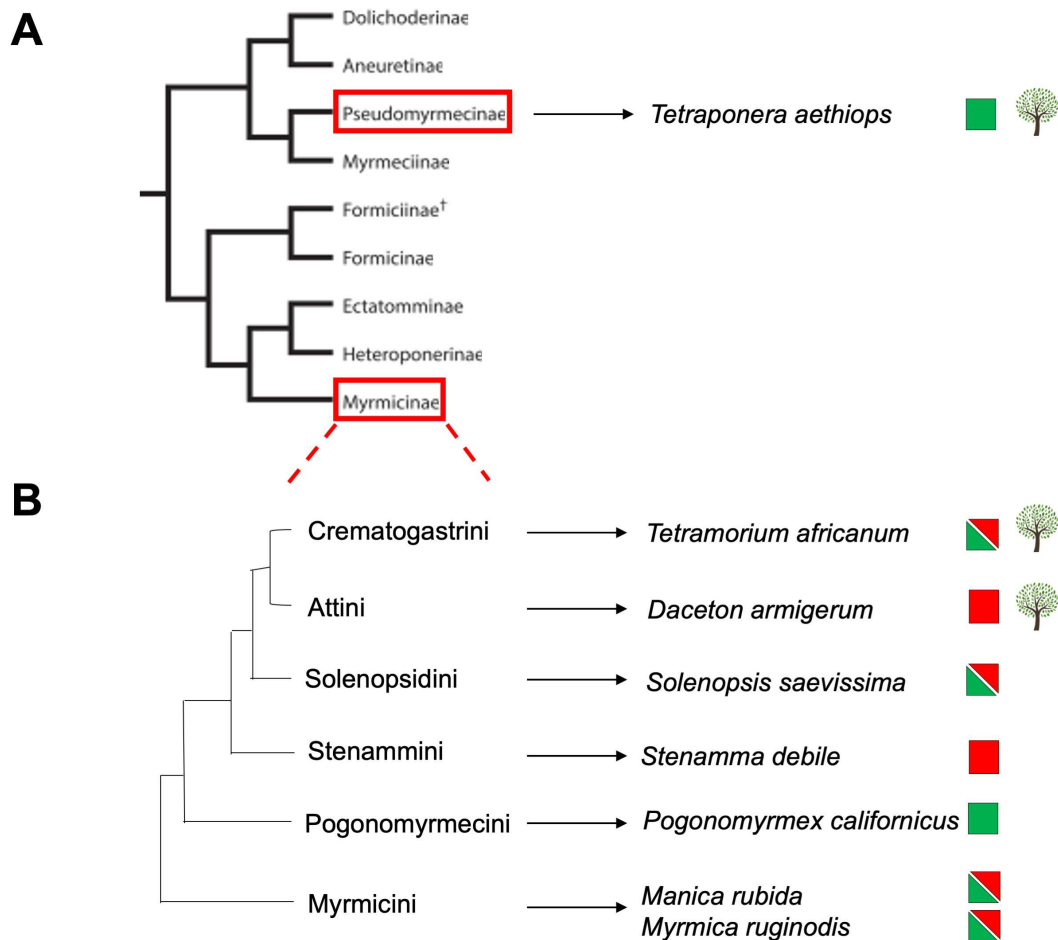
1471 **Chapitre II : Diversité moléculaire de peptides de venins de fourmis** 1472 **myrmicines**

1473
1474 L'objectif de ce chapitre est d'apporter un premier aperçu de la diversité peptidique des
1475 venins de fourmis myrmicines, une sous-famille de fourmis jusqu'à présent délaissée dans les
1476 études vénomiques. La sous-famille des *Myrmicinae* constituent de loin la sous-famille la plus
1477 importante en nombre d'espèces (i.e. 6500 décrites à ce jour). Elles représentent ainsi un clade
1478 hyperdiversifié, occupant la plupart des habitats terrestres et présentant de grandes variations
1479 dans leur écologie, la structure de leur colonie, leurs caractéristiques morphologiques et leur
1480 régime alimentaire (Ward, Brady, Fisher, & Schultz, 2015). La diversification de cette sous-
1481 famille a débuté il y a environ 100 millions d'années, et une analyse phylogénétique moléculaire
1482 a récemment permis d'organiser les espèces de *Myrmicinae* en six tribus monophylétiques (i.e.
1483 Crematogastrini, Attini, Solenopsidini, Stenammini, Pogonomyrmecini, Myrmicini) (Ward et
1484 al., 2015) (Figure 29). Malgré l'omniprésence de cette sous-famille de fourmis à travers la
1485 planète, un seul peptidome complet de venin de fourmi myrmicine a été décrit à l'heure actuelle
1486 (Touchard et al., 2018) (Cf. parties 2.5 et 3 – État de l'art). Dans ce chapitre nous avons cherché
1487 à caractériser les profils peptidiques des venins de nouvelles espèces de fourmis myrmicines en
1488 nous appuyant sur un échantillonnage rationnel prenant en compte la diversité phylogénétique.
1489 Ainsi, un ou deux représentants de chaque tribu des *Myrmicinae* ont été sélectionnés et récoltés,
1490 en plus d'une espèce de fourmi pseudomyrmecine, une sous-famille proche. (Figure 29). Les
1491 espèces ont été essentiellement choisies sur des critères d'accessibilité en termes de récoltes, et
1492 présentent une gamme variée d'écologie et d'utilisation de leur venin. Si certaines des espèces
1493 sélectionnées ont des venins essentiellement défensifs (e.g. *Tetraponera aethiops*,
1494 *Pogonomyrmex californicus*), d'autres utilisent majoritairement leur venin dans un but de
1495 prédation (i.e. *Stenamma debile*, *Daceton armigerum*) (Figure 29).

1496 La fourmi *Tetramorium africanum* représente la tribu la plus diversifiée en termes de
1497 nombre d'espèces, les Crematogastrini (Borowiec et al., 2020; Ward et al., 2015). Originaires
1498 des régions tropicales d'Afrique de l'Ouest, cette fourmi est territoriale et arboricole. Elle
1499 construit son nid à partir des feuilles de son arbre-hôte (e.g. *Lophira alata*, *Anthocleista vogelii*),
1500 qu'elle protège des herbivores notamment grâce à un comportement agressif et une piqûre
1501 défensive très douloureuse (Dejean et al., 2016).

1502 La fourmi *Daceton armigerum* est une fourmi arboricole néotropicale de la tribu des
1503 Attini. Cette espèce utilise son venin uniquement dans le but de paralyser les proies de grandes

1504 tailles difficiles à maîtriser par la seule force des mandibules. Elle possède en effet des
 1505 mandibules hypertrophiées (i.e. trap-jaw) très efficaces pour capturer des proies. Le venin est
 1506 rarement utilisé dans un but défensif et occasionne seulement une légère démangeaison
 1507 lorsqu'il inoculé à l'Homme (Dejean et al., 2012).



1508
 1509 **Figure 29** : Modèles d'études, positionnement dans la famille des Formicidae et détails de leur écologie. (A)
 1510 Position des sous-familles des *Myrmicinae* et des *Pseudomyrmecinae* dans le clade des Formicoïdes. (B)
 1511 Détail des tribus phylogénétiques composant les *Myrmicinae* et appartenances phylogénétiques des espèces
 1512 étudiées. Les fourmis arboricoles sont symbolisées par l'arbre. Les carrés rouges dénotent une utilisation
 1513 essentiellement offensive du venin (i.e. prédation) tandis que les carrés verts indiquent une utilisation
 1514 essentiellement défensive. Les carrés contenant les deux couleurs indiquent donc un usage double du venin
 1515 (i.e. prédation et défense).

1516 La fourmi de feu *Solenopsis saevissima*, représentante des Solenopsidini, est une espèce
 1517 de fourmi agressive. Très abondante dans les milieux ouverts de la forêt amazonienne (e.g.
 1518 savanes, bords de rivières), elle est considérée en Amérique du Sud comme un ravageur (Dejean
 1519 et al., 2015; Fox et al., 2010). Son venin, très douloureux et allergène pour les vertébrés, est

1520 riche en composés alcaloïdiques. Les allergènes identifiés sont des protéines et certaines études
1521 suggèrent également la présence de peptides (Dos Santos Pinto et al., 2012).

1522 La tribu des Pogonomyrmecini montre quant à elle un nombre réduit de genres,
1523 comprenant pour la plupart des espèces granivores. L'espèce sélectionnée pour cette tribu,
1524 *Pogonomyrmex californicus*, est une fourmi de taille moyenne, vivant dans les régions
1525 désertiques des États-Unis (Californie, Arizona). Les fourmis du genre *Pogonomyrmex* utilisent
1526 un venin douloureux et très toxique pour les vertébrés afin de protéger leurs réserves de graines
1527 des rongeurs (J. O. Schmidt & Blum, 1978a; P. J. Schmidt, Sherbrooke, & Schmidt, 1989). Un
1528 effet hallucinogène du venin de *P. californicus* a été noté, suscitant son utilisation dans des
1529 contextes thérapeutiques et religieux par les groupes indigènes de Californie du Sud (Groark,
1530 1996; Helmkampf, Mikheyev, Kang, Fewell, & Gadau, 2016). Cette particularité suggère que
1531 certaines toxines sont capables d'interagir avec des récepteurs du système nerveux de l'Homme.

1532 L'espèce *Stenamma debile* est une fourmi de petite taille, cryptique des milieux
1533 forestiers des régions tempérées septentrionales. Cette espèce discrète utilise son venin
1534 uniquement pour chasser et capturer de petits arthropodes de la litière (Branstetter, 2012;
1535 Lenoir, Khalil, Châline, & Hefetz, 2018).

1536 La tribu des Myrmicini est également endémique des régions tempérées Nord et
1537 regroupe les deux genres *Manica* et *Myrmica*. Ces fourmis communément appelées « fourmis
1538 rouges » en Europe, utilisent leur venin dans des contextes de défense, bien qu'il soit également
1539 utilisé pour la capture des proies (Borowiec et al., 2020; Heep, Klaus, et al., 2019; Heep,
1540 Skaljic, et al., 2019; Lenoir, Devers, Marchand, Bressac, & Savolainen, 2010).

1541 La fourmi *Tetraoponera aethiops* (*Pseudomyrmecinae*) est une espèce arboricole ayant
1542 développé une relation mutualiste avec une plante hôte (i.e. myrmécophyte), *Barteria fistulosa*
1543 (*Passifloraceae*). Cette espèce utilise ainsi essentiellement son venin pour protéger le
1544 myrmécophyte contre les herbivores vertébrés et invertébrés. La caractérisation de ce
1545 peptidome a mené à l'identification de 9 précurseurs de pseudomyrmécitoxines, dont cinq
1546 toxines linéaires, trois homodimères et un hétérodimère. Toutes les pseudomyrmécitoxines
1547 identifiées sont polycationiques. Certaines d'entre elles ont de plus montré des similarités de
1548 séquences avec des myrmécitoxines connues pour avoir des activités insecticides et
1549 nociceptives (Dekan et al., 2017; Robinson et al., 2018). L'analyse des région prépro a
1550 également montré que toutes les pseudomyrmécitoxines du venin de *T. aethiops* appartiennent
1551 à la superfamille A, précédemment décrite chez *Tetramorium bicarinatum* (Touchard et al.,
1552 2018) (Tableau III).

1553 La caractérisation des peptidomes des sept venins de fourmis myrmicines a mené à
1554 l'identification de 91 précurseurs de myrmicitoxines. Ces venins sont en majorité dominés par
1555 des peptides linéaires, amphiphiles et cationiques. Néanmoins, plusieurs myrmicitoxines
1556 présentent également un à trois ponts disulfures. Parmi celles-ci, quatre peptides possèdent un
1557 domaine de type EGF (Epidermal Growth Factor), et deux d'entre elles présentent également
1558 une O-glycosylation post-traductionnelle. Ce type de toxine n'avait jamais été décrit dans les
1559 venins de fourmis. Notre étude a également permis d'identifier pour la première fois dans cette
1560 sous-famille phylogénétique la présence de toxines dimériques dans le venin de la fourmi *D.*
1561 *armigerum*.

1562 Sur la base de leurs séquences matures, trente-sept familles de myrmicitoxines ont été
1563 définies. Cette répartition a permis de rendre compte de la diversité des toxines peptidiques
1564 présentes dans les venins des *Myrmicinae*, tout en mettant en évidence les compositions
1565 hétérogènes de ces venins (Table III). L'analyse des séquences signal a permis de répartir ces
1566 91 myrmicitoxines en 8 superfamilles de précurseurs (i.e. A₁, A₂, A₃, A₄, B₁, B₂, C₁ et C₂).

1567 Ces résultats ont conduit à la publication de deux articles, et à la préparation d'un
1568 troisième :

1569

1570 ➤ **Barassé, V.**; Touchard, A.; Téné, N.; Tindo, M.; Kenne, M.; Klopp, C.; Dejean, A.;
1571 Bonnafé, E.; Treilhou, M. The peptide venom composition of the fierce stinging ant
1572 *Tetraponera aethiops* (Formicidae: Pseudomyrmecinae). *Toxins (Basel)*. **2019**, *11*.

1573

1574 ➤ Touchard, A.; Aili, S.R.; Téné, N.; **Barassé, V.**; Klopp, C.; Dejean, A.; Kini, R.M.;
1575 Mrinalini, M.; Coquet, L.; Jouenne, T.; et al. Venom peptide repertoire of the european
1576 myrmicine ant *Manica rubida*: identification of insecticidal toxins. *J. Proteome Res.*
1577 **2020**, *19*.

1578

1579 ➤ **Barassé, V.**; Téné, N. ; Klopp, C. ;Tysklind, N. ; Lalägue, H. ; Orivel, J. ; Troispoux,
1580 V. ; Petitclerc, F. ; Tindo, M. ;Kenne, M. ; Treilhou, M. ; Bonnafé, E. ; Touchard, A.
1581 Myrmicine ant venoms : first insights into their peptidic diversity. (*in prep.*)

1582

1583 **Tableau III** : Composition des peptidomes de venins de fourmis pseudomyrmecine et myrmicines. PTMs : modifications post-traductionnelles.

Espèces	Familles de toxines matures	Monomères			Homodimères		Hétérodimères	PTMs	Superfamilles de précurseurs
		0 CC	1 CC	3 CC	1 CC	3CC	1 CC		
		<i>Tetraponera aethiops</i>	5	5			2		
<i>Manica rubida</i>	9	9	3	1				C-ter NH ₂ , O-glycosylation	A ₁ , A ₂ , A ₃ , A ₄ , B ₁ , B ₂ , C ₁ , C ₂
<i>Myrmica ruginodis</i>	8	11	10	4				C-ter NH ₂ , O-glycosylation	A ₁ , A ₂ , A ₃ , A ₄ , B ₁ , B ₂ , C ₁ , C ₂
<i>Pogonomyrmex californicus</i>	7	11	3	2				C-ter NH ₂	A ₁ , A ₂ , A ₃ , B ₂ , C ₁ , C ₂
<i>Stenamma debile</i>	5	4	2					C-ter NH ₂	A ₁ , A ₂ , A ₃ , C ₁
<i>Tetramorium africanum</i>	18	14	13					C-ter NH ₂	A ₁ , A ₂ , A ₃ , B ₁ , B ₂ , C ₁
<i>Daceton armigerum</i>	1				3		1	-	A ₃

1584

Article

The Peptide Venom Composition of the Fierce Stinging Ant *Tetraponera aethiops* (Formicidae: Pseudomyrmecinae)

Valentine Barassé ^{1,*} , Axel Touchard ¹, Nathan Téné ¹, Maurice Tindo ² , Martin Kenne ², Christophe Klopp ³, Alain Dejean ^{4,5}, Elsa Bonnafé ¹ and Michel Treilhou ¹

¹ EA-7417, Institut National Universitaire Champollion, Place de Verdun, 81012 Albi, France;

axel.touchard2@gmail.com (A.T.); nathan.tene@univ-jfc.fr (N.T.); elsa.bonnafe@univ-jfc.fr (E.B.);

michel.treilhou@univ-jfc.fr (M.T.)

² Laboratory of Animal Biology and Physiology, Faculty of Science, University of Douala, P.O. Box 24157, Douala, Cameroon; tindodouala@yahoo.com (M.T.); medoum68@yahoo.fr (M.K.)

³ Unité de Mathématique et Informatique Appliquées de Toulouse, UR0875, INRA Toulouse, 31326 Castanet-Tolosan, France; christophe.klopp@inra.fr

⁴ Ecolab, Université de Toulouse, CNRS, INPT, UPS, 31400 Toulouse, France; alain.dejean@wanadoo.fr

⁵ CNRS, UMR EcoFoG, AgroParisTech, Cirad, INRA, Université des Antilles, Université de Guyane, 97310 Kourou, France

* Correspondence: valentine.barasse@gmail.com; Tel.: +33-563-481-700

Received: 24 October 2019; Accepted: 11 December 2019; Published: 14 December 2019



Abstract: In the mutualisms involving certain pseudomyrmicine ants and different myrmecophytes (i.e., plants sheltering colonies of specialized “plant-ant” species in hollow structures), the ant venom contributes to the host plant biotic defenses by inducing the rapid paralysis of defoliating insects and causing intense pain to browsing mammals. Using integrated transcriptomic and proteomic approaches, we identified the venom peptidome of the plant-ant *Tetraponera aethiops* (Pseudomyrmecinae). The transcriptomic analysis of its venom glands revealed that 40% of the expressed contigs encoded only seven peptide precursors related to the ant venom peptides from the A-superfamily. Among the 12 peptide masses detected by liquid chromatography-mass spectrometry (LC–MS), nine mature peptide sequences were characterized and confirmed through proteomic analysis. These venom peptides, called pseudomyrmecitoxins (PSDTX), share amino acid sequence identities with myrmeciitoxins known for their dual offensive and defensive functions on both insects and mammals. Furthermore, we demonstrated through reduction/alkylation of the crude venom that four PSDTXs were homo- and heterodimeric. Thus, we provide the first insights into the defensive venom composition of the ant genus *Tetraponera* indicative of a streamlined peptidome.

Keywords: defensive venom; dimeric peptides; peptidome; *Tetraponera aethiops*

Key Contribution: The presence of dimeric scaffolds, found in the venom of all plant-ant species investigated to date, suggests that they are an important feature in defense.

1. Introduction

Venoms are biochemical arsenals developed by animals to defend themselves and/or capture prey [1]. Studies on arthropod (e.g., scorpions, spiders, centipedes, and insects) venoms show a majority of proteins and peptides which exhibit variable amino acid sequences and tridimensional structures in their mature form [2,3]. Although ants are a dominant feature in terms of number of individuals and biomass in most terrestrial ecosystems, few extensive studies have been conducted on

their venoms. Most of these studies have focused on emblematic and/or problematic species whose stings are painful and can cause allergies [4–6], with the major goal being to relieve the symptoms prompted by these venoms. Consequently, the allergenic venoms of fire ants of the genus *Solenopsis* (Myrmicinae), *Pachycondyla* spp. (Ponerinae), or Australian *Myrmecia* (Myrmeciinae) have especially been studied [5–9]. Also, the painful venoms of the bullet ant *Paraponera clavata* and of ponerine ants of the genera *Dinoponera* and *Neoponera* have been the subject of several studies which led to the isolation and characterization of tens of bioactive peptides [10,11].

In the past, the investigation of ant venoms was greatly limited by the difficulty in gathering large amounts of venom due to the small size of these insects. However, the use of new techniques to collect venoms now permits large amounts of ant venom to be quickly obtained [12,13]. Furthermore, the use of multi-omics strategies to study ant venoms recently revealed the whole peptidome of three ant species belonging to the subfamilies Ponerinae, Myrmicinae, and Myrmeciinae, allowing a high-throughput identification of novel peptides. The rise of such an integrated methodology now allows us to undertake the in-depth exploration of new ant venoms. The comprehensive inclusion of ant species from different subfamilies and with different ecologies should enhance our understanding of the molecular diversification of ant venom peptidomes and then lead to further discoveries.

To date, only six peptides have previously been identified from Pseudomyrmecinae venoms [14], but earlier studies revealed that venoms from this subfamily are rich in peptides even though the peptide composition is greatly influenced by both the hunting habits and the nesting mode of the species [15]. Thus, in order to contribute to the cataloguing of peptide toxins from the main lineage of ants, the present study focuses on the description of the venom peptidome of *Tetraoponera aethiops*, Smith, F., 1877 (Pseudomyrmecinae). *Tetraoponera aethiops* is an arboreal plant-ant involved in an obligatory mutualistic association with the myrmecophyte *Barteria fistulosa* (Passifloraceae), myrmecophytes being plants sheltering colonies of specialized “plant-ant” species in hollow structures called domatia [16]. Here, the colonies, housed in the plant’s hollow twigs, feed on honeydew exudates from coccids tended inside the domatia and on symbiotic fungi. In return, the workers fiercely protect the host myrmecophyte from competing vines, and herbivorous insects and mammals [17–19]. Known to be very painful to humans, *T. aethiops* venom is rarely used to capture prey. Instead, it is essentially employed to defend its host tree since most of the stung, paralyzed, or killed defoliating insects are discarded rather than being retrieved to be eaten [20]. We first aimed to verify if the venom of this species contains dimeric peptides as is known for the pseudomyrmecine plant-ant *Pseudomyrmex* [14,15]. Based on previous investigations on defensive venoms [1], we hypothesized that such a specialized mutualistic interaction between *T. aethiops* and *B. fistulosa* is likely to have strongly affected the ant venom composition in favor of the production of dimeric peptides to ensure host plant protection rather than prey capture.

2. Results

2.1. Mass Spectrometry of *Tetraoponera Aethiops* Venom

The initial LC–MS analysis was performed on an LCQ-ion trap Advantage mass spectrometer in order to establish the list of the masses of the most abundant peptides in *Tetraoponera aethiops* venom (Table 1). Before sequence determination, each peptide detected was tentatively named based on the initials of the genus and species followed by their molecular weight (i.e., Ta-XXXX) as described by Johnson et al. [21]. The fully sequenced peptides were then renamed in accordance with the nomenclature developed for venom peptides [22] and we used pseudomyrmecitoxin (PSDTX) to define the venom peptides of pseudomyrmecine ants [23]. The total ion chromatogram (TIC) analysis of the venom of *T. aethiops* revealed several peaks with 12 masses corresponding to peptides (Figure 1). These peptides were relatively large, exhibited molecular weights ranging from 2662 to 5774 Da, and eluted at retention times between 14.55 and 50.07 min—which is equivalent to 15 and 56% acetonitrile in the mobile phase. The relative abundance of these peptides is reported in Table 1. The venom peptidome was mostly dominated by a single peptide, U₄-PSDTX-Ta1a, which accounts for 66.41% of the venom

peptide composition. Interestingly, peaks eluting from 31 to 45 min may depend on non-peptide compounds such as large proteins (MW > 10 KDa). The peak eluting from 54 to 55 min was dominated by several low molecular weight compounds co-eluting with the twelfth peptide (U₅-PSDTX-Ta1a) found at trace levels (Table 1).

Table 1. Peptide mass fingerprinting of *Tetraponera aethiops* venom. List of peptide masses detected through LC-MS using an LCQ-ion trap Advantage mass spectrometer.

Retention Time (min)	Mass (Da)	Relative Abundance (%)	Temporary Name	Peptide Toxin
14.55	5773.80	0.50	Ta-5773	
20.50	2996.28	0.03	Ta-2996	
22.18	5458.56	0.60	Ta-5458	
22.24	2663.58	0.10	Ta-2662	U ₁ -PSDTX-Ta1a
23.00	2877.72	0.28	Ta-2875	U ₂ -PSDTX-Ta1a
23.64	5441.80	11.98	Ta-5438	U ₂ -PSDTX-Ta1b (homodimer)
25.23	5753.64	9.33	Ta-5750	U ₂ -PSDTX-Ta1a (homodimer)
25.50	4470.96	1.96	Ta-4468	U ₃ -PSDTX-Ta1a
26.00	5683.60	7.24	Ta-5680	U ₂ -PSDTX-Ta1a/U ₂ -PSDTX-Ta1c
26.75	5610.63	1.27	Ta-5608	U ₂ -PSDTX-Ta1c (homodimer)
50.07	3568.28	66.41	Ta-3566	U ₄ -PSDTX-Ta1a
54.93	3615.81	*	Ta-3615	U ₅ -PSDTX-Ta1a

* found at trace levels.

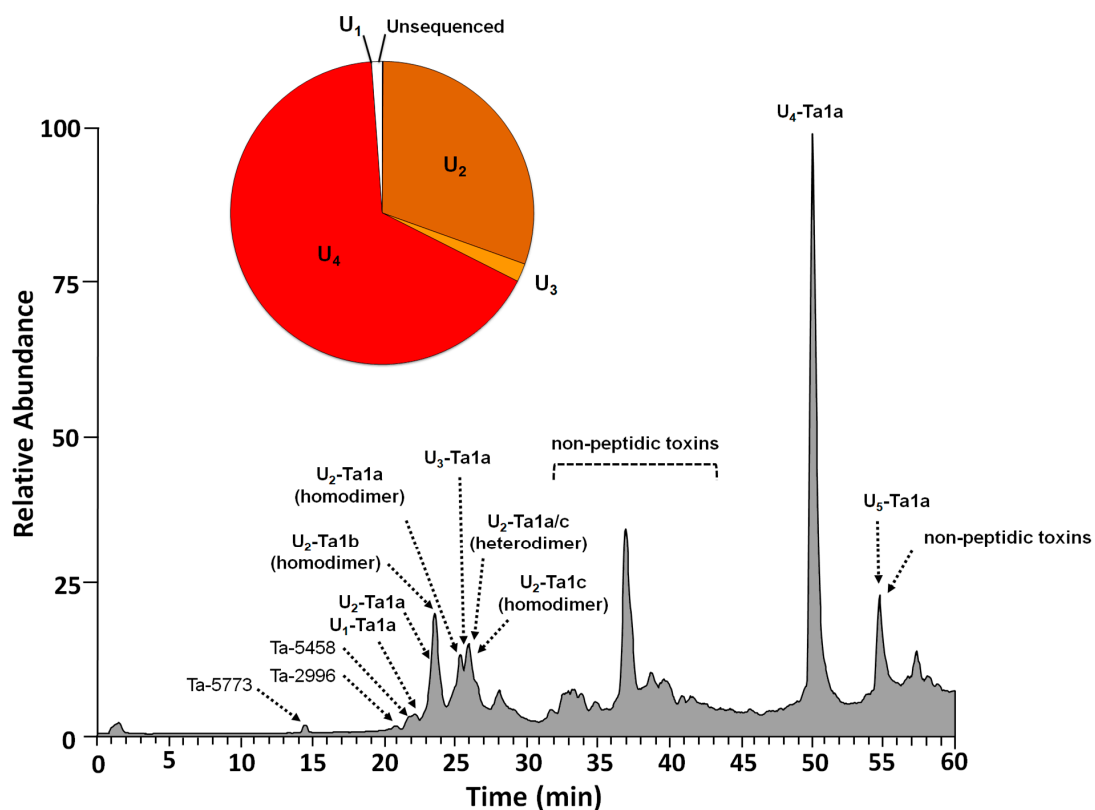


Figure 1. Positive mode total ion chromatogram (TIC) of *T. aethiops* venom using LCQ Advantage ESI mass spectrometer. Crude venom was separated by C₁₈ RP-HPLC using an ACN gradient. The mobile phase was 0.1% aqueous formic acid (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The peptides were eluted using a linear gradient from 0 to 50% of solvent B during 45 min, then from 50 to 100% during 10 min, and finally held for 5 min at a 250 $\mu\text{L min}^{-1}$ flow rate. Note that 'U₄-PSDTX-Ta1a' accounts for 66.41% of the venom peptide content.

The crude venom was submitted to LC–MS/MS using an Orbitrap mass spectrometer for a de novo sequencing which yielded 266 sequence tags with an ALC score higher than 60%. These MS/MS sequences were then used to confirm and identify the putative peptide sequences inferred from the transcriptomic analysis of the *T. aethiops* venom gland transcriptome. Orbitrap mass spectrometry analysis confirmed the presence of U₅-PSDTX-Ta1a in this venom, albeit in a low amount.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [24] via the PRIDE [25] partner repository with the dataset identifier PXD016337.

2.2. Venom Gland Transcriptome and Predictive Precursor Sequences

The RNA sequencing of the *T. aethiops* venom apparatus resulted in the acquisition of 188,907,225 demultiplexed raw reads with a length of 150-bp. They were assembled de novo using Oases graph assembler, which resulted in 16,047 contigs. Among them, 230 transcripts were expressed to a frequency higher than or equal to 100,000 hits and were subsequently addressed using the NCBI blastp program for functional annotations (Table S1). These annotated transcripts were classified into four categories (i.e., venom peptides, venom proteins, protein maturation, and others; Figure 2). This permitted us to deduce that venom peptide toxins have high transcription levels accounting for 40% of the most expressed transcripts by the venom glands (400,732 Reads Per Million (RPM)). Thus, we identified six putative venom peptide precursors. They shared an average of 52% identity, allowing us to define the following consensus sequence from the alignment of the prepro-regions: MXLSYXXLXLVXFSLAIXFXPXXXAXAXSVGMAEDAEXXALAESXANALADAXP.

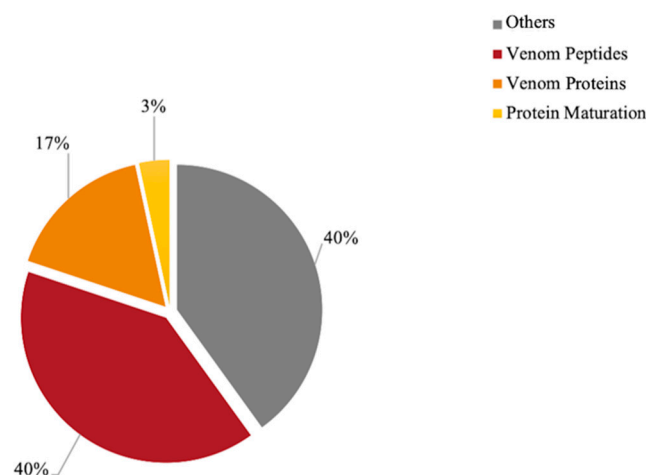


Figure 2. Proportions of addressed functions of the 230 most expressed transcripts ($\geq 100,000$ hits) from *T. aethiops* venom glands. Functional annotations were made with the NCBI blastp program. The category “Others” groups functions involved in cellular metabolism. Contigs coding for venom peptides accounted for 40% of the transcripts expressed by the venom glands. Seventeen percent of the transcripts coded for venom proteins such as venom allergens and phospholipases, and 3% of the transcripts were dedicated to protein maturation. The contigs names and functions are presented in detail in Supplementary Data (Table S1).

Tetraponera aethiops peptide precursors seemed to be related to other venom peptide precursors reported for ants. Indeed, this consensus sequence shared 43%, 35%, 52%, and 48% sequence identity with consensus prepro-sequences obtained from studies on *Odontomachus monticola* [26], *Myrmecia pilosula* [27], *Myrmecia gulosa* [28], and *Tetramorium bicarinatum* venoms [29], respectively (Figure 3). Unlike prepro-regions, mature regions of these precursors exhibited much more intra- and interspecific amino acid sequence variabilities (Figure 4). The complete cDNA sequences of venom peptide precursors from *Tetraponera aethiops* venom were submitted to GenBank (for accession numbers, see the legend to Figure 4).

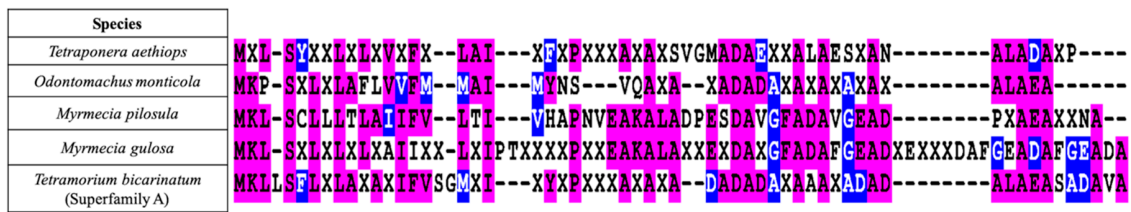


Figure 3. Alignment of consensus prepro sequences from pilosulin-like ant venom peptides. The alignment was generated with the Muscle program in Seaview version 4.6.1 and edited using BOXSHADE version 3.2. Identical residues are highlighted in magenta. Similar residues are highlighted in blue. The consensus sequence of the prepro-region of *Tetraponera aethiops* venom peptides shared 43%, 35%, 52%, and 48% with the consensus prepro-regions sequences obtained from previous studies on *Odontomachus monticola* [26], *Myrmecia pilosula* [27], *Myrmecia gulosa* [28], and *Tetramorium bicarinatum* (Superfamily A) [29], respectively.

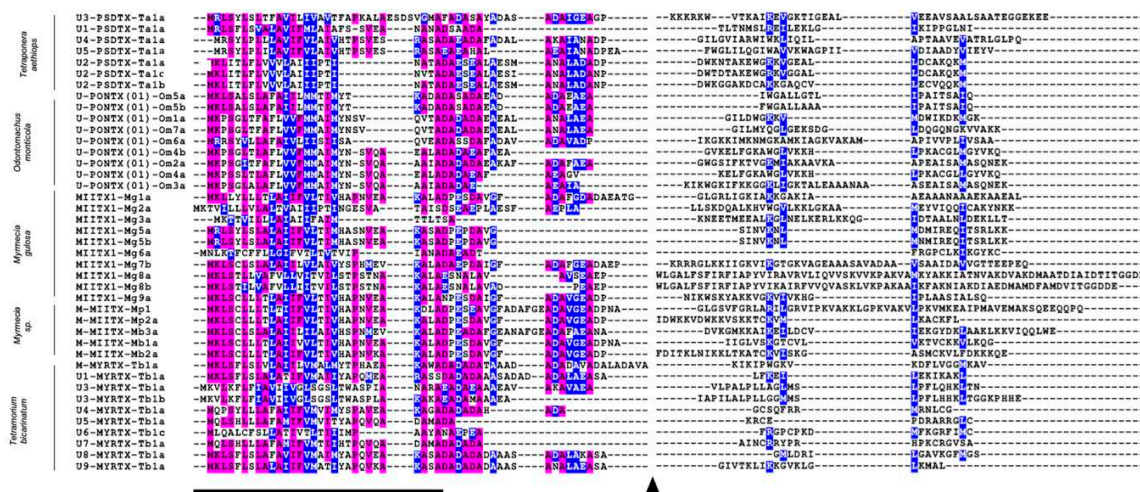


Figure 4. Alignments of *T. aethiops* venom peptide precursors with pilosuline-like peptides from *Odontomachus monticola* [26], *Myrmecia gulosa* [28], *Myrmecia pilosula* [27], and *Tetramorium bicarinatum* (Superfamily A) [29] venoms. Alignments were generated with the Muscle program in Seaview version 4.6.1 and edited using BOXSHADE version 3.2. The prepro- and mature regions were aligned separately. Identical residues are highlighted in magenta and similar residues are highlighted in blue. The black triangle indicates the cleavage site between the prepro-regions and the mature peptides. The black line marks the signal regions. Post-translational modifications are not shown. The prepro-regions showed themselves to be conserved, whereas mature peptide sequences were highly variable. *Tetraponera aethiops* venom peptide precursor cDNA sequences were submitted to GenBank, with the following accession numbers: U₁-PSDTX-Ta1a (MN607166), U₂-PSDTX-Ta1a (MN607169), U₂-PSDTX-Ta1b (MN607170), U₂-PSDTX-Ta1c (MN607168), U₃-PSDTX-Ta1a (MN607165), U₄-PSDTX-Ta1a (MN607167), and U₅-PSDTX-Ta1a (MN607171).

Several venom proteins were also expressed in *T. aethiops* venom glands including phospholipases and venom allergens accounting for 17% of the venom gland expression (164,978 RPMs). Additionally, 3% of the transcripts relied on protein maturation (33,782 RPMs) with enzymes implicated in the formation of disulfide bonds such as protein disulfide isomerase (3347 RPMs). In keeping with protein maturation, an angiotensin-converting enzyme (521 RPMs) and two types of venom Dipeptidylpeptidase IV (2356 RPMs and 736 RPMs) were identified within the transcripts. These enzymes are likely involved in the processing of bioactive peptides [30,31].

RNAseq data are available on the European Nucleotide Archive website (<https://www.ebi.ac.uk/ena>) under the following study accession number: PRJEB35699.

2.3. Molecular Features of Pseudomyrmecitoxins

The combination of LC–MS analysis, de novo sequencing-based mass spectrometry and the RNA sequencing of the venom glands allowed us to assign five linear peptide sequences (i.e., U₁-PSDTX-Ta1a, U₂-PSDTX-Ta1a, U₃-PSDTX-Ta1a, U₄-PSDTX-Ta1a and U₅-PSDTX-Ta1) to the twelve masses initially detected through LC–MS (Table 2). Additionally, the transcriptomic data led to the identification of three highly expressed peptide transcripts (i.e., 178,027; 148,472, and 144,674 RPMs) having very similar sequences with calculated masses in the range 2723–2877 Da and bearing an odd number of cysteines (i.e., one and three).

Table 2. Peptide sequences in the venom of *Tetraponera aethiops*. pI, isoelectric point. For each peptide, the transcripts frequency value (Reads Per Millions; RPMs) represents the frequency sum of all assembled contigs encoding the peptide precursor. “*” denotes C-terminal amidation.

Toxin Name	Mass (Da)	RPMs	Sequence	Features	Net Charge	Hydrophobic aa (%)	pI
U ₁ -PSDTX-Ta1a	2662.54	118,602	TLTNMSLRILEIKLGKIPPLNI	Monomer	1.0	41.67	8.26
U ₂ -PSDTX-Ta1a	2875.47	148,472	DWKNTAKEWGGKVGALLDCAKQKM *	Monomer	2.9	40.00	8.99
U ₂ -PSDTX-Ta1a	5750.94	148,472	DWKNTAKEWGGKVGALLDCAKQKM *	1 S-S Homodimer	5.8	40.00	9.23
U ₂ -PSDTX-Ta1b	5438.60	144,674	DWKGAKDCAKKAQCVCLECVQKQKM *	3 S-S Homodimer	5.6	44.00	8.83
U ₂ -PSDTX-Ta1c	5608.78	178,027	DWTDTAKEWGRKVGALLDCAKQKM *	1 S-S Homodimer	3.8	40.00	8.70
U ₂ -PSDTX-Ta1c	5680.86	178,027	DWTDTAKEWGRKVGALLDCAKQKM *	Heterodimer	4.8	40.00	9.04
U ₂ -PSDTX-Ta1a		148,472	DWKNTAKEWGGKVGALLDCAKQKM *				
U ₃ -PSDTX-Ta1a	4468.45	26,331	KKKRKVVTKAIKEVGTIGEALVEEAVSAALS AATEGGKEE	Monomer	1.0	38.10	8.27
U ₄ -PSDTX-Ta1a	3566.12	126,081	GILGVIAWIKLIQILAPTAAVEVATRLGLPQ	Monomer	2.0	60.61	10.84
U ₅ -PSDTX-Ta1a	3615.97	91,424	FWGLILQGIWAVVKWAGPIIVDIAADYVIEYV *	Monomer	1.0	71.88	4.03

Since several dimeric peptides have been found in ant venoms [32], particularly for the Pseudomyrmecinae [14,15,33], we hypothesized that these three transcripts encoded dimeric peptide subunits. Based on this hypothesis, we calculated the theoretical masses of these presumed homo/heterodimeric peptides. Indeed, the calculated masses (i.e., 5438.60; 5750.94; 5608.78; 5680.86 Da) were detected in the total ion chromatogram of the crude venom corresponding to the pseudomyrmecitoxins Ta-5438, Ta-5750, Ta-5608, Ta-5680, and eluting from 23.64 to 26.75 min (Table 1). In order to confirm the presence of these four dimeric peptides, the crude venom was reduced using dithiothreitol (DTT) and then alkylated with iodoacetamide (IA). Both reduced and reduced/alkylated venoms were submitted to LC–MS analysis on an LCQ-ion trap Advantage mass spectrometer following the same elution conditions as crude venom. The comparisons of both chromatograms and spectra before and after reduction/alkylation support our hypothesis of dimeric features. As these dimeric pseudomyrmecitoxins shared a high percentage identity (an average of 57%), we used the same subscript to denote the ‘unknown’ activity descriptor prefix (U₂). The homodimeric U₂-PSDTX-Ta1b toxin (Ta-5438) is structured by three disulfide bonds, but further investigations are required to know whether cysteines form three interchain disulfide bonds or one interchain plus two intrachain disulfide bonds. Both homodimer U₂-PSDTX-Ta1a (Ta-5750) and homodimer U₂-PSDTX-Ta1c (Ta-5608) are linked together by a single disulfide bond. The heterodimeric peptide Ta-5680 was formed of two very similar chains (i.e., U₂-PSDTX-Ta1a and U₂-PSDTX-Ta1c) linked together by one disulfide bond (Figure 5). It should also be noted that the monomer form of U₂-PSDTX-Ta1a was detected in the venom in a relatively small proportion (0.28% of the whole venom peptidome).

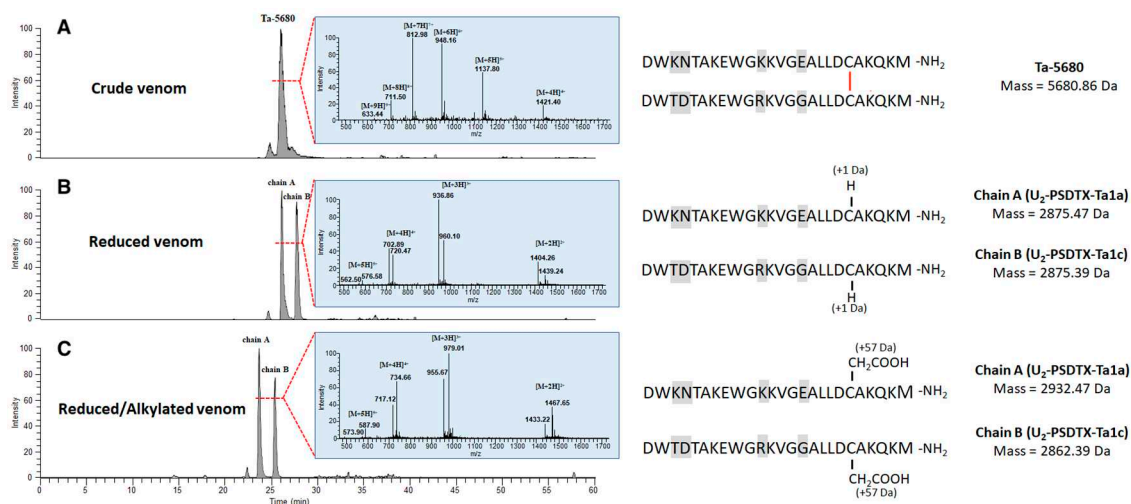


Figure 5. Identification of the dimeric features of the Ta-5680 peptides in *T. aethiops* venom. (A) Extracted-ion chromatogram and MS spectrum of the peptide Ta-5680 from the LC–MS analysis of *T. aethiops* venom before reduction/alkylation. We hypothesized that Ta-5680 is a heterodimeric peptide having the hypothetical sequence shown on the right. The distinctive residues between both monomers are highlighted in grey and the red bar represents the disulfide bond. (B) Comparison of chromatograms and spectra before and after reduction with DTT (Dithiothreitol) revealed the presence of two novel masses (2875.47 and 2805.39 Da) corresponding to both A and B chains while the Ta-5680 mass disappeared. (C) Alkylation experiment using IA (iodoacetamide) confirmed the presence of one cysteine on each alkylated monomer.

Overall, except for U₅-PSDTX-Ta1a, all the pseudomyrmecitoxins in the *T. aethiops* venom are polycationic and basic (pI ranging from 8.26 to 10.84) having a net charge ranging from 1 to 5.8. Both peptide U₄- and U₅-PSDTX-Ta1a possess a high proportion of hydrophobic residues (Table 2). Furthermore, several pseudomyrmecitoxins shared a high percentage identity with previously reported ant venom peptides. Thus, chains from both homo- and heterodimers (i.e., U₂-PSDTX-Ta1a-c) shared an average of 30% identity with the A-chain of M-MIITX-Mp2a from *Myrmecia pilosula*. In addition, two other pseudomyrmecitoxins (i.e., U₃-PSDTX-Ta1a and U₄-PSDTX-Ta1a) shared sequence identities with myrmecitoxins previously found for *Myrmecia gulosa* [28], poneratoxins found for *Neoponera goeldii* [34], and myrmecitoxins found for *Tetramorium bicarinatum* [29] (Figure 6). The structural prediction performed on the PepFold3 server suggested that all pseudomyrmecitoxins adopt secondary structures dominated by α-helices [35].

Species	%ID	Sequence
<i>Tetraponera aethiops</i>	-	U3-PSDTX-Ta1a K KKR--- K WVTKAIKEVGKTI G EALV E EA V SA A LS A ATE G --- G E K EE---
<i>Myrmecia gulosa</i>	45	MIITX1-Mg7b K RRR G L K K I I G KV I K G T G K V A G E A A S A V A D A V S A A I D A V V C T T E E P E Q
Species	%ID	Sequence
<i>Tetraponera aethiops</i>	-	U4-PSDTX-Ta1a G I L G V I--- A R W I W K L I Q I L A P T A A V E V A T R L G L P Q
<i>Tetramorium bicarinatum</i>	34	U10-MYRTX-Tb1a G L G F L A K I M G K V G M R M I K L V E A A K V A V D L S Q Q Q
<i>Myrmecia gulosa</i>	24	MIITX1-Mg1a G L G R L I--- G K I A K K G A K I A E A A A A N A A E K A E A L
<i>Neoponera goeldii</i>	19	U1-PONTX-Ng3e G L K D V W K I A G W L K K K G P G L K A A M A A A T Q-----
Species	%ID	Sequence
<i>Myrmecia pilosula</i>	-	M-MIITX-Mp2a I D W K K V D W K K V S K K T C K V M L K A C K F L-
<i>Tetraponera aethiops</i>	36	U2-PSDTX-Ta1a --- D W K N T A--- K E W G K K V G E A L L D C A K Q K M
<i>Tetraponera aethiops</i>	29	U2-PSDTX-Ta1b --- D W K G G A--- K D C A K K G A Q C V L E C V Q Q K M
<i>Tetraponera aethiops</i>	25	U2-PSDTX-Ta1c --- D W T D T A--- K E W G R K V G G A L L D C A K Q K M

Figure 6. Of *Tetraponera aethiops* mature venom peptides with similar venom peptides from ant venoms [28,29,34,36]. Alignments were generated with the Muscle program in Seaview version 4.6.1 and edited using BOXSHADE version 3.2. Conserved residues are highlighted in cyan, identical residues are highlighted in magenta and similar residues are highlighted in blue.

3. Discussion

Defensive venoms such as those of bees or fishes are arguably comprised of conserved toxins acting primarily to trigger pain and being less complex in composition than predatory venoms [1]. The venom investigation conducted in this study revealed that the defensive venom peptidome of *T. aethiops* is streamlined, containing only twelve peptides, with one being very dominant accounting for 66.41% of the overall peptidome (Figure 1, Table 1). This venom peptide diversity is obviously lower than the highly complex predatory venoms of spiders [37], scorpions [38], centipedes [39], or cone snails [40]. The venom peptidome of *T. aethiops* is also substantially less complex than the predatory ant *Tetramorium bicarinatum* venom peptidome, which is encoded by 37 peptide genes belonging to three superfamilies [29]. However, this peptide diversity is similar to those of other ants with a dual use of offensive/defensive venom such as *Myrmecia gulosa* [28] and *Odontomachus monticola* [26]. Different venom compositions have also been noted in the subfamily Pseudomyrmecinae according to a mass spectrometry-based investigation. Indeed, the predatory ground-dwelling species *Pseudomyrmex termitarius* has a venom composed of 87 linear peptides, whereas the venoms of the arboreal *P. gracilis* and the plant-ant *P. penetrator* are composed of only 23 and 26 peptides, respectively, certain of them with disulfide bonds and dimeric features [15].

Here we also demonstrate the presence of a set of homo- and heterodimeric peptides in the *T. aethiops* venom which is consistent with previous examinations of pseudomyrmecine venoms from the Paleotropical genus *Tetraoponera* [33] and the Neotropical genus *Pseudomyrmex* genera [14,15]. Interestingly, none of the dimeric pseudomyrmecitoxins characterized in *T. aethiops* venom show sequence identity with known heterodimeric pseudomyrmecitoxins found in *Pseudomyrmex triplarinus* venom (see [14]). Nevertheless, our data revealed sequence identity with a dimeric myrmeciotoxin from *Myrmecia pilosula*, M-MIITX-Mp2a: 36%, 29%, and 25% for U₂-PSDTX-Ta1a, U₂-PSDTX-Ta1b, and U₂-PSDTX-Ta1c, respectively (Figure 6; See [27,41]). Altogether, these multiple dimeric peptides found in the genera *Myrmecia*, *Pseudomyrmex* and *Tetraoponera*, suggest that dimeric scaffolds are recurrent in the phylogenetic clade Pseudomyrmecinae/Myrmeciinae, with these two subfamilies being close phylogenically [42]. Over a broader scale, dimeric venom peptides were found in the venom of several ant subfamilies (i.e., Pseudomyrmecinae, Ectatomminae, Myrmeciinae, and Ponerinae) [23,27,28,32,33,43] while this structural feature was only occasionally noted in the peptidome of other venomous animals [44–46]. Interestingly, these ant species possessing dimeric peptide toxins in their venom are also well-known for the intense pain induced by their sting. This observation raises the matter of the evolutionary advantage of such dimeric scaffolds for ant venoms that might be related to the defense of the colony and extended to the host myrmecophyte for plant-ants ([15], this study).

In terms of primary sequence, several of the pseudomyrmecitoxins described here have sequence identities with other ant venom dimeric peptides which are known for their pain-inducing properties. Actually, each chain of dimeric U₂-PSDTXs has a sequence identity with the A-chain of the M-MIITX-Mp2a, the heterodimeric pilosulin discovered in *Myrmecia pilosula* venom (Figure 6). Indeed, Dekan et al. (2017) showed that M-MIITX-Mp2a, a membrane-disrupting peptide displaying broad-spectrum antimicrobial and nociceptive properties, also induces a concentration-dependent transient increase in the intracellular Ca²⁺ in neuronal cells eliciting spontaneous pain in mice [36]. Along the same line, U₄-PSDTX-Ta1a, the major linear peptide in *T. aethiops* venom has 45% sequence identity with the major linear peptide of *Myrmecia gulosa* venom, MIITX₁-Mg1a (Figure 6). Robinson et al. (2018) showed that this peptide possesses a membrane disrupting activity capable of causing a leak in membrane ion conductance, thus altering membrane potential and triggering neuronal depolarization. On the one hand, it activates mammalian sensory neurons, which is consistent with the capacity to produce pain. On the other hand, MIITX₁-Mg1a is also able to incapacitate arthropods, demonstrating its multifunctional role [28]. Presumably, these sequence identities with pain-inducing peptides might explain the extremely painful *T. aethiops* stings even though the functional characterization of pseudomyrmecitoxins is obviously required to remove any doubt.

The defensive function of the venom is likely one of the major factors that favored the rise of sociality in hymenopterans, driving the evolution of the venom toward the protection of the colony from predators [47]. The present investigation provides another piece to the evolutive puzzle of Hymenoptera venoms where polycationic, amphiphilic, α -helical peptides, which are sometimes dimeric, seems to play a major role for colony protection.

4. Conclusions

In this study, we investigated the venom peptidome of *Tetraoponera aethiops* which is mainly used for defensive purposes. Our data revealed that the *T. aethiops* venom peptidome contains twelve venom peptides, among which nine were identified through the combination of transcriptomic and proteomic data. We hypothesized that the imposed selective pressures to deter predators have led toward the simplification of *T. aethiops* venom likely dominated by pain-inducing toxins. *Tetraoponera aethiops* venom peptides, which are either linear or dimeric, possess substantial sequence identities with myrmecitoxins previously described as both pain-inducing and insecticidal. Nevertheless, the functional characterization of *T. aethiops* venom pseudomyrmecitoxins is needed to come to a conclusion on their biological significance.

5. Materials and Methods

5.1. Collection of the Ants and Preparation of the Venom Samples

Tetraoponera aethiops workers were collected in Cameroon, near Douala (N3°53.310' E10°44.918') in 2016 and 2018. The ants' venom sacs were dissected and placed together in water containing 1% formic acid (*v/v*) and the membranes disrupted using ultrasonic waves for 2 min. Then, the samples were centrifuged for 5 min at 14,400 rpm, the supernatant was collected and dried with a speed vacuum prior to storage at $-20\text{ }^{\circ}\text{C}$ until use. Four venom samples containing 14, 19, 20, and 29 venom sacs were used for all the proteomic analyses.

5.2. Mass Spectrometry Analysis

A preliminary LC–MS analysis of the crude venom was carried out on the LCQ-Ion trap Surveyor equipped with an ESI-LC system Advantage (ThermoFisher Scientific, Courtabœuf, France). Peptides were separated using an Acclaim RSLC C₁₈ column (2.2 μm ; 2.1 \times 150 mm; ThermoFisher, France). The mobile phase was a gradient prepared from 0.1% aqueous formic acid (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The peptides were eluted using a linear gradient from 0 to 50% of solvent B during 45 min, then from 50 to 100% during 10 min, and finally held for 5 min at a 250 $\mu\text{L min}^{-1}$ flow rate. The electrospray ionization mass spectrometry detection was performed in positive mode with the following optimized parameters: the capillary temperature was set at 300 $^{\circ}\text{C}$, 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 area value of each peak corresponding to a peptide was manually integrated using the peak ion extraction function in Xcalibur software (version 4.0, ThermoFisher Scientific, Courtabœuf, France). The relative peak area indicates the contribution of each peptide to all the peptides identified in the venom, providing a measure of relative abundance.

5.3. Disulfide Bond Reduction and Alkylation

The presence of dimeric peptides in *T. aethiops* venom was determined via chemical reduction/alkylation of crude venom and subsequent LC–MS analysis. Disulfide reduction was achieved by mixing 40 μL of crude venom (six venom reservoirs) incubated with 40 μL of 100 mM ammonium bicarbonate buffer (pH 8) containing 10 mM dithiothreitol (DTT) for 30 min at 56 $^{\circ}\text{C}$. Then, 40 μL of the reduced venom was analyzed through LC–MS to identify the dimeric peptides. Finally, the remaining 40 μL of the reduced venom was alkylated by adding 1.5 μL of 0.5 M iodoacetamide (IA) for 90 min at room temperature in the dark. As chemical reduction/alkylation results in a mass

increase of 57 Da for each cysteine, the examination of mass shifts in the mass spectra of both reduced and reduced/alkylated samples permitted us to determine the number of disulfide bonds in the corresponding peptides.

5.4. *De novo Orbitrap Mass Spectrometry-Based Sequencing*

Crude venom was re-suspended in water and then desalted using ZipTip® C₁₈ (Merck Millipore, Burlington, VT, USA) after adding trifluoroacetic acid at a final concentration of 0.5%. 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 C₁₈ pre-column (2 cm, 75 µm i.d., 3 µm, 100 Å), Pepmap-RSLC Proxeon C₁₈ 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% formic acid) to 35% solvent B (99.9% acetonitrile, 0.1% formic acid) 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) and pyroglutamic acid (Glu). De novo peptide sequences with Average Local Confidence (ALC) higher than 60% were used for the peptide identifications.

5.5. *Direct Sequencing of Venom Gland RNA*

Venom glands and sacs from 20 live ant workers, anesthetized by cooling, were dissected in a PBS solution. Each tissue was immediately placed in 500 µL of TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and total RNAs were extracted afterward using the RNeasy Micro Kit (Qiagen) according to manufacturer's instructions. Contaminating genomic DNA was removed using a DNA-free kit (Applied Biosystem) according to the manufacturer's instructions. RNA quantity was evaluated using a nanodrop and a bioanalyzer (Nanodrop 2000, ThermoFisher Scientific; Agilent 2100 Bioanalyzer System). RNAseq was performed at the GeT-PlaGe core facility, INRA Toulouse, France. RNA-seq libraries were prepared according to Illumina's protocols using the Illumina TruSeq Stranded mRNA sample prep kit to analyze mRNA. Briefly, 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 experiments were performed on an Illumina HiSeq3000 using a paired-end read length of 2 × 150 pb with the Illumina HiSeq3000 sequencing kits.

5.6. *Bioinformatic Tools*

5.6.1. Contigs Quantification

The read pairs were assembled twice with drap (version 1.9.1) [48] using the de Bruijn graph assembler called oases (parameter: -dbg oases). 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) [49] 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 [50]. 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 transcripts discovered in the venom gland transcriptome, the appellation Reads Per Million (RPM) was chosen over the traditionally used Transcripts count Per Million (TPM) value. Indeed, TPM calculation takes into account the length of the transcript, or in absence of reference genome, of the contig. However, for contigs containing the venom peptides open reading frames (ORF), the assembler often generates overextended contigs (see Table S1). Thus, the expression rate of the short venom peptides transcripts would be underestimated with TPM [51]. So we calculated the RPM value for each transcript of interest in two steps: (i) by dividing the number of aligned reads for each contig by the total number of million reads aligned for the sample, and (ii) by summing up the obtained values for each contig encoding the transcript when several contigs represent the same peptide.

5.6.2. Precursor Identification and Mature Sequences

RNAseq data were translated using a translate program command lines (emboss package, command line: transeq) in order to obtain the potential Open Reading Frames. 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 (ncbi-blast-2.6.0+ package, command line: blastp, parameter: -matrix PAM30) with adapted parameters for short sequences, 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. The isoelectric points, net charges and percentage of hydrophobic amino acids were calculated using ExPASy (https://web.expasy.org/compute_pi/), PepCalc (<https://pepcalc.com/>), and Peptide Property Calculator V3.1 (<https://www.biosyn.com/peptidepropertycalculator/peptidepropertycalculator.aspx>). Secondary structure predictions were conducted with the PEP-FOLD3 server [35]. 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 and sequence identity percentages were calculated with the EMBL-EBI server and the MUSCLE program with the default parameters [52]. They were then edited using Seaview version 4.6.1 [53] and BOXSHADE version 3.2 (https://embnet.vital-it.ch/software/BOX_form.html).

5.6.3. Annotation of Most Expressed Contigs

Open reading frames (≥ 100 amino-acids length), found by translating RNAseq data, were extracted from the most abundant contigs 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).

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-6651/11/12/732/s1>, Table S1: Addressing table of major contigs expressed by *Tetraponera aethiops* venom glands.

Author Contributions: A.T. and M.T. (Michel Treilhou) conceptualized this project. A.T., N.T., M.T. (Maurice Tindo), and M.K. collected the ants. V.B., A.T., N.T., C.K., and E.B. performed the experiments and data analysis. V.B. and A.T. wrote the original draft. All authors contributed to writing and theoretical discussions.

Funding: This research received no external funding.

Acknowledgments: We are grateful for Andrea Yockey for proofreading the manuscript. V.B. was the recipient of a PhD fellowship from the French Ministry of Scientific Research. This study was performed in collaboration with the GeT core facility, Toulouse, France (<http://get.genotoul.fr>), and was aided by the France Génomique National institutional infrastructure, funded as part of the “Investissement d’avenir” program managed by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09). Ant samples were collected under the authorization of the Cameroon Ministry of Scientific Research and Innovation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Casewell, N.R.; Wüster, W.; Vonk, F.J.; Harrison, R.A.; Fry, B.G. Complex cocktails: The evolutionary novelty of venoms. *Trends Ecol. Evol.* **2013**, *28*, 219–229. [[CrossRef](#)]
2. Daly, N.L.; Wilson, D. Structural diversity of arthropod venom toxins. *Toxicon* **2018**, *152*, 46–56. [[CrossRef](#)]
3. Walker, A.A.; Robinson, S.D.; Yeates, D.K.; Jin, J.; Baumann, K.; Dobson, J.; Fry, B.G.; King, G.F. Entomo-venomics: The evolution, biology and biochemistry of insect venoms. *Toxicon* **2018**, *154*, 15–27. [[CrossRef](#)]
4. Blank, S.; Seismann, H.; Bockisch, B.; Braren, I.; Cifuentes, L.; McIntyre, M.; Ruhl, D.; Ring, J.; Bredehorst, R.; Ollert, M.W.; et al. Identification, recombinant expression, and characterization of the 100 kDa high molecular weight hymenoptera venom allergens Api m 5 and Ves v 3. *J. Immunol.* **2010**, *184*, 5403–5413. [[CrossRef](#)]
5. Hoffman, D.R. Ant venoms. *Curr. Opin. Allergy Clin. Immunol.* **2010**, *10*, 342–346. [[CrossRef](#)]
6. Wanandy, T.; Wilson, R.; Gell, D.; Rose, H.E.; Gueven, N.; Davies, N.W.; Brown, S.G.A.; Wiese, M.D. Towards complete identification of allergens in Jack Jumper (*Myrmecia pilosula*) ant venom and their clinical relevance: An immunoproteomic approach. *Clin. Exp. Allergy* **2018**, *48*, 1222–1234. [[CrossRef](#)]
7. Nelder, M.P.; Paysen, E.S.; Zungoli, P.A.; Benson, E.P. Emergence of the introduced ant *Pachycondyla chinensis* (Formicidae: Ponerinae) as a public health threat in the southeastern United States. *J. Med. Entomol.* **2006**, *43*, 1094–1098. [[CrossRef](#)]
8. Srisong, H.; Sukprasert, S.; Klaynongsruang, S.; Daduang, J.; Daduang, S. Identification, expression and characterization of the recombinant Sol g 4.1 protein from the venom of the tropical fire ant *Solenopsis geminata*. *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2018**, *24*, 23. [[CrossRef](#)]
9. Wanandy, T.; Honda-Okubo, Y.; Davies, N.W.; Rose, H.E.; Heddle, R.J.; Brown, S.G.A.; Woodman, R.; Petrovsky, N.; Wiese, M.D. Pharmaceutical and preclinical evaluation of Advax adjuvant as a dose-sparing strategy for ant venom immunotherapy. *J. Pharm. Biomed. Anal.* **2019**, *172*, 1–8. [[CrossRef](#)]
10. Cologna, C.T.; dos S. Cardoso, J.; Jourdan, E.; Degueudre, M.; Upert, G.; Gilles, N.; Uetanabaro, A.P.T.; Costa Neto, E.M.; Thonart, P.; de Pauw, E.; et al. Peptidomic comparison and characterization of the major components of the venom of the giant ant *Dinoponera quadriceps* collected in four different areas of Brazil. *J. Proteomics* **2013**, *94*, 413–422. [[CrossRef](#)]
11. Cologna, C.T.; Rodrigues, R.S.; Santos, J.; de Pauw, E.; Arantes, E.C.; Quinton, L. Peptidomic investigation of *Neoponera villosa* venom by high-resolution mass spectrometry: Seasonal and nesting habitat variations. *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2018**, *24*, 6. [[CrossRef](#)]
12. 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. A simple, rapid method for the extraction of whole fire ant venom (Insecta: Formicidae: Solenopsis). *Toxicon* **2013**, *65*, 5–8.
13. Aili, S.R.; Touchard, A.; Petitclerc, F.; Dejean, A.; Orivel, J.; Padula, M.P.; Escoubas, P.; Nicholson, G.M. Combined peptidomic and proteomic analysis of electrically stimulated and manually dissected venom from the south american bullet ant *Paraponera clavata*. *J. Proteome Res.* **2017**, *16*, 1339–1351. [[CrossRef](#)]
14. Pan, J.; Hink, W.F. Isolation and characterization of myrmexins, six isoforms of venom proteins with anti-inflammatory activity from the tropical ant, *Pseudomyrmex triplarinus*. *Toxicon* **2000**, *38*, 1403–1413. [[CrossRef](#)]
15. Touchard, A.; Labrière, N.; Roux, O.; Petitclerc, F.; Orivel, J.; Escoubas, P.; Koh, J.M.S.; Nicholson, G.M.; Dejean, A. Venom toxicity and composition in three *Pseudomyrmex* ant species having different nesting modes. *Toxicon* **2014**, *88*, 67–76. [[CrossRef](#)]
16. Rico-Gray, V.; Oliveira, P.S. *The Ecology and Evolution of Ant-Plant Interactions*; University of Chicago Press: Chicago, IL, USA, 2007.
17. Yumoto, T.; Maruhashi, T. Pruning behavior and intercolony competition of *Tetraponera* (*Pachysima*) *aethiops* (Pseudomyrmecinae, Hymenoptera) in *Barteria fistulosa* in a tropical forest, Democratic Republic of Congo. *Ecol. Res.* **1999**, *14*, 393–404. [[CrossRef](#)]
18. Blatrix, R.; Djité-Lordon, C.; Mondolot, L.; la Fisca, P.; Voglmayr, H.; Mckey, D. Plant-ants use symbiotic fungi as a food source: New insight into the nutritional ecology of ant-plant interactions. *Proc. R. Soc. B Biol. Sci.* **2012**, *279*, 3940–3947. [[CrossRef](#)]

19. Lee, D.W. *Nature's Fabric: Leaves in Science and Culture*; The University of Chicago Press: Chicago, IL, USA, 2017.
20. Dejean, A.; Djiéto-Lordon, C.; Orivel, J. The plant ant *Tetraponera aethiops* (Pseudomyrmecinae) protects its host myrmecophyte *Barteria fistulosa* (Passifloraceae) through aggressiveness and predation. *Biol. J. Linn. Soc.* **2008**, *93*, 63–69. [[CrossRef](#)]
21. Johnson, S.R.; Copello, J.A.; Evans, M.S.; Suarez, A.V. A biochemical characterization of the major peptides from the venom of the giant neotropical hunting ant *Dinoponera australis*. *Toxicon* **2010**, *55*, 702–710. [[CrossRef](#)]
22. King, G.F.; Gentz, M.C.; Escoubas, P.; Nicholson, G.M. A rational nomenclature for naming peptide toxins from spiders and other venomous animals. *Toxicon* **2008**, *52*, 264–276. [[CrossRef](#)]
23. Touchard, A.; Aili, S.R.; Fox, E.G.P.; Escoubas, P.; Orivel, J.; Nicholson, G.M.; Dejean, A. The biochemical toxin arsenal from ant venoms. *Toxins* **2016**, *8*, 30. [[CrossRef](#)]
24. Deutsch, E.W.; Csordas, A.; Sun, Z.; Jarnuczak, A.; Perez-Riverol, Y.; Ternent, T.; Campbell, D.S.; Bernal-Llinares, M.; Okuda, S.; Kawano, S.; et al. The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* **2017**, *45*, 1100–1106. [[CrossRef](#)] [[PubMed](#)]
25. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*, D442–D450. [[CrossRef](#)] [[PubMed](#)]
26. Kazuma, K.; Masuko, K.; Konno, K.; Inagaki, H. Combined venom gland transcriptomic and venom peptidomic analysis of the predatory ant *Odontomachus monticola*. *Toxins* **2017**, *9*, 323. [[CrossRef](#)] [[PubMed](#)]
27. Davies, N.W.; Wiese, M.D.; Brown, S.G.A. Characterisation of major peptides in “jack jumper” ant venom by mass spectrometry. *Toxicon* **2004**, *43*, 173–183. [[CrossRef](#)]
28. Robinson, S.D.; Mueller, A.; Clayton, D.; Starobova, H.; Hamilton, B.R.; Payne, R.J.; Vetter, I.; King, G.F.; Undheim, E.A.B. A comprehensive portrait of the venom of the giant red bull ant, *Myrmecia gulosa*, reveals a hyperdiverse hymenopteran toxin gene family. *Sci. Adv.* **2018**, *4*, eaau4640. [[CrossRef](#)] [[PubMed](#)]
29. Touchard, A.; Téné, N.; Chan Tchi Song, P.; Lefranc, B.; Leprince, J.; Treilhou, M.; Bonnafé, E. Deciphering the molecular diversity of an ant venom peptidome through a venomomics approach. *J. Proteome Res.* **2018**, *17*, 3503–3516. [[CrossRef](#)]
30. Kreil, G.; Haiml, L.; Suchanek, G. Stepwise cleavage of the pro part of promelittin by dipeptidylpeptidase IV. *Eur. J. Biochem.* **1980**, *111*, 49–58. [[CrossRef](#)] [[PubMed](#)]
31. Hurst, D.; Rylett, C.M.; Isaac, R.E.; Shirras, A.D. The drosophila angiotensin-converting enzyme homologue Ance is required for spermiogenesis. *Dev. Biol.* **2003**, *254*, 238–247. [[CrossRef](#)]
32. Aili, S.R.; Touchard, A.; Escoubas, P.; Padula, M.P.; Orivel, J.; Dejean, A.; Nicholson, G.M. Diversity of peptide toxins from stinging ant venoms. *Toxicon* **2014**, *92*, 166–178. [[CrossRef](#)]
33. Touchard, A.; Koh, J.M.S.; Aili, S.R.; Dejean, A.; Nicholson, G.M.; Orivel, J.; Escoubas, P. The complexity and structural diversity of ant venom peptidomes is revealed by mass spectrometry profiling. *Rapid Commun. Mass Spectrom.* **2015**, *29*, 385–396. [[CrossRef](#)] [[PubMed](#)]
34. Orivel, J.; Redeker, V.; Le Caer, J.P.; Krier, F.; Revol-Junelles, A.M.; Longeon, A.; Chaffotte, A.; Dejean, A.; Rossier, J. Ponericins, new antibacterial and insecticidal peptides from the venom of the ant *Pachycondyla goeldii*. *J. Biol. Chem.* **2001**, *276*, 17823–17829. [[CrossRef](#)] [[PubMed](#)]
35. Lamiable, A.; Thévenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tufféry, P. PEP-FOLD3: Faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* **2016**, *44*, W449–W454. [[CrossRef](#)]
36. Dekan, Z.; Headey, S.J.; Scanlon, M.; Baldo, B.A.; Lee, T.H.; Aguilar, M.I.; Deuis, J.R.; Vetter, I.; Elliott, A.G.; Amado, M.; et al. Δ -Myrtoxin-Mp1a is a helical heterodimer from the venom of the jack jumper ant that has antimicrobial, membrane-disrupting, and nociceptive activities. *Angew. Chem. Int. Ed.* **2017**, *56*, 8495–8499. [[CrossRef](#)]
37. Smith, J.J.; Herzig, V.; King, G.F.; Alewood, P.F. The insecticidal potential of venom peptides. *Cell. Mol. Life Sci.* **2013**, *70*, 3665–3693. [[CrossRef](#)] [[PubMed](#)]
38. Ma, Y.; He, Y.; Zhao, R.; Wu, Y.; Li, W.; Cao, Z. Extreme diversity of scorpion venom peptides and proteins revealed by transcriptomic analysis: Implication for proteome evolution of scorpion venom arsenal. *J. Proteomics* **2012**, *75*, 1563–1576. [[CrossRef](#)]

39. Undheim, E.A.B.; Fry, B.G.; King, G.F. Centipede venom: Recent discoveries and current state of knowledge. *Toxins* **2015**, *7*, 679–704. [[CrossRef](#)] [[PubMed](#)]
40. Olivera, B.M.; Ramilo, C.A.; Abogadie, F.C.; Cruz, L.J.; Woodward, S.R.; Hillyard, D.R.; Rivier, J.; Clark, C.; Corpuz, G.P.; Mena, E.E. Diversity of conus neuropeptides. *Science* **1990**, *249*, 257–263. [[CrossRef](#)]
41. Wanandy, T.; Gueven, N.; Davies, N.W.; Brown, S.G.A.; Wiese, M.D. Pilosulins: A review of the structure and mode of action of venom peptides from an Australian ant *Myrmecia pilosula*. *Toxicon* **2015**, *98*, 54–61. [[CrossRef](#)] [[PubMed](#)]
42. Borowiec, M.L.; Rabeling, C.; Brady, S.G.; Fisher, B.L.; Schultz, T.R.; Ward, P.S. Compositional heterogeneity and outgroup choice influence the internal phylogeny of the ants. *Mol. Phylogenet. Evol.* **2019**, *134*, 111–121. [[CrossRef](#)] [[PubMed](#)]
43. Tani, N.; Kazuma, K.; Ohtsuka, Y.; Shigeri, Y.; Masuko, K.; Konno, K.; Inagaki, H. Mass spectrometry analysis and biological characterization of the predatory ant *Odontomachus monticola* venom and venom sac components. *Toxins* **2019**, *11*, 50. [[CrossRef](#)] [[PubMed](#)]
44. Loughnan, M.; Nicke, A.; Jones, A.; Schroeder, C.I.; Nevin, S.T.; Adams, D.J.; Alewood, P.F.; Lewis, R.J. Identification of a novel class of nicotinic receptor antagonists. *J. Biol. Chem.* **2006**, *281*, 24745–24755. [[CrossRef](#)] [[PubMed](#)]
45. Santos, A.D.; Imperial, J.S.; Chaudhary, T.; Beavis, R.C.; Chait, B.T.; Hunsperger, J.P.; Olivera, B.M.; Adams, M.E.; Hillyard, D.R. Heterodimeric structure of the spider toxin ω -agatoxin IA revealed by precursor analysis and mass spectrometry. *J. Biol. Chem.* **1992**, *267*, 20701–20705. [[PubMed](#)]
46. Zamudio, F.Z.; Arévalo, C.; Conde, R.; Arévalos, C.; Becerril, B.; Martin, B.M.; Valdivia, H.H.; Possani, L.D. The mechanism of inhibition of ryanodine receptor channels by imperatoxin I, a heterodimeric protein from the scorpion *Pandinus imperator*. *J. Biol. Chem.* **1997**, *272*, 11886–11894. [[CrossRef](#)] [[PubMed](#)]
47. Schmidt, J.O. Evolutionary responses of solitary and social Hymenoptera to predation by primates and overwhelmingly powerful vertebrate predators. *J. Hum. Evol.* **2014**, *71*, 12–19. [[CrossRef](#)] [[PubMed](#)]
48. Cabau, C.; Escudié, F.; Djari, A.; Guiguen, Y.; Bobe, J.; Klopp, C. Compacting and correcting Trinity and Oases RNA-Seq de novo assemblies. *PeerJ* **2017**, *5*, e2988. [[CrossRef](#)]
49. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)]
50. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
51. Hsieh, P.H.; Oyang, Y.J.; Chen, C.Y. Effect of de novo transcriptome assembly on transcript quantification. *Sci. Rep.* **2019**, *9*, 8304. [[CrossRef](#)]
52. Chojnacki, S.; Cowley, A.; Lee, J.; Foix, A.; Lopez, R. Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic Acids Res.* **2017**, *45*, W550–W553. [[CrossRef](#)]
53. Gouy, M.; Guindon, S.; Gascuel, O. SeaView Version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **2010**, *27*, 221–224. [[CrossRef](#)] [[PubMed](#)]



Venom Peptide Repertoire of the European Myrmicine Ant *Manica rubida*: Identification of Insecticidal Toxins

Axel Touchard,* Samira R. Aili, Nathan Téné, Valentine Barassé, Christophe Klopp, Alain Dejean, R. Manjunatha Kini, Mrinalini, Laurent Coquet, Thierry Jouenne, Benjamin Lefranc, Jérôme Leprince, Pierre Escoubas, Graham M. Nicholson, Michel Treilhou,[#] and Elsa Bonnafé[#]



Cite This: <https://dx.doi.org/10.1021/acs.jproteome.0c00048>



Read Online

ACCESS |



Metrics & More



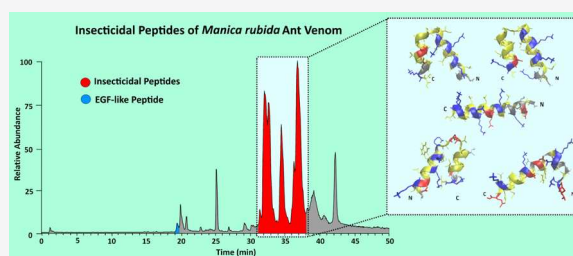
Article Recommendations



Supporting Information

ABSTRACT: Using an integrated transcriptomic and proteomic approach, we characterized the venom peptidome of the European red ant, *Manica rubida*. We identified 13 “myrmecitoxins” that share sequence similarities with previously identified ant venom peptides, one of them being identified as an EGF-like toxin likely resulting from a threonine residue modified by *O*-fucosylation. Furthermore, we conducted insecticidal assays of reversed-phase HPLC venom fractions on the blowfly *Lucilia caesar*, permitting us to identify six myrmecitoxins (i.e., U₃, U₁₀, U₁₃, U₂₀-MYRTX-Mri1a, U₁₀-MYRTX-Mri1b, and U₁₀-MYRTX-Mri1c) with an insecticidal activity. Chemically synthesized U₁₀-MYRTX-Mri1a, -Mri1b, -Mri1c, and U₂₀-MYRTX-Mri1a irreversibly paralyzed blowflies at the highest doses tested (30–125 nmol·g⁻¹). U₁₃-MYRTX-Mri1a, the most potent neurotoxic peptide at 1 h, had reversible effects after 24 h (150 nmol·g⁻¹). Finally, U₃-MYRTX-Mri1a has no insecticidal activity, even at up to 55 nmol·g⁻¹. Thus, *M. rubida* employs a paralytic venom rich in linear insecticidal peptides, which likely act by disrupting cell membranes.

KEYWORDS: glycosylated toxin, peptidome, polycationic α -helix, predation, reversible neurotoxicity, pyroglutamate



INTRODUCTION

The ecological success of ants in terrestrial ecosystems is partially due to their propensity to evolve and use a broad array of bioactive molecules to communicate, defend their colony, and capture their prey.^{1,2} Although ant venoms exhibit an extraordinary diversity of toxins ranging from low molecular mass compounds to large proteins, small linear peptides predominate.^{3,4} Most of the investigated linear ant venom peptides are membrane-active antimicrobial peptides (AMPs),^{5–9} even though some were also reported as insecticidal toxins.¹⁰ Insecticidal assays have been conducted only on peptides isolated from a few ant venoms, such as paraponeritoxin (formerly poneratoxin) and poneritoxins isolated from the venom of *Paraponera clavata* (Paraponerinae) and *Anochetus emarginatus* (Ponerinae).^{11,12}

The limited number of investigations conducted on ant venoms is mainly due to the low venom yield per individual, rendering sampling difficult. Yet, integrative “-omics” approaches have recently permitted researchers to overcome these hurdles by using a combination of venom gland transcriptomics and proteomics. Jointly termed “venomics”, such strategies permitted the comprehensive overview of the venom composition for four ant species.^{13–16} In the ant subfamily Myrmicinae, the most speciose with more than 6000 species described, the venom composition is extremely variable with alkaloids predominant in some tribes.^{17,18} Nevertheless,

recent proteomics analyses have revealed that the venoms of four myrmicine genera (i.e., *Tetramorium*, *Pogonomyrmex*, *Myrmica*, and *Manica*) are peptide-rich.^{19–22} Further investigation into the venom composition of other myrmicine species is therefore necessary to understand the intrinsic chemical properties of these venoms and gather information on the relationships between ant phylogeny and venom composition.

The myrmicine ant *Manica rubida*, the focal species of this study, is the largest stinging species in Western Europe known for its aggressive behavior associated with a painful sting. Furthermore, it is a generalist predator of small arthropods and its venom is very effective in paralyzing invertebrate prey.^{23,24} Previous mass spectrometry-based studies on *M. rubida* venom revealed the presence of ca. 200 molecular masses corresponding to peptides, some of them structured by one or three disulfide bonds;⁴ only one linear peptide has been sequenced so far.²⁰

Received: January 31, 2020

Published: March 17, 2020

In this study, we aimed to identify the peptides secreted by the venom gland of *M. rubida* and to investigate their insecticidal activity. Combining transcriptomic and proteomic methods, we identified 13 venom peptide precursors expressed in the venom glands and their encoding mature peptides. Most of the venom peptides described herein share sequence identities with peptides characterized in the venoms of *Tetramorium bicarinatum* (another Myrmicinae) and *Myrmecia gulosa* (a Myrmeciinae from Australia). Finally, we identified five novel insecticidal venom peptides which paralyze blowflies.

EXPERIMENTAL SECTION

Sample Collection

Live *M. rubida* workers were collected from Saint-Martin-Vésubie, Alpes Maritimes, France (N 44° 05' 20", E 7° 18' 47"). Colonies were maintained at 25 °C and fed three times a week with fresh mealworms and an aqueous honey solution (1:1 v/v). The ant venom sacs were dissected and pooled in water with 10% (v/v) acetonitrile (ACN) and the membranes were disrupted ultrasonically for 2 min. Then, samples were centrifuged for 5 min at 14 400 rpm, and the supernatant was collected and dried in a centrifuge evaporator prior to storage at –20 °C until use. Multiple venom samples of 100 venom sacs each were prepared to conduct the proteomic analyses. Independently, 200 ants were dissected for transcriptomic analyses. Their venom sacs plus venom glands were pooled and homogenized in 200 μ L TRIzol reagent (Ambion, Life Technologies, France) prior to storage at –80 °C until further use.

Mass Spectrometry

A preliminary LC-MS analysis of the crude venom was carried out on an LCQ-Ion trap Surveyor equipped with an ESI-LC system Advantage (Thermo Fisher Scientific, MA, USA) as for the venom of *Tetramorium bicarinatum* (for more detail see Touchard et al.).¹⁵ Briefly the peptides were eluted using a linear gradient of ACN on a C₁₈ column (Luna 5 μ m; 2 \times 150 mm; Phenomenex, CA, USA) at a flow rate of 250 μ L·min^{–1}. Then, the electrospray ionization mass spectrometry (ESI-MS) detection was performed in positive mode and the relative abundance of each peptide in the venom was determined using the peak ion extraction function in the Xcalibur software version 4.0 (Thermo Fisher Scientific, MA, USA).

Peptide Purification and Edman Degradation-Based Sequencing

Peptides from *M. rubida* venom were isolated as previously reported for the poneritoxins isolated from the *A. emarginatus* venom (see detail in Touchard et al.).¹² Five isolated peptides were selected for Edman degradation. Two peptides (i.e., U₆-MYRTX-Mri1a and U₁₇-MYRTX-Mri1b) were subjected to Edman degradation on a gas-phase sequencer model ABI 492 (Applied Biosystems, CA, USA). The phenylthiohydantoin (PTH) amino acid standard kit was used and reconstituted according to the manufacturer's instructions. The procedures and reagents were used as recommended by the manufacturer. The PTH amino acid derivatives generated at each sequence cycle were identified and quantified online with an Applied Biosystems Model 140C HPLC system using the Applied Biosystems Model 610A data analysis system for protein sequencing. The three other isolated peptides (i.e., U₁₂-MYRTX-Mri1a, U₁₈-MYRTX-Mri1a, and U₂₀-MYRTX-Mri1a) were loaded onto a precycled Biobrene Plus-coated glass filter.

The N-terminal sequences were then determined by introducing the filter disc into a Procise P494 automated protein sequencer (Applied Biosystems, CA, USA) and runs of Edman degradation were carried out.

De Novo Mass Spectrometry-Based Sequencing

Crude venom was resuspended in water and then desalted using a ZipTip C₁₈ (Merck Millipore, VT, USA) after adding TFA at a final concentration of 0.5% v/v. Then, the venom sample was subjected to *de novo* sequencing using a Q-Exactive Plus mass spectrometer coupled to a Proxeon 1000 Nano-LC (Thermo Fisher Scientific, MA, USA). Peptides were separated with the following parameters: Acclaim PepMap100 C₁₈ precolumn (2 cm, 75 μ m i.d., 3 μ m, 100 Å), Pepmap-RSLC Proxeon C₁₈ column (50 cm, 75 μ m i.d., 2 μ m, 100 Å), 300 nL·min^{–1} flow rate, a 98 min gradient from 95% solvent A (water, 0.1% formic acid) to 35% solvent B (ACN, 0.1% formic acid) 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 350–1550 *m/z*. Fragments were obtained by high collision-induced dissociation (HCD) activation with a collisional energy of 27%. MS/MS 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 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) and pyroglutamic acid (Glu). *De novo* peptide sequences with average local confidence (ALC) levels higher than 60% were used for peptide identification. Sequence tags from MS/MS were searched against the *de novo* transcriptome database of *M. rubida* (see next section). Then, the final validation of the peptide sequences was achieved by matching the masses measured by mass spectrometry with those predicted from transcriptomic and MS/MS sequencing data with a mass tolerance of 0.05 Da.

Venom Gland Transcriptomics

Total RNA was extracted following the TRIzol manufacturer's instructions (Thermo Fisher Scientific, MA, USA). Total RNA was cleaned using an RNeasy Micro Kit (Qiagen, The Netherlands) following the instructions of the RNeasy Micro Handbook (12/2007, p 56) with on-column DNase I treatment and then stored at –80 °C. RNA purity and quantity was verified using the NanoDrop 2000 (Thermo Fisher Scientific, MA, USA). Total RNA was dried with RNastable reagent (Biomatrix, CA, USA) and shipped to the Department of Biological Sciences, National University of Singapore for transcriptomic analysis. The dried sample was resuspended in 41 μ L water and RNA quality and quantity were further assessed using an Agilent 2100 Bioanalyzer. Using ~850 ng of total RNA, bead-based Poly-A tail selection was performed to purify the mRNA. Then, a cDNA library was constructed using the NEBNext Ultra Directional Library Prep Kit according to the manufacturer's protocol. Fragment size distribution of the library was verified using the Agilent 2100 Bioanalyzer. 250 bp Paired End sequencing was performed on an Illumina HiSeq 2500 platform on 1/14th of a lane. The read pairs were assembled with Drap (version 1.9.1)²⁵ with two de Bruijn graph assemblers: Oases_0.2.8 (default parameter) and Trinity-v2.3.2 (parameter: –dbg trinity). Drap 1.91 includes three read pre-processing steps before assembly with the following parameters: (1) read trimming and filtering (version

TrimGalore-0.4.5) trim_galore—length 25—quality 10—stringency 3; (2) read filtering (version fastq_illumina_filter-0.1) filter_illumina -q 10 -t 33 -e -m 25; (3) read normalization (version trinityrnaseq-Trinity-v2.3.2) insilico_read_normalization.pl—max_cov 50—pairs_together. 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)²⁶ using default parameters, and the alignment files were sorted, compressed, and indexed with SAMtools (version: 1.3.1) using default parameters.²⁷ 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 value (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 is then multiplied by 1 million to generate TPM. The two transcript databases (Oases and Trinity) were translated in all six frames using the tranSEQ program (emboss package 6.6.0.0) and converted to BLAST databases using makeblastdb. The *de novo* predictive peptide sequences were submitted to command-line BLAST 2.6.0+ program with adapted parameters for short sequences (-M PAM30) in order to determine the exact mature peptide and precursor sequences. Precursors and mature sequences were aligned using the Muscle program in SeaView 1:4.6.4–1 and edited with BoxShade v3.31 C (beta, 970507). Signal sequence and transmembrane domain were predicted with the Phobius program available at <http://phobius.sbc.su.se/>.

Fmoc Solid Phase Peptide Synthesis

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). Preloaded 4-hydroxymethyl-phenoxymethyl-copolystyrene-1%-divinylbenzene resins (HMP) were obtained from Life Technologies (CA, USA). *N,N*-Diisopropylethylamine (DIEA), piperidine, TFA, triisopropylsilane (TIS), *tert*-butylmethyl ether (TBME) were supplied by Sigma-Aldrich (MO, USA). *N*-methylpyrrolidone (NMP), dimethylformamide (DMF), dichloromethane (DCM), and acetonitrile were obtained from Fisher Scientific (MA, USA). Peptides were synthesized by Fmoc solid phase methodology on a Liberty microwave assisted automated peptide synthesizer (CEM, NC, USA) using the standard manufacturer's procedures at 0.1 mmol scale. All Fmoc-amino acids (0.5 mmol, 5 equiv) were coupled (25 W, 75 °C, 300 s) except for histidine (0 W, 50 °C, 120 s and 25 W, 50 °C, 240 s) on either HMP or Rink amide resin, by in situ activation with HBTU (0.5 mmol, 5 equiv) and DIEA (1 mmol, 10 equiv) before Fmoc removal with a 20% piperidine in DMF (35 W, 75 °C, 30 s and 35 W, 75 °C, 180 s). After completion of the chain assembly, peptides were deprotected and cleaved from the resin by adding 10 mL of the mixture TFA/TIS/H₂O (9.5:0.25:0.25) for 120 min at room temperature. After filtration, crude peptides were washed three times by precipitation in TBME followed by centrifugation (4500 rpm, 15 min). The synthetic peptides were purified by RP-HPLC on a 21.2 × 250 mm Kinetex XB-C₁₈ (5 μm, 100 Å) or Jupiter C₁₈ (5 μm, 300 Å) column (Phenomenex, CA, USA)

using a linear gradient (10–70%, 10–80%, 20–70% or 30–70% over 45 min) of ACN/0.1%TFA (99.9:0.1) at a flow rate of 10 mL/min. The purified peptides were then characterized through MALDI-TOF mass spectrometry on an UltrafleXtreme (Bruker, MA, USA) in reflector mode using α -cyano-4-hydroxycinnamic acid as a matrix. Analytical RP-HPLC, performed on a 4.6 × 250 mm Kinetex XB-C₁₈ (5 μm, 100 Å) or Jupiter C₁₈ (5 μm, 300 Å) column, indicated that the purity of all the peptides was >99.9%.

Venom Fractionation and Insecticidal Assays

The fractionation of crude venom was performed using a Finnigan Spectra SYSTEM HPLC (pump PC4000 and AS3000) equipped with a DAD-UV 6000LP detector and a C₁₈ column (Luna 5 μm; 2 × 150 mm; Phenomenex, CA, USA), controlled by ChromQuest 5.0 software (Thermo Fisher Scientific, MA, USA). RP-HPLC was performed at a flow rate of 1 mL·min⁻¹ with the same solvent system as ESI-LC system Advantage (see [Mass Spectrometry of the Experimental Section](#)). The fractions were collected each minute and vacuum-dried prior to MS and insecticidal assays. Insecticidal assays were conducted on *Lucilia caesar* blowflies, a model insect often used for the insecticidal screening of venom peptides,²⁸ and thus allowing comparison with other venom toxins activity and potency. Each C₁₈ RP-HPLC fraction was dissolved in 16 μL phosphate buffered saline (PBS) solution and manually injected into the lateral thoracic region of the blowflies using a 10 μL Hamilton syringe (Hamilton Company, NV, USA). A volume of 2 μL was injected per blowfly. Three flies were injected for each fraction and negative controls (PBS). Paralytic activity and lethality were determined after 5 min, 1 and 24 h. Flies that showed no sign of paralysis (no sign of movement dysfunction) were categorized as unaffected; otherwise, they were noted as affected. Flies were considered dead if they did not react at all after mechanical stimulation. The peptide content of each fraction was determined by analyzing 10 μL of resuspended fractions on an LCQ-Ion trap Surveyor equipped with an Advantage ESI-LC system (Thermo Fisher Scientific, MA, USA). The peptides were eluted using the same column and elution system as the crude venom (see [Mass Spectrometry of the Experimental Section](#)) using a linear gradient from 0 to 50% B over 17 min, from 50% to 100% B over 2 min and then held for 1 min at 100% B.

Insecticidal assays with synthetic peptides were performed as described by Bende et al.²⁹ and peptides were diluted in a PBS solution at several concentrations and injected into blowflies. Injections were made using a 1.0 mL Hamilton Syringe (1000 Series Gastight, Hamilton Company, NV, USA) with a fixed 29 G needle fitted to an Arnold hand microapplicator (Burkard Manufacturing Co. Ltd., England). Each fly received 1 μL of peptide-PBS solution and was individually housed in a 2 mL tube provided with 5 μL of 5% sucrose solution. The paralytic activity and lethality were assessed 1 and 24 h after injection. For each toxicity assay and for the appropriate control (PBS solution; *n* = 10 insects), we used up to ten doses ([Supplementary Table S1](#)) of each peptide (*n* = 10 flies per dose). The assay was repeated three times. Dose–response data were analyzed as detailed by Guo et al.²⁸ Briefly, the percentage of affected flies was plotted against toxin concentration and the resultant dose–response curve fitted using a Logistic function. This was then used to interpolate the respective median paralytic dose (PD₅₀) and the median lethal

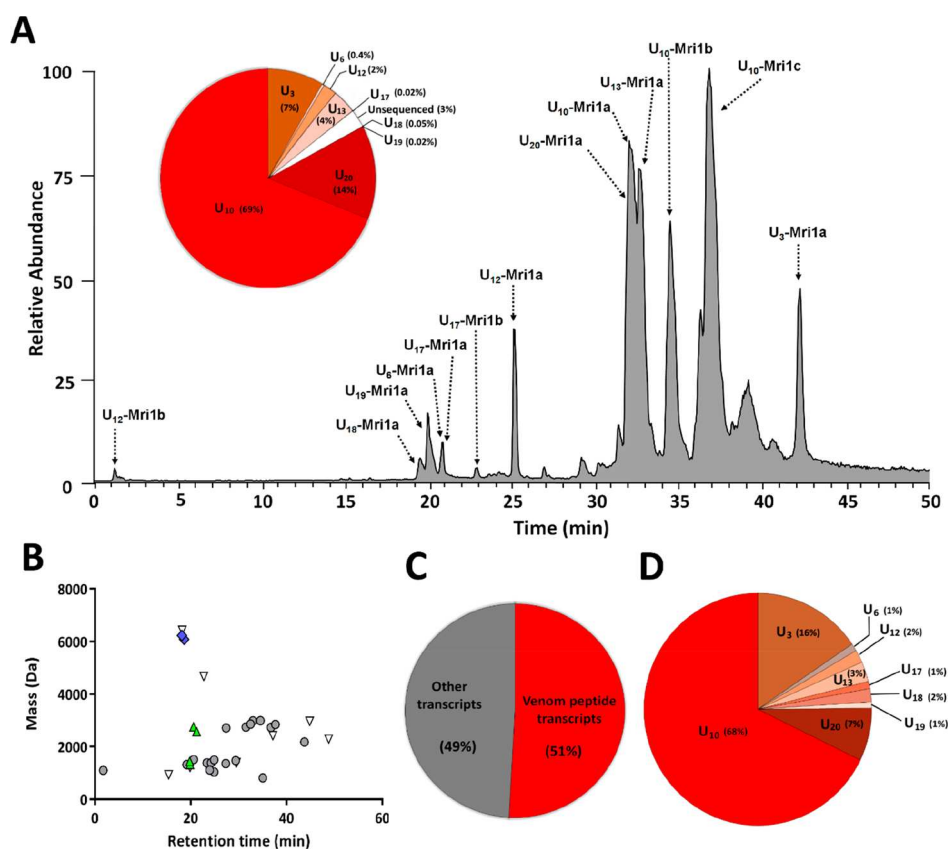


Figure 1. Venom peptidome for *M. rubida*. (A) Total ion chromatogram (TIC) for *M. rubida* venom. Peptides were eluted through RP-HPLC on a C₁₈ column using an ACN gradient of 1.1% min⁻¹ over 45 min and analyzed using an LCQ Advantage mass spectrometer. The circle graph is the relative abundance in the venom of each group of peptides (U_x) determined for each individual mass in the TIC (see [Supplementary Table S2](#)). The venom peptidome is mainly dominated by U₁₀ peptides that account for 69% of the venom peptide content. (B) The 2D venom peptide landscape shows the LC-MS venom profile for *M. rubida* which is mostly dominated by linear peptides. Gray circles represent peptides devoid of disulfide bridges, green triangles indicate peptides with one disulfide bridge and blue diamonds denote peptides with three disulfide bridges. White triangles represent unsequenced peptides. (C) Pie graph showing the proportion of venom peptide transcripts expressed in the *M. rubida* venom gland based on TPM values. Identified venom peptide transcripts represent 51% of the total venom gland transcript expression. (D) Pie graph showing the proportion of each peptide family expression based on TPM values (see [Supplementary Table S4](#)). U₁₀-MYRTX-Mri1a, -Mri1b, and -Mri1c and U₂₀-MYRTX-Mri1a peptide transcripts accounted for 75% of the total peptide transcript expression.

dose (LD₅₀) values using nonlinear regression analysis. The curves were fitted using Prism 7.0 (GraphPad Software, CA, USA).

Availability of Supporting Data

The MS data were deposited in the ProteomeXchange Consortium via the PRIDE partner repository (<https://www.ebi.ac.uk/pride/archive/>) with the data set identifier PXD015675. The transcriptome raw data are available through The European Nucleotide Archive (ENA) under study number PRJEB34828. All toxin sequences were deposited in Genbank with accession numbers MN765034 to MN765046.

Toxin Nomenclature

Novel myrmecitoxin sequences were named following the standard nomenclature for animal venom peptides,³⁰ and we used myrmecitoxins (MYRTX) to define the venom peptides of myrmecine ants.¹⁸ Since the genus/species descriptor “Mr” has been used to define myrmecitoxins from *Myrmica rubra* (UniProt accession number: PODLS0), we propose “Mri” to distinguish myrmecitoxins from *M. rubida*. We chose not to follow the genus/species descriptor “MANr” used by Heep and colleagues to name peptides from *M. rubida*,²⁰ since the

rational nomenclature system recommends keeping the descriptor as short as possible.

RESULTS

Venom Peptidome Characterization

As per Touchard et al.,¹⁵ the masses of peptides present in the *M. rubida* venom were listed through LC-MS on an LCQ-ion trap mass spectrometer (Figure 1A, B). The LC-MS analysis showed 33 masses corresponding to peptides and their relative abundance in the venom ([Supplementary Table S2](#)). The LC-MS/MS of the crude venom was then performed using a Q-exactive Orbitrap mass spectrometer which yielded fragmentation spectra for *de novo* sequencing using Peaks software. This resulted in 509 sequence tags with an ALC score higher than or equal to 80% ([Supplementary Table S3](#)). In addition, five peptides were isolated using C₁₈ RP-HPLC and submitted to Edman degradation for N-terminal sequencing. Initially, the peptides selected to perform Edman sequencing were those containing cysteines (i.e., U₆-MYRTX-Mri1a, U₁₇-MYRTX-Mri1b and U₁₈-MYRTX-Mri1a) though two additional linear peptides isolated in HPLC fractions (U₁₂-MYRTX-Mri1a and U₂₀-MYRTX-Mri1a) were also processed. The N-terminal Edman sequencing of both U₆-MYRTX-Mri1a and U₂₀-

Table 1. Myrmicotoxin Sequences in the Venom of *M. rubida*

peptide	sequence	PTMs ^a	mass calculated (Da)	mass measured (Da)	TPM ^b
U ₃ -MYRTX-Mri1a	GLPLLALLMTLPFIQHAI TN	C-ter NH ₂	2174.2649	2174.2720	78 932 ^c
U ₆ -MYRTX-Mri1a	IIGPCPKKPIGIVC	1 S-S	1434.8091	1434.8104	5296 ^c
U ₁₀ -MYRTX-Mri1a	GFKSMLAKAALKILKAVAPAAAAIADKI	C-ter NH ₂	2850.7245	2850.7212	193 522
U ₁₀ -MYRTX-Mri1b	GIKDALAKIWKILKAEVPTVAAA IENKV	C-ter NH ₂	2987.7899	2987.7941	7623
U ₁₀ -MYRTX-Mri1c	GVGSLAKAALKILKIVAPAAAEVIANKI	C-ter NH ₂	2840.7943	2840.7956	143 717
U ₁₂ -MYRTX-Mri1a	IDPKVLES LV	C-ter NH ₂	1110.6649	1110.6654	8923 ^c
U ₁₂ -MYRTX-Mri1b	IPPKAIKSLQ	C-ter NH ₂	1092.7019	1092.6543	178
U ₁₃ -MYRTX-Mri1a	DKPGQAKKIGLFDQIDKAAAAFMKLFEE	C-ter NH ₂	2978.6052	2978.6051	14 184 ^c
U ₁₇ -MYRTX-Mri1a	HIIVAPCREGYVMVGNVCVEEY	1 S-S	2541.1491	2541.1502	4357 ^c
U ₁₇ -MYRTX-Mri1b	QVWVPCNPRSKKTTDDAGICRNTY	1 S-S/N-ter pyro-Q	2744.3163	2744.3211	722 ^c
U ₁₇ -MYRTX-Mri1c	YIIVAPCREGYVMVGNVCVEEY	1 S-S	2567.1535	2567.1705	— ^d
U ₁₈ -MYRTX-Mri1a	NHDPCPPQYAEALCLNGGT CFSVTIMGSDNYNCICAPGFRGWRCQEKDLDHPVNQ	3 S-S/O-linked fucose	6223.6608	6223.6751	9359
U ₁₉ -MYRTX-Mri1a	IDSAAIATLQGGTV	C-ter NH ₂	1314.7143	1314.7193	4380
U ₂₀ -MYRTX-Mri1a	GIMESLKQLSAKAEELIKKLLAKKA	—	2739.6296	2739.6298	37 466 ^c
CL24Contig3/4_1	TAEATAEATAKALAEAFEAIGTRA	—	2434.2339	— ^e	23 367 ^c

^aPTMs, post-translational modifications. ^bTPM, transcripts per millions. ^cSum of more than one transcript. ^dNot detected in the transcriptome. ^eNot validated through proteomics.

