

Article

# Venom Gland Transcriptomic and Proteomic Analyses of the Enigmatic Scorpion *Superstitionia donensis* (Scorpiones: Superstitioniidae), with Insights on the Evolution of Its Venom Components

Carlos E. Santibáñez-López <sup>1</sup>, Jimena I. Cid-Urbe <sup>1</sup>, Cesar V. F. Batista <sup>2</sup>, Ernesto Ortiz <sup>1,\*</sup> and Lourival D. Possani <sup>1,\*</sup>

<sup>1</sup> Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Apartado Postal 510-3, Cuernavaca, Morelos 62210, Mexico; santibanezlo@wisc.edu (C.E.S.-L.); jcidu@ibt.unam.mx (J.I.C.-U.)

<sup>2</sup> Laboratorio Universitario de Proteómica, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Apartado Postal 510-3, Cuernavaca, Morelos 62210, Mexico; fbatista@ibt.unam.mx

\* Correspondence: erne@ibt.unam.mx (E.O.); possani@ibt.unam.mx (L.D.P.); Tel.: +52-777-329-1647 (E.O.); +52-777-317-1209 (L.D.P.)

Academic Editor: Richard J. Lewis

Received: 25 October 2016; Accepted: 1 December 2016; Published: 9 December 2016

**Abstract:** Venom gland transcriptomic and proteomic analyses have improved our knowledge on the diversity of the heterogeneous components present in scorpion venoms. However, most of these studies have focused on species from the family Buthidae. To gain insights into the molecular diversity of the venom components of scorpions belonging to the family Superstitioniidae, one of the neglected scorpion families, we performed a transcriptomic and proteomic analyses for the species *Superstitionia donensis*. The total mRNA extracted from the venom glands of two specimens was subjected to massive sequencing by the Illumina protocol, and a total of 219,073 transcripts were generated. We annotated 135 transcripts putatively coding for peptides with identity to known venom components available from different protein databases. Fresh venom collected by electrostimulation was analyzed by LC-MS/MS allowing the identification of 26 distinct components with sequences matching counterparts from the transcriptomic analysis. In addition, the phylogenetic affinities of the found putative calcins, scorpines, La1-like peptides and potassium channel  $\kappa$  toxins were analyzed. The first three components are often reported as ubiquitous in the venom of different families of scorpions. Our results suggest that, at least calcins and scorpines, could be used as molecular markers in phylogenetic studies of scorpion venoms.

**Keywords:** enzymes; motifs; phylogenetic analysis; toxins; transcriptome

## 1. Introduction

Despite the large number of studies available in the scorpion venom literature, concerning venom components and identification of their activities, only twelve scorpion families of the twenty recognized extant families [1,2] are currently studied (Table 1). While most of the studied scorpions belong to the family Buthidae, an increasing number of species from other families (i.e., Bothriuridae, Caraboctonidae, Hormuridae, Scorpionidae, Scorpiopidae, Urodacidae and Vaejovidae) are drawing the attention of researchers. In recent years, transcriptomic analyses of the venom gland of several scorpion species have been published, increasing our knowledge on the biodiversity of venom peptides, and allowing us to focus on the evolution of the genes coding for them (e.g., [1,3–6]). More recently,

RNA-Seq has become the technology of choice in the study of venom gland transcriptomes, because it is a low-cost sequencing technology capable of producing millions of sequences at once [6], including those of the transcripts coding for several putative toxins or venom components that may not be easily detected in the venom for reasons including low expression levels, fast turnover, etc. (e.g., the difference between the number of scorpines found in the transcriptome and proteome of *Urodacus yaschenkoi* [6,7]).

**Table 1.** Venom studies, cDNA and/or transcriptomic analysis from the current scorpion families recognized by [1–10]. \* Denotes manuscript submitted for publication, now under revision.

Family	Venom Studies Available	cDNA or Transcriptome Analysis Available
Akravidae	No	No
Bothriuridae	No	Yes
Buthidae	Yes	Yes
Caraboctonidae	Yes	Yes
Chactidae	No	No
Chaerilidae	Yes	Yes
Diplocentridae	No	No
Euscorpiidae	Under revision *	Under revision *
Hemiscorpiidae	Yes	No
Heteroscorpionidae	No	No
Hormuridae	Yes	Yes
Iuridae	No	No
Pseudochactidae	No	No
Scorpionidae	Yes	Yes
Scorpiopidae	Yes	Yes
Superstitioniidae	This study	This study
Troglotayosidae	No	No
Typhlochactidae	No	No
Urodacidae	Yes	Yes
Vaejovidae	Yes	Yes

Among the eight neglected scorpion families, Superstitioniidae Stahnke, 1940, stands out. This family, along with family Akravidae Levi, 2007 (a possible extinct family), are the only two monotypic scorpion families, meaning that each contains only one genus and one species (see [1]). The phylogenetic position of the family Superstitioniidae within the Tree of Life of scorpions suggests that it is closely related to the family Typhlochactidae Mitchell, 1971, a family of troglobitic scorpions endemic to Mexico [1], with both families included within the superfamily Chactoidea [8]; therefore, distantly related to buthid scorpions. The Superstitioniidae is geographically isolated in Arizona and the Baja Peninsula from its closest Typhlochactidae relatives in Eastern Mexico. The taxonomy and systematics of the family Superstitioniidae remains undisputed (see [8]). *Superstitionia donensis* Stahnke, 1940, the Superstition Mountains Scorpion, is a small (reaching a length of 30 mm in adults), shiny and spotted scorpion that inhabits arid deserts with sparse plant cover [9].

Given its uniqueness and phylogenetic position within the order Scorpiones (i.e., distantly related to buthids and closely related to troglobite scorpions, plus highly endemic), the scorpion species *S. donensis* is a perfect candidate to study its venom. In the present contribution, we analyze the venom gland transcriptome of *S. donensis*. We report 135 annotated venom transcripts, among which we found sequences that putatively code for toxins, plus other peptides and venom-specific proteins. We also identified 26 components for which the sequences determined through mass spectrometry analysis correspond to translated sequences found in the transcriptome. This work enriches our knowledge on venom peptides from unexplored scorpion families, and allows us to fill more pieces in the jigsaw puzzle of the scorpion venom evolution.

### Calcins, Scorpines, La1-Like and Potassium Channel $\kappa$ Toxins in Scorpion Venoms

While most of the families of sodium and potassium channel toxins are the best-studied components of scorpion venoms, other constituents have recently received major attention. Among them are calcins, scorpines, La1-like peptides and the potassium channel  $\kappa$  toxins, which have been recently described for several species and have been associated with a number of envenomation effects produced by scorpion stings.

Calcins, toxins affecting calcium channels, are structurally characterized by an inhibitor cystine knot (ICK) motif, making them different from the sodium, chloride and potassium channel toxins [10–12]. They were first discovered in the venom of the buthid scorpions: *Hottentotta hottentotta* (= *Buthotus hottentotta*) in 1991 [13]; *Hottentotta judaicus* (= *Buthotus judaicus*) in 1996 [14]; and *Mesobuthus martensii* in 1997 [15]. The first reported peptide with affinity to ryanodine receptors (RyR) from non-buthid scorpions was found in the venom of *Pandinus imperator* in 1992 [16]. Since then, calcins have been found in the venom of several species belonging to nine families, out of the 11 thus far studied (see Table 2). Ma et al. [3] analyzed the phylogenetic affinities of calcins (those available at that time) and their results showed the differences between buthid and non-buthid calcins, and that maurocalcin had an independent origin from the calcins of the rest of the Scorpionidae species.

**Table 2.** Presence of calcins, scorpines, La1 like peptides and potassium channel  $\kappa$  toxins in the eleven scorpion families with venom or transcriptomic studies as indicated in Table 1.

Family	Calcins	Scorpines	La1-Like Peptides	Potassium Channel $\kappa$ Toxins
Bothriuridae	No	Yes	Yes	No
Buthidae	Yes	Yes	Yes	No
Caraboctonidae	Yes	Yes	Yes	No
Chaerilidae	Yes	Yes	Yes	No
Hemiscorpiidae	No	No	No	No
Hormuridae	No	Yes	Yes	Yes
Scorpionidae	Yes	Yes	Yes	Yes
Scorpiopidae	Yes	Yes	Yes	No
Superstitioniidae	This study	This study	This study	This study
Urodacidae	Yes	Yes	Yes	No
Vaejovidae	Yes	Yes	Yes	Yes

Scorpines are peptides with dual activity: they are cytolytic (or antimicrobial), but also contain a potassium channel-blocking domain [17]. The first scorpine discovered was isolated from the venom of *Pandinus imperator* [18]. Subsequently, a wide phylogenetic distribution of these peptides in other scorpion families was established [19] (Table 2). Recently, this subfamily was shown to have an independent origin from the rest of the potassium channel toxins [5].

La1-like, long-chain peptides with eight cysteines [11], are also ubiquitous in scorpion venoms. They have been described in the venom of species belonging to 10 out of the 11 families studied so far (Table 2); and interestingly their function remains unknown. Ma et al. [3] also revised the phylogenetic status of these peptides, and found the presence of two main clades, plus four La1-like peptides from the venom of *Scorpiops jendeki* clustered independently from their phylogenetic origin, which would suggest multiple gene duplications. Later, Sunagar et al. [4] revised again these peptides, though not including the sequences obtained in the previous analysis (i.e., [3]). Their results showed that these peptides have multiple origins, and therefore do not mirror the scorpion phylogenetics (i.e., follows the proposed phylogenetic history of scorpions).

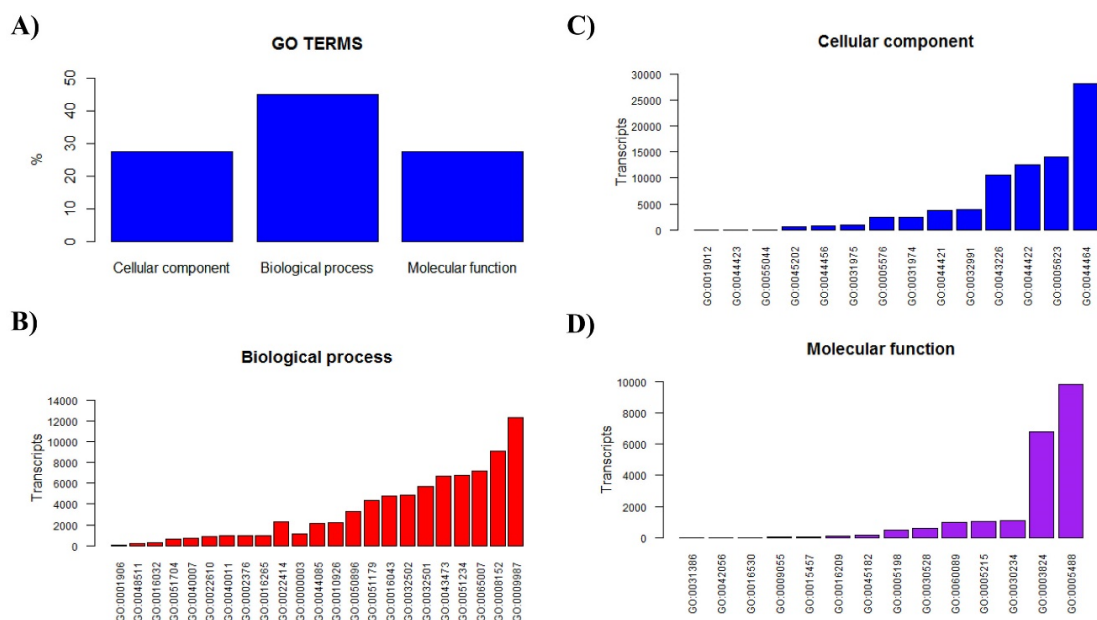
Finally, and unlike the other components mentioned before, the potassium channel  $\kappa$  toxins subfamily has not been found in many of the scorpion venoms studied thus far. These toxins, with a distinct CS $\alpha\alpha$  motif [17] and low activity on potassium channels have been described in only a few species belonging to three scorpion families (Table 2).

In the present contribution, we revise the status of these components (calcins, scorpines, La1-like peptides and potassium channel  $\kappa$  toxins) using phylogenetic analyses under Bayesian inference. We propose motifs for those clades recovered with high posterior probabilities. This will contribute to establish a more complete and stable classification, and will help to categorize newly discovered components from different scorpion venoms in the future.

## 2. Results and Discussion

### 2.1. *S. donensis* Venom Gland Global Transcriptomic Analysis

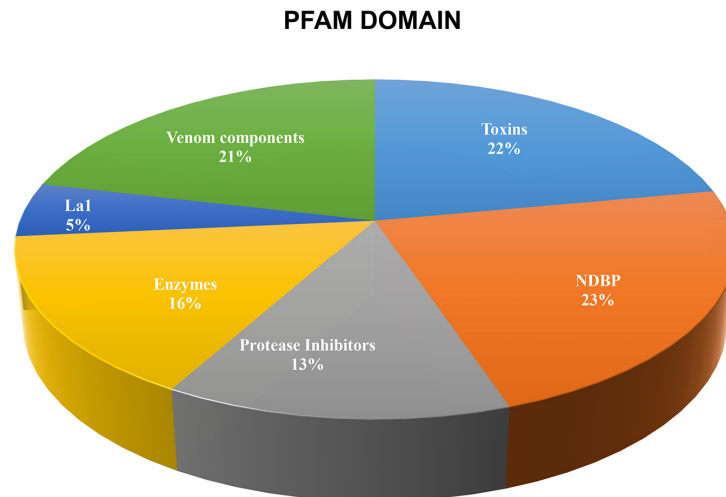
After sequencing, assembly and cleaning, 16,145,663 reads were obtained corresponding to 219,073 transcripts. From them, a total of 45,979 were identified matching sequences in databases, with an N50 of 468 bp. A subgroup of 9930 was annotated, of which 1719 matched known arthropod sequences. Few transcripts (120) were classified as having identity to annotated genes or transcripts from arachnids, in particular, 62 were from scorpions; however, this seems to be the result of incompleteness of the databases against which the sequences are compared. Figure 1 shows the most abundant GO-term categories found in the transcriptome analysis of the venom gland of *S. donensis*.



