toxins

MDPI

Article

Venomics of the Central European Myrmicine Ants *Myrmica rubra* and *Myrmica ruginodis*

Sabine Hurka [1,2], Karina Brinkrolf [3], Rabia Özbek [4], Frank Förster [3], André Billion [4], John Heep [1], Thomas Timm [5], Günter Lochnit [5], Andreas Vilcinskas [1,2,4] and Tim Lüddecke [2,4,*]

1 Institute for Insect Biotechnology, Justus Liebig University Giessen, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany; sabine.hurka@innere.med.uni-giessen.de (S.H.); john.heep@web.de (J.H.); andreas.vilcinskas@ime.fraunhofer.de (A.V.)
2 LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Senckenberganlage 25, 60325 Frankfurt, Germany
3 Bioinformatics and Systems Biology, Justus Liebig University Giessen, Heinrich-Buff-Ring 58, 35392 Giessen, Germany; karina.brinkrolf@bio.uni-giessen.de (K.B.); frank.foerster@computational.bio.uni-giessen.de (F.F.)
4 Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Ohlebergsweg 12, 35392 Giessen, Germany; ozbekrabia@gmail.com (R.Ö.); andre.billion@ime.fraunhofer.de (A.B.)
5 Institute of Biochemistry, Justus Liebig University Giessen, Friedrichstraße 24, 35392 Giessen, Germany; thomas.timm@biochemie.med.uni-giessen.de (T.T.); guenter.lochnit@biochemie.med.uni-giessen.de (G.L.)
* Correspondence: tim.lueddecke@ime.fraunhofer.de

Abstract: Animal venoms are a rich source of novel biomolecules with potential applications in medicine and agriculture. Ants are one of the most species-rich lineages of venomous animals. However, only a fraction of their biodiversity has been studied so far. Here, we investigated the venom components of two myrmicine (subfamily Myrmicinae) ants: *Myrmica rubra* and *Myrmica ruginodis*. We applied a venomics workflow based on proteotranscriptomics and found that the venoms of both species are composed of several protein classes, including venom serine proteases, cysteine-rich secretory protein, antigen 5 and pathogenesis-related 1 (CAP) superfamily proteins, Kunitz-type serine protease inhibitors and venom acid phosphatases. Several of these protein classes are known venom allergens, and for the first time we detected phospholipase A1 in the venom of *M. ruginodis*. We also identified two novel epidermal growth factor (EGF) family toxins in the *M. ruginodis* venom proteome and an array of additional EGF-like toxins in the venom gland transcriptomes of both species. These are similar to known toxins from the related myrmicine ant, *Manica rubida*, and the myrmecine (subfamily Myrmeciinae) Australian red bulldog ant *Myrmecia gullosa*, and are possibly deployed as weapons in defensive scenarios or to subdue prey. Our work suggests that *M. rubra* and *M. ruginodis* venoms contain many enzymes and other high-molecular-weight proteins that cause cell damage. Nevertheless, the presence of EGF-like toxins suggests that myrmicine ants have also recruited smaller peptide components into their venom arsenal. Although little is known about the bioactivity and function of EGF-like toxins, their presence in myrmicine and myrmecine ants suggests they play a key role in the venom systems of the superfamily Formicoidea. Our work adds to the emerging picture of ant venoms as a source of novel bioactive molecules and highlights the need to incorporate such taxa in future venom bioprospecting programs.

Keywords: insect venom; proteotranscriptomics; biodiscovery; allergens; phospholipase A1; EGF-like toxins

Key Contribution: We describe the venom components of two ant species representing the overlooked subfamily Myrmicinae and have identified novel putative toxins with an EGF-like fold.

check for updates

Citation: Hurka, S.; Brinkrolf, K.; Özbek, R.; Förster, F.; Billion, A.; Heep, J.; Timm, T.; Lochnit, G.; Vilcinskas, A.; Lüddecke, T. Venomics of the Central European Myrmicine Ants *Myrmica rubra* and *Myrmica ruginodis*. *Toxins* **2022**, *14*, 358. https://doi.org/10.3390/toxins14050358

Received: 22 April 2022
Accepted: 19 May 2022
Published: 21 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Ants (Formicidae) are a species-rich family within the hyperdiverse insect order Hymenoptera. They emerged in the Cretaceous about 130 million years ago, and have diversified into more than 14,000 species distributed across 17 subfamilies [1,2]. They have conquered all continents and represent arguably one of the most successful lineages among terrestrial animals [3]. Although most ant species are found in the tropics, remarkable diversity is also present in more temperate regions [1]. For example, 108 genera are native to central Europe, most of which are members of the subfamily Myrmicinae (52% of European ant species at time of writing) [1]. The evolutionary success of ants is based upon an unprecedented array of ecological innovations, such as their eusociality, foraging strategies [3], and the venom systems present in several ant subfamilies [4].

Venoms are complex mixtures of proteins, peptides and organic molecules, and are found throughout the animal kingdom [5]. Venomous animals inject venom into other animals to disrupt important physiological processes [6]. The detrimental effects of venom mainly reflect the activities of proteins and peptides collectively described as toxins. Across all animal groups, venom serves the three major functions of predation, defense and competitor deterrence [6], as well as many secondary functions, such as sexual communication, prey tracking and maintenance of the immune system [7]. Venom toxins have been subject to millions of years of natural selection, becoming refined molecular weapons that act with unmatched selectivity and potency [5]. The venom system of ants is found at the anterior end of the abdomen and is derived from an ancient ovipositor [8]. Ants primarily use their venoms to overpower prey and defend their colony against threats [4]. Ant venoms are mainly composed of polypeptides and small organic molecules, particularly alkaloids [4,9–11].

Although there are several venomous ant lineages, only a few ant venoms have been studied in detail [12–18]. This is primarily because small arthropods deliver only miniscule amounts of venom, so it may be necessary to sample hundreds of specimens to accumulate sufficient amounts of crude venom for traditional analytical platforms [4,19,20]. This challenge has been addressed by the new research field of venomics, where modern systems biology approaches are used to disentangle venom components, often by combining proteomics and transcriptomics in an approach known as proteotranscriptomics [21]. Given the specificity and sensitivity of such modern platforms, small samples are sufficient for analysis and venoms can now be explored across the entire animal kingdom [21].

The application of venomics provides valuable insights into the chemical diversity of venoms from many underrepresented groups of venomous animals, especially arthropods, but ant venoms have still been largely overlooked [22]. Only a handful of proteotranscriptomics studies have been conducted on ant venoms, so a significant knowledge gap remains [23–27]. Furthermore, the studies conducted thus far have mostly focused on enigmatic species such as the bullet ant (*Paraponera clavata*) and the red bulldog ant (*Myrmecia gullosa*), two atypically large ant species with potent nociceptive venoms [15,22]. Proteotranscriptomics has also been applied to *Manica rubida*, one of the larger species of myrmicine ants from central Europe [14]. Beyond these prominent species, there has been little progress in the investigation of ant venoms using modern approaches.

To fill this important knowledge gap, we investigated the venoms of *Myrmica rubra* and *Myrmica ruginodis*, two smaller myrmicine ants that are ubiquitous in central Europe [1]. We applied our modern venomics workflow to identify the venom components, including novel toxins of the epidermal growth factor (EGF)-like family. We discuss our findings in relation to data available from other ant species, including *M. rubida*. This work advances our understanding of ant venoms and can be used as the basis for further studies on the venomics of ants and other arthropod lineages.

2. Results

2.1. Verification of Species Identity

We conducted initial morphological examinations to ensure that the ant specimens were correctly identified. We measured the diameter of the scapus base and the scapus length–head width ratio to distinguish *M. rubra* and *M. ruginodis* from other Myrmicinae [25]. We also considered features such as the petiolus shape, postpetiolus posture and propodeal spines to differentiate between *M. rubra* and *M. ruginodis*. As a second line of evidence, we retrieved the mitochondrial *cytochrome oxidase 1* (*CO1*) gene, a common barcoding gene for animals (Supplementary Tables S1 and S2) and compared them to available barcodes for species of Formicidae. The retrieved *CO1* sequences matched published *M. rubra* and *M. ruginodis* barcodes, thus validating our morphological examination.

2.2. Proteomic and Transcriptomic Landscapes

We used our bespoke proteotranscriptomics workflow (Figure 1) to investigate the venom components of both ant species. This generated 69,929,174 sequencing reads for *M. rubra* and 69,617,668 for *M. ruginodis* (69,645,529 and 69,318,655, respectively, after trimming and quality control). The clustered assemblies contained 512,855 (*M. rubra*) and 466,158 (*M. ruginodis*) unique contigs. We mapped the trimmed reads onto the two assemblies, revealing minimum mapping rates of 96.30%. BUSCO results showed that at least 94.2% of the expected hymenopteran genes were present (*M. rubra* C:95.5% [S:8.8%, D:86.7%], F:1.6%, M:2.9%, n:5991; *M. ruginodis* C:94.2% [S:8.8%, D:85.4%], F:2.0%, M:3.8%, n:5991).

De novo transcriptomes for venom glands are known to overestimate the diversity of toxin transcripts and produce false-positive sequences [28–30]. This phenomenon is likely to be amplified by the use of multiple assemblers. Therefore, we added proteomic verification to our analysis and used our clustered assembly as a database for peptide searches. For the peptide searches, Mascot assigned the spectra to 2085 (*M. rubra*) and 1695 (*M. ruginodis*) different open reading frames (ORFs). We removed 1797 (*M. rubra*) and 1314 (*M. ruginodis*) spectra from the dataset using our filtering protocol (Section 5.6). All remaining hits were covered by at least two peptides. When combining data from both species, the molecular weight of venom components ranged from 2.6 to 377.8 kDa, with sequence lengths of 21–3392 amino acids. We annotated 255 of the 288 candidates in *M. rubra* and 348 of the 381 candidates in *M. ruginodis* using BLAST.

2.3. Venom Composition of Myrmica rubra

We identified 44 proteins in *M. rubra* venom with similarities to known venom proteins, and these were assigned to three protein families. The vast majority (37/44 or ~84% of total protein diversity) belonged to the venom serine protease or S1 protease family (hereafter VSPs). Another four (~9%) belonged to the venom acid phosphatase family, and the remaining three (~7%) were related to the CAP superfamily (cysteine-rich secretory proteins, antigen 5 and pathogenesis-related 1 proteins). VSPs were not only the most diverse protein family among the *M. rubra* venom candidates but also the most abundant. The sum of all VSP members expressed in transcripts per million (TPM) was 76.7% of all contigs in the *M. rubra* venom gland, followed by the CAP proteins (18.5%) and the venom acid phosphatases (4.8%).

