Identifying candidate genes involved in the regulation of growth using comparative transcriptome analysis (*Androctonus crassicauda*)

Fatemeh Salabi (✉ f.salabi@rvsri.ac.ir)

Razi Vaccine and Serum Research Institute

Hedieh Jafari

Razi Vaccine and Serum Research Institute

Research Article

Keywords: Scorpion, Venomous, Transcriptomic analyses, *Androctonus crassicauda*, Growth stages

Posted Date: February 14th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1285424/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Scorpions are venomous arthropod belonging to the arachnida, which considered as living fossils. Recently, transcriptomic analyses of scorpion venom gland were carried out by high-throughput sequencing and further analysis, but so far, the transcriptional analysis of venom gland of scorpion in different growth stages (adult and juvenile) have not been examined. Here, an Illumina-based sequencing study was designed to detect the candidate genes associated with growth development of scorpions of *A. crassicauda*.

Results

A total of 209951 mRNA transcripts were identified from total RNA-seq data, of which 963 transcripts were differentially expressed (DE) in adult and juvenile scorpions ($p < 0.01$). Overall, we identified *397 out of 558 up-regulated transcripts and 269 out of 405 down-regulated transcripts with sequence similarity to annotated genes from various animals*. GO and KEGG enrichment analyses indicated that the metabolic, thermogenesis, cytoskeleton, estrogen signaling, GnRH signaling, growth hormone signaling and melanogenesis pathways were affected by two different growth conditions and the results suggested that the DE genes related to those pathways are important biomarker genes associated with scorpions development and maturation, in which they may be important in future studies, including Chs, Elovl, MYH, RDX, ACTN, VCL, PIP5K, PP1C, FGFR, GNAS, EGFR, CREB, CoA, PLCB, CALM, CACNA, PKA and CAMK genes.

Conclusions

These findings broadened our knowledge about the differences between adult and juvenile scorpion's venom and opens new perspectives about the application of comparative transcriptome analysis to identify the special biomarkers.

Background

Scorpions are venomous predatory arthropod animals belonging to the arachnida, which have gained more attentions from researchers not only because of their venom component, but also because they are ancient terrestrial animals that are considered as living fossils (Dunlop, 2010; Quintero-Hernández et al., 2015). In previous, scorpion studies focused on the morphological and phylogenetical analysis, while for decades, most studies have focused on isolation and biochemical characterization of specific venom components.

Scorpion venom is complex mixture of biogenic amines, low molecular weight (MW) peptides and high MW proteins, mucoproteins and other components which produced by the columnar epithelial cells of venom gland (Al-Asmari et al., 2016). The number of unique natural compounds of mostly animal's

venom is highly variable, instance in scorpion's venom, it has been contained from 300 to over 1000 unique molecular entities (Quintero-Hernández et al., 2013) but it has not been fully proven yet. Up until now, the protein components of scorpion venom are divided into two categories: small peptides (<10 kDa) and large proteins (>10 kDa). Small peptides are the most studied scorpion venom components, which are mainly exhibited neurotoxins properties while the large protein derivatives of scorpion venom have shown enzymatic activities (Zhang and Zhang, 2016). Mainly due to the heterogeneity, complexity and high variability of venom composition, as well as low resolution of current venom protein separation methods, the amount of discovered venom peptides is much less than the expected amount (Vetter et al., 2011; Ahmadi, Shirin, et al. 2020; Ma et al., 2012). The emergence of advanced techniques such as high-throughput RNA-seq, promises the venom peptides sequence evolution and novel venom components prediction. RNA-Seq or transcriptome sequencing is one of the currently developed methods that utilized for novel gene identification, biology, medicine and molecular phylogenetic analysis especially in non-model organisms (Wen et al., 2015). Nowadays, implying of transcriptomic analysis has become essential in venomomics studies, as this analysis may represent promising lead for discovering of new venom components.