**Figure 1.** (A) Distribution of annotated sequences from the venom gland transcriptome of *S. donensis* according to Gene Ontology (GO) terms. The category designated by GO as “Biological process” was the most diverse. (B–D) Distribution of the most represented categories within each GO term (GO numbers shown).

### 2.2. The Repertoire of Venom-Specific Transcripts in *S. donensis*

Following the scheme presented in recent venom gland transcriptome analyses [6,11], we report here 135 sequences putatively coding for the following known venom peptides and proteins (Figure 2):



**Figure 2.** Relative proportion (expressed as percentages) of the Pfam domains of the 135 annotated transcripts, which putatively code for venom components found in the venom gland transcriptome analysis of *S. donensis*. The category Toxins includes putative Na<sup>+</sup>, K<sup>+</sup> and Ca<sup>2+</sup> toxin channels peptides; the category NDBPs (Non-Disulfide-Bridged Peptides) includes all possible NDBPs peptides even when no Pfam domain was found; the category Protease Inhibitors includes Ascaris-Type and Kunitz-Type inhibitors; the category La1 includes putative La1-type peptides; the category Enzymes includes all possible peptides with venom enzymatic activity; and the category Other Venom Components includes putative venom proteins and possible CAP peptides.

### 2.2.1. Toxins

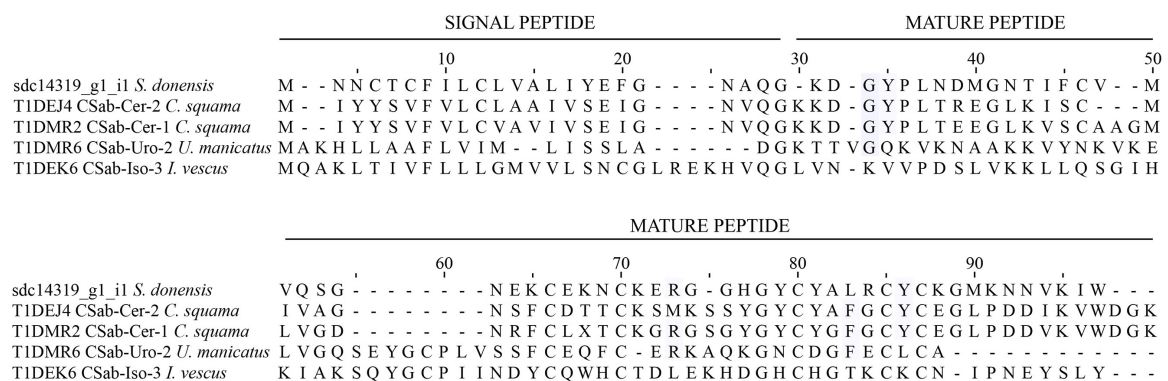
Scorpion venoms are composed mainly by two distinct types of fractions: the toxic and cytolytic peptides and the nontoxic [20], which includes a complex mixture of different enzymes. In addition they might contain carbohydrates, lipids, free amines, nucleotides and other components with unknown function [5]. The toxic fraction of scorpion venom is historically the most studied one, due to the scorpion's clinical relevance. These toxins affect sodium, potassium, calcium and chloride channels; therefore, they have been employed as tools to study the physiology, or the three-dimensional level of the molecular structure of these channels (e.g., [21,22]). In the transcriptome analysis of the venom gland of *S. donensis*, we found 30 transcripts putatively coding for ion channel-acting toxins, representing 22% of all transcripts (Figure 2). This is congruent with other transcriptomic analyses from non-Buthidae species like the members of families Caraboctonidae [23]; Urodacidae [6]; and Vaejovidae [11].

#### Sodium Channel Toxins

The sodium channel toxins are modifiers of the gating mechanism of the channel [24], and are responsible for the neurotoxic symptoms during envenomation [5,25]. Therefore, these toxins (along with those affecting potassium channels) are by far the best studied in scorpion venoms. There are currently more than 520 sequences of toxins (or putative toxins) listed in the InterProt database [26]. Most of them (99%) belong to buthid scorpions (for recent reviews, see [5,17]). Several non-Buthidae transcriptomic analyses (e.g., [6,11,23]) have reported the presence of a low number of sodium channel toxin-coding transcripts. We, unlike other studies (e.g., [6,10]), found eight transcripts for this kind of toxin: (a) one putatively coding for a component labeled sdc10528\_g1\_i1 with 43% identity with the precursor of Toxin To9, deduced from a cDNA cloned from *Tityus obscurus* [27]; (b) component sdc14462\_g1\_i1 with 54% identity with the mature chain of Altitoxin, obtained from the venom of *Parabuthus transvaalicus* [28]; (c) component sdc14462\_g1\_i2 with 57% identity with the mature chain of Birtoxin, also from the venom of *P. transvaalicus* [29]; (d) component sdc15193\_g1\_i1 with 61% identity with the precursor of Toxin CII7, deduced from a cDNA cloned from *Centruroides limpidus*



(Uniprot accession number: P59865); (e) component sdc16570\_g1\_i1 with 44% identity with the precursor of Toxin TdNa8, deduced from a cDNA cloned from *Tityus discrepans* [30,31]; (f) component sdc21236\_g1\_i1 with 42% identity with the precursor of the Neurotoxin LmNaTx30, deduced from the transcriptome analysis of the venom gland of *Lychas mucronatus* [32]; (g) component sdc10528\_g1\_i1 with 43% identity with the precursor of the Toxin Pg8, isolated from the venom (but also deduced from the cloned cDNA) of *Parabuthus granulatus* [33]; and (h) sdc14319\_g1\_i1 with 48% identity with the precursor Csab-Cer-2 deduced from the transcriptomic analysis of the venom gland of *Cercophonius squama* [4]. In Figure 3 we show only some representative examples of these Na<sup>+</sup>-channel peptides. In addition to this, one sequence had hits with 50% identity with the precursor of the Lipolysis activating peptide 1 (LPV1) alpha chain, isolated (and deduced from a cloned cDNA) from the venom of *Buthus tunetanus* (= *Buthus occitanus tunetanus*) [34]. These peptides share sequence identity with the sodium channel toxins; however, unlike them, LPV1s lack a cysteine in their sequence resulting in a reduced number of disulfide bridges and distinct interchain bridges [34].



**Figure 3.** Sequence alignment of components with identity with sodium channel toxins (cysteine-stabilized  $\alpha/\beta$  motif, CS  $\alpha\beta$ , indicated as CSab in the toxin names) found in the transcriptome analysis of the venom gland of *S. donensis* and those that were similar. Uniprot entry numbers precede the toxins' names: (a) Component sdc14319\_g1\_i1, translated ORF; (b) CSab-Cer-2 from *Ce. squama*; (c) CSab-Cer-1 from *Ce. squama*; (d) CSab-Uro-2 from *Urodacus manicatus*; and (e) CSab-Iso-3 from *Isometroides vespucis*.

### Potassium channel toxins

Peptides acting on K<sup>+</sup> channels vary in size from 23 to up to 64 amino acids and are classified based on primary sequence and disulfide bond connectivity [5,35]. Along with the sodium channel toxins mentioned above, potassium channel toxins are also well studied. In the InterProt database, there are nearly 230 sequences coding for short and long potassium channel toxins [36,37]. Most of them (78%) belong to buthid scorpions. Four families have been recognized for these toxins: the  $\alpha$ -,  $\beta$ - and  $\gamma$ - families, stabilized by the CS $\alpha\beta$  motif, and the  $\kappa$ - family [5].

Our results were consistent with previous scorpion venom gland transcriptomic analyses (e.g., that of *U. yaschenkoi* [6]). We found 11 sequences showing identity with seven different potassium channel toxins (members of  $\alpha$ - and  $\kappa$ - families). Ten sequences coding for six putative  $\alpha$ -KTxs were found: (a) components sdc13860\_g1\_i2 and sdc14273\_g1\_i2 had hits with 39% and 48% identity, respectively, with the precursor of the potassium channel toxin  $\alpha$ KTx 6.7 deduced from a cDNA cloned from the venom of *Opisthophthalmus carinatus* [38]; (b) components sdc13860\_g1\_i1 and sdc10141\_g1\_i1 had hits with 42% and 46% identity, respectively, with the precursor of the potassium channel toxin  $\alpha$ KTx 6.10 also deduced from a cDNA cloned from the venom of *O. carinatus* [38] (Figure 4 shows the alignment of the sequences with identity with members of the  $\alpha$ KTx subfamily 6); (c) components sdc26193\_g1\_i1 and sdc13949\_g1\_i1 had hits with 60% and 70% identity, respectively, with the precursor of the Toxin LmKTx 8 deduced from a cDNA cloned from *Lychas muronatus* [39]; (d) components sdc9772\_g1\_i2

and sdc9772\_g1\_i1 had hits with 40% and 42% identity, respectively, with the precursor of a potassium channel toxin deduced from a cDNA cloned from *U. yaschenkoi* [40]; (e) component sdc14273\_g1\_i1 had hits with 44% identity with the precursor of the potassium channel toxin  $\alpha$ KTx 12.5 deduced from the transcriptome analysis of the venom gland of *L. mucronatus* [32]; and (f) component sdc13973\_g1\_i2 had hits with 46% identity with the precursor of a potassium channel toxin named Tbah02745 deduced from the transcriptome analysis of the venom gland of *Tityus bahiensis* [41].