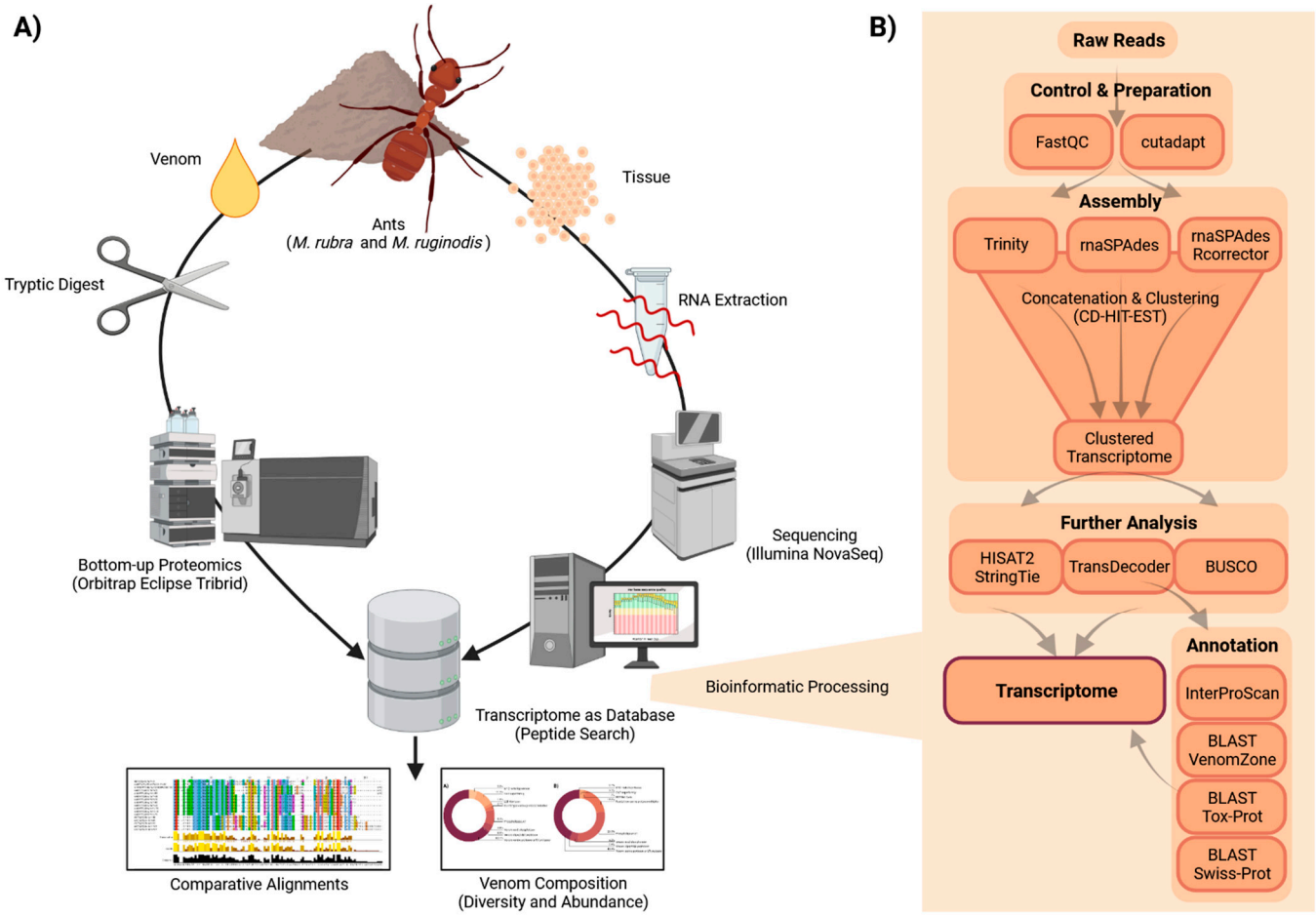

Figure 1. Overview of our proteotranscriptomics workflow. (**A**) Proteomics workflow: Crude venom was collected, digested and analyzed in a bottom–up proteomics approach using an Orbitrap Eclipse Tribrid MS and the transcriptome database. (**B**) Transcriptomics workflow: RNA was sequenced on an Illumina NovaSeq system, and the raw sequencing data were preprocessed and assembled using multiple algorithms. The concatenated dataset was further analyzed and annotated based on different sources of information. The resulting ORFs were used as a database for the proteomics experiment. Transcripts validated at the proteome level were used for the subsequent analysis of venom components in both species.

2.4. Venom Composition of Myrmica ruginodis

We identified 113 proteins in *M. ruginodis* venom with similarities to known venom proteins, and these were assigned to eight protein families (Figure 2). VSPs were again the most diverse (72/113, or ~64% of total protein diversity), followed by CAP proteins (13/113, 11.5%) and phospholipase A1 (PLA) (11/113, ~10%). Six proteins (~5%) were assigned to the M12 metalloproteases, five (~4%) to the venom acid phosphatases, and three (~3%) to the Kunitz-type serine protease inhibitors. Another two proteins (~2%) were members of the EGF-like family of toxins, and one (~1%) was a putative S9 dipeptidyl peptidase. VSPs were again the most abundant components, accounting for 43.4% of all transcripts in the *M. ruginodis* venom glad. The PLAs were next (28.5%), followed by Kunitz-type serine protease inhibitors (18.0%), and the remainder were minor components: CAP proteins (2.5%), EGF-like toxins (1.7%) and venom dipeptidyl peptidase (0.6%). The least abundant protein family was the M12 metalloprotease, accounting for 0.1% of all transcripts in the *M. ruginodis* venom gland.

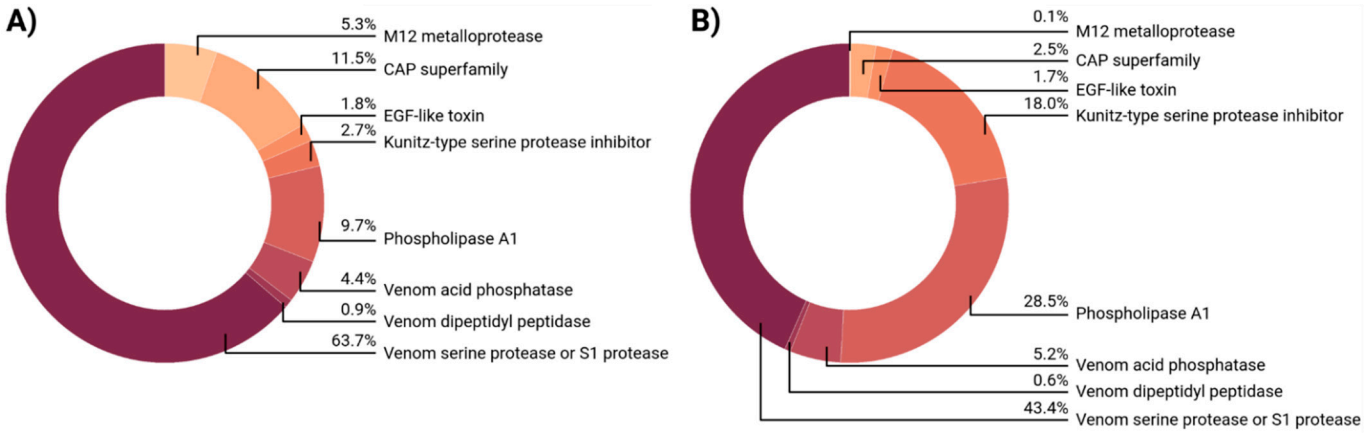

Figure 2. Venom composition of *M. ruginodis* by venom protein family. (**A**) Protein diversity based on absolute numbers of represented sequences. (**B**) Abundance based on transcripts per million (TPM).

2.5. Diversity and Characteristics of EGF-Like Toxins in Myrmicine Ants

The venom proteome of *M. ruginodis* contained two members of the EGF-like toxin family, whereas none were found in the *M. rubra* proteome. We therefore interrogated the transcriptomes of both species for additional EGF-like toxins by performing a BLAST search against the transcriptomic datasets. We identified three additional EGF-like toxin precursors in *M. ruginodis* (making five in total) and four such sequences in *M. rubra*. SignalP analysis and alignments with known ant EGF-like toxins revealed that these are secreted proteins that are expressed as prepeptides with an adjacent signal peptide but no propeptide between the EGF-like domain and the signal peptide. The predicted mature toxins were 49–56 amino acids in length with predicted molecular masses of 5.2–6.0 kDa (Table 1). All corresponding contigs showed higher TPM values or were tracked only in samples from venom gland tissue and not in body tissue samples.

Table 1. Characteristics of EGF-like toxins identified in *M. rubra* and *M. ruginodis* venom glands. The lengths of precursor and mature toxin sequences are shown in amino acids, where # denotes the number of AAs, alongside the predicted molecular weights (MW) and isoelectric points (pI) of the mature toxins. Toxins marked with * were detected by proteomic analysis.

Species	Toxin	# AA Precursor	# AA Mature Toxin	MW (kDa)	pI (pH)
M. rubra	U-MYRTX-Mrub1a	79	49	5.4	5.29
	U-MYRTX-Mrub1b	86	56	6.0	4.80
	U-MYRTX-Mrub1c	76	46	5.2	4.59
	U-MYRTX-Mrub1d	81	51	5.7	5.78
M. ruginodis	U-MYRTX-Mrug1a *	79	49	5.4	5.41
	U-MYRTX-Mrug1b *	79	49	5.4	5.41
	U-MYRTX-Mrug1c	86	56	6.0	4.80
	U-MYRTX-Mrug1d	81	51	5.6	4.72
	U-MYRTX-Mrug1e	79	49	5.3	4.76

To gain further insight into the diversity and function of ant EGF-like toxins, we analyzed their similarity to related sequences in silico. We compared the mature sequences of all putative EGF-like toxins in both species to all known EGF-like toxins from other ants. This alignment revealed an overall similar architecture, featuring a conserved six-cysteine backbone and several conserved sites, particularly within the signal peptide. The propeptide, found in plesiotypic and nontoxic EGF hormones, was not present in the EGF-like toxins. Despite the overall similarity among the sequences, we observed some differences in the inter-cysteine spacing and length.

We constructed a maximum likelihood phylogenetic tree to find unrecognized differences between the proteins and to gain insight into their evolutionary history and relationship. The tree contained three major clades of EGF-like toxins, each exclusively comprising toxins from one subfamily of ants. The MYRTX-clade contained all EGF-like toxins from Myrmicinae (taxonomically represented by *M. rubra*, *M. ruginodis* and *M. rubida*) and achieved 77% bootstrap support. The MIITX-clade was classified as a sister group of the myrmicine toxins and contained EGF-like toxins from Myrmeciinae (*M. gulosa* and *Myrmecia chrysogaster*). It received 97% bootstrap support. Finally, the ECTX-clade was classified as a sister group to all remaining ant EGF-like toxins, and comprised ectatommine toxins found only in *Rhytidoponera metallica*. It also received 100% bootstrap support. Although EGF-like toxins from Myrmeciinae and Ectatomminae formed a single clade and are thus similar at the sequence level, a much higher degree of sequence diversity was indicated for members of the MYRTX-clade. Here, three potential subclades were identified which we named MYRTX-clades A, B and C. MYRTX-clade A contains four EGF-like toxins identified in *M. ruginodis* and two identified in *M. rubra* (U-MYRTX-Mrub1a). MYRTX-clade B has a single member belonging to *M. rubra* (U-MYRTX-Mrub1c). MYRTX-clade C contains the remaining EGF-like toxins from *M. rubra* and *M. ruginodis*, and U18-MYRTX-Mri1a contains the remaining EGF-like toxins from *M. rubida*. However, some of these shallow clades within the MYRTX lineage received only marginal bootstrap support values and the relationships between them must be interpreted with caution.