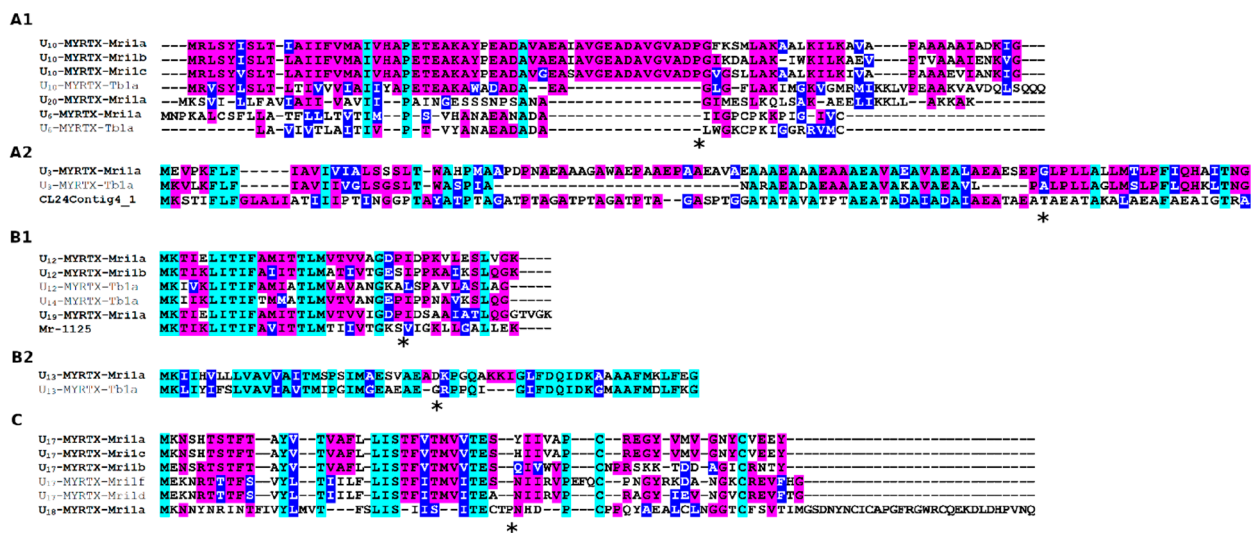


Figure 2. *M. rubida* venom peptide precursor alignments in superfamilies -A1, -A2, -B1, -B2, and -C. Sequence alignments were achieved using Muscle with SeaView 1:4.5.4.8–2 and edited with BoxShade 3.3.1–9. Gaps were added to enhance the alignments and align cleavage sites between mature and pre-prosequences. The *Tetramorium bicarinatum* precursors (MYRTX-Tb1x) for each superfamily are also shown (see Touchard et al., 2018).¹⁵ Identical residues are shown in magenta, similar residues are in blue and conserved residues are highlighted in cyan. “*” Indicates the cleavage site releasing mature peptides.

MYRTX-Mri1a yielded a 14-residue sequence IIGPCPKKPIGIVC and a 25-residue sequence GIMESLKQLSAKAEELIKKLLAKKA, respectively. The calculated monoisotopic masses of 1434.8091 Da for U₆-MYRTX-Mri1a and 2739.6296 Da for U₂₀-MYRTX-Mri1a are consistent with the measured masses of 1434.8104 and 2739.6298 Da, respectively. The Edman sequencing of U₁₂-MYRTX-Mri1a yielded a 10-residue sequence (IDPKVLES LV) having a theoretical mass of 1110.6488 Da while the native peptide measured mass was 1110.6649 Da. This measured mass is consistent with a C-terminal amidated peptide based on a predicted mass of 1110.6649 Da. The Edman degradation sequencing of U₁₈-MYRTX-Mri1a provided a partial 19-residue N-terminal sequence of NHDPCPPQYAEALCLNGG while the Edman

sequencing of U₁₇-MYRTX-Mri1b revealed two peptides present in equal amounts and having partial sequences of VPCNPRSKKTTDDXXICXN and XDDAGICRNTY (Supplementary Figure S1).

To characterize the toxins and their expression levels secreted by the venom glands of *M. rubida*, the mRNA transcriptome was sequenced using an Illumina Hiseq with a resultant ca. 29 million reads. *De novo* assembly with Trinity and Oases yielded 28 149 and 16 047 contigs with a median contig size of 1500 and 1893 bp, respectively. A combined database from both assemblies was used to search for peptide sequences generated from the proteomic analyses (LC-MS/MS and Edman degradation). Additionally, the signal sequences of the identified peptide toxins were searched

Peptide	Sequence	% ID	% S	Species
U ₃ -MYRTX-Mr1a	GLP-----LLLLMLLPFQHAIT-----N★	100	100	<i>Manica rubida</i>
U ₃ -MYRTX-Tb1a	VLP--ALPFLLAGMSLPFLOHKLIT-----N★	61	74	<i>Tetramorium bicarinatum</i>
U ₃ -MYRTX-Tb1b	LAPLIALPFLLAGMSLPFLHKLITGKPHHE★	29	42	<i>Tetramorium bicarinatum</i>
U ₃ -MYRTX-Tb1c	LAPVALLLSGLSLSLPFLHKLITGTMHHE★	29	42	<i>Tetramorium bicarinatum</i>
δ-PPOTX-Pc1a	FLP-----LILGSLMLPFWIQ--AIIHQQR--★	48	52	<i>Paraponera clavata</i>
U ₆ -MYRTX-Mr1a	IIGCPKPK--IG--IYC	100	100	<i>Manica rubida</i>
U ₆ -MYRTX-Tb1a	IWGKCPKI--GGRRTMC	40	60	<i>Tetramorium bicarinatum</i>
U ₆ -MYRTX-Tb1b	IWGKCPKI--GGRRTMC	47	60	<i>Tetramorium bicarinatum</i>
U ₆ -MYRTX-Tb1c	FRGCPKDMFRGRITMC	47	53	<i>Tetramorium bicarinatum</i>
MIITX ₁ -Mg6a	FRGPKLKI--KG--YKC	46	46	<i>Myrmecia gulosa</i>
U ₁₀ -MYRTX-Mr1a	GFKS---LAKAAIKILKAAPAAAAIAAK----★	100	100	<i>Manica rubida</i>
U ₁₀ -MYRTX-Mr1b	GKID---ALAKI--WKILKAEVPTVAATAENK----★	57	64	<i>Manica rubida</i>
U ₁₀ -MYRTX-Mr1c	GVGS---LAKAAIKILKIVAPAAAEVIAK----★	76	83	<i>Manica rubida</i>
MIITX ₁ -Mg1a	GICR---LICKAKKGAATAAPAAANFAKRAEA★	36	48	<i>Myrmecia gulosa</i>
U ₁₀ -MYRTX-Tb1a	GLGFLAKIMCKVGMRIKIVPEAAKVAVALQLSQQ	19	42	<i>Tetramorium bicarinatum</i>
U ₁₂ -MYRTX-Mr1a	IDPKVLESIV★	100	100	<i>Manica rubida</i>
U ₁₂ -MYRTX-Tb1a	ESPMVASTLA★	50	60	<i>Tetramorium bicarinatum</i>
U ₁₂ -MYRTX-Mr1a	IDPKVLESIV★	80	90	<i>Myrmica rubra</i>
odvp4	IDPKVQSLI★	60	100	<i>Orancistrocerus drewseni</i>
EPVP3S	INPKVQSLI★	60	90	<i>Eumenes pomiformis</i>
U ₁₂ -MYRTX-Mr1b	IPPKAQKSLQ★	50	70	<i>Manica rubida</i>
U ₁₄ -MYRTX-Tb1a	IPPNAQKSLQ★	40	60	<i>Tetramorium bicarinatum</i>
U ₁₃ -MYRTX-Mr1a	DKPGQAKKIGFDQIDKMAAFMKLFE★	100	100	<i>Manica rubida</i>
U ₁₃ -MYRTX-Tb1a	R---PPOIGFDQIDKMAAFMDLAK★	61	74	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Mr1a	HTIVAF-----CREG--YV-MGNVGVVEE--Y	100	100	<i>Manica rubida</i>
U ₁₇ -MYRTX-Mr1b	QIHWVP-----CNPSSKKT-DDAGCRNT--Y	21	25	<i>Manica rubida</i>
U ₁₇ -MYRTX-Tb1a	TIINAPN-----RCPPG--HVV-VKGCRTA--★	39	43	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Tb1b	TIIDVPL-----QCPG--TVK-VGNICRVI--F	38	46	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Tb1c	TIIDVPL-----QCPG--KVK-VGNICRVI--F	42	46	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Tb1d	NIIRVP-----CRAG--YI-EINQVCREMET★	43	52	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Tb1e	NIIRKAPLF-----PCPNG--YIRDYKDCREHII-E★	33	41	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Tb1f	NIIRVPEF-----QCPG--YRKDANGCREVFEH★	30	33	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Tb1g	HTIDTRSR-----LCPG--SRRSTTCRITV--	24	28	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Tb1h	HTIRVE-----CRAG--YK-EIRGICRRLT★	39	48	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Tb1i	RVLDARE-----RCPSG--YQMDGSGCRKIFGR	22	30	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Tb1j	YIIRVPEFPPPTCPG--ETM-VGKICRVIY--★	29	32	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Tb1k	YIIRVPEF--PCPNG--YMRDYEPCREHIF-E	33	41	<i>Tetramorium bicarinatum</i>
U ₁₇ -MYRTX-Tb1l	DIIDVPLR-VSKCPG--SRMSIGCCRRVSKR	23	30	<i>Tetramorium bicarinatum</i>
secapin1	YIIDVPP---RCPPG--SKF-VKNICRVIIV-P	28	36	<i>Apis mellifera</i>
secapin2	YIIDVPP---RCPPG--SKF-VKNICRVIIV-P	28	28	<i>Polistes hebraeus</i>
secapin3	YIINVEP---RCPPG--SKF-VKNICRVIIV-P	28	36	<i>Vespa maculifrons</i>
U ₁₈ -MYRTX-Mr1a	N---HDPPPPQYEAALCINGGTCESWTIIGSDIYNGITCAPGIRGIRCOEQLDHFVNC	100	100	<i>Manica rubida</i>
MIITX ₂ -Mg1a	DISDYDPCSDLLK-DYCLHG-DCHFKEIN--SPACRCYTGIVGSRCEHIHNN-----	28	44	<i>Myrmecia gulosa</i>

Figure 3. Myrmecitoxin alignments in *M. rubida* venom with other Hymenoptera venom peptides.^{5,8,9,12,26,27} Resulting alignments using the T-Coffee alignment program were edited with BoxShade 3.3.1–9. Identical residues are shown in black while similar residues are highlighted in gray. Both percentage identity (% ID) and similarity (% S) are relative to the first peptide of each group. Red stars denote C-terminal amidated peptides, the threonine residues inside boxes with red dashed lines are O-glycosylated and N-terminal pyrrolutamate is represented in boxes with solid red lines.

against transcriptomics databases in order to find other venom peptides in the *M. rubida* venom sharing the same signal sequence. The validation of mature peptide sequences was verified manually using mass spectrometry data by matching theoretical masses with those measured (Table 1). This approach permitted us to identify 14 unique peptide precursors (Figure 2). It is worth noting that the presence of one putative venom peptide (CL24Contig3/4_1) was not confirmed by our proteomic data and was therefore not considered a genuine venom peptide toxin of *M. rubida*. One additional venom peptide (U₁₇-MYRTX-Mr1c) was identified proteomically but was not found in the transcriptome and was integrated into the venom peptidome as a probable component of the venom. The venom peptides accounted for ca. 51% of the total venom gland expression (Figure 1C and Supplementary Table S4).

The identified mature peptide lengths ranged from 10 to 51 residues and accounted for 97.4% of the venom peptide content. According to the proteomic data, the venom peptidome of *M. rubida* was dominated by four myrmecitoxins (i.e., U₁₀-MYRTX-Mr1a, -Mr1b, -Mr1c, and U₂₀-MYRTX-

Mr1a) accounting for 83% of the venom peptide content (Figure 1) while the expression rate (in TPM) of the transcripts encoding these four peptides represented 75% of the total expression of the transcripts encoding venom peptides (Figure 1D). Of the identified peptides, nine were linear, four had one disulfide bond, and one had three disulfide bonds.

M. rubida Venom Polypeptide Composition

A total of 13 myrmecitoxin precursors and 14 mature myrmecitoxins were identified in the *M. rubida* venom. The myrmecitoxin precursors reported here have been classified into five precursor superfamilies (i.e., superfamily-A1, -A2, -B1, -B2, and -C) corresponding to the same precursors described in the venom of *T. bicarinatum* (see Touchard et al. 2018)¹⁵ based on their pre-propeptide sequence identity (Figure 2). Briefly, superfamily-A1 is linked to aculeatoxins, a vast gene superfamily encoding most of the peptides currently described in hymenopteran venoms,¹⁴ whereas superfamilies-A2, -B1, and -B2 contain peptides that, thus far, are only described in *T. bicarinatum* venom, and superfamily-C includes secapins, a

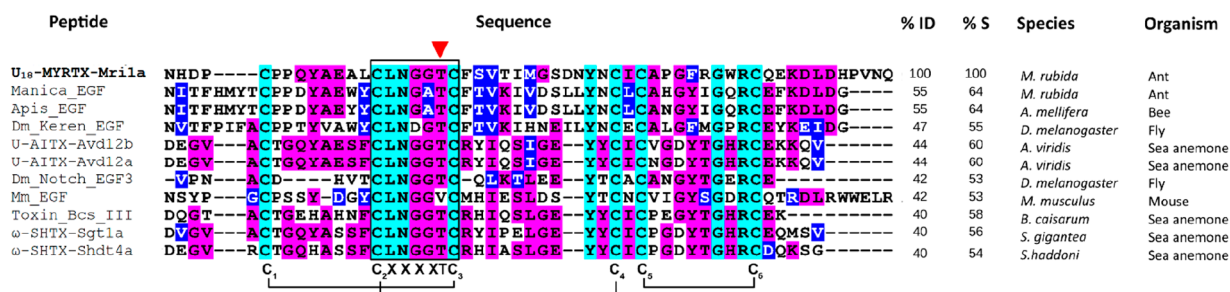


Figure 4. Multiple sequence alignment of EGF-like peptide toxins and EGF domains from *Mus musculus*, *Drosophila melanogaster*, *Apis mellifera*, and *M. rubida* EGF-like along with the EGF3 domain of *Mus musculus* Neurogenic locus Notch. The alignments resulting from the Muscle alignment program were edited with BOXSHADE 3.3.1–9. Identical residues are highlighted in magenta. Similar residues in the peptide sequences are highlighted in blue while conserved residues are shown in cyan. Both percentage identity (% ID) and similarity (% S) are relative to the U₁₈-MYRTX-Mri1a. The residues within the black boxes are the consensus sequence for the addition of *O*-fucose by *O*-fucosyltransferase 1 (POFUT1) and the red inverted triangle indicates the modified residue. EGF-like toxins from sea anemone venoms have the consensus sequence for the addition of *O*-fucose by POFUT1 but this modification was not reported for these peptides (accession number: ω-SHTX-Sgt1a, Q76CA1; Toxin Bcs III 15.09, P86468; ω-SHTX-Shdt4a, BAG12826/B1BSJ0; U-AITX-Avd12b, P0DMZ0; U-AITX-Avd12a, P0DMY9; mouse EGF, P01132; Dm_Notch_EGF3, P07207; Dm_Keren_EGF, Q9VVJ6; Apis_EGF, GAZV01019061.1). Manica_EGF was isolated after blast searches on the *M. rubida* venom gland transcriptome with Keren as the subject sequence. The disulfide connectivity was determined through Prosite-ProRule annotation for EGF domain (PRU00076).

class of cysteine-containing peptides reported in the venoms of wasps, bees and *T. bicarinatum*.^{9,24}

On the basis of the amino acid sequence alignments, we classified the mature venom peptides from *M. rubida* into ten groups. Since the molecular target of the myrmecitoxins described herein has not yet been identified, we used the activity descriptor “U” indicative of a toxin with a unknown pharmacological target. Several mature peptides from *M. rubida* shared sequence identities with toxins from *T. bicarinatum* venom, and consequently we used the same subscript to indicate the unknown activity descriptor prefix (U_x) for myrmecitoxins with at least 40% sequence similarity (Figure 3).

The peptide U₃-MYRTX-Mri1a shared a sequence identity with ant venom peptides from *T. bicarinatum* (Myrmicinae) and *P. clavata* (Paraponerinae) (61% and 48% identities with U₃-MYRTX-Tb1a and paraponeritoxin (δ-PPOTX-Pc1a), respectively).^{5,9} The molecular target of U₃-MYRTX-Tb1a has not yet been determined but δ-PPOTX-Pc1a inhibits the inactivation of sodium channels. As noted for the venom of *T. bicarinatum*, U₃-MYRTX-Mri1a seems to be very sensitive to proteases with several fragments detected in the crude venom through MS (Supplementary Table S3). In the case where the crude venom was kept for several minutes at room temperature, the U₃ peptide completely disappeared from the chromatograms which indicates a rapid degradation. This peptide (including fragments) accounted for ca. 7% of the entire peptidome but this percentage is likely underestimated due to the rapid degradation of this toxin. This peptide is also encoded by one of most expressed transcripts in venom glands (78 932 TPM; Table 1) and presumably has an important functional role.

U₆-MYRTX-Mri1a is a 14-residue peptide containing two cysteines and sharing more than 40% sequence identity with U₆ peptides from *T. bicarinatum* and 46% with MIITX₁-Mg6a from *My. gulosa*.^{14,15} These peptides possessed a net positive charge due to several cationic residues, and were also rich in aliphatic amino acids.

The alignment of the sequence of U₁₀-MYRTX-Mri1a, -Mri1b, and -Mri1c revealed strictly conserved residues that were classified into the same isotoxin group. These peptides

have substantial sequence identities and similarities with U₁₀-MYRTX-Tb1a and MIITX₁-Mg1a from *T. bicarinatum* and *My. gulosa*, respectively. They are rich in both cationic and anionic amino acids, having a positive net charge and a high alanine residue content. Two other linear polycationic peptides, U₁₃-MYRTX-Mri1a and U₂₀-MYRTX-Mri1a, are present in the venom with a relative abundance of ca. 4% and 14% in the venom peptidome, respectively.

U₁₂-MYRTX-Mri1a and Mri1b were linear decapeptides and shared more than 60% sequence similarity with venom peptides from *T. bicarinatum* and *Myrmica rubra*.^{9,12} These peptides were weakly expressed in the venom glands and their pharmacological target has not yet been determined. However, insecticidal activity has already been noted on aphids for U₁₂-MYRTX-Mri1a²⁰ as well as the peptide U₁₂-MYRTX-Mri1a from *Myr. rubra*.¹⁹ The peptide U₁₉-MYRTX-Mri1a was linear without any sequence similarity with other ant venom peptides.

As described for both *T. bicarinatum* and bee venoms,^{9,24} scapin peptides were found in *M. rubida* (U₁₇-MYRTX-Mri1a, -Mri1b, and -Mri1c), albeit in relatively low abundance since these peptides account for less than 0.02% of the peptide content. The MS and MS/MS data are consistent with a N-terminal pyroglutamate modification for the scapin U₁₇-MYRTX-Mri1b (Supplementary Figure S1). Among superfamily C precursors, the mature peptide U₁₈-MYRTX-Mri1a containing three disulfide bonds is predicted by ScanProsite to have an EGF domain (Epidermal Growth Factor) which is a peculiar structural domain for animal venoms only found in the venoms of sea anemones and one ant (Figure 4). The EGF-like toxin U₁₈-MYRTX-Mri1a from *M. rubida* exhibits a sequence with a missing 146 Da mass unit in comparison to the MS measurement which is consistent with an additional fucose glycan. Moreover, the sequence -C₁₄LNGGTC₂₀- of U₁₈-MYRTX-Mri1a is consistent with *O*-fucosylation on threonine 19 between the second and the third conserved cysteine of the EGF domain. Indeed, the -C₂XXXXS/TC₃- is a consensus sequence of threonine/serine *O*-fucosylation found in repeated EGF domains of Neurogenic locus Notch protein and its ligands.^{31,32} The *O*-fucosylation of Notch EGF domain is necessary for its function^{33,34} and maturation to the cell

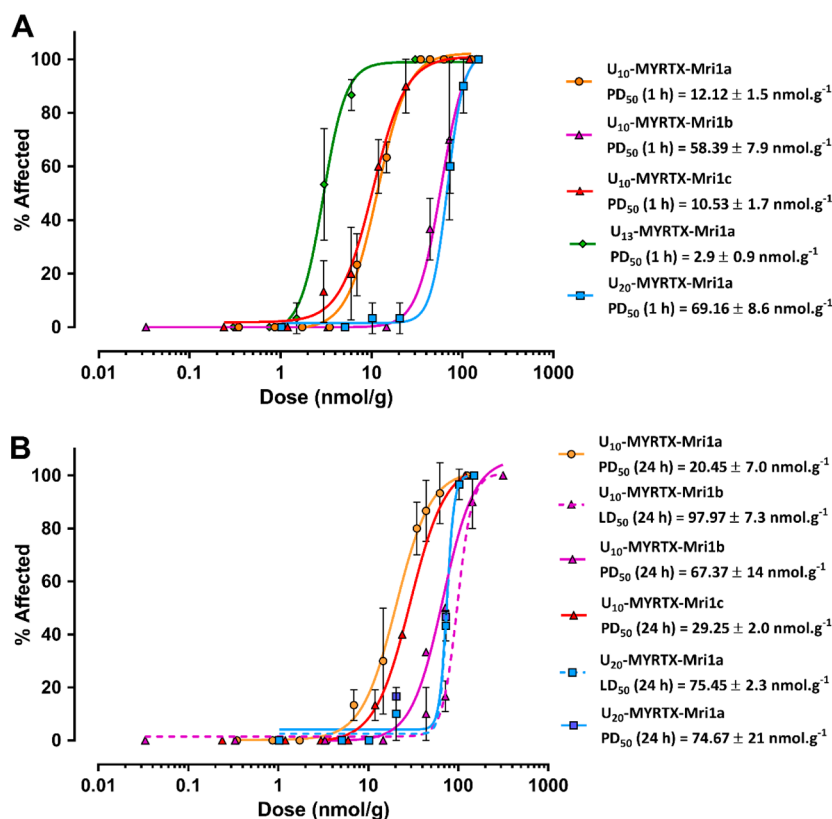


Figure 5. Dose–response curves for *Lucilia caesar* blowflies injected with U₁₀-MYRTX-Mri1a, -Mri1b, -Mri1c, U₁₃-MYRTX-Mri1a, and U₂₀-MYRTX-Mri1a, 1 h (A) and 24 h (B) following injection. Values represent the mean ± SE of three experiments. Note that after 24 h, most flies injected with U₁₃-MYRTX-Mri1a were unaffected in the dose range tested, while most of the flies injected with high doses of U₁₀-MYRTX-Mri1a and -Mri1c were paralyzed.

surface.³⁵ This reaction is carried out by the *O*-fucosyltransferase 1 (POFUT1), which is expressed in *M. rubida* venom glands albeit at very low levels. This enzyme is therefore specific to both Notch and Notch ligand EGF domains. However, it is possible that another glucosyl transferase could be involved in U₁₈-MYRTX-Mri1a fucosylation.³³

Insecticidal Activity of the Venom Peptides

Manica rubida is a predatory ant species that subdues prey using its venom. The injection in blowfly ($n = 5$) of 2 μ L of crude venom, equivalent to a single worker venom reservoir, showed an immediate and irreversible paralytic activity leading to death in 24 h. Then, the venom of 40 ant workers was fractionated on a RP-HPLC column into 60 fractions (one fraction per minute) and injected into blowflies. Of these 60 fractions, 14 displayed paralytic and insecticidal activities during the three time periods monitored (i.e., 5 min, 1 h, and 24 h). Further LC-MS of the active fractions revealed six different putative neurotoxic peptides (Supplementary Table S5) which were chemically synthesized for further toxicity testing (i.e., U₃-MYRTX-Mri1a, U₁₀-MYRTX-Mri1a, -Mri1b, -Mri1c, U₁₃-MYRTX-Mri1a, and U₂₀-MYRTX-Mri1a). Although the fractions with U₁₂-MYRTX-Mri1a showed no insecticidal activity, this peptide was also synthesized and tested on blowflies since this peptide was reported as being insecticidal on aphids.²⁰ The injection of synthetic U₁₀-Mri1a, -Mri1b, -Mri1c, U₁₃-Mri1a, and U₂₀-Mri1a resulted in a rapid contractile paralysis that incapacitates flies within minutes (immediate paralysis for the highest doses), the PD₅₀ values measured 1 h after injection were 12.1 ± 1.5 nmol.g⁻¹, 58.4 ±

7.9 nmol.g⁻¹, 10.5 ± 1.7 nmol.g⁻¹, 2.9 ± 0.9 nmol.g⁻¹, and 69.2 ± 8.6 nmol.g⁻¹ ($n = 3$), respectively (Figure 5A). The paralysis induced by all peptides was reversible for the lowest and moderate doses after 24 h following the injection, but was irreversible or lethal for the higher doses (except for U₁₃-Mri1a, for which only 10% mortality was noted at 150 nmol.g⁻¹, whereas 60% of the flies were noted as normal) (Figure 5B). Consequently, the lethality of these five neurotoxic peptides is weak with LD₅₀(24 h) values estimated to be 98.0 ± 7.3 nmol.g⁻¹ and 75.4 ± 2.3 nmol.g⁻¹ ($n = 3$) for U₁₀-Mri1b and U₂₀-Mri1a, respectively. Despite a high amino acid sequence identity with U₁₀-Mri1b, both U₁₀-Mri1a and -Mri1c were in most cases not lethal after 24 h in the dose range tested even though the paralysis was irreversible (paralyzed flies were monitored over 48 h). The PD₅₀ values measured 24 h after injection were 20.5 ± 7.0 nmol.g⁻¹ and 29.3 ± 2.0 nmol.g⁻¹ ($n = 3$) for U₁₀-Mri1a and U₁₀-Mri1c, respectively. No paralytic or lethal effects of U₃-MYRTX-Mri1a were observed within the dose range tested (0.05 to 54.6 nmol.g⁻¹) 48 h following injection. Therefore, the biological effects of the insecticidal fraction containing U₃-MYRTX-Mri1a (F43; Supplementary Table S5) is likely linked to the unidentified protein which coelutes with U₃-MYRTX-Mri1a. Although Heep and colleagues have reported insecticidal activity on aphids for U₁₂-MYRTX-Mri1a,²⁰ we did not observe neurotoxic activities on blowflies at the highest dose of 203 nmol.g⁻¹ (1 μ L of 3.8 mmol L⁻¹ injected/fly).

DISCUSSION

Integrative “-omics” methodologies have recently been applied to four ant species providing a comprehensive characterization of their venom peptidomes.^{13–16} Here, we employed a venomomics strategy to characterize the venom peptidome of the red ant *M. rubida* for which a previous MS-based analysis revealed hundreds of masses corresponding to peptides, among which some were structured by one or three disulfide bonds.⁴ Interestingly, while we only obtained 13 transcripts encoding myrmecitoxin peptides, we found that 450 of the 509 sequence tags from the MS data (ca. 80%) were linked to the 13 identified transcripts (see [Supplementary Table S3](#)). The striking discrepancy between peptide count from mass spectrometry data and the number of venom precursors is likely due to peptide degradation in the crude venom. Indeed, it is now well established that ant venom peptidomes, including *M. rubida* venom, are mainly composed of linear and polycationic peptides which generally have low stability and are very susceptible to degradation by proteases.^{36,37} The multiple fragments generated from peptide degradation may explain the high number of masses detected by highly sensitive mass spectrometry. Furthermore, several MS investigations have already reported intact peptides plus multiple truncated/fragmented forms in ant venoms.^{13,15,38} Thus, the peptidic venom composition of *M. rubida* appeared to be relatively simple in comparison to nonhymenopteran venoms such as those of cone snails, spiders, or scorpions. However, such “simplicity” was also noted for the venom peptidome of ants belonging to other subfamilies such as *My. gulosa* (Myrmeciniinae) and *Odontomachus monticola* (Ponerinae), whereas the peptidome of *T. bicarinatum* (also a Myrmeciniinae) is comparatively complex.^{13–15} Aside from the peptides U₁₉- and U₂₀-Mri1a, the myrmecitoxins described here present similarities with those of other myrmecine species, supporting the idea that ants from the same clade exhibit similar venom profiles.

Insecticidal Peptides of *M. rubida* Venom

Although insecticidal peptides have been reported for a few ant venoms, this study is the first that deals with the insecticidal arsenal of an ant species belonging to the subfamily Myrmeciniinae. Indeed, the myrmecotoxins U₁₀-Mri1a, -Mri1b, -Mri1c, U₁₃-Mri1a, and U₂₀-Mri1a, which dominated the venom peptidome of *M. rubida*, have a paralytic effect on blowflies, this effect being irreversible at high doses. Nevertheless, each of these toxins has only a weakly potent insecticidal activity, having LD₅₀ (24 h), higher than 75 nmol·g⁻¹ that is 325-fold higher than the most potent insecticidal spider venom peptide (i.e., β -Diguetoxin-Dc1a)²⁹ which was tested under similar conditions. Therefore, the rapid death observed for blowflies after the injection of *M. rubida* crude venom is suggestive of a synergic action of insecticidal peptides as noted for ponerinins from the venom of the ponerine ant *Neoponera goeldii*.¹⁰ The five *M. rubida* insecticidal peptides, rich in cationic and aliphatic residues, are predicted to have a secondary structure forming a full-length α -helix or two α -helices separated by a loop (noted using the PepFold3 server;³⁹ [Supplementary Figure S2](#)). The helix wheel projections also showed amphipathic features with distinguishable hydrophobic and hydrophilic domains, suggesting that these myrmecitoxins are membrane-active as is known for most insecticidal linear peptides described in ant venoms. Note that similar peptides (e.g., MIITX₁-Mg1a) also paralyze

arthropods and elicit pain in mammals via pore formation in neuronal cell membranes.¹⁴ Thus, offensive (insecticidal activity) and defensive (triggering pain in vertebrates) functions are combined and are favored by the abundance of these peptides in the *M. rubida* venom peptidome. Furthermore, the antimicrobial properties of membrane-active venom peptides^{9,10} are involved in protecting the colony from infection linked to the storage and consumption of captured prey.

EGF-like Venom Peptide

The *M. rubida* venom peptidome also revealed the presence of an EGF-like toxin (U₁₈-MYRTX-Mri1a) sharing sequence similarities with the EGF-like toxins described in the venoms of the ant *My. gulosa*¹⁴ and several sea anemones,^{40–42} indicating that this scaffold has been retained in several invertebrate clades. The EGF-like toxin isolated from the venom of the sea anemone *Stichodactyla gigantea* has a potent paralytic effect on crabs (ED₅₀ of 215 μ g/kg) and exhibits a weak EGF activity on human A431 cells (i.e., 500–1000 times less potency than the human EGF).⁴² However, the insecticidal assays we conducted using the *M. rubida* venom fractions suggest that U₁₈-MYRTX-Mri1a is not toxic for blowflies. Considering that EGFs are involved in gut epithelial stem cell proliferation during development or after injury,^{43,44} the U₁₈-MYRTX-Mri1a peptide might be implicated in the renewal of both the venom gland and the sac epithelia, both being constantly exposed to insecticidal peptides that presumably trigger cell membrane disruption. This hypothesis is further supported by the highest sequence identity percentage (47% identity) shared by mature U₁₈-MYRTX-Mri1a with the EGF-like domain of Keren. Indeed, apart from its role in developmental and olfactory learning, like *M. musculus* EGF, the Keren peptide is involved in gut epithelium renewal.^{45,46} Indeed, other growth factors have also been reported as components of other animal venoms. For instance, the Nerve Growth Factors (“NGFs”), which are involved in the maintenance and proliferation of neurons, were noted in snake venoms with an activity similar to that of mouse NGFs (no toxic effects on prey).^{47–49} Also, the venom NGF of the snake *Naja kaouthia* is a potent inhibitor of metalloproteinases, suggesting a protective role.⁵⁰ Nevertheless, the potential protective role of the *M. rubida* U₁₈-MYRTX-Mri1a against cytolytic peptides remains to be demonstrated.

CONCLUDING REMARKS

Similar to other hymenopteran venoms, *M. rubida* venom is mostly composed of small polycationic and amphiphilic peptides. These peptides are the main components responsible for the insecticidal activity of the venom leading to the rapid and irreversible contractile paralysis of insects. Remarkably, one peptide isolated and characterized in this study is structured by three disulfide bonds. This is an EGF-like toxin with an *O*-fucose modification, a post-translational modification never before described in animal venoms. We hypothesized that this unusual venom peptide protects the venom apparatus from the cytolytic components of the venom. Several other peptides from the *M. rubida* venom share sequence identities with other ant venom toxins, suggesting a conserved function across ant venoms, but their biological role remains unknown. In conclusion, this study, combined with venom investigations conducted on other ant clades, confirms

that several stinging ants have evolved small polycationic, amphiphilic peptides in order to subdue their prey.

■ ASSOCIATED CONTENT

SI Supporting Information

Supporting Information The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00048>.

Table S1: Concentration of each synthetic myrmecitoxin used for insecticidal assays; Table S2: Peptide mass fingerprint of *M. rubida* venom; Table S3: The list of sequence tags yielded from Peaks software interpretation of MS/MS spectra; Table S4: Most expressed toxins from the venom apparatus of *M. rubida*; Table S5: Insecticidal activity of *M. rubida* venom fractions on blowflies; Figure S1: Integrative methodology for the sequencing of U₁₇-MYRTX-Mri1b; Figure S2: *De novo* structure prediction of insecticidal myrmecitoxins of *M. rubida* venom (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Axel Touchard – Équipe BTSB-EA 7417, Université de Toulouse, Institut National Universitaire Jean-François Champollion, 81012 Albi, France; orcid.org/0000-0002-7766-0088; Phone: +(33)5 63 48 64 32; Email: axel.touchard2@gmail.com

Authors

Samira R. Aili – Neurotoxin Research Group, School of Life Sciences, University of Technology Sydney, Ultimo, NSW 2007, Australia

Nathan Téné – Équipe BTSB-EA 7417, Université de Toulouse, Institut National Universitaire Jean-François Champollion, 81012 Albi, France

Valentine Barassé – Équipe BTSB-EA 7417, Université de Toulouse, Institut National Universitaire Jean-François Champollion, 81012 Albi, France

Christophe Klopp – Unité de Mathématique et Informatique Appliquées de Toulouse, UR0875, INRA Toulouse, 31326 Castanet-Tolosan, France

Alain Dejean – CNRS, UMR EcoFoG, AgroParisTech, CIRAD, INRAE, Université des Antilles, Université de la Guyane, 97310 Kourou, France; Ecolab, Université de Toulouse, CNRS, INPT, UPS, 31000 Toulouse, France

R. Manjunatha Kini – Protein Science Laboratory, Department of Biological Sciences, Faculty of Science and Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, 117543, Singapore

Mrinalini – Protein Science Laboratory, Department of Biological Sciences, Faculty of Science, National University of Singapore, 117543, Singapore

Laurent Coquet – CNRS UMR 6270, Normandie University, UNIROUEN, PISSARO, 76130 Mont-Saint-Aignan, France

Thierry Jouenne – CNRS UMR 6270, Normandie University, UNIROUEN, PISSARO, 76130 Mont-Saint-Aignan, France

Benjamin Lefranc – Inserm U 1239, Normandie University, UNIROUEN, Plate-forme de Recherche en Imagerie Cellulaire de Normandie (PRIMACEN), 76000 Rouen, France

Jérôme Leprince – Inserm U 1239, Normandie University, UNIROUEN, Plate-forme de Recherche en Imagerie Cellulaire

de Normandie (PRIMACEN), 76000 Rouen, France;

orcid.org/0000-0002-7814-9927

Pierre Escoubas – VenomeTech, 06560 Valbonne, France

Graham M. Nicholson – Neurotoxin Research Group, School of Life Sciences, University of Technology Sydney, Ultimo, NSW 2007, Australia; orcid.org/0000-0002-4277-4296

Michel Treilhou – Équipe BTSB-EA 7417, Université de Toulouse, Institut National Universitaire Jean-François Champollion, 81012 Albi, France

Elsa Bonnafé – Équipe BTSB-EA 7417, Université de Toulouse, Institut National Universitaire Jean-François Champollion, 81012 Albi, France

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jproteome.0c00048>

Author Contributions

#Michel Treilhou and Elsa Bonnafé contributed equally. AT, EB, MT, RMK, PE, and GMN jointly directed this work. AT, SRA, NT, LC, BL, JL, CK, Mrinalini, and EB generated the data. AT and EB drafted the manuscript and interpreted the results. AT, MT, EB, and AD wrote the paper. All authors contributed important intellectual content to the study and have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The PISSARO Platform is cofunded by The European Union and Région Normandie through the European Regional Development Fund (ERDF). This research was cofunded by the Southeast Asian Biodiversity Genomics Center, National University of Singapore, Grant No. R-154-000-648-646. We would like to thank Jean-Pierre Andrieu from the IBS platform of the Partnership for Structural Biology and the Institut de Biologie Structurale in Grenoble (PSB/IBS), for assistance and access to the Protein Sequencing Facility. We are grateful to Andrea Yockey for proofreading the manuscript. The collection of *Manica rubida* workers in France is subject to a declarative procedure (NOR: TREL1820249A/S4) for non-commercial use at the competent administrative authority, in accordance with Article L.412-7 of the French environmental code.

■ REFERENCES

- (1) Cerdá, X.; Dejean, A. Predation by ants on arthropods and other animals. In *Predation in the Hymenoptera: An Evolutionary Perspective*; Polidori, C., Ed.; TransWorld Research Network: Kerala, India, 2011; pp 39–78.
- (2) Wilson, E. O. *Success and Dominance in Ecosystems: The Case of the Social Insects*; Ecology Institute: Oldendorf/Luhe, Germany, 1990; Vol. 2.
- (3) Aili, S. R.; Touchard, A.; Escoubas, P.; Padula, M. P.; Orivel, J.; Dejean, A.; Nicholson, G. M. Diversity of peptide toxins from stinging ant venoms. *Toxicon* **2014**, *92*, 166–178.
- (4) Touchard, A.; Koh, J. M. S.; Aili, S. R.; Dejean, A.; Nicholson, G. M.; Orivel, J.; Escoubas, P. The complexity and structural diversity of ant venom peptidomes is revealed by mass spectrometry profiling. *Rapid Commun. Mass Spectrom.* **2015**, *29*, 385–396.
- (5) Rifflet, A.; Gavalda, S.; Téné, N.; Orivel, J.; Leprince, J.; Guilhaudis, L.; Génin, E.; Vétillard, A.; Treilhou, M. Identification and characterization of a novel antimicrobial peptide from the venom of the ant *Tetramorium bicarinatum*. *Peptides* **2012**, *38* (2), 363–370.

- (6) Pluzhnikov, K. A.; Kozlov, S. A.; Vassilevski, A. A.; Vorontsova, O. V.; Feofanov, A. V.; Grishin, E. V. Linear antimicrobial peptides from *Ectatomma quadridens* ant venom. *Biochimie* **2014**, *107*, 211–215.
- (7) Zelezetsky, I.; Pag, U.; Antcheva, N.; Sahl, H. G.; Tossi, A. Identification and optimization of an antimicrobial peptide from the ant venom toxin pilosulin. *Arch. Biochem. Biophys.* **2005**, *434* (2), 358–64.
- (8) Tani, N.; Kazuma, K.; Ohtsuka, Y.; Shigeri, Y.; Masuko, K.; Konno, K.; Inagaki, H. Mass spectrometry analysis and biological characterization of the predatory ant *Odontomachus monticola* venom and venom sac components. *Toxins* **2019**, *11* (1), 50.
- (9) Cologna, C. T.; Cardoso, J. d. S.; Jourdan, E.; Degueldre, M.; Upert, G.; Gilles, N.; Uetanabaro, A. P. T.; Costa Neto, E. M.; Thonart, P.; de Pauw, E. Peptidomic comparison and characterization of the major components of the venom of the giant ant *Dinoponera quadriciceps* collected in four different areas of Brazil. *J. Proteomics* **2013**, *94*, 413–422.
- (10) Orivel, J.; Redeker, V.; Le Caer, J. P.; Krier, F.; Revol-Junelles, A. M.; Longeon, A.; Chaffotte, A.; Dejean, A.; Rossier, J. Ponericsins, new antibacterial and insecticidal peptides from the venom of the ant *Pachycondyla goeldii*. *J. Biol. Chem.* **2001**, *276* (21), 17823–9.
- (11) Piek, T.; Duval, A.; Hue, B.; Karst, H.; Lapied, B.; Mantel, P.; Nakajima, T.; Pelhate, M.; Schmidt, J. O. Poneratoxin, a novel peptide neurotoxin from the venom of the ant, *Paraponera clavata*. *Comp. Biochem. Physiol., C: Comp. Pharmacol.* **1991**, *99* (3), 487–495.
- (12) 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. Isolation and characterization of a structurally unique β -hairpin venom peptide from the predatory ant *Anochetus emarginatus*. *Biochim. Biophys. Acta, Gen. Subj.* **2016**, *1860* (11A), 2553–2562.
- (13) Kazuma, K.; Masuko, K.; Konno, K.; Inagaki, H. Combined venom gland transcriptomic and venom peptidomic analysis of the predatory ant *Odontomachus monticola*. *Toxins* **2017**, *9* (10), 323.
- (14) Robinson, S. D.; Mueller, A.; Clayton, D.; Starobova, H.; Hamilton, B. R.; Payne, R. J.; Vetter, I.; King, G. F.; Undheim, E. A. A comprehensive portrait of the venom of the giant red bull ant, *Myrmecia gulosa*, reveals a hyperdiverse hymenopteran toxin gene family. *Sci. Adv.* **2018**, *4* (9), No. eaau4640.
- (15) Touchard, A.; Téné, N.; Song, P. C. T.; Lefranc, B.; Leprince, J.; Treilhou, M.; Bonnafé, E. Deciphering the molecular diversity of an ant venom peptidome through a venomics approach. *J. Proteome Res.* **2018**, *17* (10), 3503–3516.
- (16) Barassé, V.; Touchard, A.; Téné, N.; Tindo, M.; Kenne, M.; Klopp, C.; Dejean, A.; Bonnafé, E.; Treilhou, M. The peptide venom composition of the fierce stinging ant *Tetraponera aethiops* (Formicidae: Pseudomyrmecinae). *Toxins* **2019**, *11* (12), 732.
- (17) Morgan, E. D. Chemical sorcery for sociality: exocrine secretions of ants (Hymenoptera: Formicidae). *Myrmecol. News* **2008**, *11*, 79–90.
- (18) Touchard, A.; Aili, S. R.; Fox, E. G. P.; Escoubas, P.; Orivel, J.; Nicholson, G. M.; Dejean, A. The biochemical toxin arsenal from ant venoms. *Toxins* **2016**, *8* (1), 30.
- (19) Heep, J.; Klaus, A.; Kessel, T.; Seip, M.; Vilcinskas, A.; Skaljic, M. Proteomic analysis of the venom from the ruby ant *Myrmica rubra* and the isolation of a novel insecticidal decapeptide. *Insects* **2019**, *10* (2), 42.
- (20) Heep, J.; Skaljic, M.; Grotmann, J.; Kessel, T.; Seip, M.; Schmidtberg, H.; Vilcinskas, A. Identification and functional characterization of a novel insecticidal decapeptide from the myrmicine ant *Manica rubida*. *Toxins* **2019**, *11* (10), 562.
- (21) Schmidt, J. O.; Blum, M. S. Biochemical constituents on venom of Harvester ant, *Pogonomyrmex badius*. *Comp. Biochem. Physiol., C: Comp. Pharmacol.* **1978**, *61* (1), 239–247.
- (22) von Sicard, N. A.; Candy, D. J.; Anderson, M. The biochemical composition of venom from the pavement ant (*Tetramorium caespitum* L.). *Toxicon* **1989**, *27* (10), 1127–33.
- (23) Blatrix, R.; Galkowski, C.; Lebas, C.; Wegnez, P. *Guide des Fourmis de France*; Delachaux et Niestlé, 2013.
- (24) Monnin, T.; Espadaler, X.; Lenoir, A. *Guide des Fourmis de France*; Belin, 2014.
- (25) Cabau, C.; Escudé, F.; Djari, A.; Guiguen, Y.; Bobe, J.; Klopp, C. Compacting and correcting Trinity and Oases RNA-Seq de novo assemblies. *PeerJ* **2017**, *5*, No. e2988.
- (26) Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **2010**, *26* (5), 589–595.
- (27) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25* (16), 2078–2079.
- (28) Guo, S.; Herzig, V.; King, G. F. Dipteran toxicity assays for determining the oral insecticidal activity of venoms and toxins. *Toxicon* **2018**, *150*, 297–303.
- (29) Bende, N. S.; Dziemborowicz, S.; Mobli, M.; Herzig, V.; Gilchrist, J.; Wagner, J.; Nicholson, G. M.; King, G. F.; Bosmans, F. A distinct sodium channel voltage-sensor locus determines insect selectivity of the spider toxin Dc1a. *Nat. Commun.* **2014**, *5*, 4350–4350.
- (30) King, G. F.; Gentz, M. C.; Escoubas, P.; Nicholson, G. M. A rational nomenclature for naming peptide toxins from spiders and other venomous animals. *Toxicon* **2008**, *52* (2), 264–276.
- (31) Harvey, B. M.; Rana, N. A.; Moss, H.; Leonardi, J.; Jafar-Nejad, H.; Haltiwanger, R. S. Mapping sites of O-glycosylation and fringe elongation on *Drosophila* Notch. *J. Biol. Chem.* **2016**, *291* (31), 16348–16360.
- (32) Müller, J.; Rana, N. A.; Serth, K.; Kakuda, S.; Haltiwanger, R. S.; Gossler, A. O-fucosylation of the notch ligand mDLL1 by POFUT1 is dispensable for ligand function. *PLoS One* **2014**, *9* (2), No. e88571.
- (33) Okajima, T.; Irvine, K. D. Regulation of notch signaling by o-linked fucose. *Cell* **2002**, *111* (6), 893–904.
- (34) Shi, S.; Stanley, P. Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (9), 5234–5239.
- (35) Ishio, A.; Sasamura, T.; Ayukawa, T.; Kuroda, J.; Ishikawa, H. O.; Aoyama, N.; Matsumoto, K.; Gushiken, T.; Okajima, T.; Yamakawa, T.; et al. O-fucose monosaccharide of *Drosophila* Notch has a temperature-sensitive function and cooperates with O-glucose glycan in Notch transport and Notch signaling activation. *J. Biol. Chem.* **2015**, *290* (1), 505–519.
- (36) Kim, H.; Jang, J. H.; Kim, S. C.; Cho, J. H. De novo generation of short antimicrobial peptides with enhanced stability and cell specificity. *J. Antimicrob. Chemother.* **2014**, *69* (1), 121–132.
- (37) Bernard, J. M.; Kara, P.; Lisa Cencia, R.; Phillip W, G. Degradation of naturally occurring and engineered antimicrobial peptides by proteases. *Adv. Biosci. Biotechnol.* **2011**, *2011*, 404–408.
- (38) Cologna, C. T.; Rodrigues, R. S.; Santos, J.; de Pauw, E.; Arantes, E. C.; Quinton, L. Peptidomic investigation of *Neoponera villosa* venom by high-resolution mass spectrometry: seasonal and nesting habitat variations. *J. Venomous Anim. Toxins Incl. Trop. Dis.* **2018**, *24* (1), 6.
- (39) Lamiable, A.; Thévenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tufféry, P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* **2016**, *44* (W1), W449–W454.
- (40) Honma, T.; Kawahata, S.; Ishida, M.; Nagai, H.; Nagashima, Y.; Shiomi, K. Novel peptide toxins from the sea anemone *Stichodactyla haddoni*. *Peptides* **2008**, *29* (4), 536–544.
- (41) Kozlov, S.; Grishin, E. The mining of toxin-like polypeptides from EST database by single residue distribution analysis. *BMC Genomics* **2011**, *12* (1), 88.
- (42) Shiomi, K.; Honma, T.; Ide, M.; Nagashima, Y.; Ishida, M.; Chino, M. An epidermal growth factor-like toxin and two sodium channel toxins from the sea anemone *Stichodactyla gigantea*. *Toxicon* **2003**, *41* (2), 229–236.
- (43) Sato, C.; Zhao, G.; Ilagan, X. G. An overview of notch signaling in adult tissue renewal and maintenance. *Curr. Alzheimer Res.* **2012**, *9* (2), 227–240.

- (44) Schneider, M.; Al-Shareffi, E.; Haltiwanger, R. S. Biological functions of fucose in mammals. *Glycobiology* **2017**, *27* (7), 601–618.
- (45) Buchon, N.; Broderick, N. A.; Kuraishi, T.; Lemaitre, B. *Drosophila* EGFR pathway coordinates stem cell proliferation and gut remodeling following infection. *BMC Biol.* **2010**, *8* (1), 152.
- (46) Jiang, H.; Grenley, M. O.; Bravo, M.-J.; Blumhagen, R. Z.; Edgar, B. A. EGFR/Ras/MAPK signaling mediates adult midgut epithelial homeostasis and regeneration in *Drosophila*. *Cell stem cell* **2011**, *8* (1), 84–95.
- (47) Trummal, K.; Tõnismägi, K.; Paalme, V.; Järvekülg, L.; Siigur, J.; Siigur, E. Molecular diversity of snake venom nerve growth factors. *Toxicon* **2011**, *58* (4), 363–368.
- (48) Guo, L.-y.; Zhu, H.; Zhu, J.-f.; Jing, N.-h.; Feng, L.; Zhou, Y.-c. Identification of a serine protease with nerve growth promoting activity from snake venom. *NeuroReport* **1998**, *9* (16), 3577–3581.
- (49) Paalme, V.; Trummal, K.; Samel, M.; Tõnismägi, K.; Järvekülg, L.; Vija, H.; Subbi, J.; Siigur, J.; Siigur, E. Nerve growth factor from *Vipera lebetina* venom. *Toxicon* **2009**, *54* (3), 329–336.
- (50) Wijeyewickrema, L. C.; Gardiner, E. E.; Gladigau, E. L.; Berndt, M. C.; Andrews, R. K. Nerve growth factor inhibits metalloproteinase-disintegrins and blocks ectodomain shedding of platelet glycoprotein VI. *J. Biol. Chem.* **2010**, *285* (16), 11793–11799.

1639 **Article 3**

1640 In preparation for *Toxins*

1641

1642 **Myrmicine ant venoms: first insights into their peptidic diversity**

1643

1644 *Valentine Barassé^{1*}, Nathan Téné¹, Christophe Klopp², Niklas Tysklind⁴, Hadrien Lalägue⁴,*
1645 *Jérôme Orivel⁴, Valérie Troispoux⁴, Frédérick Petitclerc⁴, Martin Kenne³, Maurice Tindo³,*
1646 *Michel Treilhou¹, Elsa Bonnafé^{1#}, Axel Touchard^{1,4#}*

1647

1648 ¹ EA-7417, Institut National Universitaire Champollion, Place de Verdun, 81012 Albi,
1649 France ; nathan.tene@univ-jfc.fr (N.T.) ; elsa.bonnafé@univ-jfc.fr (E.B.) ;
1650 michel.treilhou@univ-jfc.fr (M.T.)

1651 ² Unité de Mathématique et Informatique Appliquées de Toulouse, UR0875, INRA
1652 Toulouse, Castanet-Tolosan, France ; christophe.klopp@inra.fr (C.K.)

1653 ³ Laboratory of Animal Biology and Physiology, Faculty of Science, University of Douala,
1654 Cameroon, P.O.Box. 24157 Douala, Cameroon; tindodouala@yahoo.com (M.T.);
1655 medoum68@yahoo.fr (M.K.)

1656 ⁴ CNRS, UMR EcoFoG, AgroParisTech, Cirad, INRA, Université des Antilles, Université
1657 de Guyane, 97310 Kourou, France; axel.touchard2@gmail.com (A.T.) ;
1658 niklas.tysklind@ecofog.gf (N.T.) ; hadrien.lalague@ecofog.gf (H.L.) ;
1659 jerome.orivel@ecofog.gf (J.O.) ; valerie.troispoux@ecofog.gf (V.T.) ;
1660 frederick.petitclerc@ecofog.gf (F.P.)

1661 [#] These authors contributed equally to this study.

1662

1663 Corresponding author:

1664 * EA-7417, Institut National Universitaire Champollion, Place de Verdun, 81012 Albi,
1665 France. Phone: +(33)5 63 48 17 00. Email: valentine.barasse@gmail.com

1666 **Abstract (200/200 words):**

1667 Among the stinging ants, *Myrmicinae* represents the most speciose phylogenetic
1668 subfamily. Venom composition already described for these venomous organisms is extremely
1669 variable with alkaloids predominant in some tribes. Nevertheless, proteomics studies revealed
1670 that some myrmicine ant venoms are peptide-rich. Using integrated transcriptomic and
1671 proteomic approaches, we characterized the venom peptidomes of six ants belonging to the
1672 different phylogenetic tribes of *Myrmicinae*. We identified a total of 78 myrmicotoxins
1673 precursors which can be classified into 37 different families according to their mature
1674 sequences. Most of those peptidic toxins were linear, amphiphilic and polycationic. However,
1675 the presence of peptides containing one to three disulfide bonds was also noted. Furthermore,
1676 the venom of *Daceton armigerum* consisted in a complex of dimeric peptides. Myrmicine ant
1677 venoms showed heterogeneous peptidic compositions, with some exhibiting more diversity
1678 than others. We also observed several analogous toxins among the venom peptidomes of this
1679 study. Moreover, an analysis on signal sequences revealed that myrmicotoxin precursors are
1680 divided up to eight superfamilies, some of which have already been described in others aculeate
1681 hymenoptera. Thus, our results suggest that mature venom peptides from myrmicine ant might
1682 have derived from a reduced number of precursors, although genomic studies are needed for
1683 confirmation.

1684

1685 **Keywords:** toxin precursors; *Myrmicinae*; molecular diversity; dimeric peptides

1686

1687 **Key Contribution:** This study provides novel insights into both the peptidic diversity and the
1688 evolution of myrmicine ant venoms.

1689

1690 **1. Introduction**

1691 Venoms are fine-tuned biochemical arsenals mainly used by animals to defend
1692 themselves and/or capture preys (Casewell et al., 2013). Most investigations conducted on
1693 arthropod venoms (e.g. scorpions, spiders, centipedes, and insects) revealed mixtures of toxins
1694 with peptides as the dominant components with variable amino acid sequences and
1695 tridimensional structures (Daly & Wilson, 2018; Walker, Robinson, et al., 2018). Although ants
1696 are dominant terrestrial venomous animals in terms of species richness and biomass, their
1697 venom peptides have been less studied in comparison to others venomous arthropods. The small
1698 size of these insects and consequently the difficulty in gathering large amounts of venom partly
1699 explains that ant venoms were overlooked. Recently, proteotranscriptomic investigations of ant

1700 venoms provide several comprehensive peptidomes in the subfamilies *Pseudomyrmecinae*,
1701 *Paraponerinae*, *Myrmicinae*, *Ponerinae* and *Myrmeciinae* (Aili et al., 2020; Barassé et al.,
1702 2019; Kazuma et al., 2017; Robinson et al., 2018; Touchard, Aili, et al., 2020; Touchard et al.,
1703 2018) and start to reveal the molecular diversity of these toxins. The comprehensive inclusion
1704 of ant species from different subfamilies and with different ecologies should enhance our
1705 understanding of the molecular diversification of ant venom peptidomes and then lead to further
1706 discoveries. Within Formicidae, the subfamily of *Myrmicinae* is the most speciose phylogenetic
1707 ant subfamily with actually 48% of described species (Antweb, consulted on 16/06/20). The
1708 venom composition already described for these ants is extremely variable with alkaloids
1709 predominant in some tribes (Morgan, 2008; Touchard, Aili, et al., 2016). Nevertheless,
1710 proteomic analyses have revealed that the venoms of four genera (i.e., *Tetramorium*,
1711 *Pogonomyrmex*, *Myrmica*, and *Manica*) are peptide-rich (Heep, Klaus, et al., 2019; Heep,
1712 Skaljic, et al., 2019; J. O. Schmidt & Blum, 1978b; von Sicard, Candy, & Anderson, 1989).
1713 Further investigations into the venom composition of other myrmicine species is therefore
1714 necessary to understand the intrinsic chemical properties of these venoms and gather
1715 informations on the relationships between ant phylogeny and venom composition. *Myrmicinae*
1716 are indeed a hyperdiverse clade occupying most of the terrestrial habitats and exhibiting wide
1717 variation in ecology behavior, colony structure and diet (Blaimer, Ward, Schultz, Fisher, &
1718 Brady, 2018; Ward et al., 2015).

1719 This study aimed to draw first insights into the peptidic diversity of myrmicine ant
1720 venoms. We thus combined transcriptomics and proteomics to decipher venoms from ants
1721 belonging to the six tribes composing this phylogenetic subfamily: *Myrmica ruginodis*
1722 (*Myrmicini*), *Pogonomyrmex californicus* (*Pogonomyrmecini*), *Stenamma debile*
1723 (*Stenammini*), *Solenopsis saevissima* (*Solenopsidini*), *Daceton armigerum* (*Attini*) and
1724 *Tetramorium africanum* (*Crematogastrini*). We also compared the venom composition of these
1725 ants with previously described ant venoms.

1726

1727 **2. Materials and Methods**

1728 *2.1. Collection and preparation of venom samples*

1729 Ant workers were collected from different locations in France, French Guiana, Cameroon,
1730 and USA. Ant venom reservoirs were dissected and pooled in a solution containing 10%
1731 acetonitrile (ACN)/ ultrapure water (v/v) (Supplementary Table I), and the membranes were
1732 disrupted using ultrasonic waves for 2 min. The samples were then centrifuged for 5 min at

1733 14,400 rpm, and the supernatant was collected and dried using a speed vacuum prior to storage
1734 at $-20\text{ }^{\circ}\text{C}$ until proteomic analysis.

1735

1736 *2.2. Mass spectrometry analysis*

1737 A preliminary LC–MS analysis of crude venoms was carried out on the LCQ-Ion trap
1738 Advantage equipped with an ESI-LC system Accela (ThermoFisher Scientific, Courtabœuf,
1739 France). Peptides were separated using an Acclaim RSLC C_{18} column ($2.2\text{ }\mu\text{m}$; $2.1 \times 150\text{ mm}$;
1740 ThermoFisher, France). The mobile phase was a gradient prepared from 0.1% formic acid
1741 (FA)/water (v/v) (solvent A) and 0.1% FA/ACN (v/v) (solvent B). The peptides were eluted
1742 using a linear gradient from 0 to 50% of solvent B over 45 min, then from 50 to 100% over 10
1743 min, and finally held for 5 min at a $250\text{ }\mu\text{L}\cdot\text{min}^{-1}$ flow rate. The electrospray ionization mass
1744 spectrometry detection was performed in positive mode with the following optimized
1745 parameters: the capillary temperature was set at $300\text{ }^{\circ}\text{C}$, the spray voltage was 4.5 kV, and the
1746 sheath gas and auxiliary gas were set at 50 and 10 psi, respectively. The acquisition range was
1747 from 100 to 2000 m/z . The area value of each peak corresponding to a peptide was manually
1748 integrated using the peak ion extraction function in Xcalibur software (version 4.0,
1749 ThermoFisher Scientific, Courtabœuf, France). The relative peak area indicates the contribution
1750 of each peptide to all the peptides identified in the venom, providing a measure of relative
1751 abundance.

1752 Reduction of disulfide bonds was achieved by mixing $30\mu\text{L}$ of crude venom with $30\mu\text{L}$ of
1753 $100\text{ mol}\cdot\text{L}^{-1}$ ammonium bicarbonate buffer (pH 8) containing $10\text{ mmol}\cdot\text{L}^{-1}$ dithiothreitol (DTT)
1754 followed by an incubation for 30 min at $56\text{ }^{\circ}\text{C}$. Then, the reduced venom was alkylated by
1755 adding $10\text{ }\mu\text{L}$ of $50\text{ mol}\cdot\text{L}^{-1}$ iodoacetamide (IA) for 15 min at room temperature in the dark. As
1756 chemical reduction/alkylation results in a mass increase of 57 Da for each cysteine, the
1757 examination of mass shifts in the mass spectra of reduced/alkylated samples permitted us to
1758 determine the presence and the number of disulfide bonds in the corresponding peptides.

1759 Hexanic extraction of venoms was conducted by mixing an aqueous solution of crude
1760 venom ($500\text{ }\mu\text{L}$) with $500\mu\text{L}$ of hexane. After phase separation, the hexane phase was discarded
1761 and the operation was repeated twice. The aqueous phase was analyzed through LC-MS
1762 according to the protocol described earlier.

1763

1764 *2.3. De novo orbitrap mass spectrometry-based sequencing*

1765 Each crude venom was re-suspended in water and then desalted using ZipTip[®] C_{18}
1766 (Merck Millipore, Burlington, VT, USA) after adding trifluoroacetic acid (TFA) at a final

1767 concentration of 0.5%. Then, the venom sample was subjected to *de novo* sequencing using a
1768 Q-Exactive Plus mass spectrometer coupled to a Nano-LC Proxeon 1000 (ThermoFisher
1769 Scientific, Waltham, MA, USA). Peptides were separated through chromatography with the
1770 following parameters: Acclaim PepMap100 C₁₈ pre-column (2 cm, 75 μm i.d., 3 μm, 100 Å),
1771 Pepmap-RSLC Proxeon C₁₈ column (50 cm, 75 μm i.d., 2 μm, 100 Å), 300 nL min⁻¹ flow rate,
1772 a 98 min gradient from 95% solvent A (water, 0.1% FA) to 35% solvent B (99.9% ACN, 0.1%
1773 FA) for a total time of 2 h. Peptides were analyzed in the Orbitrap cell, at a resolution of
1774 120,000, with a mass range of *m/z* 350–1550. Fragments were obtained through high collision-
1775 induced dissociation (HCD) activation with a collisional energy of 27%. Data were acquired in
1776 the Orbitrap cell in a Top20 mode, at a resolution of 17,500. For the identification step, all MS
1777 and MS/MS data were processed with an in-house Peaks software (BSI, version 6.0) to perform
1778 *de novo* sequencing. The mass tolerance was set to 10 ppm for precursor ions and 0.02 Da for
1779 fragments. The following modifications were allowed: oxidation (Met) and pyroglutamic acid
1780 (Glu). *De novo* peptide sequences with Average Local Confidence (ALC) higher than 60% were
1781 used for the peptide identifications.

1782

1783 2.4. *Venom gland transcriptomics*

1784 Venom apparatus (venom glands and reservoir) of ant workers were dissected in a
1785 Phosphate Buffered Saline solution (PBS) and immediately placed in 500 μL of TRIzol reagent
1786 (Invitrogen, Carlsbad, CA, USA) and store at -80°C prior to RNAs extraction. Given the small
1787 size of *Stenamma debile* workers combined to the difficulty to collect workers in large numbers,
1788 the transcriptome database for this species was performed with the whole ant tissues
1789 (Supplementary Table II). Total RNAs were extracted with the RNeasy Micro Kit (Qiagen)
1790 following the manufacturer's instructions. Contaminating genomic DNA was removed using a
1791 DNA-free kit (Applied Biosystem) according to the manufacturer's instructions.

1792 RNA quantity and quality were assessed using a nanodrop and a bioanalyzer (Nanodrop
1793 2000, ThermoFisher Scientific; Agilent 2100 Bioanalyzer System). RNAseq was performed at
1794 the GeT-PlaGe core facility, INRA Toulouse, France. RNA-seq libraries were prepared
1795 according to Illumina's protocols using the Illumina TruSeq Stranded mRNA sample prep kit
1796 to analyze mRNA. Briefly, mRNA was selected using poly-T beads. Then, the RNA was
1797 fragmented to generate double stranded cDNA and adaptors were ligated to be sequenced.
1798 Eleven cycles of PCR were applied to amplify the libraries. Library quality was assessed using
1799 a Fragment Analyser, and the libraries were quantified through qPCR using the Kapa Library

1800 Quantification Kit. RNA-seq experiments were performed on an Illumina HiSeq3000 using a
1801 paired-end read length of 2×150 pb with the Illumina HiSeq3000 sequencing kits.

1802

1803 *2.5. Bioinformatic tools*

1804 *2.5.1. Contig quantification*

1805 The read pairs were assembled twice with drap (version 1.9.1) (Cabau et al., 2017) using
1806 the de Bruijn graph assemblers called Oases and Trinity (parameters: `-dbg oases/trinity`). The
1807 assembly metrics were produced with the `assemblathon_stats.pl` scripts. Raw reads were
1808 aligned on the contigs with `bwa mem` (version 0.7.12-r1039) (Heng Li & Durbin, 2010) using
1809 the default parameters and the alignment files were sorted, compressed, and indexed with
1810 `samtools view, sort, and index` (version: 1.3.1) using the default parameters (H. Li et al., 2009).
1811 The quantification files were generated with `samtools idxstats` (version: 1.3.1), giving us the
1812 length of each contig in base pairs along with the number of hits, corresponding to the number
1813 of sequences from RNAseq reads which aligned on a given contig. To calculate the expression
1814 rate of each contig, we calculated the transcripts per million value (TPM) by dividing the
1815 number of aligned reads for each contig by the contig length, then dividing this value by the
1816 ratio of counts to contig length for all contig. This value was then multiplied by 1 million to
1817 generate TPM.

1818

1819 *2.5.2. Precursors identifications and mature sequences*

1820 RNAseq data were translated using a translate program command lines (`embooss`
1821 package, command line: `transeq`) in order to obtain the potential Open Reading Frames (ORFs).
1822 Then, the fragments of sequences obtained during the *de novo* Orbitrap mass spectrometry-
1823 based sequencing were aligned against these data by using the command-line NCBI BLAST
1824 program (`ncbi-blast-2.6.0+` package, command line: `blastp`, parameter: `-matrix PAM30`) with
1825 adapted parameters for short sequences, allowing us to find the complete peptide sequences and
1826 the name of the contigs on which they aligned. The masses of mature peptide sequences,
1827 obtained from these different approaches, were systematically verified using the peptide mass
1828 program from ExPASy portal (<https://expasy.org>) and compared to those obtained through
1829 mass spectrometry. Signal sequences and transmembrane domains were predicted with the
1830 `phobius` program available at <http://phobius.sbc.su.se/>. Sequence similarities were searched for
1831 using the NCBI BLAST program presented in the Uniprot server with the default parameters.
1832 Alignments were achieved using the `Muscle` program in `Seaview` version 4.6.1 (Gouy,
1833 Guindon, & Gascuel, 2010), and edited using `BOXSHADE` version 3.2 ([110](https://embnet.vital-</p></div><div data-bbox=)

1834 it.ch/software/BOX_form.html).. Sequence identity and similarity percentages were calculated
1835 with the software infoalign from the EMBOSS suite of bioinformatic tools (Rice, Longden, &
1836 Bleasby, 2000). We also performed a Hierarchical Cluster Analysis (HCA) on signal sequences
1837 of myrmecitoxins identified in this study and those previously defined in myrmecine ant venom
1838 peptidomes (Touchard, Aili, et al., 2020; Touchard et al., 2018). Briefly, signal sequences were
1839 predicted with SignalP 5.0 (Almagro Armenteros et al., 2019). Multiple alignments were then
1840 achieved with the ClustalW program and pairwise distances were computed using MEGAX
1841 version 10.1.7 with default parameters (Kumar, Stecher, Li, Knyaz, & Tamura, 2018). HCA
1842 was then performed using Ward's method with the R software (R Core Team, 2017).

1843

1844 *2.5.3. Annotation of most expressed contigs*

1845 Open reading frames (≥ 100 amino-acids length), found by translating RNAseq data,
1846 were extracted from the most abundant contigs and then submitted to the NCBI BLAST
1847 program against the Uniprot refseq protein database on the computational cluster of the
1848 Genotoul bioinformatic facility, INRA Toulouse, France (ncbi-blast-2.6.0 + package, command
1849 line: blastp, parameter: -matrix BLOSUM62).

1850

1851 *2.6. Toxin nomenclature*

1852 Novel myrmecitoxin sequences were named following the standard nomenclature for
1853 animal venom peptides (G. F. King et al., 2008), and we used myrmecitoxin (MYRTX) to define
1854 the venom peptides from the subfamily Myrmecinae and genus/species descriptors as follow;
1855 *Myrmica ruginodis* (Mru), *Pogonomyrmex californicus* (Pc), *Stenamma debile* (Sd), *Daceton*
1856 *armigerum* (Da) and *Tetramorium africanum* (Ta) (Touchard, Aili, et al., 2016). In cases of
1857 sequence similarities, we also named the novel myrmecitoxins according to the same subscripts
1858 used to denote the 'unknown' activity descriptor prefixes previously defined for myrmecine ant
1859 venom peptides (U₁-U₂₀), and forming mature myrmecitoxin families (Touchard, Aili, et al.,
1860 2020; Touchard et al., 2018).

1861

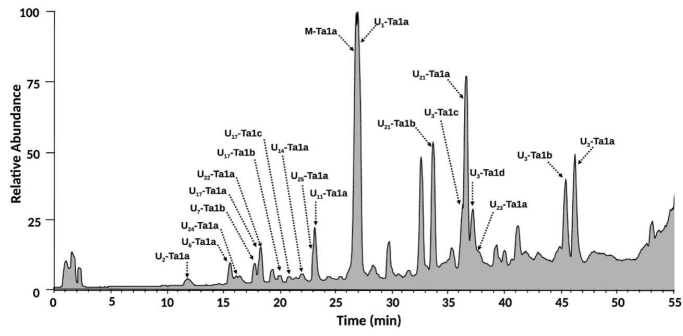
1862 **3. Results**

1863 *3.1. Proteotranscriptomics data*

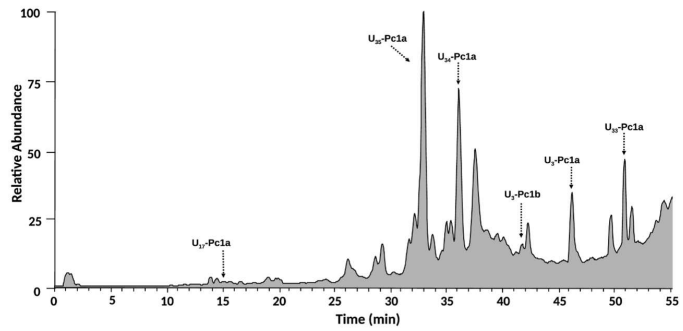
1864 A preliminary LC-MS analysis was performed on an LCQ-ion trap Advantage mass
1865 spectrometer in order to establish the lists of the masses corresponding to peptides in each crude
1866 venom. Before sequence determination, each peptide detected was tentatively named based on
1867 the abbreviations specified in part 2.6, followed by their molecular weight as described by

1868 Johnson et al. (Johnson et al., 2010). The fully sequenced peptides were then renamed in
1869 accordance with the nomenclature previously described in the part 2.6 (Figure 1). Although the
1870 venom peptides structured by disulfide bonds are less abundant within ant venoms than in
1871 marine snail, spider or scorpion venoms, some studies revealed the presence of venom peptides
1872 with one, two and three disulfide bonds (Aili et al., 2014; Pan & Hink, 2000). Consequently, a
1873 chemical reduction/alkylation was performed on every crude venom from this study, excepted
1874 *Tetramorium africanum*, *Stenamma debile* and *Solenopsis saevissima*, followed by a LC-MS
1875 analysis.

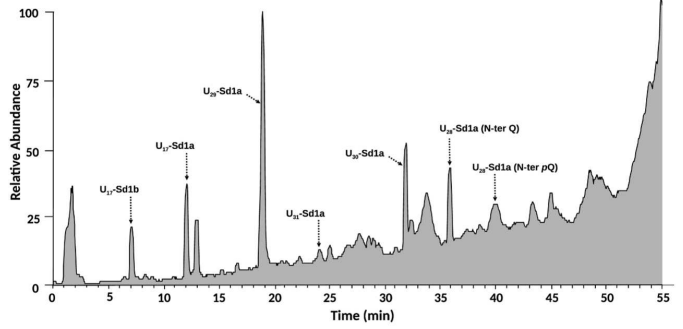
Tetramorium africanum



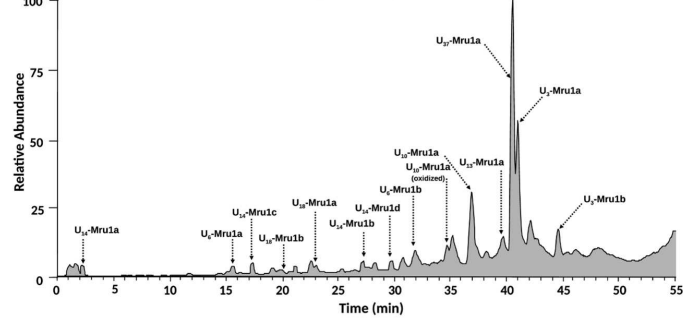
Pogonomyrmex californicus



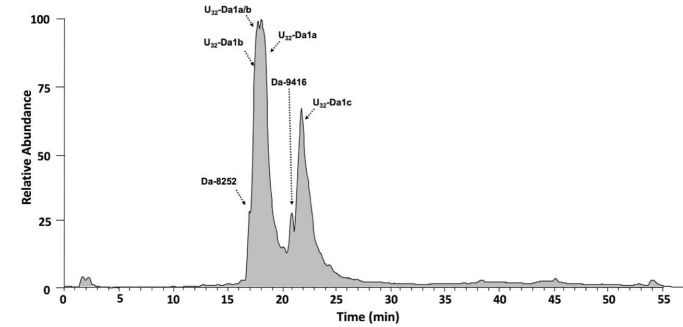
Stenamma debile



Myrmica ruginodis



Daceton armigerum



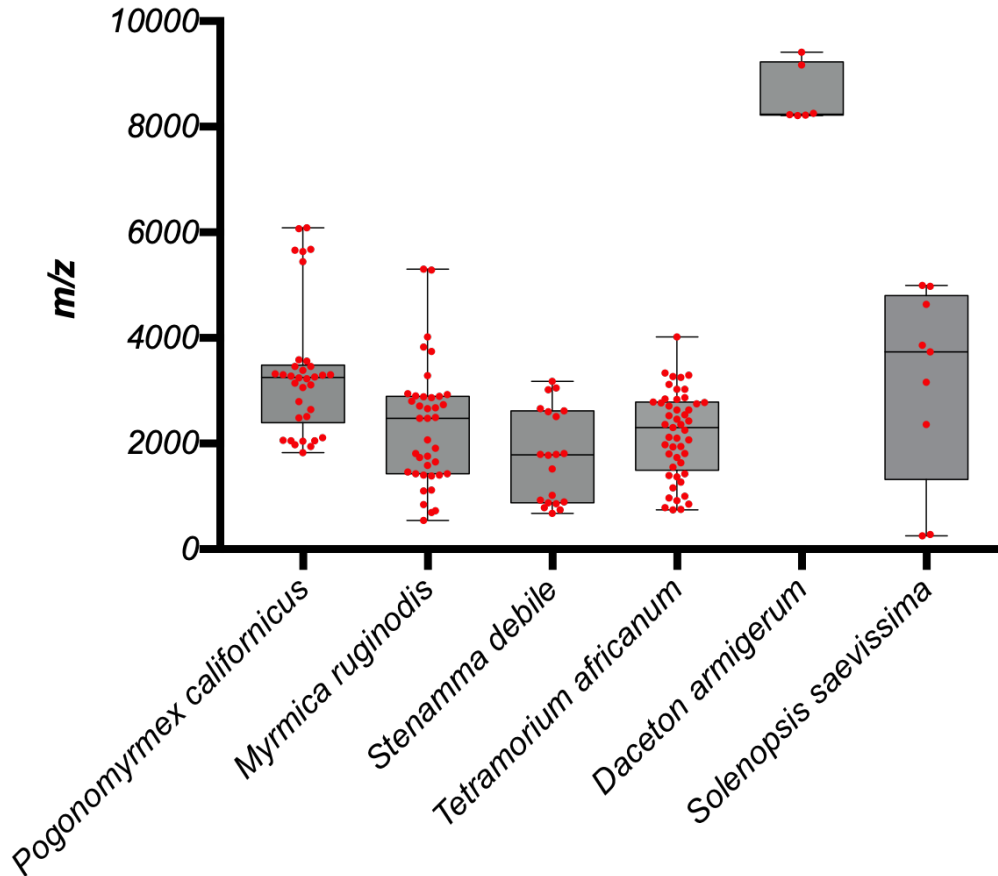
1877 **Figure 1:** Total ion chromatograms (TIC) of myrmicine ant crude venoms. Peptides were eluted through RP-
1878 HPLC on a C18 column using a linear H₂O/ACN gradient at a flow rate of 250 $\mu\text{L}\cdot\text{min}^{-1}$.

1879

1880 A total of 155 masses corresponding to peptides were detected by LC-MS analysis in
1881 myrmicine ant venoms (Figure 2; Supplementary tables III-XIII). *Tetramorium africanum*
1882 venom exhibited the highest number of masses (i.e. 49), which were distributed relatively
1883 evenly between 1.84 and 46.24 min, ranging from 744 and 3,338 Da (Supplementary Table III
1884 – Figure 2). *Stenamma debile* venom exhibited a total of 20 masses eluting mostly between
1885 12.01 and 39.77 min. Most of them were lower than 2,500 Da, although three masses ranging
1886 from 3,000 to 3,200 Da were detected (Supplementary Table V – Figure 2). *Myrmica ruginodis*
1887 and *Pogonomyrmex californicus* venoms showed 39 and 34 masses corresponding to peptides,
1888 eluting from 2.28 to 44.61 min and from 11.77 to 51.62 min, respectively (Supplementary
1889 Tables IX and XI). Although most of the masses were comprised between 2,500 and 3,600 Da,
1890 several masses ranging from 4,000 and 6,086 Da were detected in those venoms (Figure 2).

1891 *Solenopsis sp.* venoms are known to be rich in alkaloids. We thus performed an hexanic
1892 extraction and the peptidic fraction was submitted to LC-MS analysis (Figure S1). Peptides
1893 were detected in *S. saevissima* venom, ranging from 2,361 to 4,979 Da and eluting from 21.91
1894 to 41.98 min. This venom was predominantly composed of alkaloids, with two masses (m/z
1895 253.00-279.22) representing about 90% of the venom (Supplementary Table XIII – Figures 2
1896 and S1). *Daceton armigerum* venom exhibited six masses corresponding to peptides which
1897 were greater than 8,000 Da, eluting from 17.01 to 21.87 min (Supplementary Table VII - Figure
1898 2).

1899



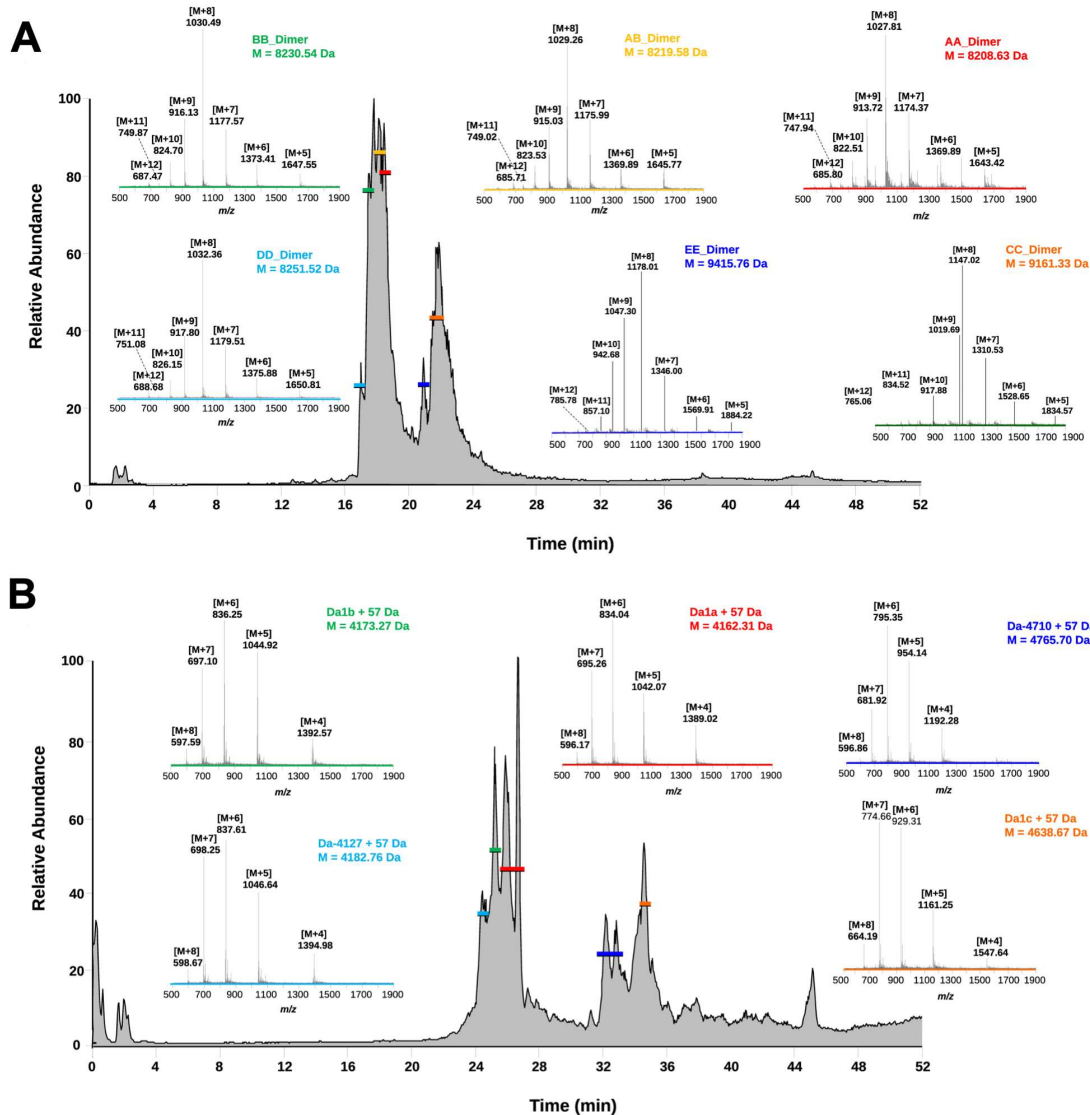
1900
 1901 **Figure 2:** Repartition of masses detected by LC-MS analysis in the six myrmicine ant venoms. The bottom
 1902 and up end of each box represent the minimal and the maximal mass detected, respectively, while the line
 1903 inside each box represents the mean mass. Each red point represents a detected mass.

1904 The LC-MS/MS analysis of the six crude venoms was then performed using a Q-exactive
 1905 orbitrap mass in order to achieve *de novo* sequencing. This resulted in a total of 6,314 sequence
 1906 tags with an ALC score higher or equal to 60%, ranging from 145 to 2,898 sequence tags for *S.*
 1907 *debile* to *T. africanum* venoms, respectively.

1908 To characterize the myrmecitoxins, the mRNA transcriptome of venom glands was
 1909 sequenced using an Illumina Hiseq, resulting in a total of 319,437,422 raw reads, ranging from
 1910 19,389,866 to 82,055,072 for *D. armigerum* and *T. africanum*, respectively. *De novo* assembly
 1911 with Trinity and Oases yielded a total of 181,237 and 102,277 contigs for all venoms, with a
 1912 mean contig size of 1,880 and 2,174 bp, respectively. Both assembly methods gave various
 1913 number of contigs for each venom glands transcriptome, ranging from 13,701 for *D. armigerum*
 1914 (Oases) to 48,353 for *S. debile* (Trinity) (Supplementary table XIV). For each myrmicine ant,
 1915 a combined database from both assemblies was used to search for peptide sequences generated
 1916 from LC-MS/MS analyses. Additionally, the signal sequences of the previously identified
 1917 myrmecitoxins were searched against transcriptomic databases in order to find peptides sharing

1918 the same signal sequences. This permitted us to deduce that most of the venom peptide toxins
1919 have high transcription levels accounting from 16% to 79% of the most expressed transcripts by
1920 the venom glands of *D. armigerum* and *T. africanum*, respectively (Supplementary Tables XV-
1921 XVIII). The mature myrmecitoxins sequences were verified manually using mass spectrometry
1922 data by matching theoretical masses with those measured. This approach permitted us to infer
1923 19 putative sequences and to confirm 59 myrmecitoxin precursors, including four dimeric
1924 peptides.

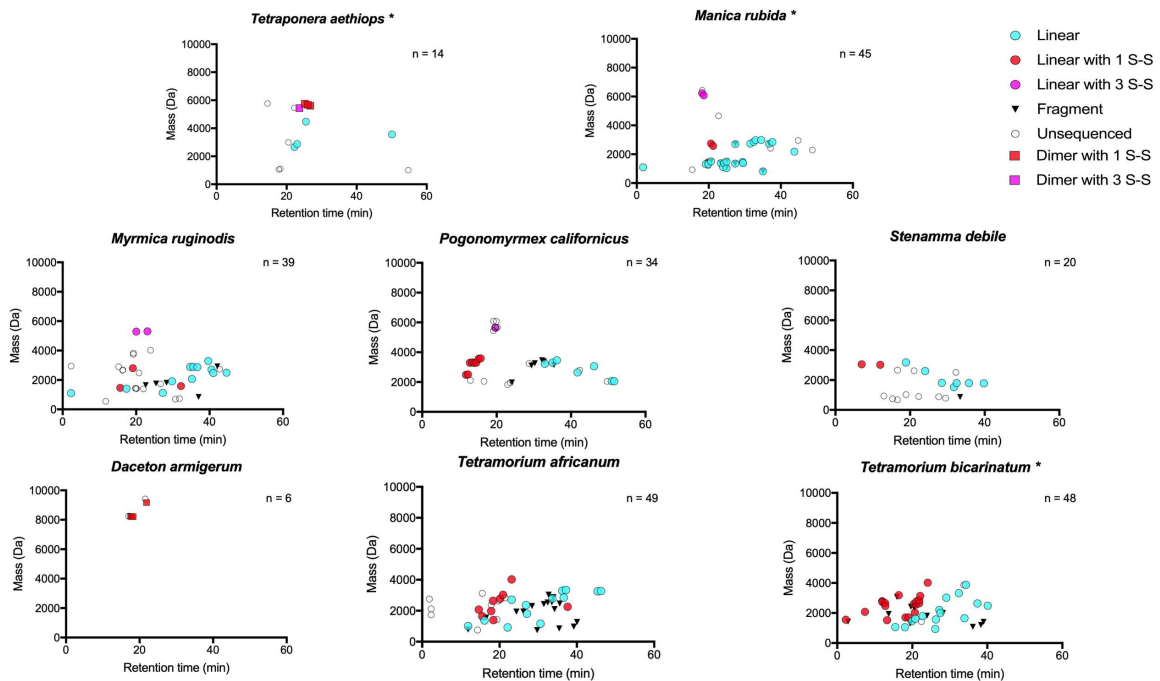
1925 *D. armigerum* venom exhibited indeed a streamlined peptidome with 6 detected masses
1926 which were greater than 8,000 Da. Additionally, the transcriptomic data led to the identification
1927 of three transcripts encoding peptides with calculated masses ranging from 4,104 to 4,581 Da,
1928 and exhibiting only one cysteine (Supplementary table V). Several dimeric peptides have been
1929 observed in ant venoms, particularly for arboreal pseudomyrmecine ants (Barassé et al., 2019;
1930 Touchard et al., 2014; Touchard, Mendel, et al., 2020). Thus, we hypothesized that *D.*
1931 *armigerum* venom is composed of a complex of dimeric peptides, and that those transcripts
1932 encoded dimeric peptides subunits. Based on this hypothesis, we calculated the theoretical
1933 masses of the presumed homo/heterodimeric peptides from the three identified transcripts. Four
1934 calculated masses matched with those detected in the total ion chromatogram of the crude
1935 venom, corresponding to the peptides Da-8231, Da-8220, Da-8209 and Da-9161 (i.e.
1936 dimer_BB, heterodimer_AB, homodimers AA and CC, respectively) (Figure 3 – A). The
1937 reduction/alkylation of the crude venom followed by a LC-MS analysis allowed us to identify
1938 a total of five monomers, including those encoded by the three transcripts (i.e. -Da1a, -Da1b
1939 and -Da1c) (Figure 3 – B).



1940
 1941 **Figure 3:** Identification of *Daceton armigerum* dimeric peptides. (A) Total ion chromatogram and MS
 1942 spectra of hypothesized dimeric peptides from the LC-MS analysis of *D. armigerum* crude venom. (B) Total
 1943 ion chromatogram and MS spectra from the LC-MS analysis of *D. armigerum* reduced/alkylated venom.

1944 *T. africanum*, *S. debile*, *P. californicus* and *M. ruginodis* venoms were dominated by linear
 1945 monomeric peptides. Those four venoms also exhibited disulfide-bonded peptides, with a
 1946 majority containing one disulfide bond. Myrmecitoxins containing three disulfide bonds were
 1947 also noted within *M. ruginodis* and *P. californicus* venoms (Figure 4).

1948



1949

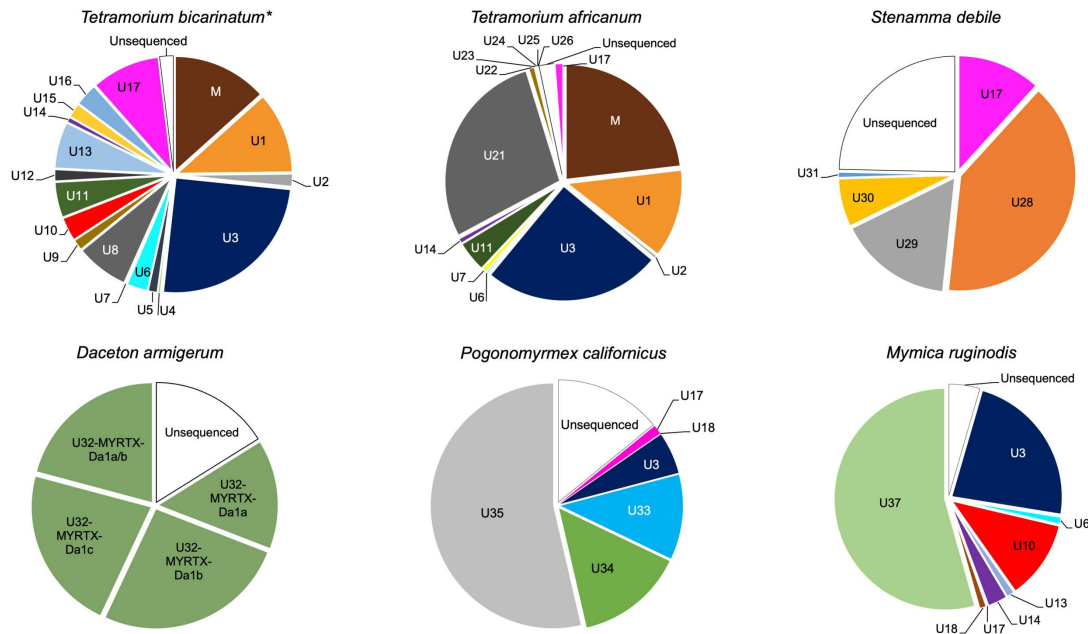
1950 **Figure 4:** Landscapes of myrmicine and pseudomyrmecine ant venoms. “*” indicates previously published
 1951 data, which were obtained with a similar methodology from (Barassé et al., 2019; Touchard, Aili, et al., 2020;
 1952 Touchard et al., 2018).

1953

1954 3.2. Molecular features of mature myrmecitoxins

1955 Here, the myrmicine ant venom peptidomes exhibited heterogeneous compositions,
 1956 with some venoms displaying more mature toxin families than others. For instance, eighteen
 1957 mature myrmecitoxin families were identified in *T. africanum* venom peptidome whereas *D.*
 1958 *armigerum* venom peptides were gathered in a single family (Figure 5). Among the mature
 1959 myrmecitoxin families which were previously defined in *Tetramorium bicarinatum* and *Manica*
 1960 *rubida* venoms (U_1 - U_{20}) (Touchard, Aili, et al., 2020; Touchard et al., 2018), twelve were
 1961 observed in the different venoms of this study, ranging from one to nine for *S. debile* and *T.*
 1962 *africanum*, respectively. Furthermore, we identified a total of seventeen new mature
 1963 myrmecitoxin families (U_{21} - U_{37}) (Figure 5).

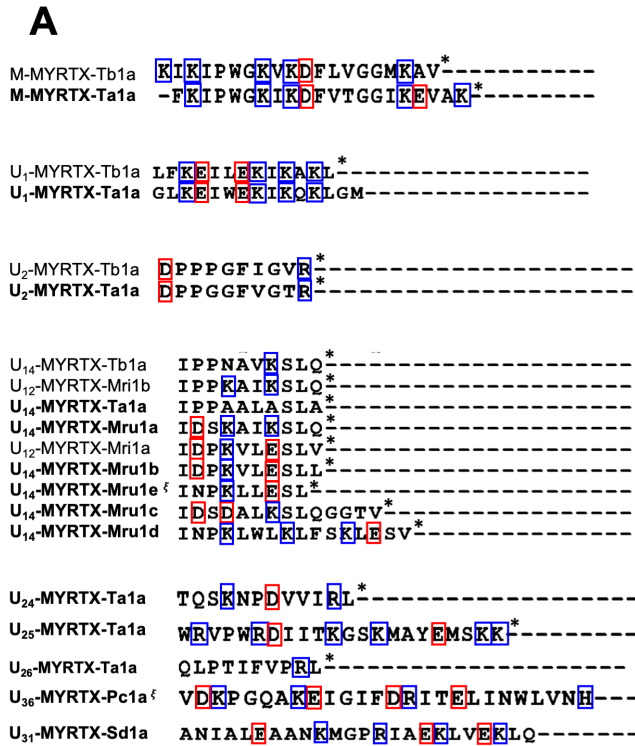
1964 Several venom peptides exhibited post-translational modifications (PTMs) which are
 1965 essential to their activity, such as C-terminal amidation, fucosylation or the formation of
 1966 disulfide bonds. In the first place, we thus distinguished linear and disulfide-bonded peptides
 1967 to expose the diversity contained in myrmicine ant venoms.



1968
 1969 **Figure 5:** Relative abundances of mature myrmecitoxin families found in myrmecine ant venom peptidomes.
 1970 “ * ” denotes previously described venom peptidomes, which were characterized with a similar methodology
 1971 from (Touchard et al., 2018).

1972 *3.2.1. Linear myrmecitoxins*

1973 Several mature myrmecitoxin families, such as M, U₁, U₁₄, U₂₄, U₂₅, U₂₆ and U₃₁, gather
 1974 linear amphiphilic peptides which are polycationic, containing therefore several arginine and
 1975 lysine residues (Figure 6 – A). Those short myrmecitoxins (i.e. 10-21 amino acids long)
 1976 constitute a total of 37% of *T. africanum* venom peptidome, whereas they represent 1 to 3% of
 1977 *S. debile* and *M. ruginodis* venoms, respectively (Figure 5). A linear amphiphilic neutral peptide
 1978 was also identified in *T. africanum* venom (i.e. U₂), although it represents less than 1% of the
 1979 venom peptidome (Figure 6 – A). Two families gather linear amphiphilic peptides from 24 to
 1980 31 amino acids long that exhibit a negative net charge (i.e. U₂₁ and U₂₉), due to the presence of
 1981 aspartic and glutamic acids (Figure 6 – B). Those myrmecitoxins represent 17% and 28% of *S.*
 1982 *debile* and *T. africanum* venom peptidomes, respectively.



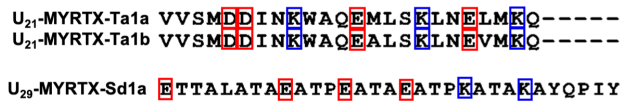
%ID	%S	Species
-	-	<i>Tetramorium bicarinatum</i>
62	76	<i>Tetramorium africanum</i>

%ID	%S	Species
-	-	<i>Tetramorium bicarinatum</i>
60	60	<i>Tetramorium africanum</i>

%ID	%S	Species
-	-	<i>Tetramorium bicarinatum</i>
70	80	<i>Tetramorium africanum</i>

%ID	%S	Species
-	-	<i>Tetramorium bicarinatum</i>
80	90	<i>Manica rubida</i>
60	70	<i>Tetramorium africanum</i>
60	70	<i>Myrmica ruginodis</i>
40	60	<i>Manica rubida</i>
40	60	<i>Myrmica ruginodis</i>
44	67	<i>Myrmica ruginodis</i>
43	57	<i>Myrmica ruginodis</i>
19	19	<i>Myrmica ruginodis</i>

B



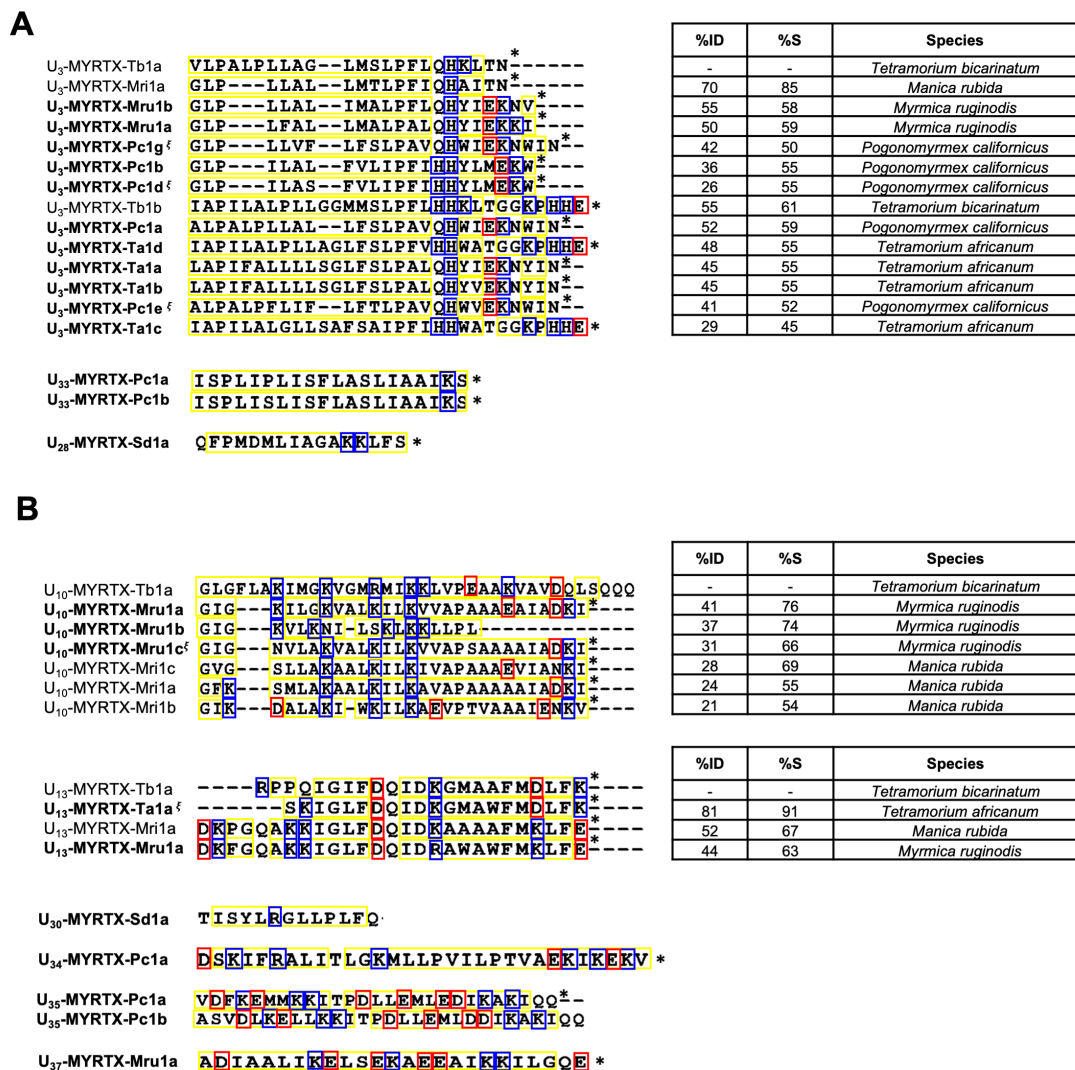
1983

1984 **Figure 6:** Linear cationic, neutral (A) and anionic (B) myrmecitoxins in *Tetramorium africanum* (Ta),
1985 *Pogonomyrmex californicus* (Pc) and *Myrmica ruginodis* (Mru) venoms. Myrmecitoxins identified in this
1986 study were compared to venom peptides from *T. bicarinatum* and *M. rubida* (Touchard, Aili, et al., 2020;
1987 Touchard et al., 2018), and are highlighted in bold. Percentage identity (%ID) and similarity (%S) are relative
1988 to the first peptide of each family. “^ξ” denotes putative peptides. “*” denotes a C-terminal amidation.
1989 Alignments were generated with the Muscle program in Seaview version 4.6.1 and edited using BOXSHADE
1990 version 3.2 (Gouy et al., 2010). Positively charged residues are surrounded in blue whereas negatively
1991 charged residues are surrounded in red.