		10	20	30	
sdc10141_g1_i1 <i>S. donensis</i>	<u>MNKT</u> <b>LIF</b> <u>SML</u> <b>LL</b> <u>IT</u> <b>I</b> - - MMTIG - - <u>DADGE</u> - - - - <b>RCSSS</b>				
sdc13860_g1_i1 <i>S. donensis</i>	<u>MTKLF</u> <b>I</b> - - <u>VLLV</u> <b>STVI</b> <u>AMT</u> <b>IVPQ</b> <u>VDA</u> <b>SVG</b> - - - - <b>RCTAS</b>				
sdc13860_g1_i2 <i>S. donensis</i>	<u>MTKLF</u> <b>I</b> - - <u>VLLV</u> <b>STVI</b> <u>AMT</u> <b>IVPK</b> <u>VDA</u> <b>SGE</b> - - - - <b>KCRNS</b>				
sdc14273_g1_i2 <i>S. donensis</i>	<u>MRKFL</u> <b>I</b> - - <u>VLLV</u> <b>LTVLE</b> <u>MAI</u> <b>VPQ</b> <u>VDA</u> <b>SGE</b> <b>PP</b> <u>IWP</u> <b>CR</b> <u>TS</u>				
C5J896 <i>Oc. cayaporum</i>	<u>MNLK</u> <b>LA</b> - - <u>LVLL</u> <b>LTVIN</b> <u>VGML</u> <b>PGAT</b> <u>SNGN</u> <b>IKTD</b> <u>IK</u> <b>CYRN</b>				
H2CYS1 <i>P. cavimanus</i>	<u>MNAKF</u> <b>AY</b> - <u>VLL</u> <b>LL</b> <u>TV</u> <b>-</b> - <u>MTLV</u> <b>Y</b> - - <u>DAAG</u> <b>Q</b> <u>IS</u> <b>IN</b> <u>HRC</u> <b>GG</b> <u>S</u>				
Q6XLL6 <i>Op. carinatus</i>	<u>MNAKF</u> <b>IL</b> - <u>LLL</u> <b>V</b> <u>VTT</u> <b>-</b> - <u>TLL</u> <b>LP</b> - - <u>DAKGA</u> <b>E</b> <u>II</u> <b>-</b> - <b>RC</b> <u>SG</u> <b>T</b>				
Q6XLL5 <i>Op. carinatus</i>	<u>MNAKF</u> <b>I</b> - - <u>LLL</u> <b>V</b> <u>VTT</u> <b>-</b> - <u>MML</u> <b>LP</b> - - <u>DTKGA</u> <b>E</b> <u>VI</u> <b>-</b> - <b>RC</b> <u>SG</u> <b>S</b>				
Q6XLL7 <i>Op. carinatus</i>	<u>MNAKF</u> <b>IL</b> - <u>LLL</u> <b>V</b> <u>VTT</u> <b>-</b> - <u>TIL</u> <b>LP</b> - - <u>DTQGA</u> <b>E</b> <u>VI</u> <b>-</b> - <b>K</b> <u>CR</u> <b>TP</b>				
Q6XLL8 <i>Op. carinatus</i>	<u>MNAKF</u> <b>IL</b> - <u>LLL</u> <b>V</b> <u>VTT</u> <b>-</b> - <u>TML</u> <b>LP</b> - - <u>DTQGA</u> <b>E</b> <u>VI</u> <b>-</b> - <b>K</b> <u>CR</u> <b>TP</b>				
P0DL37 <i>U. yaschenkoi</i>	<u>MNAKL</u> <b>IY</b> - <u>LLL</u> <b>V</b> <u>VTT</u> <b>-</b> - <u>MML</u> <b>TF</b> - - <u>DTTQA</u> <b>GD</b> <u>I</u> <b>-</b> - <b>K</b> <u>CS</u> <b>GT</b>				
		40	50	60	
sdc10141_g1_i1 <i>S. donensis</i>	<b>T</b> <u>YCN</u> <b>S</b> <u>YCW</u> <b>KK</b> <u>A</u> <b>K</b> <u>C</u> <b>T</b> <u>R</u> <b>G</b> <u>K</u> <b>C</b> <u>I</u> <b>N</b> <u>K</u> <b>E</b> <u>C</u> <b>K</b> <u>C</u> <b>Y</b> <u>N</u> <b>C</b> <u>G</u> <b>R</b> <u>G</u> <b>D</b>				
sdc13860_g1_i1 <i>S. donensis</i>	<b>K</b> <u>Q</u> <b>C</b> <u>T</u> <b>A</b> <u>H</u> <b>C</b> <u>R</u> <b>M<u>R</u><b>G</b> <u>E</u><b>-</b> <u>S</u><b>H</b> <u>G</u><b>E</b> <u>C</u><b>E</b> <u>N</u><b>R</b> <u>R</u><b>C</b> <u>K</u><b>C<u>Y</u><b>-</b> - - - -</b></b>				
sdc13860_g1_i2 <i>S. donensis</i>	<b>A</b> <u>Q</u> <b>C</b> <u>K</u> <b>D</b> <u>I</u> <b>C</b> <u>R</u> <b>A<u>E</u><b>T</b> <u>G</u><b>-</b> <u>G</u><b>Q</b> <u>G</u><b>R</b> <u>C</u><b>M</b> <u>N</u><b>S</b> <u>K</u><b>C</b> <u>K</u><b>C</b> <u>F</u><b>-</b> - - - <u>V</u><b>G</b></b>				
sdc14273_g1_i2 <i>S. donensis</i>	<b>K</b> <u>Q</u> <b>C</b> <u>T</u> <b>D</b> <u>P</u> <b>C</b> <u>R</u> <b>A<u>G</u><b>G</b> <u>G</u><b>-</b> <u>S</u><b>H</b> <u>G</u><b>K</b> <u>C</u><b>M<u>N</u><b>G</b> <u>K</u><b>C<u>R</u><b>C<u>Y</u><b>-</b> - - - -</b></b></b></b>				
C5J896 <i>Oc. cayaporum</i>	<b>S</b> <u>H</u> <b>C<u>N</u><b>F</b> <u>H</u><b>C<u>E</u><b>K<u>S</u><b>S<u>Y</u><b>C<u>Q</u><b>G</b> <u>S</u><b>K<u>C</u><b>V<u>R</u><b>K<u>R</u><b>C<u>N</u><b>C<u>Y</u><b>N<u>C</u><b>P<u>L</u><b>-</b> -</b></b></b></b></b></b></b></b></b></b></b></b>				
H2CYS1 <i>P. cavimanus</i>	<b>S</b> <u>E</u> <b>C</b> <u>R</u> <b>K<u>P</u><b>C</b> <u>A</u><b>R<u>A</u><b>T</b> <u>G</u><b>Y<u>S</u><b>S</b> <u>S</u><b>K<u>C</u><b>I<u>N</u><b>G</b> <u>S</u><b>C<u>V</u><b>C<u>Y</u><b>G</b> <u>S</u><b>G</b> - - -</b></b></b></b></b></b></b>				
Q6XLL6 <i>Op. carinatus</i>	<b>R</b> <u>E</u> <b>C</b> <u>Y</u> <b>A</b> <u>P</u> <b>C<u>Q</u><b>K<u>L</u><b>T</b> <u>G</u><b>C<u>L</u><b>N</b> <u>A</u><b>K<u>C</u><b>M<u>N</u><b>K</b> <u>A</u><b>C<u>K</u><b>C<u>Y</u><b>G</b> <u>C</u><b>V</b> - - -</b></b></b></b></b></b></b>				
Q6XLL5 <i>Op. carinatus</i>	<b>K</b> <u>Q</u> <b>C<u>Y</u><b>G</b> <u>P</u><b>C<u>K</u><b>Q<u>T</u><b>G</b> <u>C</u><b>T<b>N</b> <u>S</u><b>K<u>C</u><b>M<u>N</u><b>K<u>V</u><b>C<u>K</u><b>C<u>Y</u><b>G</b> <u>C</u><b>G</b> - - -</b></b></b></b></b></b></b></b></b>				
Q6XLL7 <i>Op. carinatus</i>	<b>K</b> <u>D</u> <b>C<u>A</u><b>D</b> <u>P</u><b>C<u>R</u><b>K<u>Q</u><b>T<b>G</b> <u>C</u><b>P<b>H</b> <u>A</u><b>K</b> <u>C</u><b>M<u>N</u><b>K</b> <u>T</u><b>C</b> <u>R</u><b>C<u>H</u><b>R</b> <u>C</u><b>G</b> - - -</b></b></b></b></b></b></b>				
Q6XLL8 <i>Op. carinatus</i>	<b>K</b> <u>D</u> <b>C<u>A</u><b>D</b> <u>P</u><b>C<u>R</u><b>K<u>Q</u><b>T<b>G</b> <u>C</u><b>P<b>H</b> <u>G</u><b>K</b> <u>C</u><b>M<u>N</u><b>R</b> <u>T</u><b>C</b> <u>R</u><b>C<u>N</u><b>R</b> <u>C</u><b>G</b> - - -</b></b></b></b></b></b></b>				
P0DL37 <i>U. yaschenkoi</i>	<b>R</b> <u>Q</u> <b>C<u>W</u><b>G</b> <u>P</u><b>C<u>K</u><b>K<u>Q</u><b>T<b>T<b>C<b>T<b>N</b> <u>S</u><b>K<u>C</u><b>M<u>N</u><b>G</b> <u>K</u><b>C<u>K</u><b>C<u>Y</u><b>G</b> <u>C</u><b>V<b>G</b> - -</b></b></b></b></b></b></b></b></b></b></b></b>				

**Figure 4.** Amino acid sequences of the translated transcripts showing identity with the  $\alpha$ KTx subfamily 6, found in the transcriptome analysis of the venom gland of *S. donensis*, aligned to similar sequences. Unitprot entry numbers precede the species' names: C5J896 (potassium channel toxin  $\alpha$ KTx 6.16); H2CYS1 ( $\alpha$ KTx-like peptide); Q6XLL6 (potassium channel toxin  $\alpha$ KTx 6.9); Q6XLL5 (Potassium channel toxin  $\alpha$ KTx 6.10); Q6XLL7 (potassium channel toxin  $\alpha$ KTx 6.8); Q6XLL8 (Potassium channel toxin  $\alpha$ KTx 6.7); and P0DL37 (potassium channel toxin  $\alpha$ KTx 6.21). The predicted signal peptide is underlined and the mature peptide is in bold.

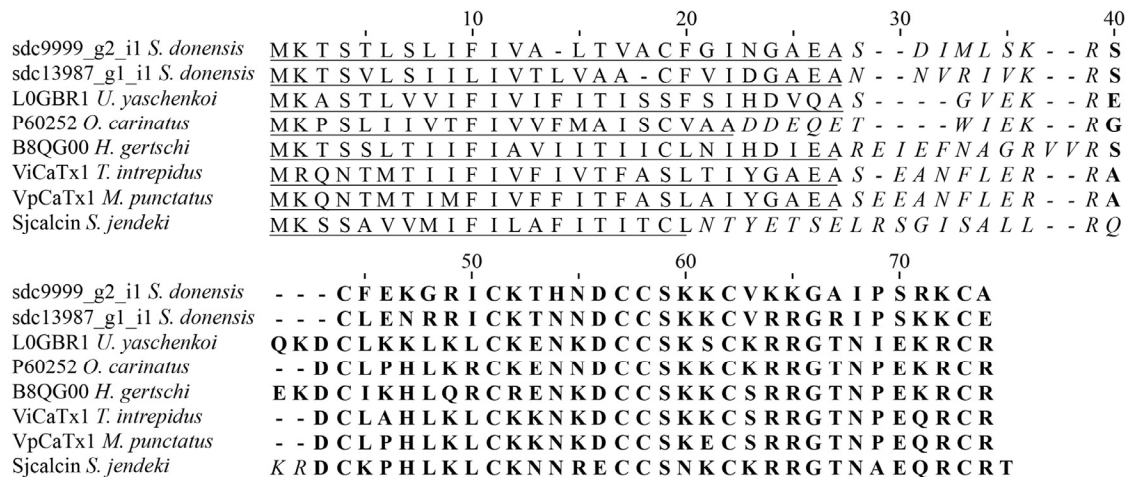
Finally, one component (sdc14251\_g2\_i1) had hits with 58% identity with the precursor of the potassium channel  $\kappa$ KTx 5.1 deduced from a cDNA cloned (and later isolated) from *Heterometrus laoticus* [42].

### Scorpine-like peptides

These are 59–75-amino acid-long peptides stabilized by three disulfide bridges [19]. As mentioned before, they were originally classified as members of the  $\beta$ -KTx family (e.g., [19]), although they were also considered “orphan” peptides because their function was not completely identified at that time. These peptides possess two different structural and functional domains: one domain with antimicrobial and cytotoxic activity and another with potassium channel-blocking activity [17]. However, recent phylogenetic analysis of scorpion toxins [5] and unpublished analysis considering only the potassium channel-blocking domain showed independent origin from the rest of the potassium channel toxin families; therefore, they should be considered as an independent subfamily. Seven sequences encoding putative scorpines were found in our analysis: (a) component sdc34997\_g1\_i1 with 59% identity with the precursor of the Hge scorpine deduced from the transcriptome analysis of the venom gland of *Hoffmannihadrurus gertschi* (= *Hadrurus gertschi*) [23]; (b) components sdc2871\_g1\_i1, sdc14222\_g4\_i1, sdc14222\_g4\_i2 and sdc20456\_g1\_i1 with 59%, 53%, 64% and 64% identity, respectively,





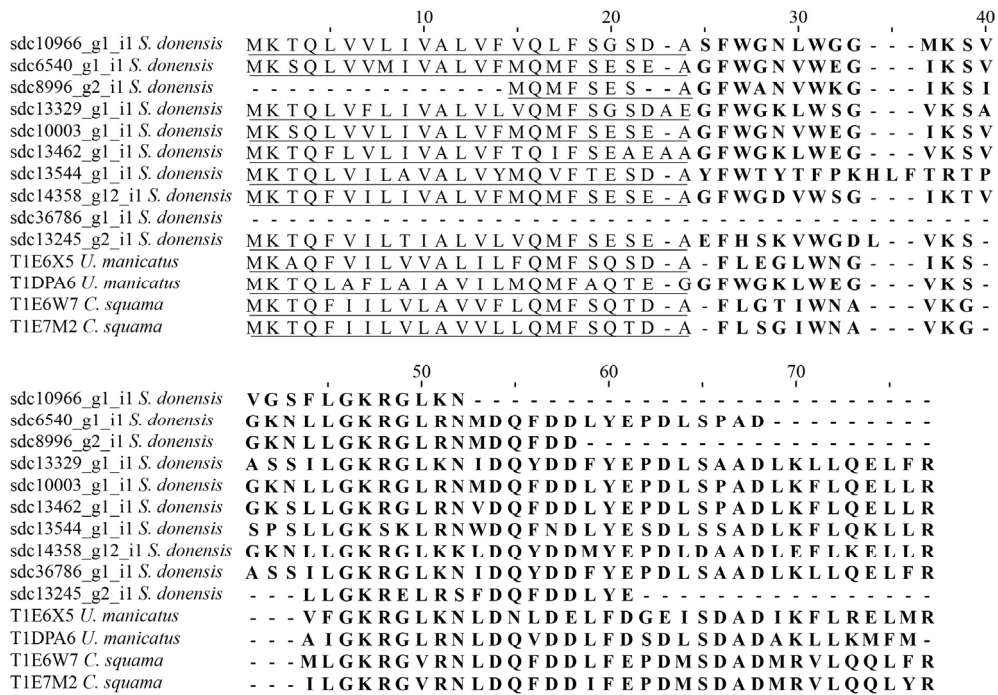


**Figure 6.** Sequence alignment of components with identity with calcins found in the transcriptome analysis of the venom gland of *S. donensis* and those that were similar. The transcripts were translated to generate the peptidic precursor sequences. Unitprot entry numbers in brackets. Components sdc9999\_g2\_i1 and sdc13987\_g1\_i1; Calcium channel toxin like 20 from *Urodacus yaschenkoi* (L0GBR1); Hadrucalcin from *Hoffmanniadrurus gertschi* (B8QG00); ViCaTx1 from *Thorellius intrepidus* [11];  $\beta$ -KTx-like peptide LaIT2 from *Liocheles australasiae* (C7G3K3); Antimicrobial peptide scorpine-like 2 *Urodacus yaschenkoi* (L0GCW2); and Opiscorpine 3 from *Op. carinatus* (Q5WQZ7). The predicted signal peptide is underlined; the mature peptide is in bold and the propeptide is in italics.

### 2.2.2. Non-Disulfide-Bridged Peptides (NDBPs)

Although often scarce in toxins, the venom of the non-Buthidae scorpions is usually rich in another large group of peptides characterized by the lack of cysteines, called the Non-Disulfide-Bridged Peptides (NDBPs). They have been shown to display multiple biological activities, including antimicrobial, cytolytic, anti-inflammatory, among others [46], therefore attracting major attention as potential clinical and therapeutic candidates in drug development [47]. The NDBPs appear to be the most abundant components in the venom of the non-buthid scorpions [11]. In our analysis, we found 31 transcripts (23% of all transcripts reported here; Figure 2) with identities to those encoding for 19 previously reported NDBPs. Among those peptides, only one sequence (sdc4413\_g1\_i1) had hits with 58% identity with the precursor of the venom antimicrobial peptide 6 deduced from a cDNA cloned from *Mesobuthus eupeus* [48,49], an amphipathic peptide exhibiting extensive cytolytic activity found in a buthid scorpion.

Among the rest of the NDBPs found in this study (but see Supplementary Table S1 for the complete list), two components, sdc24668\_g1\_i1 and sdc36542\_g1\_i1, had hits with 60% and 68% identity, respectively, with the precursor of the amphipathic peptide CT1 deduced from a cDNA cloned from *Vaejovis mexicanus* [50]; and four transcripts had hits (with identities greater than 56%) with several precursors named amphipathic peptides CT2 from different scorpion species (all deduced from cDNA) such as those cloned from *Scorpiops tibetanus* [51]; *Mesomexovis subcristatus* (= *Vaejovis subcristatus*) [50], and *Vaejovis smithi* [50]. Four more components (see Supplementary Table S2) had hits with two different amphipathic peptides deduced from cDNA cloned from *Mesomexovis punctatus* (= *Vaejovis punctatus*) [52]; ten components had hits with identities between 42% and 60% with four different CYLIP peptides deduced from the transcriptome analyses of the venom glands of *U. manicatus* and *Ce. squama* [4] (Figure 7). Other components had hits with several NDBPs reported from the venom gland of several scorpion species including *O. madagascariensis*, *Ho. gertschi*, *Heterometrus petersii* and *V. mexicanus* (see Supplementary Table S1).