Finally, we employed a strategy related to that of Eagles and colleagues [31] by initiating a BLAST search to identify patterns of similarity between ant EGF-like toxins and known EGF hormones across the animal kingdom (Table 2). We found that five major subgroups of EGF-like toxins exist in ants. Interestingly, the similarity subgroups revealed by this approach corresponded to clades within the EGF-like toxin phylogeny. The first similarity subgroup comprised the two peptides in the MIITX-clade, which are highly similar to vertebrate heparin-binding EGF (HBEGF) hormones. The second similarity subgroup comprised the members of the ECTX-clade, which are similar to vertebrate betacellulin and epiregulin. The remaining three similarity subgroups featured EGF-like toxins from the MYRTX-clade (myrmicine ants). The most diverse of these three groups contained EGF-like toxins resembling vertebrate transforming growth factor α (TGFα) hormones and comprised six members: two from *M. rubra* (U-MYRTX-Mrub1a and U-MYRTX-Mrub1d) and four from *M. ruginodis* (U-MYRTX-Mrug1a, 1b, 1d and 1e), representing MYRTX-clade A. The second group contained a single peptide (U-MYRTX-Mrub1c) related to vertebrate epiregulin, corresponding to MYRTX-clade B. The remaining three EGF-like toxins (U18-MYRTX-Mri1a, U-MYRTX-Mrub1b and U-MYRTX-Mrug1c) resemble insect Spitz-like proteins and mostly represent MYRTX-clade C.

Table 2. List of ant EGF-like toxins and closely related EGF hormones sorted by clade. The top BLAST hit according to our criteria is shown along with its similarity (percentage) to each toxin. Toxins marked with an * were discovered in this study.

Toxin	BLAST Result ID	Similarity (%)	Clade	EGF-Type
ECTX(02)-Rm1a	RXN00400.1	48.0	ECTX-clade	Vertebrate betacellulin-like
ECTX(02)-Rm1b	RXN00400.1	46.0	ECTX-clade	Vertebrate betacellulin-like
ECTX(02)-Rm1c	XP_032961198.1	50.0	ECTX-clade	Vertebrate betacellulin-like
ECTX(02)-Rm1d	XP_042333623.1	46.3	ECTX-clade	Vertebrate epiregulin-like
ECTX(02)-Rm1e	XP_036083747.1	41.4	ECTX-clade	Vertebrate epiregulin-like
MIITX(02)-Mc1a	XP_041957410.1	62.2	MIICTX-clade	HBEGF-like
U-MIITX(02)-Mg1a	XP_038648772.1	65.8	MIICTX-clade	HBEGF-like
U-MYRTX-Mrub1a *	XP_041634697.1	57.1	MYRTX-clade A	Vertebrate TGF-like
U-MYRTX-Mrub1d *	XP_041634697.1	57.1	MYRTX-clade A	Vertebrate TGF-like
U-MYRTX-Mrug1a *	XP_041634697.1	55.3	MYRTX-clade A	Vertebrate TGF-like
U-MYRTX-Mrug1b *	XP_041634697.1	51.1	MYRTX-clade A	Vertebrate TGF-like
U-MYRTX-Mrug1d *	XP_041634697.1	57.1	MYRTX-clade A	Vertebrate TGF-like

Table 2. *Cont.*

Toxin	BLAST Result ID	Similarity (%)	Clade	EGF-Type
U-MYRTX-Mrug1e *	XP_041634697.1	57.1	MYRTX-clade A	Vertebrate TGF-like
U-MYRTX-Mrub1c *	XP_019908008.1	58.8	MYRTX-clade B	Vertebrate epiregulin-like
U-MYRTX-Mrub1b *	XP_016916180.1	63.0	MYRTX-clade C	Insect Spitz-like
U-MYRTX-Mrug1c *	XP_016916180.1	63.0	MYRTX-clade C	Insect Spitz-like
U18-MYRTX-Mri1a	XP_016916180.1	63.0	MYRTX-clade C	Insect Spitz-like

3. Discussion

3.1. Serine Proteases and Kunitz-Type Serine Protease Inhibitors

VSPs dominated the venom profile of both *M. rubra* and *M. ruginodis* and share similarities with known members of this class from other hymenopteran genera, particularly *Apis*, *Bombus* and *Polistes*. Like other serine proteases, VSPs hydrolyze peptide bonds and cause protein degradation, often triggering cytotoxic or hemotoxic effects [32]. However, few VSPs from insect venom have been analyzed in detail. One exception is Bi-VSP from the venom of *Bombus ignites*, which is similar to several of the *M. rubra* and *M. ruginodis* VSPs. Bi-VSP is a multifunctional serine protease that activates the phenoloxidase cascade and causes lethal melanization in insects, but in mammalian blood it has coagulotoxic effects by activating prothrombin and cleaving fibrinogen [33]. Given the similarity between Bi-VSP and several VSPs in *M. rubra* and *M. ruginodis*, it is possible that at least some of the latter possess similar insecticidal and coagulotoxic activities. Interestingly, Bi-VSP acts in concert with a co-injected Kunitz-type serine protease inhibitor (Bi-KTI), which inhibits plasmin as part of a multipronged attack on the coagulation cascade [34]. Although we found no evidence for the presence of Kunitz-type serine protease inhibitors in *M. rubra* venom at the proteomic level, three members of this class were present in the venom proteome of *M. ruginodis*, and they were the third most abundant component in this sample based on transcriptome data. Sequence comparisons revealed that Kunitz-type serine protease inhibitors in myrmicine ants are closely related to those found in the parasitoid wasp *Pimpla hypochondriaca* (UniProtKB Q8T0W4) and the bumblebee *Bombus terrestris* (Figure 3) [35]. In particular, the *B. terrestris* protein (Bt-KTI, UniProtKB D8KY58) is similar to Bi-KTI (UniProtKB G3LH89) from *Bombus ignitus* and can likewise inhibit plasmin [36]. It is therefore possible that the combinatorial mode of action between VSPs and Kunitz-type serine protease inhibitors observed in *B. ignitus* venom may also be prevalent in myrmicine ants [34]. However, the venom profiles of other myrmicine ants must be resolved and complemented with bioactivity studies to confirm whether the venom components work in this cooperative manner.

Figure 3. Sequence similarities between Kunitz-type serine protease inhibitors from *M. ruginodis* and other hymenopteran species (*Bombus terrestrix*, *Bombus ignitus* and *Pimpla hypochondriaca*). Conserved cysteine residues are shown in blue.

3.2. Venom Allergens

The CAP family is another important group of venom components in *M. rubra* and *M. ruginodis*, although particularly in the latter, where they contribute significantly to the venom profile. CAP proteins are found in most animal venoms and are functionally very diverse [6,37]. For example, cone snail CAP proteins may function as proteolytic enzymes, whereas snake CAP proteins are neurotoxins [6]. In ants and other hymenopterans, CAP proteins are clinically relevant because they are major allergens that underlie the reactions

to bee stings and ant bites [38]. We identified three CAP proteins in the venom of *M. rubra* and 13 in *M. ruginodis*. Interestingly, most were highly similar to allergen 3 from the black imported fire ant (*Solenopsis richteri*), another myrmicine species. The CAP allergens in *Solenopsis* venoms can cause severe, sometimes fatal anaphylaxis in humans and are highly cross-reactive [10,39–41].

Another important group of hymenopteran venom allergens is the PLAs, which hydrolyze phosphatidylcholine and cause inflammation as well as local tissue damage. Some PLAs are described collectively as allergen 1 and exacerbate the allergic response to ant bites. The analysis of *M. ruginodis* venom in the 1960s revealed a surprising lack of PLAs despite their presence in other hymenopteran venoms [42], leading to predictions that future studies would reveal their presence also in *M. ruginodis* [43]. As predicted, we can report for the first time that PLA is also present in *M. ruginodis* venom, and indeed this is the first ever description of PLAs in this species. We identified 11 PLAs, accounting for 18.5% of the transcripts in the venom gland and thus representing the second most abundant protein family in *M. ruginodis* venom. Some of these proteins resembled a PLA from *Dinoponera quadriceps*, a ponerine ant, but most showed greater similarity to Sol i 1, an allergen from *Solenopsis invicta*, and are therefore also likely to trigger allergic reactions.

In summary, potential venom allergens representing the CAP and PLA families are present in the venoms of *M. rubra* and *M. ruginodis*. The candidates resemble proteins from related species that induce allergic reactions and sometimes severe anaphylaxis. In Germany, neither *M. rubra* nor *M. ruginodis* are known to cause significant allergic reactions, although both species are common and synanthropic [44]. However, dangerous allergic reactions have been reported in the USA, where *M. rubra* is considered an invasive pest [45]. Therefore, although *M. rubra* and *M. ruginodis* are not dangerous per se, stings may lead to occasional emergencies and should be treated with caution by the affected individual.

3.3. EGF-Like Toxins

The 113 venom components in *M. ruginodis* included two peptides with an EGF-like motif. These form part of a diverse family of metazoan peptide hormones that are also venom components in anemones and ants [14,22,31,46,47]. EGF-like peptides therefore provide another example of the "toxipotent" nature of peptide hormones and their potential to become weaponized as venom components [48–53].

Although we identified only two EGF-like peptides in the *M. ruginodis* proteome, the transcripts were among the top 30 most abundant. Given that not all mRNAs are constantly translated into proteins, we screened the venom gland transcriptomes of *M. rubra* and *M. ruginodis* for additional EGF-like toxins that were not detected in the proteome. This revealed four *M. rubra* toxins (U-MYRTX-Mrub1a, 1b, 1c and 1d) as well as three additional toxins from *M. ruginodis* (U-MYRTX-Mrug1c, 1d and 1e). This is a key finding that supports the general but lineage-specific presence of EGF-like toxins in formicoid ant venoms. Several EGF-like toxins have recently been described in the ant subfamilies Myrmeciinae (*M. gullosa* and *M. chrysogaster*), Ectatomminae (*R. metallica*), Myrmicinae (*Myrmica sulconodus, M. rubra, M. ruginodis, M. rubida, Pogonomyrmex californicus* and *Pogonomyrmex barbatus*) and Formicinae (*Formica aquilonia*) [14,22,31]. With the exception of the three peptides from *M. gulosa, M. chrysogaster* and *M. rubida*, all the EGF-like toxins were retrieved from whole-body transcriptomes or genomes, so it is unclear if they are translated into venom proteins [31]. In our study, we found that nine EGF-like toxins were expressed in the venom gland, two of which were also detected at the proteomic level. This suggests that EGF-like toxins are important venom components in Myrmicinae, and supports their presence in other formicoid venoms. However, transcriptome data may overestimate toxin diversity [30] and only two of the nine EGF-like toxins were verified at the protein level. Therefore, we recommend that future studies investigate the presence of EGF-like toxins in the venoms of different ants at the proteome level.