The first scorpion venom gland cDNA library which constitutes collections of cDNA sequences cloned into vectors was generated in 1989 for the scorpion *A. australis* (Bougis et al. 1989). Thereafter, transcriptomic analyses of scorpion venom gland have been greatly enhanced by the advent of high-throughput sequencing. In Iran, based on cladistic morphological analysis, the genus *Androctonus* can be phylogenetically divided into 5 species; *A. crassicauda*, *A. baluchicus*, *A. amurexi*, *A. robustus* and *A. finitimus* (Yagmur et al., 2016; Mirshamsi et al., 2011; Jalali and Rahim, 2014). The scorpion of *A. crassicauda* is one of the most venomous members of the Buthidae family which widely distributed in the Middle East, including Iran, Iraq, Turkey, Jordan and Arabia (Crucitti, 1999). Up to date, lack of transcriptomic information is evident in *A. crassicauda* and the available transcriptome data published previously on this species mostly concentrate on a few species in this genus such as *A. bicolor* (Zhang et al., 2015) and *A. mauretanicus* (Oukkache et al., 200). Furthermore, unfortunately, the comparative transcriptomic analysis of different growth stages of the scorpion has been rarely investigated over the years. Therefore, here we have reported the transcriptome analyses of the venom gland of *A. crassicauda* scorpions in two different growth stages: juvenile and adult, with insights on the identification of the biomarkers for scorpion development. For this purpose, we investigated the expression profiling of deregulated mRNAs in adult and juvenile *A. crassicauda* scorpions. Finally, the potential functions of *differentially deregulated mRNAs were annotated by analysis* of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG).

Methods

Sample collection and cDNA library construction

A. crassicauda scorpions were collected from the deserts of Baghmalek on south west of Khuzestan province, Iran. After classification of specimens according to morphological characteristics (Koch,

1977), the scorpions were grouped based on the gender and age-class as suggested by Kaltsas et al. (2008). Thereafter, the telsons of adult and juvenile *A. crassicauda* were removed 72 h after being milked by electrical stimulation. For total RNA extraction, the telsons were powdered in liquid nitrogen and total RNA was extracted with RNeasy Animal Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. Then, RNA Integrity Number (RIN) was evaluated for all samples. Only RNA samples with a RIN > 7 were used for cDNA library construction. All cDNA libraries were sequenced with an Illumina HiSeq 2000 platform (Illumina, San Diego, CA, USA) at Macrogen Co (Macrogen, Seoul, Korea) and 150 bp paired-end reads were produced.

Assembly of read-sequences and bioinformatics analysis

After cleaning, trimming and assessing the quality of raw reads by means of the Trimmomatic and FastQC program (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>), clean reads were de novo assembled into contigs using Trinity software, version 2.0.3 (Grabherr et al., 2011, Haas et al., 2013) [with optimized parameters](#). Afterwards, in order to identify transcripts harboring any protein domains *in the* De novo assembled transcripts/contigs, we used annotation pipeline with *five search strategies*; Trinotate (<http://trinotate.sourceforge.net>), Blastp (<http://blast.ncbi.nlm.nih.gov/Blast>), Blastx (<http://blast.ncbi.nlm.nih.gov/Blast>), HMMER/pfam (<http://hmmer.janelia.org/software>) and *species-specific annotation filter*. The main step consists of the *identification of all* putative homologous sequences with sequence similarity searching against Swissprot, NCBI non-redundant (Nr), UniProtKB/TrEMBL and Pfam protein domain databases with an *E-value* threshold of 10^{-3} . *Furthermore, we performed species-specific annotation, because we were not expecting to identify many known homologs of proteins in venomous animals and animals with close phylogenetic affinity with scorpion in the public databases. Therefore, the species-specific annotations were done using following procedure.* We first created databases using the scorpion, tick and spider specific protein sequences (<https://www.uniprot.org/>) and all manually reviewed venom proteins and toxins from the venomous animals (<https://www.uniprot.org/program/Toxins>). Then the scorpion datasets were annotated manually with these sequences using blastx. Ultimately, all of the annotated and identified coding transcripts were merged and known as mRNA transcripts.

Quantification of gene expression levels of mRNAs

RSEM software (Li and Dewey, 2011) was used to estimate the gene expression levels in terms of FPKM (fragments per kilobase of transcript sequence per million base pairs sequenced). RSEM is an abundance estimation method for quantifying gene and isoform expression levels in species with or without sequenced genomes. In this study, the sequence length and expected FPKM values were used to characterization of mRNAs in two growth stages of the *A. crassicauda* (mature and juvenile). Gene expression levels in terms of FPKM were quantified using RSEM software.

Differential expression analysis of mRNA

Differential expression analysis of mRNAs transcripts was determined using the EdgeR package (Robinson et al., 2010). For biological replicates, transcripts or genes with a $\log_2$ (fold change) > 1 or $\log_2$ (fold change) < -1 and $q\text{-value} < 0.01$ were defined as mRNAs transcripts that were differentially expressed (DE) between the two groups, as determined by R Package-EdgeR.