**Figure 7.** Sequence alignment of components with identity with Non-Disulfide-Bridged Peptides found in the transcriptome analysis of the venom gland of *S. donensis*. The sequences derived from transcripts were translated to show the precursor peptidic sequences. For comparative purposes other known sequences are included (Unitprot entry numbers in brackets): CYLIP-Uro-1 and CYLIP-Uro-3 from *U. manicatus* (T1E6X5 and T1DPA6, respectively); and CYLIP-Cer-2 and CYLIP-Cer-3 from *Ce. squama* (T1E6W7 and T1E7M2, respectively). The predicted signal peptide is underlined and the mature peptide is in bold.

### 2.2.3. La1-Like Peptides

La1-like peptides have been found more and more frequently in the scorpion venoms. The original La1 peptide was isolated from the venom of *Liocheles australasiae* as its most abundant component [53]. The family of scorpion venom peptides defined by La1 [54] consists of long chain (73–100 aa) peptides stabilized by four disulfide bridges [6,11]. The biological activity of these peptides or their role in the venom remains unknown. In our analysis, seven sequences encoding six putative mature La1-like peptides were found: (a) sdc12897\_g1\_i1 with 48% identity with the precursor of La1-like protein 13 deduced from the transcriptomic analysis of the venom gland of *Urodoacus yaschenkoi* [6]; (b) sdc14036\_g1\_i1 with 56% identity with the precursor of La1-like protein 15 deduced from the transcriptome analysis of the venom gland of the same species; (c) sdc10164\_g1\_i1 and sdc13004\_g1\_i1 with 40%–41% identity with the precursor of a putative secreted protein deduced from cDNA cloned from the venom of *Opisthacanthus cayaporum* [55]; (d) sdc7328\_g1\_i1 with 53% identity with the precursor of the SV-SVC-Cer 1 deduced from the transcriptome analysis of *Ce. squama* [4]; (e) sdc14589\_g1\_i1 with 47% identity with the precursor of a Toxin like protein 14 deduced from cDNA and a transcriptomic analysis of the venom gland of *U. yaschenkoi* [6,7]; and (f) sdc5116\_g1\_i1 with 42% identity with the precursor of the venom protein 7 deduced from cDNA cloned from *Mesobuthus eupeus* (with Uniprot accession number E4VP44). These results were consistent with those reported in the transcriptome analyses of two species of *Scorpiops* [3–56], and of *U. yaschenkoi* [6].

### 2.2.4. Enzymes

Enzymes have been reported to be abundant components of scorpion venoms [6]. Interestingly, and not in line with the findings in other transcriptome analyses (e.g., [6]), where sequences coding for

enzymes represented about a third of the transcripts, our results show that only 16% (21 transcripts; Figure 2) putatively code for enzymes, including phospholipases, hyaluronidases and peptidases. Within these transcripts we report five transcripts with identities with one putative phospholipase deduced for one scorpion species, and six transcripts with identity with two putative hyaluronidases deduced from the venom of two scorpion species (but see Supplementary Table S1 for the complete list). Among the phospholipases, we found five transcripts with 42%–55% identity with the precursor of the Phospholipase A2 deduced from the transcriptome analysis of the venom gland of *Ho. gertschi* [23]. The six transcripts with identities with two putative hyaluronidases found were: (a) three transcripts with 42%–53% identity with the precursor of the Hyaluronidase 1 deduced from cDNA cloned from the venom of *Mesobuthus martensii* [57]; and (b) three transcripts with 46% identity (all) with the precursor of the Hyaluronidase 2 deduced from cDNA cloned from the venom of *Tityus serrulatus* [58].

### 2.2.5. Protease Inhibitors

Two different types of protease inhibitors have been isolated from the venom of scorpions.

*Ascaris-type*: These serine protease inhibitors proposed to be modulators of protease activity, have been found in different organisms protecting against toxins and other components from unwanted degradation [59]. They have a common structure with short  $\beta$  strands stabilized by five disulfide bridges [59,60]. Unlike other venom gland transcriptomes reported [6,11] where less than ten sequences were reported, 18 sequences encoding seven different ascaris-type peptides were obtained from the transcriptome analysis of the venom gland of *S. donensis*. Among these peptides (see Supplementary Table S1 for the complete list), we found 16 transcripts with identities with three putative Ascaris-type peptides deduced from cDNA and transcriptome analyses of three scorpion species including: the precursor of a cysteine-rich venom protein and a putative salivary secreted serine protease inhibitor deduced from cDNA cloned from *Pandinoides cavimanus* (= *Pandinus cavimanus*) [61]; and the precursor of the venom peptide SjAPI and SjAPI 2 deduced from cDNA cloned from *Scorpiops jendeki* [59]. Two transcripts had hits with peptides found in the saliva of a mosquito (*Stegomyia albopicta*) and a mite (*Rhipicephalus pulchellus*), respectively.

*Kunitz-type*: These peptides are known for their protease inhibitor activity (as trypsin inhibitors), but they also block (very weak) the Kv1.3 potassium channels [62]. In the transcriptome analysis of the venom gland of *S. donensis* we found two sequences: (a) sdc12570\_g1\_i1 had hits with 71% identity with the precursor of Kunitz-type serine protease inhibitor Hg1 deduced from the transcriptome of *Ho. gertschi* [23,62]; and (b) sdc31500\_g1\_i1 with 47% identity with the precursor of the peptide HW11c39 deduced from the cDNA cloned from the tarantula *Haplopelma schmidti* [63].

### 2.2.6. Other Venom Components

*CAP superfamily*: The proteins and peptides included in this superfamily have been more frequently found in the venom of scorpions. They are cysteine-rich secretory proteins, with extracellular endocrine or paracrine functions; or they might act as proteases or protease inhibitors [64]. Within this category, we also include allergens, peptides found in the venom of several arthropods (insects, arachnids and myriapods). We report 13 sequences encoding seven putative CAP peptides (see also Supplementary Table S1). Component sdc13900\_g1\_i1 had hits with 32% identity with the precursor of CAP-Iso-1 deduced from the transcriptome analysis of *Isometroides vescus* [4]. Three components had hits with 24% to 27% identity with the precursor of CAP-Lyc-1 deduced from the transcriptome analysis of the venom gland of *Lychas buchari* [4]. Two components had hits with 54% and 79% identity with the precursor CAP-Uro-1 deduced from the transcriptome analysis of *U. manicatus* [4]. Two components had hits with 32% identity (both) with the precursor Tbah00853 deduced from the transcriptome analysis of *T. bahiensis* [41]; and two components had hits with 59% identity (both) with the precursor of a putative cysteine-rich secretory peptide deduced from the transcriptome analysis of *Hottentotta judaicus* [65].

*Venom components:* Other transcripts potentially coding for venom proteins not covered in the above-described categories represent 21% of those annotated in the transcriptome analysis (Figure 2, but see also Supplementary Table S1). We found four sequences with identities ranging from 42% to 46% with the precursor of venom insulin-like growth factor binding protein 1 deduced from cDNA cloned from *Mesobuthus martensii* (only submitted to GenBank); three sequences with 44%–51% identity with the precursor of Tbah01400 deduced from the transcriptome analysis of the venom gland of *Tityus bahiensis* [41]; one sequence with 28% identity with the precursor of venom protein 29 and one sequence with 45% identity with the precursor of venom protein 302, both deduced from the transcriptome analysis of the venom gland of *Lychas mucronatus* [32]. We also found one sequence with 44% identity with the precursor of the orphan peptide AbOp5 deduced from the transcriptome analysis of *Androctonus bicolor* [66]; and one sequence with 49% identity with the precursor of Tbah02469 deduced from the transcriptome analysis of *T. bahiensis* [41]. Finally, two components had hits with 25% identity (both) with the precursor of venom allergen 5 deduced from the genome analysis of the spider *Stegodyphus mimosaroum* (only submitted to GenBank), and one component had hits with 55% identity with the precursor of putative scp tpx 1 ag5 pr1 deduced from the transcriptome analysis of the mite *Ixodes ricinus* [67].

Several transcripts putatively coding for precursors with an odd number of cysteines were found. Most of them contained a cysteine within the signal peptides (e.g., sdC14319\_g1\_i1). The presence of odd cysteines in signal peptide sequences has previously been observed, with the *Gaussia luciferase* (UniProt Q9BLZ2) signal peptide as a classic example. Since the signal peptide is cleaved by a signal peptidase upon translocation of the nascent secretory proteins to the endoplasmic reticulum, the odd cysteines play no roll in the folding or biological function of the mature secreted protein. We found a few transcripts potentially coding for proteins with odd cysteines within their mature sequence, two for a hyaluronidase (e.g., sdC14647\_g1\_i1) and three for CAP peptides (e.g., sdC3852\_g1\_i1). Proteins with an odd number of cysteines can potentially form dimers or link to other proteins. Whether this is the case for the here-reported transcripts remains to be demonstrated. The remaining deduced protein sequences with odd cysteine numbers (e.g., sdC14619\_g1\_i1) correspond to incomplete CDS, therefore the missing sequence most probably also contains an odd number of cysteines, which results in an even total.

### 2.3. Amino Acid Sequence Determination of Venom Components

The soluble fraction obtained from the venom was analyzed by LC-MS/MS, which resulted in the sequencing of 26 proteins/peptides with identity to translated transcript from the RNA-Seq (Table 3; Supplementary Table S2). Eight toxins (including four sodium channel toxins and two potassium channel toxins), two enzymes, four La1-like peptides, two CAP peptides, and six Non Disulfide Bridged Peptides were identified by mass spectrometry.

### 2.4. Phylogenetic Affinities of the Calcins, Scorpines, La1-Like Peptides and Potassium Channel $\kappa$ Toxins Found in the Transcriptomic Analysis of the Venom Gland of *S. donensis*

#### 2.4.1. Calcins

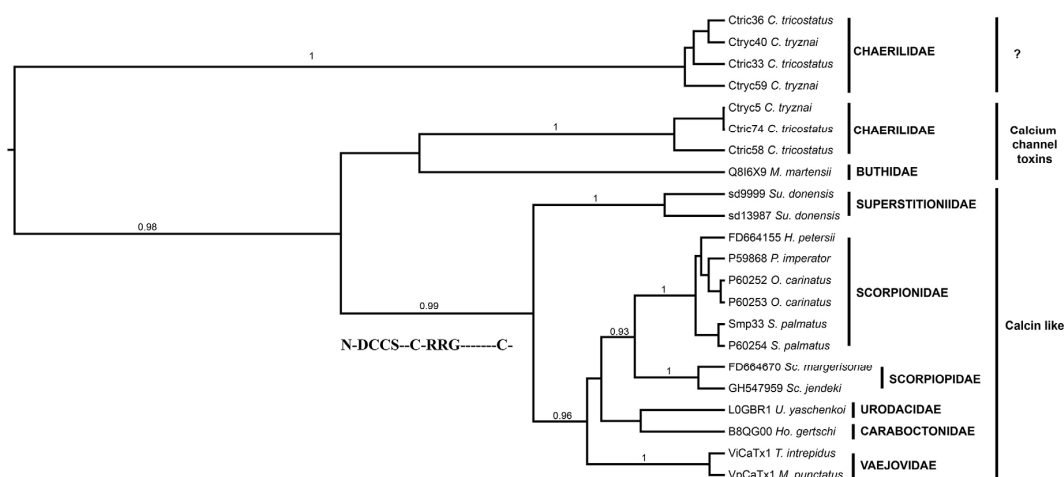
The result from the phylogenetic analysis of 22 putative and confirmed calcins reported for 14 scorpion species showed the presence of two “groups” of calcins. One clade grouped most of the sequences (18), whereas the other grouped only four sequences from two species, members of one scorpion family (Chaerilidae). The two components found in the *S. donensis* transcriptome analysis were grouped together and were related to the remaining of the calcins of the non-buthid scorpions in a clade named Calcin-like 1 (Figure 8). This clade is a sister group to the one formed by the toxin BmCa1 (Q8I6X9) from the venom of *Mesobuthus martensii* (Buthidae) plus three putative calcins found in two species from the genus *Chaerilus* (Chaerilidae). This pattern reflects the phylogenetic relationships between the two scorpion’s parvorders: Buthoida (Buthidae and Chaerilidae) and



Iuroida (Caraboctonidae, Scorpionidae, Scorpiopidae, Superstitioniidae, Urodacidae and Vaejovidae). One sequence motif was found in the signal peptide of 14 of the sequences grouped in Calcin-like 1 (see Supplementary Figure S1).

**Table 3.** Transcripts coding for components present in the venom, as validated by mass spectrometry. Abbreviations: # Pep. = number of sequenced peptide fragments corresponding to a given transcript; Seq. Cov. = sequence coverage; NaTx = sodium channel toxins; KTx = potassium channel toxins; Sc = Scorpine-like peptides; NDBP = Non-Disulfide-Bridged Peptide.