1992 Most of the venoms of this study also contain polycationic peptides which are
1993 hydrophobic (Figure 7). For instance, the U₃ family is represented in every venom peptidome
1994 of this study, except in *S. debile* and *D. armigerum* venoms. This family is abundant in *T.*
1995 *africanum* and *M. ruginodis* venoms (ca 25%), and represents about 6% of *P. californicus*
1996 venom peptidome (Figure 5). Peptides from the U₃ family are cationic and share common
1997 biochemical features: a hydrophobic leucine-rich N-terminal region and charged residues in
1998 their C-terminal part. More peptides with similar distributions of hydrophobic and charged

1999 amino acids are also observed in *P. californicus* (i.e. U₃₃) and *S. debile* (i.e. U₂₈) venom
 2000 peptidomes (Figure 7 – A), accounting for 11 and 41% of their content, respectively (Figure 5).

2001 Five more families gathering hydrophobic peptides were observed in myrmicine ant
 2002 venoms (i.e. U₁₀, U₁₃, U₃₀, U₃₄ and U₃₇), except in *D. armigerum*. Most of them are cationic,
 2003 due to the presence of several lysine residues which are distributed all along the peptidic chain.
 2004 Indeed, only U₃₅ and U₃₇ peptides, with several glutamic and aspartic acid residues, exhibit a
 2005 negative net charge (Figure 7 – B). Those hydrophobic peptides are relatively abundant in *M.*
 2006 *ruginodis*, *P. californicus* and *S. debile* venoms, as they account for a total of 67%, 68% and
 2007 7% of the venom peptidomes, respectively (Figure 5).

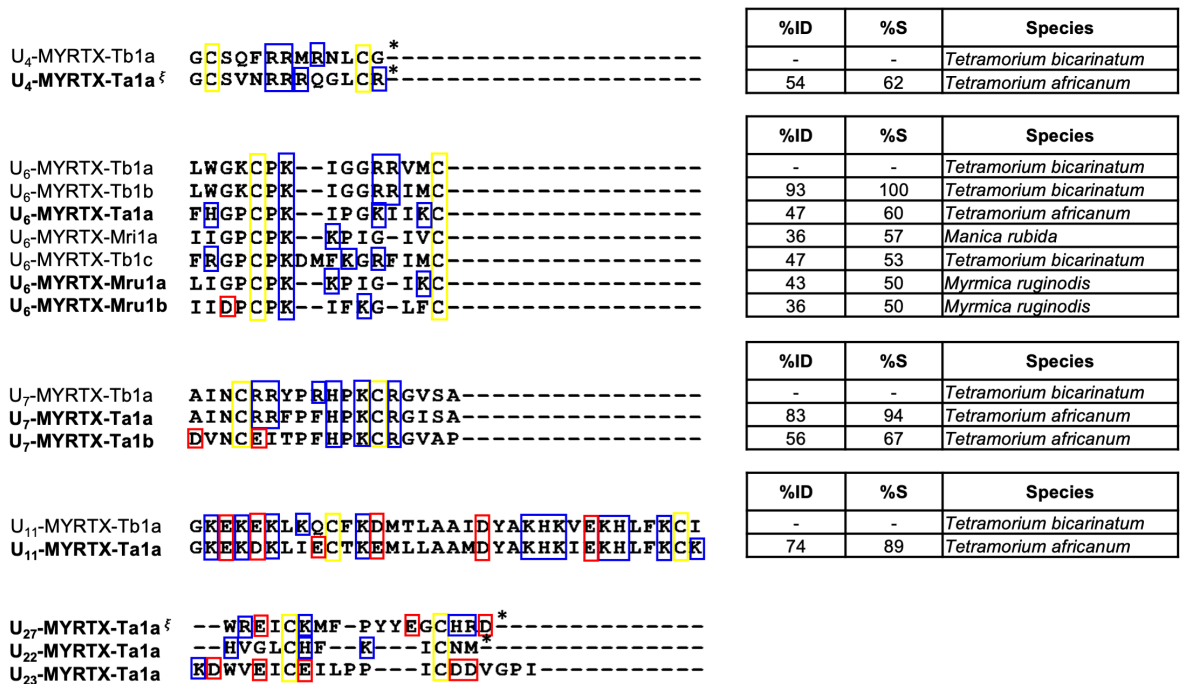


2008 **Figure 7:** Hydrophobic myrmicotoxins and repartition of their charged amino acids. (A) The charged residues
 2009 are mainly located in the N-terminal region. (B) The charged residues are distributed all along the peptidic
 2010 chain. Myrmicotoxins identified in this study were compared to venom peptides from *T. bicarinatum* and *M.*
 2011 *rubida* (Touchard, Aili, et al., 2020; Touchard et al., 2018), and are highlighted in bold. Percentage identity
 2012 (%ID) and similarity (%S) are relative to the first peptide of each family. “^ξ” denotes putative peptides. “*”
 2013

2014 denotes a C-terminal amidation. Alignments were generated with the Muscle program in Seaview version
 2015 4.6.1 and edited using BOXSHADE version 3.2 (Gouy et al., 2010). Positively charged residues are
 2016 surrounded in blue whereas negatively charged residues are surrounded in red. Hydrophobic residues are
 2017 surrounded in yellow.

2018 3.2.2. Disulfide-bonded peptides

2019 Seven families from *T. africanum* and *M. ruginodis* venoms gather peptides from 13 to
 2020 34 amino acids long, containing one intrachain disulfide bond due to the presence of a pair of
 2021 cysteines (i.e. U₄, U₆, U₇, U₁₁, U₂₂, U₂₃ and U₂₇). The most part also contain several arginine,
 2022 lysine and histidine residues, and are therefore polycationic. Indeed, only U₂₃-MYRTX-Ta1a
 2023 exhibits a negative net charge, due to several aspartic acid residues (Figure 8). The seven
 2024 families of disulfide-bonded peptides are represented in *T. africanum* venom, each accounting
 2025 for about 1% of the venom peptidome. However, only U₆ peptides were identified in *M.*
 2026 *ruginodis* venom, with a similar relative abundance (i.e. about 1%).



2027
 2028 **Figure 8:** Myrmecitoxins containing one disulfide bond. The venom peptides identified in this study were
 2029 compared to myrmecitoxins from *T. bicarinatum* and *M. rubida* (Touchard, Aili, et al., 2020; Touchard et al.,
 2030 2018), and are highlighted in bold. Percentage identity (%ID) and similarity (%S) are relative to the first
 2031 peptide of each family. “ ξ ” denotes putative peptides. “*” denotes C-terminal amidation. Alignments were
 2032 generated with the Muscle program in Seaview version 4.6.1 and edited using BOXSHADE version 3.2
 2033 (Gouy et al., 2010). Cysteines are surrounded in yellow. Positively charged residues are surrounded in blue
 2034 whereas negatively charged residues are surrounded in red.

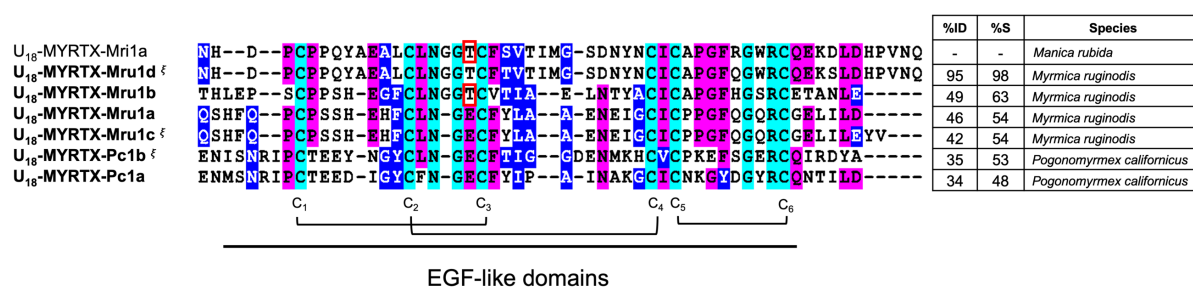
2035 The U₁₇ peptides constitute another family of peptides containing one disulfide bond.
 2036 They are cationic, hydrophilic and from 23 to 32 amino acids long (Figure 9). Sharing similar
 2037 mature sequences with scapins from hymenoptera venoms (Touchard et al., 2018), the
 2038 consensus sequence of U₁₇ peptides from myrmicine ant venoms is defined as
 2039 XIIXXPXXXXXNTCNPXGXXXXXXGXCRXXXX. The U₁₇ family is observed in every
 2040 venom of this study, except for *D. armigerum*, with relative abundances around 1% for *T.*
 2041 *africanum*, *P. californicus* and *M. ruginodis* venom peptidomes. On the other hand, it represents
 2042 12% of *S. debile* venom (Figure 5). Several U₁₇ peptides begin with the pattern ‘HII-’ and
 2043 couldn’t be validated in proteomics, in particular those inferred from *M. ruginodis* venom gland
 2044 transcriptome. Indeed, the calculated masses systematically exhibit a missing 26 Da mass unit
 2045 in comparison with the MS measurements. The same missing mass unit is observed for U₁₇
 2046 peptides from *P. californicus* venom, also beginning with a histidine. Although their calculated
 2047 masses were validated by proteomics data, each of them is also followed by a mass
 2048 corresponding to an addition of 26 Da in a close retention time (Supplementary table VII),
 2049 suggesting a PTM resulting in the gain of a 26 Da.

U₁₇-MYRTX-Tb1a **TIINAPNR**-----**CPPGH-VVV--KGR**CRIA*-----
 U₁₇-MYRTX-Ta1b **YIIRVPIQ**-----**CPPGK-VKV--GNRCRIVY**-----
 U₁₇-MYRTX-Tb1c **TIIDVPIQ**-----**CPPGK-VKV--GNRCRVIF**-----
 U₁₇-MYRTX-Mru1c^ξ **HIINVPIQ**-----**CPPGK-VRV--GNRCRDVGRV**-----
 U₁₇-MYRTX-Ta1a **GIINAPDR**-----**CPDGY-KRA--GNQCRKVI***-----
 U₁₇-MYRTX-Ta1d **SVINVPIQ**-----**CPPGT-IQV--GKRCRETF**-----
 U₁₇-MYRTX-Mru1b^ξ **HIIVFPEN**-----**CPPGQKP--RGNCROEY**-----
 U₁₇-MYRTX-Mru1f^ξ **HIIVPERPPK**-----**CPPGQDR--RGRCRMVP**-----
 U₁₇-MYRTX-Mru1a **AIIDAPRN**-----**CPPGHDIDH--AGDCVEIFE**-----
 U₁₇-MYRTX-Mru1d^ξ **HIIVPDRN**-----**CPPGQRKDN--HGNCRDIA***-----
 U₁₇-MYRTX-Tb1b **TVIDVPIQ**-----**CPSGT-VKV--GNKCRVIF**-----
 U₁₇-MYRTX-Mri1a **HIIVAP**-----**CREGY-VMV--GNYCVEEY**-----
 U₁₇-MYRTX-Mri1c **YIIVAP**-----**CREGY-VMV--GNYCVEEY**-----
 U₁₇-MYRTX-Tb1e **NIKAPLFP**-----**CPNGYIRDY--KGDCREIEE**-----
 U₁₇-MYRTX-Sd1b **YIIDAPSRK**-----**CPEGSRST--QGE CRTTSR***-----
 U₁₇-MYRTX-Mru1g^ξ **HIIVPELQ**-----**CPPGQKRDR--QGRCRKVFN**-----
 U₁₇-MYRTX-Mru1h^ξ **HIIVLPNRTTNT** **CPPGQKDN--YGNCRKIA**-----
 U₁₇-MYRTX-Tb1h **HIIRVP**-----**CRAGY-KEI--RGRCRKILT***-----
 U₁₇-MYRTX-Tb1d **HIIRVP**-----**CRAGY-IEV--NGVCREVFT***-----
 U₁₇-MYRTX-Mru1e^ξ **HIIVIDKN**-----**CPPGQREDH--HGHCRI**V-----
 U₁₇-MYRTX-Pc1a **HIIQVP**-----**CLPGY-VKVGKDGVCREAFKFKPGQRP***
 U₁₇-MYRTX-Sd1a **NIITVPERP**-----**CPSGQRKDS--RGKCRQVLS**-----
 U₁₇-MYRTX-Pc1c **HIIQVP**-----**CRDGY-VKD--NGACIPEY**-----
 U₁₇-MYRTX-Pc1b **HIIQVP**-----**CRDGY-VMV--NGVCREVFNEKDEE**-----
 U₁₇-MYRTX-Ta1e^ξ **NIKVP**-----**CRAGY-RKI--NGVCRKIYR***-----
 U₁₇-MYRTX-Ta1c **GIIRIPELK**-----**CPDGYKDA--LGVCREIFT***-----
 U₁₇-MYRTX-Tb1f **NIIRVPEFQ**-----**CPNGYRKDA--NGKCREVFH***-----
 U₁₇-MYRTX-Mri1b **QIVWVP**-----**CNPRS-KKTDDAGICRNTY**-----
 U₁₇-MYRTX-Tb1g **HVIDTRSRL**-----**CPEGSRST--TGECKTV**-----

%ID	%S	Species
-	-	<i>Tetramorium bicarinatum</i>
54	58	<i>Tetramorium africanum</i>
54	67	<i>Tetramorium bicarinatum</i>
50	54	<i>Tetramorium africanum</i>
46	58	<i>Tetramorium africanum</i>
46	63	<i>Myrmica ruginodis</i>
44	48	<i>Myrmica ruginodis</i>
43	50	<i>Myrmica ruginodis</i>
42	50	<i>Myrmica ruginodis</i>
42	46	<i>Tetramorium bicarinatum</i>
42	63	<i>Manica rubida</i>
41	50	<i>Manica rubida</i>
41	50	<i>Tetramorium bicarinatum</i>
41	44	<i>Myrmica ruginodis</i>
41	52	<i>Stenamma debile</i>
41	44	<i>Myrmica ruginodis</i>
40	40	<i>Myrmica ruginodis</i>
39	52	<i>Tetramorium bicarinatum</i>
39	48	<i>Tetramorium bicarinatum</i>
38	42	<i>Myrmica ruginodis</i>
38	41	<i>Myrmica ruginodis</i>
37	44	<i>Pogonomymex californicus</i>
36	41	<i>Stenamma debile</i>
36	43	<i>Pogonomymex californicus</i>
35	44	<i>Pogonomymex californicus</i>
33	37	<i>Tetramorium africanum</i>
33	41	<i>Tetramorium africanum</i>
29	33	<i>Tetramorium bicarinatum</i>
28	44	<i>Manica rubida</i>
54	58	<i>Tetramorium bicarinatum</i>

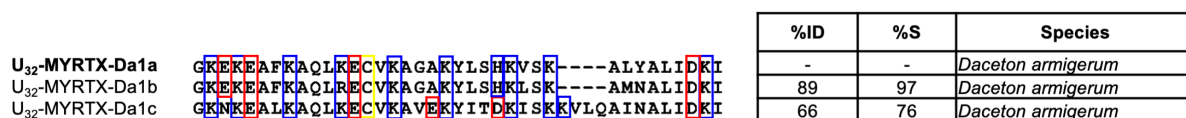
2050
 2051 **Figure 9:** Mature myrmicotoxins of the U₁₇ family. Myrmicotoxins identified in this study were compared to
 2052 U₁₇ venom peptides from *T. bicarinatum* and *M. rubida* (Touchard, Aili, et al., 2020; Touchard et al., 2018),
 2053 and are highlighted in bold. Percentage identity (%ID) and similarity (%S) are relative to the first peptide.
 2054 “ξ” denotes putative peptides. “*” denotes C-terminal amidations. Alignments were generated with the
 2055 Muscle program in Seaview version 4.6.1 and edited using BOXSHADE version 3.2 (Gouy et al., 2010).
 2056 Cysteines are surrounded in yellow.

2057 Six myrmecitoxins containing three disulfide bonds were identified in *P. californicus* and
 2058 *M. ruginodis* venoms, accounting for less than 1% of each venom peptidome (i.e. U₁₈) (Figure
 2059 5). They also were predicted by ScanProsite to have an EGF-like domain (Epidermal Growth
 2060 Factor), a specific structural domain for animal venoms which was only observed in the venoms
 2061 of sea anemones and in *M. rubida* venom until now (Madio, Undheim, & King, 2017; Touchard,
 2062 Aili, et al., 2020). Furthermore, the toxin U₁₈-MYRTX-Mru1b exhibited a missing 146 Da mass
 2063 unit in its calculated mass, corresponding to an additional fucose glycan (Supplementary Table
 2064 XII – Figure 10) and in accordance with the presence of the consensus site observed in its
 2065 primary sequence -C₂XXXXS/TC₃-.
 2066



2067
 2068 **Figure 10:** EGF-like myrmecitoxins (i.e. U₁₈). “ξ” denotes putative peptides. Myrmecitoxins identified in
 2069 this study were compared to U₁₈ venom peptides from *M. rubida* (Touchard, Aili, et al., 2020), and are
 2070 highlighted in bold. Percentage identity (%ID) and similarity (%S) are relative to the first peptide.
 2071 Alignments were generated with the Muscle program in Seaview version 4.6.1 (Gouy et al., 2010), and edited
 2072 using BOXSHADE version 3.2. Conserved residues are highlighted in cyan, identical residues are
 2073 highlighted in magenta and similar residues are highlighted in blue. The threonine residues which are
 2074 modified by the addition of a fucose are surrounded in red.

2075 The last mature myrmecitoxin family groups the dimeric peptides of *D. armigerum* venom.
 2076 These amphiphilic and polycationic dimers are gathered in a single family (i.e. U₃₂) accounting
 2077 for 84% of the venom peptidome (Figure 5). The chains composing those toxins are from 37 to
 2078 41 amino acids long and contain one cysteine allowing the formation of an interchain disulfide
 2079 bond. They possess very similar sequences with a mean of 78% sequence identity (Figure 11).
 2080 However, they display no similarity with previously reported venom peptides.



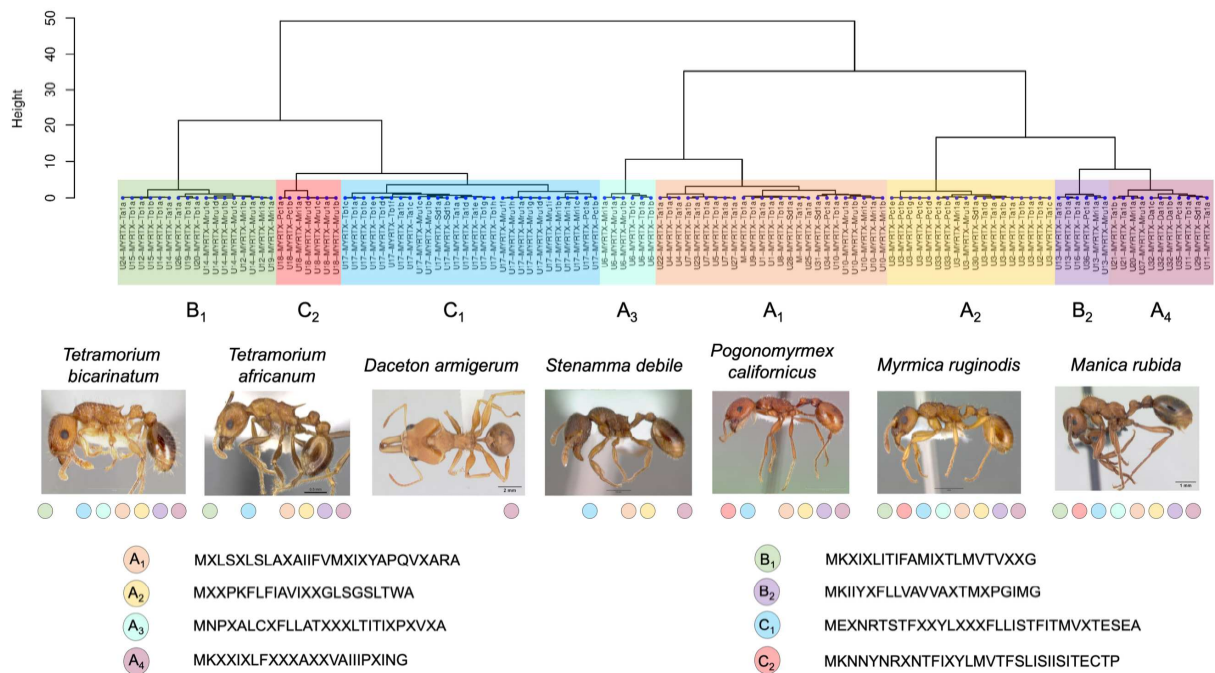
2081
 2082 **Figure 11:** Peptidic chains of dimeric toxins from *D. armigerum* venom. Percentage identity (%ID) and
 2083 similarity (%S) are relative to the first peptide of each family. Alignments were generated with the Muscle
 2084 program in Seaview version 4.6.1 (Gouy et al., 2010), and edited using BOXSHADE version 3.2. Cysteines

2085 are surrounded in yellow. Positively charged residues are surrounded in blue whereas negatively charged
 2086 residues are surrounded in red.

2087

2088 3.3. Superfamilies of myrmecitoxins precursors

2089 In order to determine to which superfamily of precursors each myrmecitoxin belongs, a
 2090 Hierarchical Cluster Analysis (HCA) based on a total of 118 signal sequences was performed.
 2091 This permitted us to define eight groups (Figure 12). As we included *T. bicarinatum* and *M.*
 2092 *rubida* signal sequences in the analysis (Touchard, Aili, et al., 2020; Touchard et al., 2018), we
 2093 initially named the clusters according to the superfamilies of precursors previously defined in
 2094 these ant venom peptidomes which were based on prepropeptide regions (i.e. A₁, A₂, B₁, B₂
 2095 and C) (Touchard, Aili, et al., 2020; Touchard et al., 2018). The focus on signal sequences, and
 2096 the addition of new myrmecitoxin precursors led to the division of superfamilies A and C into
 2097 four additional superfamilies (i.e. A₃, A₄, C₁ and C₂) (Figure 12).
 2098



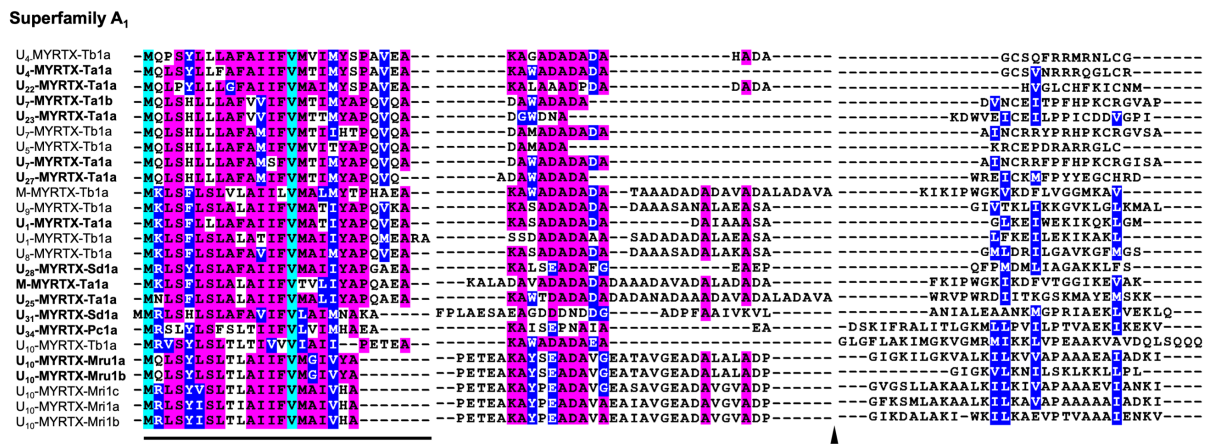
2099

2100 **Figure 12:** Superfamilies of myrmecitoxin precursors. We included every venom peptide precursor from
 2101 myrmicine ant venoms, including those defined in *T. bicarinatum* and *M. rubida* venom peptidomes
 2102 (Touchard, Aili, et al., 2020; Touchard et al., 2018). Superfamilies were defined using a hierarchical cluster
 2103 analysis based on signal sequences. Multiple alignments were achieved with ClustalW and pairwise distances
 2104 were computed with MEGAX software (Kumar et al., 2018). Hierarchical cluster analysis was performed
 2105 with the R software (R Core Team, 2017). The colored circles and their position under the photos refer to
 2106 superfamilies of myrmecitoxin precursors found. Every ant photography is extracted from AntWeb
 2107 (<https://www.antweb.org>).

2108 3.3.1. Superfamily A₁

2109 The superfamily A₁ was represented in every myrmicine ant venom studied so far
 2110 (Touchard, Aili, et al., 2020; Touchard et al., 2018), except for *D. armigerum*. The superfamily
 2111 A₁ was made up of 25 signal sequences, reflecting fifteen mature myrmicotoxin families (i.e.
 2112 M, U₁, U₄, U₅, U₇, U₈, U₉, U₁₀, U₂₂, U₂₃, U₂₅, U₂₇, U₂₈, U₃₁ and U₃₄). The signal sequences
 2113 defined a 28 amino acids long consensus sequence:
 2114 MXLSXLSLAXAIIIFVMXIXYAPQVXARA (Figures 12 and 13), and shared from 54 (i.e. M-
 2115 MYRTX-Tb1a) to 73% (i.e. U₈ and U₉-MYRTX-Tb1a) sequence identity with it. The signal
 2116 sequences of U₃₄-MYRTX-Pc1a and U₁₀-MYRTX-Tb1a were more dissimilar as they shared
 2117 39 and 42% sequence identity with the consensus sequence, respectively.

2118 Most of the superfamily A₁ propeptides bore the consensus motif KAXADADADA
 2119 with an additional C-terminal extension which alanine-rich (from 0 to 22 amino acids). The
 2120 mature parts had quite divergent primary sequences with no strictly conserved features (Figure
 2121 13).



2122 **Figure 13:** Superfamily A₁ of myrmicotoxins precursors. The myrmicotoxin precursors identified in this study
 2123 were compared to those observed in *T. bicarinatum* and *M. rubida* venom peptidomes (Touchard, Aili, et al.,
 2124 2020; Touchard et al., 2018), and are highlighted in bold. “ξ” denotes putative peptides. Alignments were
 2125 generated with the Muscle program in Seaview version 4.6.1(Gouy et al., 2010), and edited using
 2126 BOXSHADE version 3.2. Conserved residues are highlighted in cyan, identical residues are highlighted in
 2127 magenta and similar residues are highlighted in blue. Black lines indicate signal sequences where black
 2128 triangles indicate the cleavage site releasing the mature peptides. Signal, propeptide and mature sequences
 2129 were aligned separately.

2131 3.3.2. Superfamily A₂

2132 As observed for the superfamily A₁, the superfamily A₂ was identified in every
 2133 myrmicine ant venom studied until now (Touchard, Aili, et al., 2020; Touchard et al., 2018),
 2134 except for *D. armigerum* (Figure 12), and gathered 18 signal sequences, representing four

2135 mature myrmecitoxin families (i.e. U₂, U₃, U₃₀ and U₃₃). They shared from 65 (i.e. U₃-MYRDX-
2136 Pc1b) to 82% (i.e. U₂-MYRDX-Ta1a) sequence identity with the consensus sequence:
2137 MXXPKFLFIAVIXXGLSGSLTWA (i.e. 23 amino acids long) (Figures 12 and 14).

2138 The majority of the superfamily A₂ propeptides showed the motif XPLA₂ in their N-
2139 terminal part, coupled with the pattern AXAXAEAXAXAXAXAEAXAEAXXEA in their C-
2140 terminal part. Consequently, the propeptides regions were quite long, as they contained 27 to
2141 62 amino acids. Most of the precursor mature parts show similar sequences, as expected by the
2142 gathering of the whole U₃ family (Figure 14). However, the precursors encoding distinct mature
2143 myrmecitoxin families did not show sequence similarity.



2144
2145 **Figure 14:** Superfamily A₂ of myrmecitoxins precursors. The myrmecitoxin precursors identified in this study
2146 were compared to those observed in *T. bicarinatum* and *M. rubida* venom peptidomes (Touchard, Aili, et al.,
2147 2020; Touchard et al., 2018), and are highlighted in bold. “ ξ “ denotes putative peptides. Alignments were
2148 generated with the Muscle program in Seaview version 4.6.1 (Gouy et al., 2010), and edited using
2149 BOXSHADE version 3.2. Conserved residues are highlighted in cyan, identical residues are highlighted in
2150 magenta and similar residues are highlighted in blue. Black lines indicate signal sequences whereas black
2151 triangles indicate the cleavage site releasing the mature peptides. Signal, propeptide and mature sequences
2152 were aligned separately.

3.3.3. Superfamily A₃

2155 The superfamily A₃ was redefined here as it gathered the U₆ venom peptide precursors.
2156 In the studies of *T. bicarinatum* and *M. rubida* venom peptidomes, the most part was previously
2157 classified in the superfamily A₁, except for U₆-MYRDX-Tb1c who was classified in the
2158 superfamily A₃ (Touchard, Aili, et al., 2020; Touchard et al., 2018). Here, the superfamily A₃
2159 included the U₆ precursors of *T. bicarinatum*, *M. rubida* and *M. ruginodis* venoms. The
2160 consensus signal sequence was 27 amino acids long and was defined as
2161 MNPXALCXFLLATXXXLTITIXPVXA (Figures 12 and 16), with which the precursors
2162 shared from 52 (i.e. U₆-MYRDX-Tb1a) to 70% (i.e. U₆-MYRDX-Mru1b) sequence identity.

2163 The propeptide sequences were 8 amino acids long and were very conserved, following the
 2164 consensus sequence NAEAXADA, except for the propeptide U₆-MYRTX-Tb1c which was 6
 2165 amino acids long (Figure 15).
 2166

Superfamily A₃

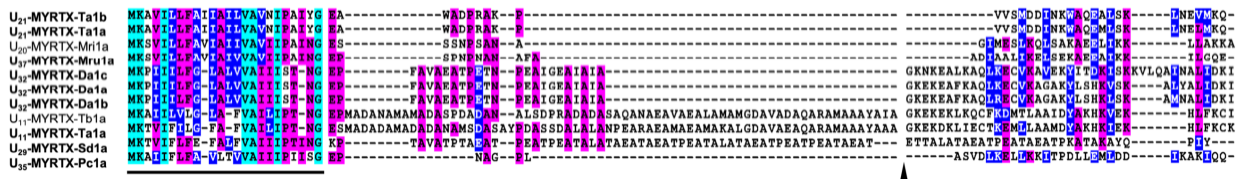
U ₆ -MYRTX-Mri1a	M N P K A L C S F L L A T F L L L T V T I M P S V H -	N A E A N A D A	I I G P C P K K P I G ---	I V C
U ₆ -MYRTX-Mru1a	M N P K A L C S F L L A T F L L L T V T I T P L V R A	N A E A N A D A	L I G P C P K K P I G ---	I K C
U ₆ -MYRTX-Mru1b	M N P K A L C S F L L A T F L L L T T I T P S V R A	N A E A N A D A	L I D P C P K I F K G ---	L F C
U ₆ -MYRTX-Tb1c	M L - Q A L C F S L L A T I V T L T I I M P A A Y A	N A E P -- ε A	F R G P C P K D M F K G R F I M C	
U ₆ -MYRTX-Tb1a	M L Q Q H L V V F L L A V I V T L T I T I V P T V H A	N A E A D A D A	L W G K C P K --	I G G R R V M C
U ₆ -MYRTX-Tb1b	M N P T L C C F L L A V I V T L T I T I V P T V H A	N A E A D A D A	L W G K C P K --	I G G R R I M C

2167
 2168 **Figure 15:** Superfamily A₃ of myrmicitoxins precursors. The myrmicitoxin precursors identified in this study
 2169 were compared to those observed in *T. bicarinatum* and *M. rubida* venom peptidomes (Touchard, Aili, et al.,
 2170 2020; Touchard et al., 2018), and are highlighted in bold. “ε” denotes putative peptides. Alignments were
 2171 generated with the Muscle program in Seaview version 4.6.1 (Gouy et al., 2010), and edited using
 2172 BOXSHADE version 3.2. Conserved residues are highlighted in cyan, identical residues are highlighted in
 2173 magenta and similar residues are highlighted in blue. Black lines indicate signal sequences whereas black
 2174 triangles indicate the cleavage site releasing the mature peptides. Signal, propeptide and mature sequences
 2175 were aligned separately.

3.3.4. Superfamily A₄

2176
 2177 The superfamily A₄ was defined as it gathered previous members belonging to the
 2178 superfamilies “A₁ and A₂” identified in *T. bicarinatum* and *M. rubida* venoms (i.e. U₁₁-
 2179 MYRTX-Tb1a and U₂₀-MYRTX-Mri1a) (Touchard, Aili, et al., 2020; Touchard et al., 2018).
 2180 This superfamily was represented in every myrmicine ant venom and included 11 signal
 2181 sequences, which defined a 24 amino acids long consensus sequence,
 2182 MKXXIXLFXXXAXXVAIIPXING. Seven mature myrmicitoxin families (i.e. U₁₁, U₂₀, U₂₁,
 2183 U₂₉, U₃₂, U₃₅ and U₃₇) and the signal sequences shared from 50 (i.e U₂₁) to 65 % (i.e. U₂₉-
 2184 MYRTX-Sd1a) sequence identity with the consensus sequence. The propeptide regions were
 2185 alanine-rich and exhibited various length, ranging from 7 to 68 amino acids, for U₃₅-MYRTX-
 2186 Pc1a and U₁₁ peptides, respectively. The mature sequences exhibited quite divergent sequences
 2187 with no strictly conserved features (Figure 16).

Superfamily A₄



2188

2189 **Figure 16:** Superfamily A₄ of myrmecitoxins precursors. The myrmecitoxin precursors identified in this study
 2190 were compared to those observed in *T. bicarinatum* and *M. rubida* venom peptidomes (Touchard, Aili, et al.,
 2191 2020; Touchard et al., 2018), and are highlighted in bold. “ξ” denotes putative peptides. Alignments were
 2192 generated with the Muscle program in Seaview version 4.6.1 (Gouy et al., 2010), and edited using
 2193 BOXSHADE version 3.2. Conserved residues are highlighted in cyan, identical residues are highlighted in
 2194 magenta and similar residues are highlighted in blue. Black lines indicate signal sequences whereas black
 2195 triangles indicate the cleavage site releasing the mature peptides. Signal, propeptide and mature sequences
 2196 were aligned separately.

2197

3.3.5. Superfamilies B

2198

2199 The superfamilies B₁ and B₂ were previously described in *T. bicarinatum* and *M. rubida*
 2200 venom peptidomes in which they exhibited a total of five (i.e. U₁₂, U₁₄, U₁₅, U₁₉ and U₂₀-
 2201 MYRTX-Tb1a) and two (i.e. U₁₃ and U₁₆) mature myrmecitoxin families, respectively
 2202 (Touchard, Aili, et al., 2020; Touchard et al., 2018). In our study, those superfamilies were
 2203 grouped in distinct clusters, in accordance with the previous definitions (Touchard, Aili, et al.,
 2204 2020; Touchard et al., 2018).

2205

2206 The U₁₄ precursors identified in *M. ruginodis* and *T. africanum* venom peptidomes were
 2207 gathered with their counterparts in the superfamily B₁, as were U₂₄ and U₂₆ precursors from *T.*
 2208 *africanum* venom. Thus, the superfamily B₁ gathered a total of 17 signal sequences, defining a
 2209 23 amino acids long consensus sequences MKXIXLITIFAMIXTLMVTVXXG, with which
 2210 they shared from 57 (i.e. U₁₄-MYRTX-Ta1a) to 78% (i.e. U₁₂- and U₁₉-MYRTX-Mri1a)
 2211 sequence identity (Figures 12 et 17). Those precursors exhibited short propeptide regions from
 2212 4 to 2 amino acids long. Similarly to what has been observed for the superfamily A₃, the
 2213 precursors encoding distinct mature myrmecitoxin families did not show sequence similarity.

2213

2214 The U₁₃ precursors observed in *M. ruginodis* and *T. africanum* venom peptidomes were
 2215 clustered in the superfamily B₂, as was the U₃₆ precursor from *P. californicus* venom.
 2216 Therefore, this superfamily contained 6 signal sequences, forming the consensus sequence
 2217 MKIIYXFLLVAVVAXTMXPGIMG (i.e. 23 amino acids long), with which they shared from
 70 to 74% sequence identity (Figure 12). Most of the propeptide regions were from 6 to 13

2218 amino acids long and exhibited the conserved motif EAXXAEA. As for the mature parts of
 2219 these precursors, they all exhibited the conserved pattern IGXF in their N-terminal region,
 2220 although they belong to three distinct mature myrmecitoxin families (Figure 17).

Superfamily B₁

U ₁₅ -MYRTX-Tb1b	MKIV KLITIFAMIATLMVAVANG --EA --VF TPDQ IKAMIKRH---
U ₁₂ -MYRTX-Tb1a	MKIV KLITIFAMIATLMVAVANG --KA --- LS PAVLASLA-----
U ₁₅ -MYRTX-Tb1a	MKIV KLITIFAMIATLMVAVANG --EA --- LT ADQLKAIKRH---
U₂₄-MYRTX-Ta1a	MKI I KLITIFAMIATLM IT VAVANG --KP -TQSK NP DVVIRL-----
U ₁₄ -MYRTX-Tb1a	MKI I KLITIF TMM ATLMVAVANG -- E --- TP NAVKS LQ -----
U ₁₄ -MYRTX-Ta1a	MKIV KLITIFAM MATLIT AVANG --KA --- TP PAALASLA-----
U₂₆-MYRTX-Ta1a	MKTL KLITIFAMIA-L MIT MVAA --- --- Q PTIFV PRL -----
U ₁₉ -MYRTX-Tb1a	MKSIN LITIFAMIATLMATMIAG --EA ---ARSRLK IGRM GR---
U ₂₀ -MYRTX-Tb1a	MKSIN LITIFAMIATLMATMIAG --EA ARVSDCTMFTSK LKL LIPV
U ₁₄ -MYRTX-Mru1e [§]	MKTIE LITIFAMIT TTW MVTVMA-TPDP --- IN PK L ESL-----
U ₁₄ -MYRTX-Mru1d	MKTIE LITIFAMIT TTW MVTVMA-TPDP --- IN PKLW LK LF S KL ES V
U ₁₄ -MYRTX-Mru1b	MKTIE LITIFAMIT TTW MVTVMA-TPDP --- ID PKV L ESL L -----
U ₁₄ -MYRTX-Mru1c	MKTIE LITIFAMIT TTW MVTVMA-TPDP --- ID SDAL K SL Q GGTV--
U ₁₂ -MYRTX-Mri1b	MKTI KLITIFAI IT TLMAT IV TVG --ES --- TP PKA K SL Q -----
U ₁₄ -MYRTX-Mru1a	MKTIE LITIFAMIT TTW MV TV IVTVG --DP --- ID SKA K SL Q -----
U ₁₂ -MYRTX-Mri1a	MKTIE LITIFAMIT TL MV TV VAVG --DP --- ID PKV L ESL V -----
U ₁₉ -MYRTX-Mri1a	MKTIE LITIFAMIT TL MV TV VVIG --DP --- ID SA A IA T L Q GGTV--

Superfamily B₂

U ₁₃ -MYRTX-Ta1a	MKI I YVFL VAV IAAVTMIPGIMG EAE AEAESEHGKA -----SK I GL F DQID K GM A WFMD L FK
U ₁₃ -MYRTX-Tb1a	MKLI Y F SLVAVVIVTMIPGIMG EAE AE G -----RPPQ I GL F DQID K GM A WFMD L FK
U ₁₆ -MYRTX-Tb1a	MKLI Y F FLVAVIAV TM IPGIMG EA -----GDEAGK-PK I GL F HD V NKA T EW L L K Q T K
U₃₆-MYRTX-Pc1a	MKI I Y A FL L AVVAF-MGSG I MA EPL AE A IA-----VDKP Q Q A KE I GL F DRIT E L N W L V N H--
U ₁₃ -MYRTX-Mri1a	MKI I HVLLVAVVA IT MSE S IMA ESV AE A -----DKP Q Q A KK I GL F DQID K AA A AF M KL F E
U ₁₃ -MYRTX-Mru1a	MKI I HVLLVAVVA IT AMSP G IMA ES AE A -----DK F Q Q AK I GL F DQID R AW A WF M KL F E

2221
 2222 **Figure 17:** Superfamilies B of myrmecitoxins precursors. The myrmecitoxin precursors identified in this
 2223 study were compared to those observed in *T. bicarinatum* and *M. rubida* venom peptidomes (Touchard, Aili,
 2224 et al., 2020; Touchard et al., 2018), and are highlighted in bold. “[§]” denotes putative peptides. Alignments
 2225 were generated with the Muscle program in Seaview version 4.6.1 (Gouy et al., 2010), and edited using
 2226 BOXSHADE version 3.2. Conserved residues are highlighted in cyan, identical residues are highlighted in
 2227 magenta and similar residues are highlighted in blue. Black lines indicate signal sequences whereas black
 2228 triangles indicate the cleavage site releasing the mature peptides. Signal, propeptide and mature sequences
 2229 were aligned separately.

3.3.6. Superfamilies C

2230
 2231 A single superfamily C was previously defined in *T. bicarinatum* and *M. rubida* venom
 2232 peptidomes, containing both U₁₇ and U₁₈ peptides (Touchard, Aili, et al., 2020; Touchard et al.,
 2233 2018). The HCA allowed us to distinguish two separate clusters gathering those mature
 2234 myrmecitoxins families, respectively (Figure 12). The superfamily C₁ grouped the 28 signal
 2235 sequences of U₁₇ precursors and was represented in every myrmecine ant venom peptidome,
 2236

2237 except for *D. armigerum* (Figure 17). The signal sequences shared from 52 (i.e. U₁₇-MYRTX-
 2238 Pc1a) to 71% (i.e. U₁₇-MYRTX-Mru1b, -Mru1c, -Sd1a, -Sd1b) sequence identity with the
 2239 defined consensus sequence of 33 amino acids long:
 2240 MEXNRTSTFXXYLXXXFLLISTFITMVXTESEA. The superfamily C₂ gathered 7 signal
 2241 sequences of U₁₈ precursors from *P. californicus*, *M. ruginodis* and *M. rubida* venom
 2242 peptidomes (Touchard, Aili, et al., 2020) (Figure 17). The signal sequences shared from 52 (i.e.
 2243 U₁₈-MYRTX-Pc1A, -Pc1b) to 94% (i.e. U₁₈-MYRTX-Mru1a) sequence identity with the
 2244 consensus sequence MKNNYNRXNTFIXYLMVTFSLIISITECTP, which is 32 amino acids
 2245 long. For both of these superfamilies, the signal sequences were not followed by any propeptide
 2246 regions.

Superfamily C₁

```

U17-MYRTX-Tb1c MEKHRTNIFSYYLMITFLLISIFITMVMSDGEA TITDVP IQ----CPPGK-VKV--GNRCRVHF-----
U17-MYRTX-Tb1a MEKHRTNIFSYYLMITFLLISIFITMVMSDGEA TIINAPNR----CPPGH-VVV--KGRERIA-----
U17-MYRTX-Tb1b MEKHRTNIFSYYLMITFLLISIFITMVMSDGEA TVIDVP IQ----CPSGT-VKV--GNRCRVHF-----
U17-MYRTX-Tb1e MEKNRTTFFSVYLTITLFLISTFITMVITES-- NIKAPLFP----CPNGYIRDY--KGDREHTE-----
U17-MYRTX-Tb1d MEKNRTTFFSVYLTITLFLISTFITMVITEA-- NIIKRVV----CRAGY-IEV--NGVCREVFT-----
U17-MYRTX-Tb1f MEKNRTTFFSVYLTITLFLISTFITMVITES-- NIIKRVV----CRAGY-IEV--NGVCREVFT-----
U17-MYRTX-Ta1b MEKNRTTFFSVYLTITLFLISTFITMVITES-- NIIKRVV----CRAGY-IEV--NGVCREVFT-----
U17-MYRTX-Ta1c MEKNRTTFFSVYLTITLFLISTFITMVITES-- NIIKRVV----CRAGY-IEV--NGVCREVFT-----
U17-MYRTX-Mru1cξ MERNHTSTFFSHYLVTFVLISTFITMVVITES-- GIIRVPELQ----CPDGYKDA--LVGCRVFT-----
U17-MYRTX-Mru1bξ MGNRTSTFFTYLMTFLLISTFITMVGHTES-- HIIVFPFN----CPPGQKP--RGNCRQVY-----
U17-MYRTX-Sd1a MEKGRISMFSYTLMTFLLISTFITMVVITES-- NIIIVPERP----CPSGQRDS--RGCRCVLS-----
U17-MYRTX-Sd1b MEKGRISMFSYTLMTFLLISTFITMVVITES-- YIIDAPSRK----CPEGSRST--QGCRTTSR-----
U17-MYRTX-Ta1a MEKNRNIFSTYLMITFLLISTFITMVISES-- GIINAPDR----CPDGYKDA--GNCRKVI-----
U17-MYRTX-Ta1d MEKNRTYIFSTYLMITFLLISTFITMVISES-- SIVNVP IQ----CPPGT-IQV--GKRCRET-----
U17-MYRTX-Ta1e MEKNRTSIFSTYLMITFLLISTFITMVITEA-- NIIKVP----CRAGY-RKI--NGVCRKLYR-----
U17-MYRTX-Tb1g MEKSRITSTFFSYLMTFLLISTFITMVITEA-- HVIDTRSSL--CPEGSRST--TGCKRTV-----
U17-MYRTX-Tb1h MEK-RVSTFFSYLMTFLLISTFITMVITEA-- HIIRVP----CRAGY-KEI--RGRCKRLT-----
U17-MYRTX-Mru1fξ MESNCISTLWASLMIAFLLISTFITMVVSES-- HIIVVPERPK--CPPGQEDR--RGRCRMVP-----
U17-MYRTX-Mru1hξ MESNCISTLWASLMVAFLLISTFITMVVSES-- HIIVLNNRNTNT--CPPGQKDN--YGNCRKIA-----
U17-MYRTX-Mru1aξ MENNRISTLWASLMVAFLLISTFITMVVSES-- AIIDAPRN----CPPGHDIDH--AGDCVEHFE-----
U17-MYRTX-Mru1gξ MESNRISTLWASLMVAFLLISTFITMVVSES-- HIIVVPELQ----CPPGQKDR--QGRCRVFN-----
U17-MYRTX-Mru1dξ MESNRISTLWASLMVAFLLISTFITMVVSES-- HIIVVPELQ----CPPGQKDR--QGRCRVFN-----
U17-MYRTX-Mru1eξ MESNRISTLWASLMVAFLLISTFITMVVSES-- QIIVVPRK--CNERSKTD--DAGICRRTV-----
U17-MYRTX-Mru1b MEKNSHTSTFFAYVTVAFLLISTFITMVVITES-- HIIVVP----CREGY-VMV--GNVCEEVY-----
U17-MYRTX-Mru1a MEKNSHTSTFFAYVTVAFLLISTFITMVVITES-- HIIVVP----CREGY-VMV--GNVCEEVY-----
U17-MYRTX-Mru1c MEKNSHTSTFFAYVTVAFLLISTFITMVVITES-- YITVAP----CREGY-VMV--GNVCEEVY-----
U17-MYRTX-Pc1a MENRYVNFPTTYLMTFLLISTFITMAVITES-- HITQVP----CLPGY-VKVGKDVCREAFKFKPGQRP-----
U17-MYRTX-Pc1b MKACYINIFPTTYLMTFLLISTFITMAVITES-- HITQVP----CLPGY-VMV--NGVCEEVYNEKDEE-----
  
```

Superfamily C₂

```

U18-MYRTX-Pc1a M-N--NNANKFIIYITIISSLLSMVLIKECMT ENMSNRIPTTEED-IQYCFN-GECFYFP--A-INAAGCIENKQYDYRCNTIILD-----
U18-MYRTX-Pc1bξ M-N--NNANKFIIYITIISSLLSMVLIKECMT ENMSNRIPTTEEY-NGYCLN-GECFRIQ--GDNMKNHGVCPKEFSGERCOIRDYA-----
U18-MYRTX-Mru1aξ MKNNYNINTEIIVYLMVTFSLIISITICTEPT NH--D--PCFPQYAEALCLNGTCTSVTIMG--SDNYNCICAPGFGQWRCKEKLDFHPVNO-----
U18-MYRTX-Mru1dξ MKN--NRNNTPIIYLMVTFSLIISITICTEPT NH--D--PCFPQYAEALCLNGTCTSVTIMG--SDNYNCICAPGFGQWRCKEKLDFHPVNO-----
U18-MYRTX-Mru1cξ MKN--NRNNTPIIYLMVTFSLIISITICTEPT QSHFQ--PCPSSH-EFECFN-GECFYLA--A-ENEIGCICPPGFGQRCGELLILQYV-----
U18-MYRTX-Mru1aξ MKN--NRNNTPIIYLMVTFSLIISITICTEPT THLEP--SCPSSH-EFECFN-GECFYLA--E-LNTYACICAPGFGHRCETANLQ-----
U18-MYRTX-Mru1bξ MKN--NRNNTPIIYLMVTFSLIISITICTEPT QSHFQ--PCPSSH-EFECFN-GECFYLA--A-ENEIGCICPPGFGQRCGELLILQ-----
  
```

2247
 2248 **Figure 18:** Superfamilies C of myrmecitoxins precursors. The myrmecitoxin precursors identified in this
 2249 study were compared to those observed in *T. bicarinatum* and *M. rubida* venom peptidomes (Touchard, Aili,
 2250 et al., 2020; Touchard et al., 2018), and are highlighted in bold. “^ξ” denotes putative peptides. Alignments
 2251 were generated with the Muscle program in Seaview version 4.6.1 (Gouy et al., 2010), and edited using
 2252 BOXSHADE version 3.2. Conserved residues are highlighted in cyan, identical residues are highlighted in
 2253 magenta and similar residues are highlighted in blue. Black lines indicate signal sequences whereas black
 2254 triangles indicate the cleavage site releasing the mature peptides. Signal, propeptide and mature sequences
 2255 were aligned separately.

2256 4. Discussion

2257

2258 In this study, we identified a total of 78 myrmecitoxins precursors, which were classified
2259 into 37 families according to their mature sequences, and into 8 superfamilies based on their
2260 signal sequences. The characterized venom peptidomes showed a wide heterogeneity in their
2261 peptidic compositions, with some exhibiting more diversity than others. For instance, *T.*
2262 *africanum* venom shows the highest number of mature toxin families of this study. Similarly to
2263 the toxins observed in the predatory ant *T. bicarinatum* venom (Touchard et al., 2018), the
2264 majority of those families gather amphiphilic cationic linear peptides and some contain toxins
2265 structured by one disulfide bond. Furthermore, *T. africanum* venom peptidome exhibited
2266 several peptides which were analogous to those described in *T. bicarinatum* venom, such as M-
2267 MYRTX-Ta1a and U₁-MYRTX-Ta1a (i.e. Bicarinaline and P17 in *T. bicarinatum* venom
2268 peptidome, respectively) (Touchard et al., 2018). Bicarinaline and P17 are both linear,
2269 polycationic and α -helicoidal peptides. They possess similar sequences and yet display distinct
2270 biological activities. Indeed, Bicarinaline exerts a cytolytic action while P17 exhibits an
2271 immunomodulatory activity through the activation of a GPCR (i.e. MRGPRX2) (Benmoussa
2272 et al., 2017; Duraisamy et al., 2020; Téné et al., 2016), therefore illustrating that similar toxins
2273 in terms of primary sequences can display very different functions. *T. africanum* venom
2274 peptidome also exhibited several analogues of disulfide-bonded peptides from *T. bicarinatum*
2275 venom (i.e. U₆, U₇ and U₁₁), whose biological activities remain to be determined. However, as
2276 the presence of a disulfide bond induces a folding as well as more stability, we hypothesize that
2277 myrmicine ant venom peptides exhibiting this scaffold might specifically interact with one or
2278 several receptors. These results highlight the multifunctionality of venoms and peptidic toxins
2279 composing them.

2280 *P. californicus*, *S. debile* and *M. ruginodis* venom peptidomes also show a majority of linear
2281 peptide families and disulfide-bonded toxins, but exhibit much less myrmecitoxin families than
2282 *T. africanum* and *T. bicarinatum* venoms, similarly to *M. rubida* venom peptidome (Touchard,
2283 Aili, et al., 2020; Touchard et al., 2018). Interestingly, the studied ants from the genus
2284 *Tetramorium* use their venoms both for defense and prey capture whereas *P. californicus*, *M.*
2285 *ruginodis* and *M. rubida* use predominantly their venom for defense. Furthermore, the toxin
2286 arsenals of *P. californicus* and *M. ruginodis* are also composed of toxins containing three
2287 disulfide bonds which are part of the U₁₈ family. This family has also been described in *Manica*
2288 *rubida* venom peptidome (Touchard, Aili, et al., 2020), and seem thus to have been retained in
2289 several ant venoms. It was suggested that these peptides might be involved in the renewal of
2290 both the venom gland and the sac epithelia (Touchard, Aili, et al., 2020), but this remains to be
2291 demonstrated.

2292 In the same line, several families of mature myrmecitoxins have been observed in most of
2293 the myrmecine ant venom characterized until now. One of the most striking examples is the
2294 family U₃, which was abundant in *T. africanum*, *P. californicus* and *M. ruginodis* venom
2295 peptidomes. Those peptides were previously described in *T. bicarinatum* and *M. rubida* venom
2296 peptidomes (Touchard, Aili, et al., 2020; Touchard et al., 2018) and were shown to share
2297 sequence similarities with the δ -PPONTX-Pc1a from *Paraponera clavata* (Aili et al., 2020;
2298 Touchard et al., 2018). This toxin is known to inhibit voltage-activated sodium channels
2299 (Na_v1.7), inducing an insecticidal effect as well as pain to vertebrates (Johnson et al., 2017;
2300 Piek, Hue, et al., 1991). Concerning U₃ peptides from myrmecine ant venoms, only U₃-
2301 MYRTX-Mri1a from *M. rubida* was tested and showed no insecticidal activity (Touchard, Aili,
2302 et al., 2020), illustrating once again that venom peptides showing similar biochemical features
2303 do not necessarily display the exact same activity. This was also demonstrated for
2304 paraponeratoxins isolated from different populations of *P. clavata*. The analogue peptides only
2305 differed in their primary sequences of one amino acid in their C-terminal part, and mediated
2306 Na_v1.7 channels with different potencies (Johnson et al., 2017). In our study, U₃ peptides were
2307 not observed in *D. armigerum* and *S. debile* venoms, which both are essentially used to capture
2308 preys. Given their relative abundances in myrmecine ant venoms, with dual or essentially
2309 defensive uses, and the biological activities showed by similar toxins, U₃ peptides presumably
2310 have an important functional role for defense against vertebrates. Disulfide bonded-peptides
2311 from the U₁₇ family also seem to be essential components of myrmecine ant venoms. Indeed,
2312 this family has been found in every myrmecine ant venom peptidomes (Touchard, Aili, et al.,
2313 2020; Touchard et al., 2018), including those from this study, except for *D. armigerum*. The
2314 U₁₇ mature peptides were previously shown to share significant sequence identity with secapins
2315 (Touchard et al., 2018), which are multifunctional peptides found in the venom of numerous
2316 hymenoptera (Lee et al., 2016). However, the biological activity of U₁₇ peptides from
2317 myrmecine ant venoms has not been determined yet. Several U₁₇ peptides from *T. bicarinatum*
2318 and *M. rubida* venoms showed PTMs such as O-glycosylation and N-ter pyroglutamate
2319 modification (Touchard, Aili, et al., 2020; Touchard et al., 2018). In our study, several U₁₇
2320 peptides from *P. californicus* and *M. ruginodis* venoms showed a histidine in their N-terminal
2321 part with the transcriptomics data, and systematically possessed a missing 26 Da mass unit in
2322 their calculated mass. The calculated masses remained undetected in proteomics data for U₁₇
2323 peptides from *M. ruginodis* venom. However, both forms were detected for those identified in
2324 *P. californicus* venom peptidome, with similar relative abundances. A comparable observation
2325 was noted in the venom of *M. rubida*. Indeed, only U₁₇-MYRTX-Mri1a, with a N-terminal

2326 histidine, was visible in transcriptomics data while MS/MS data replaced the N-terminal
2327 histidine by a tyrosine (i.e. U₁₇-MYRTX-Mri1c). The exact mass difference between the two
2328 residues is 26 Da and only the N-terminal tyrosine form was confirmed by proteomics data in
2329 *M. rubida* venom (Touchard, Aili, et al., 2020). Further investigation is therefore needed to
2330 determine if we highlighted a new PTM leading to a gain of 26 Da or if this phenomenon is
2331 artefactual.

2332 Here, we also demonstrate for the first time the presence of dimeric peptides in a myrmicine
2333 ant venom. *D. armigerum* exhibits indeed the most streamlined venom peptidome observed in
2334 this study, which consists in a complex of dimeric peptides. Those homodimers and the
2335 heterodimer are both part of a unique myrmicitoxin family (i.e. U₃₂) with no similarity with any
2336 previously reported venom toxin. Dimeric peptides have been found in pseudomyrmecine,
2337 ponerine, ectatomine and myrmecine ant venoms which are well-known for the pain caused by
2338 their sting (Barassé et al., 2019; Kazuma et al., 2017; K. Pluzhnikov et al., 1999; Robinson et
2339 al., 2018; Touchard et al., 2014; Wanandy et al., 2015). However, *D. armigerum* is an arboreal
2340 ant whose sting is reported as being mildly painful nevertheless lethal to insects (J. O. Schmidt,
2341 2019b). This species uses its venom exclusively to paralyze large preys that can't be overcome
2342 by their trap-jaws, such as large myriapods (Dejean et al., 2012). Therefore, the dimeric
2343 peptides composing *D. armigerum* venom seem likely to be neurotoxic against invertebrates,
2344 although their biological activities remain to be demonstrated. The unique composition of this
2345 venom peptidome and its exclusive predatory use at last resort raise the matter of the
2346 evolutionary advantage of such dimeric toxins. We might assume their presence in *D.*
2347 *armigerum* venom peptidome may be linked to a high stability coupled with an increased target
2348 specificity.

2349 Altogether, the molecular features referring to mature myrmicitoxin families and their
2350 presence among, or exclusive to, different myrmicine ant venoms suggest a compromise
2351 between the diversity of toxins composing the different arsenals and their biochemical
2352 complexity. This might also be coupled with venom specific uses, as myrmicine ants exhibit
2353 wide variations in their ecology. However, further integrative studies of ant venom peptidomes
2354 in terms of compositions, biological activities and ecological data of biological models are
2355 needed.

2356 Moreover, the HCA analysis performed on signal sequences revealed that myrmicitoxin
2357 signal sequences constitute 8 superfamilies, therefore leading to the maturation of 37 different
2358 myrmicitoxin families across myrmicine ant venoms. Conserved signal regions of venom
2359 peptide precursors, coupled with divergent sequences of mature toxins, have already been

2360 observed in *T. bicarinatum* and *M. rubida* venom peptidomes, as well as in pseudomyrmecine,
2361 ponerine and myrmecine ant venoms (Barassé et al., 2019; Kazuma et al., 2017; Mariano et al.,
2362 2019; Robinson et al., 2018). Furthermore, this observation was recently extended to aculeate
2363 hymenoptera (Robinson et al., 2018). These results suggest that venom peptides exhibiting
2364 similar signal sequences might share a gene family and may have evolved through gene
2365 duplication and neofunctionalization (Touchard et al., 2018). However, genomic studies are
2366 needed for confirmation.

2367

2368 **5. Conclusion**

2369 In summary, we provide the first insights into the peptidic diversity of myrmicine ant
2370 venoms. The proteotranscriptomics approach employed in our study allowed the identification
2371 of 78 venom peptide precursors classified into 37 mature peptide families based on sequence
2372 similarities. Most of the myrmecitoxins identified in this study are linear and cationic. Several
2373 toxins families also showed one to three disulfide bonds, and we identified for the first time the
2374 presence of dimeric myrmecitoxins. Myrmicine ant venoms exhibited various venom
2375 compositions, with mature peptide families exclusive to one venom, and some which were
2376 encountered in multiple ones. Altogether, our results suggest a compromise between the
2377 diversity of toxins composing the different arsenals and their biochemical complexity, that
2378 might also be coupled with venom specific uses. Furthermore, the analysis of signal sequences
2379 revealed that the myrmecitoxins precursors can be classified into 8 superfamilies. This result
2380 suggests that myrmecitoxins clustered in one superfamily might belong to the same gene family,
2381 although genomic studies are need for confirmation. Besides, the molecular diversification of
2382 ant venom toxins seems to be led by a multitude of parameters, highlighting the need for more
2383 integrative studies of ant venom peptidomes.

2384

2385 **Acknowledgements:** V.B. was the recipient of a PhD fellowship from the French Ministry of
2386 Scientific Research. This study was performed in collaboration with the GeT core facility,
2387 Toulouse, France (<http://get.genotoul.fr>), and was aided by the France Génomique National
2388 institutional infrastructure, funded as part of the “Investissement d’avenir” program managed
2389 by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09). Ant samples were
2390 collected under the authorizations of both the Cameroon Ministry of Scientific Research and
2391 Innovation and the French Ministry of Ecological and Solidarity Transition, in accordance with
2392 Article 17, paragraph 2, of the Nagoya Protocol on Access and Benefit-sharing (Reference
2393 number of the permit: TREL1820249A/54).

2394 **Supplementary materials:**

2395 **Figure S1:** Total ion chromatogram (TIC) of crude venom and venom after hexanic extraction from
2396 *Solenopsis saevissima*.

2397 **Supplementary Table I:** Source and composition of ant venom samples used for proteomic analyses.

2398 **Supplementary Table II:** Source and composition of ant tissue samples used for transcriptomic
2399 analyses.

2400 **Supplementary Table III:** Peptide mass fingerprinting of *Tetramorium africanum* venom.

2401 **Supplementary Table IV:** Peptide sequences in the venom of *Tetramorium africanum*.

2402 **Supplementary Table V:** Peptide mass fingerprinting of *Stenamma debile* venom.

2403 **Supplementary Table VI:** Peptide sequences in the venom of *Stenamma debile*.

2404 **Supplementary Table VII:** Peptide mass fingerprinting of *Daceton armigerum* crude venom.

2405 **Supplementary Table VIII:** Peptide sequences in the venom of *Daceton armigerum*.

2406 **Supplementary Table IX:** Peptide mass fingerprinting of *Pogonomyrmex californicus* venom.

2407 **Supplementary Table X:** Peptide sequences in the venom of *Pogonomyrmex californicus*.

2408 **Supplementary Table XI:** Peptide mass fingerprinting of *Myrmica ruginodis* venom.

2409 **Supplementary Table XII:** Peptide sequences in the venom of *Myrmica ruginodis*.

2410 **Supplementary Table XIII:** Mass fingerprinting of *Solenopsis saevissima* venom after hexanic extraction.

2411 **Supplementary Table XIV:** Venom glands RNA sequencing data.

2412 **Supplementary Table XV:** Addressing table of major contigs expressed by *Myrmica ruginodis* venom
2413 glands.

2414 **Supplementary Table XVI:** Addressing table of major contigs expressed by *Tetramorium africanum*
2415 venom glands.

2416 **Supplementary Table XVII:** Addressing table of major contigs expressed by *Pogonomyrmex californicus*
2417 venom glands.

2418 **Supplementary Table XVIII:** Addressing table of major contigs expressed in *Daceton armigerum* venom
2419 glands.

2420

2421

2422 **Chapitre III : Lien entre les toxines peptidiques et la fonction** 2423 **immunitaire**

2424
2425 Comme mentionné dans la partie 4 du chapitre 1, les insectes ne possèdent pas de
2426 molécules impliquées dans la reconnaissance du soi et du non-soi, et impliquant une mémoire
2427 à long-terme (molécules du complexe majeur d'histocompatibilité, récepteurs T, anticorps)
2428 (Helge Schlüns & Crozier, 2009). La défense de l'organisme repose donc essentiellement sur
2429 les réactions du système immunitaire inné, incluant notamment des mécanismes de
2430 phagocytose, de nodulation, d'encapsulation ou encore la production de peptides dits
2431 « antimicrobiens » (PAMs), également appelé Host Defense Peptides (HDPs) (Satyavathi et al.,
2432 2014).

2433 Ces molécules constituent l'une des premières lignes de défense de l'immunité des
2434 insectes. Ceci est d'autant plus important pour les insectes sociaux (abeilles, fourmis, termites,
2435 etc.). Le fonctionnement d'une colonie implique en effet des échanges alimentaires (e.g.
2436 trophallaxie) ainsi que des transferts de symbiontes, ce qui, corrélé à une forte densité de
2437 population couplée à une faible diversité génétique, augmente le risque d'introduction et de
2438 transmission de pathogènes (Aili et al., 2014; Hamilton & Bulmer, 2012). La prévention et/ou
2439 l'éradication de l'infection dépend (i) des comportements adoptés dans le cadre de l'immunité
2440 sociale (e.g. allogrooming, élimination des individus infectés, désinfection du nid) (Aanen,
2441 2018; Cremer, Armitage, & Schmid-Hempel, 2007), (ii) de la rapidité de l'activation des voies
2442 de signalisation impliquées dans le système immunitaire inné et (iii) de l'efficacité des
2443 molécules effectrices de l'immunité contre les pathogènes à l'échelle individuelle (HDPs,
2444 dérivés réactifs de l'oxygène, etc.) (Satyavathi et al., 2014; Helge Schlüns & Crozier, 2009).
2445 La synthèse d'HDPs est principalement effectuée *via* l'activation des voies Toll et IMD, deux
2446 voies immunitaires connues pour répondre aux infections bactériennes (Bier & Guichard,
2447 2012). Chez les insectes, cette synthèse s'effectue au niveau des hémocytes ainsi qu'au niveau
2448 d'organes particuliers tels que le corps gras et certaines parties du tube digestif (e.g. mésentéron,
2449 tubes de Malpighi). Les molécules effectrices synthétisées sont ensuite sécrétées dans
2450 l'hémolymphe.

2451 Cependant, une autre source de molécules antimicrobiennes, notamment chez les
2452 Hyménoptères, est le venin (Moreau, 2013). Bien que cette hypothèse n'ait jamais été
2453 confirmée, il est supposé que ces molécules feraient partie d'un système de défense interne aux
2454 glandes à venin afin d'éviter la contamination lors de l'envenimation ainsi que pour prévenir

2455 l'infection des proies ramenées dans la colonie avant leur consommation (Moreau, 2013; Orivel
2456 et al., 2001). Il a néanmoins été démontré que des toxines avaient évolué plusieurs fois de
2457 manière indépendante à partir des gènes de l'immunité chez les Vertébrés (Whittington et al.,
2458 2008).

2459 Ce chapitre a pour objectif d'évaluer le lien potentiel entre les peptides de venin et la
2460 fonction immunitaire individuelle des fourmis. Nous avons donc choisi de nous consacrer à
2461 l'immunité innée de la fourmi *Tetramorium bicarinatum*, un modèle d'étude historique de
2462 l'équipe BTSB dont la composition peptidique du venin a été récemment décrite (Touchard et
2463 al., 2018). Deux peptides majoritaires de ce venin, Bicarinaline et P17, ont respectivement
2464 montré des effets antimicrobiens et immunomodulateurs chez l'Homme (Benmoussa et al.,
2465 2017; Duraisamy et al., 2020; Rifflet et al., 2012; Téné et al., 2014). Les transcrits codant pour
2466 ces peptides ont de plus été retrouvés en dehors des glandes à venin, dans l'abdomen de la
2467 fourmi. Nous posons l'hypothèse que ces peptides, voire d'autres peptides du venin, seraient
2468 des effecteurs de l'immunité dite innée de la fourmi *T. bicarinatum* au niveau individuel. Si tel
2469 est le cas, les gènes les codant devraient s'exprimer dans des organes impliqués dans l'immunité
2470 innée des insectes, et être surexprimés face à une infection bactérienne.

2471 Le travail présenté dans ce chapitre vise à étudier la relation des toxines peptidiques
2472 avec le système immunitaire de *T. bicarinatum* au travers d'approches basées sur l'étude de
2473 l'expression des gènes les codant au niveau transcriptionnel (i.e. recherche des transcrits) et
2474 traductionnel (i.e. recherche des peptides matures). Ce chapitre vise deux objectifs, se déclinant
2475 en deux parties :

- 2476 - Localiser l'expression des gènes codant les toxines peptidiques en dehors du
2477 système vulnérant,
- 2478 - Étudier la réponse des gènes codant ces toxines face à un challenge bactérien.

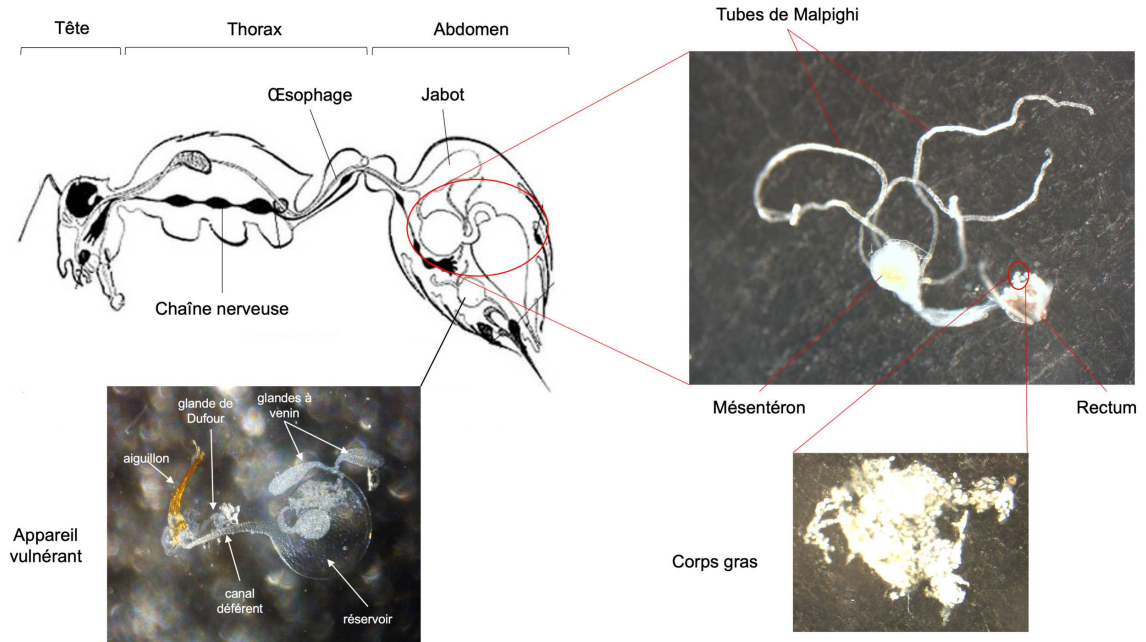
2479

2480 **1. Etude préliminaire du lien entre les peptides du venin de *Tetramorium bicarinatum* et** 2481 **l'immunité innée de la fourmi**

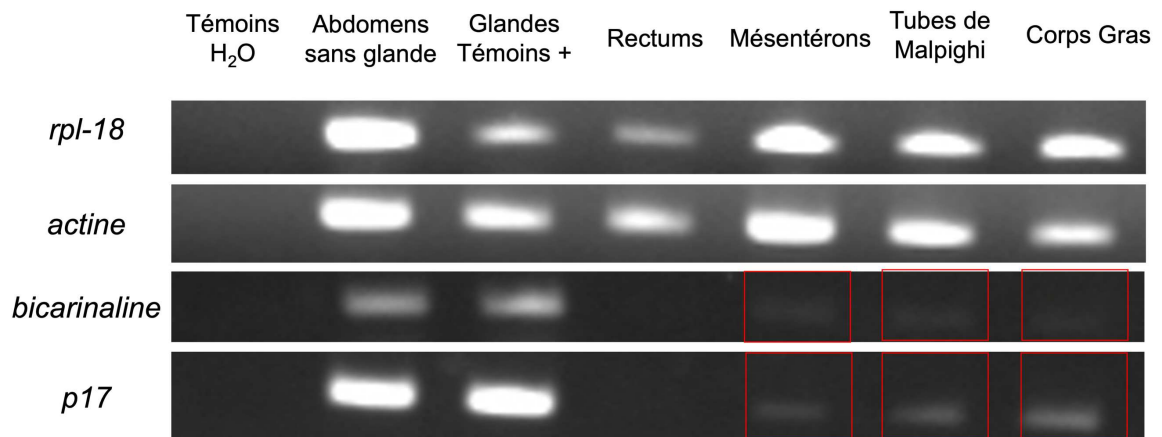
2482 Dans cette étude préliminaire, nous nous sommes attachés à vérifier l'existence d'un lien
2483 entre les peptides du venin de *T. bicarinatum* (i.e. Bicarinaline et P17) et son système
2484 immunitaire inné. Cette étude s'est déroulée lors de mon stage de Master 2 (2017).

2485 Une première analyse en RT-PCR (35 cycles) a été effectuée afin d'évaluer l'expression
2486 des gènes codant la Bicarinaline et le P17 dans différents organes placés par lots de cinq,
2487 composant l'abdomen de la fourmi (Figure 30). Cette analyse a révélé la présence de signaux

2488 d'expression dans des organes impliqués dans l'immunité innée des insectes (i.e. mésentéron,
 2489 tubes de Malpighi et corps gras) (Figure 31). La faiblesse de ces signaux peut être expliquée
 2490 par la faible quantité d'ARNs récoltés à partir d'organes isolés ainsi que par les biais dus à la
 2491 rétrotranscription et à la sensibilité de la technique de PCR classique.

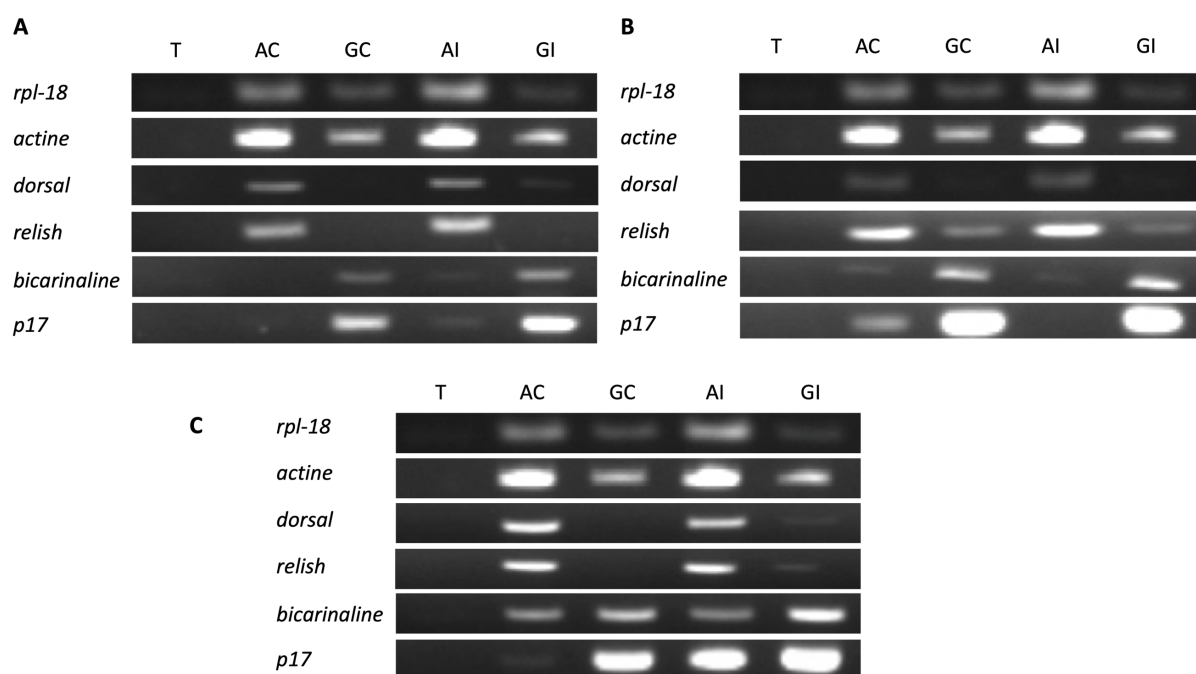


2492
 2493 **Figure 30** : Schéma général de l'anatomie interne d'une fourmi et des organes composant le tube digestif
 2494 et le système vulnérant.



2495
 2496 **Figure 31** : Expression des gènes *bicarinaline* et *p17* dans différents organes contenus dans l'abdomen de la
 2497 fourmi *Tetramorium bicarinatum*, obtenus par RT-PCR (60°C, 35 cycles). Les migrations ont été réalisées
 2498 sur des gels d'agarose 2%. Chaque échantillon est constitué d'un lot de 5 tissus. Les gènes *rpl-18* et *actine*
 2499 servent de gènes de références. Les signaux positifs d'expression dans les différents organes isolés sont
 2500 encadrés en rouge.

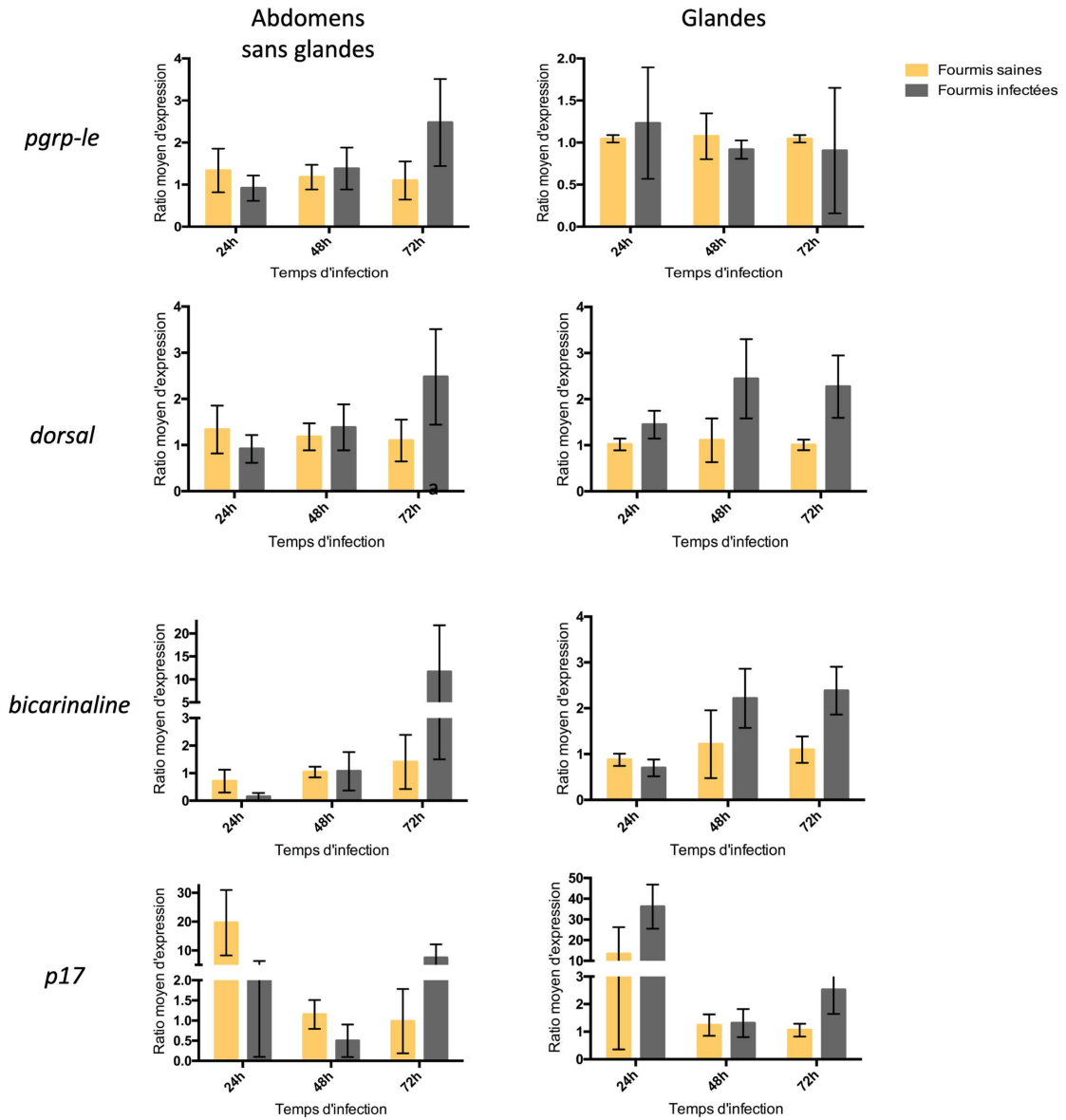
2501 La réponse des gènes codant pour ces deux myrmécitoxines, ainsi que des gènes relatifs à
 2502 l'immunité innée (i.e. *dorsal* pour la voie Toll, *relish* pour la voie IMD) a également été évaluée
 2503 suite à une infection bactérienne. L'infection a été réalisée en continue sur trois jours avec une
 2504 bactérie à Gram négatif, *Serratia marcescens*. Des extractions ARNs ont été réalisées à partir
 2505 des glandes à venin et d'abdomens dépourvus d'appareils vulnérants à trois pas de temps (i.e.
 2506 24h, 48h, 72h). Une première évaluation par RT-PCR (35 cycles) a montré la présence de
 2507 transcrits dans la majorité des échantillons de la condition « infectés », 24 et 72h après le début
 2508 de l'infection. Tout comme l'expérimentation précédente, la faiblesse de certains signaux peut
 2509 également être expliquée par la faible quantité d'ARNs récoltés ainsi que par les biais dus à la
 2510 rétrotranscription et à la sensibilité de la technique de PCR classique (Figure 32).
 2511



2512 **Figure 32** : Expression des gènes *dorsal*, *relish*, *bicarinaline* et *p17* déterminées par RT-PCR (60°C, 35
 2513 cycles) dans les abdomens sans glande à venin, et les glandes à venin de *Tetramorium bicarinatum*, à
 2514 différents pas de temps après le début de l'infection par *Serratia marcescens* : 24h (A), 48h (B) et 72h (C).
 2515 Les migrations ont été réalisées sur des gels d'agarose 2%. Chaque échantillon est constitué d'un lot de 3
 2516 tissus. Les gènes *rpl-18* et *actine* servaient de gènes de référence. T : Témoin négatif d'amplification PCR
 2517 (H₂O) ; AC : abdomens sans glande ni réservoir de la condition contrôle ; GC : glandes à venin de la
 2518 condition contrôle ; AI : abdomens sans glande ni réservoir de la condition infecté ; GI : glandes à venin de
 2519 la condition infecté.
 2520

2521 Une seconde évaluation par qPCR a été réalisée sur ces mêmes échantillons. Une tendance
 2522 à la surexpression a été observée pour l'ensemble des gènes, excepté *bicarinaline*, dans les
 2523 échantillons de glandes à venin dès 24h après le début de l'infection. Une tendance à la

2524 surexpression a également été observée dans les abdomens dépourvus d'appareils vulnérant au
 2525 pas de temps « 72h ». Les échantillons ont présenté néanmoins des ratios d'expression très
 2526 variables les uns par rapport aux autres pour un même gène (Figure 33).



2527
 2528 **Figure 33** : Expression des gènes *pgrp-le* (IMD), *dorsal* (Toll), *bicarinaline*, et *p17* déterminées par qPCR
 2529 chez des fourmis *Tetramorium bicarinatum* saines et infectées par *Serratia marcescens*. Les variations
 2530 d'expression sont données par les ratios moyens d'expression \pm SE (Erreur Standard) calculés par la méthode
 2531 du $2^{-\Delta\Delta C_T}$ (Livak & Schmittgen, 2001). Trois échantillons ont été réalisés par pas de temps et par condition.
 2532 Chaque échantillon est constitué d'un lot de 3 tissus.

2533 L'absence de surexpression à 24 h dans les abdomens sans appareil vulnérant peut être
 2534 expliquée par le contenu de l'échantillon ainsi que par le mode d'infection. En effet, lors d'une

2535 infection par voie orale, le pathogène doit traverser la barrière épithéliale avant de se retrouver
2536 dans l'hémolymphe et de déclencher une réponse immunitaire systémique (Nehme et al., 2007).
2537 Ceci a notamment été étayé par Nehme et ses collaborateurs (2007) chez *Drosophila*
2538 *melanogaster* lors d'une infection par *Serratia marcescens*. Une réponse locale par activation
2539 de la transcription du gène codant pour la Diptéricine a été obtenue dans les 24 h suivant une
2540 infection par voie orale mais une réponse systémique n'a pas été mise en évidence dans ce laps
2541 de temps (Nehme et al., 2007). Il est donc nécessaire de décomposer les échantillons
2542 d'abdomens sans glande afin d'étudier les réponses au niveau local, directement sur les organes
2543 composant le tube digestif.

2544 Les activations de transcriptions semblent également s'effectuer par vagues temps- et tissus-
2545 dépendantes comme cela a déjà été démontré pour les activités transcriptionnelles des facteurs
2546 de transcription κ B chez *D. melanogaster* (Tanenhaus, Zhang, & Yin, 2012). Ceci pourrait en
2547 partie expliquer les variations d'expression d'un échantillon à l'autre, mais il est néanmoins
2548 nécessaire de réduire le nombre d'individus par lot tout en augmentant le nombre d'échantillons
2549 pour tenter de pallier à cette variabilité.

2550 Bien que seules des tendances aient pu être dégagées, ces dernières, couplées à la présence
2551 de transcrits codant pour la Bicarinaline et P17 dans des organes impliqués dans l'immunité des
2552 insectes, fournissent des perspectives intéressantes quant au lien entre les toxines de venin et
2553 l'immunité innée individuelle de la fourmi. Au vu des variations d'expression observées pour
2554 les gènes *bicarinaline*, *p17*, *dorsal* et *pgrp-le*, il est en effet probable qu'il y ait une connexion
2555 entre ces différents acteurs.

2556

2557 **2. Localisation des peptides en dehors des appareils vulnérants**

2558 Dans cette partie, nous nous sommes attachés à évaluer l'expression de gènes codant des
2559 peptides de venin dans des organes impliqués dans l'immunité innée des insectes. Les organes
2560 sélectionnés à ce titre pour notre étude sont les corps gras et différentes parties du tube digestif
2561 (i.e. mésentéron, tubes de Malpighi, rectum). Il s'agit dans un premier temps de confirmer la
2562 localisation de la Bicarinaline et du P17 dans les différents tissus composant l'abdomen par
2563 RT-PCR. L'étude a ensuite été étendue à sept autres toxines peptidiques appartenant aux
2564 superfamilles de précurseurs A et B (i.e. U₃-MYRTX-Tb1a, U₄-MYRTX-Tb1a, U₅-MYRTX-
2565 Tb1a, U₇-MYRTX-Tb1a, U₈-MYRTX-Tb1a, U₉-MYRTX-Tb1a et U₁₆-MYRTX-Tb1a), dont
2566 les séquences signal et propeptide présentent 20 à 40% d'identité de séquences avec celles de
2567 HDPs de la drosophile (i.e. drosomycine, cécropine, metchnikowin) (Touchard et al., 2018). La

2568 seconde partie de cette étude consiste à vérifier la présence effective de la Bicarinaline et du
2569 P17 en dehors des appareils vulnérants, au moyen d'une immunohistochimie indirecte.

2570

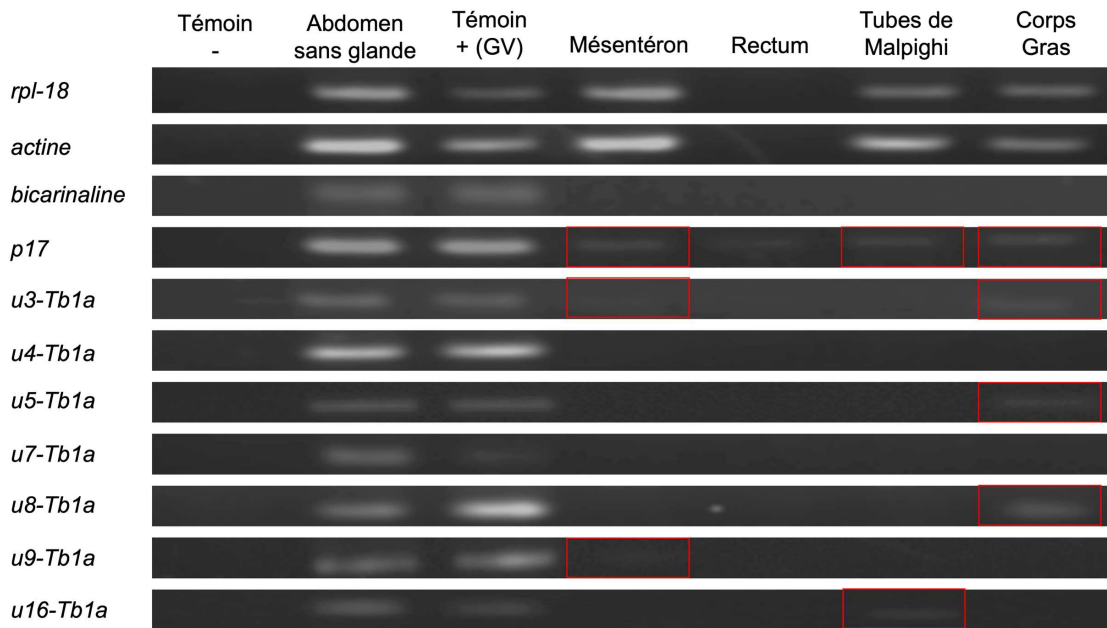
2571 *2.1. Localisation des transcrits codant certains peptides de venin dans l'abdomen de T.*
2572 *bicarinatum*

2573 Cette analyse a été réalisée sur six types de tissus : des glandes à venin, des abdomens
2574 dépourvus d'appareil vulnérant, des mésentérons, des tubes de Malpighi, des rectums et des
2575 corps gras. Chaque échantillon contenait un lot de 5 organes et cette expérimentation a été
2576 répétée 3 fois.

2577 L'électrophorèse des produits PCR a montré que les deux gènes de référence, *rpl-18* et
2578 *actine*, sont exprimés dans tous les organes à l'exception des rectums. Cet organe n'a donc pas
2579 été pris en compte dans la suite de l'étude. Les signaux d'expression visibles pour chaque gène
2580 dans les abdomens dépourvus d'appareil vulnérant confirment la présence de transcrits codant
2581 pour les toxines testées en dehors des glandes à venin (Figure 34).

2582 Des signaux d'expression ont notamment été détectés pour *p17* dans tous les organes
2583 testés. Les gènes *u3-Tb1a* et *u9-Tb1a* étaient exprimés dans les mésentérons isolés. Des signaux
2584 d'expression ont été détectés dans les tubes de Malpighi pour le gène *u16-Tb1a*, ainsi que dans
2585 les corps gras pour les gènes *u3-Tb1a*, *u5-Tb1a* et *u8-Tb1a*.

2586 Les signaux observés sont néanmoins légers et reflètent donc une faible quantité d'ADNc
2587 de départ. Ceci peut également expliquer l'absence de signaux pour les gènes *bicarinaline*, *u4-*
2588 *Tb1a* et *u7-Tb1a* dans les organes isolés. Le recours à des techniques d'analyses
2589 transcriptomiques plus sensibles a donc semblé nécessaire.



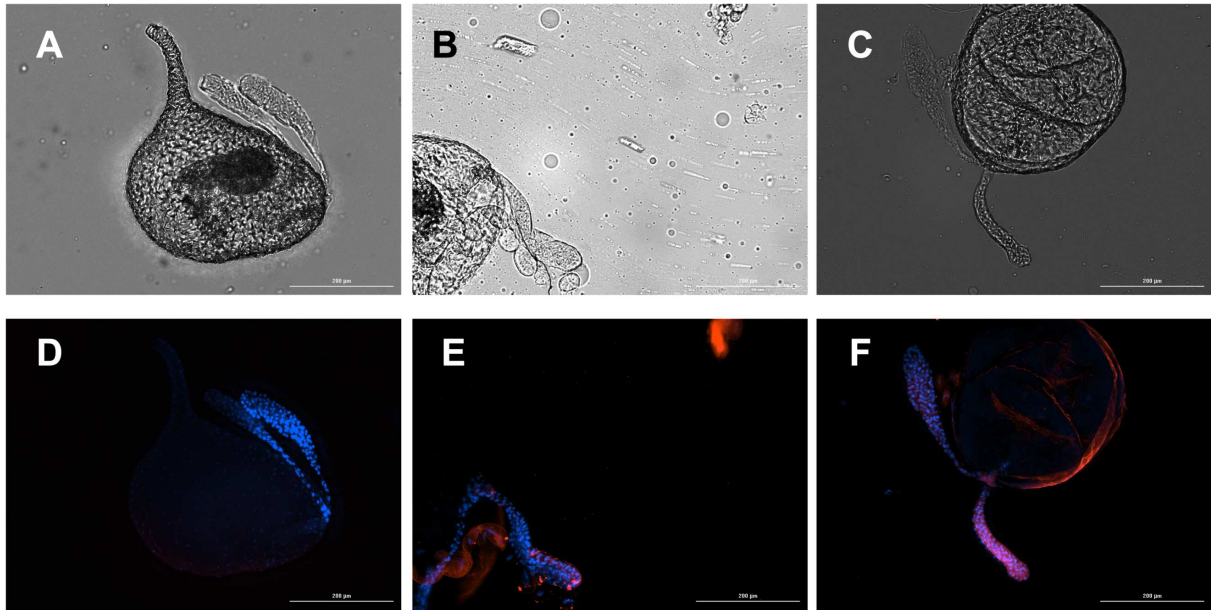
2590

2591 **Figure 34** : Expressions des gènes *rpl-18*, *actine*, *bicarinaline*, *p17*, *u3-Tb1a*, *u4-Tb1a*, *u5-Tb1a*, *u7-*
2592 *Tb1a*, *u8-Tb1a*, *u9-Tb1a* et *u16-Tb1a* déterminées par RT-PCR (60°C, 40 cycles) dans les organes de
2593 *Tetramorium bicarinatum*. Les migrations ont été réalisées sur des gels d'agarose 2%. Chaque
2594 échantillon est composé d'un lot de 5 tissus. Les signaux positifs sont encadrés en rouge.

2595 2.2. Localisation de la Bicarinaline et du P17 dans l'abdomen de *T. bicarinatum*

2596 Afin de vérifier la présence effective de la Bicarinaline et du P17 en dehors des glandes à
2597 venin, un protocole d'immunohistochimie indirecte a été appliqué sur les tissus composant
2598 l'abdomen. Elle permet de localiser les protéines au niveau cellulaire et fait intervenir deux
2599 types d'anticorps : des anticorps primaires, dirigés contre la ou les protéines ciblées, et des
2600 anticorps secondaires couplés à un fluorochrome et dirigés contre les anticorps primaires. Le
2601 fait d'utiliser ces deux sortes d'anticorps offre une meilleure sensibilité. Le protocole de
2602 production choisi a permis d'obtenir des anticorps anti-Bicarinaline et anti-P17 spécifiques des
2603 parties N-terminales, ceci afin de ne révéler les peptides qu'après leur maturation. Des signaux
2604 intracellulaires, localisés majoritairement dans des vésicules de sécrétion étaient donc attendus.
2605 Les anticorps secondaires étaient couplés au Texas Red.

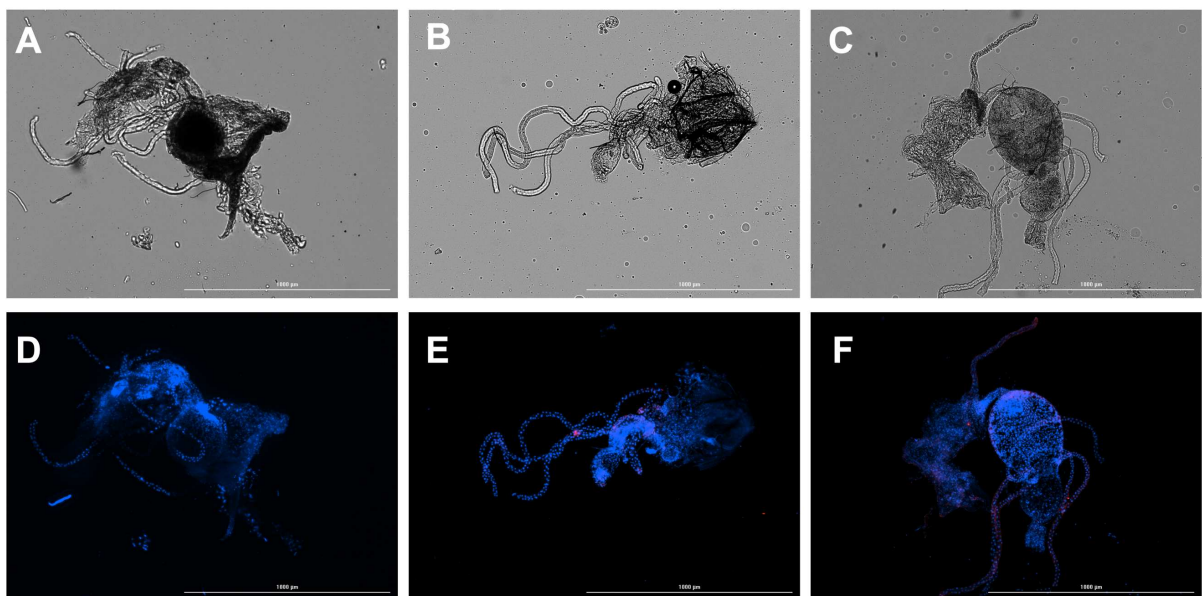
2606 Des signaux positifs (i.e. point intracellulaires) ont été détectés pour les deux peptides dans
2607 les glandes à venin, confirmant la validité des anticorps utilisés (Figure 35 - E et F). Il serait
2608 néanmoins nécessaire de préciser l'allure de ces signaux avec un objectif X40. Ces résultats
2609 nous ont permis d'effectuer les réglages de luminosité nécessaires pour la distinction des deux
2610 peptides dans les autres organes de l'abdomen de la fourmi *T. bicarinatum*, notamment par
2611 rapport à un potentiel bruit de fond.