Identification of known and novel DE venom gland proteins

The DE mRNA transcripts were searched for sequence similarity with created Uniprot protein-toxin databases from venomous animal. *Blast* was used to search against the venomous animal's protein databases ($E\text{-value } 1e-3$). Transcripts that exhibit significant hit were considered as venom proteins and they were broadly grouped in different categories according to their function.

Gene ontology and pathway enrichment analysis

Differentially expressed coding genes in both of growth stages were enriched for Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways to investigate the main functions of DE mRNA. *Further, GO and KEGG analysis also were conducted to interpret the possible functions by forming hierarchical categories based on the biological process, cellular component, and molecular function terms. False discovery rate (FDR) < 0.01 was considered as significantly enriched.*

Results

High-throughput sequencing and identification of candidate mRNAs

To investigate the changes in gene expression among different developmental stages of *A. crassicauda*, an experimental workflow was implemented (Figure 1). Briefly, paired-end sequencing of cDNA libraries was constructed. Approximately average of 200,537,746 raw reads were gained for the adult and juvenile scorpions. After filtering the raw reads and removing low quality, adapter and uncertain reads, a total of 200 million clean reads, approximately 54 Gb of high-quality data, were generated. The clean reads were assembled into 952,725 transcripts after removal of sequences shorter than 200 bp as uninformative using Trinity tool with the default parameters. In order to identify the known and novel protein coding transcripts of *A. crassicauda* transcriptome, the long open reading frames (ORFs) in the assembled transcripts were first predicted, followed by similarity searches to all known organisms in Swiss-Prot; NCBI nr; UniProtKB/TrEMBL, Pfam, UniProt proteins/toxins databases and species-specific annotation with blast. After removing *non-coding* RNAs from the assembled transcriptomes, we annotated 148992 coding sequences on the basis of blast hits to the Swiss-Prot; NCBI nr;

UniProtKB/TrEMBL, Pfam database, 38 coding sequences using TransDecoder predictions, *a total of 58045 putative transcripts with sequence similarity to annotated genes from arachnids, 2661 of which corresponded specifically to scorpions*. Using blast hits to the UniProt proteins/toxins database, we annotated an additional 215 venom proteins. All annotated sequences were merged and considered as mRNAs. Finally, out of 952,725 transcripts, 209951 were classified as mRNA transcripts. Comparison of the length of mRNA sequences of the two growth stage groups is illustrated in figure 2A. Sequence analysis suggested that mRNA candidates of both groups exhibited an equal sequence length (mean length of 838.77 bp). Furthermore, this study provided the first comprehensive analysis of the mRNA expression profiles in juvenile and adult scorpion venom glands (Figure 2B). As shown in the figure 2B, the gene expression levels differed significantly between candidate mRNAs of adult and juvenile groups. The expression of mRNAs of adult scorpions is lower compared with mRNAs of juvenile groups.

Differential expression analysis of mRNAs

To study the differential expression of mRNAs in the two developmental stages of *A. crassicauda*, *the expression levels of mRNA transcripts were measured and differential expression analysis was conducted using EdgeR. 963 mRNA transcripts were identified with significant expression level differences ($p < 0.01$), including 558 up-regulated and 405 down-regulated mRNAs (Figures 3A). Bioinformatics analysis reveals different gene expression patterns in the adult and juvenile scorpions (Figures 3B,C).*

Annotation of dysregulated mRNAs

All differentially dysregulated mRNA sequences from each groups were initially annotated on the basis of blast hits to the UniProt proteins/toxins and Swiss-Prot databases. By removing the redundancy occurring in the isoforms of the same genes, a total of 397 out of 558 up-regulated transcripts and 269 out of 405 down-regulated transcripts were identified with sequence similarity to annotated genes from various animals. Next, all identified proteins from adult and juvenile scorpions were classified according to the functionality of each of these, into following venom components; Enzymes, Toxin, Binding protein, Skeletal muscle protein, Venom peptide, Potassium channel toxin, Calcium channel toxin, Sodium channel toxin, Zinc finger protein, Insect toxin, Antimicrobial peptide, Mammal toxin and Other protein. The number of transcripts classified for each class is summarized in figures 4A,B. We further described of those venom components below.