Peptide Type	Transcript	Score	# Unique Pep.	Seq. Cov.	MW (kDa)	Protein/Accession
NaTx	sd14462_g1_i1	257.51	7	91.43%	11.7	Lipolysis activating peptide 1 alpha chain/93140443
	sd14462_g2_i2	29.7	2	29.81%	11.8	Birtoxin/20137305
	sd15193_g1_i1	20.23	1	85.71%	7.1	Toxin Cll7/31376362
	sd14462_g1_i2	118.64	1	78.10%	11.8	Altitoxin/116241245
KTx	sd13949_g1_i1	27.98	2	65.38%	5.5	Toxin KTx 8/159146538
	sd14273_g1_i2	58.09	1	76.32%	4.1	KTx 6.7/74838004
Sc	sd14222_g4_i1	296.09	3	86.90%	9.3	Hg scorpine like 2/224493299
	sd14222_g4_i2	153.55	3	76.47%	9.2	Hg scorpine like 2/224493299
NDBP	sd12606_g1_i1	59.38	2	50.00%	4	Vejovine/325515699
	sd14106_g1_i1	81.77	2	81.82%	2.5	Amp1/932534523
	sd28695_g1_i1	45.7	2	77.27%	2.4	Amp2/932534537
	sd4010_g1_i1	15.61	1	100.00%	5.2	Heterin 1/485896696
	sd13544_g1_i1	83.11	1	86.79%	6.4	CYLIP Cer 2/522802549
	sd14358_g5_i1	207.52	1	62.50%	1.8	Amphiphatic peptide CT2/384382524
	sd14358_g12_i1	211.79	1	52.38%	2.3	CYLIP Cer 2/522802549
	sd6540_g1_i1	161.85	1	52.38%	2.3	CYLIP Uro 3/522802596
La1-like	sd13004_g1_i1	111.52	4	92.68%	9.1	Putative secreted protein/240247657
	sd14036_g1_i1	162.81	2	100.00%	8.4	La1 like protein 15/430802826
	sd12897_g1_i1	50.14	2	74.32%	7.9	La1 like protein 13/430802824
	sd14589_g1_i1	53.02	1	90.24%	8.8	Toxin like protein 14/430802832
Enzymes	sd14619_g1_i3	310.92	9	87.73%	67.7	Neprilysin 1/567441193
	sd14393_g1_i1	70.24	4	76.13%	26.4	Phospholipase A2/218546750
	sd14212_g1_i1	90.57	3	74.19%	24.1	Phospholipase A2/218546750
	sd14619_g1_i1	42.67	1	78.17%	83.3	Neprilysin 1/567441193
CAP	sd3852_g1_i1	97.67	3	71.20%	43.1	CAP-Uro-1/522802590
	sd13900_g1_i1	87.21	2	54.10%	38	CAP-Iso-1/522802633



**Figure 8.** Phylogenetic tree obtained from the Bayesian analysis of 22 sequences of putative and confirmed calcins from 14 scorpion species belonging to 12 genera and eight families selected from the InterPro database and the available literature. The originally reported names are used (or the UniProt or GenBank accession codes for those lacking a name), followed by the scorpion species (see Supplementary Table S3). Posterior probabilities higher than 0.76 are indicated above the branches.



The sequences grouped within the Calcin-like 1 clade have the following motif: NNDCCKKCKRRGTNPEKRCR with an E-value of  $1.2 \times 10^{-136}$  (but see also Supplementary Figure S2). Calcins from species of the Scorpionidae, Scorpipidae and Vaejovidae families were recovered as monophyletic. The calcins from scorpions of family Scorpionidae are the most studied (e.g., impercalcin or imperatoxin); and since the six sequences from this family were recovered as monophyletic, we further found two sequence motifs for that clade (see Supplementary Figure S3).

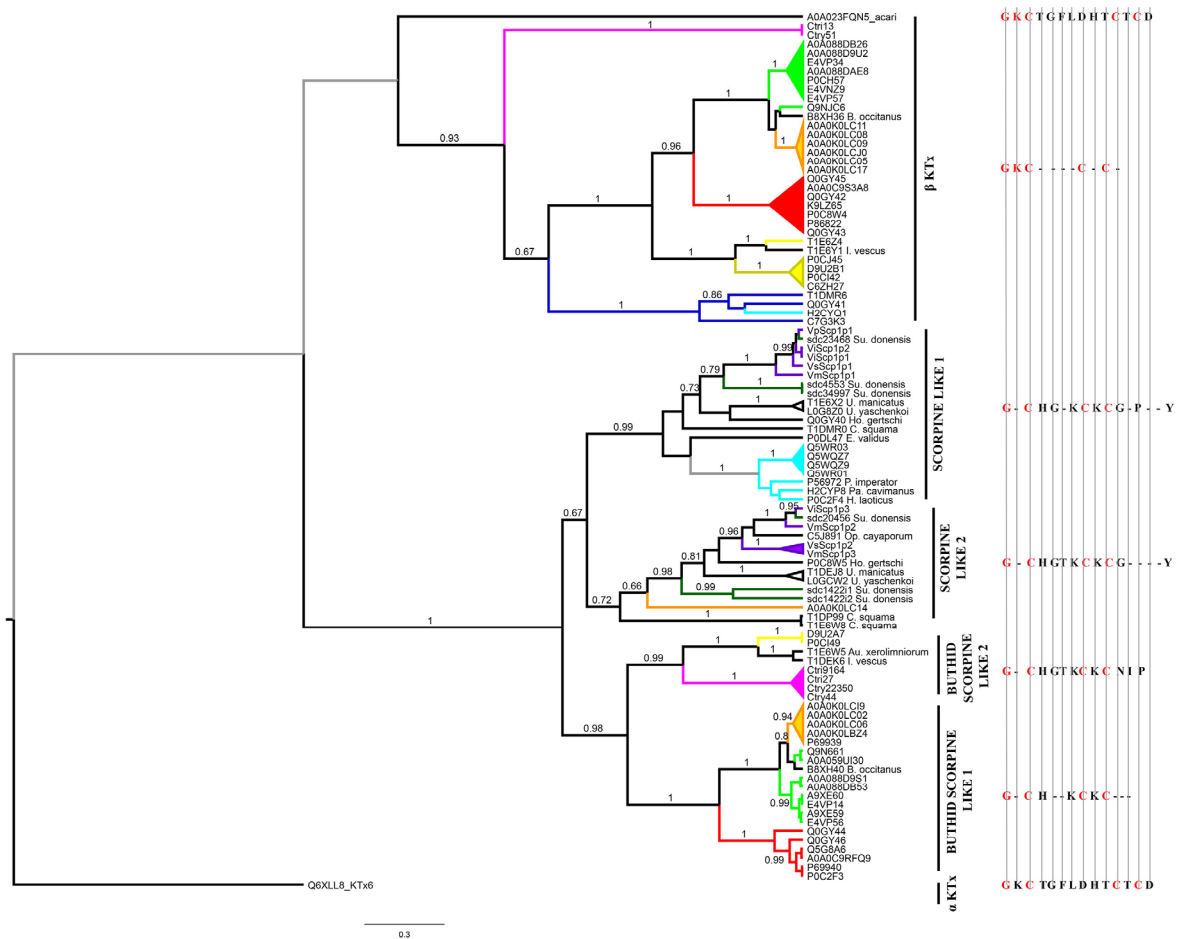
Four (out of seven) sequences from putative calcium channel toxins deduced from the transcriptome analysis of two species of *Chaerilus* were recovered as monophyletic, with high posterior probabilities. They constitute a sister group to the rest of the calcins mentioned before (Figure 8). These sequences were used as queries in BLAST searches against the UniProt database, and none had hits with calcium channel toxins, not even with scorpion calcins or with other known calcins from arthropods or mollusks. Therefore, these sequences should not be considered as true scorpion calcins unless proven otherwise (by experimental validation).

Our results were partially consistent with those presented earlier [3]. Both analyses recovered the differences between buthid and non-buthid calcins. However, the main difference between these two analyses (except for the terminal sequences used) is the fact that maurocalcin (P60254) from *Scorpio palmatus* was not grouped with the rest of the scorpionid calcins in [3]; whereas in our analysis it was recovered with the highest support within scorpionid calcins (Figure 8). Calcins have not been (or were not) found for species of the families Bothriuridae (one species in genus *Cercophonius*), Hemiscorpidae (genus *Hemiscorpius*; however, no transcriptomic data are available yet) and Hormuridae (genera *Opisthacanthus* and *Liochelis*; with no transcriptomic data available yet). However, they have been found in 12 genera of eight families; and they appear to be more common in the venom of scorpions of family Scorpionidae (thus far, four of the nine genera in the family have been studied). This suggests that calcins might be ubiquitous to scorpion venoms.

#### 2.4.2. Scorpines

The Bayesian phylogenetic analysis of 62 sequences of scorpines and putative scorpines, 34 sequences of  $\beta$ -like KTx; plus two outgroup sequences (one scorpion  $\alpha$ KTx, and one “long chain scorpion toxin” from a mite) recovered scorpines grouped into a single clade with 100% support. The  $\beta$ -like KTx clade had high posterior probabilities and it was split into three groups: one with two chaerilid  $\beta$ -like KTx sequences, another with only buthid  $\beta$ -like KTx sequences, and the last one composed by four  $\beta$ -like KTx sequences from non-buthid scorpion species, including the potassium channel toxin Hge  $\beta$ KTx (Q0GY41, Figure 9; and Supplementary Figure S4). Within the clade of scorpines, we recovered two major clades: (a) one including non-buthid species plus only one buthid sequence, with 67% of support, and subdivided further into two more clades (Scorpine-like 1 and 2) (Figure 9); and (b) the other clade including only buthid species, with 98% of support, and subdivided further into two more clades (Buthid scorpine-like 1 and 2) (Figure 9). The scorpine sequences deduced in the present contribution from *S. donensis* were included in the Scorpine-like 1 and 2 clades.

The clade Scorpine-like 1, supported by high posterior probabilities, includes the original scorpine (isolated from the *P. imperator* venom; see Supplementary Figure S5). Figure 9 shows that all scorpine sequences obtained from scorpions belonging to the same family were recovered as monophyletic, except for family Superstitioniidae, since one of our sequences (sdc23468\_g1\_i1) was recovered within the Vaejovidae family clade. With lower posterior probabilities, the clade named Scorpine-like 2 (see its motif in the Supplementary Figure S6) included one scorpine from a buthid species (AbKTx1 isolated from the venom of *Androctonus bicolor*) and 12 sequences from non-buthid scorpions. Like in Scorpine-like 1, one of our sequences (sdc20456\_g1\_i1) was grouped within Vaejovidae; and the rest of the scorpines were grouped accordingly to their familial hierarchy.



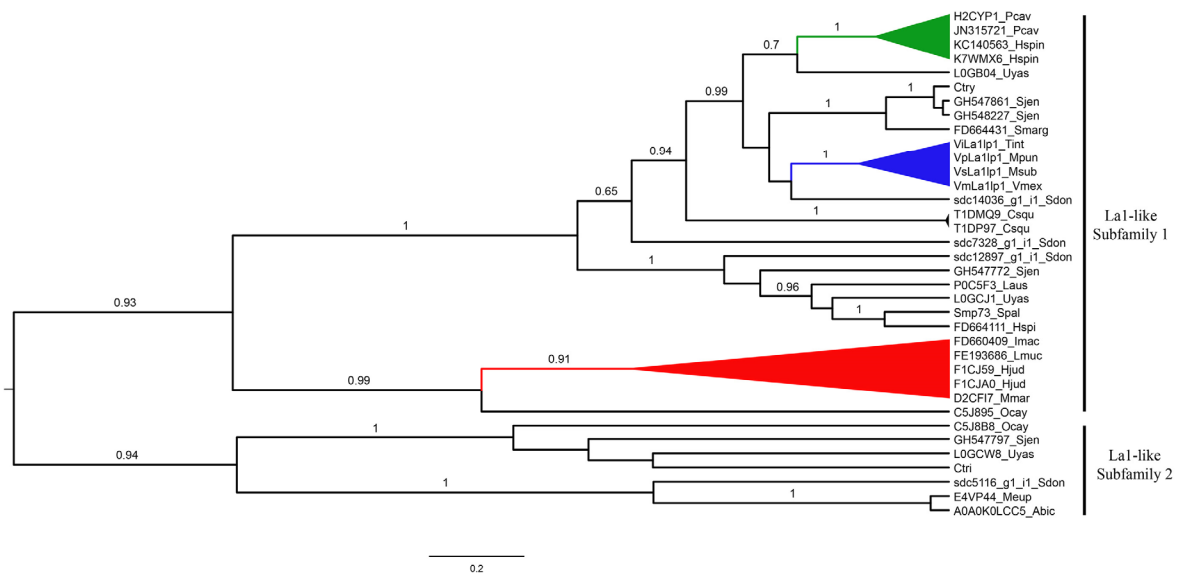
**Figure 9.** Phylogenetic tree obtained from the Bayesian analysis of 62 sequences of scorpines and putative scorpines, plus 34 sequences of  $\beta$ KTx or putative  $\beta$ KTx from 34 scorpion species of 22 genera and 10 families, and one sequence as outgroup ( $\alpha$ KTx), selected from the InterPro database and the available literature. Terminal names are composed of UniProt or GenBank accession codes and the name of the scorpion species, except for those named as in their original publications (see Supplementary Table S4). Posterior probabilities higher than 0.65 are indicated above/below branches. Clades in red represent sequences from species of genus *Tityus*; in light green sequences from species of genus *Mesobuthus*; in orange sequences from species of genus *Androctonus*; in magenta, sequences from species of genus *Chaerilus*; in yellow sequences from species of genus *Lychas*; in purple sequences from species of family Vaejovidae; in light blue sequences from species of family Scorpionidae; in dark green sequences from *S. donensis*, and in dark blue sequences from several non buthid families.

The last two clades (Buthid Scorpine-like 1 and 2; motif in the Supplementary Figure S7) were supported by high posterior probabilities and included scorpines from scorpions of parvorder Buthoida (Figure 9). Scorpines from species of genera *Androctonus*, *Chaerilus*, *Lychas* and *Tityus* were recovered as monophyletic supported by high posterior probabilities; but not those from genus *Mesobuthus* (Figure 9). As suggested by Santibáñez-López et al. [5], our results confirm that scorpines had an independent origin from the potassium channel  $\alpha$  toxins.