3.4. Evolution and Function of Venom Proteins in Myrmicine Ants

Animal venoms have evolved to serve three main functions: predation, defense and competitor deterrence [7]. Although ants use their venom systems mainly to overpower prey and defend their colony, defensive venoms largely underpin their evolutionary success [4]. The eusocial lifestyle of many ant species, with large numbers of individuals forming a single colony, presents a conspicuous target for predators [4]. Vulnerability thus led to the evolutionary innovation of chemical defense systems in ants, with two principal forms: the spraying of formic acid and the injection of venom. Although the acid-based defense system is better known to the general public, the number of species that deploy this mechanism is limited, whereas ~75% of all ant species inject venom [22,54]. Across the animal kingdom, defensive venoms are characterized by their ability to induce pain [55–57].

In agreement with this common role of ant venoms, the functional annotation of venom components in *M. rubra* and *M. ruginodis* suggests activities that may facilitate defense. For example, several protein families are associated with proteolytic and/or tissue-damaging activities, including VSPs, M12 metalloproteases and PLAs, which contribute to the painful effect of envenomation. Furthermore, the cooperative attack on the coagulation cascade by VSPs and Kunitz-type serine protease inhibitors may cause local edema and thus additional hypersensitivity.

The most interesting components identified in the venom systems of *M. rubra* and *M. ruginodis* were the EGF-like toxins, whose biological functions remain unclear. The corresponding myrmecine peptides (MIITX-clade) were recently identified as members of this venom protein family [31]. They showed no activity against insects but induced pain in mice for several days [31]. The MIITX EGF-like toxins may therefore have evolved to mimic HBEGF hormones, resulting in their remarkable algogenic effects by activating mammalian ErbB receptors. The authors also found that the other known ant EGF-like peptides resembled either vertebrate betacellulin or insect EGF-like hormones from *R. metallica*, depending on the analysis method. In contrast, the EGF-like toxin previously described in *M. rubida* was most similar to insect Spitz-like proteins, and may therefore have evolved to target insects. Our sequence analysis was largely consistent with these results, annotating the *R. metallica* (ECTX-clade) peptides as betacellulin or epiregulin, the MYRTX-clade C peptides as insect Spitz-like proteins and the *Myrmecia* peptides (MIITX-clade) as HBEGF mimics. Phylogenetic and sequence similarity analysis for the MYRTX peptides indicated a potentially diverse evolutionary history and molecular diversification. Here, functionally divergent peptides may have emerged within the Myrmicinae that share similarity with different vertebrate and insect templates (Figure 4). However, some of the internal branches of the phylogeny were only weakly supported by bootstraps and thus must be interpreted with caution. Future studies adding more EGF-like toxins from other Formicoidea and in particular from the Myrmicinae will help to clarify the relationships and strengthen the proposed hypotheses. That said, our phylogenetic tree suggests that the EGF-like toxins resembling insect Spitz-like proteins (MYRTX-clade C) were more recently recruited to the venom system of myrmicine ants, probably as weapons against other insects given their similarity to insect hormones. MYRTX-clade A appears to be derived from weaponized TGF hormones and may therefore target the corresponding vertebrate receptors, whereas the single peptide representing MYRTX-clade B is strikingly similar to vertebrate epiregulin and is also likely to target vertebrates. It shares this similarity with two peptides from the distant ECTX-clade, suggesting that convergent evolution has resulted in the recruitment of epiregulin into ant venom systems twice, once each in the subfamilies Ectatomminae and Myrmicinae (Figure 4).

Many of the identified EGF-like toxins in the MYRTX-clade may target vertebrates and thus probably act as defensive weapons. This agrees with earlier reports of painful stings (by European standards) and significant neurotoxicity caused by *M. ruginodis*. However, some EGF-like toxins have different effects depending on the type of test [31], so additional analysis with more sophisticated software and a wider range of ant venoms is recommended before drawing firm conclusions. The ability of MYRTX-clades A and B to affect vertebrates

could also be studied by isolating or synthesizing them, followed by careful functional assays.

In addition to defense, ant venom can be used for predation and minor roles such as communication, which must be considered when assigning functions to venom components. In *M. rubra* and *M. ruginodis*, the EGF-like toxins of MYRTX-clade C that resemble insect Spitz-like proteins may be used to overpower insect prey, perhaps in cooperation with VSPs. Indeed, many ant venom components may serve multiple functions depending on the context. For example, the universal molecular function of VSPs is the hydrolysis of peptide bonds, but this may fulfil various biological roles including trophic agents when injected to insect prey, spreading factors that facilitate the rapid uptake of co-injected toxins, the decomposition of killed prey to provide nutrition, and the triggering of edema and localized pain in predators. The strict division of venom components into unique functional categories may not be possible, and it may be more valuable to interpret venom exochemistry in a context-dependent manner. Indeed, the multifunctional nature of venom components has been reported for several venomous animals, including some ants [49,58–61].

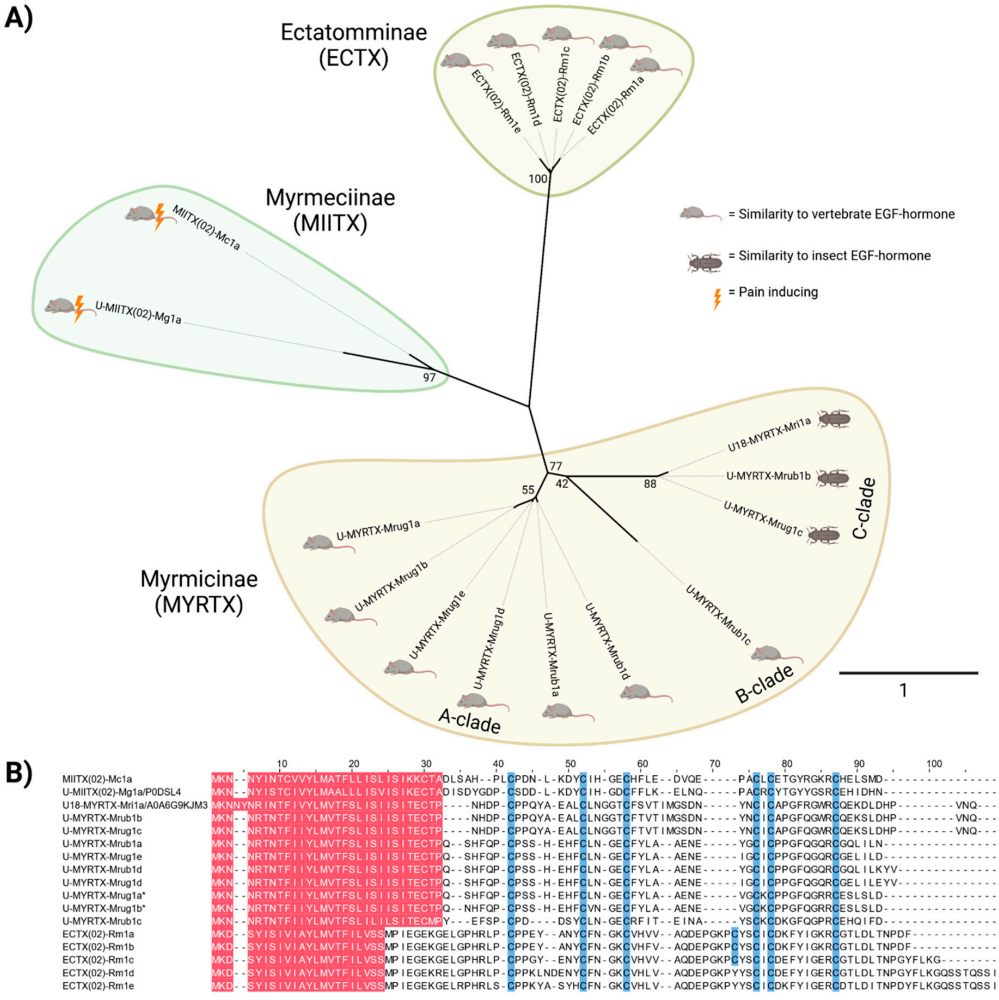

Figure 4. Relationships between all known EGF-like toxins from ants. (**A**) The unrooted phylogenetic tree includes node numbers indicating bootstrap support. The colored boxes indicate the taxonomic placement of the species in which the corresponding toxin was identified. Major clades within the myrmicine EGF-like toxins are indicated (A-clade, B-clade and C-clade). Animal images show the predicted target of each toxin. (**B**) The sequence alignment shows predicted signal peptides (adjusted SignalP prediction and information from UniProtKB) in red and conserved cysteine residues in blue. Toxins marked with an * were discovered in this study.

4. Conclusions

Venoms have contributed to the evolutionary success of many animal lineages, particularly arthropods. Ants are among the most abundant arthropod species, but many have not been studied in detail and questions remain about their evolutionary ecology and biochemistry, including the composition of their venom systems.

We applied a modern venomics workflow based on proteotranscriptomics to shed light on the venom composition of *M. rubra* and *M. ruginodis*, two common myrmicine ant species native to central Europe. We identified several protein families commonly found in ant venoms, including VSPs, Kunitz-type serine proteases, CAP superfamily protein and PLAs. Many of these components are proteolytic enzymes that may be used for predation, defense, as spreading factors, for external digestion or combinations of the above. We also identified several allergenic components in these venoms. Although neither *M. rubra* nor *M. ruginodis* is recognized as medically significant, potentially dangerous anaphylactic shock may occur following envenomation. The previous analysis of *M. ruginodis* venom revealed the conspicuous absence of phospholipases. However, by identifying several members of the PLA family, we demonstrate the presence of these important allergenic substances in *M. ruginodis* venom for the first time. We also detected previously unknown EGF-like toxins in the venom systems of *M. rubra* and *M. ruginodis,* highlighting their importance as lineage-specific ant toxins.

Our work represents a valuable contribution to the growing body of knowledge on the composition, evolutionary ecology and biochemistry of ant venom systems, and we have discussed the potential functions and interactions of the venom components. To test our hypotheses, it will be necessary to isolate or synthesize individual venom peptides and proteins for bioactivity assays. It is also important to broaden the taxonomic coverage of ant venomics, not only to include fellow species of Myrmicinae but also other subfamilies. Such studies may also lead to the discovery of novel bioactive components that can be translated into drugs and bioinsecticides.