Enzymes

Among up-regulated mRNAs, 106 recognized protein sequences were belonged to enzyme classification. Of these, the most important enzymes are proteases (metalloproteinases (8) and serine proteases (5)), synthases, nucleotidases, oxidases and aldehyde dehydrogenase (Figure 4A). Here, we further found

approximately 26% (70) of the differentially down-regulated transcripts codifying for enzymes. Of these, the most important enzymes were various venom proteinases (metalloproteinases (2) and serine proteases (12)), transferases (10), kinases (9), polymerase, fatty acid synthases, fatty acid elongase and ligases (Figure 4).

Skeletal muscle proteins

Skeletal muscle proteins are another class of dysregulated proteins. 36 skeletal proteins were up-regulated in adult group which the most important of these are Vinculin, Actin, Myosin, Tensin, Dystrophin, Kinesin, Chitin synthases, Chitinase cytoskeletal, Papilin, transmembrane proteins and other cuticular proteins. Moreover, cytoskeletal proteins of adult-specific rigid cuticular protein, cuticlin-1, cuticle protein 14 and cuticle protein 16.8 were also up-regulated in adult scorpions.

In juvenile scorpions, 34 skeletal muscle proteins were up-regulated. Most of these were including structural proteins of the cytoskeleton that maintained the tissue integrity and regulatory of microtubule dynamics. Of these, Plectin, Striatin, Ankyrin 2 and 3, Rootletin, Actin, Myosin, Titin, Talin-2 and Filamin-A cytoskeletal proteins can be mentioned.

Chanel Toxins

As shown in figure 4, a total of 12 channel toxins were up-regulated in adult scorpions, including one sodium channel toxin, five calcium channel toxins and six potassium channel toxins, will in juvenile scorpions, a total of 5 channel toxins were up-regulated that including one calcium channel toxin and four potassium channel toxins.

Toxins

Of dysregulated mRNAs, a total of 3 antimicrobial peptides, 1 mammal toxin, 5 insect toxins and 22 other toxin genes, such as neurotoxins, toxin Acra (I-2, I-1, II-1, II-2 and III-1), U-scoloptoxins and Beta-toxin BmK were significantly up-regulated in adult scorpions. *In juvenile scorpions, a total of 1 insect toxin and 9 other toxin genes especially neurotoxins were significantly up-regulated (Figure 4).*

Other

We identified number of hemolymph proteins which become very highly expressed in adult scorpions. These proteins included; Hemocyanin, Hemocytin, Techylectin-5A and Hephaestin-like proteins. Among other interesting proteins that up-regulated in adult scorpions, Papilin, Sex determination protein tasselseed-2 and Dual oxidase maturation factor 1 can either be highlighted (Figure 4). *Other interesting*

proteins that overexpressed in the juvenile scorpions are Vitellogenin, Transforming growth factor-beta-induced protein ig-h3 and Endoplasmic reticulum lectin 1 (Figure 4).

GO terms and KEGG pathways enrichment analysis of DE mRNAs

GO and KEGG pathway enrichment analyses were carried out to explore the potential function of the DE mRNAs. For the GO classification, the identified DE mRNAs were divided into three ontologies; molecular function, biological process and cellular component. The GO analysis of differentially up-regulated mRNA transcripts ($p < 0.01$), revealed that “binding”, “catalytic activity” and “protein binding” in molecular function; “membrane”, “cell” and “cell part” in cellular component and “cellular process”, “organic substance metabolic process” and “cellular metabolic process”, in biological process were highly represented (Figure 5A, Supplementary file 1A). Correspondingly, our GO analysis result indicated that the most enriched GO terms of down-regulated genes in biological processes were mainly involved in “cellular process”, “biological regulation”, “regulation of biological processes”, “regulation of cellular processes”, and “response to stimulus”. In terms of molecular function, the down-regulated mRNAs were enriched in “binding” and “cellular process”, and in cellular component they were enriched in “cell” and “cell part” (Figure 5B, Supplementary file 1B).