### 2.4.3. La1-Like Peptides

Two major clades were recovered in the phylogenetic analysis of 36 La1-like sequences from 23 scorpion species (Figure 10): the La1-like clade subfamily 1 and subfamily 2. The La1-like subfamily 1 (La1.1) clade included 29 sequences with the motif IPVGQXKXDPXXCTLYKCSXXNRRXVLXKXTC with an E-value of  $1.8 \times 10^{-307}$  (see also Supplementary Figure S8). It was further subdivided into

two clades: (a) one with all buthid La1-like peptides (red clade in Figure 10), sister group to a putative secreted protein deduced from cDNA cloned from *O. cayaporum* (UniProt accession number C5J8B8); and (b) the other clade with 23 sequences from different scorpion families. This last clade can also be divided into two subgroups, one including the original La1 from *Li. australasiae* (P0C5F3) and five more sequences from scorpions of families Scorpionidae, Scorpiopidae, Superstitioniidae and Urodacidae. The other clade included La1-like peptides from scorpions of families Bothriuridae, Chaerilidae, Scorpionidae, Scorpiopidae, Superstitioniidae, Urodacidae and Vaejoividae.

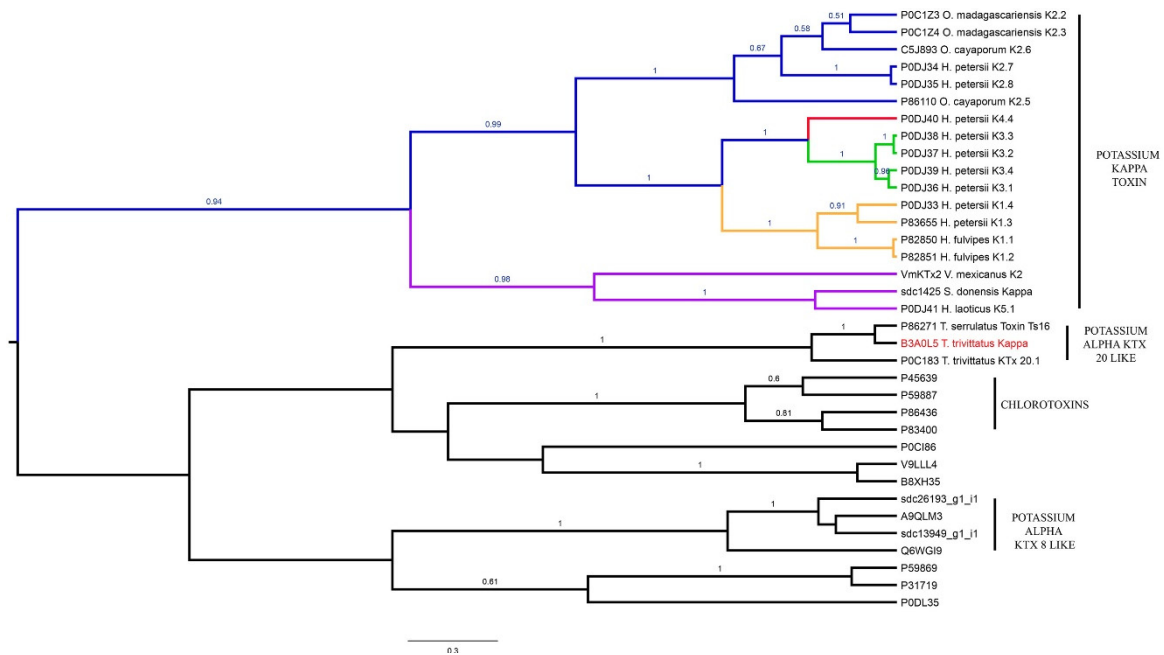


**Figure 10.** Phylogenetic tree obtained from the Bayesian analysis of 36 sequences of La1-like peptides or putative La1-like peptides from 23 scorpion species of 18 genera and nine families selected from the InterPro database and the available literature. Terminal names are composed of UniProt or GenBank accession codes and the name of the scorpion species, except for those named as in their original publications (see Supplementary Table S5). Posterior probabilities higher than 0.65 are indicated above branches. Colored clades indicate monophyletic groups of La1-like peptides from scorpions of families Buthidae (red), Scorpionidae (green) and Vaejoividae (blue).

The La1-like subfamily 2 clade included seven sequences of putative La1-like peptides from scorpions of six families (Figure 10) and the motif VTPVPPNCTLVRGRGSDCC with an E-value of  $1.08 \times 10^{-28}$  (but see Supplementary Figure S9). However, the internal relationships within this clade were not supported by high posterior probabilities (<50%), except for the clade with the two buthid SVWC-like peptides (100%). Three of the four sequences, deduced in the transcriptome analysis of the venom gland of *S. donensis* with identity with La1-like peptides, were found in the La1-like clade, but not forming a monophyletic group; whereas the other sequence was grouped within the SVWC-like peptides clade as sister group to the buthid sequences (Figure 10). These results suggest that the diversity of these peptides is greater in some families, such as Hormuridae (genus *Liocheles*) and in Superstitioniidae.

#### 2.4.4. Potassium Channel $\kappa$ Toxins ( $\kappa$ KTx)

The Bayesian phylogenetic analysis of 20  $\kappa$ KTx from eight scorpion species and 12 sequences of chlorotoxins and  $\alpha$ KTx from 11 species showed the presence of the five subfamilies (Figure 11) as proposed earlier [42]. Subfamily 1 had the motif GHGCRYSCWREGNDEETCK with an E-value of  $4.4 \times 10^{-18}$  (Supplementary Figure S10); and included four  $\kappa$ KTx deduced from cDNA cloned from the venom of *He. petersii*.



**Figure 11.** Phylogenetic tree obtained from the Bayesian analysis of 20 sequences of potassium channel  $\kappa$  toxins ( $\kappa$ KTxs) from eight scorpion species of four genera and three families; and 12 sequences of potassium channel  $\alpha$  toxins and chlorotoxins as outgroup, selected from the InterPro database and available literature (see Supplementary Table S6). Posterior probabilities are indicated above branches. Colored clades indicate the monophyletic subfamilies proposed earlier: subfamily 1 (orange); subfamily 2 (blue); subfamily 3 (green); subfamily 4 (red); and subfamily 5 (purple). The name in red shows  $\kappa$  buthithoxin.

Subfamily 2 had the motif DPCVEVCLQHTGNVKECEEAC with an E-value  $3.9 \times 10^{-34}$ ; and included two  $\kappa$ KTx from two scorpionid species of genus *Heterometrus* (*He. petersii* and *He. fulvipes*; Supplementary Figure S11). Subfamilies 3 and 4 were recovered as sister groups, while subfamily 4 is only represented by the  $\kappa$ KTx 4.1 deduced from the transcriptome analysis of *H. petersii* (P0DJ40). Subfamily 3 included four sequences also deduced from the analysis of the same scorpion species.

Finally, subfamily 5 had the motif MKVLPFLFVFLIVCVMLPTEASCTQ (in the signal peptide), but with a low E-value ( $9 \times 10^{-5}$ ). This subfamily was represented by three sequences including the  $\kappa$ KTx found in *S. donensis*, one  $\kappa$ KTx from the species *V. mexicanus* and  $\kappa$ KTx 5.1 from *H. laoticus* (the original member of this subfamily; Supplementary Figure S12).

Our results differed from those presented earlier (e.g., [5]), since we recovered the potassium channel  $\kappa$  toxins family as monophyletic. In the previous analysis [5], subfamily 5 was not recovered as closely related to the  $\kappa$ KTxs but to chlorotoxins. Of course, the main scope of that study was to explore the status of the current classification and the phylogenetic affinities of the CS $\alpha\beta$  toxins. Both analyses recovered  $\kappa$  buthithoxin Tt2b, Ts28 and Toxin Ts16 as a monophyletic group, but it was not closely related to the rest of the  $\kappa$ KTx. This is not surprising since Saucedo et al. [68] mentioned that  $\kappa$  BUTX Tt2b and Ts16 toxins have the CS $\alpha\beta$  motif, but adopt a CS $\alpha\alpha$  motif in solution. These authors elegantly discussed the 3D structure of these toxins. However, they hesitated to establish a new subfamily for these toxins. Our results are in accordance with their proposed relationship between these three toxins, but not with their proposed relationship between these toxins and the  $\kappa$ KTxs. We lack evidence to establish a new subfamily for these toxins, so we decided to include them in the  $\alpha$ KTx 20 subfamily [69].

### 3. Conclusions

The 135 transcripts annotated here highlight the differences between buthid and non-buthid scorpion venoms. The annotated transcripts constitute just a small fraction of the whole assembled

transcriptome. This reflects the lack of thorough knowledge on the toxinology and cellular biology of scorpions and prompts for a deeper investigation in the area. Future research on different unexplored scorpion families would contribute to the understanding on the diversity of venom components and their evolution. The discovery of calcins, scorpines and La1-like peptides in different scorpion families suggests the ubiquitous presence of these components in scorpion venoms. Some of them (e.g., calcins and La1-like peptides) can contribute to our knowledge on venom evolution when more families are sampled. Given that our results on the phylogenetic affinities of calcins partially mirror the phylogenetics of some scorpion families, we recommend the use of these peptides in generic phylogenetic reconstructions.

#### 4. Materials and Methods

##### 4.1. Biological Material

Scorpion specimens were collected in Ensenada Baja California, Mexico on August 2015. Permit for collection was issued by SEMARNAT (Scientific Permit FAUT-0175 granted to Oscar Francke, see acknowledgments). They were maintained in plastic boxes with water and fed with crickets. The venom was extracted by electric stimulation. Two specimens were sacrificed to process the telsons, and the other two were deposited at the Colección Nacional de Arácnidos in the Instituto de Biología of the Universidad Nacional Autónoma de México, in Mexico City.

##### 4.2. Molecular Mass Determination and Protein Identification

The venom obtained was solubilized in water and centrifuged 10,000 g for 10 min. The soluble fraction of the total scorpion venom was desalted using ZipTip C18 (Millipore, Billerica, MA, USA) and 5 µg of the desalted material was then analyzed by applying LC-MS/MS. The LC was performed using an Accela HPLC from Thermo Scientific (San Jose, CA, USA) at a flow rate of 500 nL/min (splitter 19:1). The RP-C18 column (75 µm ID × 100 mm) used was constructed in house and a gradient ranging from 5% to 70% solvent B over 120 min was applied. Solvent A was made up of 0.1% acetic acid/water, and solvent B consisted of 0.1% acetic acid/acetonitrile. Eluted venom components were electrosprayed with a nano-electrospray at a voltage of 2.4-kV into an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). MS data acquisition was carried out automatically, using a data acquisition method specifically devised for molecular mass determinations. The data were automatically deconvoluted using the Xcalibur software for each 10 min run. Adducts formed by different combinations of Na and K as wells as single and double methionine oxidations were eliminated of the mass list.

Protein identification of the components present in the crude venom was performed using tryptic digestion in solution and subsequently analysis by LC-MS. First, 45 µg of the soluble venom was reduced with DTT 10 mM, alkylated with iodoacetamide 55 mM and desalted using ZipTip C18 (Millipore, Billerica, MA, USA). Secondly, the tryptic digestion was applied into an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) using the same gradient described previously for molecular mass determination. MS data acquisition was accomplished automatically, using a method specifically devised for “de novo” sequencing. MS data were acquired with 30,000 resolutions in the FT-Orbitrap analyzer in the positive ion mode and only the five most intense doubly and triply charged ions were selected for dissociation by CID (Collision Induced Dissociation) and HCD (High-energy Collision Dissociation). Dynamic exclusion of 60 s was enabled and a pre-exclusion of 30 s was applied. Finally, the normalized collision energy was set at 35 arbitrary units, with an activation Q of 0.250 and an activation time of 30 ms for both CID and HCD. MS data were search against the cDNA database obtained from previous transcriptomic analysis using the Protein Discoverer 1.4 program (Thermo-Fisher Co., San Jose, CA, USA).



#### 4.3. RNA Extraction, RNA-Seq and Venom Gland Transcriptome Assembly

The telson from two specimens were dissected under RNase-free conditions and pooled into a single tube. RNA was isolated using the SV Total RNA Isolation System (Promega, Madison, WI, USA) following the protocol provided by the manufacturer. Briefly, the dissected telsons were manually macerated to homogeneity with a Kontes microtube pellet pestle rod (Daigger, Vernon Hills, IL, USA) in a 1.5 mL microtube with the provided RNA Lysis Buffer. After dilution with the RNA Dilution Buffer the sample was heated at 70 °C for 3 min, then centrifuged to discard all cellular debris. The cleared lysate was mixed with 95% ethanol and transferred to one of the spin baskets supplied by the kit. After washing with the RNA Wash Solution, the sample was treated with the provided DNase for 15 min and then washed twice with the RNA Wash Solution. After centrifugation, total RNA was recovered in Nuclease-Free Water. The RNA was quantitated with a Nanodrop 1000 (Thermo Scientific) and its integrity was confirmed using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

A complementary DNA (cDNA) library was constructed from the total RNA obtained, using the Illumina TruSeq Stranded mRNA Sample Preparation Kit, following the protocol provided by the supplier. Automated DNA sequencing was performed at the Massive DNA Sequencing Facility in the Institute of Biotechnology (Cuernavaca, Mexico) with a Genome Analyzer Iix (Illumina, San Diego, CA, USA), using a 72 bp paired-end sequencing scheme over cDNA fragments ranging in size of 200–400 bp. Each library consisted of two fastq files (forward and reverse reads), from which the adaptors were clipped-off. The quality of cleaned raw reads was assessed by means of the FastQC program (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>).

This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GFCD00000000. The version described in this paper is the first version, GFCD10000000.

The short reads were assembled into contigs in a de novo fashion with the Trinity software (v. 2.0.3), using the standard protocol [70], executing the strand-specific parameter and normalizing the reads. Basic statistics as the number of “genes”, transcripts and contigs were obtained by running the TrinityStats.pl script.