5. Materials and Methods

5.1. Animals

Specimens were collected from two colonies each of *M. rubra* and *M. ruginodis* in Giessen (Hesse, Germany). The ants were kept in plastic containers (18 × 14 × 12 cm) filled with soil and were maintained at ~23 °C and 40% relative humidity with a 16 h photoperiod. They were fed weekly with mealworms (*Tenebrio molitor*) and 20% sucrose solution. We examined the collected ants under a VHX-5000 microscope (Keyence, Neu-Isenburg, Germany) to verify the species using an identification key [62].

5.2. Collection of Venom and Venom Glands

We developed a noninvasive venom-sampling protocol in which ants were gently lifted by the thorax using watchmakers' forceps, allowing their abdomen to be submerged in 500 µL methanol in a 1.5-mL Eppendorf tube for 30 s. This induced the animals to release venom into the solvent. The ants were then transferred to a small plastic container and their venom glands were dissected under a light microscope. The venom glands and remaining body tissue were stored separately in TRIzol reagent (Invitrogen, Carlsbad, CA, USA). We collected samples from 27 ants in each colony of both species. Samples from the same colony were pooled and stored at –80 °C.

5.3. RNA Extraction and Sequencing

Total RNA was extracted in TRIzol reagent according to the manufacturer's instructions and was then treated with Turbo DNase (Thermo Fisher Scientific, Waltham, MA, USA) and RNA Clean and Concentrator 5 (Zymo Research, Irvine, CA, USA). Transcriptome library preparation and sequencing were outsourced to Macrogen (Seoul, Korea). Libraries were prepared using the TruSeq stranded mRNA kit including poly(A) RNA selection. Samples were sequenced on an Illumina NovaSeq system to produce 151 bp

paired-end reads. The raw sequence data have been uploaded to the NCBI database (Bio Project PRJNA807911).

5.4. Transcriptome Analysis

Read quality was controlled using FastQC v0.11.9 [63]. We removed sequencing adapters and poly(G) tails, and performed quality trimming with cutadapt v2.10 [64]. All RNA-Seq data from the venom glands and remaining body tissues was used for de novo transcriptome assembly by applying a multiple assembler strategy comprising Trinity v2.11.0 [65] with HPC GridRunner v1.0.2 [66] and a minimum assembled contig length of 30, rnaSPAdes v3.14.1 [67] and one additional rnaSPAdes assembly, based on reads corrected with Rcorrector v1.0.4 [68]. Both rnaSPAdes assemblies were run with *k*-mer sizes of 21, 33 and 55. Assemblies from all three approaches were concatenated and clustered using CD-HIT-EST v4.8.1 [69] with a sequence identity of 1. Further steps were applied to the clustered de novo transcriptome. To verify the assembly, we ran BUSCO v4.1.4 [70] with lineage dataset hymenoptera_odb10 (2020-08-05). Trimmed reads were mapped with HISAT2 v2.2.1 [71] against the assembly and counted with StringTie v2.1.6 [72] to calculate TPM values. Gene expression for each species was calculated using the mean TPM value of both colonies for each species. We used TransDecoder v5.5.0 [65] to identify ORFs with a minimum protein length of 10 amino acids for the proteomics peptide search. We also used SAMtools [73] for data conversions.

A blastp v2.10.1 [74] search against UniProtKB/Swiss-Prot v2020_06 [75], VenomZone [76] (downloaded on 11 January 2021) and UniProtKB/Swiss-Prot Tox-Prot [77] (downloaded on 15 January 2021) was performed on confirmed ORFs to identify venom toxins. The E-value was set to a maximum of 1×10^{-3} and max_target_seqs was set to the size of the query database. For each BLAST hit, we calculated the coverage of query and subject, and similarity with the BLOSUM62 matrix [78] was assessed using BioPython v1.77 [79]. Results were sorted by similarity, query and subject coverage descending for each venom candidate. The resulting top BLAST hit was used for further analysis. We screened InterProScan v5.52-86.0 [80] with all included databases. Putative venom components were assigned to the corresponding venom family.

We manually verified the identity of venom components in comparative alignments for each venom protein family using MAFFT v7.490 [81], Jalview v2.11.1.4 [82], FastTree v2.1.11 [83] and iTol v6 [84] with predictions from SignalP v6.0g [85]. Comparison sequences were collected from VenomZone (release October 2021), UniProtKB/Swiss-Prot v2021_04, and UniProtKB/TrEMBL v2021_04. We also applied alignment seeds available from Pfam v35.0 [86]. Program calls with parameters used are listed in Table S5.

5.5. Species Identification by CO1 Sequence Analysis

Overrepresented sequences identified with FastQC were used to extract the corresponding reads from the raw sequence data. These were transferred to one FASTA file per species along with their corresponding mates, without preserving the mate–pair relationship of the reads. Sequences were then dereplicated using VSEARCH v2.15.1 [87] with derep_prefix and used as a query search (blastn v2.10.1) with default settings and E-value $< 1 \times 10^{-3}$ against available Formicidae *CO1* sequences in BOLD [88] (downloaded on 30 November 2020). All hits with 100% identity and 100% query coverage were kept, including multiple matches per read. Overrepresented sequences are listed in Supplementary Table S1.

5.6. Proteomics

Our bottom–up proteomics strategy involved a mass spectrometry protocol already applied to animal venoms [48,89]. Briefly, we dissolved 10 µg of sample material in 25 mM ammonium bicarbonate containing 0.6 nM ProteasMax. We added 5 mM DTT and incubated for 30 min at 50 °C to complete the disulfide reduction, followed by modification with 10 mM iodacetamide for 30 min at 24 °C. After quenching the reaction with excess

cysteine, we added a 50:1 ratio of trypsin and digested the venom for 16 h at 37 °C. After stopping the reaction by adding 1% trifluoroacetic acid, we purified the sample using a C18-ZipTip (Merck-Millipore, Burlington, MA, USA), dried it under vacuum and redissolved the material in 10 µL 0.1% trifluoroacetic acid.

We separated the peptides on an UltiMate 3000RSLCnano device (Thermo Fisher Scientific) then injected 1 µg of the sample material into a 50 cm µPAC C18 column (Pharma Fluidics, Thermo Fisher Scientific) in 0.1% formic acid at 35 °C. The peptides were eluted in a linear gradient of 3–44% acetonitrile over 240 min. The column was then washed at a flow rate of 300 nL/min with 72% acetonitrile. The separated peptides were injected into an Orbitrap Eclipse Tribrid MS (Thermo Fisher Scientific) in positive ionization mode with the spray voltage set to 1.5 kV and a source temperature of 250 °C achieved using a TriVersa NanoMate (Advion BioSciences, Ithaca, NY, USA). We scanned the samples in data-independent acquisition mode with the following settings: scanning time = 3 s, *m/z* range = 375–1500 and resolution = 120,000. Auto-gain control was set to standard with a maximum injection time of 50 ms. The most intense ions in each cycle with a threshold ion count > 50,000 and charge states of 2–7 were selected with an isolation window of 1.6 *m/z* for higher-energy collisional dissociation (normalized collision energy = 30%). Fragment ion spectra were acquired in the linear ion trap with a rapid scan rate and normal mass range. The maximum injection time was set to 100 ms and selected precursor ions were excluded for 15 s post-fragmentation.

We used Xcalibur v4.3.73.11 and Proteome Discoverer v2.4.0.305 (both from Thermo Fisher Scientific) for data acquisition and analysis. Proteins were identified using Mascot v2.6.2 [90] by searching against the transcriptome sequences using the following settings: precursor ion mass tolerance = 10 ppm, carbamidomethylation as a global modification, methionine oxidation as a variable modification, and one missed cleavage allowed. Fragment ion mass tolerance in the linear ion trap for MS^2 detection was set to 0.8 Da and the false discovery rate was limited to 0.01 using a decoy database. For qualitative analysis, we only considered proteins that were identified with a Mascot score of at least 30 and at least two verified peptides. A comprehensive list of all venom components identified with confidence, and their characteristics and annotations, are provided in the Supplementary Materials. Proteomic raw data have been uploaded to PRIDE (PXD033537).

5.7. Analysis of EGF-Like Toxins

The *M. ruginodis* EGF-like toxins were used as blastp v2.11.0 queries against the unfiltered TransDecoder output from both species (E-value < 10, word size = 3). We kept only those hits with a query coverage of 100% which led to a resulting E-value < 2.47×10^{-24}. Then, we manually dereplicated the resulting sequences and renamed the unique ones. We adopted the Touchard nomenclature [10], which is modified from the King nomenclature for arthropod venom peptides [91]. Given that both species we analyzed were myrmicine ants, we named the EGF-like toxins myrmicitoxins (MYRTX), and assigned the prefix "U", which is used to designate toxins lacking a known molecular target.

SignalP slow–sequential mode for eukarya was used to predict signal peptides. MAFFT and Jalview were used to analyze and confirm our findings compared to the known EGF-like toxins of *Manica rubida* (UniProtKB A0A6G9KJM3) and *Myrmecia gulosa* (UniProtKB P0DSL4). We adjusted the SignalP results (*M. rubra* and *M. ruginodis*) after inspecting the multiple sequence alignment to increase the signal peptide length from 28 to 30 amino acids. Molecular weights and isoelectric points (pI) were calculated using Bio Python v1.79.

For phylogenetic analysis, sequences of all known and herein identified EGF-like toxins were aligned using MAFFT v7.505 FFT-NS-2 at default parameters. We then constructed a maximum likelihood tree with ultrafast bootstrap [92] (based on 20,000 replicates) implemented in IQ-TREE v2.1.2 [93] with ModelFinder [94] enabled (FLU + G4 model). The resulting tree was visualized on iTOL. A blastp v2.11.0 search was performed against a preformatted NCBI nonredundant protein sequence database (downloaded with BLAST

script update_blastdb.pl nr on 6 March 2022) with default settings (E-value < 10, word_size = 3, max_target_seqs = 100) to recover similarities between EGF-like toxins and known EGF hormones. The top 100 hits were sorted by similarity to the query sequence and the most similar empirically verified hit with a similar sequence length was used to hypothesize similarity between the toxin and hormone (Supplementary Table S4).

Supplementary Materials: The following supporting information can be downloaded at: http://www.mdpi.com/xxx/s1. Supplementary Table S1: Overrepresented sequences reported in FastQC during quality control, Supplementary Table S2: Results of CO1 sequence analysis, Supplementary Table S3: Comprehensive list of venom components, Supplementary Table S4: BLAST results of EGF-like toxins search, Supplementary Table S5: Program calls with parameters, Supplementary Data S01-S11: Resulting alignments from MAFFT runs used to validate the venom families as FASTA files, Data S12: Tree of EGF-like toxins used for Figure 4 in NEWICK format. Listing S1. Shell script to prepare the necessary GTF-file for the program call of StringTie. Listing S2. Python script to calculate molecular weights and isoelectrical points.