2612
 2613 **Figure 35** : Localisation de la Bicarinaline et du P17 dans les glandes à venin de *Tetramorium bicarinatum*
 2614 réalisée par immunohistochimie indirecte (grossissement X10). Les anticorps secondaires sont couplés au
 2615 Texas Red. Les photos A, B et C ont été prises en lumière blanche. Les photos D, E et F montrent les
 2616 superpositions des signaux DAPI (noyaux cellulaires) et Texas Red (Témoin anticorps secondaire (D),
 2617 Bicarinaline (E) et P17 (F)).

2618 Bien que les signaux observés soient de manière générale plus diffus que dans les glandes
 2619 à venin, la présence des deux peptides a été notée dans les tubes digestifs (Figure 36). Des
 2620 signaux positifs sont visibles pour la Bicarinaline au niveau de l'intestin inférieur (Figure 36 -
 2621 E) tandis que le P17 semble être réparti de l'œsophage au mésentéron (Figure 36 - F).

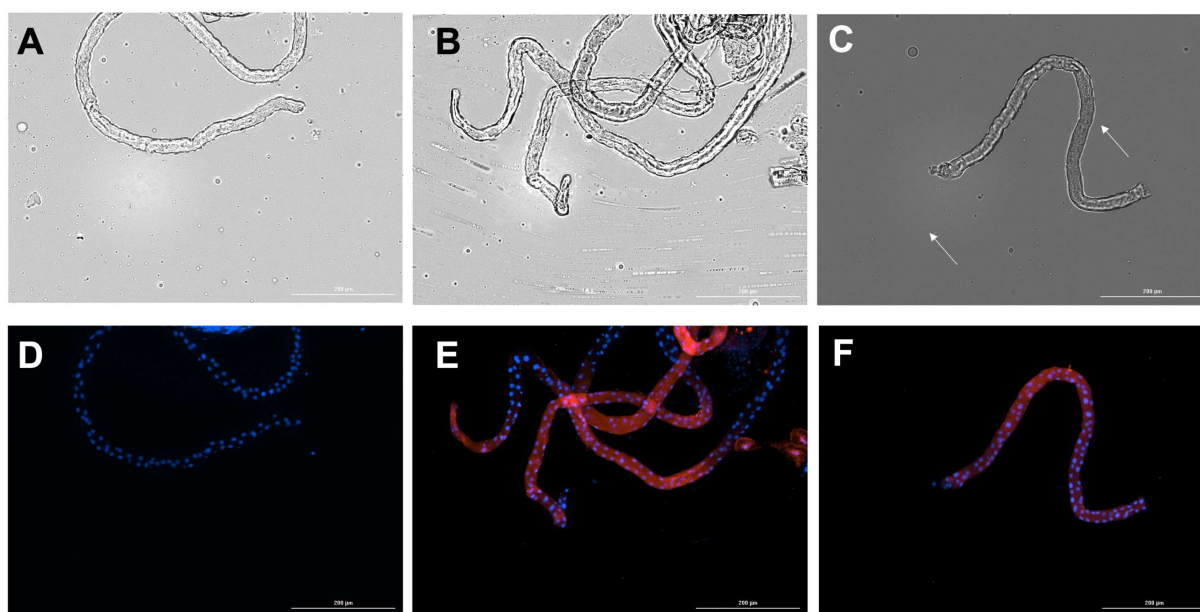
2622



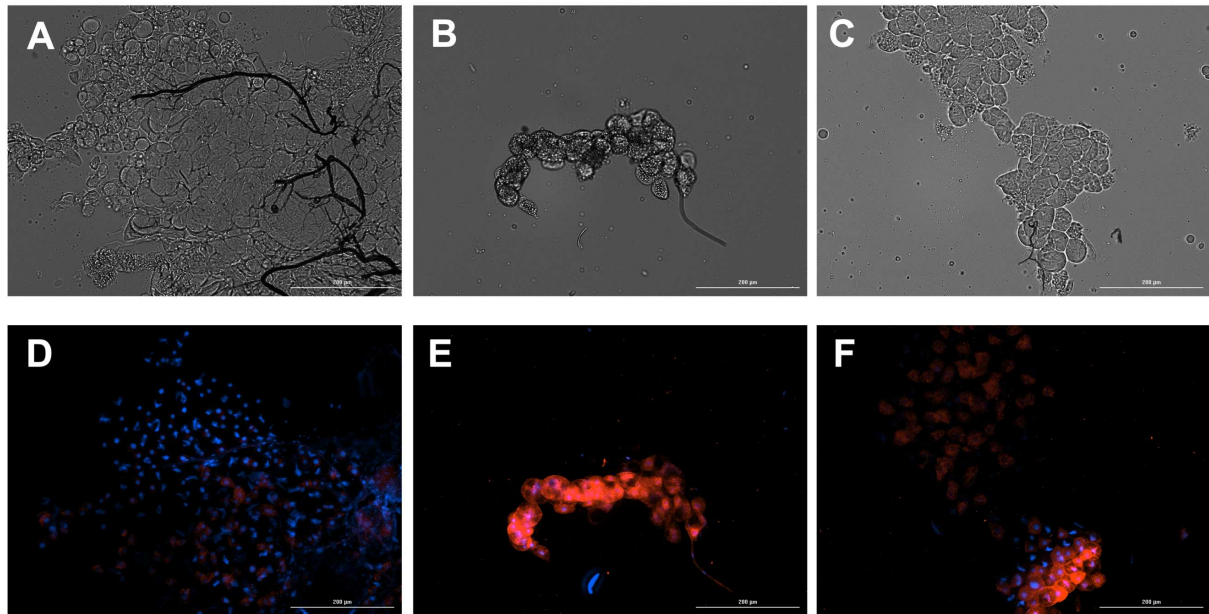
2623
 2624 **Figure 36** : Localisation de la Bicarinaline et du P17 dans les tubes digestifs de *Tetramorium bicarinatum*
 2625 réalisée par immunohistochimie indirecte (grossissement X4). Les anticorps secondaires sont couplés au

2626 Texas Red. Les photos A, B et C ont été prises en lumière blanche. Les photos D, E et F montrent les
2627 superpositions des signaux DAPI (noyaux cellulaires) et Texas Red (Témoin anticorps secondaire (D),
2628 Bicarinaline (E) et P17 (F)).

2629 De la même manière, les deux peptides semblent également présents au niveau des tubes de
2630 Malpighi (Figure 37 - E et F). Il serait néanmoins nécessaire de préciser l'allure de ces signaux
2631 à un grossissement plus important (X40) pour ces organes. En revanche, de forts signaux sont
2632 présents dans certaines cellules du corps gras (Figure 38 - E et F). Les gènes codant pour la
2633 Bicarinaline et P17 sont donc transcrits et produits dans ces organes.



2634
2635 **Figure 37** : Localisation de la Bicarinaline et du P17 dans les tubes de Malpighi de *Tetramorium bicarinatum*
2636 réalisée par immunohistochimie indirecte (grossissement X10). Les anticorps secondaires sont couplés au
2637 Texas Red. Les photos A, B et C ont été prises en lumière blanche. Les photos D, E et F montrent les
2638 superpositions des signaux DAPI (noyaux cellulaires) et Texas Red (Témoin anticorps secondaire (D),
2639 Bicarinaline (E) et P17 (F)).

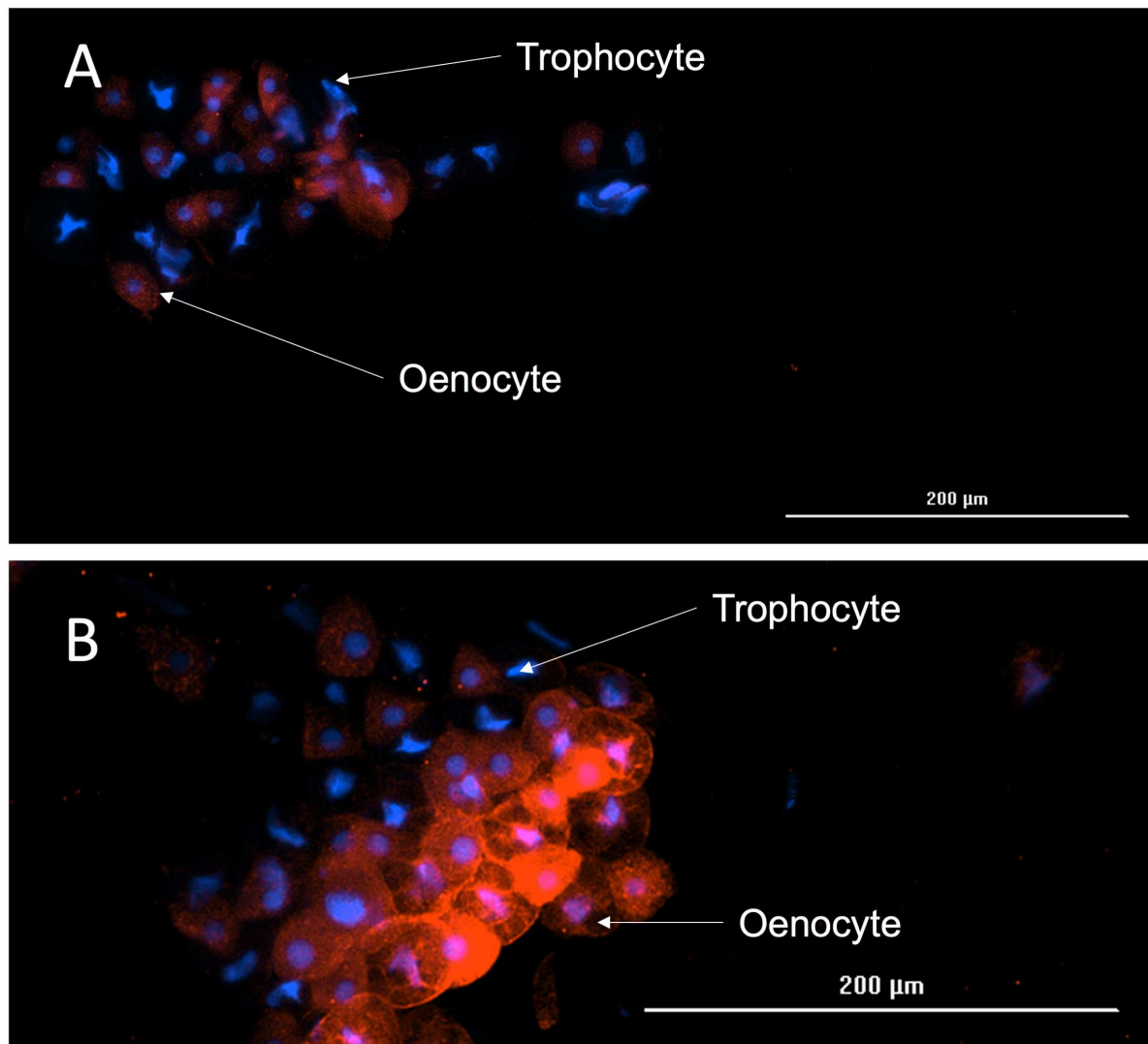


2640

2641 **Figure 38** : Localisation de la Bicarinaline et du P17 dans les corps gras de *Tetramorium bicarinatum* réalisée
 2642 par immunohistochimie indirecte (grossissement X10). Les photos A, B et C ont été prises en lumière
 2643 blanche. Les photos D, E et F montrent les superpositions des signaux DAPI (noyaux cellulaires) et Texas
 2644 Red (Témoin anticorps secondaire (D), Bicarinaline (E) et P17 (F)).

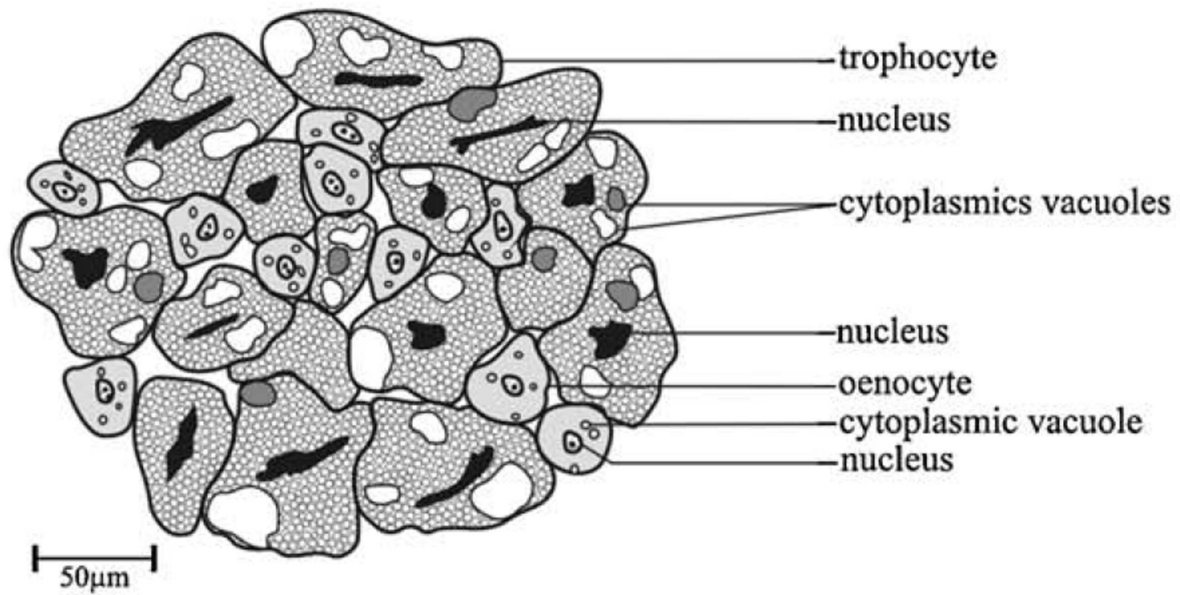
2645 Deux types cellulaires ont été distingués dans les corps gras de la fourmi *T.*
 2646 *bicarinatum* selon la forme des noyaux : les trophocytes avec des noyaux en demi-lune et les
 2647 oenocytes présentant des noyaux circulaires (Roma, Mathias, & Bueno, 2006). Les signaux
 2648 relatifs aux deux peptides testés sont systématiquement observés au niveau des oenocytes
 2649 (Figure 39).

2650



2651
 2652 **Figure 39** : Localisation de la Bicarinaline et du P17 dans les corps gras de *Tetramorium bicarinatum* réalisée
 2653 par immunohistochimie indirecte (grossissement X10). Les anticorps secondaires étaient couplés au Texas
 2654 Red. Les photos montrent les superpositions des signaux DAPI (noyaux cellulaires) et Texas Red
 2655 (Bicarinaline (A) et P17 (B)).

2656 Ces deux types cellulaires ont en effet déjà été décrits dans les corps gras des insectes, en
 2657 association avec d'autres, et sont les seuls retrouvés dans les corps gras de fourmis (Roma et
 2658 al., 2006). Les trophocytes sont d'origine mésodermique tandis que les oenocytes sont d'origine
 2659 ectodermique. Chez les fourmis myrmicines de la tribu des Attini, les oenocytes sont distribués
 2660 parmi les trophocytes (Roma, Bueno, & Camargo-Mathias, 2010) (Figure 40). Chez la fourmi
 2661 *T. bicarinatum*, deux types de distribution des oenocytes ont été observés : des cellules
 2662 relativement isolées parmi les trophocytes (Figure 39 – A), ou des amas cellulaires distincts
 2663 (Figure 39 – B). Il serait néanmoins nécessaire de confirmer la distinction de ces types
 2664 cellulaires chez *T. bicarinatum* en histologie.



2666

2667 **Figure 40** : Cellules du corps gras de la fourmi *Mycetarotes parallelus* (*Myrmicinae* : Attini). Tiré de (Roma
2668 et al., 2010).

2669 De manière générale, les oenocytes sont connus chez les insectes pour leur rôle dans le
2670 métabolisme des lipides, la détoxification et pour l'absorption de substances circulant dans
2671 l'hémolymphe. Ils participent également à la synthèse de composés cuticulaires et à l'immunité
2672 innée *via* la synthèse de lysozymes, bien que cette dernière fonction n'ait été jusqu'à présent
2673 observée que chez le moustique *Aedes aegypti* (Martins & Ramalho-Ortigão, 2012; Martins et
2674 al., 2011; Roma et al., 2010). Leur distribution dans le corps gras varie selon les modèles
2675 biologiques, tout comme leur profil d'expression protéique qui reste à ce jour peu caractérisé
2676 (Martins & Ramalho-Ortigão, 2012; Roma et al., 2010).

2677 L'ensemble des fonctions physiologiques des cellules du corps gras n'a de plus pas été
2678 décrit précisément chez les Formicidae. Les premières études cytologiques chez des fourmis de
2679 la tribu des Attini (*Myrmicinae*) ont néanmoins montré que les trophocytes possèdent toute la
2680 machinerie cellulaire nécessaire à la synthèse et au stockage des protéines (i.e. réticulum
2681 endoplasmique rugueux lamellaire et vésiculaire, appareil de Golgi, mitochondries, granules et
2682 système vacuolaire). La présence de protéines ainsi que celle de réticulum endoplasmique
2683 rugueux lamellaires et vésiculaires a également été montrée dans des oenocytes d'ouvrières
2684 *Cyphomyrmex rimosus* (*Myrmicinae*) (Roma et al., 2010, 2006).

2685 La présence de transcrits codant pour la Bicarinaline et le P17 dans le corps gras, couplée à
2686 leur détection dans les oenocytes, laissent supposer que ces cellules sont impliquées dans la
2687 production et le stockage constitutifs de ces peptides. Il serait néanmoins nécessaire de vérifier

2688 que les gènes codant pour ces peptides sont exprimés de manière spécifique au sein de ce type
2689 cellulaire pour l'affirmer. Pour ceci, une technique telle que la qPCR après isolement de ces
2690 cellules pourrait être envisagée.

2691

2692 **3. Réponse des gènes codant les peptides à une infection bactérienne**

2693 Les gènes codant pour des toxines du venin de *T. bicarinatum* sont exprimés dans des
2694 organes impliqués dans l'immunité innée des insectes (i.e. corps gras, tubes de Malpighi,
2695 mésentéron). La réponse de ces gènes a été évaluée suite à une infection orale par la bactérie
2696 *Serratia marcescens* (Gram négatif). Au niveau transcriptomique, la réponse a été analysée au
2697 moyen de la RT-qPCR et de la ddPCR, techniques plus précises que la RT-PCR. La réponse au
2698 niveau protéique a également été estimée *via* une immunohistochimie indirecte.

2699 Des expérimentations faisant suite à l'étude préliminaire nous ont permis de définir la
2700 méthode d'extraction ARN optimale (Cf. partie 4.4.2 de ce chapitre), ainsi que le nombre
2701 minimal d'organes requise pour l'obtention de signaux corrects par RT-qPCR. Les premiers
2702 résultats nous ont montré qu'il était possible de réduire le nombre d'abdomens par échantillon,
2703 mais pas le nombre d'organes isolés. Ces mêmes expérimentations nous ont également permis
2704 de définir que le pas de temps le plus pertinent pour l'étude de la réponse des gènes testés était
2705 24h après l'ingestion de l'agent pathogène.

2706 Quatre types de tissus ont donc été isolés 24h après le début de l'infection : des glandes à
2707 venin, des abdomens dépourvus d'appareil vulnérant, des corps gras et des tubes digestifs (i.e.
2708 mésentérons et tubes de Malpighi inclus).

2709 Les glandes à venin ont servi de témoins positifs pour les gènes codant les peptides. Elles
2710 ont été placées par lots de 5. Un total de deux échantillons a été constitué par condition (i.e.
2711 contrôle et infecté). Les abdomens dépourvus d'appareil vulnérant ont été placés par lots de 2
2712 tandis que les organes isolés (i.e. corps gras et tubes digestifs) ont été placés par lots de 5. Un
2713 total de vingt échantillons a été constitué par condition pour ces types de tissus.

2714

2715 *3.1. Évaluation au niveau transcriptomique par la technique de qPCR*

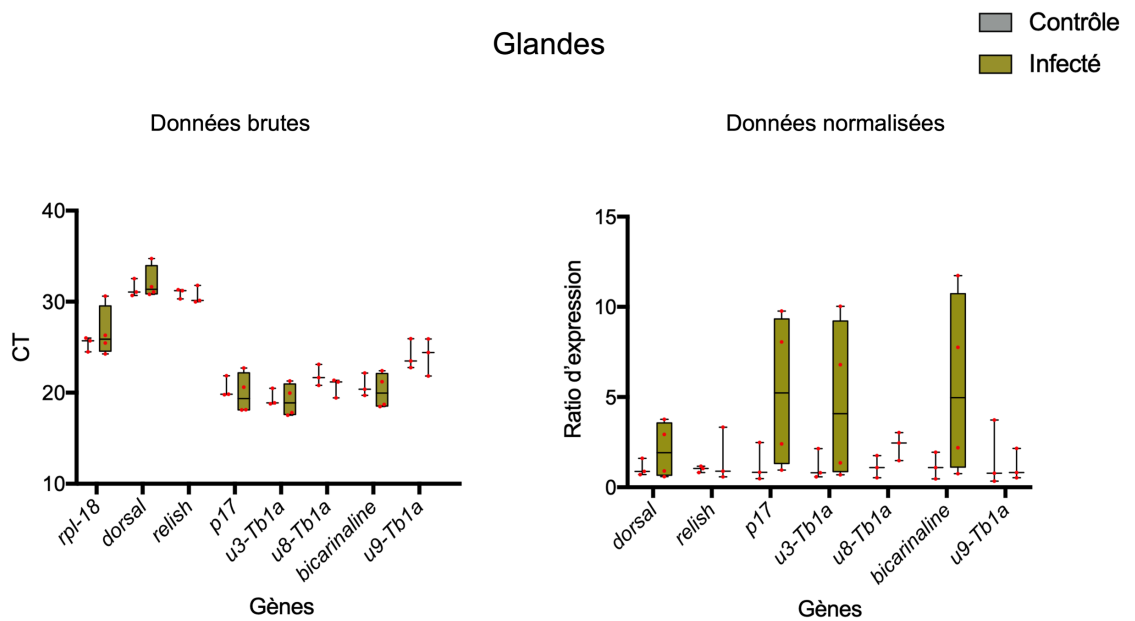
2716 La qPCR est plus précise et plus sensible que la PCR classique suivie d'une révélation
2717 sur gel d'agarose, et détermine le nombre de cycles nécessaires avant le début de la phase
2718 exponentielle de l'amplification (i.e. C_T , cycle threshold). Les échantillons ont été passés en
2719 triplicatas pour chaque gène testé. Ces gènes incluent un gène de référence (*rpl-18*), 2 gènes
2720 codant pour des facteurs NF- κ B de deux voies immunitaires des insectes réagissant aux

2721 infections bactériennes (*dorsal* pour la voie Toll et *relish* pour la voie IMD), et 5 gènes codant
2722 pour des peptides de venin (*bicarinaline*, *p17*, *u3-Tb1a*, *u8-Tb1a* et *u9-Tb1a*). Les gènes *u4-*
2723 *Tb1a*, *u5-Tb1a*, *u7-Tb1a* et *u16-Tb1a* testés précédemment ont été retirés de cette étude, les
2724 couples d’amorces n’étant pas suffisamment efficaces pour les techniques utilisées. Les
2725 séquences des transcrits codant les peptides de venin sont en effet très conservées, rendant le
2726 dessin d’amorces spécifiques particulièrement difficile.

2727 Les ratios d’expression de nos gènes d’intérêt ont été calculés grâce à la méthode de
2728 quantification relative du $2^{-\Delta\Delta CT}$ par rapport au gène de référence *rpl-18* (Livak & Schmittgen,
2729 2001).

2730 Les gènes codant pour les peptides du venin de *T. bicarinatum* sont exprimés dans les
2731 glandes à venin, avec des C_T moyens compris entre 17 et 26. Ce résultat confirme nos études
2732 des transcriptomes des glandes à venin dans lesquels les transcrits codant les peptides de venin
2733 possèdent généralement les niveaux d’expression les plus élevés. En revanche, des signaux
2734 d’expression moins élevés ont été détectés pour les gènes *dorsal* et *relish*, avec des C_T moyens
2735 compris entre 30 et 34. De manière générale, les signaux détectés pour un même gène ont
2736 présenté de grandes variabilités entre les échantillons. Le calcul des ratios d’expression a
2737 néanmoins été effectué à titre indicatif, et a montré que tous les gènes, exceptés *relish* et *u9-*
2738 *Tb1a* ont tendance à être surexprimé dans les glandes à venin des individus infectés (Figure
2739 41).

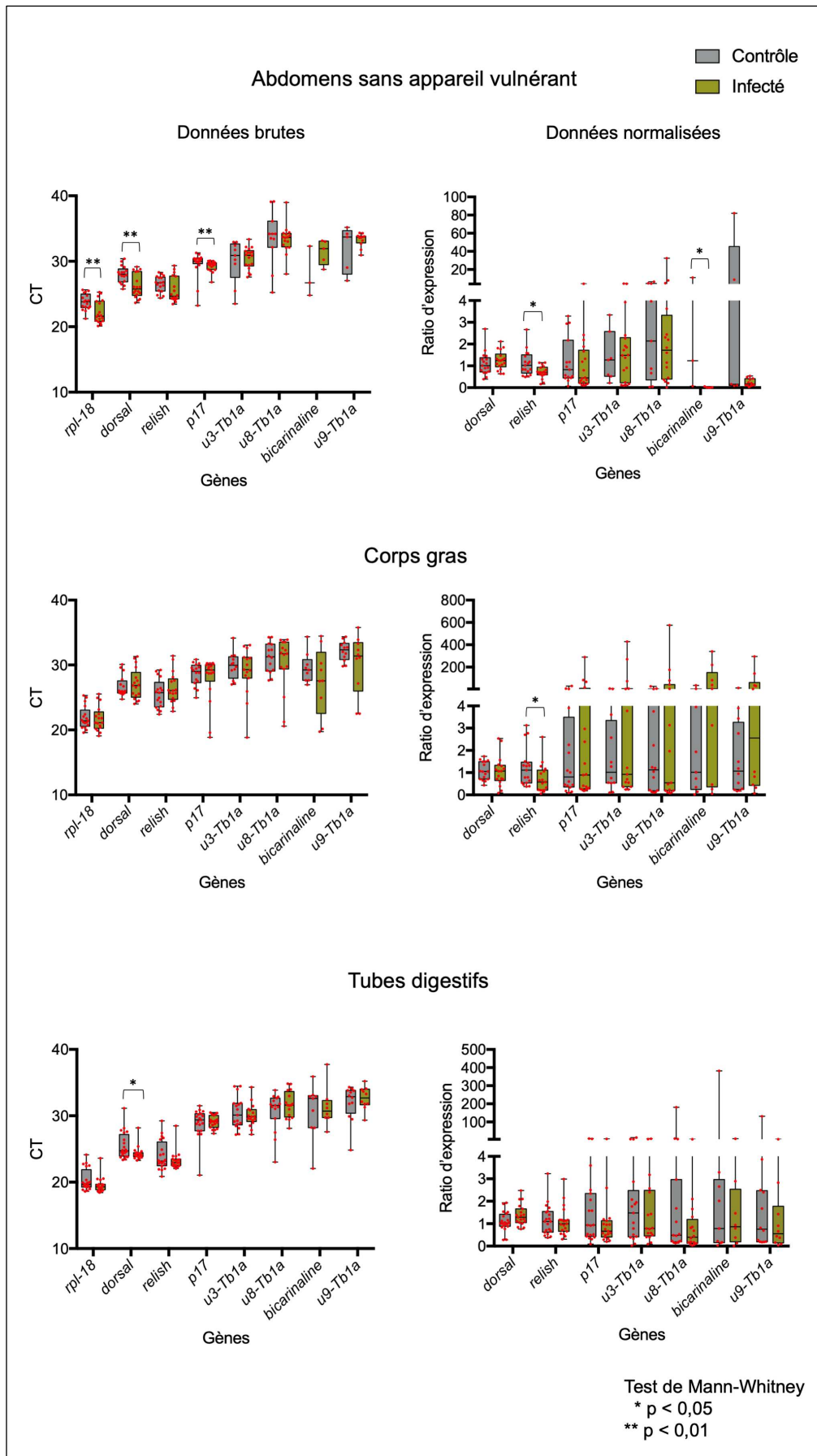
2740



2741
 2742 **Figure 41** : Répartition des C_T moyens et ratio d'expression ($2^{-\Delta\Delta C_T}$) obtenus par qPCR dans les glandes à
 2743 venin de *Tetramorium bicarinatum*. Chaque point rouge représente un échantillon (n = 4/condition).

2744

2745 Les C_T moyens des gènes codant les toxines sont généralement moins élevés dans les
 2746 tissus autres que les glandes à venin, allant de 25 pour le plus exprimé (i.e. *p17*), à 33 pour le
 2747 moins exprimé (i.e. *u8-Tb1a*). Les expressions des gènes codant les toxines, excepté *p17*, ont
 2748 d'ailleurs été détectées dans peu d'échantillons d'abdomens sains comparativement aux
 2749 individus infectés (Figure 42).



2751 **Figure 42** : Répartition des C_T moyens et ratio d'expression ($2^{-\Delta\Delta C_T}$) obtenus par qPCR dans les tissus de
 2752 *Tetramorium bicarinatum* saines ou infectées par *Serratia marcescens*. Chaque point rouge représente un
 2753 échantillon (n = 20/condition).

2754 Les analyses des C_T moyens entre les conditions ont montré que les abdomens des
 2755 individus infectés présentent des C_T moyens significativement moins élevés pour les gènes
 2756 *dorsal*, *p17* et *rpl-18* (Mann-Whitney, $P < 0,01$). Cette tendance a également été notée pour le
 2757 gène *relish* (Tableau IV, Figure 42). Dans les tubes digestifs, seul le gène *dorsal* a montré des
 2758 C_T moyens significativement moins élevés chez les individus infectés (Mann-Whitney, $P <$
 2759 $0,05$). Aucune différence significative n'a été notée entre les C_T moyens des deux conditions
 2760 pour les gènes codant les toxines dans les corps gras et les tubes digestifs (Tableau IV, Figure
 2761 42).

2762
 2763 **Tableau IV** : Valeurs de rangs (W) et probabilités (P) obtenus suite aux tests de Mann-Whitney exécutés
 2764 entre les valeurs des C_T moyens des gènes *rpl-18*, *dorsal*, *relish*, *p17*, *u3-Tb1a*, *u8-Tb1a*, *bicarinaline* et *u9-*
 2765 *Tb1a*. Les C_T ont été déterminés par RT-qPCR chez des *Tetramorium bicarinatum* saines ou infectées par
 2766 *Serratia marcescens*.

Gènes/Organes	Abdomens		Corps gras		Tubes digestifs	
	W	P	W	P	W	P
<i>rpl-18</i>	72	0,007	135	0,760	101	0,323
<i>dorsal</i>	72	0,007	139	0,929	119	0,047
<i>relish</i>	96	0,062	168	0,433	170	0,588
<i>p17</i>	56	0,002	139	0,929	154	0,452
<i>u3-Tb1a</i>	38	0,141	82	0,719	136	0,786
<i>u8-Tb1a</i>	43	0,108	105	1	138	0,495
<i>bicarinaline</i>	12	0,250	31	0,436	31	0,673
<i>u9-Tb1a</i>	32	0,879	47	0,418	62	0,923

2767
 2768 Les ratios d'expression présentent également une grande variabilité entre les échantillons
 2769 d'une même condition pour chaque type de tissu (Figure 8). Ils ont néanmoins montré que le
 2770 gène *relish* est significativement sous-exprimé dans les abdomens et les corps gras des individus
 2771 infectés (Mann-Whitney, $P < 0,05$). Le gène *bicarinaline* est également sous-exprimé de
 2772 manière significative dans les abdomens des fourmis infectées (Mann-Whitney, $P < 0,05$). Une
 2773 tendance à la surexpression a aussi été notée pour le gène *dorsal* dans les tubes digestifs des
 2774 individus infectés (Tableau V, Figure 42). Les gènes codant les toxines n'ont montré aucune

2775 différence d'expression significative entre les deux conditions dans les corps gras et les tubes
 2776 digestifs (Tableau V, Figure 42).

2777
 2778 **Tableau V** : Valeurs de rangs (W) et probabilités (P) obtenus suite aux tests de Mann-Whitney exécutés
 2779 entre les ratio d'expression (i.e. $2^{-\Delta\Delta C_T}$) des gènes de l'immunité (*dorsal* et *relish*) et codant pour les toxines
 2780 du venin de *Tetramorium bicarinatum* (*p17*, *u3-Tb1a*, *u8-Tb1a*, *bicarinaline* et *u9-Tb1a*). Les ratios
 2781 d'expression ont été déterminés par RT-qPCR chez des fourmis saines ou infectées par *Serratia marcescens*.

Gènes/Organes	Abdomens		Corps gras		Tubes digestifs	
	W	P	W	P	W	P
<i>dorsal</i>	188	0,258	126	0,736	251	0,089
<i>relish</i>	85	0,025	81	0,029	181	0,813
<i>p17</i>	106	0,126	140	0,901	173	0,478
<i>u3-Tb1a</i>	68	0,528	126	0,830	153	0,791
<i>u8-Tb1a</i>	70	0,394	109	0,881	92	0,281
<i>bicarinaline</i>	0	0,036	56	0,190	35	0,963
<i>u9-Tb1a</i>	24	0,574	78	0,254	50	0,539

2782
 2783 Au vu des différences observées dans les valeurs de C_T entre conditions pour les gènes
 2784 *dorsal* et *p17*, nous aurions dû observer des surexpressions pour ces gènes dans les abdomens
 2785 infectés grâce aux calculs des $2^{-\Delta\Delta C_T}$. Cette non-concordance entre ratio d'expression et valeurs
 2786 brutes des C_T dans les abdomens peut être expliquée par la variation des C_T du gène *rpl-18* entre
 2787 les deux conditions.

2788 Pour nos données, le gène *rpl-18* a été considéré comme étant le seul gène stable entre
 2789 les conditions par le logiciel Ref-Finder (Xie, Xiao, Chen, Xu, & Zhang, 2012). Il a également
 2790 été validé en tant que gène de référence chez la fourmi *Solenopsis invicta*, grâce à sa stabilité
 2791 qu'elle que soit la caste, le stade de développement et le tissu testé (Cheng, Zhang, He, & Liang,
 2792 2013). Cependant, les variations observées dans nos échantillons rendent ininterprétables les
 2793 ratios d'expression calculés pour les abdomens. Ceci remet également en question les ratios
 2794 d'expression observés dans les autres tissus et souligne la nécessité de s'affranchir d'un gène
 2795 de référence pour analyser nos données. Les C_T élevés observés pour les gènes codant les
 2796 toxines dans les tissus autres que les glandes à venin remettent également en question la
 2797 technique utilisée.

2798 Ces résultats suggèrent la nécessité d'utiliser une technique encore plus sensible que la
2799 qPCR, capable de détecter de faibles variations d'expression génétique en quantification
2800 absolue.

2801

2802 3.2. Évaluation au niveau transcriptomique par la technique de la ddPCR

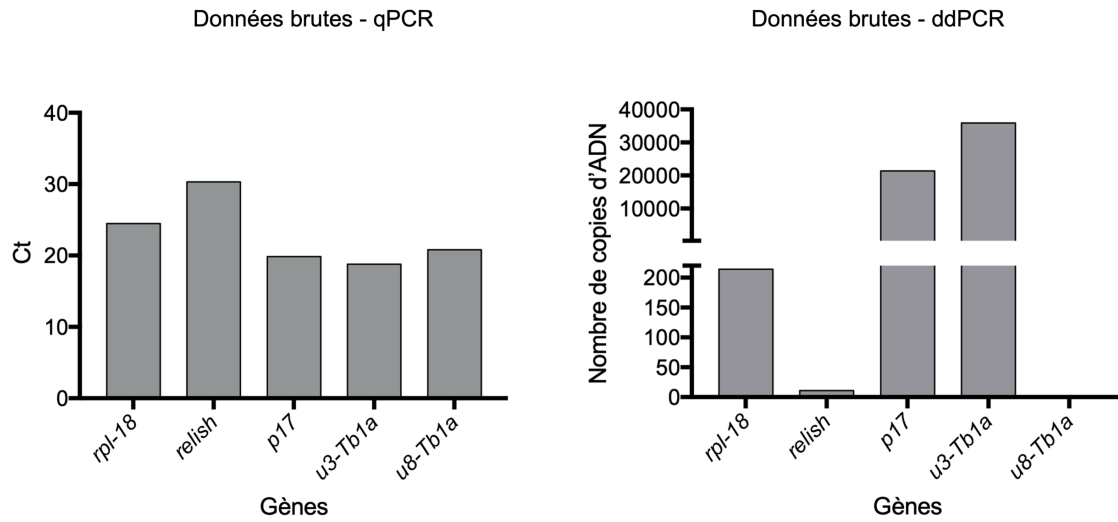
2803 Contrairement à la qPCR, la ddPCR permet la quantification absolue d'acides
2804 nucléiques dans un échantillon, sans nécessiter l'utilisation d'un gène de référence et à partir
2805 d'une faible quantité d'ARN de départ. La qPCR donne une valeur de cycle à partir de laquelle
2806 la phase exponentielle d'amplification débute (i.e. C_T) tandis que la ddPCR donne le nombre
2807 de copies obtenues en fin de réaction PCR. Cette technique permet donc (i) de détecter de
2808 faibles quantités d'ADN et (ii) de mettre en évidence de très faibles variations d'expression
2809 génétique (Hindson et al., 2011).

2810 Notre objectif a été de vérifier l'existence de faibles variations d'expression génétique
2811 dans les échantillons contrôles et infectés passés précédemment en qPCR. Cinq gènes ont été
2812 testés. Le gène *rpl-18* a été choisi dans le but de vérifier sa stabilité entre les conditions. Le
2813 gène *relish* a été sélectionné en raison de son appartenance à la voie immunitaire IMD, connue
2814 pour réagir majoritairement aux infections par des bactéries à Gram négatif (e.g. *Serratia*
2815 *marcescens*). Les gènes *p17*, *u3-Tb1a* et *u8-Tb1a* ont été testés car ils étaient exprimés dans la
2816 majorité des échantillons.

2817 Dix échantillons par type de tissu (i.e. abdomens dépourvus d'appareil vulnérant, corps
2818 gras et tubes digestifs) et par condition (i.e. contrôle et infecté) ont été choisis aléatoirement
2819 parmi les échantillons passés en qPCR. Un échantillon de glandes à venin a également été testé
2820 en tant que témoin positif de l'expression des gènes codant pour les toxines.

2821 Tous les gènes sont exprimés dans l'échantillon de glandes à venin testé, excepté *u8-*
2822 *Tb1a*. Ce gène a donc été retiré de la suite de l'étude. De manière générale, le nombre de copies
2823 retrouvées pour chaque gène dans les glandes à venin est cohérent avec les C_T observés en
2824 qPCR. Pour les gènes *p17* et *u3-Tb1a*, les nombres de copies sont supérieurs à 20 000,
2825 correspondant à des C_T inférieurs ou égaux à 20. Les gènes *rpl-18* et *relish* ont montré des
2826 nombres de copies inférieurs à 250, correspondant à des C_T supérieurs à 25 (Figure 43).

Glandes



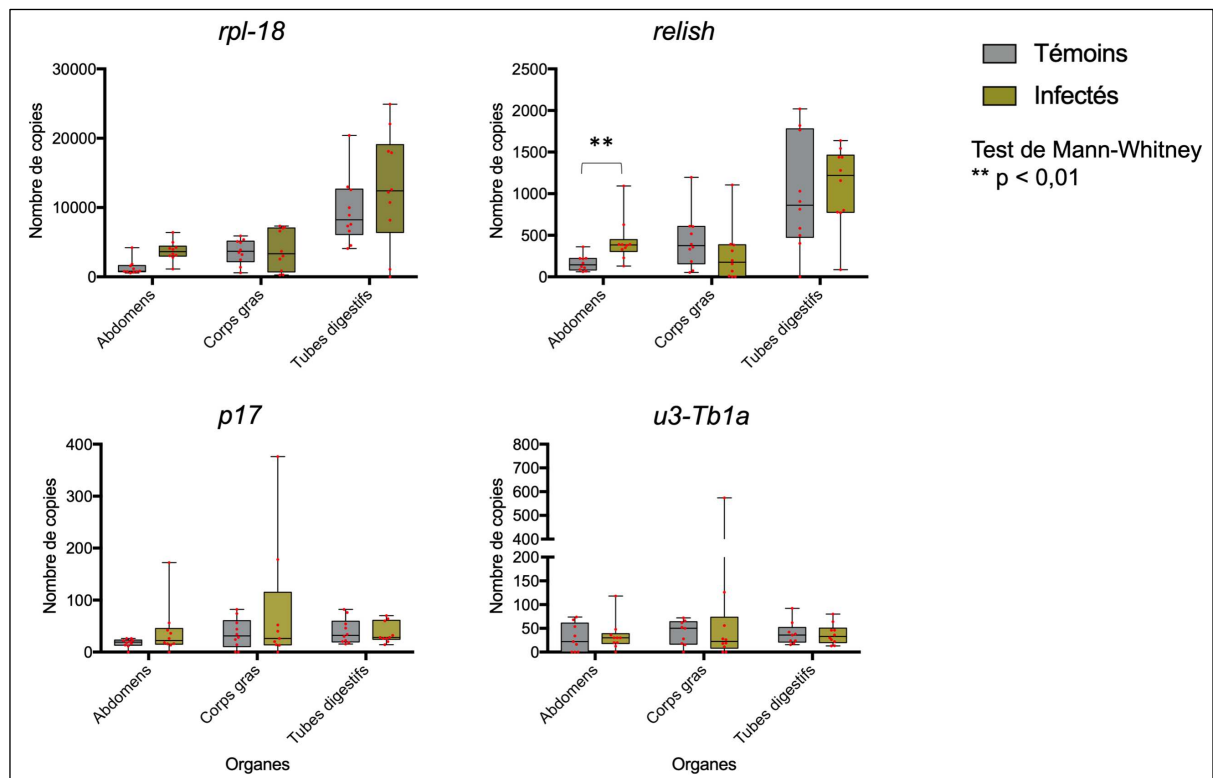
2827

2828 **Figure 43** : Comparaison des données brutes obtenues en qPCR et en ddPCR pour un échantillon de glandes
 2829 à venin.

2830 **Tableau VI** : Valeurs de rangs (W) et probabilités (P) obtenus suite aux tests de Mann-Whitney exécutés
 2831 entre les nombres de copies observées suite à une ddPCR pour les gènes *rpl-18*, *relish*, *p17* et *u3-Tb1a* chez
 2832 des *Tetramorium bicarinatum* saines ou infectées par *Serratia marcescens*.

Gènes/Organes	Abdomens		Corps gras		Tubes digestifs	
	W	P	W	P	W	P
<i>rpl-18</i>	43	0,631	42	0,842	65	0,280
<i>relish</i>	92	0,002	23,5	0,086	55	0,734
<i>p17</i>	62	0,383	45	1	47	0,880
<i>u3-Tb1a</i>	58	0,567	46	0,967	54,5	0,762

2833



2834

2835 **Figure 44** : Expression des gènes *rpl-18*, *relish*, *p17* et *u3-Tb1a* déterminées par ddPCR dans différents
 2836 organes de *Tetramorium bicarinatum* saines ou infectées par *Serratia marcescens*. Chaque point rouge
 2837 représente un échantillon.

2838 Le gène *rpl-18* s'est révélé instable entre les conditions dans les abdomens et tubes
 2839 digestifs. Bien qu'aucune différence statistique significative n'ait été notée entre les deux
 2840 conditions, les nombres moyens de copies étaient supérieurs dans ces organes chez les individus
 2841 infectés (Figure 44, Tableau VI).

2842 Les données de ddPCR ont montré que le gène *relish* est significativement surexprimé
 2843 dans les échantillons d'abdomens infectés (Mann-Whitney, $P < 0,01$), contrairement à ce qui a
 2844 été observé en qPCR. *A contrario*, les échantillons de corps gras infectés ont tendance à
 2845 présenter moins de copies pour ce gène comparativement aux individus contrôles. Ce dernier
 2846 résultat est cohérent par rapport à la sous-expression significative observée en qPCR. Aucune
 2847 différence significative pour le gène *relish* n'a cependant été observée dans les tubes digestifs,
 2848 bien que le nombre moyen de copies soit supérieur chez les individus infectés. Ce résultat
 2849 semblerait cohérent avec le fait que l'infection ait été effectuée par voie orale et que *relish* code
 2850 pour un NF- κ B impliqué dans la voie IMD. Cette voie de signalisation immunitaire est en effet
 2851 connue chez les insectes pour répondre majoritairement aux infections par des bactéries à Gram
 2852 négatif (e.g. *Serratia marcescens*) (Neyen, Bretscher, Binggeli, & Lemaitre, 2014). Cependant,
 2853 le meilleur moyen de vérifier le passage des bactéries dans le milieu intérieur et l'efficacité de

2854 l'infection orale serait d'utiliser des bactéries transformées pour exprimer la GFP (Green
2855 Fluorescent Protein) (Neyen et al., 2014).

2856 Aucune différence significative d'expression n'a également été notée entre les
2857 conditions pour les gènes *p17* et *u3-Tb1a* dans les différents tissus testés (Tableau VI). Plus de
2858 copies ont néanmoins été observées dans certains échantillons d'abdomens et de corps gras
2859 infectés pour le gène *p17* (Figure 44).

2860 De manière générale, de grandes variabilités ont été observées dans le nombre de copies
2861 estimées dans les échantillons d'une même condition pour tous les gènes testés. Ces variabilités
2862 empêchent notamment toute conclusion statistique quant à la réponse des gènes codant les
2863 toxines au niveau transcriptomique face à une infection bactérienne (Figures 42 et 44). Elles
2864 ont également été observées dans les données brutes de qPCR (i.e. C_T) et se sont donc
2865 répercutées sur les ratios d'expression relative. L'augmentation du nombre d'échantillons,
2866 couplée à une réduction du nombre d'organes par lot constituerait un moyen de contrer cette
2867 variabilité en ddPCR. Cette technique ne permettra néanmoins pas de la gommer totalement.
2868 Les premiers résultats d'immunohistochimie sur les corps gras apportent en effet un nouvel
2869 éclairage sur la cause de cette variabilité (Cf. partie 2.2). Il est possible que les échantillons
2870 présentent une hétérogénéité quant à leur contenu en oenocytes. Ces caractérisations au niveau
2871 protéique s'étant déroulées après les analyses transcriptomiques, nous n'avons pas tenu compte
2872 de ce facteur lors des dissections et préparations des échantillons pour les analyses par q- et dd-
2873 PCR.

2874

2875 3.3. Évaluation de la réponse au niveau protéique

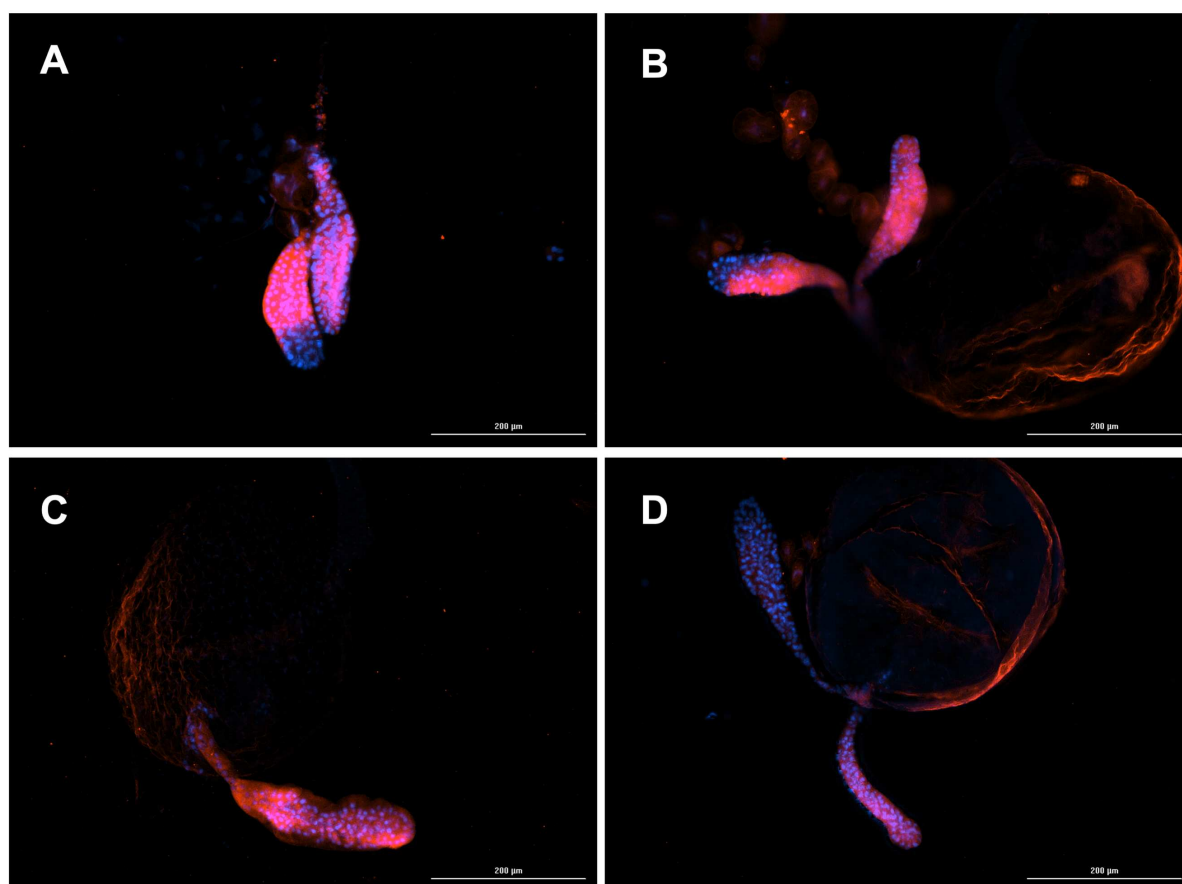
2876 Afin de vérifier l'expression des gènes codant pour la Bicarinaline et le P17 en termes
2877 de quantité intracellulaire de peptides matures suite à un challenge bactérien, un protocole
2878 d'immunohistochimie indirecte a été appliqué sur les abdomens de *T. bicarinatum* infectées ou
2879 non par la bactérie *Serratia marcescens*.

2880 Un total de 5 échantillons, contenant chacun 5 abdomens, a été constitué : un témoin
2881 relatif à l'anticorps secondaire et un échantillon de chaque condition (i.e. contrôle et infecté)
2882 pour Bicarinaline et P17. Cette expérimentation a été effectuée deux fois et un total de 50
2883 fourmis a été nécessaire à sa réalisation. Les intensités lumineuses ont été mesurées pour trois
2884 types de tissus : les glandes à venin, les corps gras et les tubes de Malpighi. Comme
2885 précédemment, les glandes à venin constituent un témoin positif quant à la présence des deux
2886 peptides.

2887 Cinq mesures ont été prises par type d'organe et par échantillon, et corrigées par
2888 soustraction de la valeur de l'intensité lumineuse des organes correspondants au témoin de
2889 fluorescence des anticorps secondaires.

2890
2891 Des signaux positifs ont été détectés pour les deux peptides dans les appareils vulnérants
2892 (i.e. glandes à venin et réservoir) (Figure 45), nous permettant de repérer la Bicarinaline et le
2893 P17 dans les autres organes. Pour chaque peptide, les intensités de ces signaux ne diffèrent pas
2894 entre les conditions (Tableau VII).

2895



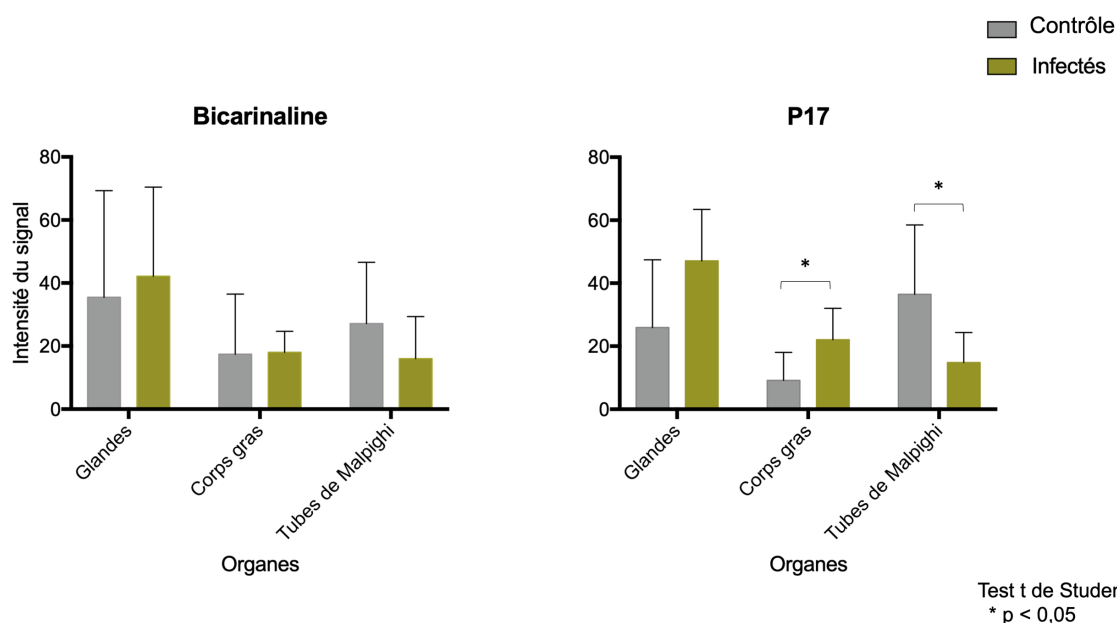
2896
2897 **Figure 45** : Localisation de la Bicarinaline et du P17 dans des glandes à venin de *Tetramorium bicarinatum*
2898 des lots contrôle (A et B) et infecté (C et D), réalisée par immunohistochimie indirecte (grossissement X10).
2899 Les anticorps secondaires étaient couplés au Texas Red. Les photos montrent les superpositions des signaux
2900 DAPI (noyaux cellulaires) et Texas Red (Bicarinaline (A et C) et P17 (B et D)).

2901

2902 **Tableau VII** : Valeurs de la statistique t, des degrés de liberté et de la probabilité obtenus suite aux tests t de
 2903 Student, exécutés entre les intensités lumineuses mesurées pour les peptides Bicarinaline et P17 dans les
 2904 organes de *Tetramorium bicarinatum* saines ou infectées par *Serratia marcescens*.

Peptides/Organes	Glandes			Corps gras			Tubes digestifs		
	t	df	P	t	df	P	t	df	P
Bicarinaline	-0,307	5,808	0,770	-0,099	8,480	0,924	1,358	15	0,194
P17	-1,777	8	0,114	-2,447	15	0,027	2,538	13	0,025

2905



2906 **Figure 46** : Intensité des signaux observés par immunohistochimie indirecte pour les peptides Bicarinaline
 2907 et P17 dans des organes de *Tetramorium bicarinatum* saines ou infectées par *Serratia marcescens*.
 2908

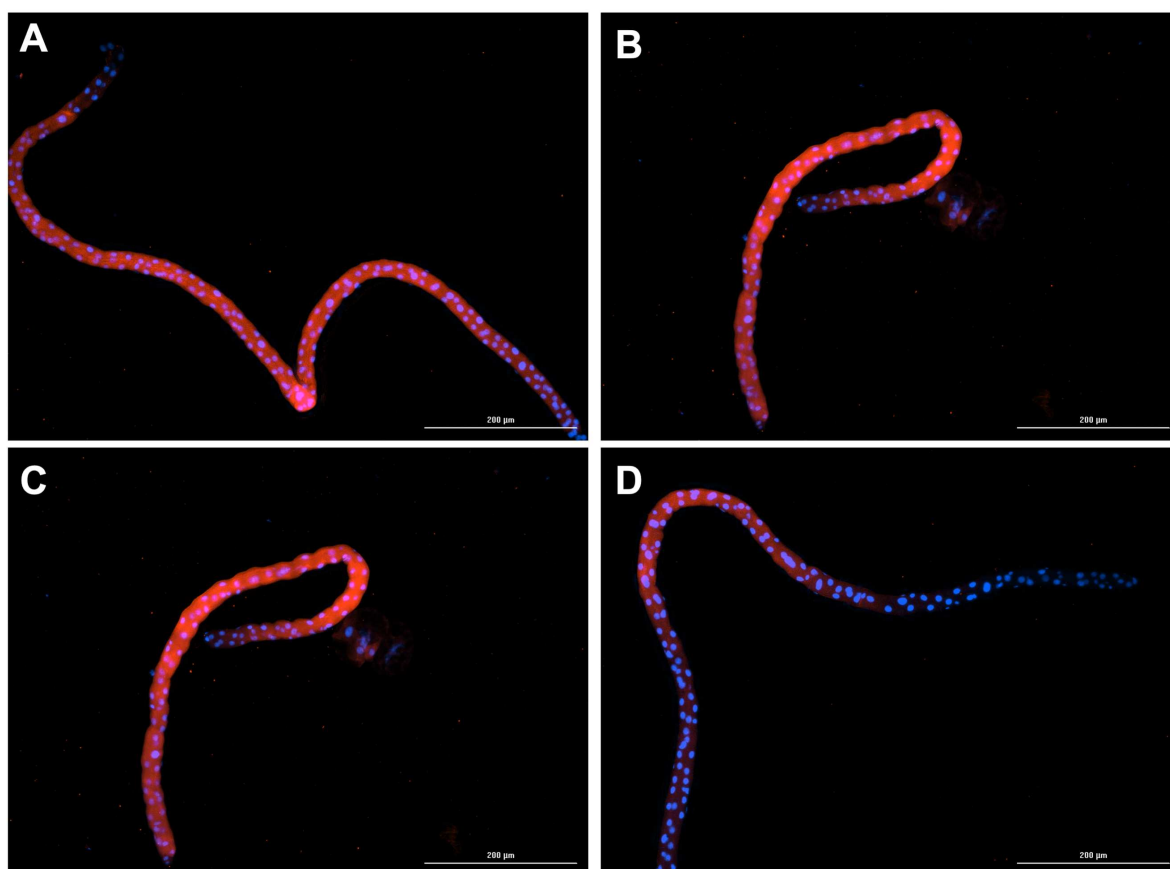
2909 La variabilité observée dans les approches transcriptomiques pour les deux peptides est
 2910 également visible dans les résultats de cette immunohistochimie indirecte, quel que soit
 2911 l'organe testé (Figure 46). P17 était néanmoins présent en plus grande quantité de manière
 2912 significative dans les corps gras des fourmis infectées (Test-t de Student, P < 0,05 ; Tableau
 2913 VII ; Figures 46 et 48), alors qu'aucune variation significative n'a été notée dans les approches
 2914 transcriptomiques pour ce gène, bien que certains échantillons de corps gras infectés aient
 2915 présenté un plus grand nombre de copies en ddPCR comparativement aux échantillons
 2916 contrôles. La présence de P17 en plus grande quantité dans les corps gras infectés démontre
 2917 néanmoins qu'il y a une production de cette toxine en réponse à une infection bactérienne.

2918 Les intensités lumineuses mesurées pour la Bicarinaline dans les corps gras et les tubes
 2919 de Malpighi n'ont présenté aucune différence significative entre les deux conditions (Tableau

2920 VII, Figure 47). La présence de transcrits du gène *bicarinaline* dans les tubes de Malpighi et
2921 les corps gras, couplée à la présence de la Bicarinaline dans ces mêmes organes chez des
2922 individus sains montrent néanmoins que ce peptide est produit et stocké de manière constitutive.
2923 Les intensités lumineuses mesurées dans les tubes de Malpighi semblaient cependant plus
2924 atténuées en moyenne chez les individus infectés (Figures 46 et 47). Le P17 a présenté des
2925 intensités lumineuses significativement moins marquées dans les tubes de Malpighi des
2926 individus infectés (Test-t de Student, Tableau VII ; $P < 0,05$; Figure 47), suggérant que ce
2927 peptide pourrait être sécrété par ces organes en réponse à un challenge bactérien.

2928 Il a déjà été démontré que des gènes codant des HDPs s'exprimaient dans les tubes de
2929 Malpighi chez *Drosophila melanogaster* (Tapadia & Verma, 2012). Le phénomène de sécrétion
2930 n'a en revanche été démontré que plus récemment pour le corps gras de la mouche *Calliphora*
2931 *vicina*. Cette étude ne s'est cependant pas intéressée aux variations éventuelles entre les types
2932 cellulaires de cet organe (Yakovlev et al., 2017).

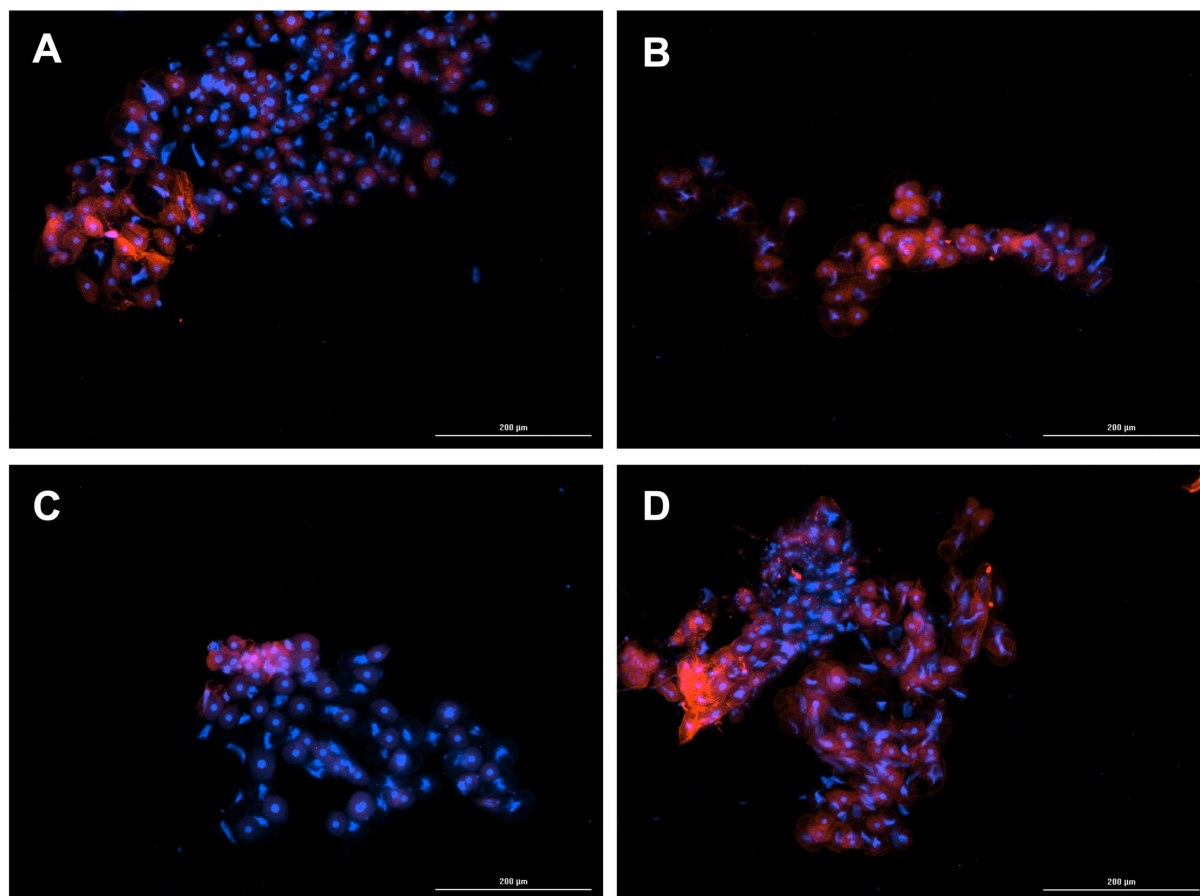
2933



2934
2935 **Figure 47 :** Localisation de la Bicarinaline et du P17 dans des tubes de Malpighi de *Tetramorium*
2936 *bicarinatum* des lots contrôle (A et B) et infecté (C et D), réalisée par immunohistochimie indirecte
2937 (grossissement X10). Les anticorps secondaires étaient couplés au Texas Red. Les photos montrent les

2938 superpositions des signaux DAPI (noyaux cellulaires) et Texas Red de la Bicarinaline (A et C) et du P17 (B
2939 et D).

2940



2941
2942 **Figure 48** : Localisation de la Bicarinaline et du P17 dans des corps gras de *Tetramorium bicarinatum* des
2943 lots contrôle (A et B) et infecté (C et D), réalisée par immunohistochimie indirecte (grossissement X10). Les
2944 anticorps secondaires étaient couplés au Texas Red. Les photos montrent les superpositions des signaux
2945 DAPI (noyaux cellulaires) et Texas Red de la Bicarinaline (A et C) et du P17 (B et D).

2946

2947 Indépendamment de l'intensité des signaux, les deux types cellulaires du corps gras (i.e.
2948 trophocytes et oenocytes) ont présenté des signaux positifs chez les individus infectés pour les
2949 deux peptides testés. *A contrario*, les peptides ont été retrouvés majoritairement dans les
2950 oenocytes chez les individus contrôles (Figure 14). Les deux peptides semblent donc produits
2951 dans les deux types cellulaires suite à un challenge bactérien. La contribution relative de ces
2952 deux types cellulaires à la synthèse d'HDPs suite à une infection microbienne n'a, à notre
2953 connaissance, jamais été étudiée chez les fourmis. Il en est de même pour les cellules composant
2954 les tubes de Malpighi.

2955 Pour résumer, nos résultats montrent que le P17 est moins présent dans les tubes de
2956 Malpighi du lot infecté, suggérant qu'il serait sécrété par ces organes en réponse à une infection
2957 bactérienne. La Bicarinaline semble suivre la même tendance pour cet organe. En revanche, le
2958 P17 est présent en plus grande quantité dans les corps gras du lot infecté. Bien qu'aucune
2959 différence significative n'ait été notée pour la Bicarinaline, les résultats d'imagerie ont montré
2960 que les deux peptides sont produits dans les deux types cellulaires du corps gras en réponse à
2961 une infection bactérienne.

2962 Pour aller plus loin, il serait tout d'abord nécessaire de mettre en place des cultures
2963 primaires des cellules du corps gras et plus particulièrement des oenocytes (Martins et al.,
2964 2011), pour ensuite mimer une infection bactérienne grâce à une incubation avec des
2965 lipopolysaccharides. Il serait ensuite possible de déterminer leurs profils d'expressions
2966 transcriptomiques respectifs par ddPCR et de confirmer la présence des deux peptides par
2967 immunohistochimie, avec une variabilité moindre que celle observée sur des échantillons
2968 prélevés par dissection. Une technique telle que le RNAseq en différentiel permettrait d'évaluer
2969 de manière précise la réponse au niveau transcriptomique des gènes codant pour les toxines du
2970 venin de *T. bicarinatum* face à un challenge bactérien. Cette technique nécessite néanmoins une
2971 grande quantité d'ARN de départ, ainsi qu'un génome de référence pour relier les transcrits
2972 détectés aux gènes d'intérêt. Au niveau protéique, un phénomène de sécrétion de la Bicarinaline
2973 et du P17 pourrait également être mis en évidence grâce à un dosage de ces derniers dans le
2974 milieu de culture cellulaire *via* des méthodes de spectrométrie de masse (Yakovlev et al., 2017).
2975

2976 4. Matériel et méthodes

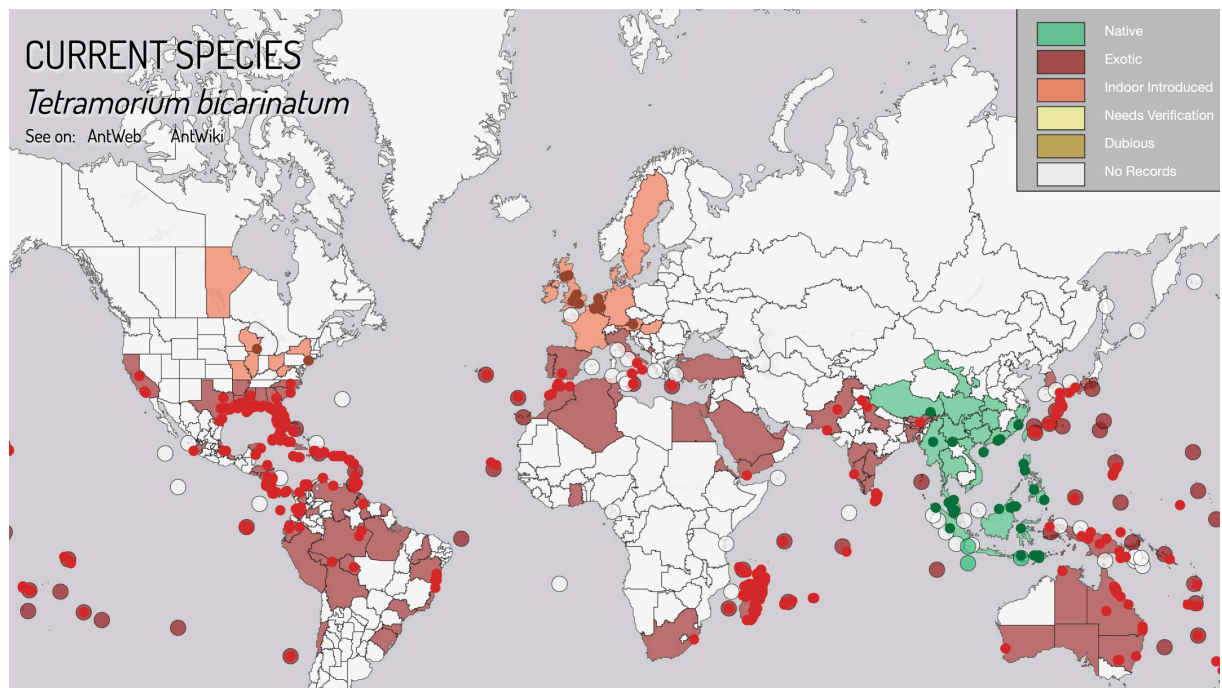
2977 4.1. *Tetramorium bicarinatum* : généralités et élevage

2978 *Tetramorium bicarinatum* (Nylander, 1846) (Hymenoptera : Formicidae) est une fourmi de
2979 la sous-famille des *Myrmicinae* dont les ouvrières mesurent approximativement 4 mm de long
2980 (Figure 49) (Wetterer, 2009).



2981
2982 **Figure 49** : *Tetramorium bicarinatum* (Hymenoptera : Formicidae), photo extraite de AntWeb le 03/08/2020
2983 (<https://www.antweb.org/specimenImages.do?name=casent0005826&project=allantwebants>).

2984 Originaire d'Asie du Sud-Est, cette fourmi dites « vagabonde » ou « tramp species » possède
2985 certaines caractéristiques propres aux espèces envahissantes telles que l'unicolonialité et la
2986 polygynie (Hölldobler & Wilson, 1990). Actuellement très répandue dans les régions tropicales
2987 et subtropicales, elle est également présente sur tous les continents, grâce à une reproduction
2988 par bouturage et une dispersion *via* les activités commerciales (Figure 50). Il s'agit d'une espèce
2989 peu exigeante qui se maintient facilement en laboratoire.
2990



2991

2992 **Figure 50** : Carte de répartition de *Tetramorium bicarinatum*, extraite de AntMap le 03/08/2020

2993 (<https://www.antmaps.org/?mode=species&species=Tetramorium.bicarinatum>).

2994

2995

2996

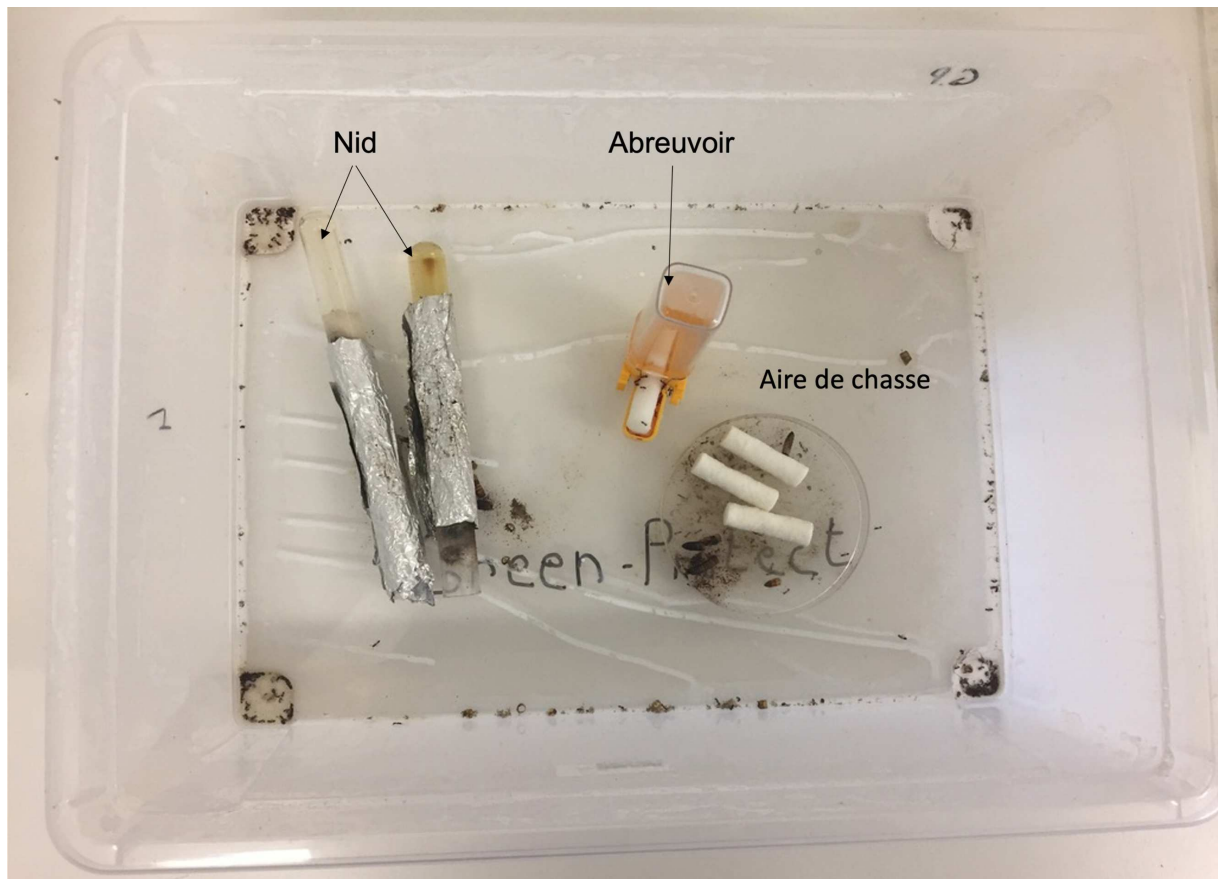
2997

2998

2999

3000

Les individus composant l'élevage de *T. bicarinatum* ont été prélevés dans une serre tropicalisée de Belgique (prélèvement et acheminement assurés par Philippe Wegnez). Un nid artificiel est maintenu au laboratoire à 25 °C et 60-70 % d'humidité (Figure 51). La salle d'élevage étant pourvue de fenêtres, la photopériode est naturelle. Les individus sont nourris 2 fois par semaine avec des vers frais de *Tenebrio molitor* (*Coleoptera* : *Tenebrionidae*) et des cotons dentaires (l = 30mm ; Ø = 10mm) imbibés de miel dilué (50 : 50 v/v).



3001
 3002 **Figure 51** : Nid artificiel de *Tetramorium bicarinatum*. Le nid est constitué de deux tubes à essais
 3003 recouverts d'aluminium, afin de recréer les conditions de sous-sol. L'aire de chasse, espace réservé à la
 3004 distribution de la nourriture et au prélèvement des fourmis, est une boîte en plastique dont les parois
 3005 sont enduites de Fluon.

3006 4.2. Dissections

3007 Des dissections ont été réalisées pour répondre aux deux objectifs de ce chapitre.

3008 Avant chaque expérimentation, les ouvrières *T. bicarinatum* ont été placées dans une boîte
 3009 de Pétri (h = 16,2mm ; \varnothing = 90mm) avec un coton dentaire (l = 30mm ; \varnothing = 10mm) imbibé d'eau
 3010 et placées à l'obscurité pendant 48h à 25°C. Les fourmis ont été endormies par un passage à -
 3011 20°C pendant 5 minutes, puis placées dans une boîte de Pétri sur glace (4°C). Elles ont ensuite
 3012 été disséquées individuellement dans un cristallisateur contenant une solution de « Phosphate
 3013 Buffered Saline » (Sigma-Aldrich) et de Tween® (Sigma-Aldrich) (PBST) sous une loupe
 3014 binoculaire (Leica). Cette solution a été changée entre chaque individu afin d'éviter toute
 3015 contamination par l'hémolymphe. Les dissections ont été réalisées avec des pinces fines
 3016 (Sigma-Aldrich : Jewelers forceps, Dumont, N° 5).

3017

3018 4.3. *Localisation des peptides en dehors des appareils vulnérants*

3019 4.3.1. *Localisation des transcrits*

3020 Vingt ouvrières de *T. bicarinatum* ont été isolées en boîte de Pétri pendant 48h puis
3021 disséquées comme précédemment décrit dans la partie 4.2. Six types de tissus ont été isolés et
3022 placés par lots de 5 : des appareils vulnérants (glandes à venin et réservoir), des mésentérons,
3023 des tubes de Malpighi, des rectums, des corps gras et des abdomens dépourvus d'appareil
3024 vulnérant. Une attention particulière a été portée à la constitution des échantillons d'abdomens,
3025 afin de s'assurer que les appareils vulnérants en aient été totalement retirés. Chaque type de
3026 tissus a été directement placé dans un tube Eppendorf de 1,5 ml contenant 250 μ L d'un tampon
3027 relatif à l'extraction des ARNs, constitué d'acide 4-Morpholineethanesulfonique (20 mM final)
3028 (Sigma-Aldrich), d'acide d'Éthylène Diamine Tetra-Acétique (pH = 8) (20 mM final) (Sigma-
3029 Aldrich) et d'hydrochlorure de Guanidine (8 M final) (Sigma-Aldrich).

3030

3031 > *Extraction et traitement de l'ARN*

3032 Les extractions et traitements des ARNs ont été réalisés immédiatement après les
3033 dissections. Les extractions ont été réalisées selon le protocole sans Trizol proposé par Yaffe et
3034 ses collaborateurs (Yaffe et al., 2012). Dans le but de supprimer l'ADN génomique
3035 contaminant, un traitement DNase a été effectué avec le kit DNA-free (DNase treatment and
3036 removal reagent, Life technologies, Ambion) selon les recommandations du fournisseur. La
3037 quantité d'ARNs contenue dans chaque échantillon a ensuite été estimée avec un
3038 spectrophotomètre (NanoDrop 2000, ThermoFisher Scientific).

3039

3040 > *Rétrotranscription et amplification par Réaction de Polymérisation en*
3041 *Chaîne (PCR)*

3042 Les rétrotranscriptions des ARNs en ADNs complémentaires ont été réalisées avec des
3043 amorces aléatoires (Random hexamers, Invitrogen) sur 80 ng d'ARN totaux pour chaque tissu,
3044 en suivant le protocole fourni avec le kit M-MLV RT (Invitrogen).

3045 Les PCRs ont été effectuées à 60°C (i.e. température d'hybridation des amorces) durant 40
3046 cycles. Deux gènes de référence ont été testés : *rpl-18* et *actine*. Ces deux gènes codent
3047 respectivement pour une protéine ribosomique appartenant à la grande sous-unité 60S du
3048 ribosome et pour une protéine impliquée dans la motilité, la structure et l'intégrité cellulaire
3049 (Cheng et al., 2013). Le gène *actine* a d'ailleurs été largement utilisé comme gène de référence
3050 dans de nombreuses études chez les insectes (Bagnall & Kotze, 2010; Cheng et al., 2013). Huit
3051 gènes codant des peptides de venin de *T. bicarinatum* ont été testés : *bicarinaline*, *p17*, *u3-*

3052 *Tb1a*, *u4-Tb1a*, *u5-Tb1a*, *u7-Tb1a*, *u8-Tb1a*, et *u9-Tb1a* et *u16-Tb1a*. La Bicarinaline est un
3053 peptide antimicrobien (Rifflet et al., 2012; Téné et al., 2014) et P17 possède des propriétés
3054 immunomodulatrices (Brevet n°FR1750280) (Benmoussa et al., 2017). Les peptides U₃, U₄,
3055 U₅, U₇, U₈ et U₉ appartiennent à la même famille de précurseurs que celle de la Bicarinaline et
3056 P17. Le peptide U₁₆ appartient à une famille de précurseurs proche de celle de la cécropine-B
3057 de *Drosophila melanogaster*, un HDP contenu dans l'hémolymphe (Touchard et al., 2018). Les
3058 couples d'amorces utilisés sont répertoriés dans le tableau VIII. Les résultats ont été visualisés
3059 via une électrophorèse sur gel d'agarose 2%.

3060

3061 **Tableau VIII** : Couples d'amorces utilisés pour la PCR. La température d'hybridation des amorces est de
3062 60 °C.

Amorces	Séquences (5'-3')	Taille des fragments attendus
<i>rpL-18 F</i>	GTATCTTGAAAGCCGGTGG	180 pb
<i>rpL-18 R</i>	CATACGTGAGGTCTAAGGGC	
<i>actine F</i>	GGCTTCTTCCTCCAGTTTGG	192 pb
<i>actine R</i>	TACAAGTCCTTGCGGATGTC	
<i>bicarinaline F</i>	GTTGGTGCTTGCCATAATCC	180 pb
<i>bicarinaline R</i>	TTTAACCTTGCCCCATGGTA	
<i>p17 F</i>	GATGCTGATGCTGCTGCTT	180 pb
<i>p17 R</i>	GTTGTTGCAGTCCTCCAGT	
<i>u3-Tb1a F</i>	TCGTAGGTCTAAGTGGCTCA	155 pb
<i>u3-Tb1a R</i>	TGGTAGACTCATCAAGCCCG	
<i>u4-Tb1a F</i>	TGCAACCATCATATCTCCTGTT	156 pb
<i>u4-Tb1a R</i>	CGCACAGGTTCCCTCATCC	
<i>u5-Tb1a F</i>	ATCGCATTTATTGTTGGCTTTTG	180 pb
<i>u5-Tb1a R</i>	TCGCTCTATCAGGTTTCGCAT	
<i>u7-Tb1a F</i>	ACGATTATCCACACACCCCA	117 pb
<i>u7-Tb1a R</i>	TTATGCACTGACTCCACGGC	
<i>u8-Tb1a F</i>	TGGCAATCATGTACGCACC	150 pb
<i>u8-Tb1a R</i>	ATCCCATGAATCCTTTTACTGCG	
<i>u9-Tb1a F</i>	ACCATATACGCACCCCAAGT	158 pb
<i>u9-Tb1a R</i>	CCCAAAGCCATTTTCAAACCC	
<i>u16-Tb1a F</i>	TGCAGTTACCATGATCCCTG	121 pb
<i>u16-Tb1a R</i>	GCCTTTTGTGTTGTTTAAAGAGCC	

3063

3064 Cette expérimentation a été répétée 3 fois. Un total de 30 fourmis a été nécessaire à sa
3065 réalisation.

3066

3067 4.3.2. Localisation de la Bicarinaline et du P17

3068 Afin de vérifier la présence effective de la Bicarinaline et du P17 en dehors des glandes à
3069 venin, un protocole d'immunohistochimie indirecte a été appliquée sur les tissus composant
3070 l'abdomen. Cette technique permet de localiser les protéines au niveau cellulaire et fait
3071 intervenir deux types d'anticorps : des anticorps primaires, dirigés contre la ou les protéines
3072 ciblées, et des anticorps secondaires couplés à un fluorochrome et dirigés contre les anticorps
3073 primaires. Le fait d'utiliser ces deux sortes d'anticorps permet une meilleure sensibilité. Les
3074 anticorps primaires utilisés étaient respectivement dirigés contre la Bicarinaline et P17 et ont
3075 été produits chez le lapin par la société Eurogentec. Ces anticorps se fixent sur les parties N-
3076 terminales des peptides matures. Les anticorps secondaires anti-lapin ont été produits chez la
3077 chèvre et couplés au Texas Red (Invitrogen).

3078 Douze *T. bicarinatum* ont été isolées en boîte de Pétri pendant 48h puis disséquées comme
3079 décrit dans la partie 4.2. Les abdomens ont été isolés du reste du corps et les tergites T1 et T2
3080 ont été retirés. Un total de quatre échantillons, contenant chacun de 3 abdomens, a été constitué :
3081 deux témoins et un échantillon pour P17 et Bicarinaline. Le témoin négatif est un échantillon
3082 dans lequel aucun anticorps n'a été ajouté. Un second témoin relatif à l'anticorps secondaire,
3083 dans lequel seul l'anticorps couplé au Texas Red a été inclus de manière à vérifier la présence
3084 et la qualité du signal de fluorescence. Après chaque dissection, les échantillons ont été placés
3085 dans une solution dite de « fixation » avant de procéder au protocole d'immunohistochimie.
3086 Cette expérimentation a été répétée 3 fois. Un total de 36 fourmis a été nécessaire à sa
3087 réalisation.

3088

3089 ➤ Protocole d'immunohistochimie indirecte

3090 Chaque échantillon a été placé dans un 1 mL de solution de fixation (paraformaldéhyde
3091 (4% final) (Acros Organics) ; PBS (1X) (Sigma-Aldrich) ; Tween (0,01% final) (Sigma-
3092 Aldrich)), puis incubé à température ambiante pendant 30 minutes. Nous avons ensuite rincé
3093 les échantillons trois fois pendant 5 min avec 1 mL d'une solution de PBST avant de les mettre
3094 en incubation dans 1 mL d'une solution de perméabilisation (PBS (1X final) ; Triton (1X final)
3095 (Sigma-Aldrich)) pendant 1 heure à température ambiante. Après une nouvelle étape de rinçage
3096 pendant 5 min dans 1 mL de PBST, les échantillons ont été incubés dans 250 µL d'une solution

3097 de blocage constituée de PBST et de Normal Goat Serum (5% final) (NGS) (Invitrogen)) à
3098 température ambiante. A la suite de cette étape, les échantillons ont été incubés avec les
3099 anticorps primaires anti-Bicarinaline et anti-P17 à une concentration de 1 : 250 à 4°C pendant
3100 une nuit à l'obscurité.

3101 Les échantillons ont été remis pendant 1 heure à température ambiante avant d'être
3102 rincés trois fois pendant 10 min dans 1 mL de PBST et d'être transférés dans 200 µL d'une
3103 solution constituée de NGS 2% final dans du PBST. Nous avons ajouté les anticorps
3104 secondaires à une concentration de 1 : 200 et laissé les échantillons en incubation à 4°C pendant
3105 une nuit à l'obscurité.

3106 Les échantillons ont été remis pendant 1 heure à température ambiante avant d'être
3107 rincés trois fois pendant 10 min dans 1 mL de PBST. Nous les avons transférés dans une
3108 solution constituée de PBS, de glycérol 50% (VWR) et de DAPI dilactate (1 µg/mL), et en
3109 incubation pendant 1 heure à température ambiante à l'obscurité. Les échantillons ont été
3110 transférés dans une solution de PBS et de glycérol 80%.

3111 De nouvelles dissections ont été réalisées sur ces échantillons directement dans la
3112 solution PBS-glycérol 80% afin d'isoler les organes composant les abdomens. Les glandes à
3113 venin, tubes digestifs (mésentéron et tubes de Malpighi) et corps gras ont été observés entre
3114 lame et lamelle au Cytation Reader 1 (BioTek).

3115

3116 *4.4. Etude de la réponse des gènes codant pour les peptides à une infection bactérienne*

3117 *4.4.1. Microbiologie et infection des fourmis par voie orale*

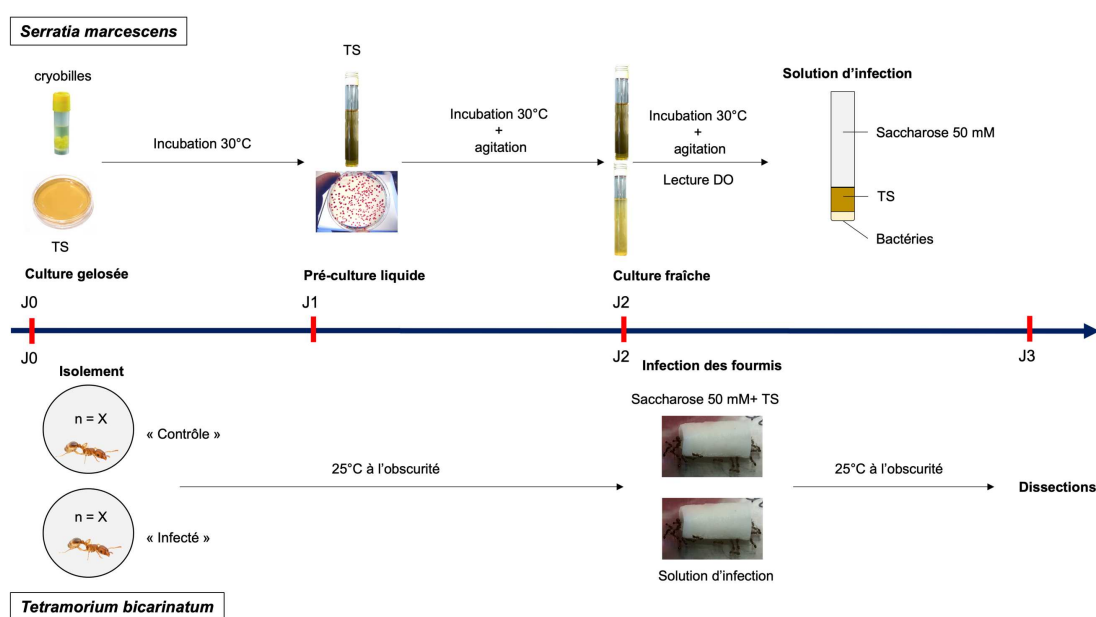
3118 Nous avons sélectionné *Serratia marcescens* (*Enterobacteriales* : *Enterobacteriaceae*)
3119 pour réaliser l'infection des fourmis. Cette bactérie à Gram négatif est considérée comme
3120 potentiellement entomopathogène du fait de sa capacité à sécréter des chitinases et donc à
3121 dégrader la cuticule pour accéder à la cavité corporelle (van Aalten et al., 2000). Elle est ainsi
3122 couramment utilisée dans l'étude de l'immunité innée des insectes (Neyen et al., 2014).

3123 Nous avons cultivé *S. marcescens* (souche : CIP103235T) selon les recommandations
3124 de l'institut Pasteur sur un milieu tryptone-soja (TS) à 30°C en condition aérobie. La culture
3125 bactérienne a été réalisée en trois phases. La première phase de culture a consisté en l'étalement
3126 d'une cryobille de *S. marcescens* sur un milieu gélosé TS suivi d'une incubation à 30°C pendant
3127 une nuit. Pour la seconde phase, dite de « pré-culture », 1 à 2 colonies ont été placées dans
3128 10 mL de milieu TS liquide sous agitation à 30°C pendant une nuit (Figure 52).

3129 La dernière phase de culture consistait à remettre les bactéries en culture dite « fraîche »
 3130 jusqu'à atteindre la phase exponentielle de croissance, phase à laquelle toutes les bactéries sont
 3131 en division. Pour ceci, 500 µL de la pré-culture liquide ont été prélevés et ajoutés à 10 mL de
 3132 milieu TS liquide. Cette préparation a ensuite été laissée sous agitation à 30°C. Une lecture de
 3133 densité optique (DO) à 600 nm a été effectuée toutes les heures à l'aide d'un lecteur de plaque
 3134 (Técan) jusqu'à ce qu'elle soit comprise entre 0,5 et 0,6. Cette DO est caractéristique de la
 3135 phase exponentielle de croissance de la souche testée de *S. marcescens* et a été déterminée
 3136 préalablement à ces travaux de thèse. La phase exponentielle de croissance pour cette souche
 3137 est atteinte en environ 4 heures et équivaut à une concentration moyenne de $1,5 \cdot 10^9$ UFC/mL
 3138 (équipe BTSSB, données non publiées).

3139 Pour préparer la solution d'infection des fourmis, le milieu TS de la culture fraîche a été
 3140 éliminé par centrifugation et remplacé par un même volume d'une solution composée d' $\frac{1}{5}$ de
 3141 milieu TS et de $\frac{4}{5}$ de solution de saccharose à 50 mM (Figure 52).

3142 Deux lots d'ouvrières *T. bicarinatum* ont été constitués : un lot « contrôle » et un lot
 3143 « infecté ». Les fourmis ont été préalablement placées dans des boîtes de Pétri à 25°C pendant
 3144 48h comme précédemment décrit dans la partie 4.2. Des cotons dentaires (l = 1 cm ; Ø = 1 cm)
 3145 imbibés de 1,5 mL de solution d'infection, classique pour le lot « infecté » et sans bactérie pour
 3146 le lot « contrôle », ont été placés au milieu des boîtes de Pétri respectives. Les lots ont ensuite
 3147 été replacés à 25°C à l'obscurité pendant 24h (Figure 52).



3148 **Figure 52 :** Microbiologie et infection par voie orale de la fourmi *Tetramorium bicarinatum* par la
 3149 bactérie *Serratia marcescens*. Le protocole d'infection se déroule sur trois jours pendant lesquels les
 3150

3151 traitements des bactéries et des fourmis sont réalisés de façon parallèle. Les bactéries sont manipulées
3152 de J0 à J2 tandis que les fourmis sont manipulées à J0, J2 et J3.

3153 4.4.2. Etude au niveau transcriptomique

3154 Soixante-dix ouvrières *T. bicarinatum* ont été isolées par condition (i.e. contrôle et
3155 infecté) puis infectées par voie orale comme décrit dans la partie 4.4.1. Vingt-quatre heures
3156 après infection, quatre types de tissus ont été isolés par dissection comme décrit dans la partie
3157 4.2 : des glandes à venin, des corps gras, des tubes digestifs et des abdomens dépourvus
3158 d'appareil vulnérant. Les tissus ont été placés par lots de manière à optimiser la quantité
3159 d'ARNs récoltés, directement dans le tampon d'extraction. Ainsi, les glandes à venin ont été
3160 placées par lots de 5 et deux échantillons ont été constitués pour chaque condition, servant de
3161 témoins positifs quant à l'expression des gènes codant les peptides testés. Les abdomens
3162 dépourvus d'appareil vulnérant ont été placés par lots de 2. Les corps gras et les tubes digestifs
3163 ont été placés par lots de 5.

3164 Cette expérimentation a été répétée deux fois, permettant de réaliser quatre échantillons
3165 de glandes à venin, et vingt échantillons pour les autres tissus, par condition. Au total, 280
3166 fourmis ont été nécessaires à la réalisation de cette expérimentation.

3167

3168 > Extraction des ARNs et rétrotranscriptions

3169 Les extractions des ARNs et traitements DNase ont été réalisés directement après les
3170 dissections via l'utilisation du RNeasy® Micro Kit (Quiagen) selon les recommandations du
3171 fournisseur. Les rétrotranscriptions des ARNs en ADNn complémentaires ont été réalisées avec
3172 des amorces aléatoires (Random hexamers, Invitrogen) sur 80 ng d'ARN suivant le protocole
3173 fourni avec le kit M-MLV RT (Invitrogen).

3174

3175 > Protocole de PCR quantitative (qPCR)

3176 Afin d'évaluer la réponse des gènes codant les peptides de venin suite à un challenge
3177 bactérien, nous avons choisi de recourir dans un premier temps à la qPCR. Les couples
3178 d'amorces utilisés ont été les mêmes que ceux utilisés pour la PCR classique et sont répertoriés
3179 dans le tableau IX. Ces amorces ont été dessinées *via* le logiciel primer3
3180 (<http://bioinfo.ut.ee/primer3-0.4.0/primer3/>) et leurs efficacités étaient comprises entre 97 et
3181 100%. Les gènes testés incluent un gène de référence (*rpl-18*), 2 gènes codant pour des facteurs
3182 NF-κB de deux voies immunitaires des insectes réagissant aux infections bactériennes (*dorsal*
3183 pour la voie Toll et *relish* pour la voie IMD), et 5 gènes codant pour des peptides de venin

3184 (*bicarinaline*, *p17*, *u3-Tb1a*, *u8-Tb1a* et *u9-Tb1a*). Les gènes *u4-Tb1a*, *u5-Tb1a*, *u7-Tb1a* et
 3185 *u16-Tb1a* testés précédemment ont été retirés de cette étude, les couples d'amorces n'étant pas
 3186 suffisamment efficaces pour cette technique. Le gène *actine* a également été retiré en raison de
 3187 son instabilité entre les conditions contrôle et infecté.

3188

3189 **Tableau IX** : Amorces utilisée pour la qPCR.

Nom de l'amorce	Séquence 5'-3'	Taille du fragment attendu (pb)
<i>rpl-18 F</i>	GTATCTTGAAAGCCGGTGG	180
<i>rpl-18 R</i>	CATACGTGAGGTCTAAGGGC	
<i>relish F</i>	GCACAAGCGACGTTTAACAGT	214
<i>relish R</i>	CGCCTTGGCATTCTCTCGTAC	
<i>dorsal F</i>	GCATACCCGGTGTCAATAGC	202
<i>dorsal R</i>	ACAACCATATTTCCCGATGG	
<i>bicarinaline F</i>	GTTGGTGCTTGCCATAATCC	180
<i>bicarinaline R</i>	TTTAACCTTGCCCCATGGTA	
<i>p17 F</i>	GATGCTGATGCTGCTGCTT	180
<i>p17 R</i>	GTTGTTGCAGTCCTCCAGT	
<i>u3-Tb1a F</i>	TCGTAGGTCTAAGTGGCTCA	155
<i>u3-Tb1a R</i>	TGGTAGACTCATCAAGCCCG	
<i>u8-Tb1a F</i>	TGGCAATCATGTACGCACC	150
<i>u8-Tb1a R</i>	ATCCCATGAATCCTTTTACTGCG	
<i>u9-Tb1a F</i>	ACCATATACGCACCCCAAGT	158
<i>u9-Tb1a R</i>	CCCAAAGCCATTTTCAAACCC	

3190

3191 Les échantillons ont été passés en triplicatas dans des plaques 384 puits avec un volume de
 3192 réaction de 6 μ L (SsoFastTM EvaGreen® Supermixes (1X final) (Biorad) ; amorces
 3193 (FORWARD & REVERSE) (0,5 μ M final) ; 2 μ L d'ADN).

3194 Les plaques 384 puits ont été analysées sur l'appareil ViiA7 (Applied Biosystem) sur la
 3195 plateforme Génome et Transcriptome (GeT) du Génomopôle de Toulouse. Les cycles seuils (C_T
 3196 pour Cycle Threshold) de chaque échantillon, déterminant le début de la phase exponentielle
 3197 de l'amplification de la PCR, ont été évalués grâce au logiciel QuantStudioTM (v 1.3) (Applied
 3198 Biosystems, Thermofisher Scientific).