The assembled contigs were used as queries to perform a BLAST analysis against UniProt (<http://www.uniprot.org>). They were then annotated with Trinotate (<https://trinotate.github.io/>, [73]). The signal peptides were predicted using the SignalP 4.0 server (<http://www.cdbb.dtu.dk/services/SignalP/>) and the propeptides were determined with the ProP 1.0 server (<http://www.cbs.dtu.dk/services/ProP/>). The theoretical molecular weights of the putative mature peptides were obtained using the ProtParam server (<http://web.expasy.org/protparam>).

#### 4.4. Multiple Sequence Alignments, Phylogenetic Analysis and Motif Search

Multiple sequence alignments of the sequences found and their similar sequences were obtained using the online version of MAFFT ver. 7.0 [71] (<http://mafft.cbrc.jp/alignment/server/>). Alignments were edited in Jalview [72] and Adobe Illustrator CS6. We retrieved 188 sequences from the InterProt database (<http://www.ebi.ac.uk/interpro/>), GenBank database (<http://www.ncbi.nlm.nih.gov/>) or the available literature corresponding to: (a) 22 calcins from 14 scorpion species in 12 genera and 8 families; (b) 96 scorpines from 34 species in 22 genera and 10 families, the sequence of Toxin 38 from a mite, one sequence of  $\alpha$ KTx from one scorpion species as outgroup; (c) 36 La1-like peptides from 23 species in 18 genera and 9 families; and (d) 20 sequences of potassium channel  $\kappa$  toxins from 8 scorpion species in 4 genera and 3 families, plus 12 sequences of potassium channel  $\alpha$  toxins and chlorotoxins from nine scorpion species as outgroups.

We constructed four matrices (aligned each separately with MAFFT ver. 7.0 [71]) as follows: (a) 22 calcins with a length of 90 amino acids; (b) 98 terminal scorpines and outgroups with a length of 134 amino acids; (c) 36 terminal-La1 like peptides with a length of 161 amino acids; and (d) 35 terminal sequences including  $\kappa$ KTx, Chlorotoxins and  $\alpha$ KTx with a length of 102 amino acids. The best fitting model of protein evolution was selected using ProtTest 3 [73,74] and the Akaike information criterion

on the basis of which the following models were selected: (a) for the calcin matrix the JTT + I + G was chosen; (b) for the scorpine matrix the LG + I + G was chosen; (c) and for the La1-like peptides and the  $\kappa$ KTxs matrices the WAG + I + G was chosen.

The phylogenetic analyses were conducted under the Bayesian inference using the algorithm implemented in the software BEAST 1.8 [75]. These analyses comprised 30 million generations, sampling every 1000 generations, and those sampled before stationarity discarded using the burn-in option in TreeAnnotator (included in BEAST 1.8 software package). The resulting topologies were edited with FigTree 1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) and Adobe Illustrator Cs6.

Considering monophyletic clades with high posterior probabilities (>76%), sets of their included sequences of amino acids were selected to establish motifs using the Multiple Em for Motif Elicitation server (MEME 4.10.0 at <http://meme-suite.org/tools/meme> [76]) and the Motif Alignment & Search Tool (MAST) to determine whether the selected motif was a unique signature or not.

The prediction of the signal peptide, propeptide and mature peptide was performed using the ArachnoServer (<http://www.arachnoserver.org/spiderP.html>); SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP/>), and ProP 1.0 Server (<http://www.cbs.dtu.dk/services/ProP/>).

**Supplementary Materials:** The following are available online at [www.mdpi.com/2072-6651/8/12/367/s1](http://www.mdpi.com/2072-6651/8/12/367/s1), Supplementary Tables S1–S6; Supplementary Figures S1–S12.

**Acknowledgments:** The authors are grateful to Oscar Francke from the Biology Institute of UNAM for support during collection of the scorpions used in this work (Scientific Permit FAUT-0175, from SEMARNAT). This work was partially supported by grant SEP-CONACyT No.237864, and IN 203416 from Dirección General de Asuntos del Personal Académico (DGAPA)—UNAM given to LDP. Carlos E. Santibáñez-López and Jimena I. Cid-Urbe are recipients of postdoctoral and PhD program scholarships from CONACyT (No. 237864 and No. 404460, respectively). The authors are acknowledged to Erika Patricia Meneses Romero for technical assistance on Mass spectrometry experiments. We are grateful to Gloria T. Vázquez Castro, Ricardo A. Grande Cano and M.S. Verónica Jiménez Jacinto at the DNA massive sequencing and the Bioinformatics Units from the Instituto de Biotecnología of UNAM for their technical support.

**Author Contributions:** C.E.S.-L., E.O. and L.D.P conceived and designed the experiments. E.O., L.D.P. and C.V.F.B. performed the experiments. C.E.S.-L., J.I.C.-U., E.O., C.V.F.B. and L.D.P analyzed the data. C.E.S.-L. performed the phylogenetic analysis. L.D.P. and C.E.S.-L. contributed reagents, materials and analysis tools. C.E.S.-L., J.I.C.-U., E.O. and L.D.P wrote the paper. All authors approved the final version of this manuscript.

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

## References

1. Santibáñez-López, C.E.; Francke, O.F.; Ureta, C.; Possani, L.D. Scorpions from Mexico: From species diversity to venom complexity. *Toxins* **2016**, *8*, 2. [[CrossRef](#)] [[PubMed](#)]
2. Sharma, P.P.; Fernández, R.; Esposito, L.A.; González-Santillán, E.; Monod, L. Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. *Proc. R. Soc. B* **2015**, *282*, 20142953. [[CrossRef](#)] [[PubMed](#)]
3. 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. Proteom.* **2012**, *75*, 1563–1576. [[CrossRef](#)] [[PubMed](#)]
4. Sunagar, K.; Undheim, E.A.; Chan, A.H.; Koludarov, I.; Muñoz-Gómez, S.A.; Antunes, A.; Fry, B.G. Evolution stings: The origin and diversification of scorpion toxin Peptide scaffolds. *Toxins* **2013**, *5*, 2456–2487. [[CrossRef](#)] [[PubMed](#)]
5. Santibáñez-López, C.E.; Possani, L.D. Overview of the Knottin scorpion toxin-like peptides in scorpion venoms: Insights on their classification and evolution. *Toxicon* **2015**, *107*, 317–326. [[CrossRef](#)] [[PubMed](#)]
6. Luna-Ramírez, K.; Quintero-Hernández, V.; Juárez-González, V.R.; Possani, L.D. Whole transcriptome of the venom gland from *Urodacus yaschenkoi* scorpion. *PLoS ONE* **2015**, *10*, e0127883. [[CrossRef](#)] [[PubMed](#)]
7. Luna-Ramírez, K.; Quintero-Hernández, V.; Vargas-Jaimes, L.; Batista, C.V.; Winkel, K.D.; Possani, L.D. Characterization of the venom from the Australian scorpion *Urodacus yaschenkoi*: Molecular mass analysis of components, cDNA sequences and peptides with antimicrobial activity. *Toxicon* **2013**, *63*, 44–54. [[CrossRef](#)] [[PubMed](#)]

8. Vignoli, V.; Prendini, L. Systematic revision of the troglomorphic North American Scorpion family Typhlochactidae (Scorpiones: Chactoidae). *Bull. Am. Mus. Nat. Hist.* **2009**, *326*, 1–94. [[CrossRef](#)]
9. Williams, S.C. Scorpions of Baja California, Mexico, and Adjacent Islands. *Occ. Pap. Calif. Acad. Sci.* **1980**, *135*, 1–127.
10. Darbon, H. Animal toxin and ion channels. *J. Soc. Biol.* **1999**, *193*, 445–450. [[PubMed](#)]
11. Quintero-Hernández, V.; Ramírez-Carreto, S.; Romero-Gutiérrez, M.T.; Valdez-Velázquez, L.L.; Becerril, B.; Possani, L.D.; Ortiz, E. Transcriptome analysis of scorpion species belonging to the *Vaejovis* genus. *PLoS ONE* **2015**, *10*, e0117188. [[CrossRef](#)] [[PubMed](#)]
12. Xiao, L.; Gurrola, G.B.; Zhang, J.; Valdivia, C.R.; SanMartín, M.; Zamudio, F.Z.; Zhang, L.; Possani, L.D.; Valdivia, H.H. Structure-function relationships of peptides forming the calcin family of ryanodine receptor ligands. *J. Gen. Phys.* **2016**, *147*, 375–394. [[CrossRef](#)] [[PubMed](#)]
13. Valdivia, H.H.; Fuentes, O.; El-Hayek, R.; Morrissette, J.; Coronado, R. Activation of the ryanodine receptor Ca<sup>2+</sup>-release channel of sarcoplasmic reticulum by a novel scorpion venom. *J. Biol. Chem.* **1991**, *266*, 19135–19138. [[PubMed](#)]
14. Morrissette, J.; Beurg, M.; Sukhareva, M.; Coronado, R. Purification and characterization of ryanotoxin, a peptide with actions similar to those of ryanodine. *Biophys. J.* **1996**, *71*, 707–721. [[CrossRef](#)]
15. Ji, Y.H.; Liu, Y.; Yu, K.; Ohishi, T.; Hoshino, M.; Mochizuki, T.; Yanaihara, N. Amino acid sequence of a novel activator of ryanodine receptor on skeletal muscle. *Chin. Sci. Bull.* **1997**, *42*, 952–956. [[CrossRef](#)]
16. Valdivia, H.H.; Kirby, M.S.; Lederer, W.J.; Coronado, R. Scorpion toxins targeted against the sarcoplasmic reticulum Ca<sup>2+</sup>-release channel of skeletal and cardiac muscle. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 12185–12189. [[CrossRef](#)] [[PubMed](#)]
17. Quintero-Hernández, V.; Jiménez-Vargas, J.M.; Gurrola, G.B.; Valdivia, H.H.; Possani, L.D. Scorpion venom components that affect ion-channels function. *Toxicon* **2013**, *76*, 328–342. [[CrossRef](#)] [[PubMed](#)]
18. Conde, R.; Zamudio, F.Z.; Rodríguez, M.H.; Possani, L.D. Scorpine, an anti-malaria and anti-bacterial agent purified from scorpion venom. *FEBS Lett.* **2000**, *471*, 165–168. [[CrossRef](#)]
19. Diego-García, E.; Schwartz, E.F.; D'Suze, G.; González, S.A.; Batista, C.V.; Garcia, B.I.; Rodríguez de la Vega, R.C.; Possani, L.D. Wide phylogenetic distribution of scorpine and long-chain  $\beta$ -KTx-like peptides in scorpion venoms: Identification of “orphan” components. *Peptides* **2007**, *28*, 31–37. [[CrossRef](#)] [[PubMed](#)]
20. Laraba-Djebari, F.; Adi-Bessalem, S.; Hammoudi-Triki, D. Scorpion Venoms: Pathogenesis and Biotherapies. In *Scorpion Venoms*; Gopalakrishnakone, P., Possani, L.D., Schwartz, E.F., Rodríguez de la Vega, R.C., Eds.; Springer: Dordrecht, The Netherlands, 2015; pp. 63–86.
21. Carbone, E.; Wanke, E.; Prestipino, G.; Possani, L.D.; Maelicke, A. Selective blockage of voltage-dependent K<sup>+</sup> channels by a novel scorpion toxin. *Nature* **1982**, *296*, 90–91. [[CrossRef](#)] [[PubMed](#)]
22. Doyle, D.A.; Morais-Cabral, J.; Pfuetzner, R.A.; Kuo, A.; Gulbis, J.M.; Cohen, S.L.; Chait, B.R.; MacKinnon, R. The structure of potassium channel: Molecular basis of K<sup>+</sup> conduction and selectivity. *Science* **1998**, *280*, 69–77. [[CrossRef](#)] [[PubMed](#)]
23. Schwartz, E.F.; Diego-García, E.; Rodríguez de la Vega, R.C.; Possani, L.D. Transcriptome analysis of the venom gland of the Mexican scorpion *Hadrurus gertschi* (Arachnida: Scorpiones). *BMC Genom.* **2007**, *8*, 119. [[CrossRef](#)] [[PubMed](#)]
24. Possani, L.D.; Becerril, B.; Delepierre, M.; Tytgat, J. Scorpion toxins specific for Na<sup>+</sup>-channels. *Eur. J. Biochem.* **1999**, *264*, 287–300. [[CrossRef](#)] [[PubMed](#)]
25. Rodríguez de la Vega, R.C.; Possani, L.D. Overview of scorpion toxins specific for Na<sup>+</sup> channels and related peptides: Biodiversity, structure-function relationships and evolution. *Toxicon* **2005**, *46*, 831–844. [[CrossRef](#)] [[PubMed](#)]
26. InterPro: Scorpion Long Chain Toxin/Defensin (IPR002061). Available online: <http://www.ebi.ac.uk/interpro/entry/IPR002061> (accessed on 17 March 2016).
27. Guerrero-Vargas, J.A.; Mourao, C.B.; Quintero-Hernández, V.; Possani, L.D.; Schwartz, E.F. Identification and phylogenetic analysis of *Tityus pachyurus* and *Tityus obscurus* novel putative Na<sup>+</sup>-channel scorpion toxins. *PLoS ONE* **2012**, *7*, e30478. [[CrossRef](#)] [[PubMed](#)]
28. Inceoglu, A.B.; Lango, J.; Pessah, I.N.; Hammock, B.D. Three structurally related, highly potent, peptides from the venom of *Parabuthus transvaalicus* possess divergent biological activity. *Toxicon* **2005**, *45*, 727–733. [[CrossRef](#)] [[PubMed](#)]