Author Contributions: Conceptualization, S.H., A.B., T.L., K.B. and A.V.; methodology, T.L., R.Ö., F.F. and G.L.; software, S.H., F.F., T.T. and K.B.; formal analysis, S.H.; investigation, S.H., F.F., J.H., K.B. and T.L.; data curation, S.H., F.F. and T.T.; writing—original draft preparation, S.H., K.B. and T.L.; writing—review and editing, all authors; visualization, S.H.; supervision, K.B. and A.V.; project administration, A.B.; funding acquisition, A.V. All authors have read and agreed to the published version of the manuscript.

Funding: The work is a result of the Animal Venomics project embedded into the Centre for Translational Biodiversity Genomics (LOEWE–TBG) and was granted to A.V. under the program "LOEWE–Landes-Offensive zur Entwicklung Wissenschaftlich-Ökonomischer Exzellenz" of the Hessian Ministry of Higher Education, Research and the Arts. We acknowledge access to resources financially supported by the BMBF grant FKZ 031A533 within the de.NBI network. The authors also acknowledge funding supporting the LOEWE Centre for Insect Biotechnology and Bioresources.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw proteomic and transcriptomic data are publicly available. Raw proteomic data have been uploaded to PRIDE (PXD033537). Raw transcriptomic data have been uploaded to the NCBI database (BioProject PRJNA807911).

Acknowledgments: We acknowledge Björn M. von Reumont for early discussions about the project; Alexander Goesmann, Oliver Rupp and Sebastian Jaenicke for providing their bioinformatics expertise in the field of high-throughput sequencing; and Richard M. Twyman for manuscript editing. Figures 1, 3 and 4 were created using BioRender.com (license YQ23SOTE2G, BK23SOSYG7, and QF23WF352V).

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

References

1. California Academy of Science AntWeb. Version 8.69.2. Available online: https://www.antweb.org/ (accessed on 2 March 2022).
2. Ward, P.S. The Phylogeny and Evolution of Ants. *Annu. Rev. Ecol. Evol. Syst.* **2014**, *45*, 23–43. [CrossRef]
3. Hölldobler, B.; Wilson, E.O. *The Ants*; Harvard University Press: Cambridge, MA, USA, 1990; ISBN 978-0-674-04075-5.
4. Schmidt, J.O. Chemistry, Pharmacology, and Chemical Ecology of Ant Venoms. In *Venoms of the Hymenoptera: Biochemical, Pharmacological and Behavioural Aspects*; Piek, T., Ed.; Academic Press: London, UK, 1986; pp. 425–508. ISBN 978-0-12-554770-3.
5. 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]
6. Fry, B.G.; Roelants, K.; Champagne, D.E.; Scheib, H.; Tyndall, J.D.A.; King, G.F.; Nevalainen, T.J.; Norman, J.A.; Lewis, R.J.; Norton, R.S.; et al. The Toxicogenomic Multiverse: Convergent Recruitment of Proteins Into Animal Venoms. *Annu. Rev. Genom. Hum. Genet.* **2009**, *10*, 483–511. [CrossRef] [PubMed]
7. Schendel, V.; Rash, L.D.; Jenner, R.A.; Undheim, E.A.B. The Diversity of Venom: The Importance of Behavior and Venom System Morphology in Understanding Its Ecology and Evolution. *Toxins* **2019**, *11*, 666. [CrossRef]
8. Robertson, P.L. A Morphological and Functional Study of the Venom Apparatus in Representatives of Some Major Groups of Hymenoptera. *Aust. J. Zool.* **1968**, *16*, 133–166. [CrossRef]

9. Fox, E.G.P.; Adams, R.M.M. On the Biological Diversity of Ant Alkaloids. *Annu. Rev. Entomol.* **2022**, *67*, 367–385. [CrossRef]
10. Touchard, A.; Aili, S.; Fox, E.; Escoubas, P.; Orivel, J.; Nicholson, G.; Dejean, A. The Biochemical Toxin Arsenal from Ant Venoms. *Toxins* **2016**, *8*, 30. [CrossRef]
11. 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]
12. 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*, 732. [CrossRef]
13. 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*, 3503–3516. [CrossRef]
14. Touchard, A.; Aili, S.R.; Téné, N.; Barassé, V.; Klopp, C.; Dejean, A.; Kini, R.M.; Mrinalini, M.; Coquet, L.; Jouenne, T.; et al. Venom Peptide Repertoire of the European Myrmicine Ant *Manica rubida*: Identification of Insecticidal Toxins. *J. Proteome Res.* **2020**, *19*, 1800–1811. [CrossRef] [PubMed]
15. Aili, S.R.; Touchard, A.; Hayward, R.; Robinson, S.D.; Pineda, S.S.; Lalagüe, H.; Vetter, I.; Undheim, E.A.B.; Kini, R.M.; Escoubas, P.; et al. An Integrated Proteomic and Transcriptomic Analysis Reveals the Venom Complexity of the Bullet Ant *Paraponera clavata*. *Toxins* **2020**, *12*, 324. [CrossRef] [PubMed]
16. 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] [PubMed]
17. 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]
18. 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]
19. Herzig, V.; King, G.F.; Undheim, E.A.B. Can We Resolve the Taxonomic Bias in Spider Venom Research? *Toxicon X* **2019**, *1*, 100005. [CrossRef]
20. Lüddecke, T.; Vilcinskas, A.; Lemke, S. Phylogeny-Guided Selection of Priority Groups for Venom Bioprospecting: Harvesting Toxin Sequences in Tarantulas as a Case Study. *Toxins* **2019**, *11*, 488. [CrossRef]
21. Drukewitz, S.H.; von Reumont, B.M. The Significance of Comparative Genomics in Modern Evolutionary Venomics. *Front. Ecol. Evol.* **2019**, *7*, 163. [CrossRef]
22. 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]
23. Ceolin Mariano, D.O.; de Oliveira, Ú.C.; Zaharenko, A.J.; Pimenta, D.C.; Rádis-Baptista, G.; Prieto-da-Silva, Á.R.d.B. Bottom-Up Proteomic Analysis of Polypeptide Venom Components of the Giant Ant *Dinoponera quadriceps*. *Toxins* **2019**, *11*, 448. [CrossRef]
24. Santos, P.P.; Games, P.D.; Azevedo, D.O.; Barros, E.; de Oliveira, L.L.; de Oliveira Ramos, H.J.; Baracat-Pereira, M.C.; Serrão, J.E. Proteomic Analysis of the Venom of the Predatory Ant *Pachycondyla striata* (Hymenoptera: Formicidae). *Arch. Insect Biochem. Physiol.* **2017**, *96*, e21424. [CrossRef] [PubMed]
25. Heep, J.; Klaus, A.; Kessel, T.; Seip, M.; Vilcinskas, A.; Skaljac, M. Proteomic Analysis of the Venom from the Ruby Ant *Myrmica rubra* and the Isolation of a Novel Insecticidal Decapeptide. *Insects* **2019**, *10*, 42. [CrossRef] [PubMed]
26. 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] [PubMed]
27. Pessoa, W.F.B.; Silva, L.C.C.; De Oliveira Dias, L.; Delabie, J.H.C.; Costa, H.; Romano, C.C. Analysis of Protein Composition and Bioactivity of *Neoponera villosa* Venom (Hymenoptera: Formicidae). *Int. J. Mol. Sci.* **2016**, *17*, 513. [CrossRef] [PubMed]
28. Holding, M.L.; Margres, M.J.; Mason, A.J.; Parkinson, C.L.; Rokyta, D.R. Evaluating the Performance of De Novo Assembly Methods for Venom-Gland Transcriptomics. *Toxins* **2018**, *10*, 249. [CrossRef]
29. von Reumont, B.M. Studying Smaller and Neglected Organisms in Modern Evolutionary Venomics Implementing RNASeq (Transcriptomics)—A Critical Guide. *Toxins* **2018**, *10*, 292. [CrossRef]
30. Smith, J.J.; Undheim, E.A.B. True Lies: Using Proteomics to Assess the Accuracy of Transcriptome-Based Venomics in Centipedes Uncovers False Positives and Reveals Startling Intraspecific Variation in *Scolopendra subspinipes*. *Toxins* **2018**, *10*, 96. [CrossRef]
31. Eagles, D.A.; Saez, N.J.; Krishnarjuna, B.; Bradford, J.J.; Chin, Y.K.-Y.; Starobova, H.; Mueller, A.; Reichelt, M.E.; Undheim, E.A.B.; Norton, R.S.; et al. A Peptide Toxin in Ant Venom Mimics Vertebrate EGF-like Hormones to Cause Long-Lasting Hypersensitivity in Mammals. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2112630119. [CrossRef]
32. Di Cera, E. Serine Proteases. *IUBMB Life* **2009**, *61*, 510–515. [CrossRef]
33. Choo, Y.M.; Lee, K.S.; Yoon, H.J.; Kim, B.Y.; Sohn, M.R.; Roh, J.Y.; Je, Y.H.; Kim, N.J.; Kim, I.; Woo, S.D.; et al. Dual Function of a Bee Venom Serine Protease: Prophenoloxidase-Activating Factor in Arthropods and Fibrin(Ogen)Olytic Enzyme in Mammals. *PLoS ONE* **2010**, *5*, e10393. [CrossRef]