3199 Les ratios d'expression de nos gènes d'intérêt ont ensuite été calculés grâce à la méthode
 3200 du $2^{-\Delta\Delta C_T}$ (Livak & Schmittgen, 2001). Il s'agit d'une méthode de quantification relative qui
 3201 permet de calculer les changements relatifs d'expression de gènes d'intérêts par rapport à un ou
 3202 plusieurs gènes de références, ici *rpl-18*, selon les conditions testées et cela sans réaliser de

3203 droite standard et sans connaissance de la valeur absolue du nombre de copies dans chaque
3204 échantillon. La formule pour calculer le $2^{-\Delta\Delta C_T}$, c'est-à-dire le ratio relatif d'expression d'un
3205 gène d'intérêt pour notre condition infecté par rapport à notre condition contrôle, est la
3206 suivante :

3207

$$3208 \quad 2^{-\Delta\Delta C_T} = 2^{-\left(C_{T \text{ gène d'intérêt}} - C_{T \text{ gène de référence}}\right)_{\text{échantillon}} - \left(C_{T \text{mean gène d'intérêt}} - C_{T \text{mean gène de référence}}\right)_{\text{contrôle}}}$$

3209

3210 Où $C_{T \text{mean}}$ est la moyenne des C_T obtenus pour une condition donnée.

3211

3212 Les conditions de normalité de d'homoscédasticité n'étant pas respectées, le traitement
3213 statistique des données a été réalisé via des tests de Mann-Whitney entre les ratios d'expression
3214 de nos deux conditions par gène et par tissus. Une comparaison des C_T des différents gènes
3215 entre les deux conditions a également été réalisée grâce à des tests de Mann-Whitney. Ces tests
3216 ont été effectués à l'aide du logiciel R (v 4.0.0) (R Core Team, 2017).

3217

3218 > Protocole de digital droplet PCR (ddPCR)

3219 La technique de la dd-PCR consiste en la séparation préalable de l'échantillon en
3220 gouttelettes (Hindson et al., 2011). Chaque gouttelette soutient une amplification de la molécule
3221 cible qu'elle contient. A la fin de la réaction, chaque gouttelette est analysée pour déterminer
3222 la fraction de gouttes « PCR-positives », déterminant ainsi la concentration d'ADN cible dans
3223 l'échantillon. Cette technique permet donc une quantification absolue en fin de réaction PCR,
3224 tout en laissant la possibilité de s'affranchir d'un gène de référence et en détectant les faibles
3225 variations d'expressions génétiques.

3226 Un échantillon de glandes à venin de fourmis non-infectées a été choisi comme témoin
3227 positif quant à l'expression des gènes codant les peptides ciblés. Parmi les échantillons passés
3228 en qPCR, dix échantillons par condition et par tissu ont été sélectionnés de manière aléatoire.
3229 Parmi les neuf gènes testés en qPCR, cinq ont été sélectionnés. Le gène de référence *rpl-18* a
3230 été testé afin de vérifier sa stabilité entre les conditions en quantification absolue. Le gène *relish*
3231 a été choisi en raison de son implication dans la voie IMD, connue pour réagir face aux bactéries
3232 à Gram négatif (e.g. *Serratia marcescens*). Les gènes *p17*, *u3-Tb1a* et *u8-Tb1a* ont été
3233 sélectionnés car ils étaient exprimés dans la majorité des échantillons testés en qPCR,
3234 contrairement aux gènes *bicarinaline* et *u9-Tb1a*.

3235 Les échantillons ont été passés dans des plaques 96 puits avec un volume de réaction de
3236 22 µL (QX200™ ddPCR™ EvaGreen Supermix (1X final) (Bio-Rad) ; amorces (FORWARD
3237 & REVERSE) (0,1 µM final) ; 2,2 µL d'ADN). Les couples d'amorces utilisées pour quantifier
3238 l'expression des gènes *rpl-18*, *relish*, *p17*, *u3-Tb1a* et *u8-Tb1a* sont les mêmes que ceux utilisés
3239 pour la qPCR (Tableau IX).

3240 Les plaques 96 puits ont été analysées sur l'appareil Q200X AutoDG (Bio-Rad) sur la
3241 plateforme Génome et Transcriptome (GeT) du Génomipôle de Toulouse. Le nombre de copies
3242 d'ADN obtenu par gène a été évalué grâce au logiciel Quantasoft™ (v 1.7).

3243 Les conditions de normalité de d'homoscédasticité n'étant pas respectées, le traitement
3244 statistique des données a été réalisé *via* des tests de Mann-Whitney pour comparer les nombres
3245 de copies d'ADN détectées dans les échantillons de deux conditions par gène et par tissus. Ces
3246 tests ont été effectués à l'aide du logiciel R (v 4.0.0) (R Core Team, 2017).

3247

3248 4.4.3. Etude au niveau protéique

3249 Afin de vérifier l'expression des gènes codant pour la Bicarinaline et le P17 au niveau
3250 protéique suite à un challenge bactérien, un protocole d'immunohistochimie indirecte a été
3251 appliqué sur les tissus d'abdomens d'individus sains et infectés par *Serratia marcescens*.

3252 Treize ouvrières *T. bicarinatum* ont été isolées par condition (i.e. contrôle et infecté)
3253 puis infectées par voie orale comme décrit dans la partie 4.4.1. Vingt-quatre heures après
3254 infection, les fourmis ont été disséquées comme précédemment décrit dans la partie 4.2. Les
3255 abdomens ont été isolés du reste du corps et les tergites T1 et T2 ont été retirés. Un total de cinq
3256 échantillons a été constitué, contenant chacun 5 abdomens de *T. bicarinatum* : un témoin relatif
3257 à l'anticorps secondaire et un échantillon par peptide et par condition. Chaque échantillon a été
3258 directement placé dans 1 mL de solution de fixation avant de procéder au protocole
3259 d'immunohistochimie indirecte décrit dans la partie 4.3.2. Cette expérimentation a été effectuée
3260 deux fois. Au total, 52 fourmis ont été nécessaires à sa réalisation.

3261 Pour chaque type de tissus, cinq photos par organe ciblé et par échantillon ont été
3262 réalisées grâce à l'appareil Cytation 1 (Biotek) et au logiciel Gen5™. Les intensités lumineuses
3263 ont été mesurées grâce au logiciel ImageJ™ pour chaque organe en prenant en compte leur
3264 surface. Pour chaque tissu, la valeur réelle d'intensité lumineuse a été calculée en lui
3265 soustrayant la valeur de l'intensité lumineuse des organes correspondants dans le témoin de
3266 fluorescence des anticorps secondaires. Les intensités lumineuses corrigées des conditions
3267 contrôle et infecté ont été comparées grâce au test t de Student. Ces tests ont été effectués à
3268 l'aide du logiciel R (v 4.0.0) (R Core Team, 2017).

3269

3270 *4.5. Protocoles utilisés lors de l'étude préliminaire*

3271 *4.5.1. Première localisation des transcrits codant pour la Bicarinaline et le P17 en*
3272 *dehors de l'abdomen*

3273 Dix ouvrières de *T. bicarinatum* ont été isolées en boîte de Pétri pendant 24h puis disséquées
3274 comme précédemment décrit dans la partie 4.2. Six types de tissus ont été isolés et placés par
3275 lots de 5 : des appareils vulnérants (glandes à venin et réservoir), des mésentérons, des tubes de
3276 Malpighi, des rectums, des corps gras et des abdomens dépourvus d'appareil vulnérant.

3277

3278 *4.5.2. Première estimation de la réponse des gènes bicarinaline et p17 suite à une*
3279 *infection bactérienne*

3280 Soixante fourmis ont été isolées et réparties en deux lots (i.e. contrôle et infecté), puis
3281 infectées comme précédemment décrit dans la partie 4.4.1. L'infection a été répétée les deux
3282 jours suivants, permettant ainsi de réaliser une infection en continu sur 72h. Aucune mortalité
3283 n'a été observée au cours de cette partie de l'étude.

3284 Des dissections ont été effectuées comme décrit dans la partie 4.2 sur 3 fourmis de
3285 chaque condition, et ce 24h, 48h et 72h après le début de l'infection. Deux types de tissus ont
3286 été isolés et placés par lots de trois pour chaque condition : des glandes à venin et des abdomens
3287 dépourvus d'appareil vulnérant. Un total de quatre échantillons a donc été obtenu par pas de
3288 temps. Cette expérimentation a été répétée 3 fois, nécessitant un total de 54 fourmis pour sa
3289 réalisation.

3290

3291 *4.5.3. Extraction, traitement de l'ARN et rétrotranscription*

3292 Les extractions et traitements de l'ARN ont été réalisés immédiatement après les
3293 dissections, et les rétrotranscriptions ont été effectuées sur 50 ng d'ARNs totaux comme
3294 précédemment décrit dans la partie 4.3.1.

3295

3296 *4.5.4. Amplification par PCR*

3297 Des PCRs ont été réalisées à 60°C pendant 35 cycles sur les échantillons des deux
3298 expérimentations. La première expérience nous a permis d'obtenir une estimation plus précise
3299 des lieux d'expression des gènes codant la Bicarinaline et le P17, tandis que la seconde
3300 expérimentation nous a permis d'effectuer une première évaluation de la réponse de ces gènes
3301 face à une infection bactérienne.

3302 Quatre gènes ont été testés dans le cadre de la première manipulation : *rpl-18* et *actine*
 3303 (i.e. gènes de référence), *bicarinaline* et *p17*. Les mêmes gènes ont été testés dans le cadre de
 3304 la seconde expérimentation, plus deux gènes impliqués dans les voies Toll et IMD : *dorsal* et
 3305 *relish*. Les couples d’amorces utilisés sont répertoriés dans le tableau X.

3306

3307 **Tableau X** : Couples d’amorces utilisés pour les PCR et la qPCR.

Amorces	Séquences	Taille des fragments attendus
<i>rpl-18</i> F	GTATCTTGAAAGCCGGTGG	180 pb
<i>rpl-18</i> R	CATACGTGAGGTCTAAGGGC	
<i>actine</i> F	GGCTTCTTCCTCCAGTTTGG	192 pb
<i>actine</i> R	TACAAGTCCTTGCGGATGTC	
<i>dorsal</i> F	GCATACCCGGTGTCAATAGC	202 pb
<i>dorsal</i> R	ACAACCATATTTCCCGATGG	
<i>relish</i> F	GGCGCTGATCCGTATTTAAG	196 pb
<i>relish</i> R	TGTCATCGGAAGTCTGACCA	
<i>bicarinaline</i> F	GTTGGTGCTTGCCATAATCC	180 pb
<i>bicarinaline</i> R	TTTAACCTTGCCCCATGGTA	
<i>p17</i> F	GATGCTGATGCTGCTGCTT	180 pb
<i>p17</i> R	GTTGTTGCAGTCCTCCAGT	

3308

3309

3310 4.5.5. Protocole de qPCR

3311 L’expression des gènes de l’immunité et ceux codant les peptides Bicarinaline et P17 a
 3312 également été évaluée par qPCR, selon le protocole décrit dans la partie 4.4.2. Les mêmes gènes
 3313 ont été testés, à l’exception de *relish* dont les couples d’amorces n’étaient pas suffisamment
 3314 efficaces pour cette technique. Ainsi, afin d’avoir un représentant de la voie IMD, le gène *pgrp-*
 3315 *le*, codant une protéine de reconnaissance du peptidoglycane, composé principal des
 3316 membranes bactériennes et activatrice de la voie IMD, a été utilisé par la suite (Bier and
 3317 Guichard 2012). Le couple d’amorces utilisé est répertorié dans le tableau XI.

3318

3319 **Tableau XI** : Couple d’amorces utilisé pour le gène *pgrp-le* en qPCR.

Amorces	Séquences	Taille du fragment attendu
<i>pgrp-le</i> F	ATAACATTGGCGGAGCTCAA	218 pb
<i>pgrp-le</i> R	ACCATCAACCGATGGAATTT	

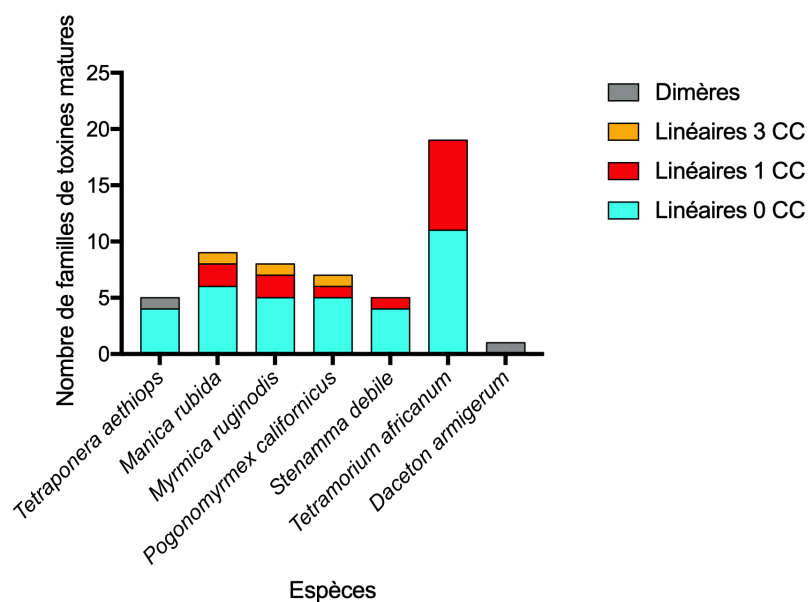
Conclusion générale / Perspectives

3320

3321

3322 Les objectifs généraux de ce travail de thèse consistaient en (i) l'apport de premiers
3323 aperçus sur la diversité moléculaire des toxines peptidiques contenues dans les venins de
3324 fourmis myrmicines, et (ii) l'exploration du lien potentiel entre les peptides du venin et le
3325 système immunitaire inné de la fourmi *T. bicarinatum* à titre individuel.

3326 Afin de répondre au premier objectif, la caractérisation des peptidomes des venins de
3327 sept fourmis appartenant à la sous-famille des Myrmicinae et d'une fourmi appartenant à la
3328 sous-famille des Pseudomyrmecinae a été effectuée au moyen d'approches
3329 protéotranscriptomiques. Ces travaux ont mené à l'identification de 100 précurseurs de toxines
3330 peptidiques (i.e. 9 pseudomyrmecitoxines et 91 myrmicitoxines). Les compositions des venins
3331 étudiés sont dominées par des peptides linéaires, cationiques et amphiphiles, excepté pour la
3332 fourmi myrmicine *Daceton armigerum* dont le venin est composé exclusivement de peptides
3333 dimériques. La présence de toxines dimériques a également été notée pour *Tetraoponera*
3334 *aethiops* (*Pseudomyrmecinae*). Les venins de fourmis myrmicines contiennent aussi plusieurs
3335 peptides à ponts disulfures intracaténaux (i.e. 1 à 3). Sur la base de leurs séquences matures,
3336 les myrmicitoxines ont été classées en 37 familles. Certaines d'entre elles ont été observées
3337 dans la plupart des venins de fourmis myrmicines caractérisées (i.e. U₃, U₁₇), suggérant
3338 l'importance de leurs rôles fonctionnels. Ces résultats nous ont permis de mettre en évidence
3339 l'hétérogénéité de la composition en toxines peptidiques des venins des fourmis myrmicines
3340 étudiées en termes de diversité structurale (Figure 53), suggérant l'existence d'un compromis
3341 entre le nombre de famille de toxines peptidiques matures présentes dans un venin donné et la
3342 complexité structurale de ces dernières. Il est également à noter que les compositions des venins
3343 pourraient varier selon leurs contextes majoritaires d'utilisation (i.e. défense et/ou prédation).
3344 La diversité des compositions de venins de fourmis myrmicines semble donc résulter d'une
3345 multitude de paramètres, soulignant le besoin d'effectuer davantage d'études intégratives sur
3346 des échantillonnages rationnels et de déterminer les activités biologiques des toxines.
3347 Cependant, cette détermination d'activité biologique devrait prendre en compte un potentiel
3348 effet synergique des peptides contenus dans un même venin.



3349
 3350 **Figure 53** : Caractéristiques structurales des familles de toxines peptidiques matures caractérisées dans les
 3351 venins de fourmis myrmicines et pseudomyrmecine.

3352 L'analyse des séquences signal des précurseurs de myrmecitoxines a permis leur
 3353 répartition en 8 superfamilles de précurseurs. Nous avons également montré que les séquences
 3354 signal et régions propeptides des toxines des venins de *Tetraponera aethiops*, *Odontomachus*
 3355 *monticola* (*Ponerinae*), *Myrmecia pilosula* et *M. gulosa* (*Myrmeciinae*) montraient des
 3356 similarités de séquence avec celles des précurseurs de toxines de la superfamille A chez
 3357 *Tetramorium bicarinatum* (Barassé et al., 2019; Touchard et al., 2018). Robinson *et al.* (2018)
 3358 ont de plus montré que les précurseurs des toxines peptidiques du venin de *M. gulosa* possèdent
 3359 des séquences signal similaires à celles de peptides contenus dans les venins d'Hyménoptères
 3360 autres que les Formicidae, tout en montrant des séquences matures divergentes et exerçant donc
 3361 potentiellement des activités biologiques différentes. Ces mêmes auteurs ont ainsi émis
 3362 l'hypothèse que la majorité des toxines de venins d'Hyménoptères aculéates appartiendraient à
 3363 une même famille multigénique (Robinson et al., 2018). Des données génomiques sont
 3364 néanmoins nécessaires pour confirmer cette hypothèse, ou à minima des analyses sur les
 3365 séquences nucléotidiques des transcrits codant pour ces précurseurs (Walker, 2020).

3366
 3367 Afin de répondre au second objectif, des techniques de biologie moléculaire (i.e. RT-
 3368 PCR, qPCR et ddPCR) ont été employées dans le but de vérifier l'expression des gènes codant
 3369 des toxines du venin de *Tetramorium bicarinatum* en dehors du système vulnérant, ainsi que
 3370 d'évaluer leur réponse suite à une infection bactérienne par voie orale. La présence des peptides
 3371 Bicarinaline et P17 a également été évaluée par immunohistochimie indirecte dans les différents

3372 organes composant l'abdomen de la fourmi suite à ce même challenge immunitaire. Ces travaux
3373 ont mené à la localisation des transcrits codant les peptides testés dans des organes connus pour
3374 être impliqués dans l'immunité innée des insectes (i.e. mésentérons, tubes de Malpighi, corps
3375 gras). La Bicarinaline et le P17 ont également été localisés dans ces organes, notamment dans
3376 un des deux types cellulaires composant le corps gras de notre modèle biologique (i.e.
3377 oenocytes). L'immunohistochimie indirecte nous a de plus permis de mettre en évidence une
3378 production de ces peptides dans les deux types cellulaires du corps gras suite à une infection
3379 bactérienne. Nous posons également l'hypothèse que ces peptides seraient sécrétés par les tubes
3380 de Malpighi en cas de challenge immunitaire, dans la mesure où les signaux traduisant la
3381 présence des peptides dans ces organes présentaient moins d'intensité chez les individus
3382 infectés.

3383 Aucune réponse nette des gènes codant les peptides du venin n'a en revanche été décelée
3384 *via* l'analyse traditionnelle des résultats de RT-qPCR (i.e. $2^{-\Delta\Delta C_T}$) suite au challenge
3385 immunitaire. Nous avons montré par la suite que le gène de référence utilisé (i.e. *rpl-18*) n'était
3386 pas stable grâce à la technique plus sensible de ddPCR. De la même manière, cette technique
3387 ne nous a pas permis d'observer de surexpression significative des gènes codant les peptides en
3388 raison d'une trop grande variabilité entre les échantillons d'une même condition. Outre le
3389 recours à des méthodes *in vitro* basées sur la culture primaire des cellules du corps gras, le
3390 séquençage du génome de *T. bicarinatum* (actuellement en cours) apparaît comme essentiel
3391 pour la poursuite de cette étude.

3392 Suite à une annotation complète, ces données nous donneraient en effet des informations
3393 sur la régulation des gènes codant les peptides testés (i.e. promoteurs), nous permettant
3394 potentiellement de les relier à une voie de signalisation cellulaire. Il sera également possible
3395 d'identifier plus précisément les composants de l'immunité innée de cette fourmi, tels que les
3396 protéines impliquées dans les voies immunitaires ainsi que les Host-Defense-Peptides circulant
3397 dans l'hémolymphe. Des approches plus ciblées seront ensuite envisageables telles que la
3398 technique du RNAseq en différentiel sur différent types de tissus.

3399 L'acquisition du génome nous permettra de plus de regarder la position des gènes codant
3400 les toxines peptidiques du venin de *T. bicarinatum*, rendant également possible la détermination
3401 des mécanismes évolutifs impliqués dans leur diversification.

Références bibliographiques

- 3403
- 3404 Aanen, D. K. (2018). Social Immunity: the disposable individual. *Current Biology*, 28(7),
3405 R322–R324. <https://doi.org/10.1016/j.cub.2018.02.050>
- 3406 Agaisse, H., & Perrimon, N. (2004). The roles of JAK/STAT signaling in *Drosophila* immune
3407 responses. *Immunological Reviews*, 198(1), 72–82. <https://doi.org/10.1111/j.0105->
3408 2896.2004.0133.x
- 3409 Aili, S. R., Touchard, A., Escoubas, P., Padula, M. P., Orivel, J., Dejean, A., & Nicholson, G.
3410 M. (2014). Diversity of peptide toxins from stinging ant venoms. *Toxicon*, 92, 166–178.
3411 <https://doi.org/10.1016/j.toxicon.2014.10.021>
- 3412 Aili, S. R., Touchard, A., Hayward, R., Robinson, S. D., Pineda, S. S., Lalagüe, H., ...
3413 Nicholson, G. M. (2020). An integrated proteomic and transcriptomic analysis reveals the
3414 venom complexity of the bullet ant *Paraponera clavata*. *Toxins*, 12(324).
3415 <https://doi.org/10.3390/toxins12050324>
- 3416 Akbari, R., Hakemi Vala, M., Hashemi, A., Aghazadeh, H., Sabatier, J. M., & Pooshang
3417 Bagheri, K. (2018). Action mechanism of melittin-derived antimicrobial peptides, MDP1
3418 and MDP2, de novo designed against multidrug resistant bacteria. *Amino Acids*, 50(9),
3419 1231–1243. <https://doi.org/10.1007/s00726-018-2596-5>
- 3420 Ali Mohammadie Kojour, M., Han, Y. S., & Jo, Y. H. (2020). An overview of insect innate
3421 immunity. *Entomological Research*, 50, 282–291. <https://doi.org/10.1111/1748-5967>
- 3422 Allcock, A. L., Arbuckle, K., Baracchi, D., Caruana, N., Casewell, N. R., Cooke, I. R., ... Zhu,
3423 S. (2017). *Evolution of venomous animals and their toxins*. (P. Gopalakrishnakone & A.
3424 Malhotra, Eds.). Springer Nature. Retrieved from <http://www.springer.com/series/13330>
- 3425 Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O.,
3426 Brunak, S., ... Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using
3427 deep neural networks. *Nature Biotechnology*, 37(4), 420–423.
3428 <https://doi.org/10.1038/s41587-019-0036-z>
- 3429 Arbuckle, K. (2017). Evolutionary context of venomous animals. In P. Gopalakrishnakone &
3430 A. Malhotra (Eds.), *Evolution of Venomous Animals and Their Toxins* (Toxinology, pp.
3431 1–36). Dordrecht: Springer.
- 3432 Ashwood, L. M., Norton, R. S., Undheim, E. A. B., Hurwood, D. A., & Prentis, P. J. (2020).
3433 Characterising functional venom profiles of Anthozoans and Medusozoans within their
3434 ecological context. *Marine Drugs*, 18(202). <https://doi.org/10.3390/md18040202>
- 3435 Babb, P. L., Lahens, N. F., Correa-Garhwal, S. M., Nicholson, D. N., Kim, E. J., Hogenesch, J.

3436 B., ... Voight, B. F. (2017). The *Nephila clavipes* genome highlights the diversity of spider
3437 silk genes and their complex expression. *Nature Genetics*, 49(6), 895–903.
3438 <https://doi.org/10.1038/ng.3852>

3439 Bagnall, N. H., & Kotze, A. C. (2010). Evaluation of reference genes for real-time PCR
3440 quantification of gene expression in the Australian sheep blowfly, *Lucilia cuprina*.
3441 *Medical and Veterinary Entomology*, 24(2), 176–181. [https://doi.org/10.1111/j.1365-
3442 2915.2010.00866.x](https://doi.org/10.1111/j.1365-2915.2010.00866.x)

3443 Bang, S. (2019). JAK/STAT signaling in insect innate immunity. *Entomological Research*, 47.
3444 <https://doi.org/10.1111/1748-5967>

3445 Baracchi, D., Francese, S., & Turillazzi, S. (2011). Beyond the antipredatory defence: honey
3446 bee venom function as a component of social immunity. *Toxicon*, 58(6–7), 550–557.
3447 <https://doi.org/10.1016/j.toxicon.2011.08.017>

3448 Baracchi, D., & Tragust, S. (2020). Venom as a component of external immune defense in
3449 Hymenoptera. In *Evolution of Venomous Animals and Their Toxins*.

3450 Barassé, V., Touchard, A., Téné, N., Tindo, M., Kenne, M., Klopp, C., ... Treilhou, M. (2019).
3451 The peptide venom composition of the fierce stinging ant *Tetraponera aethiops*
3452 (Formicidae: Pseudomyrmecinae). *Toxins*, 11(732).
3453 <https://doi.org/10.3390/toxins11120732>

3454 Baumann, K., Dashevsky, D., Sunagar, K., & Fry, B. (2018). Scratching the surface of an itch:
3455 molecular evolution of Aculeata venom allergens. *Journal of Molecular Evolution*, 86(7),
3456 484–500. <https://doi.org/10.1007/s00239-018-9860-x>

3457 Beeton, C., Pennington, M. W., Wulff, H., Singh, S., Nugent, D., Crossley, G., ... Chandy, K.
3458 G. (2005). Targeting effector memory T cells with a selective peptide inhibitor of Kv1.3
3459 channels for therapy of autoimmune diseases. *Molecular Pharmacology*, 67(4), 1369–
3460 1381. <https://doi.org/10.1124/mol.104.008193>

3461 Benmoussa, K., Authier, H., Prat, M., AlaEddine, M., Lefèvre, L., Rahabi, M. C., ... Coste, A.
3462 (2017). P17, an original host defense peptide from ant venom, promotes antifungal
3463 activities of macrophages through the Induction of C-type lectin receptors dependent on
3464 LTB4-mediated PPAR γ activation. *Frontiers in Immunology*, 8(November).
3465 <https://doi.org/10.3389/fimmu.2017.01650>

3466 Bier, E., & Guichard, A. (2012). Deconstructing host-pathogen interactions in *Drosophila*.
3467 *Disease Models & Mechanisms*, 5(1), 48–61. <https://doi.org/10.1242/dmm.000406>

3468 Blaimer, B. B., Ward, P. S., Schultz, T. R., Fisher, B. L., & Brady, S. G. (2018). Paleotropical
3469 diversification dominates the evolution of the hyperdiverse ant tribe Crematogastrini

3470 (Hymenoptera: Formicidae). *Insect Systematics and Diversity*, 2(5).
3471 <https://doi.org/10.1093/isd/ixy013>

3472 Borges, A., Lomonte, B., Angulo, Y., Acosta, H., Patiño, D., Pascale, J. M., ... Caro-lópez, J.
3473 A. (2020). Venom diversity in the neotropical scorpion genus *Tityus*: implications for
3474 antivenom design emerging from molecular and immunochemical analyses across
3475 endemic areas of scorpionism. *Acta Tropica*, 204, 105346.
3476 <https://doi.org/10.1016/j.actatropica.2020.105346>

3477 Borowiec, M. L., Moreau, C. S., & Rabeling, C. (2020). Ants: phylogeny and classification. In
3478 Starr C. (Ed.), *Encyclopedia of Social Insects* (Springer,). [https://doi.org/10.1007/978-3-](https://doi.org/10.1007/978-3-319-90306-4_155-1)
3479 [319-90306-4_155-1](https://doi.org/10.1007/978-3-319-90306-4_155-1)

3480 Bouzid, W., Verdenaud, M., Klopp, C., Ducancel, F., Noirot, C., & Vétillard, A. (2014). De
3481 Novo sequencing and transcriptome analysis for *Tetramorium bicarinatum*: a
3482 comprehensive venom gland transcriptome analysis from an ant species. *BMC Genomics*,
3483 15(987). <https://doi.org/10.1186/1471-2164-15-987>

3484 Brady, D., Grapputo, A., Romoli, O., & Sandrelli, F. (2019). Insect cecropins, antimicrobial
3485 peptides with potential therapeutic applications. *International Journal of Molecular*
3486 *Sciences*, 20(23). <https://doi.org/10.3390/ijms20235862>

3487 Branstetter, M. G. (2012). Origin and diversification of the cryptic ant genus *Stenamma*
3488 Westwood (Hymenoptera: Formicidae), inferred from multilocus molecular data,
3489 biogeography and natural history. *Systematic Entomology*, 37(3), 478–496.
3490 <https://doi.org/10.1111/j.1365-3113.2012.00624.x>

3491 Brüttsch, T., Jaffuel, G., Vallat, A., Turlings, T. C. J., & Chapuisat, M. (2017). Wood ants
3492 produce a potent antimicrobial agent by applying formic acid on tree-collected resin.
3493 *Ecology and Evolution*, 7(7), 2249–2254. <https://doi.org/10.1002/ece3.2834>

3494 Cabau, C., Escudié, F., Djari, A., Guiguen, Y., Bobe, J., & Klopp, C. (2017). Compacting and
3495 correcting Trinity and Oases RNA-Seq de novo assemblies. *PeerJ*, 5, e2988.
3496 <https://doi.org/10.7717/peerj.2988>

3497 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T.
3498 L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(421).
3499 <https://doi.org/10.1186/1471-2105-10-421>

3500 Casewell, N. R., Wüster, W., Vonk, F. J., Harrison, R. A., & Fry, B. G. (2013). Complex
3501 cocktails: the evolutionary novelty of venoms. *Trends in Ecology and Evolution*, 28(4),
3502 219–229. <https://doi.org/10.1016/j.tree.2012.10.020>

3503 Charroux, B., & Royet, J. (2010). *Drosophila* immune response: from systemic antimicrobial

3504 peptide production in fat body cells to local defense in the intestinal tract. *Fly*, 4(1), 40–
3505 47. <https://doi.org/10.4161/fly.4.1.10810>

3506 Cheng, D., Zhang, Z., He, X., & Liang, G. (2013). Validation of reference genes in *Solenopsis*
3507 *invicta* in different developmental stages, castes and tissues. *PLoS ONE*, 8(2), 1–6.
3508 <https://doi.org/10.1371/journal.pone.0057718>

3509 Chippaux, J. P., Williams, V., & White, J. (1991). Snake venom variability: methods of study,
3510 results and interpretation. *Toxicon*, 29(11), 1279–1303. [https://doi.org/10.1016/0041-](https://doi.org/10.1016/0041-0101(91)90116-9)
3511 [0101\(91\)90116-9](https://doi.org/10.1016/0041-0101(91)90116-9)

3512 Clark, G. C., Casewell, N. R., Elliott, C. T., Harvey, A. L., Jamieson, A. G., Strong, P. N., &
3513 Turner, A. D. (2019). Friends or foes? Emerging impacts of biological toxins. *Trends in*
3514 *Biochemical Sciences*, xx, 1–15. <https://doi.org/S096800041830269X>

3515 Cologna, C. T., Cardoso, J. dos S., Jourdan, E., Degueldre, M., Upert, G., Gilles, N., ...
3516 Quinton, L. (2013). Peptidomic comparison and characterization of the major components
3517 of the venom of the giant ant *Dinoponera quadriceps* collected in four different areas of
3518 Brazil. *Journal of Proteomics*, 94, 413–422. <https://doi.org/10.1016/j.jprot.2013.10.017>

3519 Cooper, D. M., Chamberlain, C. M., & Lowenberger, C. (2009). Aedes FADD: A novel death
3520 domain-containing protein required for antibacterial immunity in the yellow fever
3521 mosquito, *Aedes aegypti*. *Insect Biochemistry and Molecular Biology*, 39(1), 47–54.
3522 <https://doi.org/10.1016/j.ibmb.2008.09.011>

3523 Corporation, V. (n.d.). EPA removes bee toxicity warning statement from label of SPEAR™
3524 bioinsecticide from Vestaron Corporation. Retrieved from
3525 https://vestaron.com/assets/vestaron_beetoxicity_12.17.15.pdf

3526 Costa, A., Jan, E., Sarnow, P., & Schneider, D. (2009). The Imd pathway is involved in antiviral
3527 immune responses in *Drosophila*. *PLoS ONE*, 4(10).
3528 <https://doi.org/10.1371/journal.pone.0007436>

3529 Coutinho das Neves, R., Mortari, M. R., Ferroni Schwartz, E., Kipnis, A., & Junqueira-Kipnis,
3530 A. P. (2019). Antimicrobial and antibiofilm effects of peptides from venom of social wasp
3531 and scorpion on multidrug-resistant *Acinetobacter baumannii*. *Toxins*, 11(216).
3532 <https://doi.org/10.3390/toxins11040216>

3533 Cremer, S. (2019). Social immunity in insects. *Current Biology*, 29(11).
3534 <https://doi.org/10.1016/j.cub.2019.03.035>

3535 Cremer, S., Armitage, S. A. O., & Schmid-Hempel, P. (2007). Social Immunity. *Current*
3536 *Biology*, 17(16), 693–702. <https://doi.org/10.1016/j.cub.2007.06.008>

3537 Daly, N. L., & Wilson, D. (2018). Structural diversity of arthropod venom toxins. *Toxicon*, 152,

3538 46–56. <https://doi.org/10.1016/j.toxicon.2018.07.018>

3539 Dejean, A., Azémar, F., Céréghino, R., Leponce, M., Corbara, B., Orivel, J., & Compin, A.
3540 (2016). The dynamics of ant mosaics in tropical rainforests characterized using the Self-
3541 Organizing Map algorithm. *Insect Science*, 23(4), 630–637. <https://doi.org/10.1111/1744-7917.12208>

3542

3543 Dejean, A., Céréghino, R., Leponce, M., Rossi, V., Roux, O., Compin, A., ... Corbara, B.
3544 (2015). The fire ant *Solenopsis saevissima* and habitat disturbance alter ant communities.
3545 *Biological Conservation*, 187, 145–153. <https://doi.org/10.1016/j.biocon.2015.04.012>

3546 Dejean, A., Delabie, J. H. C., Corbara, B., Azémar, F., Groc, S., Orivel, J., & Leponce, M.
3547 (2012). The ecology and feeding habits of the arboreal trap-jawed ant *Daceton armigerum*.
3548 *PLoS ONE*, 7(6). <https://doi.org/10.1371/journal.pone.0037683>

3549 Dekan, Z., Headey, S. J., Scanlon, M., Baldo, B. A., Lee, T. H., Aguilar, M. I., ... Alewood, P.
3550 F. (2017). Δ -Myrtoxin-Mp1a is a helical heterodimer from the venom of the jack jumper
3551 ant that has antimicrobial, membrane-disrupting, and nociceptive activities. *Angewandte*
3552 *Chemie - International Edition*, 56(29), 8495–8499.
3553 <https://doi.org/10.1002/anie.201703360>

3554 Dodou Lima, H. V., de Paula Calvacante, C. S., & Radis-Baptista, G. (2020). Antifungal in
3555 vitro activity of Pilosulin- and Ponericin-Like Peptides from the giant ant *Dinoponera*
3556 *quadriceps* and synergistic effects with antimycotic drugs. *Antibiotics*, 9(354).
3557 <https://doi.org/10.3390/antibiotics9060354>

3558 Dos Santos-Pinto, J. R. A., Perez-Riverol, A., Lasa, A. M., & Palma, M. S. (2018). Diversity
3559 of peptidic and proteinaceous toxins from social Hymenoptera venoms. *Toxicon*, 148,
3560 172–196. <https://doi.org/10.1016/j.toxicon.2018.04.029>

3561 Dos Santos Pinto, J. R. A., Fox, E. G. P., Saidemberg, D. M., Santos, L. D., Da Silva
3562 Menegasso, A. R., Costa-Manso, E., ... Palma, M. S. (2012). Proteomic view of the venom
3563 from the fire ant *Solenopsis invicta* buren. *Journal of Proteome Research*, 11(9), 4643–
3564 4653. <https://doi.org/10.1021/pr300451g>

3565 Drukewitz, S. H., Bokelmann, L., Undheim, E. A. B., & von Reumont, B. M. (2019). Toxins
3566 from scratch? Diverse, multimodal gene origins in the predatory robber fly *Dasypogon*
3567 *diadema* indicate a dynamic venom evolution in dipteran insects. *GigaScience*, 8(7), 1–
3568 13. <https://doi.org/10.1093/gigascience/giz081>

3569 Drukewitz, S. H., Fuhrmann, N., Undheim, E. A. B., Blanke, A., Giribaldi, J., Mary, R., ... von
3570 Reumont, B. M. (2018). A dipteran's novel sucker punch: evolution of arthropod atypical
3571 venom with a neurotoxic component in robber flies (Asilidae, Diptera). *Toxins*, 10(1).

3572 <https://doi.org/10.3390/toxins10010029>

3573 Dufton, M. J. (1992). Venomous mammals. *Pharmacology and Therapeutics*, 53(2), 199–215.

3574 [https://doi.org/10.1016/0163-7258\(92\)90009-O](https://doi.org/10.1016/0163-7258(92)90009-O)

3575 Duraisamy, K., Singh, K., KUMar, M., Lefranc, B., Bonnafé, E., Treilhou, M., ... Chow, B. K.

3576 C. (2020). The novel venom peptide P17 induces chemotactic recruitment and

3577 differentiation of monocytes into macrophages via alternative mast cell activation. *Journal*

3578 *of Chemical Information and Modeling*. (in submission)

3579 Dutertre, S., Jin, A. H., Vetter, I., Hamilton, B., Sunagar, K., Lavergne, V., ... Lewis, R. J.

3580 (2014). Evolution of separate predation-and defence-evoked venoms in carnivorous cone

3581 snails. *Nature Communications*, 5. <https://doi.org/10.1038/ncomms4521>

3582 Ekengren, S., & Hultmark, D. (1999). *Drosophila* cecropin as an antifungal agent. *Insect*

3583 *Biochemistry and Molecular Biology*, 29(11), 965–972. [https://doi.org/10.1016/S0965-](https://doi.org/10.1016/S0965-1748(99)00071-5)

3584 [1748\(99\)00071-5](https://doi.org/10.1016/S0965-1748(99)00071-5)

3585 Evans, J. D., Aronstein, K., Chen, Y. P., Hetru, C., Imler, J. L., Jiang, H., ... Hultmark, D.

3586 (2006). Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect*

3587 *Molecular Biology*, 15(5), 645–656. <https://doi.org/10.1111/j.1365-2583.2006.00682.x>

3588 Ferrandon, D., Jung, A. C., Cricqui, M. C., Lemaitre, B., Uttenweiler-Joseph, S., Michaut, L.,

3589 ... Hoffmann, J. A. (1998). A drosomycin-GFP reporter transgene reveals a local immune

3590 response in *Drosophila* that is not dependent on the Toll pathway. *EMBO Journal*, 17(5),

3591 1217–1227. <https://doi.org/10.1093/emboj/17.5.1217>

3592 Fox, E. G. P., Bueno, O. C., Yabuki, A. T., de Jesus, C. M., Solis, D. R., Rossi, M. L., & de

3593 Lima Nogueira, N. (2010). General morphology and ultrastructure of the venom apparatus

3594 and convoluted gland of the fire ant, *Solenopsis saevissima*. *Journal of Insect Science*,

3595 10(1), 1–11. <https://doi.org/10.1673/031.010.2401>

3596 Frederickson, M. E., Greene, M. J., & Gordon, D. M. (2005). Devil's gardens' bedevilled by

3597 ants. *Nature*, 437.

3598 Fry, B. (2005). From genome to “venome”: molecular origin and evolution of the snake venom

3599 proteome inferred from phylogenetic analysis of toxin sequences and related body

3600 proteins. *Genome Research*, 15(3), 403–420. <https://doi.org/10.1101/gr.3228405>

3601 Fry, B. G., Roelants, K., Champagne, D. E., Scheib, H., Tyndall, J. D. A., King, G. F., ... de la

3602 Vega, R. C. R. (2009). The toxicogenomic multiverse: convergent recruitment of proteins

3603 into animal venoms. *Annual Review of Genomics and Human Genetics*, 10(1), 483–511.

3604 <https://doi.org/10.1146/annurev.genom.9.081307.164356>

3605 Gao, B., Dalziel, J., Tanzi, S., & Zhu, S. (2018). Meucin-49, a multifunctional scorpion venom

3606 peptide with bactericidal synergy with neurotoxins. *Amino Acids*, 50(8), 1025–1043.
3607 <https://doi.org/10.1007/s00726-018-2580-0>

3608 Gao, B., Tian, C., & Zhu, S. (2007). Inducible antibacterial response of scorpion venom gland.
3609 *Peptides*, 28(12), 2299–2305. <https://doi.org/10.1016/j.peptides.2007.10.004>

3610 Gathalkar, G., & Sen, A. (2018). Foraging and predatory activities of ants. In *The Complex*
3611 *World of Ants* (pp. 51–70). <https://doi.org/10.5772/intechopen.78011>

3612 Geng, T., Lv, D. D., Huang, Y. X., Hou, C. X., Qin, G. X., & Guo, X. J. (2016). JAK/STAT
3613 signaling pathway-mediated immune response in silkworm (*Bombyx mori*) challenged by
3614 *Beauveria bassiana*. *Gene*, 595(1), 69–76. <https://doi.org/10.1016/j.gene.2016.09.043>

3615 Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: a multiplatform graphical
3616 user interface for sequence alignment and phylogenetic tree building. *Molecular Biology*
3617 *and Evolution*, 27(2), 221–224. <https://doi.org/10.1093/molbev/msp259>

3618 Groark, K. P. (1996). Ritual and therapeutic use of “hallucinogenic” harvester ants
3619 (*Pogonomyrmex*) in native south-central California. *Journal of Ethnobiology*, 16(1), 1–29.

3620 Hamilton, C., & Bulmer, M. S. (2012). Molecular antifungal defenses in subterranean termites:
3621 RNA interference reveals in vivo roles of termicins and GNBP s against a naturally
3622 encountered pathogen. *Developmental & Comparative Immunology*, 36(2), 372–377.
3623 <https://doi.org/10.1016/j.dci.2011.07.008>

3624 Hanson, M. A., & Lemaitre, B. (2020). New insights on *Drosophila* antimicrobial peptide
3625 function in host defense and beyond. *Current Opinion in Immunology*, 62, 22–30.
3626 <https://doi.org/10.1016/j.coi.2019.11.008>

3627 Hedengren-Olcott, M., Olcott, M. C., Mooney, D. T., Ekengren, S., Geller, B. L., & Taylor, B.
3628 J. (2004). Differential activation of the NF- κ B-like factors Relish and Dif in *Drosophila*
3629 *melanogaster* by fungi and Gram-positive bacteria. *Journal of Biological Chemistry*,
3630 279(20), 21121–21127. <https://doi.org/10.1074/jbc.M313856200>

3631 Heep, J., Klaus, A., Kessel, T., Seip, M., Vilcinsk as, A., & Skaljac, M. (2019). Proteomic
3632 analysis of the venom from the ruby ant *Myrmica rubra* and the isolation of a novel
3633 insecticidal decapeptide. *Insects*, 10(42). <https://doi.org/10.3390/insects10020042>

3634 Heep, J., Skaljac, M., Grotmann, J., Kessel, T., Seip, M., Schmidtberg, H., & Vilcinsk as, A.
3635 (2019). Identification and functional characterization of a novel insecticidal decapeptide
3636 from the myrmicine ant *Manica rubida*. *Toxins*, 11(10), 1–17.
3637 <https://doi.org/10.3390/toxins11100562>

3638 Helmkamp f, M., Mikheyev, A. S., Kang, Y., Fewell, J., & Gadau, J. (2016). Gene expression
3639 and variation in social aggression by queens of the harvester ant *Pogonomyrmex*

3640 *californicus*. *Molecular Ecology*, 25(15), 3716–3730. <https://doi.org/10.1111/mec.13700>

3641 Himaya, S. W. A., & Lewis, R. J. (2018). Venomics-accelerated cone snail venom peptide
3642 discovery. *International Journal of Molecular Sciences*, 19(788).
3643 <https://doi.org/10.3390/ijms19030788>

3644 Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A.
3645 J., ... Colston, B. W. (2011). High-throughput droplet digital PCR system for absolute
3646 quantitation of DNA copy number. *Analytical Chemistry*, 83(22), 8604–8610.
3647 <https://doi.org/10.1021/ac202028g>

3648 Illumina Inc. (2011). Quality scores for next-generations sequencing: assessing sequencing
3649 accuracy using Phred quality scoring. Retrieved July 30, 2020, from
3650 [https://www.illumina.com/content/dam/illumina-](https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote_Q-Scores.pdf)
3651 [marketing/documents/products/technotes/technote_Q-Scores.pdf](https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote_Q-Scores.pdf)

3652 Imler, J., & Hoffmann, J. A. (2001). Toll receptors in innate immunity. *Trends in Cellular*
3653 *Biology*, 11(7), 304–311.

3654 Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., ... Loose, M. (2018).
3655 Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature*
3656 *Biotechnology*, 36(4), 338–345. <https://doi.org/10.1038/nbt.4060>

3657 Jimenez, R., Ikonopoulou, M. P., Lopez, J. A., & Miles, J. J. (2017). Immune drug discovery
3658 from venoms. *Toxicon*.

3659 Jin, A.-H., Dutertre, S., Dutt, M., Lavergne, V., Jones, A., Lewis, R., & Alewood, P. (2019).
3660 Transcriptomic-proteomic correlation in the predation-evoked venom of the cone snail,
3661 *Conus imperialis*. *Marine Drugs*, 17(3), 177. <https://doi.org/10.3390/md17030177>

3662 Johnson, S. R., Copello, J. A., Evans, M. S., & Suarez, A. V. (2010). A biochemical
3663 characterization of the major peptides from the venom of the giant neotropical hunting ant
3664 *Dinoponera australis*. *Toxicon*, 55(4), 702–710.
3665 <https://doi.org/10.1016/j.toxicon.2009.10.021>

3666 Johnson, S. R., Rikli, H. G., Schmidt, J. O., & Evans, M. S. (2017). A reexamination of
3667 poneratoxin from the venom of the bullet ant *Paraponera clavata*. *Peptides*, 98, 51–62.
3668 <https://doi.org/10.1016/j.peptides.2016.05.012>

3669 Judge, R. K., Henry, P. J., Mirtschin, P., Jelinek, G., & Wilce, J. A. (2006). Toxins not
3670 neutralized by brown snake antivenom. *Toxinology and Applied Pharmacology*, 213, 117–
3671 125. <https://doi.org/10.1016/j.taap.2005.09.010>

3672 Kaas, Q., Yu, R., Jin, A. H., Dutertre, S., & Craik, D. J. (2012). ConoServer: updated content,
3673 knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Research*, 40,

3674 325–330. <https://doi.org/10.1093/nar/gkr886>

3675 Kalia, J., Milescu, M., Salvatierra, J., Wagner, J., Klint, J. K., King, G. F., ... Bosmans, F.
3676 (2015). From foe to friend: using animal toxins to investigate ion channel function. *Journal*
3677 *of Molecular Biology*, 427(1), 158–175. <https://doi.org/10.1016/j.jmb.2014.07.027>

3678 Kazuma, K., Masuko, K., Konno, K., & Inagaki, H. (2017). Combined venom gland
3679 transcriptomic and venom peptidomic analysis of the predatory ant *Odontomachus*
3680 *monticola*. *Toxins*, 9(10), 1–15. <https://doi.org/10.3390/toxins9100323>

3681 Kimbrell, D. A., & Beutler, B. (2001). The evolution and genetics of innate immunity. *Nature*,
3682 2, 256–267.

3683 King, G. F. (2011). Venoms as a platform for human drugs: translating toxins into therapeutics.
3684 *Expert Opinion on Biological Therapy*, 11(11), 1469–1484.
3685 <https://doi.org/10.1517/14712598.2011.621940>

3686 King, G. F. (2019). Tying pest insects in knots: the deployment of spider-venom-derived
3687 knottins as bioinsecticides. *Pest Management Science*, 75, 2437–2445.
3688 <https://doi.org/10.1002/ps.5452>

3689 King, G. F., Gentz, M. C., Escoubas, P., & Nicholson, G. M. (2008). A rational nomenclature
3690 for naming peptide toxins from spiders and other venomous animals. *Toxicon*, 52(2), 264–
3691 276. <https://doi.org/10.1016/j.toxicon.2008.05.020>

3692 King, T. P., Jim, S. Y., & Wittkowski, K. M. (2003). Inflammatory role of two venom
3693 components of yellow jackets (*Vespula vulgaris*): a mast cell degranulating peptide
3694 mastoparan and phospholipase A1. *International Archives of Allergy and Immunology*,
3695 131(1), 25–32. <https://doi.org/10.1159/000070431>

3696 Knudsen, C., Ledsgaard, L., Ibsen Dehli, R., Ahmadi, S., Vinther Sorensen, C., & Hougaard
3697 Laustsen, A. (2019). Engineering and design considerations for next-generation snakebite
3698 antivenoms. *Toxicon*, 167, 67–75. <https://doi.org/10.1016/j.toxicon.2019.06.005>

3699 Kordis, D., & Gubensek, F. (2000). Adaptive evolution of animal toxin multigene families.
3700 *Gene*, 261(1), 43–52. [https://doi.org/10.1016/S0378-1119\(00\)00490-X](https://doi.org/10.1016/S0378-1119(00)00490-X)

3701 Korpelainen, E., Tuimala, J., Somervuo, P., Huss, M., & Wong, G. (2015). Transcriptome
3702 assembly. In N. F. Britton, X. Lin, H. M. Safer, M. V. Schneider, M. Singh, & A.
3703 Tramontao (Eds.), *RNA-seq Data Analysis* (Chapman &, pp. 85–107). Boca Raton: Taylor
3704 & Francis Group, LLC.

3705 Kozlov, S. A., Vassilevski, A. A., Feofanov, A. V., Surovoy, A. Y., Karpunin, D. V., & Grishin,
3706 E. V. (2006). Latarcins, antimicrobial and cytolytic peptides from the venom of the spider
3707 *Lachesana tarabaevi* (Zodariidae) that exemplify biomolecular diversity. *Journal of*

3708 *Biological Chemistry*, 281(30), 20983–20992. <https://doi.org/10.1074/jbc.M602168200>

3709 Kuhn-Nentwig, L. (2003). Antimicrobial and cytolytic peptides of venomous arthropods.

3710 *Cellular and Molecular Life Sciences*, 60(12), 2651–2668.

3711 <https://doi.org/10.1007/s00018-003-3106-8>

3712 Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular

3713 evolutionary genetics analysis across computing platforms. *Molecular Biology and*

3714 *Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>

3715 Lamberty, M., Zachary, D., Lanot, R., Bordereau, C., Robert, A., Hoffmann, J. A., & Bulet, P.

3716 (2001). Constitutive expression of a cysteine-rich antifungal and a linear antibacterial

3717 peptide in a termite insect. *Journal of Biological Chemistry*, 276(6), 4085–4092.

3718 <https://doi.org/10.1074/jbc.M002998200>

3719 Larsen, A., Reynaldi, F. J., & Guzmán-Novoa, E. (2019). Fundamentals of the honey bee (*Apis*

3720 *mellifera*) immune system. *Revista Mexicana De Ciencias Pecuarias*, 10(3), 705–728.

3721 <https://doi.org/10.22319/rmcp.v10i3.4785>

3722 LeBrun, E. G., Jones, N. T., & Gilbert, L. E. (2014). Chemical warfare among invaders: a

3723 detoxification interaction facilitates an ant invasion. *Science*, 343, 1014–1017.

3724 <https://doi.org/10.1890/07-0659.1>

3725 Lee, K. S., Kim, B. Y., Yoon, H. J., Choi, Y. S., & Jin, B. R. (2016). Secapin, a bee venom

3726 peptide, exhibits anti-fibrinolytic, anti-elastolytic, and anti-microbial activities.

3727 *Developmental and Comparative Immunology*, 63, 27–35.

3728 <https://doi.org/10.1016/j.dci.2016.05.011>

3729 Lemaitre, B., & Hoffmann, J. (2007). The Host Defense of *Drosophila melanogaster*. *Annual*

3730 *Review of Immunology*, 25, 697–743.

3731 <https://doi.org/10.1146/annurev.immunol.25.022106.141615>

3732 Lenoir, A., Devers, S., Marchand, P., Bressac, C., & Savolainen, R. (2010). Microgynous

3733 queens in the Palearctic ant, *Manica rubida*: dispersal morphs or social parasites? *Journal*

3734 *of Insect Science*, 10(1), 1–13. <https://doi.org/10.1673/031.010.1701>

3735 Lenoir, A., Khalil, A., Châline, N., & Hefetz, A. (2018). New chemical data on the ant

3736 *Myrmecina graminicola* (Formicidae, Myrmicinae): Unusual abundance of alkene

3737 hydrocarbons and esters. *Biochemical Systematics and Ecology*, 80(June), 39–42.

3738 <https://doi.org/10.1016/j.bse.2018.06.004>

3739 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009).

3740 The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.

3741 <https://doi.org/10.1093/bioinformatics/btp352>

- 3742 Li, Heng, & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler
3743 transform. *Bioinformatics*, 26(5), 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- 3744 Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-
3745 time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, 25, 402–408.
3746 <https://doi.org/10.1006/meth.2001.1262>
- 3747 Lourenço, A. P., Florecki, M. M., Simões, Z. L. P., & Evans, J. D. (2018). Silencing of *Apis*
3748 *mellifera* dorsal genes reveals their role in expression of the antimicrobial peptide
3749 defensin-1. *Insect Molecular Biology*, 27(5), 577–589. <https://doi.org/10.1111/imb.12498>
- 3750 Madio, B., Undheim, E. A. B., & King, G. F. (2017). Revisiting venom of the sea anemone
3751 *Stichodactyla haddoni*: omics techniques reveal the complete toxin arsenal of a well-
3752 studied sea anemone genus. *Journal of Proteomics*, 166, 83–92.
3753 <https://doi.org/10.1016/j.jprot.2017.07.007>
- 3754 Mahlapuu, M., Håkansson, J., Ringstad, L., & Björn, C. (2016). Antimicrobial peptides: an
3755 emerging category of therapeutic agents. *Frontiers in Cellular and Infection*
3756 *Microbiology*, 6(December), 194. <https://doi.org/10.3389/fcimb.2016.00194>
- 3757 Mariano, D. O. C., de Oliveira, Ú. C., Zaharenko, A. J., Pimenta, D. C., Rádis-Baptista, G., &
3758 Prieto-Da-Silva, Á. R. de B. (2019). Bottom-up proteomic analysis of polypeptide venom
3759 components of the giant ant *Dinoponera quadriceps*. *Toxins*, 11(8).
3760 <https://doi.org/10.3390/toxins11080448>
- 3761 Martins, G. F., & Ramalho-Ortigão, J. M. (2012). Oenocytes in insects. *Invertebrate Survival*
3762 *Journal*, 9(2), 139–152.
- 3763 Martins, G. F., Ramalho-Ortigão, J. M., Lobo, N. F., Severson, D. W., Mcdowell, M. A., &
3764 Pimenta, P. F. P. (2011). Insights into the transcriptome of oenocytes from *Aedes aegypti*
3765 pupae. *Memorias Do Instituto Oswaldo Cruz*, 106(3), 308–315.
3766 <https://doi.org/10.1590/S0074-02762011000300009>
- 3767 Martinson, E. O., Mrinalini, Kelkar, Y. D., Chang, C. H., & Werren, J. H. (2017). The evolution
3768 of venom by co-option of single-copy genes. *Current Biology*, 27(13), 2007–2013.e8.
3769 <https://doi.org/10.1016/j.cub.2017.05.032>
- 3770 Martinson, E. O., Wheeler, D., Wright, J., Mrinalini, Siebert, A. L., & Werren, J. H. (2014).
3771 *Nasonia vitripennis* venom causes targeted gene expression changes in its fly host.
3772 *Molecular Ecology*, 23(23), 5918–5930. <https://doi.org/10.1111/mec.12967>
- 3773 Matsumura, T., Mashiko, R., Sato, T., Itokawa, K., Maekawa, Y., Ogawa, K., ... Ato, M.
3774 (2018). Venom and antivenom of redback spider (*Latrodectus hasseltii*) in Japan.
3775 *Japanese Journal of Infectious Diseases*. <https://doi.org/10.7883/yoken.JJID.2017.291>

- 3776 Mechref, Y. (2012). Use of CID/ETD mass spectrometry to analyze glycopeptides. *Current*
3777 *Protocols in Protein Science*, 0(12). <https://doi.org/10.1002/0471140864.ps1211s68>
- 3778 Ménez, A., Stöcklin, R., & Mebs, D. (2006). “Venomics” or: the venomous systems genome
3779 project. *Toxicon*, 47(3), 255–259. <https://doi.org/10.1016/j.toxicon.2005.12.010>
- 3780 Michalek, O., Kuhn-Nentwig, L., & Pekar, S. (2019). High specific efficiency of venom of two
3781 prey-specialized spiders. *Toxins*, 11(687). <https://doi.org/10.3390/toxins11120687>
- 3782 Mir, R., Karim, S., Kamal, M. A., Wilson, C., & Mirza, Z. (2016). Conotoxins: structure,
3783 therapeutic potential and pharmacological applications. *Current Pharmaceutical Design*,
3784 22(5). <https://doi.org/10.5151/cidi2017-060>
- 3785 Mitra, A. (2013). Function of the Dufour’s gland in solitary and social Hymenoptera. *Journal*
3786 *of Hymenoptera Research*, 35, 33–58. <https://doi.org/10.3897/JHR.35.4783>
- 3787 Moreau, S. J. M. (2013). “It stings a bit but it cleans well” : venoms of hymenoptera and their
3788 antimicrobial potential. *Journal of Insect Physiology*, 59(2), 186–204.
3789 <https://doi.org/10.1016/j.jinsphys.2012.10.005>
- 3790 Moreno, M., & Giralt, E. (2015). Three valuable peptides from bee and wasp venoms for
3791 therapeutic and biotechnological use: melittin, apamin and mastoparan. *Toxins*, 7(4),
3792 1126–1150. <https://doi.org/10.3390/toxins7041126>
- 3793 Morgan, E. D. (2008). Chemical sorcery for sociality: exocrine secretions of ants
3794 (Hymenoptera: Formicidae). *Myrmecol. News*, 11(August), 79–90.
- 3795 Morgenstern, D., & King, G. F. (2013). The venom optimization hypothesis revisited. *Toxicon*,
3796 63(1), 120–128. <https://doi.org/10.1016/j.toxicon.2012.11.022>
- 3797 Nakasu, E. Y. T., Williamson, S. M., Edwards, M. G., Fitches, E. C., Gatehouse, J. A., Wright,
3798 G. A., & Gatehouse, A. M. R. (2014). Novel biopesticide based on a spider venom peptide
3799 shows no adverse effects on honeybees. *Proceedings of the Royal Society B: Biological*
3800 *Sciences*, 281(1787). <https://doi.org/10.1098/rspb.2014.0619>
- 3801 Nakhleh, J., El Moussawi, L., & Osta, M. A. (2017). *The melanization response in insect*
3802 *immunity. Advances in Insect Physiology* (1st ed., Vol. 52). Elsevier Ltd.
3803 <https://doi.org/10.1016/bs.aiip.2016.11.002>
- 3804 Nehme, N. T., Liégeois, S., Kele, B., Giammarinaro, P., Pradel, E., Hoffmann, J. A., ...
3805 Ferrandon, D. (2007). A model of bacterial intestinal infections in *Drosophila*
3806 *melanogaster*. *PLoS Pathogens*, 3(11), 1694–1709.
3807 <https://doi.org/10.1371/journal.ppat.0030173>
- 3808 Nekaris, K. A.-I., Moore, R. S., Johanna Rode, E., & Fry, B. G. (2013). Mad, bad and dangerous
3809 to know: the biochemistry, ecology and evolution of slow loris venom. *Journal of*

3810 *Venomous Animals and Toxins Including Tropical Diseases*, 19(1), 1–10.
3811 <https://doi.org/10.1186/1678-9199-19-21>

3812 Neyen, C., Bretscher, A. J., Binggeli, O., & Lemaitre, B. (2014). Methods to study *Drosophila*
3813 immunity. *Methods*, 68(1), 116–128. <https://doi.org/10.1016/j.ymeth.2014.02.023>

3814 Nixon, S. A., Dekan, Z., Robinson, S. D., Guo, S., Vetter, I., Kotze, A. C., ... Herzig, V. (2020).
3815 It takes two : dimerization is essential for the broad-spectrum predatory and defensive
3816 activities of the venom peptide Mpl1a from the jack jumper ant *Myrmecia pilosula*.
3817 *Biomedicines*, 8(185). <https://doi.org/10.3390/biomedicines8070185>

3818 Nolde, D. E., Sobol, A. G., Pluzhnikov, K. A., Grishin, E. V., & Arseniev, A. S. (1995). Three-
3819 dimensional structure of ectatomin from *Ectatomma tuberculatum* ant venom. *Journal of*
3820 *Biomolecular NMR*, 5(1), 1–13. <https://doi.org/10.1007/BF00227465>

3821 Orivel, J. (2000). *L'adaptation à la vie arboricole de la fourmi Pachycondyla goeldii*
3822 *(Hymenoptera : Ponerinae)*.

3823 Orivel, J., Redeker, V., Le Caer, J. P., Krier, F., Revol-Junelles, A. M., Longeon, A., ... Rossier,
3824 J. (2001). Ponericins, new antibacterial and insecticidal peptides from the venom of the
3825 ant *Pachycondyla goeldii*. *Journal of Biological Chemistry*, 276(21), 17823–17829.
3826 <https://doi.org/10.1074/jbc.M100216200>

3827 Pan, J., & Hink, W. F. (2000). Isolation and characterization of myrmexins, six isoforms of
3828 venom proteins with anti-inflammatory activity from the tropical ant, *Pseudomyrmex*
3829 *triplarinus*. *Toxicon*, 38(10), 1403–1413. [https://doi.org/10.1016/S0041-0101\(99\)00233-](https://doi.org/10.1016/S0041-0101(99)00233-0)
3830 0

3831 Pennington, M. W., Czerwinski, A., & Norton, R. S. (2018). Peptide therapeutics from venom:
3832 current status and potential. *Bioorganic and Medicinal Chemistry*, 26(10), 2738–2758.
3833 <https://doi.org/10.1016/j.bmc.2017.09.029>

3834 Petrunkevitch, A. (1926). Tarantula versus tarantula-hawk: a study in instinct. *Journal of*
3835 *Experimental Zoology*, 45(2), 367–397. <https://doi.org/10.1002/jez.1400450202>

3836 Piek, T., Duval, A., Hue, B., Karst, H., Lapied, B., Mantel, P., ... Schmidt, J. O. (1991).
3837 Poneratoxin, a novel peptide neurotoxin from the venom of the ant, *Paraponera clavata*.
3838 *Comparative Biochemistry and Physiology. Part C, Comparative*, 99(3), 487–495.
3839 [https://doi.org/10.1016/0742-8413\(91\)90276-Y](https://doi.org/10.1016/0742-8413(91)90276-Y)

3840 Piek, T., Hue, B., Mantel, P., Terumi, N., & Schmidt, J. O. (1991). Pharmacological
3841 characterization and chemical fractionation of the venom of the ponerine ant, *Paraponera*
3842 *clavata* (F.). *Comparative Biochemistry and Physiology. Part C, Comparative*, 99(3),
3843 481–486. [https://doi.org/10.1016/0742-8413\(91\)90275-X](https://doi.org/10.1016/0742-8413(91)90275-X)

3844 Pineda, S. S., Chaumeil, P. A., Kunert, A., Kaas, Q., Thang, M. W. C., Le, L., ... King, G. F.
3845 (2018). ArachnoServer 3.0: an online resource for automated discovery, analysis and
3846 annotation of spider toxins. *Bioinformatics*, 34(6), 1074–1076.
3847 <https://doi.org/10.1093/bioinformatics/btx661>

3848 Pla, D., Sanz, L., Sasa, M., Acevedo, M. E., Dwyer, Q., Durban, J., ... Calvete, J. J. (2017).
3849 Proteomic analysis of venom variability and ontogeny across the arboreal palm-pitvipers
3850 (genus *Bothriechis*). *Journal of Proteomics*, 152, 1–12.
3851 <https://doi.org/10.1016/j.jprot.2016.10.006>

3852 Pluzhnikov, K. A., Nolde, D. E., Tertishnikova, S. M., Sukhanov, S. V., Sobol, A. G., Torgov,
3853 M. Y., ... Grishin, E. V. (1994). Structure-activity study of the basic toxic component of
3854 venom from the ant *Ectatomma tuberculatum*. *Bioorganicheskaya Khimiya*, 20(8–9), 857–
3855 871.

3856 Pluzhnikov, K., Nosyreva, E., Shevchenko, L., Kokoz, Y., Schmalz, D., Hucho, F., & Grishin,
3857 E. (1999). Analysis of ectatomin action on cell membranes. *European Journal of*
3858 *Biochemistry*, 262(2), 501–506. <https://doi.org/10.1046/j.1432-1327.1999.00426.x>

3859 Pluzhnikov, Kirill A., Kozlov, S. A., Vassilevski, A. A., Vorontsova, O. V., Feofanov, A. V.,
3860 & Grishin, E. V. (2014). Linear antimicrobial peptides from *Ectatomma quadridens* ant
3861 venom. *Biochimie*, 107(PB), 211–215. <https://doi.org/10.1016/j.biochi.2014.09.012>

3862 Powell, M. E., Bradish, H. M., Cao, M., Makinson, R., Brown, A. P., Gatehouse, J. A., &
3863 Fitches, E. C. (2020). Demonstrating the potential of a novel spider venom-based
3864 biopesticide for target-specific control of the small hive beetle, a serious pest of the
3865 European honeybee. *Journal of Pest Science*, 93, 391–402.
3866 <https://doi.org/10.1007/s10340-019-01143-3>

3867 Prashanth, J. R., Dutertre, S., Jin, A. H., Lavergne, V., Hamilton, B., Cardoso, F. C., ... Lewis,
3868 R. J. (2016). The role of defensive ecological interactions in the evolution of conotoxins.
3869 *Molecular Ecology*, 25(2), 598–615. <https://doi.org/10.1111/mec.13504>

3870 Pull, C., Ugelvig, L., Wiesenhofer, F., Tragust, S., Schmitt, T., Brown, M., & Cremer, S.
3871 (2018). Destructive disinfection of infected brood prevents systemic disease spread in ant
3872 colonies. *eLife*, 7, 32073. <https://doi.org/10.7554/eLife.32073.001>

3873 R Core Team. (2017). R: a language and environment for statistical computing. *R. Foundation*
3874 *for Statistical Computing*. Retrieved from <https://www.r-project.org/>

3875 Radis-Baptista, G., Dodou, H. V., Prieto-da-Silva, A. R. B., Zaharenko, A. J., Nihei, K.,
3876 Inagaki, H., ... Konno, K. (2020). Comprehensive analysis of peptides and low molecular
3877 weight components of the giant ant *Dinoponera quadriceps* venom. *Biological Chemistry*.

3878 <https://doi.org/10.1515/hsz-2019-397ja-01>

3879 Rádis-Baptista, G., & Konno, K. (2020). Arthropod venom components and their potential
3880 usage. *Toxins*, *12*(82). <https://doi.org/10.3390/toxins12020082>

3881 Rice, P., Longden, L., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open
3882 Software Suite. *Trends in Genetics*, *16*(6), 276–277. <https://doi.org/10.1016/S0168->
3883 [9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2)

3884 Rifflet, A., Gavaldà, S., Téné, N., Orivel, J., Leprince, J., Guilhaudis, L., ... Treilhou, M.
3885 (2012). Identification and characterization of a novel antimicrobial peptide from the
3886 venom of the ant *Tetramorium bicarinatum*. *Peptides*, *38*(2), 363–370.
3887 <https://doi.org/10.1016/j.peptides.2012.08.018>

3888 Rima, M., Naini, S. M. A., Karam, M., Sadek, R., Sabatier, J. M., & Fajloun, Z. (2018). Vipers
3889 of the Middle East: a rich source of bioactive molecules. *Molecules*, *23*(2721).
3890 <https://doi.org/10.3390/molecules23102721>

3891 Rivel, M., Solano, D., Herrera, M., Vargas, M., Villalta, M., Segura, Á., ... Gutiérrez, J. M.
3892 (2016). Pathogenesis of dermonecrosis induced by venom of the spitting cobra, *Naja*
3893 *nigricollis*: an experimental study in mice. *Toxicon*, *119*, 171–179.
3894 <https://doi.org/10.1016/j.toxicon.2016.06.006>

3895 Robertson, P. L. (1968). A morphological and functional study of the venom apparatus in
3896 representatives of some major groups of hymenoptera. *Australian Journal of Zoology*,
3897 *16*(1), 133–166. <https://doi.org/10.1071/ZO9680133>

3898 Robinson, S. D., Mueller, A., Clayton, D., Starobova, H., Hamilton, B. R., Payne, R. J., ...
3899 Undheim, E. A. B. (2018). A comprehensive portrait of the venom of the giant red bull
3900 ant, *Myrmecia gulosa*, reveals a hyperdiverse hymenopteran toxin gene family. *Science*
3901 *Advances*, *4*(9), eaau4640. <https://doi.org/10.1126/sciadv.aau4640>

3902 Rojko, N., Dalla Serra, M., Maček, P., & Anderluh, G. (2016). Pore formation by actinoporins,
3903 cytolysins from sea anemones. *Biochimica et Biophysica Acta - Biomembranes*, *1858*(3),
3904 446–456. <https://doi.org/10.1016/j.bbamem.2015.09.007>

3905 Rollard, C., Chippaux, J.-P., & Goyffon, M. (2015). *La fonction venimeuse* (Lavoisier). Paris.

3906 Roma, G. C., Bueno, O. C., & Camargo-Mathias, M. I. (2010). Morpho-physiological analysis
3907 of the insect fat body: a review. *Micron*, *41*(5), 395–401.
3908 <https://doi.org/10.1016/j.micron.2009.12.007>

3909 Roma, G. C., Mathias, M. I. C., & Bueno, O. C. (2006). Fat body in some genera of leaf-cutting
3910 ants (Hymenoptera: Formicidae). Proteins, lipids and polysaccharides detection. *Micron*,
3911 *37*(3), 234–242. <https://doi.org/10.1016/j.micron.2005.10.012>

- 3912 Sackton, T. B., Lazzaro, B. P., Schlenke, T. A., Evans, J. D., Hultmark, D., & Clark, A. G.
3913 (2007). Dynamic evolution of the innate immune system in *Drosophila*. *Nature Genetics*,
3914 39(12), 1461–1468.
- 3915 Satyavathi, V. V., Minz, A., & Nagaraju, J. (2014). Nodulation: an unexplored cellular defense
3916 mechanism in insects. *Cellular Signalling*, 26(8), 1753–1763.
3917 <https://doi.org/10.1016/j.cellsig.2014.02.024>
- 3918 Schendel, V., Rash, L. D., Jenner, R. A., & Undheim, E. A. B. (2019). The diversity of venom:
3919 the importance of behavior and venom system morphology in understanding its ecology
3920 and evolution. *Toxins*, 11(666). <https://doi.org/10.3390/toxins11110666>
- 3921 Schlüns, H., & Crozier, R. H. (2007). Relish regulates expression of antimicrobial peptide genes
3922 in the honeybee, *Apis mellifera*, shown by RNA interference. *Insect Molecular Biology*,
3923 16(6), 753–759. <https://doi.org/10.1111/j.1365-2583.2007.00768.x>
- 3924 Schlüns, Helge, & Crozier, R. H. (2009). Molecular and chemical immune defenses in ants
3925 (Hymenoptera: Formicidae). *Myrmecological News*, 12(September), 237–249.
- 3926 Schmidt, J. O. (1987). Chemistry, pharmacology and chemical ecology of ant venoms. In T.
3927 Piek (Ed.), *Venoms of the Hymenoptera* (Academic P, pp. 425–498). London.
- 3928 Schmidt, J. O. (2019a). Pain and lethality induced by insect stings: an exploratory and
3929 correlational study. *Toxin*, 11(427). <https://doi.org/10.3390/toxins11070427>
- 3930 Schmidt, J. O. (2019b). The insect sting pain scale: how the pain and lethality of ant, wasp and
3931 bee venoms can guide the way for human benefit. *Preprints*, 2019050318.
- 3932 Schmidt, J. O., & Blum, M. S. (1978a). A harvester ant venom: chemistry and pharmacology.
3933 *Science*, 200, 1064–1066.
- 3934 Schmidt, J. O., & Blum, M. S. (1978b). The biochemical constituents of the venom of the
3935 harvester ant, *Pogonomyrmex badius*. *Comparative Biochemistry and Physiology. Part C*,
3936 *Comparative*, 61(1), 239–247. [https://doi.org/10.1016/0306-4492\(78\)90137-5](https://doi.org/10.1016/0306-4492(78)90137-5)
- 3937 Schmidt, P. J., Sherbrooke, W. C., & Schmidt, J. O. (1989). The detoxification of ant
3938 (*Pogonomyrmex*) venom by a blood factor in horned lizards (Phrynosoma). *Copeia*, 3,
3939 603–607.
- 3940 Schweitz, H., Bruhn, T., Guillemare, E., Moinier, D., Lancelin, J. M., Béress, L., & Lazdunski,
3941 M. (1995). Kalicludines and kaliseptine: two different classes of sea anemone toxins for
3942 voltage-sensitive K⁺ channels. *Journal of Biological Chemistry*, 270(42), 25121–25126.
3943 <https://doi.org/10.1074/jbc.270.42.25121>
- 3944 Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., Von Haeseler, A., &
3945 Schatz, M. C. (2018). Accurate detection of complex structural variations using single-

3946 molecule sequencing. *Nature Methods*, 15(6), 461–468. [https://doi.org/10.1038/s41592-](https://doi.org/10.1038/s41592-018-0001-7)
3947 018-0001-7

3948 Senji Laxme, R. R., Suranse, V., & Sunagar, K. (2019). Arthropod venoms: biochemistry,
3949 ecology and evolution. *Toxicon*, 158(November 2018), 84–103.
3950 <https://doi.org/10.1016/j.toxicon.2018.11.433>

3951 Servent, D. (2016). Venins : menace ou opportunité thérapeutique. In *Les conférences Cyclope*
3952 *du CEA Saclay*.

3953 Sheehan, G., Farrell, G., & Kavanagh, K. (2020). Immune priming: the secret weapon of the
3954 insect world. *Virulence*, 11(1), 238–246. <https://doi.org/10.1080/21505594.2020.1731137>

3955 Shin, J., Lee, W., & Lee, W. (2008). Structural proteomics by NMR spectroscopy. *Expert*
3956 *Review of Proteomics*, 5(4), 589–601. <https://doi.org/10.1586/14789450.5.4.589>

3957 Silva, P. M., Gonçalves, S., & Santos, N. C. (2014). Defensins: antifungal lessons from
3958 eukaryotes. *Frontiers in Microbiology*, 5(97). <https://doi.org/10.3389/fmicb.2014.00097>

3959 Strand, M. R. (2008). The insect cellular immune response. *Insect Science*, 15(1), 1–14.
3960 <https://doi.org/10.1111/j.1744-7917.2008.00183.x>

3961 Sunagar, K., Morgenstern, D., Reitzel, A. M., & Moran, Y. (2016). Ecological venomics: how
3962 genomics, transcriptomics and proteomics can shed new light on the ecology and evolution
3963 of venom. *Journal of Proteomics*, 135, 62–72. <https://doi.org/10.1016/j.jprot.2015.09.015>

3964 Szolajaska, E., Poznanski, J., Ferber, M. L., Michalik, J., Gout, E., Fender, P., ... Chroboczek,
3965 J. (2004). Poneratoxin, a neurotoxin from ant venom: Structure and expression in insect
3966 cells and construction of a bio-insecticide. *European Journal of Biochemistry*, 271(11),
3967 2127–2136. <https://doi.org/10.1111/j.1432-1033.2004.04128.x>

3968 Tafesh-Edwards, G., & Eleftherianos, I. (2020). JNK signaling in *Drosophila* immunity and
3969 homeostasis. *Immunology Letters*. <https://doi.org/10.1016/j.imlet.2020.06.017>

3970 Tanenhaus, A. K., Zhang, J., & Yin, J. C. P. (2012). In vivo circadian oscillation of dCREB2
3971 and NF-κB activity in the *Drosophila* nervous system. *PLoS ONE*, 7(10), e45130.
3972 <https://doi.org/10.1371/journal.pone.0045130>

3973 Tapadia, M. G., & Verma, P. (2012). Immune response and anti-microbial peptides expression
3974 in malpighian tubules of *Drosophila melanogaster* is under developmental regulation.
3975 *PLoS ONE*, 7(7), 1–12. <https://doi.org/10.1371/journal.pone.0040714>

3976 Téné, N., Bonnafé, E., Berger, F., Rifflet, A., Guilhaudis, L., Ségalas-Milazzo, I., ... Treilhou,
3977 M. (2016). Biochemical and biophysical combined study of bicarinalin, an ant venom
3978 antimicrobial peptide. *Peptides*, 79, 103–113.
3979 <https://doi.org/10.1016/j.peptides.2016.04.001>

- 3980 Téné, N., Roche-Chatain, V., Rifflet, A., Bonnafé, E., Lefranc, B., Leprince, J. Ô., & Treilhou,
3981 M. (2014). Potent bactericidal effects of bicarinalin against strains of the *Enterobacter* and
3982 *Cronobacter* genera. *Food Control*, 42, 202–206.
3983 <https://doi.org/10.1016/j.foodcont.2014.02.026>
- 3984 Tonk, M., Pierrot, C., Cabezas-cruz, A., Rahnamaeian, M., Khalife, J., & Vilcinskis, A. (2019).
3985 The *Drosophila melanogaster* antimicrobial peptides Mtk-1 and Mtk-2 are active against
3986 the malarial parasite *Plasmodium falciparum*. *Parasitology Research*.
- 3987 Torres, A. F. C., Huang, C., Chong, C. M., Leung, S. W., Prieto-da-Silva, Á. R. B., Havt, A.,
3988 ... Rádis-Baptista, G. (2014). Transcriptome analysis in venom gland of the predatory
3989 giant ant *Dinoponera quadriceps*: insights into the polypeptide toxin arsenal of
3990 hymenopterans. *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0087556>
- 3991 Touchard, A., Aili, S. R., Fox, E. G. P., Escoubas, P., Orivel, J., Nicholson, G. M., & Dejean,
3992 A. (2016). The biochemical toxin arsenal from ant venoms. *Toxins*, 8(1), 1–28.
3993 <https://doi.org/10.3390/toxins8010030>
- 3994 Touchard, A., Aili, S. R., Téné, N., Barassé, V., Klopp, C., Dejean, A., ... Bonnafé, E. (2020).
3995 Venom peptide repertoire of the european myrmicine ant *Manica rubida*: identification of
3996 insecticidal toxins. *Journal of Proteome Research*, 19(4), 1800–1811.
3997 <https://doi.org/10.1021/acs.jproteome.0c00048>
- 3998 Touchard, A., Brust, A., Cardoso, F. C., Chin, Y. K. Y., Herzig, V., Jin, A. H., ... Escoubas, P.
3999 (2016). Isolation and characterization of a structurally unique β -hairpin venom peptide
4000 from the predatory ant *Anochetus emarginatus*. *Biochimica et Biophysica Acta - General*
4001 *Subjects*, 1860(11), 2553–2562. <https://doi.org/10.1016/j.bbagen.2016.07.027>
- 4002 Touchard, A., Labrière, N., Roux, O., Petitclerc, F., Orivel, J., Escoubas, P., ... Dejean, A.
4003 (2014). Venom toxicity and composition in three *Pseudomyrmex* ant species having
4004 different nesting modes. *Toxicon*, 88, 67–76.
4005 <https://doi.org/10.1016/j.toxicon.2014.05.022>
- 4006 Touchard, A., Mendel, H. C., Boulogne, I., Herzig, V., & Emidio, N. B. (2020). Heterodimeric
4007 insecticidal peptide provides new insights into the molecular and functional diversity of
4008 ant venoms. *BioRxiv*. <https://doi.org/10.1101/2020.07.29.226878>
- 4009 Touchard, A., Téné, N., Chan Tchi Song, P., Lefranc, B., Leprince, J., Treilhou, M., & Bonnafé,
4010 E. (2018). Deciphering the molecular diversity of an ant venom peptidome through a
4011 venomics approach. *Journal of Proteome Research*, 17(10), 3503–3516.
4012 <https://doi.org/10.1021/acs.jproteome.8b00452>
- 4013 Tragust, S. (2016). External immune defence in ant societies (Hymenoptera: Formicidae): the

4014 role of antimicrobial venom and metapleural gland secretion. *Myrmecological News*, 23,
4015 119–128.

4016 Tragust, S., Mitteregger, B., Barone, V., Konrad, M., Ugelvig, L. V., & Cremer, S. (2013). Ants
4017 disinfect fungus-exposed brood by oral uptake and spread of their poison. *Current Biology*,
4018 23(1), 76–82. <https://doi.org/10.1016/j.cub.2012.11.034>

4019 Turillazzi, S., Mastrobuoni, G., Dani, F. R., Moneti, G., Pieraccini, G., La Marca, G., ...
4020 Dapporto, L. (2006). Dominulin A and B: Two new antibacterial peptides identified on
4021 the cuticle and in the venom of the social paper wasp *Polistes dominulus* using MALDI-
4022 TOF, MALDI-TOF/TOF, and ESI-ion trap. *Journal of the American Society for Mass*
4023 *Spectrometry*, 17(3), 376–383. <https://doi.org/10.1016/j.jasms.2005.11.017>

4024 Tzou, P., Ohresser, S., Ferrandon, D., Capovilla, M., Reichhart, J. M., Lemaitre, B., ... Imler,
4025 J. L. (2000). Tissue-specific inducible expression of antimicrobial peptide genes in
4026 *Drosophila* surface epithelia. *Immunity*, 13(5), 737–748. <https://doi.org/10.1016/S1074->
4027 7613(00)00072-8

4028 Utkin, Y. N. (2015). Animal venom studies: current benefits and future developments. *World*
4029 *Journal of Biological Chemistry*, 6(2), 28–33. <https://doi.org/10.4331/wjbc.v6.i2.28>

4030 Uvell, H., & Engström, Y. (2007). A multilayered defense against infection: combinatorial
4031 control of insect immune genes. *Trends in Genetics*, 23(7), 342–349.
4032 <https://doi.org/10.1016/j.tig.2007.05.003>

4033 van Aalten, D. M., Synstad, B., Brurberg, M. B., Hough, E., Riise, B. W., Eijsink, V. G., &
4034 Wierenga, R. K. (2000). Structure of a two-domain chitotriosidase from *Serratia*
4035 *marcescens* at 1.9-Å resolution. *Proceedings of the National Academy of Sciences of the*
4036 *United States of America*, 97(11), 5842–5847. <https://doi.org/10.1073/pnas.97.11.5842>

4037 Vassilevski, A. A., Kozlov, S. A., & Grishin, E. V. (2009). Molecular diversity of spider venom.
4038 *Biochemistry*, 74(13), 1505–1534. <https://doi.org/10.1134/S0006297909130069>

4039 Viljakainen, L. (2015). Evolutionary genetics of insect innate immunity. *Briefings in*
4040 *Functional Genomics*, 14(6), 407–412. <https://doi.org/10.1093/bfgp/evl002>

4041 von Reumont, Bjoern Marcus, Campbell, L. I., & Jenner, R. A. (2014). *Quo vadis* venomics?
4042 A roadmap to neglected venomous invertebrates. *Toxins*, 6(12), 3488–3551.
4043 <https://doi.org/10.3390/toxins6123488>

4044 von Reumont, Björn M., Campbell, L. I., Richter, S., Hering, L., Sykes, D., Hetmank, J., ...
4045 Bleidorn, C. (2014). A polychaete's powerful punch: venom gland transcriptomics of
4046 *Glycera* reveals a complex cocktail of toxin homologs. *Genome Biology and Evolution*,
4047 6(9), 2406–2423. <https://doi.org/10.1093/gbe/evu190>

- 4048 von Reumont, Björn M., Undheim, E. A. B., Jauss, R. T., & Jenner, R. A. (2017). Venomics of
4049 remipede crustaceans reveals novel peptide diversity and illuminates the venom's
4050 biological role. *Toxins*, 9(8). <https://doi.org/10.3390/toxins9080234>
- 4051 Von Reumont, Björn Marcus. (2018). Studying smaller and neglected organisms in modern
4052 evolutionary venomics implementing RNASeq (Transcriptomics)—A critical guide.
4053 *Toxins*, 10(7), 1–23. <https://doi.org/10.3390/toxins10070292>
- 4054 von Sicard, N. A. E., Candy, D. J., & Anderson, M. (1989). The biochemical composition of
4055 venom from the pavement ant (*Tetramorium caespitum* L.). *Toxicon*, 27(10), 1127–1133.
4056 [https://doi.org/10.1016/0041-0101\(89\)90006-8](https://doi.org/10.1016/0041-0101(89)90006-8)
- 4057 Walker, A. A. (2020). The evolutionary dynamics of venom toxins made by insects and other
4058 animals. *Biochemical Society Transactions*. <https://doi.org/10.42/BST20190820>
- 4059 Walker, A. A., Dobson, J., Jin, J., Robinson, S. D., Herzig, V., Vetter, I., ... Fry, B. G. (2018).
4060 Buzz Kill: function and proteomic composition of venom from the giant assassin fly
4061 *Dolopus genitalis* (Diptera: Asilidae). *Toxins*, 10(456).
4062 <https://doi.org/10.3390/toxins10110456>
- 4063 Walker, A. A., Hernández-vargas, M. J., Corzo, G., Fry, B. G., Glenn, ·, & King, F. (2018).
4064 Giant fish-killing water bug reveals ancient and dynamic venom evolution in Heteroptera.
4065 *Cellular and Molecular Life Sciences*. <https://doi.org/10.1007/s00018-018-2768-1>
- 4066 Walker, A. A., Madio, B., Jin, J., Undheim, E. A. B., Fry, B. G., & King, G. F. (2017). Melt
4067 with this kiss: paralyzing and liquefying venom of the assassin bug *Pristhesancus*
4068 *plagipennis* (Hemiptera: Reduviidae). *Molecular & Cellular Proteomics*, 16(4), 552–566.
4069 <https://doi.org/10.1074/mcp.M116.063321>
- 4070 Walker, A. A., Mayhew, M. L., Jin, J., Herzig, V., Undheim, E. A. B., Sombke, A., ... King,
4071 G. F. (2018). The assassin bug *Pristhesancus plagipennis* produces two distinct venoms
4072 in separate gland lumens. *Nature Communications*, 9(1), 1–10.
4073 <https://doi.org/10.1038/s41467-018-03091-5>
- 4074 Walker, A. A., Robinson, S. D., Hamilton, B. F., Undheim, E. A. B., & King, G. F. (2020).
4075 Deadly proteomes: a practical guide to proteotranscriptomics of animal venoms.
4076 *Proteomics*. <https://doi.org/10.1002/jssc.201200569>
- 4077 Walker, A. A., Robinson, S. D., Yeates, D. K., Jin, J., Baumann, K., Dobson, J., ... King, G. F.
4078 (2018). Entomo-venomics: the evolution, biology and biochemistry of insect venoms.
4079 *Toxicon*, 154(July), 15–27. <https://doi.org/10.1016/j.toxicon.2018.09.004>
- 4080 Wanandy, T., Gueven, N., Davies, N. W., Brown, S. G. A., & Wiese, M. D. (2015). Pilsulins:
4081 A review of the structure and mode of action of venom peptides from an australian ant

4082 *Myrmecia pilosula*. *Toxicon*, 98, 54–61. <https://doi.org/10.1016/j.toxicon.2015.02.013>

4083 Ward, P. S., Brady, S. G., Fisher, B. L., & Schultz, T. R. (2015). The evolution of myrmicine
4084 ants: phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera:
4085 Formicidae). *Systematic Entomology*, 40(1), 61–81. <https://doi.org/10.1111/syen.12090>

4086 Whittington, C. M., & Belov, K. (2007). Platypus venom: a review. *Australian Mammalogy*,
4087 29, 57–62. <https://doi.org/10.1071/AM07006>

4088 Whittington, C. M., Papenfuss, A. T., Bansal, P., Torres, A. M., Wong, E. S. W., Deakin, J. E.,
4089 ... Belov, K. (2008). Defensins and the convergent evolution of platypus and reptile
4090 venom genes. *Genome Research*, 18(6), 986–994. <https://doi.org/10.1101/gr.7149808>

4091 Williams, D. J., Habib, A. G., & Warrell, D. A. (2018). Clinical studies of the effectiveness and
4092 safety of antivenoms. *Toxicon*, 150, 1–10. <https://doi.org/10.1016/j.toxicon.2018.05.001>

4093 Williams, F. X. (1956). Life history studies of *Pepsis* and *Hemipepsis* wasps in California
4094 (Hymenoptera: Pompilidae). *Annals of the Entomological Society of America*, 49(5), 447–
4095 466. <https://doi.org/10.1093/aesa/49.5.447>

4096 Wilson, D., & Daly, N. L. (2018). Venomics: a mini-review. *High-Throughput*, 7(19).
4097 <https://doi.org/10.3390/ht7030019>

4098 Wu, K., Yang, B., Huang, W., Dobens, L., Song, H., & Ling, E. (2016). Gut immunity in
4099 lepidopteran insects. *Developmental and Comparative Immunology*, 64, 65–74.
4100 <https://doi.org/10.1016/j.dci.2016.02.010>

4101 Wu, Q., Patočka, J., & Kuča, K. (2018). Insect antimicrobial peptides, a mini review. *Toxins*,
4102 10(11), 461. <https://doi.org/10.3390/toxins10110461>

4103 Wulff, H., Christophersen, P., Colussi, P., Chandy, K. G., & Yarov-Yarovoy, V. (2019).
4104 Antibodies and venom peptides: new therapeutic modalities for ion channels. *Natural*
4105 *Reviews Drug Discovery*, 18(5), 339–357. [https://doi.org/10.1111/j.1476-](https://doi.org/10.1111/j.1476-5381.2011.01649_5.x)
4106 [5381.2011.01649_5.x](https://doi.org/10.1111/j.1476-5381.2011.01649_5.x)

4107 Xie, F., Xiao, P., Chen, D., Xu, L., & Zhang, B. (2012). miRDeepFinder: A miRNA analysis
4108 tool for deep sequencing of plant small RNAs. *Plant Molecular Biology*, 80(1), 75–84.
4109 <https://doi.org/10.1007/s11103-012-9885-2>

4110 Yacoub, T., Rima, M., Karam, M., Sabatier, J.-M., & Fajloun, Z. (2020). Antimicrobials from
4111 venomous animals: an overview. *Molecules*, 25(2402).
4112 <https://doi.org/10.3390/molecules25102402>

4113 Yaffe, H., Buxdorf, K., Shapira, I., Ein-Gedi, S., Moyal-Ben Zvi, M., Fridman, E., ... Levy, M.
4114 (2012). LogSpin: a simple, economical and fast method for RNA isolation from infected
4115 or healthy plants and other eukaryotic tissues. *BMC Research Notes*, 5(1), 45.

4116 <https://doi.org/10.1186/1756-0500-5-45>

4117 Yakovlev, A. Y., Negin, A. P., Simonenko, N. P., Gordya, N. A., Tulin, D. V., Kruglikova, A.
4118 A., & Chernysh, S. I. (2017). Fat body and hemocyte contribution to the antimicrobial
4119 peptide synthesis in *Calliphora vicina* R.-D. (Diptera: Calliphoridae) larvae. *In Vitro*
4120 *Cellular and Developmental Biology - Animal*, 53(1), 33–42.
4121 <https://doi.org/10.1007/s11626-016-0078-1>

4122 Yee, A. A., Savchenko, A., Ignachenko, A., Lukin, J., Xu, X., Skarina, T., ... Arrowsmith, C.
4123 H. (2005). NMR and X-ray crystallography, complementary tools in structural proteomics
4124 of small proteins. *Journal of the American Chemical Society*, 127(47), 16512–16517.
4125 <https://doi.org/10.1021/ja053565+>

4126 Zaidman-Rémy, A., Hervé, M., Poidevin, M., Pili-Floury, S., Kim, M. S., Blanot, D., ...
4127 Lemaitre, B. (2006). The *Drosophila* amidase PGRP-LB modulates the immune response
4128 to bacterial infection. *Immunity*, 24(4), 463–473.
4129 <https://doi.org/10.1016/j.immuni.2006.02.012>

4130 Zancolli, G., & Casewell, N. R. (2019). Venom systems as models for studying the origin and
4131 regulation of evolutionary novelties. *Molecular Biology and Evolution*, msaa133.
4132 <https://doi.org/10.1093/molbev/msaa133>

4133 Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, 415, 389–395.

4134 Zelezetsky, I., Pag, U., Antcheva, N., Sahl, H. G., & Tossi, A. (2005). Identification and
4135 optimization of an antimicrobial peptide from the ant venom toxin pilosulin. *Archives of*
4136 *Biochemistry and Biophysics*, 434(2), 358–364. <https://doi.org/10.1016/j.abb.2004.11.006>

4137 Zhang, Z. T., & Zhu, S. Y. (2009). Drosomycin, an essential component of antifungal defence
4138 in *Drosophila*. *Insect Molecular Biology*, 18(5), 549–556. [https://doi.org/10.1111/j.1365-](https://doi.org/10.1111/j.1365-2583.2009.00907.x)
4139 [2583.2009.00907.x](https://doi.org/10.1111/j.1365-2583.2009.00907.x)

4140 Ziegman, R., & Alewood, P. (2015). Bioactive components in fish venoms. *Toxins*, 7(5), 1497–
4141 1531. <https://doi.org/10.3390/toxins7051497>

4142

The Peptide Venom Composition of the Fierce Stinging Ant *Tetraponera Aethiops* (Formicidae: Pseudomyrmecinae)

Valentine Barassé, Axel Touchard, Nathan Téné, Maurice Tindo, Martin Kenne, Christophe Klopp, Alain Dejean, Elsa Bonnafé and Michel Treilhou

Table S1. Addressing table of major contigs expressed by *Tetraponera aethiops* venom glands.