29. Inceoglu, A.B.; Lango, J.; Wu, J.; Hawkins, P.; Southern, J.; Hammock, B.D. Isolation and characterization of a novel type of neurotoxic peptide from the venom of the South African scorpion *Parabuthus transvaalicus*. *Eur. J. Biochem.* **2001**, *268*, 5407–5413. [CrossRef] [PubMed]
30. Batista, C.V.F.; D'Suze, G.; Gómez-Lagunas, F.; Zamudio, F.Z.; Encarnacion, S.; Sevcik, C.; Possani, L.D. Proteomic analysis of *Tityus discrepans* scorpion venom and amino acid sequence of novel toxins. *Proteomics* **2006**, *6*, 3718–3727. [CrossRef] [PubMed]
31. D'Suze, G.; Schwartz, E.F.; García-Gómez, B.I.; Sevcik, C.; Possani, L.D. Molecular cloning and nucleotide sequence analysis of genes from a cDNA library of the scorpion *Tityus discrepans*. *Bichimie* **2009**, *91*, 1010–1019. [CrossRef] [PubMed]
32. Zhao, R.; Ma, Y.; He, Y.; Di, Z.; Wu, Y.L.; Cao, Z.J.; Li, W.X. Comparative venom gland transcriptome analysis of the scorpion *Lychas mucronatus* reveals intraspecific toxic gene diversity and new venomous components. *BMC Genom.* **2010**, *11*, 452.
33. García-Gómez, B.I.; Olamendi-Portugal, T.C.; Paniagua, J.; van der Walt, J.; Dyason, K.; Possani, L.D. Heterologous expression of a gene that codes for Pg8, a scorpion toxin of *Parabuthus granulatus*, capable of generating protecting antibodies in mice. *Toxicon* **2009**, *53*, 770–778. [CrossRef] [PubMed]
34. Soudani, N.; Gharbi-Chili, J.; Srairi-Abid, N.; Martin-El, Y.C.; Planells, R.; Margotat, A.; Torresani, J.; El Ayeb, M. Isolation and molecular characterization of LVP1 lipolysis activating peptide from scorpion *Buthus occitanus tunetanus*. *Biochim. Biophys. Acta* **2005**, *1747*, 47–56. [CrossRef] [PubMed]
35. Rodríguez de la Vega, R.C.; Possani, L.D. Current views on scorpion toxins specific for K<sup>+</sup>-channels. *Toxicon* **2004**, *43*, 865–875. [CrossRef] [PubMed]
36. InterPro: Scorpion Short Chain Toxin, Potassium Channel Inhibitor (IPR001947). Available online: <http://www.ebi.ac.uk/interpro/entry/IPR002061> (accessed on 17 March 2016).
37. InterPro: Long Chain Scorpion Toxin Family (IPR029237). Available online: <http://www.ebi.ac.uk/interpro/entry/IPR029237> (accessed on 17 March 2016).
38. Zhu, S.Y.; Huys, I.; Dyason, K.; Verdonck, F.; Tytgat, J. Evolutionary trace analysis of scorpion toxins specific for K-channels. *Proteins* **2004**, *54*, 361–370. [CrossRef] [PubMed]
39. Wu, W.; Yin, S.; Ma, Y.; Wu, Y.L.; Zhao, R.; Gan, G.; Ding, J.; Cao, Z.; Li, W. Molecular cloning and electrophysiological studies on the first K(+) channel toxin (LmKTx8) derived from scorpion *Lychas mucronatus*. *Peptides* **2007**, *28*, 2306–2312. [CrossRef] [PubMed]
40. Luna-Ramírez, K.; Bartok, A.; Restano-Cassulin, R.; Quintero-Hernández, V.; Coronas, F.; Christinseen, J.; Wright, C.E.; Panyi, G.; Possani, L.D. Structure, molecular modeling and function of Urotoxin, the first potassium channel blocker from the venom of Australian scorpion *Urodacus yaschenko*. *Mol. Pharmacol.* **2014**, *86*, 28–41. [CrossRef] [PubMed]
41. De Oliveira, U.C.; Candido, D.M.; Coronado-Dorce, V.A.; Junquiera-de-Azevedo, I.D.L. The transcriptome recipe for the venom cocktail of *Tityus bahiensis* scorpion. *Toxicon* **2015**, *95*, 52–61. [CrossRef] [PubMed]
42. Vandendriessche, T.; Kopljar, I.; Jenkins, D.P.; Diego-García, E.; Abdel-Mottaleb, Y.; Vermassen, E.; Clynen, E.; Schoofs, L.; Wulff, H.; Snyders, D.; et al. Purification, molecular cloning and functional characterization of HelaTx1 (*Heterometrus laoticus*): The first member of a new kappa KTx. *Biochem. Pharmacol.* **2012**, *83*, 1307–1317. [CrossRef] [PubMed]
43. Narasimhan, L.; Singh, J.; Humblet, C.; Guruprasad, K.; Blundell, T. Snail and spider toxins share a similar tertiary structure and 'cystine motif'. *Nat. Struct. Biol.* **1994**, *1*, 850–852. [CrossRef] [PubMed]
44. Zhu, S.Y.; Darbon, H.; Dyason, K.; Verdonck, F.; Tytgat, J. Evolutionary origin of inhibitor cysteine knot peptides. *FASEB J.* **2003**, *17*, 1765–1767. [PubMed]
45. Kozlov, S.; Malyavka, A.; McCutchen, B.; Lu, A.; Schepers, E.; Hermann, R.; Grishin, E. A novel strategy for the identification of toxinlike structures in spider venom. *Proteins* **2005**, *59*, 131–140. [CrossRef] [PubMed]
46. Almaaytah, A.; Albalas, O. Scorpion venom peptides with non disulfide bridges: A review. *Peptides* **2014**, *51*, 35–45. [CrossRef] [PubMed]
47. Ortiz, E.; Gurrola, G.B.; Schwartz, E.F.; Possani, L.D. Scorpion venom components as potential candidates for drug development. *Toxicon* **2015**, *93*, 125–135. [CrossRef] [PubMed]
48. Gao, B.; Sherman, P.; Luo, L.; Bowie, J.; Zhu, S. Structural and functional characterization of two genetically related meucin peptides highlights evolutionary divergence and convergence in antimicrobial peptides. *FASEB J.* **2009**, *23*, 1230–1245. [CrossRef] [PubMed]



49. Farajzadeh-Sheikh, A.; Jolodar, A.; Ghaemmaghami, S. Sequence characterization of cDNA sequence of encoding of an antimicrobial peptide with no disulfide bridge from the Iranian *Mesobuthus eupeus* venomous gland. *Red. Crescent. Med. J.* **2013**, *15*, 36–41. [[CrossRef](#)] [[PubMed](#)]
50. Ramírez-Carretero, S.; Quintero-Hernández, V.; Jiménez-Vargas, J.M.; Corzo, G.; Possani, L.D.; Ortiz, E. Gene cloning and functional characterization of four novel antimicrobial-like peptides from scorpions of the family Vaejovidae. *Peptides* **2012**, *34*, 290–295. [[CrossRef](#)] [[PubMed](#)]
51. Cao, L.; Li, Z.; Zhang, R.; Wu, Y.; Li, W.; Cao, Z. StCT2, a new antibacterial peptide characterized from the venom of the scorpion *Scorpiops tibetanus*. *Peptides* **2012**, *36*, 213–220. [[CrossRef](#)] [[PubMed](#)]
52. Ramírez-Carretero, S.; Jiménez-Vargas, J.M.; Rivas-Santiago, B.; Corzo, G.; Possani, L.D.; Becerril, B.; Ortiz, E. Peptides from the scorpion *Vaejovis punctatus* with broad antimicrobial activity. *Peptides* **2015**, *73*, 51–59. [[CrossRef](#)] [[PubMed](#)]
53. Miyashita, M.; Otsuki, J.; Hanai, Y.; Nakagawa, Y.; Miyagawa, H. Characterization of peptide components in the venom of the scorpion *Liocheles australasiae* (Hemiscorpiidae). *Toxicon* **2007**, *50*, 428–437. [[CrossRef](#)] [[PubMed](#)]
54. Zeng, X.C.; Nie, Y.; Luo, X.; Wu, S.; Shi, W.; Zhang, L.; Liu, Y.; Cao, H.; Yang, Y.; Zhou, J. Molecular and bioinformatical characterization of a novel superfamily of cysteine-rich peptides from arthropods. *Peptides* **2013**, *41*, 45–58. [[CrossRef](#)] [[PubMed](#)]
55. Silva, E.C.; Camargos, T.S.; Maranhao, A.Q.; Silva-Pereira, I.; Silva, L.P.; Possani, L.D.; Schwartz, E.F. Cloning and characterization of cDNA sequences encoding for new venom peptides of the Brazilian scorpion *Opisthacanthus cayaporum*. *Toxicon* **2009**, *54*, 252–261. [[CrossRef](#)] [[PubMed](#)]
56. Ma, Y.; Zhao, R.; He, Y.; Li, S.; Liu, J.; Wu, Y.; Cao, Z.; Li, W. Transcriptome analysis of the venom gland of the scorpion *Scorpiops jendeki*: Implications for the evolution of the scorpion venom arsenal. *BMC Genom.* **2009**, *10*, 290. [[CrossRef](#)] [[PubMed](#)]
57. Feng, L.; Gao, R.; Gopalakrishnakone, P. Isolation and characterization of a hyaluronidase from the venom of Chinese red scorpion *Buthus martensi*. *Comp. Biochem. Physiol.* **2008**, *148C*, 250–257. [[CrossRef](#)] [[PubMed](#)]
58. Horta, C.C.; Magalhaes, B.F.; Oliveira-Mendes, B.B.; do Carmo, A.O.; Duarte, C.G.; Felicori, L.F.; Machado-de-Avila, R.A.; Chavez-Olortegui, C.; Kalapothakis, E. Molecular, immunological, and biological characterization of *Tityus serrulatus* venom hyaluronidase: New insights into its role in envenomation. *PLoS Negl. Trop. Dis.* **2014**, *8*, e2693. [[CrossRef](#)] [[PubMed](#)]
59. Chen, Z.; Wang, B.; Yang, W.; Cao, Z.; Zhuo, R.; Li, W.; Wu, Y. SjAPI, the first functionally characterized Ascaris-Type protease inhibitor from animal venoms. *PLoS ONE* **2013**, *8*, e57529. [[CrossRef](#)] [[PubMed](#)]
60. Gronenborn, A.M.; Nilges, M.; Peanasky, R.J.; Clore, G.M. Sequential resonance assignment and secondary structure determination of the Ascaris trypsin inhibitor, a member of a novel class of proteinase inhibitors. *Biochemistry* **1990**, *29*, 183–189. [[CrossRef](#)] [[PubMed](#)]
61. Diego-Garcia, E.; Peigneur, S.; Clynen, E.; Marien, T.; Czech, L.; Schoofs, L. Molecular diversity of the telson and venom components from *Pandinus cavimanus* (Scorpionidae Latreille 1802): Transcriptome, venomics and function. *Proteomics* **2012**, *12*, 313–328. [[CrossRef](#)] [[PubMed](#)]
62. Chen, Z.Y.; Hu, Y.T.; Yang, W.S.; He, Y.W.; Feng, J.; Wang, B.; Zhao, R.M.; Ding, J.P.; Li, W.X.; Wu, Y.L. Hg1, novel peptide inhibitor specific for Kv1.3 channels from first scorpion kunitz.type potassium channel toxin family. *J. Biol. Chem.* **2012**, *287*, 13813–13821. [[CrossRef](#)] [[PubMed](#)]
63. Jiang, L.; Deng, M.; Duan, Z.; Tang, X.; Liang, S. Molecular cloning, bioinformatics analysis and functional characterization of HWTX-XI toxin superfamily from the spider *Ornithoctonus huwena*. *Peptides* **2014**, *59*, 9–18. [[CrossRef](#)] [[PubMed](#)]
64. 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](#)] [[PubMed](#)]
65. Morgenstern, D.; Rohde, B.H.; King, G.F.; Tal, T.; Sher, D.; Zlotkin, E. The tale of a resting gland: Transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. *Toxicon* **2011**, *57*, 695–703. [[CrossRef](#)] [[PubMed](#)]
66. Zhang, L.; Shi, W.; Zeng, X.C.; Ge, F.; Yang, M.; Nie, Y.; Bao, A.; Wu, S. Unique diversity of the venom peptides from the scorpion *Androctonus bicolor* revealed by transcriptomic and proteomic analysis. *J. Proteomics* **2015**, *128*, 231–250. [[CrossRef](#)] [[PubMed](#)]



67. Kotsyfakis, M.; Schwarz, A.; Erhart, J.; Ribeiro, J.M. Tissue- and time-dependent transcription in *Ixodes ricinus* salivary glands and midguts when blood feeding on the vertebrate host. *Sci. Rep.* **2015**, *5*, 9103. [[CrossRef](#)] [[PubMed](#)]
68. Saucedo, A.L.; Flóres-Solis, D.; Rodríguez de la Vega, R.C.; Ramírez-Cordero, B.; Hernández-López, R.; Cano-Sánchez, P.; Noriega-Navarro, R.; García-Valdés, J.; Coronas-Valderrama, F.; de Roodt, A.; et al. New tricks of an old pattern: Structural versatility of scorpion toxins with common cysteine spacing. *J. Biol. Chem.* **2012**, *287*, 12321–12330. [[CrossRef](#)] [[PubMed](#)]
69. Abdel-Mottaleb, Y.; Coronas, F.V.; de Roodt, A.R.; Possani, L.D.; Tytgat, J. A novel toxin from the venom of the scorpion *Tityus trivittatus*, is the first member of a new  $\alpha$ -KTX subfamily. *FEBS Lett.* **2006**, *582*, 592–596. [[CrossRef](#)] [[PubMed](#)]
70. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotech.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
71. 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](#)] [[PubMed](#)]
72. 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](#)]
73. Abascal, F.; Zardoya, R.; Posada, D. ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **2005**, *21*, 2104–2105. [[CrossRef](#)] [[PubMed](#)]
74. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. Protest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **2011**, *27*, 1164–1165. [[CrossRef](#)] [[PubMed](#)]
75. Drummond, A.J.; Suchard, M.A.; Xie, D.; Rambaut, A. Bayesian phylogenetics with beauti and the beast 1.7. *Mol. Biol. Evol.* **2012**, *29*, 1969–1973. [[CrossRef](#)] [[PubMed](#)]
76. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [[CrossRef](#)] [[PubMed](#)]



© 2016 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 (<http://creativecommons.org/licenses/by/4.0/>).