34. Choo, Y.M.; Lee, K.S.; Yoon, H.J.; Qiu, Y.; Wan, H.; Sohn, M.R.; Sohn, H.D.; Jin, B.R. Antifibrinolytic Role of a Bee Venom Serine Protease Inhibitor That Acts as a Plasmin Inhibitor. *PLoS ONE* **2012**, *7*, e32269. [CrossRef] [PubMed]
35. Parkinson, N.M.; Conyers, C.; Keen, J.; MacNicoll, A.; Smith, I.; Audsley, N.; Weaver, R. Towards a Comprehensive View of the Primary Structure of Venom Proteins from the Parasitoid Wasp *Pimpla hypochondriaca*. *Insect Biochem. Mol. Biol.* **2004**, *34*, 565–571. [CrossRef] [PubMed]
36. Qiu, Y.; Lee, K.S.; Choo, Y.M.; Kong, D.; Yoon, H.J.; Jin, B.R. Molecular Cloning and Antifibrinolytic Activity of a Serine Protease Inhibitor from Bumblebee (*Bombus terrestris*) Venom. *Toxicon* **2013**, *63*, 1–6. [CrossRef] [PubMed]
37. Gibbs, G.M.; Roelants, K.; O'Bryan, M.K. The CAP Superfamily: Cysteine-Rich Secretory Proteins, Antigen 5, and Pathogenesis-Related 1 Proteins—Roles in Reproduction, Cancer, and Immune Defense. *Endocr. Rev.* **2008**, *29*, 865–897. [CrossRef]
38. Blank, S.; Bazon, M.L.; Grosch, J.; Schmidt-Weber, C.B.; Brochetto-Braga, M.R.; Bilò, M.B.; Jakob, T. Antigen 5 Allergens of Hymenoptera Venoms and Their Role in Diagnosis and Therapy of Venom Allergy. *Curr. Allergy Asthma Rep.* **2020**, *20*, 58:1–58:13. [CrossRef]
39. Hoffman, D.R. Reactions to Less Common Species of Fire Ants. *J. Allergy Clin. Immunol.* **1997**, *100*, 679–683. [CrossRef]
40. Hoffman, D.R.; Smith, A.M.; Schmidt, M.; Moffitt, J.E.; Guralnick, M. Allergens in Hymenoptera Venom. XXII. Comparison of Venoms from Two Species of Imported Fire Ants, *Solenopsis invicta* and *richteri*. *J. Allergy Clin. Immunol.* **1990**, *85*, 988–996. [CrossRef]
41. Hoffman, D.R. Hymenoptera Venom Allergens. *Clin. Rev. Allergy Immunol.* **2006**, *30*, 109–128. [CrossRef]
42. Jentsch, J. A Procedure for Purification of Myrmica Venom: The Isolation of the Compulsive Component. In Proceedings of the VI Congress of the IUSSI, Bern, Switzerland, September 15 1969; pp. 69–75.
43. Piek, T. *Venoms of the Hymenoptera: Biochemical, Pharmacological and Behavioural Aspects*; Elsevier: Amsterdam, The Netherlands, 2013; ISBN 978-1-4832-6370-0.
44. Mebs, D. *Gifttiere: Ein Handbuch für Biologen, Toxikologen, Ärzte, Apotheker*; Wissenschaftliche Verlagsgesellschaft: Stuttgart, Germany, 1992; ISBN 978-3-8047-1219-5.
45. Arevalo, H.A.; Groden, E. European Fire Ant, Red Ant (Suggested Common Names), *Myrmica rubra* Linnaeus (Insecta: Hymenoptera: Formicidae: Myrmicinae): EENY-410/IN746, 8/2007. *EDIS* **2007**, *2007*, 1–5. [CrossRef]
46. 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*, 229–236. [CrossRef]
47. Oliveira, J.S.; Fuentes-Silva, D.; King, G.F. Development of a Rational Nomenclature for Naming Peptide and Protein Toxins from Sea Anemones. *Toxicon* **2012**, *60*, 539–550. [CrossRef] [PubMed]
48. Lüddecke, T.; von Reumont, B.M.; Förster, F.; Billion, A.; Timm, T.; Lochnit, G.; Vilcinskas, A.; Lemke, S. An Economic Dilemma between Molecular Weapon Systems May Explain an Arachno-Atypical Venom in Wasp Spiders (*Argiope bruennichi*). *Biomolecules* **2020**, *10*, 978. [CrossRef] [PubMed]
49. Lüddecke, T.; Herzig, V.; von Reumont, B.M.; Vilcinskas, A. The Biology and Evolution of Spider Venoms. *Biol. Rev.* **2022**, *97*, 163–178. [CrossRef] [PubMed]
50. Undheim, E.A.B.; Grimm, L.L.; Low, C.-F.; Morgenstern, D.; Herzig, V.; Zobel-Thropp, P.; Pineda, S.S.; Habib, R.; Dziemborowicz, S.; Fry, B.G.; et al. Weaponization of a Hormone: Convergent Recruitment of Hyperglycemic Hormone into the Venom of Arthropod Predators. *Structure* **2015**, *23*, 1283–1292. [CrossRef] [PubMed]
51. Arvidson, R.; Kaiser, M.; Lee, S.S.; Urenda, J.-P.; Dail, C.; Mohammed, H.; Nolan, C.; Pan, S.; Stajich, J.E.; Libersat, F.; et al. Parasitoid Jewel Wasp Mounts Multipronged Neurochemical Attack to Hijack a Host Brain. *Mol. Cell Proteom.* **2019**, *18*, 99–114. [CrossRef]
52. Robinson, S.D.; Li, Q.; Bandyopadhyay, P.K.; Gajewiak, J.; Yandell, M.; Papenfuss, A.T.; Purcell, A.W.; Norton, R.S.; Safavi-Hemami, H. Hormone-like Peptides in the Venoms of Marine Cone Snails. *Gen. Comp. Endocrinol.* **2017**, *244*, 11–18. [CrossRef]
53. Barnett, A. Exenatide. *Expert Opin. Pharmacother.* **2007**, *8*, 2593–2608. [CrossRef]
54. Osman, M.F.H.; Brander, J. Weitere Beiträge zur Kenntnis der chemischen Zusammensetzung des Giftes von Ameisen aus der Gattung *Formica*. *Z. Nat. B* **1961**, *16*, 749–753. [CrossRef]
55. Kazandjian, T.D.; Petras, D.; Robinson, S.D.; van Thiel, J.; Greene, H.W.; Arbuckle, K.; Barlow, A.; Carter, D.A.; Wouters, R.M.; Whiteley, G.; et al. Convergent Evolution of Pain-Inducing Defensive Venom Components in Spitting Cobras. *Science* **2021**, *371*, 386–390. [CrossRef]
56. Mouchbahani-Constance, S.; Lesperance, L.S.; Petitjean, H.; Davidova, A.; Macpherson, A.; Prescott, S.A.; Sharif-Naeini, R. Lionfish Venom Elicits Pain Predominantly through the Activation of Nonpeptidergic Nociceptors. *PAIN* **2018**, *159*, 2255–2266. [CrossRef]
57. Walker, A.A.; Robinson, S.D.; Paluzzi, J.-P.V.; Merritt, D.J.; Nixon, S.A.; Schroeder, C.I.; Jin, J.; Goudarzi, M.H.; Kotze, A.C.; Dekan, Z.; et al. Production, Composition, and Mode of Action of the Painful Defensive Venom Produced by a Limacodid Caterpillar, *Doratifera vulnerans*. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2023815118. [CrossRef] [PubMed]
58. Nixon, S.A.; Robinson, S.D.; Agwa, A.J.; Walker, A.A.; Choudhary, S.; Touchard, A.; Undheim, E.A.B.; Robertson, A.; Vetter, I.; Schroeder, C.I.; et al. Multipurpose Peptides: The Venoms of Amazonian Stinging Ants Contain Anthelmintic Ponericins with Diverse Predatory and Defensive Activities. *Biochem. Pharmacol.* **2021**, *192*, 114693. [CrossRef]
59. Fischer, M.L.; Wielsch, N.; Heckel, D.G.; Vilcinskas, A.; Vogel, H. Context-Dependent Venom Deployment and Protein Composition in Two Assassin Bugs. *Ecol. Evol.* **2020**, *10*, 9932–9947. [CrossRef]

60. Walker, A.A.; Mayhew, M.L.; Jin, J.; Herzig, V.; Undheim, E.A.B.; Sombke, A.; Fry, B.G.; Meritt, D.J.; King, G.F. The Assassin Bug *Pristhesancus plagipennis* Produces Two Distinct Venoms in Separate Gland Lumens. *Nat. Commun.* **2018**, *9*, 755. [CrossRef]
61. Dutertre, S.; Jin, A.-H.; Vetter, I.; Hamilton, B.; Sunagar, K.; Lavergne, V.; Dutertre, V.; Fry, B.G.; Antunes, A.; Venter, D.J.; et al. Evolution of Separate Predation- and Defence-Evoked Venoms in Carnivorous Cone Snails. *Nat. Commun.* **2014**, *5*, 3521. [CrossRef] [PubMed]
62. Seifert, B. *Die Ameisen Mittel- und Nordeuropas*; lutra Verlags- und Vertriebsgesellschaft: Görlitz/Tauer, Germany, 2007; ISBN 978-3-936412-03-1.
63. Andrews, S. Babraham Bioinformatics—FastQC A Quality Control Tool for High Throughput Sequence Data. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 8 February 2022).
64. Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet J.* **2011**, *17*, 10–12. [CrossRef]
65. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. *De novo* Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [CrossRef]
66. Haas, B.J. HPC GridRunner. Available online: https://github.com/HpcGridRunner/HpcGridRunner (accessed on 8 February 2021).
67. Bushmanova, E.; Antipov, D.; Lapidus, A.; Prjibelski, A.D. rnaSPAdes: A de novo Transcriptome Assembler and Its Application to RNA-Seq Data. *GigaScience* **2019**, *8*, giz100. [CrossRef]
68. Song, L.; Florea, L. Rcorrector: Efficient and Accurate Error Correction for Illumina RNA-Seq Reads. *GigaScience* **2015**, *4*, 48. [CrossRef]
69. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef]
70. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness. In *Gene Prediction: Methods and Protocols*; Kollmar, M., Ed.; Methods in Molecular Biology; Springer: New York, NY, USA, 2019; pp. 227–245. ISBN 978-1-4939-9173-0.
71. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [CrossRef] [PubMed]
72. Kovaka, S.; Zimin, A.V.; Pertea, G.M.; Razaghi, R.; Salzberg, S.L.; Pertea, M. Transcriptome Assembly from Long-Read RNA-Seq Alignments with StringTie2. *Genome Biol.* **2019**, *20*, 278. [CrossRef] [PubMed]
73. 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] [PubMed]
74. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
75. The UniProt Consortium UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef] [PubMed]
76. de Castro, E.; Jungo, F. VenomZone. Available online: https://venomzone.expasy.org/ (accessed on 11 January 2021).
77. Jungo, F.; Bairoch, A. Tox-Prot, the Toxin Protein Annotation Program of the Swiss-Prot Protein Knowledgebase. *Toxicon* **2005**, *45*, 293–301. [CrossRef] [PubMed]
78. Henikoff, S.; Henikoff, J.G. Amino Acid Substitution Matrices from Protein Blocks. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 10915–10919. [CrossRef] [PubMed]
79. Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; et al. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423. [CrossRef]
80. Mulder, N.J.; Apweiler, R. The InterPro Database and Tools for Protein Domain Analysis. *Curr. Protoc. Bioinform.* **2008**, *21*, 2.7.1–2.7.18. [CrossRef]
81. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
82. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—A Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef] [PubMed]
83. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef]
84. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]
85. Teufel, F.; Almagro Armenteros, J.J.; Johansen, A.R.; Gíslason, M.H.; Pihl, S.I.; Tsirigos, K.D.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 6.0 Predicts All Five Types of Signal Peptides Using Protein Language Models. *Nat. Biotechnol.* **2022**, 1–3. [CrossRef] [PubMed]
86. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The Protein Families Database in 2021. *Nucleic Acids Res.* **2021**, *49*, D412–D419. [CrossRef]
87. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: A Versatile Open Source Tool for Metagenomics. *PeerJ* **2016**, *4*, e2584:1–e2584:22. [CrossRef] [PubMed]

88. Ratnasingham, S.; Hebert, P.D.N. BOLD: The Barcode of Life Data System. *Mol. Ecol. Notes* **2007**, *7*, 355–364. [CrossRef]
89. von Reumont, B.M.; Lüddecke, T.; Timm, T.; Lochnit, G.; Vilcinskas, A.; von Döhren, J.; Nilsson, M.A. Proteo-Transcriptomic Analysis Identifies Potential Novel Toxins Secreted by the Predatory, Prey-Piercing Ribbon Worm *Amphiporus lactifloreus*. *Mar. Drugs* **2020**, *18*, 407. [CrossRef]
90. Perkins, D.N.; Pappin, D.J.C.; Creasy, D.M.; Cottrell, J.S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20*, 3551–3567. [CrossRef]
91. 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]
92. Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [CrossRef] [PubMed]
93. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]
94. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]

Table S1. Overrepresented sequences reported in FastQC during quality control. These sequences were used to collect raw read sequences for *CO1* sequence analysis.