KEGG analysis of the predicted DE mRNAs was used to enhance our knowledge of the biological functions of the DE mRNAs during development of *A. Crassicauda*. KEGG analyses of differentially expressed genes between the adult and juvenile groups are shown in Figure 6. According to KEGG analysis of differentially up-regulated mRNA transcripts, the most importantly enriched pathways were Metabolic pathways (44), Biosynthesis of secondary metabolites (14), Pathways in cancer (10), Regulation of actin cytoskeleton (10), Thermogenesis (10), Focal adhesion (10), Adherens junction (9), Carbon metabolism (8), Rap1 signaling pathway (8), Axon regeneration (7), Oxytocin signaling pathway (7), Endocytosis (7), PI3K-Akt signaling pathway (7), Estrogen signaling pathway (6), Calcium signaling pathway (6), GnRH signaling pathway (5), Gastric acid secretion (5), MAPK signaling pathway (5), Growth hormone synthesis, secretion and action (4), Melanogenesis (3), Fatty acid degradation (2), Regulation of lipolysis in adipocytes (1) and Insect hormone biosynthesis (1). In KEGG analyses of the down-regulated mRNAs, the most enriched pathways were Metabolic pathways (13), Pathways in cancer (5), Thermogenesis (5), Tight junction (5), Adherens junction (5), Axon regeneration (4), Focal adhesion (4), Oxytocin signaling pathway (4), MAPK signaling pathway (4), Gastric acid secretion (3), GnRH signaling pathway (3), PI3K-Akt signaling pathway (3), Calcium signaling pathway (3), Fatty acid metabolism (3), Biosynthesis of secondary metabolites (3), Melanogenesis (2), Growth hormone synthesis, secretion and action (2), Regulation of actin cytoskeleton (2), Apoptosis (2), Biosynthesis of unsaturated fatty acids (2), Fatty acid elongation (2) and Estrogen signaling pathway (1), respectively (Figure 6). Regulatory pathways of actin cytoskeleton, estrogen signaling, GnRH signaling, melanogenesis, and growth hormone synthesis, secretion and action are illustrated in supplementary files 2, 3, 4, 5 and 6 respectively.

According to the results of GO and KEGG analysis, chitin synthase (Chs), aminopeptidase and phosphodiesterase proteins were up-regulated proteins, and fatty acid elongase (Elovl) and *fatty acid synthase* proteins were the down-regulated proteins, which are the main enriched proteins in the metabolic pathway. Moreover, from up-regulated genes, the MYH, RDX, ACTN, VCL, PIP5K and PP1C, and from down-regulated genes the FGFR (Fibroblast growth factor receptor) were main enriched proteins that involved in the scorpion cytoskeleton development.

In scorpion melanogenesis pathway, the up-regulated genes of GNAS, EGFR, CALM and PLCB, and the down-regulated genes of PKA and CAMK were enriched. Furthermore, the up-regulated genes of GNAS, CREB and PLCB, and the down-regulated genes of CACNA and PKA were enriched in scorpion growth hormone signaling pathway. In scorpion estrogen signaling pathway, the up-regulated genes of GNAS, EGFR, CREB, CoA, PLCB and CALM, and the down-regulated gene of PKA were enriched. In scorpion GnRH signaling pathway, the up-regulated genes of GNAS, EGFR, CREB, PLCB and CALM, and the down-regulated gene of CACNA were the main enriched proteins. Taken together, the adult had severe genes upregulation than juvenile scorpions. In total, our results showed that the genes of Chs, MYH, RDX, ACTN, VCL, PIP5K, PP1C, GNAS, EGFR, CREB, CoA, PLCB and CALM were up-regulated, and Elovl, CACNA, PKA, CAMK and FGFR were down-regulated in adult scorpions (Supplementary files 7, 8).

Discussion

So far, a comparative study of transcriptional profiles of adult and juvenile scorpion's venom gland has not been performed. Here a comparative transcriptome analysis of the venom gland of *A. crassicauda* scorpions in two different growth stages: juvenile and adult, was done to identify the biomarkers for scorpion developments. For this purpose, the scorpion venom gland gene expression level was estimated using the FPKM value (Li and Dewey, 2011) and the differentially expressed genes during development of *A. crassicauda*, was identified based on the experimental workflow shown in Figure 1. Using this workflow, we found 963 (558 up-regulated and 405 down-regulated mRNAs) out of 209951 mRNA transcripts that were differentially expressed. To determine the function of dysregulated genes, the *up and down-regulated mRNAs* were classified into venom components *following the scheme presented in recent venom gland transcriptome analyses* (Ward et al., 2018; Rokyta and Ward, 2017; De Oliveira et al., 2015). *We have further described this classification of venom components below.*