	Length (kb)	Hits	Reads Per Million	%	Contig_ORF	Seq ID Name	E-value	Function
CL1015Contig1_1	448	118970	408.07	0.04				
CL1134Contig1_1	1124	122638	420.65	0.04	CL1134Contig1_1_6	sp Q8WRF3 RL32_APIME 60S ribosomal protein L32 OS = Apis mellifera OX = 7460 GN = RpL32 PE = 2 SV = 1	2.45e-82	Translation
CL114Contig1_1	1044	189151	648.79	0.06				
CL114Contig2_1	3807	8616798	29555.53	2.96		U2-PSDTX-Ta1a Genbank Accession Number : MN607169		Venom peptide
CL1185Contig1_1	3361	775505	2659.97	0.27				
					CL1193Contig1_1_5	sp Q3ZU95 PA1_VESGE Phospholipase A1 OS = Vespula germanica OX = 30212 PE = 2 SV = 1	7.81e-21	
CL1193Contig1_1	7145	1682413	5770.66	0.58	CL1193Contig1_1_6	sp Q8K1C7 MOT14_MOUSE Monocarboxylate transporter 14 OS = Mus musculus OX = 10090 GN = Slc16a14 PE = 2 SV = 1	2.00e-37	Phospholipase
					CL1193Contig1_1_6	sp Q3ZU95 PA1_VESGE Phospholipase A1 OS = Vespula germanica OX = 30212 PE = 2 SV = 1	2.42e-43	
CL120Contig1_4	2266	123353	423.1	0.04	CL120Contig1_4_3	sp Q1HPS0 MLR_BOMMO Myosin regulatory light chain 2 OS = Bombyx mori OX = 7091 PE = 1 SV = 1	2.16e-80	Cytoskeleton
CL1275Contig1_1	3598	18623211	63877.43	6.39		U2-PSDTX-Ta1b Genbank Accession Number : MN607170		Venom peptide

CL12Contig1_1	665	23999 4	823.18	0.08	CL12Contig1_1_6	sp Q962Q7 RS23_SPOFR 40S ribosomal protein S23 OS = Spodoptera frugiperda OX = 7108 GN = RpS23 PE = 2 SV = 1	6.30e-90	Translation
CL1321Contig1_1	5444	42983 54	14743.31	1.47				
CL134Contig1_5	2180	10389 5	356.36	0.04	CL134Contig1_5_5	sp P29341 PABP1_MOUSE Polyadenylate-binding protein 1 OS = Mus musculus OX = 10090 GN = Pabpc1 PE = 1 SV = 2	4.23e-145	RNA Maturation
CL134Contig1_6	777	16279 1	558.37	0.06	CL134Contig1_6_4	sp P20965 PABPA_XENLA Polyadenylate-binding protein 1-A OS = Xenopus laevis OX = 8355 GN = pabpc1-a PE = 1 SV = 3	8.47e-97	Transcription
CL1626Contig1_2	3783	36811 9	1262.64	0.13	CL1626Contig1_2_1	sp P31689 DNJA1_HUMAN Dnaj homolog subfamily A member 1 OS = Homo sapiens OX = 9606 GN = DNAJA1 PE = 1 SV = 2	7.89e-160	Protein Maturation
					CL1626Contig1_2_4	sp Q962Q6 RS24_SPOFR 40S ribosomal protein S24 OS = Spodoptera frugiperda OX = 7108 GN = RpS24 PE = 2 SV = 1	1.20e-71	
CL1740Contig1_1	2749	15334 801	52598.21	5.26				
CL1762Contig1_1	394	21365 8	732.84	0.07				
CL1804Contig1_1	3123	43286 324	148471.66	14.8 5		U2-PSDTX-Ta1c Genbank Accession Numbers : MN607168		Venom peptide
CL1844Contig1_1	2272	33442 6	1147.08	0.11	CL1844Contig1_1_1	sp O18640 GBLP_DROME Guanine nucleotide-binding protein subunit beta-like protein OS = Drosophila melanogaster OX = 7227 GN = Rack1 PE = 1 SV = 2	0.0	Cellular Signaling
CL1Contig165_1	4280	38373 0	1316.19	0.13				
CL1Contig387_1	1547	73881 70	25341.35	2.53		U3-PSDTX-Ta1a Genbank Accession Numbers : MN607165		Venom peptide
CL1Contig399_4	1164	10573 5	362.67	0.04	CL1Contig399_4_4	sp P08879 NDKA_DROME Nucleoside diphosphate kinase OS = Drosophila melanogaster OX = 7227 GN = awd PE = 1 SV = 3	8.15e-82	Metabolism

CL2021Contig1_1	2516	11554 9	396.33	0.04	CL2021Contig1_1_6	sp Q9VBV3 TAKT_DROME Protein takeout OS = Drosophila melanogaster OX = 7227 GN = to PE = 2 SV = 1	1.26e-08	Circadian Rythm
CL2229Contig1_1	923	37149 3	1274.22	0.13	CL2229Contig1_1_6	sp Q7KF90 RL31_SPOFR 60S ribosomal protein L31 OS = Spodoptera frugiperda OX = 7108 GN = RpL31 PE = 2 SV = 1	7.20e-62	Translation
CL2248Contig1_1	2074	12327 2	422.82	0.04				
CL235Contig1_1	5969	46548 0	1596.59	0.16	CL235Contig1_1_6	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS = Apis mellifera OX = 7460 PE = 1 SV = 1	0.0	Peptide Maturation
CL2441Contig1_1	764	18951 1	650.02	0.07	CL2441Contig1_1_3	sp P68203 RS27A_SPOFR Ubiquitin-40S ribosomal protein S27a OS = Spodoptera frugiperda OX = 7108 PE = 2 SV = 2	1.08e-77	Protein Degradation
CL2597Contig1_2	2581	12108 4	415.32	0.04	CL2597Contig1_2_6	sp Q1HDZ5 EIF3B_BOMMO Eukaryotic translation initiation factor 3 subunit B OS = Bombyx mori OX = 7091 GN = eIF3-S9 PE = 2 SV = 1	0.0	Translation
CL260Contig1_1	15246	34144 8	1171.16	0.12	CL260Contig1_1_2	sp A2T929 RXRAA_DANRE Retinoic acid receptor RXR-alpha-A OS = Danio rerio OX = 7955 GN = rxraa PE = 2 SV = 2	1.05e-142	Transcription / Cytoskeleton
					CL260Contig1_1_6	sp Q75VN3 TCTP_BOMMO Translationally-controlled tumor protein homolog OS = Bombyx mori OX = 7091 GN = Tctp PE = 2 SV = 1	3.88e-91	
CL2682Contig1_1	1709	17304 2	593.53	0.06	CL2682Contig1_1_1	sp Q95ZE8 RL14_DROVI 60S ribosomal protein L14 OS = Drosophila virilis OX = 7244 GN = RpL14 PE = 3 SV = 1	3.28e-44	Translation
CL268Contig1_1	4197	30111 7	1032.83	0.1	CL268Contig1_1_2	sp P62282 RS11_RAT 40S ribosomal protein S11 OS = Rattus norvegicus OX = 10116 GN = Rps11 PE = 1 SV = 3	4.09e-70	Unknown
					CL268Contig1_1_4	sp Q7Z3D4 LYSM3_HUMAN LysM and putative peptidoglycan-binding domain-containing protein 3 OS = Homo sapiens OX = 9606 GN = LYSMD3 PE = 1 SV = 2	2.46e-12	
CL2720Contig1_1	2730	11783 1	404.16	0.04	CL2720Contig1_1_6	sp O17389 TYB_CAEEL Thymosin beta OS = Caenorhabditis elegans OX = 6239 GN = tth-1 PE = 1 SV = 2	3.19e-15	Cytoskeleton

CL2890Contig1_1	2494	105857	363.09	0.04	CL2890Contig1_1_1	sp P55828 RS20_DROME 40S ribosomal protein S20 OS = <i>Drosophila melanogaster</i> OX = 7227 GN = RpS20 PE = 1 SV = 1	8.78e-58	Translation
					CL2890Contig1_1_4	sp P41374 IF2A_DROME Eukaryotic translation initiation factor 2 subunit 1 OS = <i>Drosophila melanogaster</i> OX = 7227 GN = eIF2alpha PE = 2 SV = 1	3.68e-146	
CL3020Contig1_1	1913	279908	960.08	0.1	CL3020Contig1_1_2	sp Q56FG6 RL5_LYSTE 60S ribosomal protein L5 OS = <i>Lysiphlebus testaceipes</i> OX = 77504 GN = RpL5 PE = 2 SV = 1	2.53e-161	Translation
CL3328Contig1_1	4478	141078	483.9	0.05	CL3328Contig1_1_5	sp Q9VFC2 SP88E_DROME Serine protease inhibitor 88Ea OS = <i>Drosophila melanogaster</i> OX = 7227 GN = Spn88Ea PE = 2 SV = 1	6.23e-62	Immunity
CL349Contig1_1	1286	222310	762.52	0.08				
CL3564Contig1_1	8161	1583033	5429.79	0.54	CL3564Contig1_1_3	sp Q3ULZ2 FHDC1_MOUSE FH2 domain-containing protein 1 OS = <i>Mus musculus</i> OX = 10090 GN = Fhdc1 PE = 1 SV = 3	1.18e-88	Cytoskeleton
CL3611Contig1_1	2022	135922	466.21	0.05	CL3611Contig1_1_1	sp P54985 PPIA_BLAGE Peptidyl-prolyl cis-trans isomerase OS = <i>Blattella germanica</i> OX = 6973 GN = CYPA PE = 2 SV = 1	3.85e-92	Protein Maturation
CL3706Contig1_1	2966	323293	1108.89	0.11	CL3706Contig1_1_2	sp Q63159 COQ3_RAT Ubiquinone biosynthesis O-methyltransferase. mitochondrial OS = <i>Rattus norvegicus</i> OX = 10116 GN = Coq3 PE = 2 SV = 2	2.12e-52	Metabolism
					CL3706Contig1_1_5	sp Q03168 ASPP_AEDAE Lysosomal aspartic protease OS = <i>Aedes aegypti</i> OX = 7159 GN = AAEL006169 PE = 1 SV = 2	1.09e-150	
CL3812Contig1_1	7028	214658	736.27	0.07	CL3812Contig1_1_6	sp B1A4F7 VDDP4_VESVU Venom dipeptidyl peptidase 4 OS = <i>Vespula vulgaris</i> OX = 7454 PE = 1 SV = 1	7.39e-125	Peptide Maturation
CL4093Contig1_1	1360	188957	648.12	0.06	CL4093Contig1_1_2	sp O96647 RL10_BOMMA 60S ribosomal protein L10 OS = <i>Bombyx mandarina</i> OX = 7092 GN = RpL10 PE = 2 SV = 1	4.54e-74	Translation

CL4128Contig1_1	2237	10249 6	351.56	0.04					
CL4129Contig1_1	1117	24621 9	844.53	0.08	CL4129Contig1_1_6	sp Q10416 HYTA_APIME Hymenoptaecin OS = Apis mellifera OX = 7460 PE = 2 SV = 1	9.76e-18	Immunity	
CL4173Contig1_1	3705	34349 7	1178.19	0.12	CL4173Contig1_1_4	sp P0CG71 UBIQ1_CAEEL Polyubiquitin-A OS = Caenorhabditis elegans OX = 6239 GN = ubq-1 PE = 3 SV = 1	0.0	Protein Degradation	
CL4485Contig1_1	3472	17752 7	608.92	0.06	CL4485Contig1_1_4	sp Q962R9 RS10_SPOFR 40S ribosomal protein S10 OS = Spodoptera frugiperda OX = 7108 GN = RpS10 PE = 2 SV = 1	4.44e-70	Translation	
CL4642Contig1_1	1666	70936 5	2433.11	0.24	CL4642Contig1_1_2	sp Q9W1C9 PEB3_DROME Ejaculatory bulb-specific protein 3 OS = Drosophila melanogaster OX = 7227 GN = EbpIII PE = 2 SV = 2	1.26e-23	Secreted Protein	
CL479Contig1_5	2010	26274 73	9012.21	0.9	CL479Contig1_5_5	sp Q9U639 HSP7D_MANSE Heat shock 70 kDa protein cognate 4 OS = Manduca sexta OX = 7130 PE = 2 SV = 1	0.0	Protein Maturation	
CL479Contig1_6	1340	22014 2	755.08	0.08	CL479Contig1_6_5	sp Q9U639 HSP7D_MANSE Heat shock 70 kDa protein cognate 4 OS = Manduca sexta OX = 7130 PE = 2 SV = 1	2.18e-106	Protein Maturation	
CL4957Contig1_1	3657	17247 4	591.58	0.06	CL4957Contig1_1_3	sp A8CAG3 RL17_PHLPP 60S ribosomal protein L17 OS = Phlebotomus papatasi OX = 29031 GN = RpL17 PE = 2 SV = 1	6.52e-100	Translation /Secretion	
					CL4957Contig1_1_6	sp Q7L1I2 SV2B_HUMAN Synaptic vesicle glycoprotein 2B OS = Homo sapiens OX = 9606 GN = SV2B PE = 1 SV = 1	4.67e-49		
CL5058Contig1_1	1235	27966 4	959.24	0.1	CL5058Contig1_1_4	sp Q02878 RL6_HUMAN 60S ribosomal protein L6 OS = Homo sapiens OX = 9606 GN = RPL6 PE = 1 SV = 3	7.07e-58	Translation	
CL521Contig1_1	852	88633 7	3040.13	0.3	CL521Contig1_1_6	sp P46782 RS5_HUMAN 40S ribosomal protein S5 OS = Homo sapiens OX = 9606 GN = RPS5 PE = 1 SV = 4	7.92e-122	Translation	
CL552Contig1_4	673	20935 7	718.09	0.07	CL552Contig1_4_6	sp Q962T5 RL24_SPOFR 60S ribosomal protein L24 OS = Spodoptera frugiperda OX = 7108 GN = RpL24 PE = 2 SV = 1	2.07e-72	Translation	
CL573Contig1_1	600	26496 2	908.82	0.09	CL573Contig1_1_2	sp Q6XIM8 RS15A_DROYA 40S ribosomal protein S15a OS = Drosophila yakuba OX = 7245 GN = RpS15Aa PE = 2 SV = 3	5.18e-84	Translation	

CL575Contig1_1	1770	193468	663.59	0.07				
CL5Contig7_4	1923	147949	507.46	0.05	CL5Contig7_4_2	sp P47830 RL27A_XENLA 60S ribosomal protein L27a OS = Xenopus laevis OX = 8355 GN = rpl27a PE = 2 SV = 2	6.25e-76	RNA Maturation /Translation
					CL5Contig7_4_4	sp Q6AXT8 SF3A2_RAT Splicing factor 3A subunit 2 OS = Rattus norvegicus OX = 10116 GN = Sf3a2 PE = 2 SV = 1	4.48e-112	
CL660Contig1_1	3987	133617	458.3	0.05	CL660Contig1_1_5	sp Q1HRV8 ELVL1_AEDAE Elongation of very long chain fatty acids protein AAEL008004 OS = Aedes aegypti OX = 7159 GN = AAEL008004 PE = 2 SV = 2	2.28e-105	Metabolism
CL669Contig1_1	16668	180230	618.19	0.06	CL669Contig1_1_6	sp B6RSP1 PKHA7_DANRE Pleckstrin homology domain-containing family A member 7 OS = Danio rerio OX = 7955 GN = plekha7 PE = 2 SV = 2	2.99e-37	Secretion
CL955Contig1_1	694	197268	676.63	0.07	CL955Contig1_1_6	sp Q8WRP6 PBGp9_SOLGI Pheromone-binding protein Gp-9 OS = Solenopsis globularia littoralis OX = 176593 GN = Gp-9 PE = 3 SV = 1	3.49e-09	Chemoreception
k25_Locus_10039_Transcript_1_1	2851	152489	523.04	0.05				
k25_Locus_108_Transcript_2_5	3260	209726	719.36	0.07	k25_Locus_108_Transcript_2_5_4	sp P29240 5NTD_DIPOM 5'-nucleotidase OS = Diplobatis ommata OX = 1870830 PE = 2 SV = 1	2.49e-71	Metabolism
					k25_Locus_108_Transcript_2_5_6	sp Q9XZ43 5NTD_LUTLO Protein 5NUC OS = Lutzomyia longipalpis OX = 7200 GN = 5NUC PE = 1 SV = 1	1.09e-34	
k25_Locus_11_Transcript_31_1	6299	5271032	18079.59	1.81	k25_Locus_11_Transcript_31_1_4	sp Q3ZU95 PA1_VESGE Phospholipase A1 OS = Vespula germanica OX = 30212 PE = 2 SV = 1	1.55e-19	Phospholipase
					k25_Locus_11_Transcript_31_1_6	sp Q68KK0 PA1_SOLIN Phospholipase A1 OS = Solenopsis invicta OX = 13686 PE = 1 SV = 1	2.07e-26	
k25_Locus_1163_Transcript_8_3	710	121774	417.68	0.04	k25_Locus_1163_Transcript_8_3_5	sp Q9NB33 RL44_OCHTR 60S ribosomal protein L44 OS = Ochlerotatus triseriatus OX = 7162 GN = RpL44 PE = 3 SV = 3	4.23e-58	Translation
k25_Locus_1168_Transcript_5_1	5138	102343	351.04	0.04	k25_Locus_1168_Transcript_5_1_1	sp Q9CR60 GOT1B_MOUSE Vesicle transport protein GOT1B OS = Mus musculus OX = 10090 GN = Golt1b PE = 1 SV = 1	7.79e-34	Metabolism

					k25_Locus_1168_Transcript_5_1_6	sp Q7PQV7 ADT2_ANOGA ADP.ATP carrier protein 2 OS = Anopheles gambiae OX = 7165 GN = AGAP002358 PE = 3 SV = 2	3.43e-173	
k25_Locus_1204_Transcript_6_1	2284	252383	865.67	0.09	k25_Locus_1204_Transcript_6_1_5	sp Q4GXG7 RL18_TIMBA 60S ribosomal protein L18 OS = Timarcha balearica OX = 79517 GN = RpL18 PE = 2 SV = 1	1.45e-105	Translation
k25_Locus_1236_Transcript_1_1	1266	412434	1414.64	0.14	k25_Locus_1236_Transcript_1_1_4	sp P41822 FRI_AEDAE Ferritin subunit OS = Aedes aegypti OX = 7159 GN = FERH PE = 1 SV = 2	1.46e-41	Metabolism
k25_Locus_126_Transcript_5_1	5493	3778623	12960.64	1.3				Metabolism
k25_Locus_1289_Transcript_17_4	1122	145672	499.65	0.05	k25_Locus_1289_Transcript_17_4_1	sp P47830 RL27A_XENLA 60S ribosomal protein L27a OS = Xenopus laevis OX = 8355 GN = rpl27a PE = 2 SV = 2	6.25e-76	Translation
k25_Locus_1373_Transcript_14_1	2718	541864	1858.59	0.19	k25_Locus_1373_Transcript_14_1_2	sp Q9V447 KRH2_DROME Krueppel homolog 2 OS = Drosophila melanogaster OX = 7227 GN = Kr-h2 PE = 1 SV = 1	6.94e-59	Protein Maturation / Translation
					k25_Locus_1373_Transcript_14_1_5	sp Q5R465 RS3_PONAB 40S ribosomal protein S3 OS = Pongo abelii OX = 9601 GN = RPS3 PE = 2 SV = 1	7.22e-52	
k25_Locus_165_Transcript_2_1	833	433100	1485.53	0.15	k25_Locus_165_Transcript_2_1_1	sp P58375 RL30_SPOFR 60S ribosomal protein L30 OS = Spodoptera frugiperda OX = 7108 GN = RpL30 PE = 3 SV = 1	3.37e-59	Translation
k25_Locus_182_Transcript_10_1	4550	136590	468.5	0.05	k25_Locus_182_Transcript_10_1_3	sp Q7KN62 TERA_DROME Transitional endoplasmic reticulum ATPase TER94 OS = Drosophila melanogaster OX = 7227 GN = TER94 PE = 1 SV = 1	0.0	Secretion
k25_Locus_1873_Transcript_1_1	957	110742	379.84	0.04	k25_Locus_1873_Transcript_1_1_3	sp Q8VZ67 Y4919_ARATH Uncharacterized zinc finger CCHC domain-containing protein At4g19190 OS = Arabidopsis thaliana OX = 3702 GN = At4g19190 PE = 2 SV = 1	0.54	Oxidative stress
					k25_Locus_1873_Transcript_1_1_6	sp Q9V3P0 PRDX1_DROME Peroxiredoxin 1 OS = Drosophila melanogaster OX = 7227 GN = Jafrac1 PE = 1 SV = 1	4.60e-105	

k25_Locus_21_Transcript_3_1	3913	58584 8	2009.45	0.2	k25_Locus_21_Transcript_3_1_5	sp P29844 BIP_DROME Endoplasmic reticulum chaperone BiP OS = Drosophila melanogaster OX = 7227 GN = Hsc70-3 PE = 1 SV = 2	0.0	Protein Maturation
k25_Locus_217_Transcript_4_2	1995	34872 4	1196.12	0.12	k25_Locus_217_Transcript_4_2_4	sp Q8TCT9 HM13_HUMAN Minor histocompatibility antigen H13 OS = Homo sapiens OX = 9606 GN = HM13 PE = 1 SV = 1	5.37e-131	Protein Maturation
k25_Locus_2186_Transcript_2_3	3394	14393 4	493.69	0.05				
k25_Locus_2191_Transcript_12_3	3159	13603 4	466.6	0.05	k25_Locus_2191_Transcript_12_3_1	sp Q9V778 ADAS_DROME Alkyldihydroxyacetonephosphate synthase OS = Drosophila melanogaster OX = 7227 GN = ADPS PE = 2 SV = 1	0.0	Metabolism
k25_Locus_254_Transcript_4_4	3407	11817 1	405.33	0.04	k25_Locus_254_Transcript_4_4_1	sp O08623 SQSTM_RAT Sequestosome-1 OS = Rattus norvegicus OX = 10116 GN = Sqstm1 PE = 1 SV = 1	9.72e-30	Protein Degradation
k25_Locus_2739_Transcript_4_1	2745	16403 3	562.63	0.06	k25_Locus_2739_Transcript_4_1_5	sp O96567 DDC_DROSI Aromatic-L-amino-acid decarboxylase OS = Drosophila simulans OX = 7240 GN = Ddc PE = 3 SV = 2	3.23e-170	Metabolism
k25_Locus_317_Transcript_5_2	2304	14700 4	504.22	0.05	k25_Locus_317_Transcript_5_2_5	sp Q9VYY4 C4G15_DROME Cytochrome P450 4g15 OS = Drosophila melanogaster OX = 7227 GN = Cyp4g15 PE = 2 SV = 1	4.21e-175	Metabolism
k25_Locus_3207_Transcript_3_1	3472	13091 4	449.03	0.04	k25_Locus_3207_Transcript_3_1_3	sp Q8MQS8 SP34_APIME Venom serine protease 34 OS = Apis mellifera OX = 7460 PE = 2 SV = 1	1.18e-90	Secreted Protein
k25_Locus_3286_Transcript_1_1	894	23134 1	793.5	0.08	k25_Locus_3286_Transcript_1_1_3	sp Q962U0 RL13A_SPOFR 60S ribosomal protein L13a OS = Spodoptera frugiperda OX = 7108 GN = RpL13A PE = 2 SV = 1	1.91e-105	Translation
k25_Locus_3385_Transcript_1_1	1257	48992 3	1680.43	0.17	k25_Locus_3385_Transcript_1_1_3	sp Q5UAP4 RSSA_BOMMO 40S ribosomal protein SA OS = Bombyx mori OX = 7091 PE = 2 SV = 1	1.39e-124	Translation
k25_Locus_3520_Transcript_10_1	3033	36303 3	1245.2	0.12	k25_Locus_3520_Transcript_10_1_2	sp Q3UST5 CP089_MOUSE UPF0764 protein C16orf89 homolog OS = Mus musculus OX = 10090 PE = 2 SV = 2	1.05e-18	Unknown
k25_Locus_3576_Transcript_3_1	1701	19106 4	655.35	0.07	k25_Locus_3576_Transcript_3_1_5	sp Q5R8Z6 MCFD2_PONAB Multiple coagulation factor deficiency protein 2 homolog OS = Pongo abelii OX = 9601 GN = MCFD2 PE = 2 SV = 1	2.41e-19	Secretion

k25_Locus_402_Transcript_3_1	554	16700 6	572.83	0.06	k25_Locus_402_Transcript_3_1_4	sp P80455 RS12_DROME 40S ribosomal protein S12 OS = Drosophila melanogaster OX = 7227 GN = RpS12 PE = 1 SV = 2	1.23e-58	Translation
k25_Locus_4241_Transcript_1_1	674	12947 9	444.11	0.04	k25_Locus_4241_Transcript_1_1_4	sp Q962Q5 RS25_SPOFR 40S ribosomal protein S25 OS = Spodoptera frugiperda OX = 7108 GN = RpS25 PE = 3 SV = 1	2.31e-43	Translation
k25_Locus_433_Transcript_9_1	4376	66650 2	2286.1	0.23		MKLITLFLVVVLAIFIRPLMPLQMRNTESFAEGSADAF AETNSDSIKI		Venom peptide
k25_Locus_45_Transcript_1_1	5655	12758 16	4376.04	0.44	k25_Locus_45_Transcript_1_1_2	sp P07709 NU6M_DROYA NADH-ubiquinone oxidoreductase chain 6 OS = Drosophila yakuba OX = 7245 GN = mt:ND6 PE = 3 SV = 2	6.22e-08	Metabolism
k25_Locus_4540_Transcript_2_1	2437	15191 2	521.06	0.05	k25_Locus_4540_Transcript_2_1_2	sp Q10714 ACE_DROME Angiotensin-converting enzyme OS = Drosophila melanogaster OX = 7227 GN = Ance PE = 1 SV = 3	0.0	Peptide Maturation
k25_Locus_4724_Transcript_2_1	971	12573 4	431.27	0.04	k25_Locus_4724_Transcript_2_1_5	sp Q962S0 RS7_SPOFR 40S ribosomal protein S7 OS = Spodoptera frugiperda OX = 7108 GN = RpS7 PE = 2 SV = 1	4.60e-100	Translation
k25_Locus_54_Transcript_3_1	2425	39087 4	1340.69	0.13	k25_Locus_54_Transcript_3_1_4 k25_Locus_54_Transcript_3_1_6	sp Q9P735 FAL1_NEUCR ATP-dependent RNA helicase fal-1 OS = Neurospora crassa (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987) OX = 367110 GN = fal-1 PE = 3 SV = 2 sp Q02748 IF4A_DROME Eukaryotic initiation factor 4A OS = Drosophila melanogaster OX = 7227 GN = eIF4A PE = 1 SV = 3	1.80e-51 8.73e-84	Translation
k25_Locus_547_Transcript_29_1	8082	14492 1	497.08	0.05	k25_Locus_547_Transcript_29_1_3	sp Q767L8 MDC1_PIG Mediator of DNA damage checkpoint protein 1 OS = Sus scrofa OX = 9823 GN = MDC1 PE = 3 SV = 1	7.17e-42	DNA Repair
k25_Locus_547_Transcript_29_2	4580	13671 2	468.92	0.05	k25_Locus_547_Transcript_29_2_4	sp P20735 GGT1_PIG Glutathione hydrolase 1 proenzyme OS = Sus scrofa OX = 9823 GN = GGT1 PE = 2 SV = 1	3.37e-97	Metabolism
k25_Locus_575_Transcript_4_1	3876	11240 9	385.56	0.04	k25_Locus_575_Transcript_4_1_2	sp Q921M4 GOGA2_MOUSE Golgin subfamily A member 2 OS = Mus musculus OX = 10090 GN = Golga2 PE = 1 SV = 3	3.25e-47	Translation

					k25_Locus_575_T ranscript_4_1_5	sp Q69CJ9 RL35_OPHHA 60S ribosomal protein L35 OS = Ophiophagus hannah OX = 8665 GN = RPL35 PE = 2 SV = 3	4.69e- 45	
k25_Locus_5867 _Transcript_4_1	4342	20252 1	694.64	0.07	k25_Locus_5867_ Transcript_4_1_5	sp P35415 MYSP1_DROME Paramyosin. long form OS = Drosophila melanogaster OX = 7227 GN = Prm PE = 1 SV = 1	0.0	Cytoskeleton
k25_Locus_737_ Transcript_1_1	1004	12314 1	422.37	0.04				
k25_Locus_7606 _Transcript_1_1	919	44160 0	1514.68	0.15	k25_Locus_7606_ Transcript_1_1_6	sp Q945U1 RS15_ELAOL 40S ribosomal protein S15 OS = Elaeis oleifera OX = 80265 GN = RPS15 PE = 2 SV = 1	4.01e- 77	Translation
k25_Locus_766_ Transcript_35_1	5458	22405 2	768.5	0.08				
k25_Locus_810_ Transcript_1_1	1343	41762 9	1432.46	0.14	k25_Locus_810_T ranscript_1_1_6	sp Q95V39 RL8_SPOFR 60S ribosomal protein L8 OS = Spodoptera frugiperda OX = 7108 GN = RpL8 PE = 2 SV = 1	3.25e- 157	Translation
k25_Locus_97_ Transcript_10_3	1564	13667 40	4687.9	0.47	k25_Locus_97_Tr anscript_10_3_5	sp Q9U639 HSP7D_MANSE Heat shock 70 kDa protein cognate 4 OS = Manduca sexta OX = 7130 PE = 2 SV = 1	0.0	Protein Maturation
k25_Locus_9770 _Transcript_4_1	684	33321 7	1142.93	0.11				
k31_Locus_1107 9_Transcript_6_ 1	2470	10521 6	360.89	0.04	k31_Locus_11079 _Transcript_6_1_ 2	sp Q58ED9 NAA20_DANRE N-alpha-acetyltransferase 20 OS = Danio rerio OX = 7955 GN = naa20 PE = 2 SV = 1	3.15e- 87	Protein Maturation
					k31_Locus_11079 _Transcript_6_1_ 3	sp Q5M8Y1 SPCS2_XENTR Probable signal peptidase complex subunit 2 OS = Xenopus tropicalis OX = 8364 GN = spcs2 PE = 2 SV = 1	1.72e- 58	
k31_Locus_1149 _Transcript_6_1	4540	25590 7	877.76	0.09	k31_Locus_1149_ Transcript_6_1_2	sp P9WQP7 3BHS_MYCTU 3 beta-hydroxysteroid dehydrogenase/Delta 5-->4-isomerase OS = Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv) OX = 83332 GN = Rv1106c PE = 1 SV = 1	4.48e- 04	Translation
					k31_Locus_1149_ Transcript_6_1_6	sp Q4GXG7 RL18_TIMBA 60S ribosomal protein L18 OS = Timarcha balearica OX = 79517 GN = RpL18 PE = 2 SV = 1	1.45e- 105	

k31_Locus_1256_Transcript_1_2	1926	13884 8	476.25	0.05	k31_Locus_1256_Transcript_1_2_5	sp P70195 PSB7_MOUSE Proteasome subunit beta type-7 OS = Mus musculus OX = 10090 GN = Psmb7 PE = 1 SV = 1	7.33e-108	Protein Degradation
k31_Locus_1299_Transcript_4_1	2874	15767 87	5408.36	0.54				
k31_Locus_142_Transcript_2_1	492	34049 3	1167.89	0.12	k31_Locus_142_Transcript_2_1_2	sp Q963B7 RL9_SPOFR 60S ribosomal protein L9 OS = Spodoptera frugiperda OX = 7108 GN = RpL9 PE = 2 SV = 1	2.22e-98	Translation
k31_Locus_1543_Transcript_1_1	2987	24572 9	842.85	0.08	k31_Locus_1543_Transcript_1_1_4	sp Q962R6 RS13_SPOFR 40S ribosomal protein S13 OS = Spodoptera frugiperda OX = 7108 GN = RpS13 PE = 2 SV = 3	1.70e-87	Translation
k31_Locus_1653_Transcript_2_1	1702	48776 4	1673.03	0.17				
k31_Locus_1695_Transcript_9_1	4325	14030 9	481.26	0.05	k31_Locus_1695_Transcript_9_1_6	sp P30151 EF1B_XENLA Elongation factor 1-beta OS = Xenopus laevis OX = 8355 GN = eef1b PE = 1 SV = 3	7.13e-75	Translation
k31_Locus_179_Transcript_1_1	5057	17493 6	600.03	0.06	k31_Locus_179_Transcript_1_1_4	sp P41824 YBOXH_APLCA Y-box factor homolog OS = Aplysia californica OX = 6500 PE = 2 SV = 1	3.61e-42	Transcription
k31_Locus_1890_Transcript_11_1	2810	75410 3	2586.57	0.26	k31_Locus_1890_Transcript_11_1_2	sp Q5BLY4 ICA_APIIME Icarapin-like OS = Apis mellifera OX = 7460 PE = 2 SV = 1	7.08e-23	Secreted Protein
k31_Locus_1920_Transcript_1_1	866	24388 4	836.52	0.08	k31_Locus_1920_Transcript_1_1_4	sp P39018 RS19A_DROME 40S ribosomal protein S19a OS = Drosophila melanogaster OX = 7227 GN = RpS19a PE = 1 SV = 3	2.43e-62	Translation
k31_Locus_21_Transcript_1_1	2877	38506 7	1320.78	0.13	k31_Locus_21_Transcript_1_1_4	sp P29844 BIP_DROME Endoplasmic reticulum chaperone BiP OS = Drosophila melanogaster OX = 7227 GN = Hsc70-3 PE = 1 SV = 2	0.0	Protein Maturation
k31_Locus_2212_Transcript_4_2	1771	29924 9	1026.42	0.1	k31_Locus_2212_Transcript_4_2_6	sp P30050 RL12_HUMAN 60S ribosomal protein L12 OS = Homo sapiens OX = 9606 GN = RPL12 PE = 1 SV = 1	2.65e-86	Translation
k31_Locus_222_Transcript_2_1	2655	10689 9	366.66	0.04	k31_Locus_222_Transcript_2_1_5	sp P13008 RS26_DROME 40S ribosomal protein S26 OS = Drosophila melanogaster OX = 7227 GN = RpS26 PE = 1 SV = 1	3.98e-58	Translation
k31_Locus_261_Transcript_21_1	2728	46402 91	15916.15	1.59	k31_Locus_261_Transcript_21_1_5	sp P35778 VA3_SOLIN Venom allergen 3 OS = Solenopsis invicta OX = 13686 PE = 1 SV = 2	1.05e-74	Venom allergen

k31_Locus_263_Transcript_6_1	2222	530403	1819.28	0.18	k31_Locus_263_Transcript_6_1_5	sp P09180 RL4_DROME 60S ribosomal protein L4 OS = <i>Drosophila melanogaster</i> OX = 7227 GN = RpL4 PE = 1 SV = 2	4.42e-168	Translation
k31_Locus_3_Transcript_5_1	3218	139260	477.66	0.05	k31_Locus_3_Transcript_5_1_6	sp B4MGF8 TMEDA_DROVI Transmembrane emp24 domain-containing protein bai OS = <i>Drosophila virilis</i> OX = 7244 GN = bai PE = 3 SV = 1	2.19e-83	Secretion
k31_Locus_3122_Transcript_1_1	809	299719	1028.03	0.1	k31_Locus_3122_Transcript_1_1_5	sp Q9W6Y0 RS30_ORYLA 40S ribosomal protein S30 OS = <i>Oryzias latipes</i> OX = 8090 GN = fau PE = 3 SV = 2	3.40e-19	Translation
k31_Locus_3178_Transcript_1_1	3098	116674	400.19	0.04	k31_Locus_3178_Transcript_1_1_4	sp P05300 LAMP1_CHICK Lysosome-associated membrane glycoprotein 1 OS = <i>Gallus gallus</i> OX = 9031 GN = LAMP1 PE = 2 SV = 1	1.54e-23	Protein Degradation
k31_Locus_3301_Transcript_10_1	3031	319224	1094.94	0.11	k31_Locus_3301_Transcript_10_1_5	sp Q3UST5 CP089_MOUSE UPF0764 protein C16orf89 homolog OS = <i>Mus musculus</i> OX = 10090 PE = 2 SV = 2	1.05e-18	Unknown
k31_Locus_3808_Transcript_1_1	1088	124126	425.75	0.04	k31_Locus_3808_Transcript_1_1_4	sp P41973 SODC_DROWI Superoxide dismutase [Cu-Zn] OS = <i>Drosophila willistoni</i> OX = 7260 GN = Sod1 PE = 3 SV = 2	8.40e-63	Oxidative stress
k31_Locus_431_Transcript_14_2	3857	102949	353.11	0.04	k31_Locus_431_Transcript_14_2_6	sp Q9VAF0 S39AD_DROME Zinc transporter ZIP13 homolog OS = <i>Drosophila melanogaster</i> OX = 7227 GN = Zip99C PE = 2 SV = 1	2.00e-52	Metabolism
k31_Locus_49_Transcript_12_1	5983	7923500	27177.53	2.72		U4-PSDIX-Ta1a Genbank Accession Numbers : MN607167		Venom peptide
k31_Locus_5020_Transcript_1_1	970	187814	644.2	0.06	k31_Locus_5020_Transcript_1_1_4	sp P46222 RL11_DROME 60S ribosomal protein L11 OS = <i>Drosophila melanogaster</i> OX = 7227 GN = RpL11 PE = 1 SV = 2	3.13e-106	Translation
k31_Locus_597_Transcript_7_1	3991	105743	3627.02	0.36	k31_Locus_597_Transcript_7_1_3	sp P29520 EF1A_BOMMO Elongation factor 1-alpha OS = <i>Bombyx mori</i> OX = 7091 PE = 2 SV = 1	0.0	Translation
k31_Locus_647_Transcript_12_3	840	133193	456.85	0.05	k31_Locus_647_Transcript_12_3_3	sp B4PEU8 RS9_DROYA 40S ribosomal protein S9 OS = <i>Drosophila yakuba</i> OX = 7245 GN = RpS9 PE = 2 SV = 1	7.76e-111	Translation
k31_Locus_66_Transcript_1_1	1672	1619399	5554.52	0.56	k31_Locus_66_Transcript_1_1_6	sp Q75VN3 TCTP_BOMMO Translationally-controlled tumor protein homolog OS = <i>Bombyx mori</i> OX = 7091 GN = Tctp PE = 2 SV = 1	1.48e-91	Cytoskeleton

k31_Locus_725_Transcript_46_1	5204	26121 6	895.97	0.09				
k31_Locus_725_Transcript_46_2	3694	77183 3	2647.38	0.26				
k31_Locus_82_Transcript_7_2	3549	60242 5	2066.31	0.21	k31_Locus_82_Transcript_7_2_2	sp Q27294 CAZ_DROME RNA-binding protein cabeza OS = Drosophila melanogaster OX = 7227 GN = caz PE = 2 SV = 2	1.71e-39	Transcription / Apoptosis
					k31_Locus_82_Transcript_7_2_6	sp Q9D2C7 BI1_MOUSE Bax inhibitor 1 OS = Mus musculus OX = 10090 GN = Tmbim6 PE = 1 SV = 1	1.59e-55	
k31_Locus_852_Transcript_8_1	3795	20186 5	692.39	0.07	k31_Locus_852_Transcript_8_1_1	sp P07602 SAP_HUMAN Prosaposin OS = Homo sapiens OX = 9606 GN = PSAP PE = 1 SV = 2	6.86e-38	Cellular Signaling
k31_Locus_853_Transcript_3_1	2089	64331 8	2206.57	0.22	k31_Locus_853_Transcript_3_1_3	sp Q8TGP1 YG123_YEAST Putative uncharacterized protein YGL123C-A OS = Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX = 559292 GN = YGL123C-A PE = 5 SV = 1	8.72e-05	Translation
					k31_Locus_853_Transcript_3_1_5	sp P31009 RS2_DROME 40S ribosomal protein S2 OS = Drosophila melanogaster OX = 7227 GN = RpS2 PE = 1 SV = 2	1.43e-131	
k31_Locus_892_Transcript_3_1	1895	33799 8	1159.33	0.12	k31_Locus_892_Transcript_3_1_6	sp P19889 RLA0_DROME 60S acidic ribosomal protein P0 OS = Drosophila melanogaster OX = 7227 GN = RpLP0 PE = 1 SV = 1	2.17e-155	Translation
k31_Locus_899_Transcript_10_2	2221	11228 4	385.13	0.04	k31_Locus_899_Transcript_10_2_5	sp Q58FK9 KAT3_RAT Kynurenine--oxoglutarate transaminase 3 OS = Rattus norvegicus OX = 10116 GN = Kyat3 PE = 2 SV = 1	5.44e-128	Protein Maturation
k31_Locus_984_Transcript_1_1	2900	12998 5	445.85	0.04	k31_Locus_984_Transcript_1_1_2	sp Q8CBY8 DCTN4_MOUSE Dynactin subunit 4 OS = Mus musculus OX = 10090 GN = Dctn4 PE = 1 SV = 1	4.98e-101	Cytoskeleton/ Translation
					k31_Locus_984_Transcript_1_1_4	sp C0HKA1 RS14B_DROME 40S ribosomal protein S14b OS = Drosophila melanogaster OX = 7227 GN = RpS14b PE = 2 SV = 1	7.47e-85	
k37_Locus_1_Transcript_1_2	4050	13660 2	468.54	0.05	k37_Locus_1_Transcript_1_2_5	sp Q4U3L0 G3P_GLOMM Glyceraldehyde-3-phosphate dehydrogenase OS = Glossina morsitans morsitans OX = 37546 GN = Gapdh PE = 2 SV = 1	0.0	Metabolism

k37_Locus_10_Transcript_14_1	4476	54127 7	1856.57	0.19	k37_Locus_10_Transcript_14_1_6	sp Q68KK0 PA1_SOLIN Phospholipase A1 OS = Solenopsis invicta OX = 13686 PE = 1 SV = 1	1.54e-51	Phospholipase
k37_Locus_1071_Transcript_2_1	2105	10374 6	355.85	0.04	k37_Locus_1071_Transcript_2_1_5	sp Q4GXG7 RL18_TIMBA 60S ribosomal protein L18 OS = Timarcha balearica OX = 79517 GN = RpL18 PE = 2 SV = 1	8.75e-70	Translation
k37_Locus_1280_Transcript_9_1	9989	13389 1	459.24	0.05	k37_Locus_1280_Transcript_9_1_6	sp Q24238 APH4_DROME Alkaline phosphatase 4 OS = Drosophila melanogaster OX = 7227 GN = Alp4 PE = 2 SV = 3	1.50e-66	Metabolism
k37_Locus_1441_Transcript_3_4	3432	13890 5	476.44	0.05	k37_Locus_1441_Transcript_3_4_6	sp Q2PQM7 IDGF4_GLOMM Chitinase-like protein Idgf4 OS = Glossina morsitans morsitans OX = 37546 GN = Idgf4 PE = 2 SV = 1	9.23e-120	Cellular Signaling
k37_Locus_1903_Transcript_4_2	4559	15433 7	529.37	0.05	k37_Locus_1903_Transcript_4_2_4	sp P28648 CD63_RAT CD63 antigen OS = Rattus norvegicus OX = 10116 GN = Cd63 PE = 1 SV = 2	3.33e-29	Cellular Signaling
k37_Locus_229_Transcript_6_1	4838	23686 7	812.45	0.08				
k37_Locus_2459_Transcript_1_1	1204	30142 9	1033.9	0.1	k37_Locus_2459_Transcript_1_1_4	sp Q962U1 RL13_SPOFR 60S ribosomal protein L13 OS = Spodoptera frugiperda OX = 7108 GN = RpL13 PE = 2 SV = 1	6.25e-94	Translation
k37_Locus_2720_Transcript_1_1	517	15861 3	544.04	0.05				
k37_Locus_3057_Transcript_4_1	4291	12819 4	439.7	0.04				
k37_Locus_391_Transcript_6_2	4093	12387 9	424.9	0.04	k37_Locus_391_Transcript_6_2_6	sp Q1HRV8 ELVL1_AEDAE Elongation of very long chain fatty acids protein AAEL008004 OS = Aedes aegypti OX = 7159 GN = AAEL008004 PE = 2 SV = 2	2.63e-105	Metabolism
k37_Locus_4169_Transcript_1_1	1410	26823 3	920.04	0.09	k37_Locus_4169_Transcript_1_1_3	sp P15357 RS27A_DROME Ubiquitin-40S ribosomal protein S27a OS = Drosophila melanogaster OX = 7227 GN = RpS27A PE = 1 SV = 2	8.86e-83	Protein Degradation
k37_Locus_4367_Transcript_1_1	1554	13859 3	475.37	0.05	k37_Locus_4367_Transcript_1_1_5	sp Q4GXU6 RS4_CARGR 40S ribosomal protein S4 OS = Carabus granulatus OX = 118799 GN = RpS4 PE = 2 SV = 1	3.47e-152	Translation
k37_Locus_45_Transcript_14_1	5685	26654 401	91424.33	9.14		U1-PSDTX-Ta1a/U4-PSDTX-Ta1a/U5-PSDTX-Ta1a Genbank Accession Numbers : MN607166/MN607167/MN607171		Venom peptide

k37_Locus_472_Transcript_5_3	2292	19416 1	665.97	0.07	k37_Locus_472_Transcript_5_3_6	sp Q08169 HUGA_APIME Hyaluronidase OS = Apis mellifera OX = 7460 PE = 1 SV = 1	1.46e-55	Metabolism
k37_Locus_472_Transcript_5_4	2275	26884 2	922.13	0.09	k37_Locus_472_Transcript_5_4_6	sp Q08169 HUGA_APIME Hyaluronidase OS = Apis mellifera OX = 7460 PE = 1 SV = 1	5.72e-59	Metabolism
k37_Locus_49_Transcript_3_1	3765	86335 7	2961.31	0.3	k37_Locus_49_Transcript_3_1_5	sp Q02748 IF4A_DROME Eukaryotic initiation factor 4A OS = Drosophila melanogaster OX = 7227 GN = eIF4A PE = 1 SV = 3	1.23e-138	Translation
k37_Locus_49_Transcript_3_2	512	32744 5	1123.13	0.11				
k37_Locus_540_Transcript_2_1	2408	21402 8	734.11	0.07	k37_Locus_540_Transcript_2_1_4	sp Q9VBV3 TAKT_DROME Protein takeout OS = Drosophila melanogaster OX = 7227 GN = to PE = 2 SV = 1	1.26e-08	Circadian Rythm
k37_Locus_544_Transcript_9_1	2629	18926 3	649.17	0.06				
k37_Locus_594_Transcript_7_2	3499	16843 3	577.72	0.06	k37_Locus_594_Transcript_7_2_5	sp Q9GPH3 ATFC_BOMMO Activating transcription factor of chaperone OS = Bombyx mori OX = 7091 GN = ATFC PE = 2 SV = 1	5.72e-26	Protein Maturation
k37_Locus_6671_Transcript_1_1	2238	40955 2	1404.76	0.14	k37_Locus_6671_Transcript_1_1_1	sp P36241 RL19_DROME 60S ribosomal protein L19 OS = Drosophila melanogaster OX = 7227 GN = RpL19 PE = 1 SV = 2	4.25e-89	Translation
k37_Locus_795_Transcript_8_1	2389	48076 9	1649.03	0.16	k37_Locus_795_Transcript_8_1_2	sp O16797 RL3_DROME 60S ribosomal protein L3 OS = Drosophila melanogaster OX = 7227 GN = RpL3 PE = 1 SV = 3	0.0	Translation
k37_Locus_87_Transcript_5_1	2675	27992 6	960.14	0.1	k37_Locus_87_Transcript_5_1_4	sp P21187 PABP_DROME Polyadenylate-binding protein OS = Drosophila melanogaster OX = 7227 GN = pAbp PE = 1 SV = 3	0.0	RNA Maturation
k37_Locus_921_Transcript_7_2	1145	29579 8	1014.58	0.1	k37_Locus_921_Transcript_7_2_4	sp C0HKA1 RS14B_DROME 40S ribosomal protein S14b OS = Drosophila melanogaster OX = 7227 GN = RpS14b PE = 2 SV = 1	7.47e-85	Translation
k37_Locus_9365_Transcript_5_1	3029	19822 8	679.92	0.07	k37_Locus_9365_Transcript_5_1_3	sp Q5ZIR1 SHLB1_CHICK Endophilin-B1 OS = Gallus gallus OX = 9031 GN = SH3GLB1 PE = 2 SV = 1	3.22e-64	Secretion / Translation

					k37_Locus_9365_Transcript_5_1_6	sp Q962R1 RS18_SPOFR 40S ribosomal protein S18 OS = Spodoptera frugiperda OX = 7108 GN = RpS18 PE = 2 SV = 1	9.35e-85	
k43_Locus_114_Transcript_2_1	2476	214200	734.7	0.07	k43_Locus_114_Transcript_2_1_6	sp Q963B7 RL9_SPOFR 60S ribosomal protein L9 OS = Spodoptera frugiperda OX = 7108 GN = RpL9 PE = 2 SV = 1	3.81e-91	Translation
k43_Locus_133_Transcript_7_1	2837	800213	2744.72	0.27	k43_Locus_133_Transcript_7_1_5	sp Q3ZU95 PA1_VESGE Phospholipase A1 OS = Vespula germanica OX = 30212 PE = 2 SV = 1	4.12e-19	Phospholipase
					k43_Locus_133_Transcript_7_1_5	sp Q68KK0 PA1_SOLIN Phospholipase A1 OS = Solenopsis invicta OX = 13686 PE = 1 SV = 1	4.11e-15	
k43_Locus_1473_Transcript_5_1	3652	407837	1398.88	0.14	k43_Locus_1473_Transcript_5_1_3	sp P14318 MP20_DROME Muscle-specific protein 20 OS = Drosophila melanogaster OX = 7227 GN = Mp20 PE = 2 SV = 2	9.42e-96	Cytoskeleton
k43_Locus_1496_Transcript_10_1	2917	196144	672.77	0.07				
k43_Locus_1586_7_Transcript_1_1	600	449378	1541.36	0.15		U2-PSDTX-Ta1a Genbank Accession Number : MN607169		Venom peptide
k43_Locus_209_Transcript_2_5	3324	118224	405.51	0.04	k43_Locus_209_Transcript_2_5_1	sp O08623 SQSTM_RAT Sequestosome-1 OS = Rattus norvegicus OX = 10116 GN = Sqstm1 PE = 1 SV = 1	9.72e-30	Protein Degradation
k43_Locus_3217_Transcript_3_3	1201	282147	967.76	0.1	k43_Locus_3217_Transcript_3_3_4	sp P30736 RL15_CHITE 60S ribosomal protein L15 OS = Chironomus tentans OX = 7153 GN = RpL15 PE = 3 SV = 3	1.14e-109	Translation
k43_Locus_3840_Transcript_4_2	2659	166658	571.64	0.06	k43_Locus_3840_Transcript_4_2_5	sp Q9NV56 MRGBP_HUMAN MRG/MORF4L-binding protein OS = Homo sapiens OX = 9606 GN = MRGBP PE = 1 SV = 1	7.48e-24	Transcription
k43_Locus_530_Transcript_3_1	1144	168563	578.17	0.06	k43_Locus_530_Transcript_3_1_4	sp Q962R9 RS10_SPOFR 40S ribosomal protein S10 OS = Spodoptera frugiperda OX = 7108 GN = RpS10 PE = 2 SV = 1	4.44e-70	Translation
k43_Locus_588_Transcript_3_1	1555	5439505	18657.45	1.87	k43_Locus_588_Transcript_3_1_3	sp P29183 LIPP_HORSE Pancreatic triacylglycerol lipase (Fragment) OS = Equus caballus OX = 9796 GN = PNLIP PE = 1 SV = 2	8.49e-31	Phospholipase

k43_Locus_684_Transcript_1_1	1148	12800 5	439.06	0.04					
k43_Locus_713_Transcript_5_1	4230	11654 5	399.75	0.04	k43_Locus_713_Transcript_5_1_1	sp Q14186 TFDP1_HUMAN OS = Homo sapiens OX = 9606 GN = TFDP1 PE = 1 SV = 1	1.18e-99	Transcription	
k43_Locus_7363_Transcript_5_1	3050	26140 1	896.6	0.09	k43_Locus_7363_Transcript_5_1_2	sp Q4GXU6 RS4_CARGR 40S ribosomal protein S4 OS = Carabus granulatus OX = 118799 GN = RpS4 PE = 2 SV = 1	2.03e-166	Translation	
					k43_Locus_7363_Transcript_5_1_5	sp A4Q9E4 TTLL2_MOUSE Probable tubulin polyglutamylase TTLL2 OS = Mus musculus OX = 10090 GN = Ttll2 PE = 2 SV = 1	9.34e-79		
k43_Locus_82_Transcript_8_5	2674	76003 3	2606.91	0.26	k43_Locus_82_Transcript_8_5_4	sp Q9XZ43 5NTD_LUTLO Protein 5NUC OS = Lutzomyia longipalpis OX = 7200 GN = 5NUC PE = 1 SV = 1	9.94e-78	Metabolism	
					k43_Locus_82_Transcript_8_5_5	sp B6EWW8 V5NTD_GLOBR Snake venom 5'-nucleotidase OS = Gloydus brevicaudus OX = 259325 PE = 2 SV = 1	1.27e-39		
k49_Locus_10848_Transcript_1_1	915	15077 7	517.16	0.05	k49_Locus_10848_Transcript_1_1_4	sp Q02878 RL6_HUMAN 60S ribosomal protein L6 OS = Homo sapiens OX = 9606 GN = RPL6 PE = 1 SV = 3	3.50e-44	Translation	
k49_Locus_11970_Transcript_2_1	2958	10816 3	371	0.04	k49_Locus_11970_Transcript_2_1_2	sp P84185 ACT5C_ANOGA Actin-5C OS = Anopheles gambiae OX = 7165 GN = Act5C PE = 2 SV = 1	0.0	Cytoskeleton	
k49_Locus_1327_Transcript_9_1	7028	31908 7	1094.47	0.11					
k49_Locus_1493_Transcript_10_1	3416	10165 4	348.67	0.03	k49_Locus_1493_Transcript_10_1_2	sp Q2T9X3 SPRY7_BOVIN SPRY domain-containing protein 7 OS = Bos taurus OX = 9913 GN = SPRYD7 PE = 2 SV = 1	1.32e-55	Protein Degradation	
k49_Locus_15244_Transcript_1_1	350	10428 1	357.68	0.04					
k49_Locus_1637_Transcript_10_1	5305	79982 9	2743.41	0.27	k49_Locus_1637_Transcript_10_1_6	sp Q24238 APH4_DROME Alkaline phosphatase 4 OS = Drosophila melanogaster OX = 7227 GN = Alp4 PE = 2 SV = 3	3.33e-121	Metabolism	

					k49_Locus_2359_	sp Q9DA39 LFG4_MOUSE Protein lifeguard 4 OS = Mus	4.06e-	
					Transcript_3_1_3	musculus OX = 10090 GN = Tmbim4 PE = 2 SV = 1	56	
k49_Locus_2359_	4026	11662	400.03	0.04	k49_Locus_2359_	sp Q6PFM4 LNPB_DANRE Endoplasmic reticulum	1.41e-	Anti-apoptosis/
_Transcript_3_1		6			Transcript_3_1_5	junction formation protein lunapark-B OS = Danio rerio	64	RE
						OX = 7955 GN = lnpcb PE = 2 SV = 2		morphology/
					k49_Locus_2359_	sp Q95WA0 RL26_LITLI 60S ribosomal protein L26 OS =	5.49e-	Translation
					Transcript_3_1_6	Littorina littorea OX = 31216 GN = RPL26 PE = 2 SV = 1	68	
k49_Locus_281_	2899	16931	5807.54	0.58	k49_Locus_281_T	sp Q02942 TRF_BLADI Transferrin OS = Blaberus	0.0	Metabolism
Transcript_3_1		64			ranscript_3_1_4	discoidealis OX = 6981 PE = 1 SV = 1		
k49_Locus_322_	1247	40416	1386.27	0.14				
Transcript_9_3		1						
k49_Locus_322_	2573	48771	1672.87	0.17	k49_Locus_322_T	sp P17892 LIPR2_MOUSE Pancreatic lipase-related	1.08e-	Phospholipase
Transcript_9_4		8			ranscript_9_4_5	protein 2 OS = Mus musculus OX = 10090 GN = Pnliprp2	86	
						PE = 1 SV = 2		
k49_Locus_397_	2724	97591	3347.39	0.33	k49_Locus_397_T	sp P54399 PDI_DROME Protein disulfide-isomerase OS	0.0	Protein
Transcript_1_1		9			ranscript_1_1_5	= Drosophila melanogaster OX = 7227 GN = Pdi PE = 2 SV		Maturation
						= 1		
k49_Locus_4240_	2005	22011	754.99	0.08	k49_Locus_4240_	sp Q963B6 RL10A_SPOFR 60S ribosomal protein L10a	2.71e-	Translation
Transcript_9_2		5			Transcript_9_2_2	OS = Spodoptera frugiperda OX = 7108 GN = RpL10A PE	119	
						= 2 SV = 1		
k49_Locus_450_	2738	10705	367.18	0.04				
Transcript_3_1		1						
k49_Locus_47_	1591	28228	968.24	0.1				
Transcript_1_1		7						
k49_Locus_5739_	7855	14951	512.82	0.05	k49_Locus_5739_	sp Q7M4F3 CUD2_SCHGR Endocuticle structural	4.13e-	Cuticule
Transcript_4_1		0			Transcript_4_1_1	glycoprotein SgAbd-2 OS = Schistocerca gregaria OX =	21	
						7010 PE = 1 SV = 1		
k49_Locus_8888_	4025	11590	397.55	0.04	k49_Locus_8888_	sp P14318 MP20_DROME Muscle-specific protein 20 OS	8.21e-	Cytoskeleton
Transcript_2_1		5			Transcript_2_1_6	= Drosophila melanogaster OX = 7227 GN = Mp20 PE = 2	65	
						SV = 2		
k49_Locus_9979_	496	17159	588.56	0.06				
Transcript_1_1		2						

k55_Locus_113_Transcript_10_1	3158	1321901	4534.11	0.45	k55_Locus_113_Transcript_10_1_4	sp Q3T126 CNIH4_BOVIN Protein cornichon homolog 4 OS = Bos taurus OX = 9913 GN = CNIH4 PE = 2 SV = 1	1.60e-39	Secretion / Phospholipase
					k55_Locus_113_Transcript_10_1_6	sp Q68KK0 PA1_SOLIN Phospholipase A1 OS = Solenopsis invicta OX = 13686 PE = 1 SV = 1	1.43e-31	
k55_Locus_12090_Transcript_2_1	2579	103030	353.39	0.04				
k55_Locus_12339_Transcript_1_1	1043	116259	398.77	0.04	k55_Locus_12339_Transcript_1_1_5	sp P41822 FRI_AEDAE Ferritin subunit OS = Aedes aegypti OX = 7159 GN = FERH PE = 1 SV = 2	1.34e-40	Metabolism
k55_Locus_1250_Transcript_14_3	1123	356892	1224.14	0.12	k55_Locus_1250_Transcript_14_3_1	sp Q5G5C4 RS3A_PERAM 40S ribosomal protein S3a OS = Periplaneta americana OX = 6978 GN = Parcxpwex01 PE = 2 SV = 1	2.43e-139	Translation
k55_Locus_1331_Transcript_6_1	4877	235948	809.3	0.08				
k55_Locus_1565_Transcript_7_1	5153	687748	2358.97	0.24	k55_Locus_1565_Transcript_7_1_6	sp Q24238 APH4_DROME Alkaline phosphatase 4 OS = Drosophila melanogaster OX = 7227 GN = Alp4 PE = 2 SV = 3	6.44e-76	Metabolism
k55_Locus_185_Transcript_11_1	1096	462590	1586.68	0.16	k55_Locus_185_Transcript_11_1_5	sp P35778 VA3_SOLIN Venom allergen 3 OS = Solenopsis invicta OX = 13686 PE = 1 SV = 2	1.15e-71	Venom allergen
k55_Locus_2271_Transcript_5_2	1733	273202	937.08	0.09	k55_Locus_2271_Transcript_5_2_2	sp Q9DA39 LFG4_MOUSE Protein lifeguard 4 OS = Mus musculus OX = 10090 GN = Tmbim4 PE = 2 SV = 1	4.66e-50	Translation
					k55_Locus_2271_Transcript_5_2_5	sp Q95WA0 RL26_LITLI 60S ribosomal protein L26 OS = Littorina littorea OX = 31216 GN = RPL26 PE = 2 SV = 1	5.49e-68	
k55_Locus_2505_Transcript_11_2	3085	221164	758.59	0.08	k55_Locus_2505_Transcript_11_2_6	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS = Apis mellifera OX = 7460 PE = 1 SV = 1	2.46e-110	Peptide Maturation
k55_Locus_2784_Transcript_7_1	3736	186655	640.22	0.06				
k55_Locus_2784_Transcript_7_2	1355	386247	1324.82	0.13				
k55_Locus_2824_Transcript_2_1	1210	317588	1089.32	0.11	k55_Locus_2824_Transcript_2_1_1	sp Q5UAP4 RSSA_BOMMO 40S ribosomal protein SA OS = Bombyx mori OX = 7091 PE = 2 SV = 1	1.39e-124	Translation

					k55_Locus_38_Transcript_120_1_1	sp Q17750 UFL1_CAEEL E3 UFM1-protein ligase 1 homolog OS = Caenorhabditis elegans OX = 6239 GN = ufl-1 PE = 3 SV = 1	0.17	
					k55_Locus_38_Transcript_120_1_3	sp Q0JKD0 GLT1_ORYSJ Glutamate synthase 1 [NADH]. chloroplastic OS = Oryza sativa subsp. japonica OX = 39947 GN = Os01g0681900 PE = 2 SV = 1	0.45	
k55_Locus_38_Transcript_120_1	10401	4377671	15015.37	1.5	k55_Locus_38_Transcript_120_1_4	sp Q09575 YRD6_CAEEL Uncharacterized protein K02A2.6 OS = Caenorhabditis elegans OX = 6239 GN = K02A2.6 PE = 4 SV = 1	2.61e-27	RNA Maturation
					k55_Locus_38_Transcript_120_1_6	sp Q09575 YRD6_CAEEL Uncharacterized protein K02A2.6 OS = Caenorhabditis elegans OX = 6239 GN = K02A2.6 PE = 4 SV = 1	2.51e-18	
					k55_Locus_38_Transcript_120_1_6	sp Q5F3X4 U5S1_CHICK 116 kDa U5 small nuclear ribonucleoprotein component OS = Gallus gallus OX = 9031 GN = EFTUD2 PE = 2 SV = 1	0.0	
k55_Locus_3981_Transcript_1_1	1058	303870	1042.27	0.1				
k55_Locus_44_Transcript_1_1	1186	336982	1155.84	0.12	k55_Locus_44_Transcript_1_1_1	sp Q75VN3 TCTP_BOMMO Translationally-controlled tumor protein homolog OS = Bombyx mori OX = 7091 GN = Tctp PE = 2 SV = 1	1.06e-85	Cytoskeleton
k55_Locus_485_Transcript_9_1	2206	308425	10578.86	1.06	k55_Locus_485_Transcript_9_1_2	sp Q6XZB0 LIPI_HUMAN Lipase member I OS = Homo sapiens OX = 9606 GN = LIPI PE = 1 SV = 3	5.93e-25	Phospholipase
k55_Locus_485_Transcript_9_2	1849	679023	2329.04	0.23				Chemoreception
k55_Locus_52_Transcript_7_1	2781	101603	348.5	0.03				
k55_Locus_552_Transcript_9_1	3315	267983	919.18	0.09	k55_Locus_552_Transcript_9_1_4	sp P0CG71 UBIQ1_CAEEL Polyubiquitin-A OS = Caenorhabditis elegans OX = 6239 GN = ubq-1 PE = 3 SV = 1	0.0	Protein Degradation
k55_Locus_5854_Transcript_3_1	7511	251601	862.99	0.09	k55_Locus_5854_Transcript_3_1_3	sp Q7M4F3 CUD2_SCHGR Endocuticle structural glycoprotein SgAbd-2 OS = Schistocerca gregaria OX = 7010 PE = 1 SV = 1	4.13e-21	Cuticule

k55_Locus_61_Transcript_6_1	2678	47812 3	1639.96	0.16	k55_Locus_61_Transcript_6_1_6	sp P29341 PABP1_MOUSE Polyadenylate-binding protein 1 OS = Mus musculus OX = 10090 GN = Pabpc1 PE = 1 SV = 2	0.0	RNA Maturation
k55_Locus_6635_Transcript_3_1	785	19174 1	657.67	0.07				
k55_Locus_6656_Transcript_3_2	845	11357 8	389.57	0.04	k55_Locus_6656_Transcript_3_2_4	sp P18101 RL40_DROME Ubiquitin-60S ribosomal protein L40 OS = Drosophila melanogaster OX = 7227 GN = RpL40 PE = 1 SV = 2	2.40e-81	Protein Degradation
k55_Locus_7083_Transcript_1_1	2980	10235 6	351.08	0.04	k55_Locus_7083_Transcript_1_1_5	sp Q9VFC2 SP88E_DROME Serine protease inhibitor 88Ea OS = Drosophila melanogaster OX = 7227 GN = Spn88Ea PE = 2 SV = 1	2.55e-61	Immunity
k55_Locus_7703_Transcript_1_1	2688	10548 62	3618.17	0.36	k55_Locus_7703_Transcript_1_1_1	sp A4IHT0 FIGL1_XENTR Fidgetin-like protein 1 OS = Xenopus tropicalis OX = 8364 GN = fign1 PE = 2 SV = 1	3.88e-134	DNA Repair
k55_Locus_8_Transcript_7_1	4349	15942 157	54681.44	5.47	k55_Locus_8_Transcript_7_1_5	sp Q3ZU95 PA1_VESGE Phospholipase A1 OS = Vespula germanica OX = 30212 PE = 2 SV = 1	2.56e-45	Phospholipase
					k55_Locus_8_Transcript_7_1_6	sp Q3ZU95 PA1_VESGE Phospholipase A1 OS = Vespula germanica OX = 30212 PE = 2 SV = 1	4.18e-45	
k55_Locus_9803_Transcript_1_1	4289	10663 2	365.75	0.04	k55_Locus_9803_Transcript_1_1_6	sp Q9UBV2 SE1L1_HUMAN Protein sel-1 homolog 1 OS = Homo sapiens OX = 9606 GN = SEL1L PE = 1 SV = 3	0.0	Protein Maturation
k55_Locus_996_Transcript_10_1	3799	41606 6	1427.1	0.14				
k61_Locus_10818_Transcript_1_1	1302	29816 0	1022.69	0.1				
k61_Locus_11296_Transcript_3_1_1	1613	51777 9	1775.98	0.18	k61_Locus_11296_Transcript_3_1_5	sp Q94624 RS6_MANSE 40S ribosomal protein S6 OS = Manduca sexta OX = 7130 GN = RpS6 PE = 2 SV = 1	2.98e-125	Translation
k61_Locus_175_Transcript_2_1	2594	35783 8	1227.38	0.12	k61_Locus_175_Transcript_2_1_2	sp P09180 RL4_DROME 60S ribosomal protein L4 OS = Drosophila melanogaster OX = 7227 GN = RpL4 PE = 1 SV = 2	3.20e-116	Translation
k61_Locus_1787_Transcript_1_1	1871	13446 6	461.22	0.05	k61_Locus_1787_Transcript_1_1_5	sp Q58FK9 KAT3_RAT Kynurenine--oxoglutarate transaminase 3 OS = Rattus norvegicus OX = 10116 GN = Kyat3 PE = 2 SV = 1	5.73e-91	Protein Maturation

k61_Locus_1873_Transcript_9_2	1132	35973 8	1233.9	0.12	k61_Locus_1873_Transcript_9_2_6	sp Q5G5C4 RS3A_PERAM 40S ribosomal protein S3a OS = Periplaneta americana OX = 6978 GN = Parcxpwex01 PE = 2 SV = 1	2.43e-139	Translation
k61_Locus_2007_Transcript_1_2	904	50234 5	1723.04	0.17	k61_Locus_2007_Transcript_1_2_6	sp Q8WQI7 RL18A_SPOFR 60S ribosomal protein L18a OS = Spodoptera frugiperda OX = 7108 GN = RpL18A PE = 2 SV = 1	6.38e-81	Translation
k61_Locus_2081_Transcript_6_1	3183	11992 3	411.33	0.04	k61_Locus_2081_Transcript_6_1_1	sp P91887 AMPN_PLUXY Aminopeptidase N OS = Plutella xylostella OX = 51655 GN = APN1 PE = 1 SV = 1	2.01e-80	Peptide Maturation
k61_Locus_2381_Transcript_1_1	3338	24358 0	835.48	0.08	k61_Locus_2381_Transcript_1_1_1	sp P12919 PDGFB_FELCA Platelet-derived growth factor subunit B OS = Felis catus OX = 9685 GN = PDGFB PE = 2 SV = 1	1.57e-05	Unknown
k61_Locus_250_Transcript_6_2	2112	20662 0	708.7	0.07	k61_Locus_250_Transcript_6_2_6	sp P29413 CALR_DROME Calreticulin OS = Drosophila melanogaster OX = 7227 GN = Calr PE = 1 SV = 2	0.0	Protein Maturation
k61_Locus_2935_Transcript_2_3	2338	33072 2	1134.37	0.11	k61_Locus_2935_Transcript_2_3_2	sp P06603 TBA1_DROME Tubulin alpha-1 chain OS = Drosophila melanogaster OX = 7227 GN = alphaTub84B PE = 1 SV = 1	0.0	Cytoskeleton
k61_Locus_3810_Transcript_2_1	4984	30465 9	1044.98	0.1	k61_Locus_3810_Transcript_2_1_6	sp Q2TBQ5 RL7A_BOVIN 60S ribosomal protein L7a OS = Bos taurus OX = 9913 GN = RPL7A PE = 2 SV = 3	7.48e-121	Translation
k61_Locus_40_Transcript_1_1	953	53161 2	1823.42	0.18				Cytoskeleton
k61_Locus_438_Transcript_8_1	3543	46369 7	1590.48	0.16	k61_Locus_438_Transcript_8_1_1	sp P0CG71 UBIQ1_CAEEL Polyubiquitin-A OS = Caenorhabditis elegans OX = 6239 GN = ubq-1 PE = 3 SV = 1	0.0	Protein Degradation
k61_Locus_4981_Transcript_2_1	1829	18384 0	630.57	0.06	k61_Locus_4981_Transcript_2_1_1	sp P14318 MP20_DROME Muscle-specific protein 20 OS = Drosophila melanogaster OX = 7227 GN = Mp20 PE = 2 SV = 2	8.73e-89	Cytoskeleton
k61_Locus_500_Transcript_6_4	2349	23908 7	820.07	0.08	k61_Locus_500_Transcript_6_4_6	sp P35502 ESTF_MYZPE Esterase FE4 OS = Myzus persicae OX = 13164 PE = 1 SV = 1	2.35e-81	Unknown
k61_Locus_582_Transcript_6_1	3659	18658 8	639.99	0.06	k61_Locus_582_Transcript_6_1_1	sp P12261 EF1G_ARTSA Elongation factor 1-gamma OS = Artemia salina OX = 85549 PE = 1 SV = 3	9.62e-179	Translation
k65_Locus_104_Transcript_8_1	3715	57846 0	1984.11	0.2	k65_Locus_104_Transcript_8_1_4	sp Q3T126 CNIH4_BOVIN Protein cornichon homolog 4 OS = Bos taurus OX = 9913 GN = CNIH4 PE = 2 SV = 1	1.60e-39	Secretion / Phospholipase

					k65_Locus_104_T ranscript_8_1_6	sp Q3ZU95 PA1_VESGE Phospholipase A1 OS = Vespula germanica OX = 30212 PE = 2 SV = 1	1.82e- 30	
k65_Locus_1147 _Transcript_2_1	3286	15671 70	5375.38	0.54	k65_Locus_1147_ Transcript_2_1_5	sp P13060 EF2_DROME Elongation factor 2 OS = Drosophila melanogaster OX = 7227 GN = EF2 PE = 1 SV = 4	0.0	Translation
k65_Locus_2088 _Transcript_11_ 1	2921	16647 2	571	0.06	k65_Locus_2088_ Transcript_11_1_ 2	sp Q9V447 KRH2_DROME Krueppel homolog 2 OS = Drosophila melanogaster OX = 7227 GN = Kr-h2 PE = 1 SV = 1	6.94e- 59	Protein Maturation / Translation
					k65_Locus_2088_ Transcript_11_1_ 5	sp Q5R465 RS3_PONAB 40S ribosomal protein S3 OS = Pongo abelii OX = 9601 GN = RPS3 PE = 2 SV = 1	7.22e- 52	
k65_Locus_26_ Transcript_2_1	1745	21805 83	7479.38	0.75		MTSDERRSYLPLLLAVIFVLAIVHTPSVESRASADAEAD AFADALAKAIANADPGILGVIARWIWKLILQILAPATAA VEVATRLGLPQK		Venom peptide
k65_Locus_4162 _Transcript_4_1	1925	20589 9	706.23	0.07				
k65_Locus_5275 _Transcript_1_1	1660	20671 1	709.02	0.07	k65_Locus_5275_ Transcript_1_1_3	sp P08570 RLA1_DROME 60S acidic ribosomal protein P1 OS = Drosophila melanogaster OX = 7227 GN = RpLP1 PE = 1 SV = 2	3.78e- 36	Translation
k65_Locus_596_ Transcript_1_1	742	19565 4	671.09	0.07	k65_Locus_596_T ranscript_1_1_5	sp O76756 RS8_APIME 40S ribosomal protein S8 OS = Apis mellifera OX = 7460 GN = RpS8 PE = 2 SV = 2	2.23e- 136	Translation
k65_Locus_7655 _Transcript_3_1	2345	33391 2	1145.31	0.11	k65_Locus_7655_ Transcript_3_1_4	sp P32100 RL7_DROME 60S ribosomal protein L7 OS = Drosophila melanogaster OX = 7227 GN = RpL7 PE = 1 SV = 2	1.90e- 103	Translation
k65_Locus_83_ Transcript_7_1	4005	11365 2	389.83	0.04				
k65_Locus_8694 _Transcript_4_1	2406	20853 97	7152.89	0.72	k65_Locus_8694_ Transcript_4_1_1	sp P35778 VA3_SOLIN Venom allergen 3 OS = Solenopsis invicta OX = 13686 PE = 1 SV = 2	1.17e- 47	Venom allergen
					k65_Locus_8694_ Transcript_4_1_1	sp P35778 VA3_SOLIN Venom allergen 3 OS = Solenopsis invicta OX = 13686 PE = 1 SV = 2	9.87e- 51	
k65_Locus_9888 _Transcript_6_2	1228	16826 7	577.15	0.06	k65_Locus_9888_ Transcript_6_2_6	sp Q6EV04 RS3A_BIPLU 40S ribosomal protein S3a OS = Biphyllus lunatus OX = 197003 PE = 2 SV = 1	1.87e- 142	Translation

k69_Locus_1020 2_Transcript_2_2	1480	13427 9	460.58	0.05					
k69_Locus_1049 _Transcript_1_1	548	19660 8	674.36	0.07					
k69_Locus_110_ Transcript_5_1	3341	19212 34	6589.81	0.66	k69_Locus_110_T ranscript_5_1_6	sp Q3ZU95 PA1_VESGE Phospholipase A1 OS = Vespula germanica OX = 30212 PE = 2 SV = 1	8.05e- 45	Phospholipase	
k69_Locus_1158 0_Transcript_1_1	727	19845 2	680.69	0.07	k69_Locus_11580 _Transcript_1_1_6	sp Q75VN3 TCTP_BOMMO Translationally-controlled tumor protein homolog OS = Bombyx mori OX = 7091 GN = Tctp PE = 2 SV = 1	1.20e- 45	Cytoskeleton	
k69_Locus_1227 2_Transcript_1_1	720	24498 3	840.29	0.08	k69_Locus_12272 _Transcript_1_1_4	sp O96647 RL10_BOMMA 60S ribosomal protein L10 OS = Bombyx mandarina OX = 7092 GN = RpL10 PE = 2 SV = 1	3.53e- 112	Translation	
k69_Locus_1235 5_Transcript_5_1	2460	29948 67	10272.36	1.03	k69_Locus_12355 _Transcript_5_1_6	sp P35778 VA3_SOLIN Venom allergen 3 OS = Solenopsis invicta OX = 13686 PE = 1 SV = 2	1.17e- 47	Venom allergen	
					k69_Locus_12355 _Transcript_5_1_6	sp P35778 VA3_SOLIN Venom allergen 3 OS = Solenopsis invicta OX = 13686 PE = 1 SV = 2	1.01e- 71		
k69_Locus_1459 3_Transcript_1_1	598	28645 7	982.54	0.1	k69_Locus_14593 _Transcript_1_1_1	sp P05389 RLA2_DROME 60S acidic ribosomal protein P2 OS = Drosophila melanogaster OX = 7227 GN = RpLP2 PE = 1 SV = 1	2.91e- 24	Translation	
k69_Locus_1496 4_Transcript_1_1	1925	13033 5	447.05	0.04	k69_Locus_14964 _Transcript_1_1_4	sp P35501 ESTE_MYZPE Esterase E4 OS = Myzus persicae OX = 13164 PE = 1 SV = 1	4.99e- 82	Unknown	
k69_Locus_1683 4_Transcript_1_1	739	65913 9	2260.84	0.23		U2-PSDTX-Ta1c Genbank Accession Numbers : MN607168		Venom peptide	
k69_Locus_2100 _Transcript_4_1	7317	13391 4	459.32	0.05	k69_Locus_2100_ Transcript_4_1_1	sp O95602 RPA1_HUMAN DNA-directed RNA polymerase I subunit RPA1 OS = Homo sapiens OX = 9606 GN = POLR1A PE = 1 SV = 2	0.0	Transcription	

k69_Locus_3195_Transcript_5_1	4866	119870	411.15	0.04	k69_Locus_3195_Transcript_5_1_5	sp Q8IZJ3 CPMD8_HUMAN C3 and PZP-like alpha-2-macroglobulin domain-containing protein 8 OS = Homo sapiens OX = 9606 GN = CPAMD8 PE = 1 SV = 2	1.14e-06	Protein Maturation
k69_Locus_4946_Transcript_1_1	2719	246026	843.87	0.08				
k69_Locus_547_Transcript_3_1	2747	23555899	80796.5	8.08				
k69_Locus_64_Transcript_10_4	3190	2483041	8516.81	0.85	k69_Locus_64_Transcript_10_4_6	sp Q9XZ43 5NTD_LUTLO Protein 5NUC OS = Lutzomyia longipalpis OX = 7200 GN = 5NUC PE = 1 SV = 1	4.98e-123	Metabolism
k69_Locus_678_Transcript_6_1	4344	165812	568.73	0.06	k69_Locus_678_Transcript_6_1_1	sp P07602 SAP_HUMAN Prosaposin OS = Homo sapiens OX = 9606 GN = PSAP PE = 1 SV = 2	6.11e-38	Cellular Signaling
k69_Locus_8284_Transcript_2_1	2255	704605	2416.79	0.24				
k69_Locus_87_Transcript_11_1	1120	845555	2900.25	0.29	k69_Locus_87_Transcript_11_1_2	sp Q9BMK4 PA2_APICC Phospholipase A2 OS = Apis cerana cerana OX = 94128 PE = 2 SV = 1	9.55e-26	Phospholipase
k69_Locus_9580_Transcript_1_1	3482	620884	2129.63	0.21	k69_Locus_9580_Transcript_1_1_1	sp B0FWD3 NU5M_AEDA NADH-ubiquinone oxidoreductase chain 5 OS = Aedes aegypti OX = 7159 GN = mt:ND5 PE = 3 SV = 1	1.83e-60	Metabolism
					k69_Locus_9580_Transcript_1_1_3	sp Q1HR20 NU4M_AEDA NADH-ubiquinone oxidoreductase chain 4 OS = Aedes aegypti OX = 7159 GN = mt:ND4 PE = 2 SV = 1	3.57e-19	

Supporting Information for Publication

Title of Paper

Characterization of the venom peptide repertoire of the red ant *Manica rubida* (Myrmicinae) and identification of insecticidal toxins

Authors

Axel Touchard^{†,*}, Samira R. Aili[‡], Nathan Téné[†], Valentine Barassé[†], Christophe Klopp[§], Alain Dejean^{||,⊥}, R. Manjunatha Kini^{*,**}, Mrinalini[‡] Laurent Coquet[∇], Thierry Jouenne[∇], Benjamin Lefranc[•], Jérôme Leprince[•], Pierre Escoubas[⊗], Graham M. Nicholson[‡], Michel Treilhou^{†,#}, Elsa Bonnafé^{†,#}

Affiliations

[†] Equipe BTSB-EA 7417, Université de Toulouse, Institut National Universitaire Jean-François Champollion, Place de Verdun, 81012, Albi, France

[‡] Neurotoxin Research Group, School of Life Sciences, University of Technology Sydney, Ultimo NSW 2007, Australia

[§] Unité de Mathématique et Informatique Appliquées de Toulouse, UR0875, INRA Toulouse, Castanet-Tolosan, France

^{||} CNRS, UMR Ecologie des forêts de Guyane (EcoFoG), Campus Agronomique, BP 316, 97379 Kourou Cedex, France

[⊥] CNRS, UMR 5245, Laboratoire Écologie Fonctionnelle et Environnement, 118 route de Narbonne, 31062 Toulouse, France

^{*} Protein Science Laboratory, Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore 117543, Singapore

^{**} Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117600, Singapore

[∇] CNRS UMR 6270, Normandie University, UNIROUEN, PISSARO, 76130 Mont-Saint-Aignan, France

[•] Inserm U 1239, Normandie University, UNIROUEN, Plate-forme de Recherche en Imagerie Cellulaire, Normandie (PRIMACEN), 76000 Rouen, France

[⊗] VenomeTech, 473 Route des Dolines — Villa 3, Valbonne 06560, France

Corresponding Authors

Dr. Axel Touchard, Equipe BTSB-EA 7417, Université de Toulouse, Institut National Universitaire Jean-François Champollion, Place de Verdun, 81012, Albi, France. Phone : +(33)5 63 48 64 32. Email: axel.touchard2@gmail.com

Table of Contents

Supplementary Table S1	Concentration of each synthetic myrmecitoxin used for insecticidal assays	Page S3
Supplementary Table S2	Peptide mass fingerprint of <i>Manica rubida</i> venom	Page S4
Supplementary Table S3	The list of sequence tags yielded from Peaks software interpretation of MS/MS spectra.	Page S5
Supplementary Figure S1	Integrative methodology used for the sequencing of U ₁₇ -MYRTX-Mri1b.	Page S13
Supplementary Table S4	Identified peptide toxin transcripts expressed in the venom transcriptome of <i>Manica rubida</i> .	Page S14
Supplementary Table S5	Insecticidal activity of <i>Manica rubida</i> venom fractions	Page S15
Supplementary Figure S2	<i>De novo</i> structure prediction of insecticidal myrmecitoxins of <i>M. rubida</i> venom	Page S16

Table S1. Concentration (mmol.L⁻¹) of each synthetic myrmecitoxin used for insecticidal assays. Each fly received 1 μ L of peptide-PBS solution.

U ₃ -MYRTX- Mri1a	U ₁₀ -MYRTX- Mri1a	U ₁₀ -MYRTX- Mri1b	U ₁₀ -MYRTX- Mri1c	U ₁₂ -MYRTX- Mri1a	U ₁₃ -MYRTX- Mri1a	U ₂₀ -MYRTX- Mri1a
1.024	2.350	2.703	2.238	3.817	2.813	2.813
0.512	1.175	1.351	0.448	0.763	1.407	2.250
0.256	0.825	0.832	0.203		0.566	1.918
0.102	0.650	0.277	0.112		0.113	0.384
0.026	0.275	0.062	0.058		0.057	0.192
0.010	0.130	0.006	0.020		0.028	0.096
	0.065		0.010		0.014	0.019
	0.032		0.004		0.006	
	0.016					
	0.006					

Table S2. Peptide mass fingerprint of *M. rubida* venom. List of peptide masses detected by LC-MS using an LCQ Advantage mass spectrometer. Bold text denotes full-length mature peptides. Masses corresponding to fragments were identified by Peaks software interpretation of MS/MS spectra. “*” Indicates isolated peptides submitted to Edman degradation. Note that U₁₇-MYRTX-Mri1c was not detected by LCQ Advantage mass spectrometer and was not included in this table.

Retention time (min)	Mass (Da)	Relative abundance (%)	Temporary name	Peptide identified in the venom
1.76	1092.71	0.053	Mr-1093	U₁₂-MYRTX-Mri1b
15.40	928.38	0.001	Mr-928	
18.16	6227.45	0.034	Mr-6227	U₁₈-MYRTX-Mri1a*
18.23	6425.95	0.006	Mr-6426	
18.62	6081.50	0.016	Mr-6082	nonglycosylated (U ₁₈ -MYRTX-Mri1a)
19.16	1314.50	0.018	Mr-1315	U₁₉-MYRTX-Mri1a
19.81	1434.91	0.348	Mr-1435	U₆-MYRTX-Mri1a*
19.86	1321.70	0.004	Mr-1322	Fragment (U ₆ -MYRTX-Mri1a)
19.86	1208.72	0.007	Mr-1209	
20.60	2745.96	0.007	Mr-2746	U₁₇-MYRTX-Mri1b*
20.64	1506.48	0.002	Mr-1506	Fragment (U ₁₀ -MYRTX-Mri1c)
21.21	2567.76	0.009	Mr-2568	U₁₇-MYRTX-Mri1a
22.73	4664.04	0.005	Mr-4664	
23.35	1379.70	0.099	Mr-1380	Fragment (U ₁₀ -MYRTX-Mri1a)
23.95	1110.80	1.975	Mr-1111	U₁₂-MYRTX-Mri1a*
24.26	1384.36	0.108	Mr-1384	Fragment (U ₃ -MYRTX-Mri1a)
24.90	1038.50	0.472	Mr-1039	Fragment (U ₃ -MYRTX-Mri1a)
24.90	1497.50	0.909	Mr-1498	Fragment (U ₃ -MYRTX-Mri1a)
27.32	1352.90	0.583	Mr-1353	Fragment (U ₁₀ -MYRTX-Mri1c)
27.37	2704.92	1.760	Mr-2705	Fragment (U ₁₃ -MYRTX-Mri1a)
29.37	1468.20	0.047	Mr-1468	
29.55	1390.86	1.514	Mr-1391	
31.53	2741.24	13.918	Mr-2741	U₂₀-MYRTX-Mri1a*
32.49	2852.32	24.303	Mr-2852	U₁₀-MYRTX-Mri1a
32.97	2979.48	1.906	Mr-2979	U₁₃-MYRTX-Mri1a
34.55	2989.02	7.784	Mr-2989	U₁₀-MYRTX-Mri1b
35.06	808.51	1.032	Mr-809	Fragment (U ₃ -MYRTX-Mri1a)
36.75	2730.50	0.475	Mr-2731	Fragment (U ₁₀ -MYRTX-Mri1c)
37.14	2413.14	0.178	Mr-2413	
37.68	2842.56	35.613	Mr-2843	U₁₀-MYRTX-Mri1c
43.75	2175.22	5.921	Mr-2175	U₃-MYRTX-Mri1a
44.82	2957.88	0.549	Mr-2958	
48.79	2287.98	0.343	Mr-2288	

Table S3. Sequence tags yielded from Peaks software interpretation of MS/MS spectra. The match with identified toxins was performed manually by comparison of each sequence tags with amino acid sequences of myrmecitoxins.

Scan	TLC	ALC (%)	m/z	z	RT	Mass	ppm	PTM	Tag	Toxin match
11207	10.8	98	589.8334	2	46.04	1177.6528	-0.4	N	LDKAAAFMKL	U ₁₃ -Mri1a
21770	13.7	98	713.9429	2	68.25	1425.8706	0.4	N	VGGSLAKAALKLW	U ₁₀ -Mri1c
11437	15.7	98	581.0359	3	46.55	1740.0872	-0.6	N	ASKAEELLKLLAKKA	U ₂₀ -Mri1a
19756	16.6	98	610.0187	3	63	1827.0327	0.9	N	GFKSMLAKAALKLLSYS	U ₁₀ -Mri1a
24468	20.5	98	700.4302	3	73.62	2098.2698	-0.4	N	GFKSMLAKAALKLLAKVAPAA	U ₁₀ -Mri1a
8672	9.8	98	443.7717	2	40.2	885.5283	0.6	N	GVGSLAKAA	U ₁₀ -Mri1c
15347	10.8	98	500.3138	2	54.73	998.6124	0.6	N	VGGSLAKAAL	U ₁₀ -Mri1c
21727	11.6	97	573.8475	2	68.19	1145.6809	-0.4	N	VGGSLAKAALF	U ₁₀ -Mri1c
13522	12.6	97	645.8925	2	51.33	1289.7708	-0.2	N	VGGSLAKAALKY	U ₁₀ -Mri1c
21818	13.6	97	686.4231	2	68.32	1370.8318	-0.1	N	VGGSLAKAALKML	U ₁₀ -Mri1c
31689	13.5	97	698.4496	2	87.69	1394.886	-0.9	N	PTGSLAKAALKLL	U ₁₀ -Mri1c
42191	13.6	97	732.4453	2	107	1462.8872	-7.6	N	EHGSLAKAALKLL	U ₁₀ -Mri1c
35845	13.5	97	744.9749	2	95.24	1487.9438	-5.8	N	KYGSLAKAALKLL	U ₁₀ -Mri1c
17777	13.6	97	762.9513	2	58.84	1523.8896	-1	N	GFKSMLAKAALKLF	U ₁₀ -Mri1a
7253	9.6	96	549.7823	2	36.23	1097.5505	-0.4	N	PFLQHALTDG	U ₃ -Mri1a
17557	9.6	96	555.8234	2	58.48	1109.6331	-0.7	N	PLDPKLVESL	U ₁₂ -Mri1a
12051	11.6	96	626.8611	2	48.38	1251.7075	0.1	N	VVTGDPLDPKVL	U ₁₂ -Mri1a
31493	13.5	96	691.4429	2	87.33	1380.8704	0.6	N	SPGSLAKAALKLL	U ₁₀ -Mri1c
16250	14.4	96	718.927	2	56.29	1435.8398	-0.3	N	LVAPAAAEVLANKLG	U ₁₀ -Mri1c
19754	12.5	96	722.3672	2	63	1442.7227	-1.9	N	MTLPFLQHALTDG	U ₃ -Mri1a
26438	15.3	96	522.6581	3	77.08	1564.9551	-1.6	N	DPGVGSLAKAALKLL	U ₁₀ -Mri1c
27267	17.2	96	660.7043	3	78.82	1979.0913	0	N	KLGLFDQLDKAAAFMLK	U ₁₃ -Mri1a
27301	17.4	96	990.5566	2	78.89	1979.0913	3.8	N	KLGLFDQLDKAAAFMKL	U ₁₃ -Mri1a
35650	21.2	96	717.7927	3	94.88	2150.3552	0.4	N	VGGSLAKAALKLLKLVAPAAF	U ₁₀ -Mri1c
19739	19.3	96	553.3193	4	62.98	2209.2478	0.1	N	GFKSMLAKAALKLVSMRPVY	U ₁₀ -Mri1a
27900	19.3	96	743.7672	3	79.96	2228.2812	-0.7	N	LGMESLKQLSAKAEELLKLL	U ₂₀ -Mri1a
30131	22.2	96	643.3917	4	84.85	2569.5239	5.3	N	MESLKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
8424	7.7	96	408.7527	2	39.52	815.4905	0.6	N	KLVAPAAF	U ₁₀ -Mri1c
9343	8.6	96	477.2684	2	41.82	952.5229	-0.6	N	GDPLDPKVL	U ₁₂ -Mri1a
19737	10.5	95	521.3376	2	62.97	1040.6594	1.3	N	VGGSLAALKLL	U ₁₀ -Mri1c
16238	8.6	95	524.2906	2	56.26	1046.5647	2	N	LDPKLVESF	U ₁₂ -Mri1a
9496	11.4	95	564.3604	2	42.49	1126.7073	-1	N	VGGSLAKAALK	U ₁₀ -Mri1c
19258	11.4	95	605.3636	2	62.21	1208.7241	-9.4	N	HTGSLAKAALK	U ₁₀ -Mri1c
14074	12.3	95	471.9607	3	52.39	1412.8601	0.1	N	ASKAEELLKLLA	U ₂₀ -Mri1a
19425	12.4	95	721.879	2	62.48	1441.7388	3.2	N	MTLPFLQHALTNG	U ₃ -Mri1a
20671	14.3	95	804.4436	2	65.74	1606.8718	0.5	N	KLGLFDQLDKAAAF	U ₁₃ -Mri1a
26221	14.3	95	561.3049	3	76.7	1680.8909	1.3	N	LFDQLDKAAAFMKL	U ₁₃ -Mri1a
19760	16.2	95	914.5253	2	63.01	1827.0327	1.8	N	KFGSMLAKAALKLLSYS	U ₁₀ -Mri1a
35996	20.9	95	805.1377	3	95.58	2412.4023	-4.5	N	LGMESLKQLSAKAEELLKLLA	U ₂₀ -Mri1a
6995	8.6	95	475.7782	2	35.61	949.5419	0	N	KAAAFMKL	U ₁₃ -Mri1a
17759	9.5	95	478.8106	2	58.82	955.6066	0.1	N	GVGSLKLVA	U ₁₀ -Mri1c
17407	9.5	95	481.787	2	58.2	961.5596	-0.1	N	VGGSLAKAF	U ₁₀ -Mri1c
4206	9.4	94	519.2667	2	24.17	1036.5188	0.1	N	ALVHAPETEA	
17109	10.3	94	587.8817	2	57.69	1173.7485	0.3	N	SLLAKAALKLF	U ₁₀ -Mri1c
16388	10.3	94	619.8607	2	56.53	1237.707	-0.1	N	GLKDALFLSFK	U ₁₀ -Mri1b
21728	9.4	94	630.8195	2	68.19	1259.626	-1.2	N	LFSKPLYDMF	
14672	11.3	94	656.3552	2	53.53	1310.6982	-1.8	N	LTPFLQHALTNG	U ₃ -Mri1a
11457	12.3	94	689.4175	2	46.58	1376.8213	-0.6	N	GFKSMLAKAALKL	U ₁₀ -Mri1a
31534	13.1	94	698.9478	2	87.41	1395.8813	-0.2	N	QAGSLLAKAALKLL	U ₁₀ -Mri1c
17782	13.1	94	491.6569	3	58.85	1471.949	-0.1	N	GFKSMLAKAALKL	U ₁₀ -Mri1c
17775	13.2	94	381.9795	4	58.84	1523.8896	-0.4	N	GFKSMLAKAALKLF	U ₁₀ -Mri1a
35671	19.8	94	694.1133	3	94.91	2079.3181	0	N	VGGSLAKAALKLLKLVAPAF	U ₁₀ -Mri1c
27134	26.2	94	913.8862	3	78.57	2738.6243	4.6	N	GFKSMLAKAALKLLKAVAPAAAAALADK	U ₁₀ -Mri1a
17133	8.5	94	470.8315	2	57.99	939.648	0.5	N	LAKAALKLL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
7067	8.4	93	521.2721	2	35.36	1040.5291	0.6	N	PFLQHALTD	U ₃ -Mri1a
15763	9.3	93	524.8113	2	55.47	1047.6077	0.3	N	LGPLLHALTN	U ₃ -Mri1a
15899	10.2	93	570.8896	2	55.93	1139.7642	0.5	N	LSLAKAALKLL	U ₁₀ -Mri1c
21888	10.2	93	570.8868	2	68.44	1139.7642	-4.4	N	SLLAKAALKLL	U ₁₀ -Mri1c
17180	11.2	93	453.9438	3	57.79	1358.8108	-0.9	N	KLLAAKALMSKW	U ₁₀ -Mri1a
35601	13	93	699.4414	2	94.79	1396.8652	2.2	N	AEGSLLAAKALKLL	U ₁₀ -Mri1c
32569	13.1	93	718.4496	2	89.2	1434.8921	-5.1	N	HTGSLAKAALKLL	U ₁₀ -Mri1c
25374	13	93	481.9819	3	75.78	1442.9224	1	N	FVGSLAKAALKLL	U ₁₀ -Mri1c

25734	13.9	93	530.9814	3	75.85	1589.9214	0.6	N	GFVESMLAKAALKLL	U ₁₀ -Mri1a
20644	14	93	536.6313	3	65.66	1606.8718	0.1	N	KLGLFDQLDKAAAAAF	U ₁₃ -Mri1a
7953	15.9	93	581.3284	3	38.16	1740.9622	0.6	N	KAEVPTVAAALENKVTT	U ₁₀ -Mri1b
35630	15.8	93	584.3549	3	94.84	1750.0393	2	N	VGVSLLAKAALKLLSYF	U ₁₀ -Mri1c
41529	16.7	93	604.7178	3	106	1811.1318	-0.1	N	VGGSLLAKAALKLLMTL	U ₁₀ -Mri1c
27077	18.7	93	946.5823	2	78.46	1891.1506	-0.3	N	VGVADPGVGSLLAKAALKLL	U ₁₀ -Mri1c
24245	16.7	93	660.7042	3	73.25	1979.0913	-0.2	N	KLLGFDQLDKAAAAFMLK	U ₁₃ -Mri1a
22410	21.4	93	616.0914	4	69.78	2460.3562	-8	N	PQGAKKLGFDQLDKAAAAFMKL	U ₁₃ -Mri1a
16750	10.1	92	542.8583	2	57.14	1083.7014	0.6	N	GSLLAALKL	U ₁₀ -Mri1c
26973	8.3	92	544.2949	2	78.22	1086.575	0.3	N	KDALFLSF	U ₁₀ -Mri1b
22133	10.1	92	570.887	2	69.12	1139.7642	-4.1	N	LSLAKAALKLL	U ₁₀ -Mri1c
14079	10.1	92	579.866	2	52.4	1157.7205	-2.6	N	SMLAKAALKLL	U ₁₀ -Mri1a
10055	12	92	590.8336	2	43.47	1179.6499	2.4	N	TPVAAALTAAPTP	U ₁₀ -Mri1b
26550	10.1	92	394.9545	3	77.3	1181.8223	16.5	N	GLLKKLLAKKA	U ₂₀ -Mri1a
11472	12	92	652.8646	2	46.6	1303.717	-1.9	N	GLMESLKAGLSAK	U ₂₀ -Mri1a
18100	12	92	711.3774	2	59.84	1420.7383	1.5	N	QDLDKAAAAFMKL	U ₁₃ -Mri1a
34349	13.8	92	740.4611	2	92.69	1478.9185	-7.3	N	QPGGSLLAKAALKLL	U ₁₀ -Mri1c
13615	12.8	92	753.9579	2	51.53	1505.9004	0.6	Y	KFGSM(+15.99)LAKAALKLL	U ₁₀ -Mri1a
25608	12.9	92	806.4514	2	75.62	1610.8855	1.7	N	LLMTLPFLQHALTN	U ₃ -Mri1a
13826	13.7	92	540.3403	3	53.4	1618.0002	-0.7	N	GFKSMLAKAALKLLK	U ₁₀ -Mri1a
21813	16.6	92	465.792	4	68.32	1859.1431	-2.3	N	GFKSMLAKAALKLLAKVA	U ₁₀ -Mri1a
27034	18.4	92	631.3918	3	78.36	1891.1506	1.6	N	VGVADPGVGSLLAKAALKLL	U ₁₀ -Mri1c
28156	16.5	92	661.4224	3	80.46	1981.2273	9.1	N	KGVPKLLKLAALKMSKW	U ₁₀ -Mri1a
19757	18.3	92	737.4241	3	63	2209.2478	1.2	N	GFKSMLAKAALKLVSMRPVY	U ₁₀ -Mri1a
35369	19.3	92	781.4626	3	94.41	2341.3652	0.3	N	GLMESLKQLSAKAEELLKLL	U ₂₀ -Mri1a
32899	21.1	92	636.1313	4	89.72	2540.4973	-0.4	N	AVMESLKQLSAKAEELLKLLAK	U ₂₀ -Mri1a
17766	22	92	648.6066	4	58.83	2590.394	1.2	N	DKPGQAKKLGFDQLDKAAAAFMK	U ₁₃ -Mri1a
28565	22.2	92	537.5212	5	81.61	2682.6079	-14	N	LMESLKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
21922	8.3	92	470.831	2	68.54	939.648	-0.6	N	LAKAALKLL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
19768	9.1	91	547.8252	2	63.02	1093.6382	-2.1	N	VLDPKLEAPL	U ₁₂ -Mri1a
16462	10.9	91	588.8186	2	56.65	1175.6221	0.5	N	LGMESLKAGLSA	U ₂₀ -Mri1a
8447	11.9	91	621.3403	2	39.59	1240.6663	-0.1	N	DLPTVAAALEGGK	U ₁₀ -Mri1b
21979	11	91	645.3485	2	68.71	1288.6848	-1.8	N	LEPQAAAAFMKL	U ₁₃ -Mri1a
21853	12.8	91	445.9628	3	68.37	1334.865	1.3	N	GVPAALAKAALKLL	U ₁₀ -Mri1c
21741	11.8	91	476.2975	3	68.21	1425.8818	-7.7	N	GVTLAKAALKLRW	U ₁₀ -Mri1c
25307	14.6	91	513.3398	3	75.07	1536.9966	0.6	N	LAGVGSLLAKAALKLL	U ₁₀ -Mri1c
21025	13.7	91	559.6636	3	66.6	1675.9846	-9.4	N	WGFKSMLAKAALKLL	U ₁₀ -Mri1a
29010	13.6	91	848.942	2	82.37	1695.8655	2.3	N	QDLDKAAAAFMKLFQ	U ₁₃ -Mri1a
16387	14.6	91	588.3126	3	56.53	1761.9182	-1.2	Y	GLM(+15.99)ESLKQLSAKAEEL	U ₂₀ -Mri1a
27955	16.3	91	589.058	3	80.04	1764.1599	-4.2	N	GVTLAKAALKLLKLVGGGL	U ₁₀ -Mri1c
25596	15.5	91	620.6874	3	75.59	1859.0437	-1.7	N	LGMESLKLSAKAEKLL	U ₂₀ -Mri1a
24401	16.3	91	663.3845	3	73.51	1987.1023	14.7	N	LGMESLKQLSAKAEELLK	U ₂₀ -Mri1a
27922	18.2	91	1115.147	2	79.99	2228.2812	-0.9	N	LGMESLKQLSAKAEELLKLL	U ₂₀ -Mri1a
6177	6.4	91	407.2394	2	32.78	812.4644	0	N	LDPKVL	U ₁₂ -Mri1a
12054	8.2	91	426.2516	2	48.38	850.4912	-2.9	N	TVAAPAKPP	
19116	8.1	90	513.7756	2	61.98	1025.5369	-0.1	N	AAAFMKLFQ	U ₁₃ -Mri1a
9738	9.9	90	376.5765	3	42.98	1126.7185	-9.5	N	RGSLLAKAALK	U ₁₀ -Mri1c
39049	10.8	90	626.3638	2	101.7	1250.7122	0.8	N	ESEGPLLLALL	U ₃ -Mri1a
13518	12.6	90	689.41	2	51.33	1376.8027	1.9	N	KLVAPAAAEVLAPQ	U ₁₀ -Mri1c
11451	11.7	90	459.9478	3	46.57	1376.8213	0.1	N	FGKSMLAKAALKL	U ₁₀ -Mri1a
17821	10.8	90	706.4575	2	58.93	1410.8962	3	N	KLDALAKLWKLL	U ₁₀ -Mri1b
13867	11.7	90	478.6354	3	52.01	1432.884	0.2	N	FKSMLAKAALKLL	U ₁₀ -Mri1a
30824	13.6	90	727.4547	2	86.17	1452.8916	2.3	N	VGGSLLAEVAALKLL	U ₁₀ -Mri1c
27317	13.5	90	733.9883	2	78.92	1465.9595	1.7	N	VGGSLLAKAALKLL	U ₁₀ -Mri1c
38111	12.6	90	741.4498	2	100	1480.8799	3.5	Y	HM(+15.99)GSLLAKAALKLL	U ₁₀ -Mri1c
26434	14.4	90	783.4821	2	77.08	1564.9551	-3.4	N	DPGVGSLLAKAALKLL	U ₁₀ -Mri1c
27098	14.4	90	528.0007	3	78.5	1580.9866	-4	N	LDGVGSLLAAKALKLL	U ₁₀ -Mri1c
27306	16.2	90	495.7801	4	78.9	1979.0913	0	N	KLGLFDQLDKAAAAFMKL	U ₁₃ -Mri1a
32346	19.8	90	608.1074	4	88.83	2428.3975	1.3	Y	GLM(+15.99)ESLKQLSAKAEELLKLLA	U ₂₀ -Mri1a
41912	25.1	90	978.2902	3	106.5	2931.8364	4.2	N	KNVYLAKAALKLLKLVAPAAAEVLAVVR	U ₁₀ -Mri1c
6476	6.3	90	389.2122	2	33.38	776.4102	-0.5	N	AVMESLK	U ₂₀ -Mri1a
15920	6.3	90	393.2433	2	55.71	784.4694	3.3	N	PKVLES	U ₁₂ -Mri1a
25470	8.1	90	428.2558	2	75.34	854.4974	-0.4	N	RLSALAGAP	
19765	8.1	90	448.3026	2	63.01	894.5902	0.6	N	KAGLLPLAL	U ₃ -Mri1a
9507	9	90	486.3156	2	42.27	970.6175	-0.8	N	SGLLAKAALK	U ₁₀ -Mri1c
19706	8.9	89	530.2718	2	62.93	1058.5283	0.7	N	VVTGDPLDPF	U ₁₂ -Mri1a
18101	10.7	89	614.3685	2	59.84	1226.7346	-9.8	N	QQGSLLAKAALK	U ₁₀ -Mri1c
19779	9.7	89	617.8912	2	63.03	1233.7695	-1.4	N	VPPELKLQVQL	