Species	Overrepresented sequence
M. rubra	CTCCATTTAAAAGACAACTGATTATGCTACCTTAGCACAGTCAATTTACT
	CTCCCAACTAAATATAAATTCATAAGTTAAAAAATCTAATTAAATCTTAT
	CTTAATCACAATTAAGATTTTTAACTTAATAACAAAATTCAAATTTTAAT
	CTTTCGTACAAATATATTTATTTTTATTAAAGATAGAAACCGATCTGGCT
	GTCCAATCTAGTAGTGCTGAGCCCAAGTTTTTCGTCGATGGAATGACTTT
	GTCCTTTCGTACAAATATATTTATTTTTATTAAAGATAGAAACCGATCTG
M. ruginodis	CTCCATTTAAAAGACAACTGATTATGCTACCTTAGCACAGTCAATTTACT
	CTCCATTTAATTTTTATCTTAATTCAACATCGAGGTCGCAATCATTTTTA
	CTCCCAACTAAATATAAATTTATAAGTTATAAAAATCTAATTAAATCTTA
	CTTTCGTACAAATATATTTATTTTTATTAAAGATAGAAACCGATCTGGCT
	GTCCAATCTAGTAGTGCTGAGCCCAAGTTTTTCGTCGATGGAATGACTTT
	GTCCTTTCGTACAAATATATTTATTTTTATTAAAGATAGAAACCGATCTG

Table S2. Results of *CO1* sequence analysis. We used overrepresented sequences to extract all corresponding raw reads with their mates, dereplicated them and ran a BLAST search against BOLD. Only hits with 100% identity and query coverage were kept. One read could match more than one record in BOLD.

	Myrmica rubra	***Myrmica ruginodis***
Reads	634,472	1,328,614
Unique reads	40,978	98,119
Unique reads after BLAST search and filtering	1,045	28,119
Matching BOLD records	21 Formicidae 160 *Myrmica* 2,343 *Myrmica rubra*	13 Formicidae 43 *Myrmica* 858 *Myrmica ruginodis*

Table S5. Tools used during the presented study arranged in alphabetically order. Generic parts are marked with curly brackets.

Tool used and comments	Version	Program call with parameters
blastn (*CO1* analysis against BOLD data)	v2.10.1	makeblastdb -in {input_bold} \ -dbtype nucl \ -parse_seqids \ -title {title} \ -out {output} blastn \ -evalue 1e-3 \ -num_threads 20 \ -db {input_bold} \ -query {input} \ -out {output} \ -outfmt "6 std slen qlen ppos qcovhsp qcovs sseq qseq" \ -task "megablast"
blastp (Tox-Prot, VenomZone, Swiss-Prot, each as query)	v2.10.1	cat {input} | \ makeblastdb -in - \ -dbtype nucl \ -out {output_prefix} \ -title {title} NUM_SEQS=$(blastdbcmd -db {assembly_as_db} -info | grep -oP "[\d,]+(?= sequences)" | tr -d ",") blastp \ -query "{db_as_query}" \ -db {assembly_as_db} \ -max_target_seqs $NUM_SEQS \ -outfmt "6 std slen qlen ppos qcovhsp qcovs sseq qseq" \ -evalue 1e-3 \ -num_threads 2 \ -word_size 2 \ -out {output}
blastp (candidates for final alignments for venome serine protease)	v2.11.0	makeblastdb -in {input_uniprot_trembl} \ -dbtype prot \ -out {output} blastp \ -query {input_query} \ -db {input_uniprot_trembl} \ -out {output} \ -outfmt "6 std slen qlen ppos qcovhsp qcovs sseq qseq sgi stitle staxids sscinames scomnames sblastnames sskingdoms salltitles" \ -evalue 10 \ -word_size 3 \ -matrix BLOSUM62 \ -max_target_seqs 500 \ -seg 'no' \ -num_threads 18
blastp (EGF-like toxins in TransDecoder-	v2.11.0	makeblastdb \ -in {input_transdecoder} \ -dbtype prot \

data and in NCBI nr database)		-out {output} blastp \ -query {input} \ -out {output} \ -outfmt "6 std slen qlen ppos qcovhsp qcovs sseq qseq" \ -evalue 10 \ -word_size 3 \ -matrix BLOSUM62 \ -max_target_seqs 500 \ -seg 'no' \ -num_threads 8 update_blastdb.pl nr update_blastdb.pl taxdb blastp \ query {input_egf_like} \ -db {input_nr} \ -out {output} \ -outfmt "6 std slen qlen ppos qcovhsp qcovs sseq qseq sgi stitle staxids sscinames scomnames sblastnames sskingdoms salltitles" \ -evalue 10 \ -word_size 3 \ -matrix BLOSUM62 \ -max_target_seqs 100 \ -seg 'no' \ -num_threads 8
BUSCO	v4.1.4	busco \ --force \ --mode tran \ --in {input} \ --out {output_prefix} \ --out_path {output_dir} \ --lineage_dataset hymenoptera_odb10.2020-08-05/hymenoptera_odb10 \ --offline \ --cpu 16
CD-HIT-EST	v4.8.1	cd-hit-est \ -i {input} \ -o {output} \ -c 1 \ -n 11 \ -p 1 \ -g 1 \ -M 64000 \ -T 32 \ -d 40
cutadapt	v2.10	cutadapt \ --pair-filter=any \ --nextseq-trim=15 \ -a "AGATCGGAAGAGCACACGTCTGAACTCCAGTCA" \ -A "AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT" \ --max-n 0 \ --minimum-length 15 \

		--too-short-output {output_shortr1} \ --too-short-paired-output {output_shortr2} \ --discard-casava -o {output_r1} \ --paired-output {output_r2} {input} \ --cores=32
HISAT2	v2.2.1	hisat2-build \ {input} \ {output_prefix} \ -p 20 hisat2 \ -x {input} \ --rna-strandness RF \ -1 {input_r1} \ -2 {input_r2} \ -p 20 \ -q \ -S {output_sam} \ --all
InterProScan	v55.52-86.0	interproscan.sh \ --input {input} \ --output-file-base {output_prefix} \ --formats TSV,XML,JSON,GFF3 \ --disable-precalc \ --goterms \ --iprlookup \ --pathways \ --seqtype p \ --tempdir {tempdir} \ --appl CDD,Coils,Gene3D,Hamap,MobiDBLite,PANTHER,Pfam,Phobius,PIRSF ,PRINTS,ProSitePatterns,ProSiteProfiles,SFLD,SMART,SUPERFAMILY,Si gnalP_EUK,SignalP_GRAM_NEGATIVE,SignalP_GRAM_POSITIVE,TI GRFAM,TMHMM \ --cpu 8
IQ-TREE	v2.1.2	iqtree \ -T 3 \ -s {input} \ -B 20000 \ --prefix {output_prefix}
MAFFT (all alignments)	v7.490	zcat {input} | \ mafft \ --localpair \ --maxiterate 1000 \ --thread 8 \ --reorder - > {output}
MAFFT (final alignments for venome serine protease)	v7.490	mafft \ --maxiterate 1000 \ --localpair \ --thread 4 \ {input} > {output}
MAFFT (Fig. 4, EGF-like toxins)	v7.505	mafft \ --retree 2 \ --reorder \ {input} > {output}

Rcorrector	v1.0.4	run_rcorrector.pl \ -t 32 \ -1 {input_left} \ -2 {input_right} \ -od {outputdir}
rnaSPAdes	v3.14.1	spades.py \ --rna \ --ss-rf \ -k 21,33,55 \ --threads 32 \ --memory 320 \ -o {output_dir} \ {input_left} \ {input_right}
SAMtools	v1.11	samtools sort \ -@ 4 \ -m 16384 \ -l 9 \ -o {output_bam_sorted} \ {input} samtools index \ -@ 4 \ -b \ {input_bam_sorted} \ {output_bam_indexed} samtools stats \ {input_bam_sorted} \ -@ 4 > {output_bam_stats} plot-bamstats \ -p {output_prefix} \ {input_bam_stats}
SignalP	v6.0g	Signalp6 \ --fastafile {input} \ --organism eukarya \ --output_dir {output_dir} \ --format all \ --mode slow-sequential
StringTie	v2.1.4	stringtie \ -o {output_transcripts} \ -p 4 \ -G {input_gtf} \ -A {output_gene_abundance} \ -C {output_coverage} \ -b {outputdir_ballgown} \ -e {input_bam} prepDE.py \ -i {outputdir_ballgown} \ -g {output_gene_count_matrix} \ -t {output_transcript_count_matrix} \ -l 145
TransDecoder	v5.5.0	TransDecoder.LongOrfs \

		-m 10 \ -t {input} \ --output_dir {output_dir}
Trinity	v2.11.0	Trinity \ --grid_exec "HPCRUNNER" \ --min_contig_length 30 \ --SS_lib_type RF \ --left {input_left} \ --right {input_right} \ --seqType fq \ --max_memory 200G \ --bflyCalculateCPU \ --CPU 10 \ --output {output_dir} \ --full_cleanup \ --normalize_max_read_cov 50 \ --NO_SEQTK
VSEARCH	v2.15.1	vsearch \ --derep_prefix \ {input} \ --output {output}

Listing S1. Shell script to prepare the necessary GTF-file for the program call of StringTie.

```
for i in {input_bam_sorted}
do
   samtools view -H $i
done | \
   sort | \
   uniq | \
   perl -ne '
      if ($_ =~ /^\@SQ\s+SN:(\S+)\s+LN:(\d+)/) {{
         print join("\t", ($1, ".", "transcript", 1, $2, 0, "+", ".","ID=$1")),"\n";
      }}
   ' > {output}
```

Listing S2. Python script to calculate molecular weights and isoelectrical points.

```
from Bio import SeqIO
from Bio.SeqUtils import molecular_weight
from Bio.SeqUtils.IsoelectricPoint import IsoelectricPoint as IP

fasta = {input_proteinlist}

for seq_record in SeqIO.parse(fasta, "fasta"):
   print(seq_record.id)
   seq = seq_record.seq.rstrip("*")
   print("MW: %0.2f" % molecular_weight(seq, "protein"))
   print("pI: %0.2f" % IP(seq).pi())
```