Enzymes

We detected 106 out of 558 up-regulated mRNAs that showed enzyme activity, including the proteases, synthases, nucleotidases, oxidases and aldehyde dehydrogenase. Scorpion venoms are well known as rich sources of different proteases (Cid-Urbe et al., 2018). Among the up-regulated proteases are metalloproteinases and serine proteases, which are important venom proteins with protein digestion, catalytic, post-translational processing of toxins activity and promote the spreading of toxins via degradation of matrix proteins activity that already described in scorpion venoms (Cid-Urbe et al., 2018;

Valdez-Velázquez et al., 2013), snake venoms (Bottrall et al., 2010) and spider venoms (Paixão-Cavalcante et al., 2006). Metalloproteinase is one of the common proteolytic enzymes of the scorpion venoms. Metalloproteinase enzymes have been reported to play critical roles in envenomation-related pathogenesis by affecting the hemostatic system, including bleeding, intravascular clotting, edema, inflammation, myonecrosis, skin damage and inflammatory reaction as an important characteristic of snake bites envenomation (Escalante et al., 2009; Mohamed Abd El-Aziz et al., 2019). Additional activities of these enzymes are in cell proliferation, differentiation and remodeling the extracellular matrix (de Oliveira et al., 2015). de Oliveira et al (2015) have been suggested that metalloproteinase plays an important role in the pathogenesis of the local tissue damage during snake envenomation and may be associated with hemorrhage by disrupting the microvessel walls. Such observations may suggest that one of the causes of envenomation-related pathogenesis damage may be related to overexpression of these enzymes. It is noteworthy that these enzymes are highly overexpressed in adult scorpions, which may indicate higher toxicity of these scorpions. On the other hand, Aldehyde dehydrogenase (ALDH) is one of the up-regulated enzymes which involved in important pathways such as Pyruvate metabolism, Fatty acid degradation and Insect hormone biosynthesis. Aldehyde dehydrogenase is a biosynthetic enzyme of the Juvenile hormones (JH) pathway that plays a critical role in insect morphogenesis, reproduction and behavior (Rivera-Perez et al., 2013). In general, the number of enzymes overexpressed in adult scorpions (106) is higher than in juvenile ones (69). This may suggested the higher enzymatic activity in adult scorpions.

Down-regulated proteins were also classified according to up-regulated proteins classification. Of differentially down-regulated mRNA, 70 out of 269 annotated transcripts found to be enzymes. Interestingly, the obtained results of differentially expressed enzymes showed that the major up-regulated synthase enzymes of adult scorpions catalyzed the synthesis of amino acid, phosphate and citrate compounds while the major synthase enzymes which overexpressed in juvenile scorpions include: fatty acid synthases and fatty acid elongase. It has been reviewed that the expression of fatty acid elongation proteins led to increases of very long chain fatty acid synthesis which widespread among insect taxa. Very long chain fatty acids linked with other compounds to produce biologically active products that are involved in the biosynthesis of cuticular waxes and they have an important role as anti-inflammatory components (Stanley-Samuelson and Dadd, 1983; Nakamura et al., 2014; Bannenberg and Serhan, 2010; Huai et al., 2015). Fatty acid elongation protein and elongation of very long chain fatty acids protein are essential in the biosynthesis of fatty acids longer than C14. These enzymes involved in the pathway fatty acid biosynthesis, which is part of lipid metabolism and may be required for normally rapid growth (Zheng et al., 2017). Our results are consistent with findings in mealworm that determined the expression of two different elongase throughout all developmental stages (Zheng et al., 2017). Of other down-regulated enzymes, Transferase enzymes can be mentioned that generally known to catalyze the transfer of L-fucose sugar, glutaminyl-peptide, acetyl, methyl, galactosamine from a donor substrate to an acceptor compound.