21327	9.8	89	627.3538	2	67.32	1252.6816	9.1	N	PFLQHALTLDV	U ₃ -Mri1a
25824	11.6	89	661.9423	2	76.02	1321.8696	0.4	N	VPSVLAKAALKLL	U ₁₀ -Mri1a
11473	11.5	89	460.2824	3	46.6	1377.8167	6.4	N	GFKSMKGKAAKLL	U ₁₀ -Mri1a
22758	12.5	89	694.437	2	71.47	1386.8599	-0.3	N	VGSSLAAKALKLF	U ₁₀ -Mri1c
6570	12.5	89	720.9065	2	33.91	1439.7983	0.1	N	KAIEVPVTAALLENK	U ₁₀ -Mri1b
34790	12.4	89	754.4584	2	94.23	1506.9172	-10	N	FYGSLLAKAALKLL	U ₁₀ -Mri1c
35112	13.3	89	754.9612	2	93.97	1507.9084	-0.3	N	SHSGSLLAKAALKLL	U ₁₀ -Mri1c
15162	14.2	89	545.2874	3	54.44	1632.8391	0.7	N	GLMESLKAGLSAKAE	U ₂₀ -Mri1a
21796	15.2	89	467.51	4	68.29	1866.0073	1.9	N	KLGLFDQLDKAAAAMFK	U ₁₃ -Mri1a
19665	16	89	393.8476	5	62.87	1964.2034	-0.8	N	PVDAKAEELLKLLAKKA	U ₂₀ -Mri1a
32415	18.7	89	692.3763	3	88.95	2074.1057	0.6	N	TAEALAKALAEFAEALGTRA	CL24Contig3/4 1
19759	17.8	89	442.8572	5	63	2209.2478	0.8	N	GFKSMLAKAALKLVSMRPVY	U ₁₀ -Mri1a
28854	18.7	89	462.8903	5	82.01	2309.4158	-0.2	N	ATRQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
27815	23.1	89	812.1389	3	79.83	2433.384	4.4	N	VGEADAVGVADPVGVSLLAKAALKLL	U ₁₀ -Mri1c
29171	20.5	89	514.921	5	82.76	2569.5569	4.6	N	PYSLKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
25113	21.3	89	671.6597	4	74.73	2682.6079	0.7	N	LMESLKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
41783	25	89	708.1867	4	106.3	2828.7214	-1.3	N	QQGSLLAKAALKLLKLVAPAAAEVLANK	U ₁₀ -Mri1c
27817	26.6	89	975.918	3	79.84	2924.7246	2.6	Y	KFGSM(+15.99)LAKAALKLLKAVAAPAAAALADKLG	U ₁₀ -Mri1a
7322	6.2	89	377.2017	2	36.45	752.3891	-0.3	N	GFKSMLA	U ₁₀ -Mri1a
7977	7.1	88	506.7653	2	38.23	1011.5137	2.3	N	PDEVFLHR	
14691	9.7	88	553.2972	2	53.58	1104.585	-4.6	N	LGMESLKAGLS	U ₂₀ -Mri1a
10385	9.7	88	568.8279	2	44.74	1135.6423	-0.9	N	GFKSMLAKAAL	U ₁₀ -Mri1a
27598	8.8	88	583.8447	2	79.44	1165.6746	0.2	N	LKDALFLSFL	
22786	10.6	88	616.392	2	70.57	1230.77	-0.4	N	GSLLAKAALKLF	U ₁₀ -Mri1c
19007	10.6	88	641.8495	2	61.78	1281.6816	2.2	N	DGPLDPKVLLESL	U ₁₂ -Mri1a
15574	10.5	88	433.2632	3	55.11	1296.7651	2	N	LDPKVLLESLVGK	U ₁₂ -Mri1a
18114	11.5	88	474.5867	3	59.87	1420.7383	0.1	N	NELDKAAAAMFKL	U ₁₃ -Mri1a
27924	13.2	88	734.9584	2	79.99	1467.9023	-0.1	N	DGVGSLLAAKALKLL	U ₁₀ -Mri1c
17762	12.3	88	508.9695	3	59.42	1523.8896	-2	N	GFKSMLAKAALKLF	U ₁₀ -Mri1a
30501	12.3	88	820.4502	2	85.81	1638.8979	-7.4	N	KDLQDFLGLKKAYT	
19647	15.9	88	492.0578	4	62.84	1964.2034	-0.7	N	PVDAKAEELLKLLAKKA	U ₂₀ -Mri1a
30520	17.5	88	675.6989	3	85.56	2024.1055	-15	N	PQGAKKDKLGLFDQLAAPGA	U ₁₃ -Mri1a
11456	17.5	88	528.3357	4	46.58	2109.3247	-5.2	N	KAGLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
32078	20.2	88	551.8492	4	88.64	2203.3667	0.6	N	VGGSLLAKAALKLLKAVLAPAAE	U ₁₀ -Mri1c
28675	19.4	88	488.7066	5	81.6	2438.4834	5.5	N	ESLKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
32888	20.2	88	509.1073	5	89.7	2540.4973	1.1	N	LGMESLKQLSAKAEELLKLLAK	U ₂₀ -Mri1a
28728	20.1	88	515.1099	5	81.73	2570.5193	-2.5	Y	NM(+15.99)TVKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
30848	22	88	457.6122	6	86.71	2739.6294	0.1	N	LGMESLKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
14112	6.1	88	415.2178	2	52.47	828.4203	0.8	N	GFKSMLF	U ₁₀ -Mri1a
18050	6.1	88	440.2518	2	59.68	878.4902	-1.2	N	VEPKFLF	
21865	7.9	88	477.76	2	68.39	953.5069	-1.4	N	ESEGPLLL	U ₃ -Mri1a
37223	8.8	88	491.3212	2	98.36	980.627	0.8	N	TAGLPLLALL	U ₃ -Mri1a
42020	8.7	87	514.3315	2	106.7	1026.6477	0.8	N	FAGLPLLALL	U ₃ -Mri1a
17223	7.9	87	529.2786	2	58.12	1056.5427	-0.1	N	MTLPFLQHA	U ₃ -Mri1a
19971	8.7	87	538.8071	2	63.7	1075.6025	-2.6	N	NNVPKLTPPP	
14101	9.5	87	386.914	3	52.45	1157.7205	-0.2	N	SMLAKAALKLL	U ₁₀ -Mri1a
42126	9.5	87	585.8624	2	106.9	1169.7092	0.9	Y	AVPLLALLM(+15.99)TL	U ₃ -Mri1a
16737	11.4	87	620.9026	2	57.38	1239.7915	-0.6	N	VGGSLLAKAALKL	U ₁₀ -Mri1c
9229	12.1	87	422.9192	3	41.55	1265.7344	1	N	LVAPAAAEVLGAGGK	U ₁₀ -Mri1c
9896	10.4	87	429.6131	3	43.33	1285.8154	1.6	N	MSLAKAALKLL	U ₁₀ -Mri1a
9104	12.2	87	689.4073	2	41.25	1376.8027	-1.9	N	KLVAAPAAEVLQAP	U ₁₀ -Mri1c
6735	12.2	87	697.9221	2	35.64	1393.8293	0.3	N	KLVAAPAAEVLANK	U ₁₀ -Mri1c
35076	12.2	87	732.4659	2	93.92	1462.9121	3.5	N	PPESLLAKAALKLL	U ₁₀ -Mri1c
41720	11.3	87	748.9018	2	106.3	1495.8035	-9.6	N	NTELPFLQHALTL	U ₃ -Mri1a
20914	13	87	776.4336	2	66.32	1550.8555	-1.8	N	VVAGDPLDPKVLLESL	U ₁₂ -Mri1a
34360	13.9	87	782.4961	2	92.7	1562.9758	1.2	N	PTAPGSLLAAKALKLL	U ₁₀ -Mri1c
34296	13	87	782.9148	2	92.6	1563.8118	2.1	N	PPTATAAAAMKFLFQ	U ₁₃ -Mri1a
20809	13	87	791.4424	2	66.06	1580.866	2.7	N	VVTGDPLDPKVLLESL	U ₁₂ -Mri1a
27433	13	87	542.6653	3	79.12	1624.9697	2.8	N	QPNKSMMLAKAALKLL	U ₁₀ -Mri1a
16629	16.6	87	664.7019	3	56.92	1991.084	0	N	NAALKKGLFDQLDKAAAAMFK	U ₁₃ -Mri1a
39500	18.3	87	686.4333	3	102.5	2056.2883	-5	N	VGGTVLAKAALKLLAKAAPHK	U ₁₀ -Mri1c
27899	17.3	87	446.6631	5	79.96	2228.2812	-0.8	N	GLMELSKQLSAKAEELLKLL	U ₂₀ -Mri1a
36795	20.9	87	762.4874	3	97.98	2284.4609	-9	N	VGGTVLAKAALKLLKLVAPAAALP	U ₁₀ -Mri1c
29000	20.1	87	643.6553	4	82.35	2570.5522	15.6	N	NFTVKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
35335	24.4	87	683.4293	4	94.36	2729.678	3.7	N	VGGSLLAKAALKLLKLVAPAAAEVLADK	U ₁₀ -Mri1c
41855	24.4	87	705.9445	4	106.4	2819.7251	8.4	N	FLALVEAAAPAVLKLKLLAKALLSGTQ	U ₁₀ -Mri1c
41848	25.4	87	712.1841	4	106.4	2844.7163	-3.2	N	NSAGSLLAKAALKLLKLVAPAAAEVLANK	U ₁₀ -Mri1c
27834	25.4	87	951.5855	3	79.86	2851.7083	9.3	N	GFKSMLAKAALKLLKAVAAPAAAALADKL	U ₁₀ -Mri1a

9528	7.9	87	457.8066	2	42.33	913.596	2.9	N	SLLAKAALK	U ₁₀ -Mri1c
4700	7.9	87	468.2507	2	27.09	934.4872	-0.4	N	FAEALGRAT	CL24Contig3/4 1
19762	7	87	470.7613	2	63.01	939.5065	1.7	N	KDALFLSF	U ₁₀ -Mri1b
16960	7	87	493.7603	2	57.45	985.5055	0.6	N	MTLPFLQH	U ₃ -Mri1a
21133	8.6	86	556.8315	2	66.84	1111.6489	-0.3	N	LDPKVLESLV	U ₁₂ -Mri1a
7976	8.6	86	599.8146	2	38.23	1197.6143	0.3	N	VSPDEVFLHR	
9669	12	86	625.3652	2	42.56	1248.719	-2.4	N	KAVAPAAAAALAPQ	U ₁₀ -Mri1a
29551	9.4	86	627.3763	2	83.62	1252.7437	-4.5	N	KNHCCLKAKKA	U ₂₀ -Mri1a
21810	11.1	86	677.4447	2	68.43	1352.8867	-8.7	N	RGSLLAALKLL	U ₁₀ -Mri1c
17896	10.3	86	684.8658	2	59.21	1367.7197	-1.9	N	PFLQHALTNNTL	U ₃ -Mri1a
14428	12.1	86	688.3845	2	53.06	1374.7542	0.2	N	VAMESLKAGLSAKA	U ₂₀ -Mri1a
42242	12.9	86	703.9517	2	107.1	1405.8909	-1.4	N	GVGSLLAGLPLALL	U ₁₀ -Mri1c
17823	10.4	86	471.3065	3	58.93	1410.8962	1.2	N	LKDALAKLWKLL	U ₁₀ -Mri1b
9271	13.7	86	479.9522	3	41.63	1436.835	-0.1	N	KLGAAPAAAAALADKLG	U ₁₀ -Mri1a
21210	11.2	86	485.2686	3	67	1452.7798	3	N	LDKAAAFMKLFQ	U ₁₃ -Mri1a
31887	12.1	86	727.4528	2	88.45	1452.9028	-8.1	N	QQGSLLAKAALKLL	U ₁₀ -Mri1c
18109	12	86	494.6636	3	59.86	1480.9817	-8.5	N	RGSLLAKAALKLLK	U ₁₀ -Mri1c
17123	12	86	745.9608	2	58.23	1489.9053	1.1	N	GFKSMLAKAALKLL	U ₁₀ -Mri1a
39826	12	86	755.4489	2	103	1508.8748	5.7	N	CEHSLLAKAALKLL	U ₁₀ -Mri1c
11693	13.7	86	782.9754	2	47.63	1563.9348	0.9	N	KLVAAPAAEVLANKLG	U ₁₀ -Mri1c
42303	14.6	86	852.9577	2	107.2	1703.8982	1.6	N	TPPAEAESEPLPLAL	U ₃ -Mri1a
9577	13.7	86	436.0295	4	44.44	1740.0872	1	N	ASKAEELKKLLAKKA	U ₂₀ -Mri1a
12014	14.6	86	595.0208	3	48.31	1782.0403	0.1	N	KLVAAPAAEVLANKLQF	U ₁₀ -Mri1c
30604	17.1	86	676.3676	3	85.73	2026.092	-5.4	N	PQGAKKGLFDGALMLAAPE	U ₁₃ -Mri1a
19758	17.2	86	682.4457	3	63	2044.3135	0.9	N	ALKLLKLVAPAAEVLAVVR	U ₁₀ -Mri1c
31766	19	86	692.448	3	88.08	2074.324	-0.9	N	VGGSLLAKAALKLLKLVAPAAA	U ₁₀ -Mri1c
24415	18.1	86	420.6614	5	73.53	2098.2698	0.5	N	GFKSMLAKAALKLLAKAVPAA	U ₁₀ -Mri1a
41981	23.3	86	877.1174	3	106.6	2628.3394	-3.4	N	DPAAEATAEALAKALAEFAEALGTRA	CL24Contig3/4 1
28522	21.4	86	1371.32	2	81.27	2740.6248	0.5	N	AVMESLKQLSAKAEKVKLLAKKA	U ₁₀ -Mri1c
40948	24.9	86	948.2719	3	105.2	2841.7781	5.5	N	VGSVLAKAALKLLKLVAPAAEVLANKL	U ₁₀ -Mri1c
41850	24.2	86	953.5998	3	106.4	2857.7441	11.7	N	DPSLLAKAALKLLKLVAPAAEVLAMKL	U ₁₀ -Mri1c
41872	24.2	86	957.2731	3	106.5	2868.7891	2.9	N	LEAQLAKAALKLLKLVAPAAEVLAVVR	U ₁₀ -Mri1c
41876	25.1	86	975.2721	3	106.5	2922.8108	-5.6	N	HTGSLLAKAALKLLKLVAPAAEVLARVV	U ₁₀ -Mri1c
41302	25	86	733.7189	4	105.6	2930.8411	1.8	N	VFGSLLAKAALKLLKLVAPAAEVLAVVR	U ₁₀ -Mri1c
41934	24.9	86	1479.415	2	106.6	2956.8052	3.5	N	DTAELLAKAALKLLKLVAPAAEVLARVV	U ₁₀ -Mri1c
27377	22.4	86	993.8811	3	79.03	2978.6091	4.2	N	FLEFLLAAHKDQLDKAAAFMKLFQ	U ₁₃ -Mri1a
7395	6	86	393.2417	2	36.54	784.4694	-0.7	N	PKVLES	U ₁₂ -Mri1a
42021	8.5	85	511.3183	2	106.7	1020.6219	0.1	N	LPLGPDALL	
38287	8.5	85	516.309	2	100.3	1030.6062	-2.7	N	EAGLFFVALL	
15348	9.3	85	596.8605	2	54.74	1191.7048	1.3	Y	AM(+15.99)LAKAALKLF	U ₁₀ -Mri1a
6823	10.2	85	422.2529	3	37.82	1263.7373	-0.2	N	GFKSMLAKAALK	U ₁₀ -Mri1a
21843	11.1	85	649.4099	2	68.36	1296.813	-6	N	VGSLLAKAALKLN	U ₁₀ -Mri1c
16713	10.2	85	670.3651	2	57.08	1338.7183	-2	N	LVHAEPTEALLF	
34059	11.1	85	685.8591	2	92.19	1369.7029	0.6	N	EAPGPSLGVHFLF	
21816	11.9	85	457.9498	3	68.32	1370.8318	-3	N	GVGSLLAKAALKML	U ₁₀ -Mri1c
27313	11.9	85	489.6609	3	78.92	1465.9707	-6.6	N	GVTLAKAALKLRL	U ₁₀ -Mri1c
12028	12.8	85	377.4889	4	48.59	1505.9292	-1.8	N	KLVAAPAAEVLARVV	U ₁₀ -Mri1c
30036	13.6	85	768.4792	2	84.65	1534.9446	-0.4	N	PSGPGSLLAKAALKLL	U ₁₀ -Mri1c
7503	12.7	85	770.4398	2	38	1538.8555	6.3	N	KAETVPVAALEDLL	U ₁₀ -Mri1b
8693	12.8	85	414.2706	4	43.4	1653.0552	-1	N	KAAEELKKLLAKKA	U ₂₀ -Mri1c
34793	15.4	85	856.9402	2	93.46	1711.8628	1.8	N	PGEAEATAEALAKALAEA	CL24Contig3/4 1
19763	13.5	85	430.5121	4	63.01	1718.0164	1.6	N	GVFESMLAKAALKLLK	U ₁₀ -Mri1a
27979	15.4	85	883.5851	2	80.07	1765.1553	0.3	N	VGSLLAKAALKLLKVKG	U ₁₀ -Mri1c
27635	16.1	85	620.6763	3	79.51	1859.0073	-0.1	N	EMSLGLKAGLSASAEGLL	U ₂₀ -Mri1a
27888	17.8	85	645.394	3	79.94	1933.1724	-6.3	N	AAAAPAVKLLKAAAALADKLG	U ₁₀ -Mri1a
11428	17	85	422.872	5	46.54	2109.3247	-0.4	N	KAGLGTCAEELKKLLAKAK	U ₂₀ -Mri1a
27723	20.3	85	760.1038	3	79.68	2277.2942	-2	N	ADEAVGVADPGVGSLLAKAALKLL	U ₁₀ -Mri1c
28031	18.6	85	574.3477	4	80.17	2293.3818	-8.8	N	RGPAAVAKALLKAAKALMSKW	U ₁₀ -Mri1a
26385	19.5	85	518.114	5	77.05	2585.5188	5.8	Y	M(+15.99)ESLKQLSAKAEELKKLLAKKA	U ₂₀ -Mri1a
27895	24.6	85	945.2446	3	79.95	2832.7026	3.4	N	GFKSMLAKAALKLLKAVAPAAAAALLPTP	U ₁₀ -Mri1a
39518	23.9	85	967.2623	3	102.5	2898.7898	-8.5	N	WARLAKAALKLLKLVAPAAEVLANKLG	U ₁₀ -Mri1c
32966	23	85	731.4402	4	89.81	2921.7349	-1	N	VGPVMESLKQLSAKAEELKKLLAKKA	U ₂₀ -Mri1a
41785	25.5	85	746.2218	4	106.4	2980.8528	1.8	N	PQAGSLLAKAALKLLKLVAPAAEVLAVVR	U ₁₀ -Mri1c
34368	23.9	85	997.2668	3	92.72	2988.7625	5.4	N	VAKDALAKLWKLLKAEVPTVAALEDLL	U ₁₀ -Mri1b
8098	7.6	85	408.2526	2	38.76	814.4912	-0.6	N	VGGSLLA	U ₁₀ -Mri1c
11450	6.8	85	463.759	2	46.57	925.5021	1.4	N	TLPFLQHA	U ₃ -Mri1a
14687	7.6	85	500.8125	2	53.57	999.6116	-1.2	N	LLGPVLKNF	
15554	8.4	84	520.7749	2	55.07	1039.5371	-1.8	N	LGMTYSKAGL	

6008	7.6	84	520.7833	2	32.98	1039.545	6.7	N	PFLQHALTN	U ₃ -Mri1a
21722	9.2	84	576.8095	2	68.17	1151.5974	6.1	N	AHGALFPLTEP	
7954	8.4	84	400.2117	3	38.16	1197.6143	-0.8	N	VSPDEVFLHR	
29462	9.2	84	606.371	2	83.42	1210.7219	4.6	N	TVHCKLLAKKA	U ₂₀ -Mri1a
17951	9.2	84	627.3392	2	59.62	1252.6638	0	N	PFLQHALTMVP	U ₃ -Mri1a
22913	10.1	84	632.8466	2	70.82	1263.6824	-2.9	N	LPKVESLDGLHG	
17147	10	84	429.612	3	58	1285.8154	-0.9	N	KSMLAKAALKLL	U ₁₀ -Mri1a
13049	10.9	84	674.8784	2	50.39	1347.7407	1.2	N	LGPLKKVPCCPP	U ₆ -Mri1a
17698	11.7	84	690.4166	2	58.72	1378.8184	0.2	N	LVAPAAAELVALKN	U ₁₀ -Mri1c
22987	10.9	84	463.294	3	71.71	1386.8711	-7.8	N	GSLLAOKALKLRF	U ₁₀ -Mri1c
34658	10.9	84	722.9664	2	93.21	1443.9065	8.2	N	YVELLAKAALKLL	U ₁₀ -Mri1c
23389	12.6	84	725.9722	2	71.75	1449.9282	1.2	N	PGVGSLLAKAALKLL	U ₁₀ -Mri1c
33640	12.6	84	740.9648	2	91.31	1479.9136	1.1	N	PQGGSLLAOKALKVK	U ₁₀ -Mri1c
32456	12.7	84	750.4554	2	89.03	1498.8904	4	N	QCAGSLLAOKALKLL	U ₁₀ -Mri1c
33872	12.7	84	764.4526	2	92.31	1526.9031	-8.1	N	QDSGSLLAOKALKLL	U ₁₀ -Mri1c
23109	12.6	84	530.9816	3	71.29	1589.9214	0.9	N	FGKSMADLAALKLL	U ₁₀ -Mri1a
30980	12.6	84	800.96	2	86.46	1599.9019	2.2	N	QNAESMLAKAALKLL	U ₁₀ -Mri1a
25412	12.6	84	803.965	2	75.24	1605.9163	-0.5	N	YEVGSMLAKAALKLL	U ₁₀ -Mri1a
16427	13.4	84	423.2667	4	57.16	1689.0374	0.2	N	KFGSMLAKAALKLLKA	U ₁₀ -Mri1a
18457	14.2	84	572.6976	3	60.87	1715.0596	6.6	N	LLKLVAPAAAEVLLTPP	U ₁₀ -Mri1c
42297	14.4	84	863.9498	2	107.2	1725.8823	1.6	N	PPTAEAESEFAAPLLAL	
11431	15.1	84	867.5523	2	46.54	1733.0925	-1.4	N	KLLKAVAAPAAAALALNK	U ₁₀ -Mri1a
9583	13.5	84	581.0366	3	43.2	1740.0872	0.5	N	TGKAEELLKLLAKKA	U ₂₀ -Mri1a
21277	14.3	84	582.9808	3	67.19	1745.9233	-1.6	N	GLMESLKAGLSAKAEEL	U ₂₀ -Mri1a
21284	13.4	84	873.9689	2	67.21	1745.9233	0	N	GLMESLKQLSAKAEEL	U ₂₀ -Mri1a
37331	20.2	84	762.4918	3	98.57	2284.4609	-3.3	N	VGVASLAKAALLKLVAAAPALP	U ₁₀ -Mri1c
42281	17.6	84	1146.147	2	107.5	2290.2866	-3.2	Y	LPLAVALLDVM(+15.99)TMLLALLPHM	
19559	19.4	84	772.5018	3	62.71	2314.4827	0.5	N	KAAALKLLKLVAPAAAEVLAVVR	U ₁₀ -Mri1c
29230	19.3	84	643.6364	4	82.9	2570.5193	-1.1	Y	LM(+15.99)SNKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
19643	21	84	861.2208	3	62.83	2580.6206	7.8	N	KAAALKLLKLVAPAAAEVLANKLKH	U ₁₀ -Mri1c
37736	22.7	84	678.4332	4	99.33	2709.6995	1.6	N	VGGSLLAOKALKLLKLVAAAPAAELKKH	U ₁₀ -Mri1c
37465	23.4	84	678.6786	4	98.82	2710.7087	-8.5	N	VGGSLLAOKLVGVKLLKLVAPAAAEVLVPV	U ₁₀ -Mri1c
27225	21.1	84	460.278	6	79.25	2755.6243	0	Y	GLM(+15.99)ESLKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
32585	21	84	552.3301	5	89.48	2756.6162	-0.8	N	FGNESLKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
41156	24.3	84	711.4554	4	105.4	2841.7781	5.1	N	VGSVLAOKALKLLKLVAAAPAEVLAVKQ	U ₁₀ -Mri1c
34907	21.9	84	719.6818	4	93.64	2874.6726	8.8	N	PHAMESLKQLTGKAEELLKLLAKKA	U ₂₀ -Mri1a
27921	22.7	84	729.431	4	79.99	2913.6936	0.5	N	SSGLMETVKQLGTGKAEELLKLLAKKA	U ₂₀ -Mri1a
16663	6.7	84	412.7832	2	56.98	823.5531	-1.6	N	AVPLLALK	U ₃ -Mri1a
25442	7.5	84	428.2696	2	75.28	854.5225	2.4	N	SVLGALAKP	
17162	6.7	84	450.7353	2	57.77	899.4575	-1.6	N	FGKSMLAF	U ₁₀ -Mri1a
25421	7.6	84	469.326	2	75.25	936.6371	0.3	N	VAPLLALLK	U ₃ -Mri1a
41931	8.4	84	498.3095	2	106.6	994.6063	-1.8	N	GEGPLLLALL	U ₃ -Mri1a
14825	7.4	83	507.2967	2	53.87	1012.5804	-1.6	N	EVLKPDLSL	
17815	8.3	83	515.2772	2	58.91	1028.5331	6.5	N	FLAFAGGFLS	
15957	7.4	83	518.289	2	55.78	1034.5647	-1.2	N	LDPKLSYSL	U ₁₂ -Mri1a
12727	9.2	83	556.3411	2	49.81	1110.6648	2.5	N	PTVAAALEKVL	U ₁₀ -Mri1b
12040	10	83	567.8474	2	48.61	1133.6807	-0.4	N	KLVAAPAAALSPP	U ₁₀ -Mri1c
42304	9.1	83	577.8641	2	107.2	1153.7144	-0.5	N	VAPLLALLMTL	U ₃ -Mri1a
10118	9.9	83	590.8572	2	43.66	1179.6975	2.1	N	TPVAAALTARPL	U ₁₀ -Mri1b
19098	10.8	83	620.8429	2	61.95	1239.6709	0.3	N	LPDSAAVDAVVLA	
17663	10.9	83	622.3666	2	58.66	1242.7295	-8.7	N	QGGSGSLLAOKALK	U ₁₀ -Mri1c
23062	10	83	632.8345	2	71.21	1263.6646	-8	N	LPDLKSLGHMGP	
25407	9.1	83	433.6109	3	75.49	1297.8123	-1	N	KDALAKLWKLL	U ₁₀ -Mri1b
28740	10	83	435.9305	3	81.75	1304.7671	1.9	N	TACMKKLLAKAK	
42234	11.7	83	668.4343	2	107.1	1334.8538	0.2	N	VGSGLLGLPLLALL	U ₃ -Mri1a
25394	11.6	83	722.4649	2	75.21	1442.9224	-4.9	N	FVGSLLAAKALKLL	U ₁₀ -Mri1c
28420	10.8	83	484.6205	3	82.01	1450.8506	-7.4	N	LEEATHKLLAKKA	U ₂₀ -Mri1a
23266	10.8	83	486.9728	3	71.55	1457.897	-0.2	N	KNFELAKAALKLL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
36772	13.3	83	762.9838	2	97.45	1523.9399	8.6	N	VGGSLLAOKALKLLGN	U ₁₀ -Mri1c
9934	13.2	83	520.9619	3	43.51	1559.8672	-2.1	N	KAEVPTVAAAALVHGTP	U ₁₀ -Mri1b
24373	12.4	83	795.9678	2	73.47	1589.9214	-0.2	N	LGDFSMLAKAALKLL	U ₁₀ -Mri1a
31264	12.5	83	534.9997	3	87.93	1601.9724	3.1	N	KMTKLACAKAALKLL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
17742	13.2	83	535.9467	3	58.79	1604.8157	1.6	N	SSAHLGLSDALGHLDL	
31480	12.4	83	809.9659	2	87.3	1617.9197	-1.4	Y	M(+15.99)AELSMMLAKAALKLL	U ₁₀ -Mri1a
13816	12.5	83	405.5071	4	54.77	1618.0002	-0.5	N	KGFSMLAKAALKLLK	U ₁₀ -Mri1a
27943	15	83	608.7231	3	80.02	1823.1494	-1.1	N	SKAVSLVLGKAGLLLE	
41998	14.2	83	924.5468	2	106.7	1847.0742	2.6	N	LPGLLALLMTLPFLQHA	U ₃ -Mri1a
28059	15.7	83	521.3126	4	80.22	2081.2407	-9.4	N	AVAGPRKRKLAOKALMSKW	U ₂₀ -Mri1a

28253	16.6	83	539.0842	4	80.69	2152.2996	3.7	N	GFKSFPKAAALKLLAKAVRH	U ₁₀ -Mri1a
38678	20.8	83	816.4318	3	101	2446.2703	1.3	N	TAEATAEALAKALAEFAEALGRTA	CL24Contig3/4 1
28783	19	83	429.2603	6	81.84	2569.5205	-0.8	N	DVPKQYVSAKAEELLKLLAKKA	U ₂₀ -Mri1a
38061	23.2	83	678.6885	4	99.91	2710.7087	5.9	N	VGGSLLAKVGVKLLKLVAPAAAEVLVVP	U ₁₀ -Mri1c
27854	22.3	83	705.1473	4	79.89	2816.562	-0.7	N	QPGAKKDKLPDKYGAAATAAAAFMKLL	U ₁₃ -Mri1a
37589	23.3	83	565.9295	5	99.06	2824.6135	-0.9	N	VGGSLLAKAALKLLKLVAPAAAEVDMF	U ₁₀ -Mri1c
34210	21.6	83	710.9187	4	92.7	2839.6567	-3.9	N	QAAMESLKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
35597	24.8	83	725.9533	4	94.79	2899.7837	0.2	N	VGSLGLAKAALKLLKLVAPAAAEVLADKLG	U ₁₀ -Mri1c
40936	23.2	83	737.9741	4	105	2947.8538	4.6	N	KNALLRHAALKLLKLVAPAAAEVLAVVR	U ₁₀ -Mri1c
41910	25	83	1067.006	3	106.5	3197.988	2.2	N	VYELLAKAALKLLKLVAPAAAEVLLSLKH	U ₁₀ -Mri1c
41952	26.7	83	1081.687	3	106.6	3242.0508	-3.7	N	KYGSLAKAALKLLKLVAPAAAEVLAVLPVVL	U ₁₀ -Mri1c
14986	6.6	83	486.7695	2	54.17	971.5228	1.7	N	PFLQHALF	U ₃ -Mri1a
8881	7.4	82	517.7811	2	40.77	1033.5479	-0.2	Y	GLM(+15.99)ESLKQL	U ₂₀ -Mri1a
7269	7.4	82	520.288	2	36.02	1038.5498	11.1	N	PFLQHALVD	U ₃ -Mri1a
15819	7.4	82	528.3515	2	55.55	1054.6902	-1.6	N	ALAKLWKLL	U ₁₀ -Mri1b
23022	8.2	82	555.8135	2	71.11	1109.6121	0.3	N	LGKDALFLSF	
9114	9.8	82	567.8453	2	41.27	1133.6809	-4.2	N	LAVKAPAAATVPP	
6037	10.7	82	570.3257	2	32.02	1138.6345	2	N	LGAPAAAAALADK	U ₁₀ -Mri1a
11356	10.6	82	582.3441	2	46.42	1162.6709	2.4	N	LGPAAAAALALQP	U ₁₀ -Mri1c
5412	9	82	390.5676	3	29.86	1168.6816	-0.5	N	LDPKALKSLQG	U ₁₂ -Mri1a
25095	9.9	82	448.2756	3	74.71	1341.8167	-8.6	N	MTPNKKLLAKKA	U ₂₀ -Mri1a
11958	9.9	82	448.2811	3	48.17	1341.823	-1	N	SAKAEELLKLL	U ₂₀ -Mri1a
11432	10.6	82	681.8857	2	46.54	1361.7554	1.2	N	TEPEALVHALVAL	
21108	9.8	82	692.8741	2	66.78	1383.7219	8.5	N	MTLPFLQHALDV	U ₃ -Mri1a
18397	9.9	82	693.3692	2	61.76	1384.7173	4.8	N	MTLPFLQHALTN	U ₃ -Mri1a
33380	11.5	82	698.4228	2	90.74	1394.8318	-0.5	N	SAPCPLAKAALKLL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
29747	10.6	82	484.9543	3	84.04	1451.8357	3.6	N	MTFACKLLAKKA	U ₂₀ -Mri1a
31494	12.3	82	735.4512	2	89.27	1468.8977	-6.6	N	QSGSLLAKAALKLL	U ₁₀ -Mri1c
22556	10.7	82	749.4172	2	70.09	1496.8062	9.2	N	PFLQHALTMLLDV	U ₃ -Mri1a
13997	12.3	82	507.6566	3	52.26	1519.9448	2	N	KLVAPAAAEELLAVVR	U ₁₀ -Mri1c
8754	9.9	82	510.9321	3	40.43	1529.7725	1.3	N	LWEEDDKEVLVR	
24540	11.5	82	514.645	3	73.74	1540.9089	2.7	N	GLRSALAKLEGKLV	U ₁₀ -Mri1b
31517	11.4	82	784.9762	2	87.38	1567.9448	-4.4	N	QAAKDALAKLWKLL	U ₁₀ -Mri1b
26223	13.1	82	528.0037	3	76.7	1580.9978	-5.3	N	QAGGSLLAKAALKLLK	U ₁₀ -Mri1c
37670	13.1	82	796.5092	2	99.2	1591.0183	-9.1	N	NKLGLLKLAAKALAAP	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
20117	11.5	82	802.3663	2	64.15	1602.7026	9.7	Y	MLTLCCHTM(+15.99)HALDV	
21147	13.1	82	579.0104	3	67.55	1734.0112	-0.9	N	EYVGSMLAKAALKLLK	U ₁₀ -Mri1a
36271	14.8	82	874.0851	2	96.32	1746.1494	3.6	N	LKLGLLKLAALKALVGAPA	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
41995	14	82	616.6993	3	106.7	1847.0742	1	N	LGLPLALLMTLPFLQHA	U ₃ -Mri1a
36498	16.4	82	993.6407	2	96.85	1985.2764	-4.8	N	PAAAVLKLLKLAALKALLPQG	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
26505	17.2	82	678.4222	3	77.22	2032.2407	2.1	N	VGGSLLKLVAPAAAEVLARVV	U ₁₀ -Mri1c
12035	16.4	82	683.746	3	48.61	2048.2034	6.2	N	KLVAPAAAEVLANKLFPVPT	U ₁₀ -Mri1c
37671	17.2	82	686.4356	3	99.2	2056.2883	-1.6	N	AAPAVVRVALKLAALKLPGQ	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
37925	17.9	82	686.451	3	99.67	2056.3135	8.6	N	APAAVLKALLKLAALKALVGAP	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
30434	16.5	82	699.3937	3	85.4	2095.1423	8	N	PQGAKKGLFDQLDKAAAPQ	U ₁₃ -Mri1a
28430	17.3	82	538.8259	4	81.07	2151.2964	-10	N	APAAVAKLLAKLAALKALMSKW	U ₁₀ -Mri1a
37233	18.9	82	729.4624	3	98.38	2185.356	4.4	N	EAVAGPVVKLLKLAALKALVGAPA	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
40396	20.6	82	762.4864	3	104	2284.4246	5.6	N	VGVASLAKAALKLLVAAPVAPAAALP	U ₁₀ -Mri1c
32591	16.4	82	596.5857	4	89.24	2382.3027	4.6	Y	M(+15.99)LRTRKQLDKAAAFAFMKLFQ	U ₁₃ -Mri1a
28659	19.7	82	528.5303	5	81.57	2637.6267	-4.4	N	QLKSARTLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
25072	19.7	82	537.5293	5	74.68	2682.6079	0.8	N	LMESLKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
36869	21.3	82	672.1842	4	97.67	2684.7043	1.3	N	KNALLTALKLLKLVAPAAAEVLARVV	U ₁₀ -Mri1c
37002	23	82	678.692	4	97.94	2710.7087	11.2	N	VGSGLLAKVGVKLLKLVAPAAAEVLVVP	U ₁₀ -Mri1c
41792	23	82	690.1826	4	106.4	2756.689	4.5	N	PSGSLAKAALKLLKLVAPAAAEVLANK	U ₁₀ -Mri1c
36295	23	82	706.9519	4	96.87	2823.7676	3.9	N	NKALLVEAAPAVLKLLKLAALKALGQPL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
37323	23.9	82	565.9708	5	98.55	2824.7881	10.5	N	VGGSLLAKAALKLLKLVAPAAAEVKVVPV	U ₁₀ -Mri1c
41863	23.8	82	949.2441	3	106.5	2844.7163	-2.1	N	NTGGSLLAKAALKLLKLVAPAAAEVLANK	U ₁₀ -Mri1c
27102	23.7	82	950.8964	3	78.51	2849.6562	3.9	N	QPPPMLAKAALKLLKVAAPAAAAALAEQ	U ₁₀ -Mri1a
21792	22.3	82	726.897	4	68.28	2903.5579	0.4	N	LSDQPANAKKGLFDQLDKAAAFAFMKLL	U ₁₃ -Mri1a
41925	22.1	82	1010.943	3	106.5	3029.8367	-9.4	N	DLYLLAKRVKLLKLVADLNPVLPALN	
27869	23	82	609.9587	5	79.91	3044.7341	7.5	N	CTTGLMESLKQLTGKEAELLKLLAKKA	U ₂₀ -Mri1a
30274	27.8	82	794.7291	4	85.14	3174.8643	7.3	N	LKDALAAAAGAAAAPAVAALKLFAAAAAALAVKW	
32629	24.6	82	837.9678	4	89.29	3347.8381	1.2	N	AVMESLKQLSAKVEKLLKLLAAQKPCDCYC	U ₂₀ -Mri1a
17341	6.6	82	426.787	2	58.08	851.5593	0.3	N	AVPLLALR	
5413	6.6	82	429.2675	2	29.86	856.5204	-0.1	N	KKLPGLVC	
8181	6.6	82	442.7581	2	38.73	883.5015	0.2	N	AELDPKVL	
5643	6.6	82	454.7512	2	30.56	907.4916	-4.2	N	PFLQHAVP	U ₃ -Mri1a
32813	7.4	82	454.776	2	89.58	907.5378	-0.3	N	LPGDPLLAL	U ₃ -Mri1a

7649	9	82	456.7614	2	37.19	911.5076	0.7	N	ALATGGGAVPV	
34503	7.4	82	457.7917	2	92.95	913.5637	5.7	N	FAGLPLLAL	U ₃ -Mri1a
12271	8.2	82	475.7861	2	48.89	949.5596	-2.1	N	VLKAAPAPS	
12848	8.2	82	500.2714	2	50.03	998.5284	-0.1	N	TPVAAAEVLE	
37287	8.1	81	513.3167	2	98.48	1024.6168	1.9	N	ESGLPLLALL	U ₃ -Mri1a
11427	7.3	81	521.7634	2	46.54	1041.5032	8.8	N	WPQASGLFH	
15651	7.2	81	526.7766	2	55.26	1051.5371	1.5	N	GLMESLKQF	U ₂₀ -Mri1a
15383	7.3	81	537.2929	2	54.79	1072.5706	0.7	N	PFLQHALTF	U ₃ -Mri1a
10021	7.3	81	550.265	2	43.38	1098.5134	1.9	N	WYGGEVKNF	
15912	8.9	81	380.9286	3	55.69	1139.7642	-0.2	N	SLLAKALAKLL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
21812	8.9	81	599.3994	2	68.82	1196.7856	-1.2	N	KELAKAALKLL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
11497	9.7	81	605.3275	2	47.42	1208.6409	-0.3	N	LGPLKPCCPGP	U ₆ -Mri1a
12108	9.8	81	611.8544	2	48.52	1221.6968	-2	N	VVDGAPLDKPV	U ₁₂ -Mri1a
16633	10.6	81	624.8693	2	56.93	1247.7124	9.3	N	LVEAAPAVLPPT	U ₁₀ -Mri1c
12655	10.5	81	447.2577	3	49.66	1338.7395	8.7	N	VTPEVAAALENLL	
25358	11.3	81	684.4513	2	75.16	1366.8911	-2.2	N	AVSGLLAKAALKLL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
36763	10.6	81	685.9382	2	97.43	1369.8518	7.4	N	RGSLLAKAALKRS	U ₁₀ -Mri1c
23617	10.5	81	691.385	2	72.15	1380.75	4	N	SLEVKPLDPGDLV	
17779	10.5	81	705.8885	2	58.85	1409.7554	5	N	LAVHAEPTEALLF	
33177	12.2	81	490.9716	3	90.3	1469.9004	-4.9	N	GSLLAALKLGLCLL	U ₁₀ -Mri1c
9076	12.9	81	497.6367	3	41.2	1489.8867	1	N	KLVAAPAAEVLALPGA	U ₁₀ -Mri1c
9827	12.2	81	503.3124	3	42.96	1506.9133	1.4	N	KLVAAPAAEVLAKVQ	U ₁₀ -Mri1c
21791	11.3	81	505.9834	3	68.28	1514.9185	6.5	N	RTALKAAALFDLL	
11181	12.2	81	513.9648	3	46.23	1538.8667	3.8	N	RSALPTVAAALEDLL	U ₁₀ -Mri1c
31466	11.3	81	523.6517	3	87.28	1567.9448	-7.3	N	QAAKDALAKLWKLL	U ₁₀ -Mri1b
37517	12.2	81	789.5012	2	98.91	1577.0027	-9.4	N	NKLLLAKAALLNAP	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
27960	13	81	789.9666	2	80.05	1577.9326	-8.9	N	PAAQQLLAKAALPC	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
42022	11.3	81	790.4496	2	106.7	1578.8911	-4.1	Y	M(+15.99)TM(+15.99)MLLAKAALKLL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
24975	12.1	81	542.3094	3	74.52	1623.9092	-1.7	N	SMLTSMMLAKAALKLF	U ₁₀ -Mri1a
19773	13	81	573.6794	3	63.02	1718.0198	-2	Y	M(+15.99)ALTSMLAKAALKLLK	U ₁₀ -Mri1a
18203	14.6	81	597.7102	3	60.11	1790.1028	3.4	N	LLKLVAPAAAEVLANKLG	U ₁₀ -Mri1c
12990	12.2	81	449.9973	4	50.26	1795.958	1.1	N	SHLDLDRYPLDRLA	
13854	13.8	81	472.0482	4	51.99	1884.1594	2.2	N	ESLLRSMMLAKAALKLLK	U ₁₀ -Mri1a
19683	14.6	81	655.7418	3	62.89	1964.2034	0.1	N	AKAPVDEELLKLLAKKA	U ₂₀ -Mri1a
19603	15.4	81	659.096	3	62.77	1974.2605	2.9	N	LKLLKLVAPAAAEVLALKN	U ₁₀ -Mri1c
19718	16.3	81	512.0857	4	62.95	2044.3135	0.1	N	LAKLLKLVAPAAEVLARVV	U ₁₀ -Mri1c
37232	16.3	81	1025.627	2	98.38	2049.2236	7.3	N	LPLGSELALFKKATPAAPLL	
36330	17.1	81	686.4387	3	96.47	2056.3135	-9.2	N	VGVTLAKAALKLGLLALLVH	U ₁₀ -Mri1c
35439	16.3	81	543.5672	4	94.52	2170.2571	-8	N	DGSLKQLSAKAEELLKLL	U ₂₀ -Mri1a
28594	17.9	81	488.9044	5	81.41	2439.4788	2.9	N	LSEKNSAKAEELLKLLAKKA	U ₂₀ -Mri1a
28505	17.8	81	502.3054	5	81.24	2506.4668	9.5	N	EMHKQLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
42086	19.3	81	1309.229	2	106.8	2616.4568	-4.8	N	LPGECLPLALLMTLMLLQHAQQA	U ₃ -Mri1a
36437	21.8	81	674.1699	4	96.71	2692.6616	-4.1	N	VGGSLLAKAALKLLKLVAPAAEVVYK	U ₁₀ -Mri1c
36622	21.8	81	674.4127	4	97.13	2693.6279	-2.3	N	VGGTVLAKAALKLLVLAPAAAEVLMF	U ₁₀ -Mri1c
28298	22	81	681.413	4	80.79	2721.6089	5.1	N	DKKAAAAPAVAKALLKLAALKMSKW	U ₁₀ -Mri1a
29163	20.2	81	392.3832	7	82.75	2739.637	-2	N	LLMVEMGLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
28481	19.5	81	392.5258	7	81.19	2740.636	-2.3	N	KLQSEMRKSAKAEELLKLLAKKA	U ₂₀ -Mri1a
36131	22.6	81	1421.392	2	95.97	2840.7803	-3.9	N	RRTLAKAALQLLKLVAAAPAQVLKGGAL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
28426	23.5	81	718.1878	4	81.06	2868.6887	11.6	N	PAAAAHKAAPAVAKALLKLAALKMSKW	U ₁₀ -Mri1a
27451	21.1	81	580.1536	5	79.16	2895.7305	0.4	N	GLMESLKQLSAKAEELLKLLAKKRA	U ₂₀ -Mri1a
29278	20.9	81	488.7992	6	83	2926.7629	-3.8	N	QLKHACHKLSAKAEELLKLLAKKA	U ₂₀ -Mri1a
36474	23.5	81	980.5971	3	96.79	2938.7944	-8.5	N	DLKGLAKAALKLLKLVAPAAAEVLANPQL	U ₁₀ -Mri1c
29053	23.5	81	983.9131	3	82.47	2948.7439	-8.9	N	KNNGLHAKAALKLLKAVAPAAAAALWLQP	U ₁₀ -Mri1a
35823	24.2	81	1001.956	3	95.18	3002.8372	2.7	N	VGGSLLAKAALKLLKLVAPAAAEVLANFRV	U ₁₀ -Mri1c
40941	24.2	81	1036.988	3	105.2	3107.916	8.6	N	KDLGLAKAALKLLKLVAPAAAEVLGEKHLK	U ₁₀ -Mri1c
41847	23.6	81	1042.316	3	106.4	3123.9109	5.3	N	DPSLLAKAALKLLKLVANVPEVLALQLRQ	U ₁₀ -Mri1c
24424	25	81	818.9393	4	73.55	3271.7273	0.3	N	EPAAEADKAPQGKGLGLFDQLDKAAAFMKL	U ₁₃ -Mri1a
9327	7.3	81	411.7447	2	41.76	821.4759	-1.2	N	AAAALHVVA	
17630	6.5	81	417.2664	2	58.6	832.517	1.6	N	GLPLLALH	U ₃ -Mri1a
14767	7.3	81	437.7477	2	54.07	873.4807	0.1	N	LDSAALATL	U ₁₉ -Mri1a
17770	6.5	81	458.7777	2	58.83	915.5436	-3.1	N	CGARRLLK	
8579	6.5	81	470.8026	2	39.95	939.5905	0.1	N	AVPLKKVW	
34960	7.3	81	478.7957	2	93.73	955.5776	-0.6	Y	VAPLLALLM(+15.99)	U ₃ -Mri1a
10111	6.5	81	498.267	2	43.64	994.5157	3.7	N	LMESLKQF	U ₂₀ -Mri1a
7348	7.3	81	499.2816	2	36.27	996.5491	-0.6	N	LPDPKVLES	U ₁₂ -Mri1a
9270	8.8	80	503.3112	2	41.63	1004.6018	6.1	N	KLVAAPAAAPP	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
16977	7.2	80	510.2762	2	57.47	1018.5369	1	N	GLMESLCEL	U ₂₀ -Mri1a
5487	7.2	80	511.774	2	31.02	1021.5345	-1	N	PFLQHAVPN	U ₃ -Mri1a

6381	8	80	549.2908	2	34.28	1096.5664	0.6	N	PFLQHALTNG	U ₃ -Mri1a
13192	11.2	80	617.8704	2	50.68	1233.7192	5.7	N	LGAPAAAAALGRP	U ₁₀ -Mri1a
13167	11.1	80	626.3747	2	50.63	1250.7346	0.2	N	GLAPAAAAALALNK	U ₁₀ -Mri1a
19457	8.8	80	627.337	2	62.54	1252.6638	-3.4	N	PFLQHALTMVP	U ₃ -Mri1a
20484	8.8	80	627.3531	2	66.33	1252.6816	8.1	N	PFLQHALTLDV	U ₃ -Mri1a
14095	8.8	80	627.8455	2	53.58	1253.6768	-0.2	N	LTPFLQHALTN	U ₃ -Mri1a
17787	8.8	80	627.8417	2	58.86	1253.6768	-6.3	N	PFLQHALTNL	U ₃ -Mri1a
22034	10.5	80	649.4125	2	68.87	1296.813	-2	N	VGSLAKAALKLN	U ₁₀ -Mri1c
14026	11.3	80	669.3553	2	53.37	1336.6987	-1.9	N	LDSAALATHAASLP	U ₁₉ -Mri1a
9109	12.1	80	681.9083	2	41.26	1361.803	-0.6	N	KLGAAPAAAAALPQ	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
30178	11.2	80	690.4334	2	84.96	1378.8547	-1.8	N	PGDGLAKAALKLL	U ₁₀ -Mri1c
11275	10.4	80	710.896	2	47.19	1419.7722	3.7	N	VLHATTSHTLVLE	
22342	10.4	80	740.902	2	69.61	1479.7908	-0.8	N	PFLQHALTNLPVM	U ₃ -Mri1a
25405	11.2	80	495.6567	3	75.23	1483.949	-0.4	N	VGPKYLAKAALKLL	U ₁₀ -Mri1a/ U ₁₀ -Mri1c
18302	10.4	80	745.8246	2	60.99	1489.636	-0.9	Y	TMLPM(+15.99)M(+15.99)M(+15.99)CSALT	
22138	11.2	80	765.3831	2	69.13	1528.7507	0.6	N	LESEAAELETDP	
32996	12	80	514.9934	3	89.87	1541.9578	0.5	Y	M(+15.99)VVGSLAKAALKLL	U ₁₀ -Mri1c
13848	12	80	810.0093	2	51.98	1618.0002	2.3	N	KGFSMLAKAALKLL	U ₁₀ -Mri1a
15181	14.4	80	466.3032	4	54.47	1861.1763	4	N	LKLLGLKAPAAEVLANK	U ₁₀ -Mri1c
18290	12.7	80	932.0182	2	60.34	1862.0012	11.1	N	VTFLELQHALTMSFL	U ₃ -Mri1a
42105	17.5	80	782.4635	3	106.8	2344.3591	4.1	N	LPLAVGLALLMTLPFLQHALVD	U ₃ -Mri1a
42130	17.6	80	787.1362	3	106.9	2358.3748	5	N	LPLALGLALLMTLPFLQHALVD	U ₃ -Mri1a
29144	17.6	80	488.9004	5	82.98	2439.4788	-5.4	N	SLQKQDSAKAEELLKLLAKKA	U ₂₀ -Mri1a
28640	18.3	80	511.3092	5	82.46	2551.5034	2.5	N	MEAKAPALSAKQWQLKLLAKKA	U ₂₀ -Mri1a
17765	18.4	80	519.0859	5	58.82	2590.394	-0.4	N	DKPRLLGLDKFDQDKAAAFAFMK	U ₁₃ -Mri1a
27820	21.7	80	873.1839	3	79.84	2616.5398	-3.8	N	DKALDLAAAAAPAVAKLLKLAALKALPC	U ₁₀ -Mri1a
25684	20.1	80	561.7057	5	76.72	2803.4875	1.7	N	TNQPRPMKGLGFDQDKAAAFAFMK	U ₁₃ -Mri1a
41184	23.3	80	569.3648	5	105.4	2841.7781	3.4	N	VGSGLLAKAALKLLKVAAPAAEVLAKVQ	U ₁₀ -Mri1c
41846	22.4	80	715.4533	4	106.4	2857.7554	10.1	Y	KNALLVEAAVVLLKLVAAAPAAAM(+15.99)VLAVVR	U ₁₀ -Mri1c
36593	21.5	80	1470.894	2	97.07	2939.8235	-17	N	RRERKAALKLLKLVAPAAAEVLASRLP	U ₁₀ -Mri1c
27165	24	80	752.2055	4	78.63	3004.7986	-1.9	N	GFKSMLAKAALKLLKVALHAAAAAPT	U ₁₀ -Mri1a
41889	24.7	80	1007.93	3	106.5	3020.8364	-23	N	HTGSLAVVSLLLKLVAPAAAEVLGGKLP	U ₁₀ -Mri1c
32920	24.1	80	1016.271	3	89.75	3045.7913	0	N	LKGDALAKLGRLLKAEVPTVAAALNKVG	U ₁₀ -Mri1b
32073	24	80	795.9706	4	88.38	3179.8313	7	N	PSANAGLMESLKQLASKAEELLKLLAKKA	U ₂₀ -Mri1a
30769	23.2	80	815.1953	4	86.04	3256.7542	-0.7	N	PQGAKKLGFDQDKAAAFAFMKLEFLRNH	U ₁₃ -Mri1a
38887	26.3	80	1125.371	3	101.4	3373.095	-1.5	N	VGSGLLAKAALKLLKVAAPAAEVLANKLQKF	U ₁₀ -Mri1c
29204	7.2	80	426.7814	2	82.84	851.548	0.2	N	VGAVPLLAL	U ₃ -Mri1a
8006	6.4	80	451.2971	2	39.32	900.5797	-0.1	N	LPVAKKVF	
15603	8	80	500.7789	2	55.16	999.5423	1	N	LECPGLKAGL	

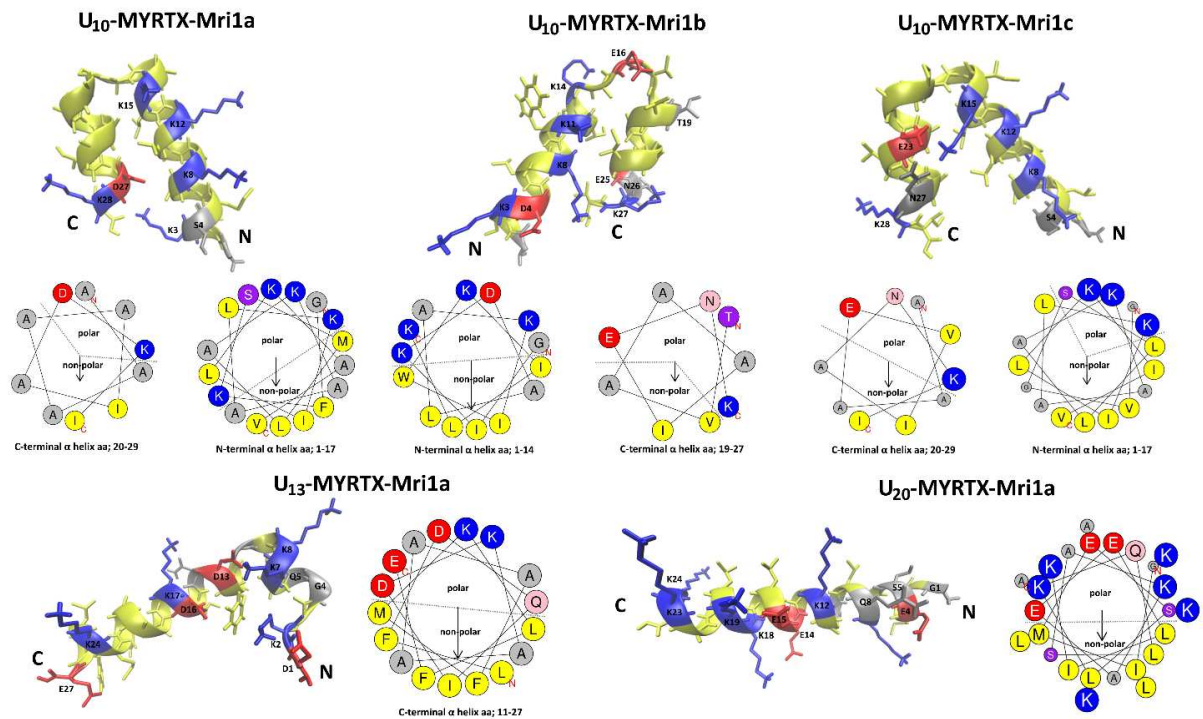
Table S4. Identified peptide toxin transcripts expressed in the venom transcriptome of *Manica rubida*.

Toxin	Contig_name	TPM
U ₁₀ -Myrmecitoxin-Mri1a	CL4Contig1_1	193,522
U ₁₀ -Myrmecitoxin-Mri1c	TRINITY_DN10134_c4_g1_i2_1	143,717
	TRINITY_DN88_c0_g1_i1_1	55,388
U ₃ -Myrmecitoxin-Mri1a	TRINITY_DN9765_c0_g4_i2_1	21,514
	TRINITY_DN2643_c0_g2_i1_1	1,937
	CL82Contig3_1	93
U ₂₀ -Myrmecitoxin-Mri1a	TRINITY_DN9416_c1_g1_i1_1	29,959
	CL1Contig199_2	4,881
	TRINITY_DN9416_c1_g1_i2_1	2,626
	CL33Contig2_1	2,818
U ₁₃ -Myrmecitoxin-Mri1a	TRINITY_DN10038_c6_g3_i3_1	5,595
	TRINITY_DN10038_c6_g3_i6_1	5,771
U ₁₈ -Myrmecitoxin-Mri1a	TRINITY_DN9460_c2_g2_i1_1	9,359
	TRINITY_DN10131_c4_g16_i3_1	3,380
U ₁₂ -Myrmecitoxin-Mri1a	CL8Contig5_1	4,809
	CL8Contig4_1	734
U ₁₀ -Myrmecitoxin-Mri1b	CL1636Contig1_2	7,623
U ₆ -Myrmecitoxin-Mri1a	TRINITY_DN4219_c0_g1_i1_1	1,678
	TRINITY_DN4685_c0_g1_i1_1	3,618
U ₁₉ -Myrmecitoxin-Mri1a	TRINITY_DN3084_c0_g1_i5_1	4,380
	CL291Contig1_1	264
	TRINITY_DN10022_c2_g12_i11_1	905
U ₁₇ -Myrmecitoxin-Mri1a	TRINITY_DN10022_c2_g12_i2_1	1,144
	TRINITY_DN10022_c2_g12_i14_1	971
	TRINITY_DN10022_c2_g12_i4_1	683
	TRINITY_DN10022_c2_g12_i1_1	390
	CL3004Contig1_1	337
U ₁₇ -Myrmecitoxin-Mri1b	CL1211Contig1_1	168
	TRINITY_DN10022_c2_g12_i10_1	217
U ₁₂ -Myrmecitoxin-Mri1b	CL899Contig1_1	178

Table S5. Insecticidal activity of *Manica rubida* venom fractions on blowflies 5 min, 1 h and 24 h post injection. Each fraction was injected into three flies.

Fraction	Peptide	5 min	1 h	24 h
F 26	Fragment (U ₁₃ -MYRTX-Mri1a)	PPP	PPP	DDD
F 30	U ₂₀ -MYRTX-Mri1a	PPP	PPP	DDD
F 31	U ₂₀ -MYRTX-Mri1a, U ₁₃ -MYRTX-Mri1a	PPP	PPP	DPN
F 32	U ₁₀ -MYRTX-Mri1a, U ₁₃ -MYRTX-Mri1a	PPP	PPP	DDD
F 33	U ₁₀ -MYRTX-Mri1a	NNN	PPP	DDD
F 34	U ₁₀ -MYRTX-Mri1a	PPP	PPP	DDD
F 35	U ₁₀ -MYRTX-Mri1a, U ₁₀ -MYRTX-Mri1b	PPP	PPP	DDD
F 36	U ₁₀ -MYRTX-Mri1b, U ₁₀ -MYRTX-Mri1c	NPP	PPP	DDD
F 37	U ₁₀ -MYRTX-Mri1c	PPN	PPN	DDD
F 38	U ₁₀ -MYRTX-Mri1c	PPP	PPP	DDD
F 39	U ₁₀ -MYRTX-Mri1c	PPP	PPP	DDD
F 40	U ₁₀ -MYRTX-Mri1c	PPP	PPP	DDD
F 41	U ₁₀ -MYRTX-Mri1c	PPP	PPP	DDD
F 43	U ₃ -MYRTX-Mri1a, unidentified protein (MW >10KDa)	PPP	PPP	DDD

N = normal, P = paralyzed, D = dead



Supplementary Figure S2. *De novo* structure prediction of insecticidal myrmecitoxins of *M. rubida* venom. These peptides are thought to be dominated by an amphiphilic and polycationic α -helix. Secondary structure prediction was performed on the PepFold3 server.

Supporting Information for Publication

Title of Paper

Myrmicine ant venoms: first insights into their peptidic diversity.

Authors:

Valentine Barassé^{1*}, Nathan Téné¹, Christophe Klopp², Niklas Tysklind⁴, Hadrien Lalague⁴, Jérôme Orivel⁴, Valérie Troispoux⁴, Frédérick Petitclerc⁴, Martin Kenne³, Maurice Tindo³, Michel Treilhou¹, Elsa Bonnafé^{1#}, Axel Touchard^{4#}

Affiliations:

- 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, INRA Toulouse, Castanet-Tolosan, France ; christophe.klopp@inra.fr (C.K.)
- ³ 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.)
- ⁴ CNRS, UMR EcoFoG, AgroParisTech, Cirad, INRA, Université des Antilles, Université de Guyane, 97310 Kourou, France; axel.touchard2@gmail.com (A.T) ; niklas.tysklind@ecofog.gf (N.T.) ; hadrien.lalague@ecofog.gf (H.L.) ; jerome.orivel@ecofog.gf (J.O.) ; valerie.troispoux@ecofog.gf (V.T.) ; frederick.petitclerc@ecofog.gf (F.P.)

The authors contributed equally to this study.

Corresponding author:

- **Valentine Barassé:** EA-7417, Institut National Universitaire Champollion, Place de Verdun, 81012 Albi, France. Phone: +(33)5 63 48 17 00. Email: valentine.barasse@gmail.com

Table of Contents

Figure S1: Total ion chromatogram (TIC) of crude venom and venom after hexanic extraction from *Solenopsis saevissima*. S17

Supplementary Table I: Source and composition of ant venom samples used for proteomic analyses.	S3
Supplementary Table II: Source and composition of ant tissue samples used for transcriptomic analyses.	S3
Supplementary Table III : Peptide mass fingerprinting of <i>Tetramorium africanum</i> venom.	S4
Supplementary Table IV : Peptide sequences in the venom of <i>Tetramorium africanum</i> .	S6
Supplementary Table V : Peptide mass fingerprinting of <i>Stenamma debile</i> venom.	S7
Supplementary Table VI : Peptide sequences in the venom of <i>Stenamma debile</i> .	S8
Supplementary Table VII : Peptide mass fingerprinting of <i>Daceton armigerum</i> crude venom.	S8
Supplementary Table VIII : Peptide sequences in the venom of <i>Daceton armigerum</i> .	S9
Supplementary Table IX : Peptide mass fingerprinting of <i>Pogonomyrmex californicus</i> venom.	S9
Supplementary Table X : Peptide sequences in the venom of <i>Pogonomyrmex californicus</i> .	S11
Supplementary Table XI : Peptide mass fingerprinting of <i>Myrmica ruginodis</i> venom.	S12
Supplementary Table XII : Peptide sequences in the venom of <i>Myrmica ruginodis</i> .	S13
Supplementary Table XIII : Mass fingerprinting of <i>Solenopsis saevissima</i> venom after hexanic extraction.	S15
Supplementary Table XIV : Venom glands RNA sequencing data.	S16
Supplementary Table XV: Addressing table of major contigs expressed by <i>Myrmica ruginodis</i> venom glands.	S18
Supplementary Table XVI : Addressing table of major contigs expressed by <i>Tetramorium africanum</i> venom glands.	S23
Supplementary Table XVII : Addressing table of major contigs expressed by <i>Pogonomyrmex californicus</i> venom glands.	S26
Supplementary Table XVIII : Addressing table of major contigs expressed in <i>Daceton armigerum</i> venom glands.	S38

Supplementary Table I: Source and composition of ant venom samples used for proteomic analyses.

Species	Collection sites	# Venom reservoirs	Proteomic analysis
<i>Tetramorium africanum</i>	Manengole, Cameroon	52 venom sacs	Mass profiling
		50 venom sacs	MS/MS
<i>Daceton armigerum</i>	Kourou, French Guyana	22 venom sacs	Mass profiling
		24 venom sacs	MS/MS
<i>Stenamma debile</i>	Le Grez, France	22 venom sacs	Mass profiling
		12 venom sacs	MS/MS
<i>Pogonomyrmex californicus</i>	Tempe, Arizona (USA)	10 venom sacs	Mass profiling
		11 venom sacs	MS/MS
<i>Solenopsis saevissima</i>	Kourou, French Guyana	31 venom sacs	Mass profiling
		32 venom sacs	MS/MS
<i>Myrmica ruginodis</i>	Lacaune, France	28 venom sacs	Mass profiling
		25 venom sacs	MS/MS

Supplementary Table II: Source and composition of ant tissue samples used for transcriptomic analyses.

Species	Sample composition
<i>Myrmica ruginodis</i>	100 venom apparatus
<i>Pogonomyrmex californicus</i>	70 venom apparatus
<i>Stenamma debile</i>	10 ants
<i>Solenopsis saevissima</i>	130 venom apparatus
<i>Daceton armigerum</i>	91 venom apparatus
<i>Tetramorium aculeatum</i>	100 venom apparatus
<i>Tetramorium africanum</i>	100 venom apparatus

Supplementary Table III : Peptide mass fingerprinting of *Tetramorium africanum* venom. List of peptide masses detected through LC–MS using an LCQ-ion trap Advantage mass spectrometer.

Retention Time (min)	Mass (Da)	Relative abundance (%)	Temporary Name	Peptide Toxin
1.84	2753.19	0.14	Ta-2753	
2.3	2120.50	0.46	Ta-2121	
2.33	1735.84	0.08	Ta-1736	
11.89	788.41	0.18	Ta-788	U ₂ -MYRTX-Ta1a (cleaved)
11.89	1000.49	0.31	Ta-1001	U ₂ -MYRTX-Ta1a
14.3	757.94	0.04	Ta-758	
14.66	2070.07	0.01	Ta-2070	U ₇ -MYRTX-Ta1a
15.56	1634.90	0.34	Ta-1635	U ₆ -MYRTX-Ta1a
15.56	3119.13	0.16	Ta-3119	
16.11	1367.77	0.18	Ta-1368	U ₂₄ -MYRTX-Ta1a
16.51	1557.86	0.32	Ta-1558	U ₃ -MYRTX-Ta1a (cleaved)
17.84	1979.95	0.69	Ta-1980	U ₇ -MYRTX-Ta1b
17.84	2362.08	0.26	Ta-2362	
18.3	2640.35	0.58	Ta-2640	U ₁₇ -MYRTX-Ta1a
18.4	1397.65	0.85	Ta-1398	U ₂₂ -MYRTX-Ta1a
19.31	1431.25	0.47	Ta-1431	
19.74	2639.38	0.03	Ta-2639	U ₁₇ -MYRTX-Ta1d
20.14	2768.56	0.29	Ta-2769	U ₁₇ -MYRTX-Ta1b
20.91	3030.61	0.31	Ta-3031	U ₁₇ -MYRTX-Ta1c
21.44	2839.83	0.23	Ta-2840	
22.11	921.55	0.77	Ta-922	U ₁₄ -MYRTX-Ta1a
23.16	2708.43	0.11	Ta-2708	U ₂₅ -MYRTX-Ta1a
23.16	4017.12	4.36	Ta-4017	U ₁₁ -MYRTX-Ta1a
24.39	1934.01	0.22	Ta-1934	U ₃ -MYRTX-Ta1b (cleaved)

24.68	1816.03	0.13	Ta-1816	U ₁ -MYRTX-Ta1a (oxidized)
26.21	1948.01	0.18	Ta-1948	U ₃ -MYRTX-Ta1a (cleaved)
26.86	2359.36	23.13	Ta-2359	M-MYRTX-Ta1a
27.13	1800.03	12.32	Ta-1800	U ₁ -MYRTX-Ta1a
28.42	2303.18	0.16	Ta-2303	U ₂₁ -MYRTX-Ta1b (cleaved)
29.74	743.46	1.91	Ta-744	U ₃ -MYRTX-Ta1b (cleaved)
30.62	1164.71	0.07	Ta-1165	U ₂₆ -MYRTX-Ta1a
30.85	2791.39	0.07	Ta-2791	U ₂₁ -MYRTX-Ta1b (oxidized)
31.47	2431.31	0.27	Ta-2431	U ₃ -MYRTX-Ta1a (cleaved)
32.48	2531.38	2.22	Ta-2531	U ₃ -MYRTX-Ta1b (cleaved)
32.7	3032.53	7.62	Ta-3033	U ₂₁ -MYRTX-Ta1b (cleaved)
33.54	2544.39	2.31	Ta-2544	U ₃ -MYRTX-Ta1a (cleaved)
33.68	2775.39	6.88	Ta-2775	U ₂₁ -MYRTX-Ta1b
33.91	2867.19	0.27	Ta-2867	
34.26	2106.05	0.05	Ta-2106	U ₂₁ -MYRTX-Ta1a (cleaved)
35.45	856.54	0.39	Ta-857	U ₃ -MYRTX-Ta1b (cleaved)
35.57	2462.22	0.40	Ta-2462	U ₂₁ -MYRTX-Ta1a (cleaved)
36.24	3299.79	2.29	Ta-3300	U ₃ -MYRTX-Ta1c
36.61	2849.41	12.92	Ta-2849	U ₂₁ -MYRTX-Ta1a
37.18	3337.84	1.83	Ta-3338	U ₃ -MYRTX-Ta1d
37.62	2251.11	0.21	Ta-2251	U ₂₃ -MYRTX-Ta1a
39.22	969.62	0.88	Ta-970	U ₃ -MYRTX-Ta1a (cleaved)
40.05	1273.84	0.67	Ta-1274	U ₃ -MYRTX-Ta1d (cleaved)
45.37	3255.83	4.38	Ta-3256	U ₃ -MYRTX-Ta1b
46.24	3269.84	7.05	Ta-3270	U ₃ -MYRTX-Ta1a

Supplementary Table IV : Peptide sequences in the venom of *Tetramorium africanum*. PTMs. post-translational modifications. TPMs. transcripts per millions. “*” denotes C-terminal amidation. “ξ” indicates the peptide was not validated through proteomics. “p” The TPM value is the sum of the frequency from more than one assembled contig encoding the same transcript.

Toxin name	Mass (Da)	TPMs	Sequence	PTMs
M-MYRTX-Ta1a	2359.36	46,405.86	FKIPWGKIKDFVTGGIKEVAK*	
U1-MYRTX-Ta1a	1800.03	149,799.21	GLKEIWEKIKQKLG	
U2-MYRTX-Ta1a	1000.49	113,947.14	DPPGGFVGTR*	
U3-MYRTX-Ta1a	3269.84	9,082.74	LAPIFALLLSGLFSLPALQHYIEKKNYIN*	
U3-MYRTX-Ta1b	3255.83	-	LAPIFALLLSGLFSLPALQHYVEKKNYIN*	
U3-MYRTX-Ta1c	3299.79	11,100.69	IAPILALGLLSAFSAIPFIHHWATGGKPHHE*	
U3-MYRTX-Ta1d	3337.84	55,025.36	IAPILALPLLGLFSLPFVHHWATGGKPHHE*	
U4-MYRTX-Ta1a	1500.76 ξ	2,715.685	GCSVNRRRQGLCR*	1 S-S
U6-MYRTX-Ta1a	1634.90	1,792.98	FHGPCPKIPGKIIC	1 S-S
U7-MYRTX-Ta1a	2070.07	1,937.06	AINCRRFPFHPKCRGISA	1 S-S
U7-MYRTX-Ta1b	1979.95	4,031.09	DVNCEITPFHPKCRGVAP	1 S-S
U11-MYRTX-Ta1a	4017.12	7,902.77	GKEKDKLIECTKEMLLAAMDYAKHKIEKHLFKCK	1 S-S
U13-MYRTX-Ta1a	2488.25 ξ	4,469.8	SKIGLFDQIDKGMAWFMDLKF*	
U14-MYRTX-Ta1a	921.55	437.82	IPPAALASLA*	
U17-MYRTX-Ta1a	2640.35	1,046.62	GIINAPDRCPDGYKRAGNQCRKVI*	1 S-S
U17-MYRTX-Ta1b	2768.56	4,805.32	YIIRVPIQCPPGKVKVGNRCRIVY	1 S-S
U17-MYRTX-Ta1c	3030.61	4,352.09	GIIRIPELKCPDGYKKDALGVCREIFT*	1 S-S
U17-MYRTX-Ta1d	2639.38	8,877.64	SVINVPIQCPPGTIQVGKRCRETF	1 S-S
U17-MYRTX-Ta1e	2716.54 ξ	1,508.21	NIKVPCRAGYRKINGVCRKIYR*	1 S-S
U21-MYRTX-Ta1a	2849.41	6,704.71	VVSMDDINKWAQEMLSKLNELMKQ	
U21-MYRTX-Ta1b	2775.39	-	VVSMDDINKWAQEALSKLNEVMKQ	
U22-MYRTX-Ta1a	1397.65	42,166.97	HVGLCHFKICNM*	1 S-S

U ₂₃ -MYRTX-Ta1a	2251.11	3,921.32	KDWVEICEILPPICDDVGPI	1 S-S
U ₂₄ -MYRTX-Ta1a	1367.77	678.29	TQSKNPDVVIRL*	
U ₂₅ -MYRTX-Ta1a	2708.43	2,351.59	WRVPWRDIITKGSKMAYEMSKK*	
U ₂₆ -MYRTX-Ta1a	1164.71	337.17	QLPTIFVPRL*	Pyro-Q
U ₂₇ -MYRTX-Ta1a	2228.97 ξ	3,923.8	WREICKMFPYYEGCHRD*	1 S-S

Supplementary Table V : Peptide mass fingerprinting of *Stenamma debile* venom. List of peptide masses detected through LC-MS using an LCQ-ion trap Advantage mass spectrometer.

Retention time (min)	Mass (Da)	Relative abundance (%)	Temporary name	Peptide toxin
7.06	3050.49	4.00	Sd-3051	U ₁₇ -MYRTX-Sd1b
12.01	3021.59	8.17	Sd-3022	U ₁₇ -MYRTX-Sd1a
12.97	932.57	4.08	Sd-933	
15.3	744.32	0.34	Sd-744	
16.62	682.1	0.63	Sd-682	
16.64	2665.35	0.24	Sd-2666	
18.93	1021.47	9.25	Sd-1022	
18.93	3179.59	16.57	Sd-3180	U ₂₉ -MYRTX-Sd1a
21.09	2623.99	0.44	Sd-2624	
22.32	895.21	0.20	Sd-895	
24.05	2606.46	0.86	Sd-2607	U ₃₁ -MYRTX-Sd1a
27.68	879.23	1.69	Sd-879	
28.51	1810.94	4.46	Sd-1811	U ₂₈ -MYRTX-Sd1a (Oxidized)
29.53	786.28	0.94	Sd-787	
31.75	1519.88	7.05	Sd-1520	U ₃₀ -MYRTX-Sd1a
32.29	2513.14	1.74	Sd-2492	

32.47	1793.94	3.55	Sd-1794	U ₂₈ -MYRTX-Sd1a (Pyro-Q/Oxidized)
33.41	863.15	9.79	Sd-863	U ₂₈ -MYRTX-Sd1a (cleaved/Pyro-Q)
35.79	1794.94	18.85	Sd-1795	U ₂₈ -MYRTX-Sd1a
39.77	1777.94	7.12	Sd-1778	U ₂₈ -MYRTX-Sd1a (Pyro-Q)

Supplementary Table VI : Peptide sequences in the venom of *Stenamma debile*. PTMs. post-translational modifications. TPMs. transcripts per millions (sum of more than one transcript). "*" C-terminal amidation.

Toxin name	Mass (Da)	Sequence	PTMs
U ₁₇ -MYRTX-Sd1b	3050.49	YIIDAPSRKCPGSRSTQGE CRTTSR*	1 S-S
U ₁₇ -MYRTX-Sd1a	3021.59	NIITVPERPCPSGQRKDSRGKCRQVLS	1 S-S
U ₂₈ -MYRTX-Sd1a	1793.93	QFPMDMLIAGAKKLF*	
U ₂₉ -MYRTX-Sd1a	3179.60	ETALATAEATPEATAEATPKATAKAYQPIY	
U ₃₀ -MYRTX-Sd1a	1519.88	TISYLRGLLPLFQ	
U ₃₁ -MYRTX-Sd1a	2606.46	ANIALEAANKMGPRIAEKLVKQLQ	

Supplementary Table VII : Peptide mass fingerprinting of *Daceton armigerum* crude venom. List of peptide masses detected through LC-MS using an LCQ-ion trap Advantage mass spectrometer.

Retention time (min)	Mass (Da)	Relative abundance (%)	Temporary name	Peptide toxin
17.01	8251.52	8.53	Da-8252	
17.57	8230.54	26.08	Da-8231	U ₃₂ -MYRTX-Da1b
17.88	8219.58	20.84	Da-8220	U ₃₂ -MYRTX-Da1a/b
18.28	8208.63	14.77	Da-8209	U ₃₂ -MYRTX-Da1a
21.54	9415.76	7.59	Da-9416	
21.87	9161.33	22.19	Da-9161	U ₃₂ -MYRTX-Da1c

Supplementary Table VIII : Peptide sequences in the venom of *Daceton armigerum*. PTMs. post-translational modifications. TPMs. transcripts per millions. “p” The TPM value is the sum of the frequency from more than one assembled contig encoding the same transcript.

Toxin name	Mass (Da)	TPMs	Sequence	Features
U ₃₂ -MYRTX-Da1a	8208.63	141,219.69 ^p	GKEKEAFKAQLKECVKAGAKYLSHKVSKALYALIDKI	Homodimer
U ₃₂ -MYRTX-Da1b	8230.54	7,442.10	GKEKEAFKAQLRECVKAGAKYLSHKLSKAMNALIDKI	Homodimer
U ₃₂ -MYRTX-Da1c	9161.33	6,313.82	GKNKEALKAQLKECVKAVEKYITDKISKKVLQAINALIDKI	Homodimer
U ₃₂ -MYRTX-Da1a/b	8219.58	141,219.69 ^p 7,442.10	GKEKEAFKAQLKECVKAGAKYLSHKVSKALYALIDKI GKEKEAFKAQLRECVKAGAKYLSHKLSKAMNALIDKI	Heterodimer

Supplementary Table IX : Peptide mass fingerprinting of *Pogonomyrmex californicus* venom. List of peptide masses detected through LC-MS using an LCQ-ion trap Advantage mass spectrometer.

Retention time (min)	Mass (Da)	Relative abundance (%)	Temporary name	Peptide toxin
11.77	2489.40	0.16	Pc-2489	U ₁₇ -MYRTX-Pc1c
12.31	2514.17	0.12	Pc-2516	U ₁₇ -MYRTX-Pc1c (+ 26 Da)
12.82	3292.92	0.04	Pc-3292	U ₁₇ -MYRTX-Pc1b (Oxidized)
12.94	2112.12	0.02	Pc-2112	
13.41	3317.53	0.03	Pc-3319	U ₁₇ -MYRTX-Pc1b (Oxidized; +26 Da)
14	3275.53	0.91	Pc-3278	U ₁₇ -MYRTX-Pc1b
14.55	3301.53	0.68	Pc-3303	U ₁₇ -MYRTX-Pc1b (+26 Da)
15.16	3562.92	0.23	Pc-3563	U ₁₇ -MYRTX-Pc1a
15.71	3588.92	0.07	Pc-3591	U ₁₇ -MYRTX-Pc1a (+26 Da)
16.62	2053.08	0.18	Pc-2053	
19.15	5449.6	0.11	Pc-5450	
19.15	6086.44	0.27	Pc-6086	

19.66	5633.46	0.09	Pc-5636	U ₁₈ -MYRTX-Pc1a
19.66	5676.72	0.08	Pc-5677	
19.91	6070.6	0.36	Pc-6071	
20.21	5660.36	0.30	Pc-5660	
22.94	1825.14	0.03	Pc-1826	
23.66	1946.88	0.09	Pc-1947	
24.12	1980.02	0.07	Pc-1981	U ₃ -MYRTX-Pc1a (cleaved)
28.7	3249.90	1.75	Pc-3249	
29.37	3109.80	3.79	Pc-3110	U ₃₅ -MYRTX-Pc1c
30.29	3260.03	0.42	Pc-3262	U ₃₄ -MYRTX-Pc1a (cleaved)
32.2	3464,80	1.51	Pc-3468	U ₃₅ -MYRTX-Pc1a (cleaved/Oxidized)
32.74	3390.78	8.85	Pc-3391	U ₃₅ -MYRTX-Pc1a (cleaved)
32.94	3235.12	34.58	Pc-3233	U ₃₅ -MYRTX-Pc1a
35.04	3306,87	4.72	Pc-3308	U ₃₅ -MYRTX-Pc1b
35.43	3148,80	3.96	Pc-3151	U ₃₅ -MYRTX-Pc1b (cleaved)
36.21	3462,10	13.85	Pc-3462	U ₃₄ -MYRTX-Pc1a
41.74	2648,49	1.09	Pc-2650	U ₃ -MYRTX-Pc1b
42.31	2799.50	2.69	Pc-2800	
46.18	3065.74	4.36	Pc-3071	U ₃ -MYRTX-Pc1a
49.71	2049.8	3.38	Pc-2050	
50.95	2065,28	8.80	Pc-2065	U ₃₃ -MYRTX-Pc1a
51.62	2055,26	2.40	Pc-2056	U ₃₃ -MYRTX-Pc1b

Supplementary Table X: Peptide sequences in the venom of *Pogonomyrmex californicus*. PTMs, post-translational modifications. TPMs, transcripts per millions. "*" C-terminal amidation. "ξ" Not validated through proteomics. "-" Not detected in the transcriptome after the final assembly. "b" The TPM value is the sum of the frequency from more than one assembled contig encoding the same transcript.

Toxin name	Mass (Da)	TPMs	Sequence	PTMs
U ₃ -MYRTX-Pc1a	3065.74	-	ALPALPLLALLFSLPAVQHWIEKNWIN*	
U ₃ -MYRTX-Pc1b	2648,49	-	GLPILALFVLIPFIHHYLMEKW*	
U ₃ -MYRTX-Pc1d	2622.44 ξ	191,922.9 2	GLPILASFVLIPFIHHYLMEKW*	
U ₃ -MYRTX-Pc1e	3175.76 ξ	32,210.36	ALPALPFLIFLFTLPAVQHWVEKNWIN*	
U ₃ -MYRTX-Pc1g	2832.57 ξ	15,834,74	GLPLLVFLFSLPAVQHWIEKNWIN*	
U ₁₇ -MYRTX-Pc1a	3562.92	698,35	HIIQVPCLPGYVKVGKDGVCREAFKFKPGQRP*	1 S-S
U ₁₇ -MYRTX-Pc1b	3275.53	1,936.25	HIIQVPCRDGYVMVNGVCREVFNEKDEE	1 S-S
U ₁₇ -MYRTX-Pc1c	2488.16	-	HIIQVPCRDGYVKDNGACIPEY	1 S-S Deamidation
U ₁₈ -MYRTX-Pc1a	5633,46	394.84	ENMSNRIPCTEEDIGYCFNGECFYIPAINAKGCICNKGYDGYRCQNTILD	3 S-S Dehydration
U ₁₈ -MYRTX-Pc1b	5828,51 ξ	354.17	ENISNRIPCTEEYNGYCLNGECFTIGGDENMKHCVCPCKEFSGERCQIRDYA	3 S-S
U ₃₃ -MYRTX-Pc1a	2065.28	64,414.01	ISPLIPLISFLASLIAAIKS*	
U ₃₃ -MYRTX-Pc1b	2055.26	33,954.24	ISPLISLISFLASLIAAIKS*	
U ₃₄ -MYRTX-Pc1a	3462.10	55,877.39	DSKIFRALITLGKMLLPVILPTVAEKIKEKV*	
U ₃₅ -MYRTX-Pc1a	3232.71	159,799.0 7	VDFKEMMKKITPDLLEMLEDIKAKIQQ*	
U ₃₅ -MYRTX-Pc1b	3306.87	-	ASVDLKELLKKITPDLLEMLDDIKAKIQQ*	
U ₃₆ -MYRTX-Pc1a	3004.61 ξ	6,004.13	VDPGQAQKEIGIFDRITELINWLVNH	

Supplementary Table XI : Peptide mass fingerprinting of *Myrmica ruginodis* venom. List of peptide masses detected through LC–MS using an LCQ-ion trap Advantage mass spectrometer.

Retention time (min)	Mass (Da)	Relative abundance (%)	Temporary name	Peptide toxin
2.28	1101.64	0.15	Mru-1102	U ₁₄ -MYRTX-Mru1a
2.3	2948.88	0.20	Mru-2949	
11.75	544.12	0.18	Mru-544	
15.26	2907.60	0.31	Mru-2908	
15.68	1463.85	0.61	Mru-1464	U ₆ -MYRTX-Mru1a
16.41	2675.10	0.08	Mru-2675	
16.36	2665.65	0.07	Mru-2666	
17.38	1401.73	1.31	Mru-1402	U ₁₄ -MYRTX-Mru1c
19.12	2801.27	0.22	Mru-2801	U ₁₇ -MYRTX-Mru1a
19.26	3829.77	0.24	Mru-3830	
19.26	3742.95	0.12	Mru-3743	
19.8	1430.43	0.06	Mru-1430	
19.91	1427.46	0.04	Mru-1427	
20.05	5289.26	0.14	Mru-5289	U ₁₈ -MYRTX-Mru1b
20.14	1407.46	0.08	Mru-1408	
20.77	2478.66	0.07	Mru-2479	
21.99	1389.42	0.06	Mru-1389	
22.63	1652.93	1.44	Mru-1653	U ₃ -MYRTX-Mru1a (cleaved)
23.07	5300.32	0.92	Mru-5300	U ₁₈ -MYRTX-Mru1a
23.92	4019.36	0.12	Mru-4019	
25.42	1766.01	0.33	Mru-1766	U ₃ -MYRTX-Mru1a (cleaved)
26.64	1735.62	0.03	Mru-1736	
27.26	1124.67	0.75	Mru-1125	U ₁₄ -MYRTX-Mru1b

28.28	1813.98	0.89	Mru-1814	U ₃ -MYRTX-Mru1b (cleaved)
29.79	1914.13	0.71	Mru-1914	U ₁₄ -MYRTX-Mru1d
30.71	695.26	0.63	Mru-695	
31.82	729.40	0.69	Mru-729	
32.22	1590.85	0.50	Mru-1591	U ₆ -MYRTX-Mru1b
34.73	2884.81	1.07	Mru-2885	U ₁₀ -MYRTX-Mru1a (Oxidized)
35.2	2074.40	2.80	Mru-2074	U ₁₀ -MYRTX-Mru1b
35.56	2898.81	2.43	Mru-2899	U ₁₀ -MYRTX-Mru1a (Oxidized/ Methylated)
36.69	2868.81	5.28	Mru-2869	U ₁₀ -MYRTX-Mru1a
37	842.53	4.46	Mru-843	U ₃ -MYRTX-Mru1a (cleaved)
39.64	3286.70	1.20	Mru-3287	U ₁₃ -MYRTX-Mru1a
40.51	2708.52	50.96	Mru-2709	U ₃₇ -MYRTX-Mru1a
41.02	2477.44	12.77	Mru-2477	U ₃ -MYRTX-Mru1a
42.1	2926.62	3.39	Mru-2927	U ₃₈ -MYRTX-Mru1a (cleaved)
42.75	2735.73	1.54	Mru-2735	
44.61	2492.30	3.17	Mru-2492	U ₃ -MYRTX-Mru1b

Supplementary Table XII : Peptide sequences in the venom of *Myrmica ruginodis*. PTMs. post-translational modifications. TPMs. transcripts per millions. "*" C-terminal amidation. "ξ" Not validated through proteomics. "-" Not detected in the transcriptome. "p" The TPM value is the sum of the frequency from more than one assembled contig encoding the same transcript.

Toxin name	Mass (Da)	TPMs	Sequence	PTMs
U ₃ -MYRTX-Mru1a	2477.44	100,084.03	GLPLFALLMALPALQHYIEKKI*	
U ₃ -MYRTX-Mru1b	2492.30	19,980.38	GLPLLALIMALPFLQHYIEKNV	

U ₆ -MYRTX- Mru1a	1463.85	2,879.34 [‡]	LIGPCPKKPIGIKC	1 S-S
U ₆ -MYRTX- Mru1b	1590.8	58,024.59 [‡]	IIDPCPKIFKGLFC	1 S-S
U ₁₀ -MYRTX- Mru1a	2868.81	83,363.96	GIGKILGKVALKILKVVAPAAAEAIADKI*	
U ₁₀ -MYRTX- Mru1b	2074.40	-	GIGKVLKNILSKLKKLLPL	
U ₁₀ -MYRTX- Mru1c	2812.75 ^ξ	1,822,750	GIGNVLAKVALKILKVVAPSAAAAIADKI*	
U ₁₃ -MYRTX- Mru1a	3286.70	4,440.72 [‡]	DKFGQAKKIGLFDQIDRAWAWFMKLEF*	
U ₁₄ -MYRTX- Mru1a	1101.64	417.18	IDSKAIKSLQ*	
U ₁₄ -MYRTX- Mru1b	1124.67	17,118.51	IDPKVLESLL*	
U ₁₄ -MYRTX- Mru1c	1401.73	4,544.89 [‡]	IDSALKSLQGGTV*	
U ₁₄ -MYRTX- Mru1d	1914.13	14,22.39	INPKLWLKLFKLESV*	
U ₁₇ -MYRTX- Mru1a	2801.27	1,288.04	AIIDAPRNCPPGHDIDHAGDCVEIFE	1 S-S
U ₁₇ -MYRTX- Mru1b	2673.31 ^ξ	209.00 [‡]	HIIVFPNCPPGQKPRGNCRQEY	1 S-S
U ₁₇ -MYRTX- Mru1c	2905.58 ^ξ	209.00 [‡]	HIINVPIQCPPGKVRVGNRCRDVGRV	1 S-S
U ₁₇ -MYRTX- Mru1d	2946.47 ^ξ	510.81	HIIVPDRNCPPGQRKDNHGNCRDIA	1 S-S
U ₁₇ -MYRTX- Mru1e	2769.38 ^ξ	1,139.63	HIIVIDKNCPPGQREDHHGHCRVI*	1 S-S

U ₁₇ -MYRTX- Mru1f	3275.73 ξ	758.24	HIIIVPERPPKCPPGQERDRRGRCRMVP	1 S-S
U ₁₇ -MYRTX- Mru1g	3169.70 ξ	690.48	HIIIVPELQCPPGQKRDRQGRCRKVFN*	1 S-S
U ₁₇ -MYRTX- Mru1h	3388.72 ξ	5,315.62	HIIVLPNRNTTNTCPPGQKKDNYGNCRKIA	1 S-S
U ₁₈ -MYRTX- Mru1a	5300.32	2,172.43 ^b	QSHFQPCSSHEHFCLNGECFYLAANEIGCICPPGFQGGQRCGELILD	3 S-S / 1 Pyro- Q
U ₁₈ -MYRTX- Mru1b	5289.26	206.96	THLEPSCPPSHEGFCLNGGTCVTIAELNTYACICAPGFHGSRCETANLE	3 S-S / 1 Fucose
U ₁₈ -MYRTX- Mru1c	5593.47 ξ	2,172.43 ^b	QSHFQPCSSHEHFCLNGECFYLAANEIGCICPPGFQGGQRCGELILEYV	3 S-S
U ₁₈ -MYRTX- Mru1d	6181.63 ξ	1,204.75 ^b	NHDPCPPQYAEALCLNGGTCFTVTIMGSDNYNCICAPGFQGWRCQEKSLDHP VNQ	3 S-S / O- glycosylation
U ₃₇ -MYRTX- Mru1a	2708.52	7,396.18 ^b	ADIAALIKELSEKAEEAIKKILGQE*	

Supplementary Table XIII : Mass fingerprinting of *Solenopsis saevissima* venom after hexanic extraction. List of peptide masses detected through LC–MS using an LCQ-ion trap Advantage mass spectrometer.

Retention time (min)	Mass (Da)	Relative abundance (%)	Temporary name
21.91	3162.78	0.45	Ss-3163
25.24	4993.96	< 0.01	Ss-4994
27.21	2361.08	1.25	Ss-2361
30.85	3866.08	1.55	Ss-3866
30.85	3737.48	0.22	Ss-3738
40.29	4634.6	4.95	Ss-4635
41.98	4979.44	1.53	Ss-4979
47.95	253.00	8.48	Alkaloid1

49.23

279.22

81.56

Alkaloid2

Supplementary Table XIV : Venom glands RNA sequencing data.

Species	Raw reads	Assembler	Contigs	Mean length (bp)
<i>Myrmica ruginodis</i>	79,786,762	Trinity	27,951	2,069
		Oases	16,066	2,581
<i>Daceton armigerum</i>	19,389,866	Trinity	28,581	1,662
		Oases	13,701	2,050
<i>Pogonomyrmex californicus</i>	63,958,136	Trinity	32,277	1,341
		Oases	25,722	1,335
<i>Stenammas debile</i>	74,247,606	Trinity	48,353	2,804
		Oases	25,710	2,995
<i>Tetramorium africanum</i>	82,055,072	Trinity	44,075	1,524
		Oases	21,077	1,907

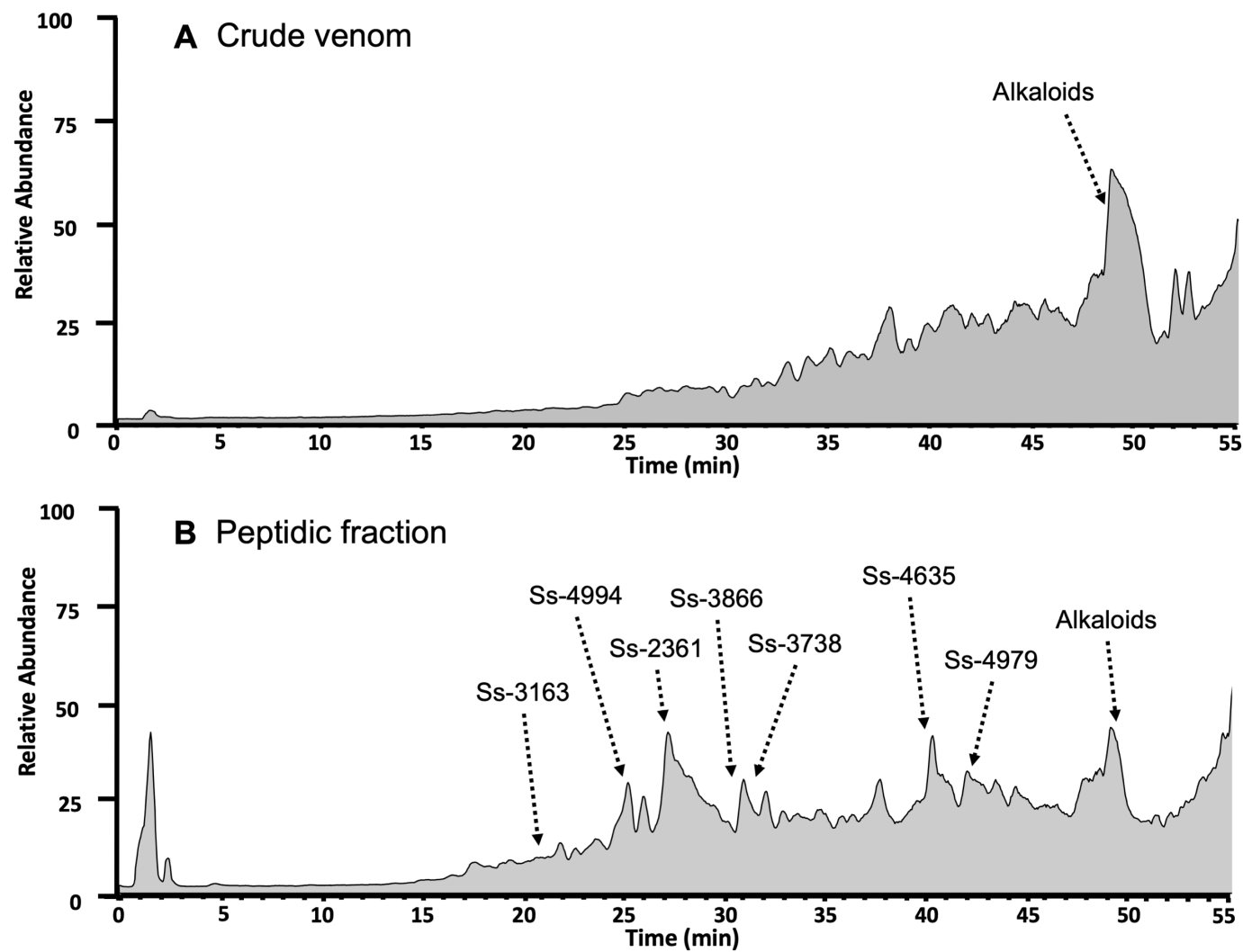


Figure S1: Total ion chromatograms (TIC) of crude venom (A) and venom after hexanic extraction (B) from *Solenopsis saevissima*. Peptides were eluted through RP-HPLC on a C18 column using a linear H₂O/ACN gradient at a flow rate of 250 $\mu\text{L}\cdot\text{min}^{-1}$.

Supplementary Table XV: Addressing table of major contigs expressed by *Myrmica ruginodis* venom glands.