Skeletal muscle proteins

Furthermore, 36 out of 558 up-regulated mRNAs that encode many skeletal muscle proteins were recognized. Of these proteins, the most important are Vinculin, Actin, Myosin, Tensin, Dystrophin, Kinesin, Chitin synthases, Chitinase cytoskeletal, Papilin, transmembrane proteins, adult-specific rigid cuticular protein, cuticlin-1, cuticle protein 14 and cuticle protein 16.8. Insect cuticles form an exoskeleton that has a rigid structure due to the presence of chitin and sclerotized proteins. Chitin is a vital polysaccharide of the cuticular exoskeletons and the peritrophic matrix of arthropods that synthesized by a complex family of Chitin synthases and degraded via Chitinases. Chitin considered as primarily structural scaffold proteins supporting the cuticles of the epidermis and trachea as well as the gut epithelium peritrophic matrices. Insect development and morphology forces them to periodically replace their old cuticles with new ones during molting. Therefore, in order to coordinate synthesis and degradation of the cuticle, insects consistently require strict control of the participating component during molting such as Chitin synthases, Chitinase, Actin, Myosin and Kinesin. To achieve this goal, insects regularly produce Chitin synthases, Chitinase enzymes, Actin, Myosin and Kinesin during development (Merzendorfer and Zimoch., 2003; Schuster et al., 2012; Reynolds and Samuels, 1996). Furthermore, Vinculin and Kinesin have been reported to up-regulate upon differentiation and they may be involved in the differentiation of adult muscle cells (Gomes and Cadot, 2017; Han et al., 2017; Ginkel and Wordeman., 2000). Chitinase also investigated to has neurotoxic effects include an important role in the pathogenesis of inflammatory and allergic disease conditions (Matute-Blanch et al., 2020; Mack et al., 2015).

Dystrophin, a member of the β -spectrin/ α -actinin protein family, is a cytoskeletal protein that located at the muscle sarcolemma (Koenig et al., 1988). Dystrophin is part of a protein complex that works together to maintain the muscle fibers by connects the cytoskeleton to the basal lamina and protect them from injury (Blake et al., 2002). Based on this study results, adult skeletal muscle consists of the highly cytoskeletal proteins that strengthen the muscle fibers and regulate muscle cell activity. Accordingly, Ginkel and Wordeman (2000) reported the presence of Kinesin in differentiating and adult skeletal muscles.

In juvenile scorpions, 34 skeletal muscle proteins out of 269 annotated transcripts were up-regulated. The most important ones are Plectin, Striatin, Ankyrin 2 and 3, Rootletin, Actin, Myosin, Titin, Talin-2 and Filamin-A. To date, scorpion skeletal muscle protein's function has not been explored. In eukaryotic cells, Ankyrins are intracellular adaptor proteins critical for the expression and targeting of cytoskeletal elements that serve as the interface between the plasma membrane and cytoskeleton (El Refaey and Mohler, 2017). Previous studies have suggested an important role for Ankyrin protein during embryonic development (Daniel et al., 2019). Similar to our results, Daniel et al (2019) have been showed that both forms of Ankyrin were ubiquitously expressed early in development stages.

Plectin, a large size (>500 kDa) cytolinker protein, is abundantly expressed in a wide range of cells and tissues, most prominently in muscle, brain and stratified squamous epithelia (Wiche et al., 1983). It acts as a multi-functional cytoskeletal intermediate filament (IF)-associated protein and signaling scaffold (Castanón et al., 2013). There is little information about the expression of this protein in different stages

of life, especially in insects, and comparing its expression levels of different life stages is required further study.

Striatin, a multivalent scaffold protein, was found to act as regulator of calcium homeostasis (Abu-Zaid et al., 2012). In rat, the highest expression of *Striatin* occurs at neonatal stages and the lowest in adults (Nader et al., 2017) as we also observed for *A. Crassicauda*.

Rootletin has been shown to be one of linker protein that is the only structural constituent of the ciliary rootlet (Yang et al., 2002). Ciliary rootlet is a prominent fibrous cytoskeletal structure that originating from basal bodies of ciliated cells and links the base of the cilium to the cell body (Chen et al., 2015; Fawcett and Porter, 1954). Previous studies have been reported the presence of sensory organs (sensory hairs) in various parts of the scorpion cuticle, including the pedipalps, legs, pecten and venom gland (Barth and Wadepuhl, 1975; Ahmed, 2015; Knowlton, and Gaffin, 2011), but no study has been done on their structural proteins in scorpions so far. The function of *Rootletin* has been already investigated in the fruitfly. The previous studies have been showed that *Rootletin* is essential for mechanosensory function and ciliary rootlet formation in chordotonal sensory neurons and it could organize the ciliary rootlet to achieve neuron sensory function that affecting various behaviors related to mechanosensation and chemosensation in *Drosophila* (Chen et al., 2015; Styczynska-Soczka and Jarman, 2015). Considering the various vital physiological roles of cilia such as; chemoreception, mechanoreception, hygromoreception and thermoreception in arthropods (Keil, 2012); motility, sensory and signalling roles in insects (Pratt et al., 2016); chemosensory and mechanosensory in scorpions (Barth and Wadepuhl, 1975; Foelix and Schabronath, 1983a,b) the study of the expression of this protein seems to be necessary. Our knowledge about function and expression pattern of this protein as well as the roles of sensory cilia in the scorpions is limited and needs more attentions. Although interestingly, the results of this study showed that this protein is expressed in two investigated stages of scorpion life cycle and the obtained results indicated an increase in its expression in juvenile scorpions compared to adult ones, however, the expression pattern of *Rootletin* in scorpions requires further studies.