Contig Name	TPMs	Contig ORF	Seq ID Name	E-value	Function
CL1068Contig1_1	2337.68				
CL1111Contig1_1	100,084.41		U ₁₀ -MYRTX-Mru1b		Venom peptide
CL1186Contig1_1	4,286.44	CL1186Contig1_1_1	sp Q962U0 RL13A_SPOFR 60S ribosomal protein L13a OS=Spodoptera frugiperda OX=7108 GN=RpL13A PE=2 SV=1	4.08E-107	Translation
CL1339Contig1_1	844.92	CL1339Contig1_1_2	sp Q8BT14 CNOT4_MOUSE CCR4-NOT transcription complex subunit 4 OS=Mus musculus OX=10090 GN=Cnot4 PE=1 SV=2	6.81E-127	Transcription
		CL1339Contig1_1_4	sp Q68KK0 PA1_SOLIN Phospholipase A1 OS=Solenopsis invicta OX=13686 PE=1 SV=1	1.14E-71	Venom allergen
CL181Contig1_1	10,776.66	CL181Contig1_1_3	sp Q9UL36 ZN236_HUMAN Zinc finger protein 236 OS=Homo sapiens OX=9606 GN=ZNF236 PE=2 SV=2	0.000193	Transcription
CL1902Contig1_1	1,137.65				
CL1Contig401_1	950.32				
CL1Contig424_1	719.57	CL1Contig424_1_4	sp P98165 VLDLR_CHICK Very low-density lipoprotein receptor OS=Gallus gallus OX=9031 GN=VLDLR PE=1 SV=1	3.16E-16	Metabolism
		CL1Contig424_1_6	sp P28175 LFC_TACTR Limulus clotting factor C OS=Tachypleus tridentatus OX=6853 PE=1 SV=1	4.36E-33	Host Defense
CL1Contig49_1	605.99	CL1Contig49_1_1	sp O62589 GD_DROME Serine protease gd OS=Drosophila melanogaster OX=7227 GN=gd PE=1 SV=2	2.38E-29	Metabolism
CL3021Contig1_1	996.04	CL3021Contig1_1_3	sp Q02241 KIF23_HUMAN Kinesin-like protein KIF23 OS=Homo sapiens OX=9606 GN=KIF23 PE=1 SV=3	3.11E-8	Metabolism
		CL3021Contig1_1_6	sp Q963B7 RL9_SPOFR 60S ribosomal protein L9 OS=Spodoptera frugiperda OX=7108 GN=RpL9 PE=2 SV=1	3.62E-113	Translation
		CL3021Contig1_1_6	sp Q9VFE6 RRP15_DROME RRP15-like protein OS=Drosophila melanogaster OX=7227 GN=CG3817 PE=1 SV=1	5.86E-33	RNA Maturation
CL3151Contig1_1	4476.93		U ₁₄ -MYRTX-Mru1c		Venom peptide
CL3360Contig1_1	5,239.96	CL3360Contig1_1_2	sp P15357 RS27A_DROME Ubiquitin-40S ribosomal protein S27a OS=Drosophila melanogaster OX=7227 GN=RpS27A PE=1 SV=2	7.84E-82	Translation
CL4015Contig1_1	1,349.14	CL4015Contig1_1_4	sp P29520 EF1A_BOMMO Elongation factor 1-alpha OS=Bombyx mori OX=7091 PE=2 SV=1	0	Translation
CL4093Contig1_1	119,676.43				

CL410Contig1_1	2,083.71	CL410Contig1_1_1	sp Q8TGM7 ART2_YEAST Putative uncharacterized protein ART2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=ART2 PE=5 SV=1	1.96E-14	
CL4476Contig1_1	9,205.87	CL4476Contig1_1_5	sp P32429 RL7A_CHICK 60S ribosomal protein L7a OS=Gallus gallus OX=9031 GN=RPL7A PE=2 SV=2	2.89E-121	Translation
CL45Contig1_1	1,871.42	CL45Contig1_1_1	sp Q3KQG9 PPAT_XENLA Testicular acid phosphatase homolog OS=Xenopus laevis OX=8355 GN=acp4 PE=2 SV=1	7.43E-21	Metabolism
		CL45Contig1_1_1	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase Acph-1 OS=Apis mellifera OX=7460 PE=1 SV=1	1E-3	Venom allergen
		CL45Contig1_1_2	sp A6H730 PPAP_BOVIN Prostatic acid phosphatase OS=Bos taurus OX=9913 GN=ACPP PE=2 SV=1	1.89E-6	Metabolism
		CL45Contig1_1_5	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase Acph-1 OS=Apis mellifera OX=7460 PE=1 SV=1	1.21E-104	Venom allergen
CL511Contig1_1	1,112.82	CL511Contig1_1_5	sp P36362 CHIT_MANSE Endochitinase OS=Manduca sexta OX=7130 PE=2 SV=1	0	Metabolism
CL51Contig1_1	1,997.87				
CL729Contig1_1	10,516.03	CL729Contig1_1_5	sp O76756 RS8_APIME 40S ribosomal protein S8 OS=Apis mellifera OX=7460 GN=RpS8 PE=2 SV=2	1.77E-125	Translation
CL976Contig1_1	8,253.62				
k25_Locus_1189_Transcript_1_1	1,887.15	k25_Locus_1189_Transcript_1_1_2	sp Q95WA0 RL26_LITLI 60S ribosomal protein L26 OS=Littorina littorea OX=31216 GN=RPL26 PE=2 SV=1	5.49E-68	Translation
		k25_Locus_1189_Transcript_1_1_5	sp Q9DA39 LFG4_MOUSE Protein lifeguard 4 OS=Mus musculus OX=10090 GN=Tmbim4 PE=2 SV=1	1.54E-53	Anti-apoptosis
k25_Locus_1941_Transcript_5_1	534.93				
k25_Locus_2391_Transcript_5_1	36,400.91				
k25_Locus_2822_Transcript_2_1	3,894.72		U6-MYRTX-Mru1a		Venom peptide
k25_Locus_3473_Transcript_1_1	2,594.72	k25_Locus_3473_Transcript_1_1_6	sp Q95V39 RL8_SPOFR 60S ribosomal protein L8 OS=Spodoptera frugiperda OX=7108 GN=RpL8 PE=2 SV=1	1.56E-157	Translation
k25_Locus_466_Transcript_10_1	820.64	k25_Locus_466_Transcript_10_1_3	sp Q68KK0 PA1_SOLIN Phospholipase A1 OS=Solenopsis invicta OX=13686 PE=1 SV=1	1.65E-71	Venom phospholipase
		k25_Locus_466_Transcript_10_1_6	sp O95628 CNOT4_HUMAN CCR4-NOT transcription complex subunit 4 OS=Homo sapiens OX=9606 GN=CNOT4 PE=1 SV=3	9.08E-127	Transcription
k25_Locus_469_Transcript_10_1	1,534.73	k25_Locus_469_Transcript_10_1_1	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase Acph-1	4.24E-6	Venom allergen

			OS=Apis mellifera OX=7460 PE=1 SV=1		
		k25_Locus_469_Transcript_10_1_2	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase Acph-1 OS=Apis mellifera OX=7460 PE=1 SV=1	1E-3	Venom allergen
		k25_Locus_469_Transcript_10_1_3	sp Q3KQG9 PPAT_XENLA Testicular acid phosphatase homolog OS=Xenopus laevis OX=8355 GN=acp4 PE=2 SV=1	7.43E-21	Venom allergen
		k25_Locus_469_Transcript_10_1_6	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase Acph-1 OS=Apis mellifera OX=7460 PE=1 SV=1	2.31E-58	Venom allergen
		k25_Locus_469_Transcript_10_1_6	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase Acph-1 OS=Apis mellifera OX=7460 PE=1 SV=1	6.35E-26	Venom allergen
k25_Locus_688_Transcript_1_1	2,673.28	k25_Locus_688_Transcript_1_1_6	sp P09180 RL4_DROME 60S ribosomal protein L4 OS=Drosophila melanogaster OX=7227 GN=RpL4 PE=1 SV=2	9.62E-167	Translation
k31_Locus_2738_Transcript_1_1	4,063.54	k31_Locus_2738_Transcript_1_1_6	sp Q962U1 RL13_SPOFR 60S ribosomal protein L13 OS=Spodoptera frugiperda OX=7108 GN=RpL13 PE=2 SV=1	1.24E-92	Translation
k31_Locus_303_Transcript_6_1	2,776.65	k31_Locus_303_Transcript_6_1_5	sp P36362 CHIT_MANSE Endochitinase OS=Manduca sexta OX=7130 PE=2 SV=1	4.16E-99	Metabolism
		k31_Locus_303_Transcript_6_1_6	sp P36362 CHIT_MANSE Endochitinase OS=Manduca sexta OX=7130 PE=2 SV=1	2.57E-72	Metabolism
k31_Locus_3464_Transcript_1_1	1,352.61	k31_Locus_3464_Transcript_1_1_5	sp Q38PS2 NU1M_MAMPR NADH-ubiquinone oxidoreductase chain 1 OS=Mammuthus primigenius OX=37349 GN=MT-ND1 PE=3 SV=1	1.32E-10	Metabolism
		k31_Locus_3464_Transcript_1_1_5	sp Q70Y29 NU1M_MYRTR NADH-ubiquinone oxidoreductase chain 1 OS=Myrmecophaga tridactyla OX=71006 GN=MT-ND1 PE=3 SV=1	1.04E-19	Metabolism
k31_Locus_3573_Transcript_1_1	3,781.27	k31_Locus_3573_Transcript_1_1_6	sp Q962R2 RS17_SPOFR 40S ribosomal protein S17 OS=Spodoptera frugiperda OX=7108 GN=RpS17 PE=2 SV=3	4.23E-35	Translation
k31_Locus_48_Transcript_9_1	3,645.56				
k31_Locus_750_Transcript_1_1	443.50	k31_Locus_750_Transcript_1_1_1	sp P82147 L2EFL_DROME Protein lethal(2)essential for life OS=Drosophila melanogaster OX=7227 GN=l(2)efl PE=1 SV=1	1.13E-14	
k31_Locus_795_Transcript_15_1	554.42	k31_Locus_795_Transcript_15_1_1	sp P13060 EF2_DROME Elongation factor 2 OS=Drosophila melanogaster OX=7227 GN=EF2 PE=1 SV=4	0	Translation
k31_Locus_813_Transcript_8_1	678.38	k31_Locus_813_Transcript_8_1_4	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS=Apis mellifera OX=7460 PE=1 SV=1	0	Peptide maturation
		k31_Locus_813_Transcript_8_1_6	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS=Apis mellifera OX=7460 PE=1 SV=1	1.49E-39	Peptide maturation
k37_Locus_1036_Transcript_1_1	2,033.54	k37_Locus_1036_Transcript_1_1_1	sp A8CAG3 RL17_PHLPP 60S ribosomal protein L17 OS=Phlebotomus papatasi OX=29031 GN=RpL17 PE=2 SV=1	1.36E-101	Translation

k37_Locus_1386_Transcript_2_1	1,891.30	k37_Locus_1386_Transcript_2_1_2	sp P62909 RS3_RAT 40S ribosomal protein S3 OS=Rattus norvegicus OX=10116 GN=Rps3 PE=1 SV=1	4.52E-130	Translation
		k37_Locus_1386_Transcript_2_1_5	sp Q9V447 KRH2_DROME Krueppel homolog 2 OS=Drosophila melanogaster OX=7227 GN=Kr-h2 PE=1 SV=1	2.67E-61	Protein maturation
k37_Locus_1583_Transcript_1_1	3,447.70	k37_Locus_1583_Transcript_1_1_5	sp P46782 RS5_HUMAN 40S ribosomal protein S5 OS=Homo sapiens OX=9606 GN=RPS5 PE=1 SV=4	7.48E-122	Translation
k37_Locus_263_Transcript_1_1	5,600.36		U ₁₄ -MYRTX-Mru1c		Venom peptide
k37_Locus_35_Transcript_13_1	2,679.89	k37_Locus_35_Transcript_13_1_3	sp Q8TGM7 ART2_YEAST Putative uncharacterized protein ART2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=ART2 PE=5 SV=1	1.96E-14	
k43_Locus_1312_Transcript_28_1	28,412.38		U ₃ -MYRTX-Mru1a		Venom peptide
k43_Locus_1332_Transcript_1_1	1,962.86	k43_Locus_1332_Transcript_1_1_4	sp Q02748 IF4A_DROME Eukaryotic initiation factor 4A OS=Drosophila melanogaster OX=7227 GN=eIF4A PE=1 SV=3	0	Host Defense
k43_Locus_2044_Transcript_1_1	2,787.90	k43_Locus_2044_Transcript_1_1_5	sp Q9U3U0 RLA0_CERCA 60S acidic ribosomal protein P0 OS=Ceratitidis capitata OX=7213 GN=RpLP0 PE=3 SV=1	9.53E-160	Translation
k43_Locus_4249_Transcript_1_1	2,334.73	k43_Locus_4249_Transcript_1_1_4	sp P55833 RS27_HOMAM 40S ribosomal protein S27 OS=Homarus americanus OX=6706 GN=RPS27 PE=3 SV=2	9.34E-36	Translation
k43_Locus_708_Transcript_7_1	443.25	k43_Locus_708_Transcript_7_1_4	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS=Apis mellifera OX=7460 PE=1 SV=1	2.31E-33	Protein maturation
		k43_Locus_708_Transcript_7_1_5	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS=Apis mellifera OX=7460 PE=1 SV=1	9.17E-134	Protein maturation
		k43_Locus_708_Transcript_7_1_5	sp B1A4F7 VDDP4_VESVU Venom dipeptidyl peptidase 4 OS=Vespula vulgaris OX=7454 PE=1 SV=1	6.7E-40	Protein maturation
k43_Locus_927_Transcript_3_1	2,244.74	k43_Locus_927_Transcript_3_1_6	sp P29341 PABP1_MOUSE Polyadenylate-binding protein 1 OS=Mus musculus OX=10090 GN=Pabpc1 PE=1 SV=2	0	Transcription
k49_Locus_334_Transcript_10_1	2,043.84	k49_Locus_334_Transcript_10_1_1	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase Acph-1 OS=Apis mellifera OX=7460 PE=1 SV=1	0.001	Venom allergen
		k49_Locus_334_Transcript_10_1_2	sp Q3KQG9 PPAT_XENLA Testicular acid phosphatase homolog OS=Xenopus laevis OX=8355 GN=acp4 PE=2 SV=1	7.43E-21	Venom allergen
		k49_Locus_334_Transcript_10_1_3	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase Acph-1 OS=Apis mellifera OX=7460 PE=1 SV=1	4.24E-6	Venom allergen
		k49_Locus_334_Transcript_10_1_5	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase Acph-1 OS=Apis mellifera OX=7460 PE=1 SV=1	5.26E-97	Venom allergen
k49_Locus_35_Transcript_15_1	1,681.90	k49_Locus_35_Transcript_15_1_3	sp Q96P53 WDFY2_HUMAN WD repeat and FYVE domain-containing protein 2 OS=Homo sapiens OX=9606 GN=WDFY2 PE=1 SV=2	1.2E-136	Metabolism

k55_Locus_185_Transcript_38_1	30,880.37		U ₃₇ -MYRTX-Mru1a		Venom peptide
k55_Locus_2820_Transcript_6_1	7,813.26	k55_Locus_2820_Transcript_6_1_5	sp B0W6N3 EIF3A_CULQU Eukaryotic translation initiation factor 3 subunit A OS=Culex quinquefasciatus OX=7176 GN=eIF3-S10 PE=3 SV=2	0	Translation
k55_Locus_487_Transcript_2_1	1,183.92	k55_Locus_487_Transcript_2_1_5	sp P29520 EF1A_BOMMO Elongation factor 1-alpha OS=Bombyx mori OX=7091 PE=2 SV=1	0	Translation
k55_Locus_7250_Transcript_1_1	2,092.26	k55_Locus_7250_Transcript_1_1_2	sp Q922J9 FACR1_MOUSE Fatty acyl-CoA reductase 1 OS=Mus musculus OX=10090 GN=Far1 PE=1 SV=1	3.93E-96	Metabolism
		k55_Locus_7250_Transcript_1_1_5	sp Q5G5C4 RS3A_PERAM 40S ribosomal protein S3a OS=Periplaneta americana OX=6978 GN=Parcpwex01 PE=2 SV=1	6.85E-143	Translation
		k55_Locus_7250_Transcript_1_1_5	sp P9WQ49 FAD21_MYCTU Putative fatty-acid-CoA ligase FadD21 OS=Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv) OX=83332 GN=fadD21 PE=1 SV=1	8E-3	Metabolism
k55_Locus_781_Transcript_76_1	7,060.17		U ₆ -MYRTX-Mru1b		Venom peptide
k61_Locus_2195_Transcript_3_1	2,519.21		U ₁₇ -MYRTX-Mru1h		Venom peptide
k61_Locus_2828_Transcript_1_1	5,315.64				
k61_Locus_31_Transcript_6_1	1,320.87	k61_Locus_31_Transcript_6_1_5	sp Q96P53 WDFY2_HUMAN WD repeat and FYVE domain-containing protein 2 OS=Homo sapiens OX=9606 GN=WDFY2 PE=1 SV=2	1.2E-136	Metabolism
		k61_Locus_31_Transcript_6_1_6	sp B0VYY2 COX5A_NYCCO Cytochrome c oxidase subunit 5A, mitochondrial OS=Nycticebus coucang OX=9470 GN=COX5A PE=2 SV=1	1.4E-47	Metabolism
k61_Locus_6941_Transcript_4_1	1,037.24	k61_Locus_6941_Transcript_4_1_1	sp Q5BLY4 ICA_APIME Icarapin-like OS=Apis mellifera OX=7460 PE=2 SV=1	6.44E-21	Venom protein
k65_Locus_2783_Transcript_1_1	3,951.92		U ₁₃ -MYRTX-Mru1a		Venom peptide
k69_Locus_2037_Transcript_1_1	1,790.05				
k69_Locus_303_Transcript_5_1	785.20				
k69_Locus_3235_Transcript_2_1	2,442.86				
k69_Locus_3763_Transcript_5_1	852.86	k69_Locus_3763_Transcript_5_1_4	sp B1A4F7 VDDP4_VESVU Venom dipeptidyl peptidase 4 OS=Vespa vulgaris OX=7454 PE=1 SV=1	2.42E-133	Protein maturation
		k69_Locus_3763_Transcript_5_1_6	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS=Apis mellifera OX=7460 PE=1 SV=1	5.28E-85	Protein maturation
k69_Locus_421_Transcript_19_1	340.22	k69_Locus_421_Transcript_19_1_2	sp Q8MQS8 SP34_APIME Venom serine protease 34 OS=Apis mellifera OX=7460 PE=2 SV=1	2.17E-69	Protein maturation

		k69_Locus_421_Transcript_19_1_3	sp Q8BTI8 SRRM2_MOUSE Serine/arginine repetitive matrix protein 2 OS=Mus musculus OX=10090 GN=Srrm2 PE=1 SV=3	0.0000243	RNA Maturation
		k69_Locus_421_Transcript_19_1_6	sp Q9VKJ9 C2D1_DROME Coiled-coil and C2 domain-containing protein 1-like OS=Drosophila melanogaster OX=7227 GN=(2)gd1 PE=1 SV=1	8.27E-144	Transcription
k69_Locus_509_Transcript_1_1	1,731.35	k69_Locus_509_Transcript_1_1_4	sp P29520 EF1A_BOMMO Elongation factor 1-alpha OS=Bombyx mori OX=7091 PE=2 SV=1	0	Translation
k69_Locus_8400_Transcript_2_1	5,434.82	k69_Locus_8400_Transcript_2_1_3	sp Q75VN3 TCTP_BOMMO Translationally-controlled tumor protein homolog OS=Bombyx mori OX=7091 GN=Tctp PE=2 SV=1	2.95E-91	Metabolism
k69_Locus_8834_Transcript_1_1	58,024.81				

Supplementary Table XVI: Addressing table of major contigs expressed by *Tetramorium africanum* venom glands.

Contig Name	TPMs	Contig ORF	Seq ID Name	E-value	Function
CL108Contig1_1	368.12	CL108Contig1_1_5	sp A3DC27 RSGI2_HUNT2 Anti-sigma-I factor RsgI2 OS=Hungateiclostridium thermocellum (strain ATCC 27405 / DSM 1237 / JCM 9322 / NBRC 103400/ NCIMB 10682 / NRRL B-4536 / VPI 7372) OX=203119 GN=rsgI2 PE=1 SV=1	4.79E-6	Metabolism
CL1458Contig1_1	1,508.21		U ₁₇ -MYRTX-Ta1e		Venom peptide
CL160Contig1_1	1,792.98		U ₆ -MYRTX-Ta1a		Venom peptide
CL164Contig1_1	17,899.26		U ₂ -MYRTX-Ta1a/U ₃ -MYRTX-Ta1b		Venom peptide
CL179Contig1_1	2,148.80				
CL17Contig2_1	1,147.34		U ₄ -MYRTX-Ta1a		Venom peptide
CL17Contig3_1	3,136.69		U ₄ -MYRTX-Ta1a		
CL3868Contig1_1	729.22	CL3868Contig1_1_2	sp P58743 S26A5_HUMAN Prestin OS=Homo sapiens OX=9606 GN=SLC26A5 PE=1 SV=1	2.07E-100	Metabolism
CL459Contig1_1	1,456.14		U ₇ -MYRTX-Ta1b		Venom peptide
CL4634Contig1_1	13,523.30				
CL4869Contig1_1	5,907.92				
CL5482Contig1_1	5,808.34				
CL5522Contig1_1	1,224.34	CL5522Contig1_1_5	sp P35778 VA3_SOLIN Venom allergen 3 OS=Solenopsis invicta OX=13686 PE=1 SV=2	2.14E-38	Venom allergen

		CL5522Contig1_1_6	sp P0DSI3 VA3_DINQU Venom allergen 3 homolog OS=Dinoponera quadriceps OX=609295 PE=2 SV=1	3.25E-28	Venom allergen
CL588Contig1_1	8,094.94		U ₁₇ -MYRTX-Ta1d		Venom peptide
CL617Contig1_1	4,805.32		U ₁₇ -MYRTX-Ta1b		Venom peptide
CL653Contig1_1	1,894.19				
CL87Contig1_1	2,351.59		U ₂₅ -MYRTX-Ta1a		Venom peptide
k25_Locus_653_Transcript_26_6	4,937.03				
k31_Locus_1121_Transcript_37_1	25,205.99		U ₂₂ -MYRTX-Ta1a		Venom peptide
k31_Locus_467_Transcript_20_1	3,878.95	k31_Locus_467_Transcript_20_1_5	sp P0DSI2 PA1_DINQU Phospholipase A1 OS=Dinoponera quadriceps OX=609295 PE=2 SV=1	2.39E-60	Venom protein
k31_Locus_663_Transcript_19_1	2,046.71		U ₇ -MYRTX-Ta1a		
k31_Locus_683_Transcript_25_1	3,740.35				
k31_Locus_683_Transcript_25_2	2,510.27				
k31_Locus_683_Transcript_25_7	2,338.19				
k31_Locus_778_Transcript_27_1	2,816.59		U ₁₇ -MYRTX-Ta1c		Venom peptide
k31_Locus_941_Transcript_5_1	13,409.41		U ₂₁ -MYRTX-Ta1a / U ₁₁ -MYRTX-Ta1a / U ₁₃ -MYRTX-Ta1a		Venom peptide
k37_Locus_18_Transcript_477_1	782.71		U ₁₇ -MYRTX-Ta1d		Venom peptide
k37_Locus_2356_Transcript_10_1	73,264.21		U ₂ -MYRTX-Ta1a / U ₃ -MYRTX-Ta1d		
k37_Locus_266_Transcript_1_1	1,842.83				
k37_Locus_4283_Transcript_1_1	16,960.97		U ₂₂ -MYRTX-Ta1a		Venom peptide
k43_Locus_102_Transcript_4_1	76,679.09		M-MYRTX-Ta1a		
k43_Locus_14252_Transcript_1_1	2,872.25	k43_Locus_14252_Transcript_1_1_6	sp Q9U3U0 RLA0_CERCA 60S acidic ribosomal protein P0 OS=Ceratitis capitata OX=7213 GN=RpLP0 PE=3 SV=1	2.27E-159	Translation
k43_Locus_1699_Transcript_37_1	49,972.40		U ₂ -MYRTX-Ta1a		Venom peptide
k43_Locus_211_Transcript_11_1	325.61	k43_Locus_211_Transcript_11_1_4	sp A3DC27 RSGI2_HUNT2 Anti-sigma-I factor RsgI2	1.6E-2	Metabolism

			OS=Hungateclostridium thermocellum (strain ATCC 27405 / DSM 1237 / JCM 9322 / NBRC 103400 / NCIMB 10682 / NRRL B-4536 / VPI 7372) OX=203119 GN=rsgI2 PE=1 SV=1		
k43_Locus_856_Transcript_6_1	6,865.94		U ₁₁ -MYRTX-Ta1a/ U ₂₁ -MYRTX-Ta1b / U ₁₃ -MYRTX-Ta1a		Venom peptide
k49_Locus_1553_Transcript_12_1	36,786.51		U ₃ -MYRTX-Ta1d /U ₂ -MYRTX-Ta1a		Venom peptide
k49_Locus_2984_Transcript_13_2	2,205.24	k49_Locus_2984_Transcript_13_2_3	sp P0DSI2 PA1_DINQU Phospholipase A1 OS=Dinoponera quadriceps OX=609295 PE=2 SV=1	1.42E-59	Venom protein
k49_Locus_554_Transcript_11_1	1,827.42		U ₇ -MYRTX-Ta1b / U ₇ -MYRTX-Ta1a		Venom peptide
k55_Locus_11_Transcript_13_1	429.46		U ₁₇ -MYRTX-Ta1e		Venom peptide
k55_Locus_174_Transcript_1_1	637.63	k55_Locus_174_Transcript_1_1_2	sp P07707 NU4M_DROYA NADH-ubiquinone oxidoreductase chain 4 OS=Drosophila yakuba OX=7245 GN=mt :ND4 PE=3 SV=1	1.08E-39	Metabolism
k55_Locus_513_Transcript_10_1	1,557.40		U ₂₃ -MYRTX-Ta1a		Venom peptide
k55_Locus_683_Transcript_2_1	8,066.32		M-MYRTX-Ta1a		Venom peptide
k61_Locus_1435_Transcript_6_1	9,082.74		U ₃ -MYRTX-Ta1a		Venom peptide
k61_Locus_187_Transcript_13_1	405.20	k61_Locus_187_Transcript_13_1_4	sp A3DC27 RSGI2_HUNT2 Anti-sigma-I factor RsgI2 OS=Hungateclostridium thermocellum (strain ATCC 27405 / DSM 1237 / JCM 9322 / NBRC 103400 / NCIMB 10682 / NRRL B-4536 / VPI 7372) OX=203119 GN=rsgI2 PE=1 SV=1	4.36E-6	Metabolism
k61_Locus_2243_Transcript_6_1	2,363.92		U ₂₃ -MYRTX-Ta1a		Venom peptide
k61_Locus_874_Transcript_8_1	1,087.53				
k65_Locus_1263_Transcript_4_1	2,589.81		U ₁ -MYRTX-Ta1a		Venom peptide
k65_Locus_14_Transcript_5_1	1,619.97		U ₁₇ -MYRTX-Ta1c		
k65_Locus_2350_Transcript_3_1	1,931.24		U ₇ -MYRTX-Ta1b		Venom peptide
k65_Locus_8717_Transcript_3_1	10,923.04		U ₃ -MYRTX-Ta1c		Venom peptide
k69_Locus_16_Transcript_8_1	1,451.04		U ₁₇ -MYRTX-Ta1c		Venom peptide
k69_Locus_191_Transcript_7_1	447.04				

k69_Locus_2236_Transcript_5_1	5,186.60	k69_Locus_2236_Transcript_5_1_4	sp Q7JQ07 MOS1T_DROMA Mariner Mos1 transposase OS=Drosophila mauritiana OX=7226 GN=mariner\T PE=1 SV=1	8.13E-155	Transcription
k69_Locus_5034_Transcript_1_1	5,411.41		U ₃ -MYRTX-Ta1c		Venom peptide
k69_Locus_898_Transcript_2_1	14,7209.37		U ₁ -MYRTX-Ta1a		Venom peptide

Supplementary Table XVII : Addressing table of major contigs expressed by *Pogonomyrmex californicus* venom glands.

Name	TPMs	Contig ORF	Seq ID Name	E-value	Function
CL1024Contig1_1	75,492.15		U34-MYRTX-Pc1a		Venom peptide
CL103Contig1_1	30.03	CL103Contig1_1_2	sp P98158 LRP2_RAT Low-density lipoprotein receptor-related protein 2 OS=Rattus norvegicus OX=10116 GN=Lrp2 PE=1 SV=1	0	Metabolism
		CL103Contig1_1_4	sp Q5XIN3 MIPT3_RAT TRAF3-interacting protein 1 OS=Rattus norvegicus OX=10116 GN=Traf3ip1 PE=1 SV=1	1.19E-4	Metabolism
		CL103Contig1_1_4	sp Q5RC80 RBM39_PONAB RNA-binding protein 39 OS=Pongo abelii OX=9601 GN=RBM39 PE=2 SV=1	1.74E-154	Transcription
CL1171Contig1_1	523.14				
CL131Contig2_1	64.36	CL131Contig2_1_3	sp A0A291PQF1 UGT5_DACCO UDP-glycosyltransferase UGT5 OS=Dactylopius coccus OX=765876 GN=UGT5 PE=1 SV=1	6.23E-10	Protein maturation
CL15Contig2_1	112.47	CL15Contig2_1_5	sp Q9R049 AMFR_MOUSE E3 ubiquitin-protein ligase AMFR OS=Mus musculus OX=10090 GN=Amfr PE=1 SV=2	2.2E-148	Metabolism
CL161Contig1_1	252.98				
CL161Contig2_1	212.25				
CL1660Contig1_1	85.26				
CL1739Contig1_1	103.53	CL1739Contig1_1_1	sp C3YWU0 FUCO_BRAFL Alpha-L-fucosidase OS=Branchiostoma floridae OX=7739 GN=BRAFLDRAFT_56888 PE=3 SV=2	2.62E-162	Metabolism
		CL1739Contig1_1_2	sp C3YWU0 FUCO_BRAFL Alpha-L-fucosidase OS=Branchiostoma floridae OX=7739 GN=BRAFLDRAFT_56888 PE=3 SV=2	2.73E-44	Metabolism
		CL1739Contig1_1_6	sp Q96KP4 CNDP2_HUMAN Cytosolic non-specific dipeptidase OS=Homo sapiens OX=9606 GN=CNDP2 PE=1 SV=2	0	Metabolism
CL1823Contig1_1	57.37	CL1823Contig1_1_1	sp Q7LBC6 KDM3B_HUMAN Lysine-specific demethylase 3B OS=Homo sapiens OX=9606 GN=KDM3B PE=1 SV=2	0	Metabolism
		CL1823Contig1_1_2	sp P39703 YAA4_YEAST Putative uncharacterized protein YAL004W	6.5E-22	

			OS= <i>Saccharomyces cerevisiae</i> (strain ATCC 204508 / S288c) OX=559292 GN=YAL004W PE=5 SV=2		
		CL1823Contig1_1_5	sp P29844 BIP_DROME Endoplasmic reticulum chaperone BiP OS= <i>Drosophila melanogaster</i> OX=7227 GN=Hsc70-3 PE=1 SV=2	0	Protein maturation
CL204Contig1_1	6203.22				
CL216Contig1_1	49.10	CL216Contig1_1_3	sp O15013 ARHGA_HUMAN Rho guanine nucleotide exchange factor 10 OS= <i>Homo sapiens</i> OX=9606 GN=ARHGEF10 PE=1 SV=4	1.73E-141	Metabolism
		CL216Contig1_1_6	sp Q15643 TRIPB_HUMAN Thyroid receptor-interacting protein 11 OS= <i>Homo sapiens</i> OX=9606 GN=TRIP11 PE=1 SV=3	1.05E-39	Metabolism
CL244Contig1_1	751.72	CL244Contig1_1_3	sp Q962Q7 RS23_SPOFR 40S ribosomal protein S23 OS= <i>Spodoptera frugiperda</i> OX=7108 GN=RpS23 PE=2 SV=1	6.37E-90	Translation
CL506Contig1_1	77.10	CL506Contig1_1_4	sp Q2F637 1433Z_BOMMO 14-3-3 protein zeta OS= <i>Bombyx mori</i> OX=7091 GN=14-3-3zeta PE=2 SV=2	4.27E-160	Metabolism
CL5424Contig1_1	99.52	CL5424Contig1_1_4	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS= <i>Apis mellifera</i> OX=7460 PE=1 SV=1	0.001	Protein maturation
		CL5424Contig1_1_5	sp B1A4F7 VDDP4_VESVU Venom dipeptidyl peptidase 4 OS= <i>Vespula vulgaris</i> OX=7454 PE=1 SV=1	8.09E-31	Protein maturation
		CL5424Contig1_1_6	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS= <i>Apis mellifera</i> OX=7460 PE=1 SV=1	6.86E-166	Protein maturation
CL613Contig1_1	402.24				
CL628Contig1_1	40.08	CL628Contig1_1_1	sp Q7SXP0 S22BB_DANRE Vesicle-trafficking protein SEC22b-B OS= <i>Danio rerio</i> OX=7955 GN=sec22bb PE=2 SV=1	3.83E-82	Protein maturation
		CL628Contig1_1_5	sp Q9NWK9 BCD1_HUMAN Box C/D snoRNA protein 1 OS= <i>Homo sapiens</i> OX=9606 GN=ZNHIT6 PE=1 SV=1	1.08E-14	Translation
		CL628Contig1_1_6	sp E9QAT4 SC16A_MOUSE Protein transport protein Sec16A OS= <i>Mus musculus</i> OX=10090 GN=Sec16a PE=1 SV=1	1.48E-114	Protein maturation
CL62Contig1_1	829.14	CL62Contig1_1_5	sp P0CG71 UBIQ1_CAEEL Polyubiquitin-A OS= <i>Caenorhabditis elegans</i> OX=6239 GN=ubq-1 PE=3 SV=1	0	Metabolism
CL6782Contig1_1	501.48				
CL6869Contig1_1	219.80	CL6869Contig1_1_4	sp P05031 DDC_DROME Aromatic-L-amino-acid decarboxylase OS= <i>Drosophila melanogaster</i> OX=7227 GN=Ddc PE=1 SV=4	0	Metabolism
		CL6969Contig1_1_4	sp Q8T8R1 Y3800_DROME CCHC-type zinc finger protein CG3800 OS= <i>Drosophila melanogaster</i> OX=7227 GN=CG3800 PE=1 SV=1	5.03E-40	Translation
CL6927Contig1_1	351.25				

CL6969Contig1_1	104.64				
CL7032Contig1_1	714.28	CL7032Contig1_1_4	sp P29520 EF1A_BOMMO Elongation factor 1-alpha OS=Bombyx mori OX=7091 PE=2 SV=1	0	Translation
CL7153Contig1_1	426.91	CL7153Contig1_1_3	sp P82147 L2EFL_DROME Protein lethal(2)essential for life OS=Drosophila melanogaster OX=7227 GN=l(2)efl PE=1 SV=1	3.25E-46	Metabolism
CL715Contig1_1	26.57	CL715Contig1_1_5	sp G5E8K5 ANK3_MOUSE Ankyrin-3 OS=Mus musculus OX=10090 GN=Ank3 PE=1 SV=1	0	Transcription
CL7245Contig1_1	90.94	CL7245Contig1_1_4	sp B6VQA1 DIMM_DROME Protein dimmed OS=Drosophila melanogaster OX=7227 GN=dimm PE=1 SV=1	1.94E-30	Transcription
CL7399Contig1_1	1,171.86				
CL73Contig2_1	3,81.75	CL73Contig2_1_2	sp P0C5Q0 YL54F_YEAST Putative uncharacterized protein YLR154W-F OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=YLR154W-F PE=5 SV=1	6.66E-5	
		CL73Contig2_1_6	sp P46413 GSHB_RAT Glutathione synthetase OS=Rattus norvegicus OX=10116 GN=Gss PE=1 SV=1	3.22E-30	Metabolism
CL7560Contig1_1	135.82	CL7560Contig1_1_2	sp Q9Y600 CSAD_HUMAN Cysteine sulfinic acid decarboxylase OS=Homo sapiens OX=9606 GN=CSAD PE=1 SV=2	6.22E-163	Metabolism
CL7618Contig1_1	121.02	CL7618Contig1_1_2	sp Q8BGX0 TRI23_MOUSE E3 ubiquitin-protein ligase TRIM23 OS=Mus musculus OX=10090 GN=Trim23 PE=2 SV=1	2.08E-152	Metabolism
		CL7618Contig1_1_6	sp P29413 CALR_DROME Calreticulin OS=Drosophila melanogaster OX=7227 GN=Calr PE=1 SV=2	0	Protein maturation
CL7696Contig1_1	1,450.61				
CL7902Contig1_1	566.94	CL7902Contig1_1_5	sp P62752 RL23A_RAT 60S ribosomal protein L23a OS=Rattus norvegicus OX=10116 GN=Rpl23a PE=2 SV=1	7.86E-53	Translation
CL7969Contig1_1	441.86	CL7969Contig1_1_5	sp E0X9F6 THAP1_THAPI Allergen Tha p 1 OS=Thaumatococcus panyocarpa OX=208016 PE=1 SV=1	1.68E-36	Venom allergen
CL7Contig1_1	240.54	CL7Contig1_1_5	sp P34854 NU5M_ANOGA NADH-ubiquinone oxidoreductase chain 5 OS=Anopheles gambiae OX=7165 GN=mt:ND5 PE=3 SV=1	3.21E-63	Metabolism
		CL7Contig5_1_4	sp P34854 NU5M_ANOGA NADH-ubiquinone oxidoreductase chain 5 OS=Anopheles gambiae OX=7165 GN=mt:ND5 PE=3 SV=1	3.21E-63	Metabolism
		CL7Contig5_1_5	sp Q34048 NU4M_CERCA NADH-ubiquinone oxidoreductase chain 4 OS=Ceratitidis capitata OX=7213 GN=ND4 PE=3 SV=2	1.52E-22	Metabolism
CL7Contig5_1	374.81				
CL8006Contig1_1	155.10	CL8006Contig1_1_6	sp Q05974 RAB1A_LYMST Ras-related protein Rab-1A OS=Lymnaea stagnalis OX=6523 GN=RAB1A PE=2 SV=1	1.79E-117	Protein maturation

CL8116Contig1_1	764.67	CL8116Contig1_1_2	sp Q95ZE8 RL14_DROVI 60S ribosomal protein L14 OS=Drosophila virilis OX=7244 GN=RpL14 PE=3 SV=1	1.25E-22	Translation
CL8284Contig1_1	38.48	CL8284Contig1_1_5	sp Q5T5U3 RHG21_HUMAN Rho GTPase-activating protein 21 OS=Homo sapiens OX=9606 GN=ARHGAP21 PE=1 SV=2	7.31E-91	Metabolism
CL9614Contig1_1	527.87	CL9614Contig1_1_4	sp Q962R1 RS18_SPOFR 40S ribosomal protein S18 OS=Spodoptera frugiperda OX=7108 GN=RpS18 PE=2 SV=1	9.45E-85	Translation
CL964Contig1_1	805.42	CL964Contig1_1_1	sp P05389 RLA2_DROME 60S acidic ribosomal protein P2 OS=Drosophila melanogaster OX=7227 GN=RpLP2 PE=1 SV=1	3.52E-23	Translation
k25_Locus_1215_Transcript_6_1	81.19	k25_Locus_1215_Transcript_6_1_4	sp Q9DC11 PXDC2_MOUSE Plexin domain-containing protein 2 OS=Mus musculus OX=10090 GN=Plxdc2 PE=1 SV=1	2.5E-75	Metabolism
k25_Locus_1230_Transcript_1_1	171.98	k25_Locus_1230_Transcript_1_1_6	sp P41822 FRI_AEDAE Ferritin subunit OS=Aedes aegypti OX=7159 GN=FERH PE=1 SV=2	2.41E-42	Host defense
k25_Locus_12664_Transcript_1_1	370.04	k25_Locus_12664_Transcript_1_1_4	sp Q4GXU6 RS4_CARGR 40S ribosomal protein S4 OS=Carabus granulatus OX=118799 GN=RpS4 PE=2 SV=1	5.47E-167	Translation
k25_Locus_12814_Transcript_3_1	1,066.10	k25_Locus_12814_Transcript_3_1_5	sp P35502 ESTF_MYZPE Esterase FE4 OS=Myzus persicae OX=13164 PE=1 SV=1	7.69E-86	Host defense
k25_Locus_1290_Transcript_5_1	521.21	k25_Locus_1290_Transcript_5_1_4	sp P09180 RL4_DROME 60S ribosomal protein L4 OS=Drosophila melanogaster OX=7227 GN=RpL4 PE=1 SV=2	6.98E-170	Translation
k25_Locus_141_Transcript_1_1	35.43	k25_Locus_141_Transcript_1_1_4	sp Q9U943 APLP_LOCFMI Apolipoproteins OS=Locusta migratoria OX=7004 PE=1 SV=2	0	Metabolism
k25_Locus_1447_Transcript_8_1	99.03				
k25_Locus_1480_Transcript_2_1	574.42				
k25_Locus_1556_Transcript_1_1	76.31	k25_Locus_1556_Transcript_1_1_2	sp B4JYU5 TMEDA_DROGR Transmembrane emp24 domain-containing protein bai OS=Drosophila grimshawi OX=7222 GN=bai PE=3 SV=2	4.79E-85	Metabolism
		k25_Locus_1556_Transcript_1_1_5	sp O01735 YC91_CAEEL Uncharacterized MFS-type transporter C09D4.1 OS=Caenorhabditis elegans OX=6239 GN=C09D4.1 PE=3 SV=2	8.14E-119	Metabolism
k25_Locus_1936_Transcript_1_1	79.51	k25_Locus_1936_Transcript_1_1_2	sp O88307 SORL_MOUSE Sortilin-related receptor OS=Mus musculus OX=10090 GN=Sortl1 PE=1 SV=3	0	Metabolism
k25_Locus_1994_Transcript_5_1	144.50				
k25_Locus_201_Transcript_13_1	138.15	k25_Locus_201_Transcript_13_1_6	sp B1A4F7 VDDP4_VESVU Venom dipeptidyl peptidase 4 OS=Vespula vulgaris OX=7454 PE=1 SV=1	4.4E-59	Peptide maturation
		k25_Locus_201_Transcript_13_1_6	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS=Apis mellifera OX=7460 PE=1 SV=1	3.7E-60	Peptide maturation

		k25_Locus_201_Transcript_13_1_6	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS=Apis mellifera OX=7460 PE=1 SV=1	7.68E-79	Peptide maturation
k25_Locus_2163_Transcript_2_1	167.49	k25_Locus_2163_Transcript_2_1_5	sp P36362 CHIT_MANSE Endochitinase OS=Manduca sexta OX=7130 PE=2 SV=1	2.59E-138	Metabolism
		k25_Locus_2163_Transcript_2_1_5	sp P36362 CHIT_MANSE Endochitinase OS=Manduca sexta OX=7130 PE=2 SV=1	1.83E-22	Metabolism
k25_Locus_2444_Transcript_9_1	72.79	k25_Locus_2444_Transcript_9_1_6	sp Q7G192 ALDO2_ARATH Indole-3-acetaldehyde oxidase OS=Arabidopsis thaliana OX=3702 GN=AAO2 PE=1 SV=2	7.39E-163	Metabolism
k25_Locus_2518_Transcript_3_1	116.46	k25_Locus_2518_Transcript_3_1_5	sp Q8VDJ3 VIGLN_MOUSE Vigilin OS=Mus musculus OX=10090 GN=Hdlbp PE=1 SV=1	0	Metabolism
k25_Locus_2595_Transcript_2_1	104.12	k25_Locus_2595_Transcript_2_1_5	sp Q9DBG3 AP2B1_MOUSE AP-2 complex subunit beta OS=Mus musculus OX=10090 GN=Ap2b1 PE=1 SV=1	0	Protein maturation
k25_Locus_2747_Transcript_4_1	432.10	k25_Locus_2747_Transcript_4_1_4	sp B4PEU8 RS9_DROYA 40S ribosomal protein S9 OS=Drosophila yakuba OX=7245 GN=RpS9 PE=2 SV=1	7.84E-111	Translation
k25_Locus_2909_Transcript_4_1	1,284.87	k25_Locus_2909_Transcript_4_1_5	sp P29341 PABP1_MOUSE Polyadenylate-binding protein 1 OS=Mus musculus OX=10090 GN=Pabpc1 PE=1 SV=2	0	Translation
k25_Locus_2955_Transcript_10_1	946.95	k25_Locus_2955_Transcript_10_1_4	sp Q9VRK8 PLBL_DROME Putative phospholipase B-like lamina ancestor OS=Drosophila melanogaster OX=7227 GN=lama PE=2 SV=3	5.5E-66	Metabolism
k25_Locus_3019_Transcript_10_1	97.01	k25_Locus_3019_Transcript_10_1_5	sp Q8BGM5 BEST2_MOUSE Bestrophin-2 OS=Mus musculus OX=10090 GN=Best2 PE=2 SV=1	1.94E-116	
k25_Locus_310_Transcript_6_1	76.09	k25_Locus_310_Transcript_6_1_5	sp O96005 CLPT1_HUMAN Cleft lip and palate transmembrane protein 1 OS=Homo sapiens OX=9606 GN=CLPTM1 PE=1 SV=1	1.57E-175	Metabolism
		k25_Locus_310_Transcript_6_1_6	sp Q5R8Y6 TM9S2_PONAB Transmembrane 9 superfamily member 2 OS=Pongo abelii OX=9601 GN=TM9SF2 PE=2 SV=1	0	Protein maturation
k25_Locus_3218_Transcript_10_1	78.48	k25_Locus_3218_Transcript_10_1_2	sp Q27294 CAZ_DROME RNA-binding protein cabeza OS=Drosophila melanogaster OX=7227 GN=caz PE=2 SV=2	1.67E-38	Translation
		k25_Locus_3218_Transcript_10_1_4	sp Q9IA79 BI1_PAROL Probable Bax inhibitor 1 OS=Paralichthys olivaceus OX=8255 GN=tmbim6 PE=2 SV=1	2.81E-53	Metabolism
k25_Locus_344_Transcript_6_1	294.67	k25_Locus_344_Transcript_6_1_4	sp O15173 PGRC2_HUMAN Membrane-associated progesterone receptor component 2 OS=Homo sapiens OX=9606 GN=PGRMC2 PE=1 SV=1	2.82E-42	Metabolism
k25_Locus_3500_Transcript_4_1	397.76	k25_Locus_3500_Transcript_4_1_6	sp Q7M4F4 CUD1_SCHGR Endocuticle structural glycoprotein SgAbd-1 OS=Schistocerca gregaria OX=7010 PE=1 SV=1	5.78E-22	Metabolism
k25_Locus_3564_Transcript_1_1	451.18	k25_Locus_3564_Transcript_1_1_6	sp Q963B7 RL9_SPOFR 60S ribosomal protein L9 OS=Spodoptera frugiperda OX=7108 GN=RpL9 PE=2 SV=1	9.55E-112	Translation
k25_Locus_358_Transcript_8_1	68.91	k25_Locus_358_Transcript_8_1_5	sp Q5FW12 TIPRL_XENTR TIP41-like protein OS=Xenopus tropicalis OX=8364 GN=tiprl PE=2 SV=1	8.5E-65	Metabolism

		k25_Locus_358_Transcript_8_1_6	sp P29747 CREBA_DROME Cyclic AMP response element-binding protein A OS=Drosophila melanogaster OX=7227 GN=CrebA PE=1 SV=2	8.47E-38	Metabolism
k25_Locus_3667_Transcript_6_1	95.64	k25_Locus_3667_Transcript_6_1_6	sp Q91V92 ACLY_MOUSE ATP-citrate synthase OS=Mus musculus OX=10090 GN=Acly PE=1 SV=1	0	Metabolism
k25_Locus_374_Transcript_240_1	12.60	k25_Locus_374_Transcript_240_1_3	sp Q9VFP2 RDX_DROME Protein roadkill OS=Drosophila melanogaster OX=7227 GN=rdx PE=1 SV=2	0	Metabolism
		k25_Locus_374_Transcript_240_1_4	sp Q03001 DYST_HUMAN Dystonin OS=Homo sapiens OX=9606 GN=DST PE=1 SV=4	4.82E-113	Metabolism
		k25_Locus_374_Transcript_240_1_4	sp Q03001 DYST_HUMAN Dystonin OS=Homo sapiens OX=9606 GN=DST PE=1 SV=4	0	Metabolism
		k25_Locus_374_Transcript_240_1_6	sp P30427 PLEC_RAT Plectin OS=Rattus norvegicus OX=10116 GN=Plec PE=1 SV=2	2.99E-98	Metabolism
k25_Locus_4036_Transcript_3_1	92.25	k25_Locus_4036_Transcript_3_1_4	sp Q7KN62 TERA_DROME Transitional endoplasmic reticulum ATPase TER94 OS=Drosophila melanogaster OX=7227 GN=TER94 PE=1 SV=1	0	Protein maturation
k25_Locus_4255_Transcript_4_1	230.58	k25_Locus_4255_Transcript_4_1_4	sp P47199 QOR_MOUSE Quinone oxidoreductase OS=Mus musculus OX=10090 GN=Cryz PE=1 SV=1	3.05E-4	Translation
k25_Locus_4362_Transcript_8_1	40.45	k25_Locus_4362_Transcript_8_1_6	sp P22700 ATC1_DROME Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type OS=Drosophila melanogaster OX=7227 GN=SERCA PE=1 SV=2	0	Metabolism
k25_Locus_4634_Transcript_5_1	239.77				
k25_Locus_4686_Transcript_1_1	141.53	k25_Locus_4686_Transcript_1_1_5	sp P31401 VATB_MANSE V-type proton ATPase subunit B OS=Manduca sexta OX=7130 GN=VHA55 PE=2 SV=1	0	Metabolism
k25_Locus_4895_Transcript_3_1	301.37	k25_Locus_4895_Transcript_3_1_5	sp P12261 EF1G_ARTSA Elongation factor 1-gamma OS=Artemia salina OX=85549 PE=1 SV=3	0	Translation
k25_Locus_499_Transcript_44_1	359.64	k25_Locus_499_Transcript_44_1_2	sp P0C5Q0 YL54F_YEAST Putative uncharacterized protein YLR154W-F OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=YLR154W-F PE=5 SV=1	6.66E-5	
k25_Locus_50_Transcript_10_1	632.71	k25_Locus_50_Transcript_10_1_5	sp Q9VLW2 KI26L_DROME Kinesin-like protein CG14535 OS=Drosophila melanogaster OX=7227 GN=CG14535 PE=2 SV=2	9.12E-154	Metabolism
k25_Locus_503_Transcript_5_1	3,237.56				
k25_Locus_5123_Transcript_3_1	443.51	k25_Locus_5123_Transcript_3_1_5	sp Q2PQM7 IDGF4_GLOMM Chitinase-like protein Idgf4 OS=Glossina morsitans morsitans OX=37546 GN=Idgf4 PE=2 SV=1	9.53E-94	Metabolism
k25_Locus_5472_Transcript_4_1	204.82	k25_Locus_5472_Transcript_4_1_1	sp P32429 RL7A_CHICK 60S ribosomal protein L7a OS=Gallus gallus OX=9031 GN=RPL7A PE=2 SV=2	3.91E-122	Translation

		k25_Locus_5472_Transcript_4_1_4	sp Q5ZLD7 VPS53_CHICK Vacuolar protein sorting-associated protein 53 homolog OS=Gallus gallus OX=9031 GN=VPS53 PE=2 SV=1	0	Metabolism
k25_Locus_5834_Transcript_1_1	333.98	k25_Locus_5834_Transcript_1_1_6	sp P54399 PDI_DROME Protein disulfide-isomerase OS=Drosophila melanogaster OX=7227 GN=Pdi PE=2 SV=1	0	Protein maturation
k25_Locus_587_Transcript_10_1	35.79	k25_Locus_587_Transcript_10_1_2	sp Q2KJG3 SYNC_BOVIN Asparagine-tRNA ligase. Cytoplasmic OS=Bos taurus OX=9913 GN=NARS PE=2 SV=3	0	Translation
		k25_Locus_587_Transcript_10_1_5	sp Q8TDJ6 DMXL2_HUMAN DmX-like protein 2 OS=Homo sapiens OX=9606 GN=DMXL2 PE=1 SV=2	0	Protein maturation
k25_Locus_60_Transcript_1_1	287.10	k25_Locus_60_Transcript_1_1_3	sp O13532 YL217_YEAST Putative uncharacterized protein YLR217W OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=YLR217W PE=5 SV=1	4.77E-5	
		k25_Locus_60_Transcript_1_1_6	sp P54985 PPIA_BLAGE Peptidyl-prolyl cis-trans isomerase OS=Blattella germanica OX=6973 GN=CYP A PE=2 SV=1	3.93E-92	Protein maturation
k25_Locus_622_Transcript_1_1	168.75	k25_Locus_622_Transcript_1_1_6	sp P17336 CATA_DROME Catalase OS=Drosophila melanogaster OX=7227 GN=Cat PE=1 SV=2	0	Host defense
k25_Locus_638_Transcript_7_1	194.38		U36-MYRTX-Pc1a		Venom peptide
k25_Locus_673_Transcript_10_1	173.12	k25_Locus_673_Transcript_10_1_5	sp P19111 PPBI_BOVIN Intestinal-type alkaline phosphatase OS=Bos taurus OX=9913 GN=ALPI PE=1 SV=2	7.38E-34	Metabolism
		k25_Locus_673_Transcript_10_1_5	sp Q24238 APH4_DROME Alkaline phosphatase 4 OS=Drosophila melanogaster OX=7227 GN=Alp4 PE=2 SV=3	1.63E-57	Metabolism
k25_Locus_7244_Transcript_1_1	457.91	k25_Locus_7244_Transcript_1_1_4	sp P32100 RL7_DROME 60S ribosomal protein L7 OS=Drosophila melanogaster OX=7227 GN=Rpl7 PE=1 SV=2	2.69E-89	Translation
k25_Locus_745_Transcript_21_1	219.43	k25_Locus_745_Transcript_21_1_5	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase AcpH-1 OS=Apis mellifera OX=7460 PE=1 SV=1	5.67E-19	Venom allergen
		k25_Locus_745_Transcript_21_1_6	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase AcpH-1 OS=Apis mellifera OX=7460 PE=1 SV=1	3.88E-45	Venom allergen
k25_Locus_789_Transcript_4_1	173.80	k25_Locus_789_Transcript_4_1_5	sp Q80U59 K0232_MOUSE Uncharacterized protein KIAA0232 OS=Mus musculus OX=10090 GN=Kiaa0232 PE=1 SV=2	7.42E-14	
k25_Locus_80_Transcript_9_1	30.85	k25_Locus_80_Transcript_9_1_3	sp Q8NF91 SYNE1_HUMAN Nesprin-1 OS=Homo sapiens OX=9606 GN=SYNE1 PE=1 SV=4	6.91E-57	Metabolism
k25_Locus_8262_Transcript_1_1	320.63	k25_Locus_8262_Transcript_1_1_6	sp Q4VIT4 PDIA3_CHLAE Protein disulfide-isomerase A3 OS=Chlorocebus aethiops OX=9534 GN=PDIA3 PE=1 SV=1	3.35E-129	Protein maturation
k25_Locus_8271_Transcript_1_1	217.55	k25_Locus_8271_Transcript_1_1_5	sp O35704 SPTC1_MOUSE Serine palmitoyltransferase 1 OS=Mus musculus OX=10090 GN=Sptlc1 PE=1 SV=2	3.23E-124	Metabolism
k25_Locus_843_Transcript_1_1	965.27				

k31_Locus_1168_Transcript_1_1	166.56	k31_Locus_1168_Transcript_1_1_6	sp P41822 FRI_AEDAE Ferritin subunit OS=Aedes aegypti OX=7159 GN=FERH PE=1 SV=2	2.41E-42	Metabolism
k31_Locus_1208_Transcript_7_1	106.18				
k31_Locus_12495_Transcript_1_1	883.01	k31_Locus_12495_Transcript_1_1_4	sp Q6EV04 RS3A_BIPLU 40S ribosomal protein S3a OS=Biphyllus lunatus OX=197003 PE=2 SV=1	3.44E-131	Translation
k31_Locus_13257_Transcript_3_1	339.78	k31_Locus_13257_Transcript_3_1_6	sp Q2KIK3 SIM14_BOVIN Small integral membrane protein 14 OS=Bos taurus OX=9913 GN=SMIM14 PE=3 SV=1	8.72E-14	Metabolism
k31_Locus_133_Transcript_13_1	46.60	k31_Locus_133_Transcript_13_1_3	sp O15013 ARHGA_HUMAN Rho guanine nucleotide exchange factor 10 OS=Homo sapiens OX=9606 GN=ARHGEF10 PE=1 SV=4	1.55E-141	Host defense
		k31_Locus_133_Transcript_13_1_6	sp Q15643 TRIPB_HUMAN Thyroid receptor-interacting protein 11 OS=Homo sapiens OX=9606 GN=TRIP11 PE=1 SV=3	7.7E-40	Protein maturation
k31_Locus_1424_Transcript_4_1	665.37	k31_Locus_1424_Transcript_4_1_5	sp Q5BLY4 ICA_APIME Icarapin-like OS=Apis mellifera OX=7460 PE=2 SV=1	1.9E-20	Venom protein
k31_Locus_1690_Transcript_1_1	507.48	k31_Locus_1690_Transcript_1_1_4	sp Q68KK0 PA1_SOLIN Phospholipase A1 OS=Solenopsis invicta OX=13686 PE=1 SV=1	1.64E-70	Venom allergen
k31_Locus_2082_Transcript_4_1	292.38	k31_Locus_2082_Transcript_4_1_5	sp P36362 CHIT_MANSE Endochitinase OS=Manduca sexta OX=7130 PE=2 SV=1	9.5E-174	Metabolism
k31_Locus_2397_Transcript_7_1	128.83	k31_Locus_2397_Transcript_7_1_1	sp Q1HDZ5 EIF3B_BOMMO Eukaryotic translation initiation factor 3 subunit B OS=Bombyx mori OX=7091 GN=eIF3-S9 PE=2 SV=1	0	Translation
		k31_Locus_2397_Transcript_7_1_4	sp Q5ZHN9 PGPS1_CHICK CDP-diacylglycerol—glycerol-3-phosphate 3-phosphatidyltransferase. Mitochondrial OS=Gallus gallus OX=9031 GN=PGS1 PE=2 SV=1	1.92E-106	Metabolism
k31_Locus_2480_Transcript_5_1	126.90	k31_Locus_2480_Transcript_5_1_6	sp P62154 CALM_LOCMI Calmodulin OS=Locusta migratoria OX=7004 PE=1 SV=2	4.93E-94	Metabolism
k31_Locus_2701_Transcript_2_1	269.72	k31_Locus_2701_Transcript_2_1_6	sp P13060 EF2_DROME Elongation factor 2 OS=Drosophila melanogaster OX=7227 GN=EF2 PE=1 SV=4	0	Translation
k31_Locus_2821_Transcript_1_1	232.67	k31_Locus_2821_Transcript_1_1_4	sp O74503 UAF30_SCHPO Upstream activation factor subunit spp27 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812 GN=spp27 PE=1 SV=1	1.44E-20	Transcription
k31_Locus_2869_Transcript_7_1	830.20		U17-MYRTX-Pc1b		Venom peptide
k31_Locus_3104_Transcript_1_1	1055.57	k31_Locus_3104_Transcript_1_1_4	sp P62909 RS3_RAT 40S ribosomal protein S3 OS=Rattus norvegicus OX=10116 GN=Rps3 PE=1 SV=1	1.11E-130	Translation
k31_Locus_3140_Transcript_7_1	102.29	k31_Locus_3140_Transcript_7_1_1	sp Q27294 CAZ_DROME RNA-binding protein cabeza OS=Drosophila melanogaster OX=7227 GN=caz PE=2 SV=2	1.67E-38	Translation

		k31_Locus_3140_Transcript_7_1_6	sp Q9IA79 BI1_PAROL Probable Bax inhibitor 1 OS=Paralichthys olivaceus OX=8255 GN=tmbim6 PE=2 SV=1	2.81E-53	Metabolism
k31_Locus_3320_Transcript_8_1	277.06	k31_Locus_3320_Transcript_8_1_6	sp Q9Y600 CSAD_HUMAN Cysteine sulfinic acid decarboxylase OS=Homo sapiens OX=9606 GN=CSAD PE=1 SV=2	5.09E-164	Metabolism
k31_Locus_359_Transcript_3_1	177.69	k31_Locus_359_Transcript_3_1_6	sp Q02942 TRF_BLADI Transferrin OS=Blaberus discoidalis OX=6981 PE=1 SV=1	0	Metabolism
k31_Locus_3661_Transcript_4_1	50.82	k31_Locus_3661_Transcript_4_1_1	sp P48814 ADH1_CERCA Alcohol dehydrogenase 1 OS=Ceratitidis capitata OX=7213 GN=ADH1 PE=3 SV=1	3.49E-14	Metabolism
		k31_Locus_3661_Transcript_4_1_4	sp Q8ND56 LS14A_HUMAN Protein LSM14 homolog A OS=Homo sapiens OX=9606 GN=LSM14A PE=1 SV=3	1.43E-48	Translation
k31_Locus_3796_Transcript_9_1	1321.44	k31_Locus_3796_Transcript_9_1_4	sp Q964R0 CP6K1_BLAGE Cytochrome P450 6k1 OS=Blattella germanica OX=6973 GN=CYP6K1 PE=2 SV=1	7.86E-146	Metabolism
k31_Locus_4256_Transcript_5_1	832.68	k31_Locus_4256_Transcript_5_1_4	sp Q9U3U0 RLA0_CERCA 60S acidic ribosomal protein P0 OS=Ceratitidis capitata OX=7213 GN=RpLP0 PE=3 SV=1	2.64E-161	Translation
k31_Locus_436_Transcript_4_1	54.69	k31_Locus_436_Transcript_4_1_4	sp Q29GR8 MOEH_DROPS Moesin/ezrin/radixin homolog 1 OS=Drosophila pseudoobscura pseudoobscura OX=46245 GN=Moe PE=3 SV=3	0	Metabolism
k31_Locus_4459_Transcript_3_1	2,379.72	k31_Locus_4459_Transcript_3_1_1	sp P35778 VA3_SOLIN Venom allergen 3 OS=Solenopsis invicta OX=13686 PE=1 SV=2	2.75E-76	Venom allergen
k31_Locus_458_Transcript_29_1	116.12	k31_Locus_458_Transcript_29_1_1	sp P0C5Q0 YL54F_YEAST Putative uncharacterized protein YLR154W-F OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=YLR154W-F PE=5 SV=1	6.66E-5	
k31_Locus_4995_Transcript_2_1	213.96				
k31_Locus_5153_Transcript_1_1	580.68		U17-MYRTX-Pc1a		Venom peptide
k31_Locus_555_Transcript_3_1	153.40	k31_Locus_555_Transcript_3_1_5	sp Q9CZW4 ACSL3_MOUSE Long-chain-fatty-acid-CoA ligase 3 OS=Mus musculus OX=10090 GN=AcsL3 PE=1 SV=2	0	Metabolism
k31_Locus_5900_Transcript_6_1	99.10				
k31_Locus_7820_Transcript_5_1	171.15	k31_Locus_7820_Transcript_5_1_6	sp Q964R0 CP6K1_BLAGE Cytochrome P450 6k1 OS=Blattella germanica OX=6973 GN=CYP6K1 PE=2 SV=1	2.11E-111	Metabolism
k31_Locus_837_Transcript_10_1	140,113.05		U35-MYRTX-Pc1a		Venom peptide
k31_Locus_9915_Transcript_1_1	368.99	k31_Locus_9915_Transcript_1_1_6	sp A8CAG3 RL17_PHLPP 60S ribosomal protein L17 OS=Phlebotomus papatasi OX=29031 GN=RpL17 PE=2 SV=1	1.54E-101	Translation
k37_Locus_1437_Transcript_11_1	22.64	k37_Locus_1437_Transcript_11_1_1	sp Q14692 BMS1_HUMAN Ribosome biogenesis protein BMS1 homolog OS=Homo sapiens OX=9606 GN=BMS1 PE=1 SV=1	8.59E-131	Translation

		k37_Locus_1437_Transcript_11_1_1	sp Q10751 ACE_CHICK Angiotensin-converting enzyme (Fragment) OS=Gallus gallus OX=9031 GN=ACE PE=2 SV=1	1.17E-150	Metabolism
		k37_Locus_1437_Transcript_11_1_4	sp Q8BTM8 FLNA_MOUSE Filamin-A OS=Mus musculus OX=10090 GN=Flna PE=1 SV=5	3.9E-56	Metabolism
		k37_Locus_1437_Transcript_11_1_6	sp P21333 FLNA_HUMAN Filamin-A OS=Homo sapiens OX=9606 GN=FLNA PE=1 SV=4	2.41E-83	Metabolism
k37_Locus_1633_Transcript_10_1	100.32	k37_Locus_1546_Transcript_4_1_3	sp P31403 VATL_MANSE V-type proton ATPase 16 kDa proteolipid subunit OS=Manduca sexta OX=7130 GN=VHA16 PE=2 SV=1	1.45E-81	Metabolism
k37_Locus_17208_Transcript_1_1	305.05	k37_Locus_1633_Transcript_10_1_1	sp P82147 L2EFL_DROME Protein lethal(2)essential for life OS=Drosophila melanogaster OX=7227 GN=l(2)efl PE=1 SV=1	3.94E-15	Host defense
k37_Locus_1783_Transcript_3_1	59.40	k37_Locus_17208_Transcript_1_1_5	sp O18640 GBLP_DROME Guanine nucleotide-binding protein subunit beta-like protein OS=Drosophila melanogaster OX=7227 GN=Rack1 PE=1 SV=2	0	Translation
		k37_Locus_1783_Transcript_3_1_1	sp P48603 CAPZB_DROME F-actin-capping protein subunit beta OS=Drosophila melanogaster OX=7227 GN=cpb PE=2 SV=1	7.52E-167	Metabolism
k37_Locus_220_Transcript_3_1	123.81	k37_Locus_1783_Transcript_3_1_6	sp P62925 IF5A_SPOFR Eukaryotic translation initiation factor 5A OS=Spodoptera frugiperda OX=7108 GN=eIF-5A PE=1 SV=1	2.67E-97	Translation
k37_Locus_235_Transcript_7_1	25.71	k37_Locus_220_Transcript_3_1_4	sp P82147 L2EFL_DROME Protein lethal(2)essential for life OS=Drosophila melanogaster OX=7227 GN=l(2)efl PE=1 SV=1	1.09E-34	Host defense
k37_Locus_3064_Transcript_4_1	885.57		U17-MYRTX-Pc1b		Venom peptide
k37_Locus_4142_Transcript_12_1	68.07	k37_Locus_4142_Transcript_12_1_4	sp Q9W0A0 DRPR_DROME Protein draper OS=Drosophila melanogaster OX=7227 GN=drpr PE=1 SV=3	0	Host defense
k37_Locus_434_Transcript_24_1	284.63	k37_Locus_434_Transcript_24_1_2	sp P0C5Q0 YL54F_YEAST Putative uncharacterized protein YLR154W-F OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=YLR154W-F PE=5 SV=1	6.66E-5	
k37_Locus_4784_Transcript_3_1	72.33				
k37_Locus_4799_Transcript_3_1	356.68	k37_Locus_4799_Transcript_3_1_5	sp Q6XHI1 THIO2_DROYA Thioredoxin-2 OS=Drosophila yakuba OX=7245 GN=Trx-2 PE=3 SV=1	3.8E-28	Metabolism
k37_Locus_4979_Transcript_3_1	316.11	k37_Locus_4979_Transcript_3_1_5	sp Q2PQM7 IDGF4_GLOMM Chitinase-like protein Idgf4 OS=Glossina morsitans morsitans OX=37546 GN=Idgf4 PE=2 SV=1	1.65E-61	Metabolism
k37_Locus_522_Transcript_5_1	69.41	k37_Locus_522_Transcript_5_1_2	sp Q99J09 MEP50_MOUSE Methylosome protein 50 OS=Mus musculus OX=10090 GN=Wdr77 PE=1 SV=1	7.24E-29	Metabolism
		k37_Locus_522_Transcript_5_1_5	sp Q9W0S7 SND1_DROME Staphylococcal nuclease domain-containing protein 1 OS=Drosophila melanogaster OX=7227 GN=Tudor-SN PE=1 SV=1	0	Transcription
k37_Locus_5370_Transcript_3_1	354.71	k37_Locus_5370_Transcript_3_1_6	sp P13060 EF2_DROME Elongation factor 2 OS=Drosophila melanogaster OX=7227 GN=EF2 PE=1 SV=4	0	Translation

k37_Locus_562_Transcript_3_1	14,337.79		U35-MYRTX-Pc1b		Venom peptide
k37_Locus_6464_Transcript_1_1	544.36				
k37_Locus_6510_Transcript_3_1	227.69	k37_Locus_6510_Transcript_3_1_2	sp Q9GPH3 ATFC_BOMMO Activating transcription factor of chaperone OS=Bombyx mori OX=7091 GN=ATFC PE=2 SV=1	5.5E-24	Transcription
k37_Locus_7415_Transcript_1_1	233.11	k37_Locus_7415_Transcript_1_1_2	sp P08962 CD63_HUMAN CD63 antigen OS=Homo sapiens OX=9606 GN=CD63 PE=1 SV=2	3.29E-29	Host defense
k37_Locus_7727_Transcript_12_1	162.55	k37_Locus_7727_Transcript_12_1_4	sp P51659 DHB4_HUMAN Peroxisomal multifunctional enzyme type 2 OS=Homo sapiens OX=9606 GN=HSD17B4 PE=1 SV=3	0	Metabolism
k37_Locus_801_Transcript_9_1	197.61				
k37_Locus_878_Transcript_3_1	63.45	k37_Locus_878_Transcript_3_1_5	sp P19096 FAS_MOUSE Fatty acid synthase OS=Mus musculus OX=10090 GN=Fasn PE=1 SV=2	0	Metabolism
k37_Locus_9285_Transcript_4_1	65.33	k37_Locus_9285_Transcript_4_1_1	sp P48814 ADH1_CERCA Alcohol dehydrogenase 1 OS=Ceratitis capitata OX=7213 GN=ADH1 PE=3 SV=1	3.49E-14	Metabolism
k37_Locus_996_Transcript_1_1	223.36	k37_Locus_9285_Transcript_4_1_5	sp Q8ND56 LS14A_HUMAN Protein LSM14 homolog A OS=Homo sapiens OX=9606 GN=LSM14A PE=1 SV=3	1.43E-48	Transcription
k43_Locus_1261_Transcript_5_1	90.96	k37_Locus_996_Transcript_1_1_4	sp O13035 SAP_CHICK Prosaposin OS=Gallus gallus OX=9031 GN=PSAP PE=1 SV=1	9.28E-30	Metabolism
k43_Locus_158_Transcript_1_1	508.90	k43_Locus_1580_Transcript_1_1_4	sp P47911 RL6_MOUSE 60S ribosomal protein L6 OS=Mus musculus OX=10090 GN=Rpl6 PE=1 SV=3	8.92E-54	Translation
k43_Locus_17484_Transcript_1_1	583.66	k43_Locus_17484_Transcript_1_1_6	sp P46782 RS5_HUMAN 40S ribosomal protein S5 OS=Homo sapiens OX=9606 GN=RPS5 PE=1 SV=4	6.81E-122	Translation
k43_Locus_18786_Transcript_1_1	666.99	k43_Locus_18786_Transcript_1_1_2	sp Q95V32 RS6_SPOFR 40S ribosomal protein S6 OS=Spodoptera frugiperda OX=7108 GN=RpS6 PE=2 SV=1	4.71E-124	Translation
k43_Locus_1931_Transcript_3_1	268.65	k43_Locus_1931_Transcript_3_1_5	sp P36362 CHIT_MANSE Endochitinase OS=Manduca sexta OX=7130 PE=2 SV=1	9.5E-174	Metabolism
k43_Locus_1965_Transcript_19_1	68.06	k43_Locus_1965_Transcript_19_1_2	sp A0A291PQF1 UGT5_DACCO UDP-glycosyltransferase UGT5 OS=Dactylopius coccus OX=765876 GN=UGT5 PE=1 SV=1	6.23E-10	Peptide maturation
k43_Locus_2002_Transcript_2_1	14,661.25		U3-MYRTX-Pc1e		Venom peptide
k43_Locus_21989_Transcript_1_1	379.98	k43_Locus_21989_Transcript_1_1_5	sp Q962U0 RL13A_SPOFR 60S ribosomal protein L13a OS=Spodoptera frugiperda OX=7108 GN=RpL13A PE=2 SV=1	2.13E-72	Translation
k43_Locus_2691_Transcript_5_1	174.07				
k43_Locus_4209_Transcript_8_1	168.13	k43_Locus_4209_Transcript_8_1_4	sp Q8BYR8 S41A2_MOUSE Solute carrier family 41 member 2 OS=Mus musculus OX=10090 GN=Slc41a2 PE=1 SV=1	2.18E-81	Metabolism

k43_Locus_431_Transcript_14_1	238.95	k43_Locus_431_Transcript_14_1_1	sp P0C5Q0 YL54F_YEAST Putative uncharacterized protein YLR154W-F OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=YLR154W-F PE=5 SV=1	6.66E-5	
k43_Locus_5470_Transcript_2_1	65451.55				
k43_Locus_556_Transcript_1_1	26,546.81				
k43_Locus_7390_Transcript_3_1	405.53	k43_Locus_7390_Transcript_3_1_3	sp P08962 CD63_HUMAN CD63 antigen OS=Homo sapiens OX=9606 GN=CD63 PE=1 SV=2	3.91E-32	Host defense
k49_Locus_1030_Transcript_1_1	53.41	k49_Locus_1030_Transcript_1_1_1	sp Q29L43 MON2_DROPS Protein MON2 homolog OS=Drosophila pseudoobscura pseudoobscura OX=46245 GN=mon2 PE=3 SV=2	0	Protein maturation
k49_Locus_1379_Transcript_16_1	1,254.42	k49_Locus_1379_Transcript_16_1_5	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase AcpH-1 OS=Apis mellifera OX=7460 PE=1 SV=1	4.63E-106	Venom allergen
k49_Locus_14301_Transcript_1_1	13,919.13		U36-MYRTX-Pc1a		Venom peptide
k49_Locus_159_Transcript_39_1	182.74	k49_Locus_159_Transcript_39_1_2	sp B1A4F7 VDDP4_VESVU Venom dipeptidyl peptidase 4 OS=Vespula vulgaris OX=7454 PE=1 SV=1	6.92E-9	Peptide maturation
		k49_Locus_159_Transcript_39_1_2	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS=Apis mellifera OX=7460 PE=1 SV=1	0	Peptide maturation
k49_Locus_2546_Transcript_1_1	136.53	k49_Locus_2546_Transcript_1_1_1	sp Q9QUR2 DCTN4_RAT Dynactin subunit 4 OS=Rattus norvegicus OX=10116 GN=Dctn4 PE=1 SV=1	4.82E-98	Protein maturation
		k49_Locus_2546_Transcript_1_1_5	sp C0HKA1 RS14B_DROME 40S ribosomal protein S14b OS=Drosophila melanogaster OX=7227 GN=RpS14b PE=2 SV=1	7.55E-85	Translation
k49_Locus_3004_Transcript_3_1	550.03		U18-MYRTX-Pc1a		Venom peptide
k49_Locus_4858_Transcript_1_1	514.25	k49_Locus_4858_Transcript_1_1_6	sp Q8TCT9 HM13_HUMAN Minor histocompatibility antigen H13 OS=Homo sapiens OX=9606 GN=HM13 PE=1 SV=1	7.4E-138	Protein maturation
k55_Locus_13572_Transcript_1_1	355.99	k55_Locus_13572_Transcript_1_1_4	sp Q7PQV7 ADT2_ANOGA ADP.ATP carrier protein 2 OS=Anopheles gambiae OX=7165 GN=AGAP002358 PE=3 SV=2	2.95E-173	Metabolism
k55_Locus_2091_Transcript_3_1	139.93				
k55_Locus_3230_Transcript_3_1	199,776.57		U3-MYRTX-Pc1d		Venom peptide
k55_Locus_5233_Transcript_1_1	341.30	k55_Locus_5233_Transcript_1_1_2	sp Q32NQ8 RNF10_XENLA RING finger protein 10 OS=Xenopus laevis OX=8355 GN=rnf10 PE=2 SV=1	3.41E-91	Transcription
		k55_Locus_5233_Transcript_1_1_3	sp P41755 DHE2_ACHKL NAD-specific glutamate dehydrogenase OS=Achlya klebsiana OX=4767 PE=3 SV=1	1.52E-42	Metabolism

		k55_Locus_5233_Transcript_1_1_6	sp Q9U639 HSP7D_MANSE Heat shock 70 kDa protein cognate 4 OS=Manduca sexta OX=7130 PE=2 SV=1	0	Host defense
k55_Locus_600_Transcript_5_1	12.98	k55_Locus_600_Transcript_5_1_5	sp Q03001 DYST_HUMAN Dystonin OS=Homo sapiens OX=9606 GN=DST PE=1 SV=4	0	Metabolism
k55_Locus_6017_Transcript_1_1	146.37	k55_Locus_6017_Transcript_1_1_3	sp Q16P20 CISY2_AEDAE Probable citrate synthase 2. Mitochondrial OS=Aedes aegypti OX=7159 GN=AAEL011789 PE=3 SV=1	0	Metabolism
k55_Locus_9791_Transcript_1_1	505.48	k55_Locus_9791_Transcript_1_1_5	sp Q962U1 RL13_SPOFR 60S ribosomal protein L13 OS=Spodoptera frugiperda OX=7108 GN=RpL13 PE=2 SV=1	1.48E-88	Translation
k61_Locus_10402_Transcript_1_1	2,364.73				
k61_Locus_1164_Transcript_1_21_1	1,026.06	k61_Locus_1164_Transcript_1_21_1_6	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase Acph-1 OS=Apis mellifera OX=7460 PE=1 SV=1	1.22E-105	Venom allergen
k61_Locus_121_Transcript_1_17_1	174.89	k61_Locus_121_Transcript_1_17_1_6	sp B2D0J4 VDPP4_APIME Venom dipeptidyl peptidase 4 OS=Apis mellifera OX=7460 PE=1 SV=1	0	Peptide maturation
k61_Locus_1387_Transcript_1_2_1	12,4881.32				
k61_Locus_173_Transcript_1_1	341.75				
k61_Locus_1813_Transcript_1_1	118.55				
k61_Locus_2144_Transcript_1_4_1	314.24		U17-MYRTX-Pc1b		Venom peptide
k61_Locus_424_Transcript_1_7_1	233.62				
k61_Locus_430_Transcript_1_1	659.09				
k61_Locus_6299_Transcript_1_1	122.65	k61_Locus_6299_Transcript_1_1_2	sp Q16P20 CISY2_AEDAE Probable citrate synthase 2. Mitochondrial OS=Aedes aegypti OX=7159 GN=AAEL011789 PE=3 SV=1	0	Metabolism
k65_Locus_1108_Transcript_1_29_1	18.87	k65_Locus_1108_Transcript_1_29_1_1	sp G5E8K5 ANK3_MOUSE Ankyrin-3 OS=Mus musculus OX=10090 GN=Ank3 PE=1 SV=1	0	Transcription
k65_Locus_1117_Transcript_1_4_1	121.99	k65_Locus_1117_Transcript_1_4_1_3	sp Q9R049 AMFR_MOUSE E3 ubiquitin-protein ligase AMFR OS=Mus musculus OX=10090 GN=Amfr PE=1 SV=2	2.2E-148	Metabolism

Supplementary Table XVIII : Addressing table of major contigs expressed in *Daceton armigerum* venom glands.

Contig Name	TPMs	Contig ORF	Seq ID Name	E-value	Function
CL1058Contig1_1	385.52				

CL114Contig1_1	969.15	CL114Contig1_1_3	sp Q91PP5 POLG_RTSVT Genome polyprotein OS=Rice tungro spherical virus (strain Vt6) OX=337080 PE=1 SV=1	6.26E-24	Transcription
CL114Contig2_1	561.49	CL114Contig2_1_5	sp Q91PP5 POLG_RTSVT Genome polyprotein OS=Rice tungro spherical virus (strain Vt6) OX=337080 PE=1 SV=1	6.35E-33	Transcription
CL1189Contig1_1	309.25				
CL1260Contig1_1	433.93				
CL1331Contig1_1	988.51				
CL1378Contig1_1	318.00				
CL183Contig1_1	325.39				
CL19Contig1_1	284.52				
CL19Contig4_1	804.76				
CL259Contig1_3	1,131.75				
CL2Contig6_1	420.77				
CL2Contig7_1	330.37				
CL315Contig1_1	315.23				
CL338Contig1_1	396.39	CL338Contig1_1_6	sp Q6UP17 POLN_EOPV Non-structural polyprotein OS=Ectropis obliqua picorna-like virus OX=240555 PE=1 SV=1	1.16E-15	Transcription
CL365Contig1_1	9,155.40	CL365Contig1_1_1	sp O13067 CND2_XENLA Condensin complex subunit 2 OS=Xenopus laevis OX=8355 GN=ncaph PE=1 SV=1	5.44E-57	Metabolism
		CL365Contig1_1_5	sp Q0J2B5 APM1C_ORYSJ Aminopeptidase M1-C OS=Oryza sativa subsp. japonica OX=39947 GN=Os09g0362500 PE=2 SV=2	1.95E-4	Metabolism
CL396Contig1_1	16,535.34				
CL42Contig4_1	250.22				
CL42Contig5_1	224.96				
CL511Contig1_1	33,333.90				
CL511Contig1_2	41,848.84	CL511Contig1_2_5	sp Q8TGM7 ART2_YEAST Putative uncharacterized protein ART2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=ART2 PE=5 SV=1	2.33E-14	
CL51Contig3_1	590.70				
CL620Contig1_1	612.96	CL620Contig1_1_1	sp Q91PP5 POLG_RTSVT Genome polyprotein OS=Rice tungro spherical virus (strain Vt6) OX=337080 PE=1 SV=1	2.42E-10	Transcription

CL620Contig1_2	3,804.30	CL620Contig1_2_3	sp Q9DSN9 POLN_ABPVR Replicase polyprotein OS=Acute bee paralysis virus (strain Rothamsted) OX=1217067 GN=ORF1 PE=4 SV=1	3.07E-41	Transcription
CL630Contig1_1	13,264.70				
CL63Contig2_1	141,924.80		U32-MYRTX-Da1a		Venom peptide
CL66Contig1_1	21,188.69				
CL66Contig2_1	167,208.06				
CL76Contig1_1	765.71	CL76Contig1_1_1	sp Q91PP5 POLG_RTSVT Genome polyprotein OS=Rice tungro spherical virus (strain Vt6) OX=337080 PE=1 SV=1	1.38E-51	Transcription
CL866Contig1_1	392.78				
CL917Contig1_1	39,215.83				
CL917Contig1_3	418.16				
CL949Contig1_1	739.50		U32-MYRTX-Da1a		Venom peptide
CL963Contig1_1	392.34				
TRINITY_DN11065_c0_g1_i1_1	470.38				
TRINITY_DN21142_c0_g1_i1_1	1,283.38	TRINITY_DN21142_c0_g1_i1_1_2	sp O96647 RL10_BOMMA 60S ribosomal protein L10 OS=Bombyx mandarina OX=7092 GN=RpL10 PE=2 SV=1	1.48E-127	Translation
TRINITY_DN21571_c0_g1_i1_1	1,349.17	TRINITY_DN21571_c0_g1_i1_1_6	sp Q9VQ62 NPC2_DROME NPC intracellular cholesterol transporter 2 homolog a OS=Drosophila melanogaster OX=7227 GN=Npc2a PE=2 SV=1	5.41E-15	Metabolism
TRINITY_DN235_c0_g1_i1_1	2,099.21	TRINITY_DN235_c0_g1_i1_1_1	sp Q5G5C4 RS3A_PERAM 40S ribosomal protein S3a OS=Periplaneta americana OX=6978 GN=Parcxpwex01 PE=2 SV=1	6.85E-143	Translation
TRINITY_DN235_c0_g1_i2_1	1,317.66	TRINITY_DN235_c0_g1_i2_1_3	sp Q6EV04 RS3A_BIPLU 40S ribosomal protein S3a OS=Biphyllus lunatus OX=197003 PE=2 SV=1	2.59E-131	Translation
TRINITY_DN23662_c0_g1_i1_1	1,625.12		U32-MYRTX-Da1c		Venom peptide
TRINITY_DN2978_c0_g1_i2_1	303.85				
TRINITY_DN3162_c0_g1_i1_1	219.28				
TRINITY_DN3926_c0_g1_i2_1	732.57				
TRINITY_DN3928_c0_g1_i1_1	1,594.39				

TRINITY_DN4030_c0_g1_i1_1	438.71	TRINITY_DN4030_c0_g1_i1_1_4	sp Q02748 IF4A_DROME Eukaryotic initiation factor 4A OS=Drosophila melanogaster OX=7227 GN=eIF4A PE=1 SV=3	0	Translation
TRINITY_DN4030_c0_g1_i2_1	313.30	TRINITY_DN4030_c0_g1_i2_1_6	sp Q02748 IF4A_DROME Eukaryotic initiation factor 4A OS=Drosophila melanogaster OX=7227 GN=eIF4A PE=1 SV=3	0	Translation
TRINITY_DN4101_c0_g1_i1_1	366.55				
TRINITY_DN4111_c0_g1_i1_1	328.25				
TRINITY_DN4120_c0_g1_i1_1	258.17				
TRINITY_DN4249_c0_g1_i1_1	1,268.03	TRINITY_DN4249_c0_g1_i1_1_1	sp Q8MUR4 RL13A_CHOPR 60S ribosomal protein L13a OS=Choristoneura parallela OX=106495 GN=RpL13A PE=2 SV=1	3.89E-107	Translation
TRINITY_DN4348_c0_g1_i1_1	326.24				
TRINITY_DN4416_c0_g1_i2_1	680.63	TRINITY_DN4416_c0_g1_i2_1_4	sp Q9U3U0 RLA0_CERCA 60S acidic ribosomal protein P0 OS=Ceratitits capitata OX=7213 GN=RpLP0 PE=3 SV=1	9.75E-159	Translation
TRINITY_DN4468_c0_g1_i2_1	318.45				
TRINITY_DN4519_c0_g1_i1_1	430.41				
TRINITY_DN4631_c0_g1_i1_1	236.12				
TRINITY_DN4656_c0_g1_i2_1	447.85				
TRINITY_DN4754_c0_g1_i1_1	358.79				
TRINITY_DN4754_c0_g1_i2_1	237.69				
TRINITY_DN4845_c0_g1_i1_1	225.96				
TRINITY_DN5009_c0_g1_i2_1	405.79				
TRINITY_DN5068_c0_g1_i1_1	608.26				
TRINITY_DN5122_c0_g1_i1_1	280.77				
TRINITY_DN5137_c0_g1_i1_1	361.14				
TRINITY_DN5239_c0_g1_i1_1	489.27				
TRINITY_DN5243_c0_g1_i1_1	361.30				

TRINITY_DN5401_c0_g1_i1_1	455.62				
TRINITY_DN5569_c0_g1_i1_1	8162.46				
TRINITY_DN5794_c0_g1_i1_1	217.27				
TRINITY_DN5825_c0_g1_i1_1	386.36	TRINITY_DN5825_c0_g1_i1_1_3	sp P13060 EF2_DROME Elongation factor 2 OS=Drosophila melanogaster OX=7227 GN=EF2 PE=1 SV=4	0	Translation
TRINITY_DN5881_c0_g1_i2_1	407.19				
TRINITY_DN5889_c0_g1_i1_1	444.78				
TRINITY_DN5998_c0_g1_i1_1	816.90				
TRINITY_DN5998_c0_g1_i2_1	930.76				
TRINITY_DN6008_c0_g2_i1_1	403.55				
TRINITY_DN6008_c0_g2_i2_1	2208.07	TRINITY_DN6008_c0_g2_i2_1_4	sp Q94624 RS6_MANSE 40S ribosomal protein S6 OS=Manduca sexta OX=7130 GN=RpS6 PE=2 SV=1	4.94E-124	Translation
TRINITY_DN6037_c0_g1_i1_1	556.56				
TRINITY_DN6083_c0_g2_i2_1	290.56				
TRINITY_DN6088_c0_g1_i1_1	4293.59				
TRINITY_DN6252_c0_g1_i1_1	692.80				
TRINITY_DN641_c0_g1_i2_1	327.99				
TRINITY_DN6414_c0_g1_i1_1	575.05				
TRINITY_DN6414_c0_g1_i3_1	484.29				
TRINITY_DN6456_c0_g1_i1_1	1,070.14				
TRINITY_DN6532_c0_g1_i1_1	1,494.28	TRINITY_DN6532_c0_g1_i1_1_3	sp Q95V39 RL8_SPOFR 60S ribosomal protein L8 OS=Spodoptera frugiperda OX=7108 GN=RpL8 PE=2 SV=1	1.56E-157	Translation
TRINITY_DN6646_c1_g1_i1_1	239.69				
TRINITY_DN6917_c0_g1_i2_1	878.32				

TRINITY_DN6928_c0_g1_i1_1	474.39				
TRINITY_DN6980_c0_g1_i1_1	944.65	TRINITY_DN6980_c0_g1_i1_1_1	sp P32100 RL7_DROME 60S ribosomal protein L7 OS=Drosophila melanogaster OX=7227 GN=RpL7 PE=1 SV=2	5.63E-101	Translation
TRINITY_DN7103_c1_g2_i1_1	316.89				
TRINITY_DN7260_c0_g1_i1_1	370.42				
TRINITY_DN7260_c0_g2_i1_1	1,532.15	TRINITY_DN7260_c0_g2_i1_1_5	sp P31009 RS2_DROME 40S ribosomal protein S2 OS=Drosophila melanogaster OX=7227 GN=RpS2 PE=1 SV=2	1.89E-131	Translation
TRINITY_DN7291_c0_g1_i1_1	406.05				
TRINITY_DN7310_c0_g1_i8_1	219.99				
TRINITY_DN7364_c1_g1_i1_1	228.01				
TRINITY_DN7494_c0_g2_i1_1	1,868.46	TRINITY_DN7494_c0_g2_i1_1_2	sp P09180 RL4_DROME 60S ribosomal protein L4 OS=Drosophila melanogaster OX=7227 GN=RpL4 PE=1 SV=2	3.44E-173	Translation
TRINITY_DN7521_c0_g1_i4_2	426.46				
TRINITY_DN7534_c1_g1_i2_1	454.73				
TRINITY_DN7595_c0_g1_i1_1	1,112.07	TRINITY_DN7595_c0_g1_i1_1_4	sp P47911 RL6_MOUSE 60S ribosomal protein L6 OS=Mus musculus OX=10090 GN=Rpl6 PE=1 SV=3	2.36E-55	Translation
TRINITY_DN7597_c0_g1_i1_1	1,209.49	TRINITY_DN7597_c0_g1_i1_1_2	sp Q0PXX8 RSSA_DIACI 40S ribosomal protein SA OS=Diaphorina citri OX=121845 PE=2 SV=1	2.42E-125	Translation
TRINITY_DN7608_c0_g1_i3_1	293.26				
TRINITY_DN7641_c0_g1_i2_1	622.10				
TRINITY_DN7641_c0_g1_i2_2	818.84	TRINITY_DN7641_c0_g1_i2_1_2	sp P29520 EF1A_BOMMO Elongation factor 1-alpha OS=Bombyx mori OX=7091 PE=2 SV=1	0	Translation
TRINITY_DN7672_c0_g1_i1_1	236.84				
TRINITY_DN7721_c0_g1_i1_1	267.15				
TRINITY_DN7804_c0_g4_i1_1	209.45				
TRINITY_DN7871_c6_g2_i1_1	525.75	TRINITY_DN7871_c6_g2_i1_1_5	sp Q6PDY2 AEDO_MOUSE 2-aminoethanethiol dioxygenase OS=Mus musculus OX=10090 GN=Ado PE=1 SV=2	3.59E-26	Metabolism

TRINITY_DN7873_c0_g1_i2_1	349.91				
TRINITY_DN7904_c0_g2_i1_2	313.34				
TRINITY_DN7905_c0_g1_i5_1	9,375.46	TRINITY_DN7905_c0_g1_i5_1_1	sp P70645 BLMH_RAT Bleomycin hydrolase OS=Rattus norvegicus OX=10116 GN=Blmh PE=1 SV=1	1.22E-163	Host defense
TRINITY_DN7944_c0_g6_i1_1	261.65				
TRINITY_DN7954_c0_g1_i1_2	513.60				
TRINITY_DN7989_c0_g1_i1_1	918.70	TRINITY_DN7989_c0_g1_i1_1_3	sp Q2TBQ5 RL7A_BOVIN 60S ribosomal protein L7a OS=Bos taurus OX=9913 GN=RPL7A PE=2 SV=3	1.17E-121	Translation
TRINITY_DN8124_c3_g3_i2_1	440.96				
TRINITY_DN8132_c0_g1_i1_1	1,263.16				
TRINITY_DN8132_c0_g1_i5_1	1,288.92				
TRINITY_DN8132_c0_g1_i6_1	1,918.12				
TRINITY_DN8132_c0_g3_i1_2	430.35				
TRINITY_DN8138_c3_g7_i1_1	309.42				
TRINITY_DN8173_c2_g1_0_i1_1	73,677.50				
TRINITY_DN8173_c2_g4_i1_1	2,515.81				
TRINITY_DN8173_c2_g8_i1_1	33,807.83				
TRINITY_DN8173_c3_g3_i2_1	5,7166.36				
TRINITY_DN8190_c0_g1_i10_1	227.49				
TRINITY_DN8190_c0_g1_i13_1	249.32				
TRINITY_DN8213_c7_g5_i2_1	320.78				
TRINITY_DN8220_c4_g1_i2_1	476.49				
TRINITY_DN8286_c7_g1_i1_1	1,367.45	TRINITY_DN8286_c7_g1_i1_1_4	sp Q5BLY5 ACPH1_APIME Venom acid phosphatase Acph-1 OS=Apis mellifera OX=7460 PE=1 SV=1	1.73E-92	Venom allergen

TRINITY_DN8286_c7_g1_i3_1	449.80				
TRINITY_DN8298_c6_g2_i2_1	267.27				
TRINITY_DN8357_c4_g4_i1_1	420.65				
TRINITY_DN8363_c1_g1_i2_1	510.95				
TRINITY_DN8363_c1_g4_i8_1	242.89				
TRINITY_DN8368_c2_g1_i1_1	17,112.72				
TRINITY_DN8407_c0_g1_i1_2	648.21				
TRINITY_DN8420_c2_g6_i1_1	420.49				
TRINITY_DN8422_c0_g1_i1_1	230.72				
TRINITY_DN8452_c4_g2_i1_1	18,809.08				
TRINITY_DN8482_c4_g4_i2_1	317.33				
TRINITY_DN8485_c1_g7_i2_1	818.56				
TRINITY_DN8492_c0_g1_i9_1	560.74				
TRINITY_DN8494_c4_g1_i1_1	457.72	TRINITY_DN8494_c4_g1_i1_1_5	sp Q75VN3 TCTP_BOMMO Translationally-controlled tumor protein homolog OS=Bombyx mori OX=7091 GN=Tctp PE=2 SV=1	7.8E-83	
TRINITY_DN8494_c4_g1_i2_1	507.00	TRINITY_DN8494_c4_g1_i2_1_5	sp Q75VN3 TCTP_BOMMO Translationally-controlled tumor protein homolog OS=Bombyx mori OX=7091 GN=Tctp PE=2 SV=1	1.9E-90	
TRINITY_DN8503_c0_g1_i1_1	1,484.80				
TRINITY_DN8517_c3_g1_i1_1	523.43				
TRINITY_DN8556_c0_g1_i9_1	240.75				
TRINITY_DN8568_c0_g1_i1_1	503.83				
TRINITY_DN8568_c0_g1_i2_1	380.81				
TRINITY_DN8653_c1_g1_i1_1	211.87				

TRINITY_DN8653_c1_g1_i12_1	223.70				
TRINITY_DN8654_c4_g2_i1_1	992.33				
TRINITY_DN8654_c4_g4_i1_1	6,529.57				
TRINITY_DN8681_c5_g1_i14_1	12,130.05				
TRINITY_DN8681_c5_g1_i6_1	15,300.94				
TRINITY_DN8720_c0_g1_i1_1	2,229.90				
TRINITY_DN8738_c2_g2_i1_1	257.88				
TRINITY_DN8749_c1_g1_i1_1	545.65				
TRINITY_DN8765_c1_g3_i1_1	257.52				
TRINITY_DN8778_c1_g1_i17_2	436.26				
TRINITY_DN8823_c3_g1_i1_1	293.18				
TRINITY_DN8829_c2_g2_i1_1	227.42				
TRINITY_DN8845_c2_g1_i1_1	8,320.91				
TRINITY_DN8851_c8_g1_i1_1	3,600.40	TRINITY_DN8851_c8_g1_i1_1_1	sp Q36423 NU4LM_LOCFMI NADH-ubiquinone oxidoreductase chain 4L OS=Locusta migratoria OX=7004 GN=ND4L PE=3 SV=1	1.00E-3	
		TRINITY_DN8851_c8_g1_i1_1_2	sp Q34048 NU4M_CERCA NADH-ubiquinone oxidoreductase chain 4 OS=Ceratitidis capitata OX=7213 GN=ND4 PE=3 SV=2	3.26E-7	
		TRINITY_DN8851_c8_g1_i1_1_2	sp P18931 NU4M_DROME NADH-ubiquinone oxidoreductase chain 4 OS=Drosophila melanogaster OX=7227 GN=mt:ND4 PE=3 SV=3	3.42E-23	
		TRINITY_DN8851_c8_g1_i1_1_2	sp P18932 NU5M_DROME NADH-ubiquinone oxidoreductase chain 5 OS=Drosophila melanogaster OX=7227 GN=mt:ND5 PE=3 SV=4	1.68E-38	
		TRINITY_DN8851_c8_g1_i1_1_3	sp P24875 NU1M_ASCSU NADH-ubiquinone oxidoreductase chain 1 OS=Ascaris suum OX=6253 GN=ND1 PE=3 SV=2	7.8E-12	
		TRINITY_DN8851_c8_g1_i1_1_4	sp Q37600 COX3_PYLLI Cytochrome c oxidase subunit 3 OS=Pyliella littoralis OX=2885 GN=COX3 PE=3 SV=1	1.55E-16	
TRINITY_DN8851_c8_g3_i1_1	13,990.26				

TRINITY_DN8854_c4_g4 _i1_1	23,304.49				
TRINITY_DN927_c0_g2_ i1_1	484.04				

Études de peptides de venins de fourmis : diversité moléculaire et lien avec la fonction immunitaire

Résumé :

Les venins d'animaux sont des bibliothèques naturelles de composés bioactifs optimisés au cours de l'évolution, appelés toxines. Les venins de nombreux animaux restent néanmoins inexploités, notamment ceux des insectes. Plusieurs études portant sur les venins de fourmis ont révélé que ces venins étaient riches en peptides. La caractérisation du peptidome du venin de *Tetramorium bicarinatum* a également permis de constater que, malgré la diversité de peptides matures, ces derniers se classent en 3 grandes familles de précurseurs dont certaines ont déjà été décrites chez d'autres hyménoptères. Il est de plus apparu que des gènes codant certains d'entre eux s'expriment en dehors du système vulnérant. Ces résultats posent les questions des mécanismes impliqués dans la diversification des toxines peptidiques de venins de fourmis, ainsi que leur rôle en dehors de la fonction venimeuse. Pour répondre à ces problématiques, la première partie de ce travail de thèse a consisté en la caractérisation *via* des approches protéotranscriptomiques, des venins de 7 espèces de fourmis appartenant aux différentes tribus phylogénétiques de la sous-famille des *Myrmicinae*, et du venin d'une espèce appartenant à une sous-famille proche, les *Pseudomyrmecinae*. Cent toxines peptidiques aux structures variées ont ainsi été identifiées et classées en 8 superfamilles de précurseurs. La seconde partie a consisté en l'exploration du lien entre les toxines peptidiques du venin de *T. bicarinatum* et son immunité innée *via* des méthodes de biologie moléculaire et cellulaire. La présence de transcrits codant certains peptides a été vérifiée dans des organes impliqués dans l'immunité innée (i.e. corps gras, tubes digestifs). L'expression des gènes les codant a également été évaluée suite à une infection bactérienne. Il a ainsi été montré que les transcrits codant les peptides de venin sélectionnés sont présents dans les organes testés, et que certains sont produits dans les corps gras en réponse à une infection bactérienne. Ces résultats confirment l'existence d'un lien entre les peptides de venin et l'immunité innée de la fourmi *T. bicarinatum*, bien que des études complémentaires soient nécessaires.

Abstract:

Animal venoms are natural libraries of bioactive compounds, called toxins, which have been fine-tuned through the course of evolution. However, numerous venomous organisms are still neglected, especially venomous insects. Several studies of ant venoms revealed that they were peptide-rich. Furthermore, the characterization of the ant *Tetramorium bicarinatum* venom peptidome revealed that, despite the diversity of mature peptides, they belonged to 3 superfamilies of precursors, some of which have already been described in other aculeate hymenoptera. This study also observed that genes encoding some of them were expressed outside the venom apparatus. These results raise questions about the mechanisms involved in the diversification of peptide toxins from ant venoms, as well as their role apart from the venomous function. To address these issues, the first part of this thesis work consisted in the characterization *via* proteotranscriptomics approaches of 7 venoms from ants belonging to the different phylogenetic tribes of the *Myrmicinae* subfamily, and of the venom of one species belonging to a close subfamily, the *Pseudomyrmecinae*. A total of 100 peptide toxins with various structures were thus identified and classified into 8 precursor superfamilies. The second part explored the link between peptide toxins of *T. bicarinatum* venom and its innate immunity *via* molecular and cellular biology methods. The presence of transcripts encoding certain peptides have been verified in organs which are involved in innate immunity of insects (i.e. fat bodies, digestive tracts). The expression of the genes encoding them has also been evaluated following a bacterial infection. It has thus been shown that the transcripts encoding the selected venom peptides are present in the organs tested, and that some are produced in fat bodies in response to a bacterial infection. These results confirm the existence of a link between the venom peptides and the innate immunity of the ant *T. bicarinatum*, although further studies are needed.