Titin is a large scaffold protein in striated muscle physiology (Tonino et al., 2017). In previous literature, *Titin* is also called mini-titin, titin/twitchin, titinlike, connectin, etc (Hooper et al., 2005). Due to springlike nature, *Titin* is suggested to fulfill multiple functions. It causes passive elasticity of muscle, maintains the thick filaments in the center of the sarcomere, provides much of muscle passive resistance to stretch and elasticity and determines myofilament distensibility. In addition, it regulates the activity of other sarcomere proteins and preserves myosin molecules in place. Due to its elasticity, this protein may help determine the length of the sarcomere during development (Powers et al 2014; Hooper et al., 2005), which may be the reason for the upregulation of these proteins in juvenile scorpions.

Talins are large, cytoskeletal proteins found in animals that link the actin cytoskeleton to the plasma membrane in several different multicomponent adhesion complexes (Singiser and McCann, 2006; Conti et al., 2009). *Talin-2* is crucial for myoblast fusion, multicellular differentiation and morphogenesis during

skeletal muscle development (Tsujioka et al., 1999; Tsujioka et al., 2004; Singiser and McCann, 2006). We found that Talin-2 is highly expressed in juvenile scorpions.

Dystroglycan is a laminin binding glycoprotein that serves as a transmembrane linker between intracellular cytoskeleton and extracellular matrix protein (Durbeej et al., 1998; Hara and Campbell, 2014). Recently, it was suggested that Dystroglycan is expressed in developing epithelial cells and may be important for the maintenance of tissue integrity (Ekblom, 1996; Durbeej et al., 1998). Dystroglycan is also required for the formation of a basement membrane at early stages of development (Henry and Campbell, 1998). In this study we found that Dystroglycan is over expressed in juvenile scorpions.

Chanel Toxins

Totally, 12 and 5 channel toxins were up-regulated in adult and juvenile scorpions respectively. Interestingly, as shown in fig 4, the majority of the differentially expressed channel toxins codifying for potassium channel toxins. Potassium channels, a family of neurotoxins, subiquitously exist in excitable and nonexcitable cells. Potassium channels can be divided into four major classes based on their functional and structure: calcium activated, voltage-gated, inward rectifier, and two-pore K⁺ channels (Martin-Eauclaire and Bougis, 2012). Most of the putative potassium channel toxins described here have similarities to voltage-gated potassium channel-blocking group and may involve in scorpion envenomation. adult scorpions express the higher level of potassium channel toxins. Therefore, causes of more symptoms following envenomation by adult scorpion venom may be due to increased expression of potassium channel toxins.

Toxins

By comparing the overexpressed toxin proteins, we found that the number of up-regulated toxins in adult groups is more abundant than juvenile scorpions. As shown in figure 4, and Supplementary file 7 and 8, several insects, mammal, antimicrobial specific venom toxins and channel toxins have been overexpressed in adult scorpion venom. However, no mammal or antimicrobial specific venom toxins were found in juvenile scorpions. These results may suggest that the venom of adult scorpions is more toxic to insects and mammals than juvenile ones.

Other

Of up-regulated mRNAs of adult scorpions, we have identified hemolymph proteins including; Hemocyanin, Hemocytin, Techylectin-5A and Hephaestin-like proteins. Techylectin-5A is one of the most overexpressed proteins in adult scorpions. Recently studies suggested that Tachylectins plays an important role as pattern recognition molecules in the innate immune response system and blood coagulation ((Gokudan et al., 1999). Homocyanins, a large copper-binding metalloproteins in the

hemolymph of invertebrates, acts as a respiratory pigment in the transport of oxygen and responsible for the blue coloration of arthropods hemolymph (Hosseinzadeh et al., 2019; Jaenicke et al., 2012). The present studies showed that, even though the levels of Hemocyanin mRNA were found in all samples but the highest level of Hemocyanin mRNA expression occurred in the adult scorpions. Contrary to the results obtained in this study, it has been reported that despite a moderate amount of Hemocyanin was found in all developmental stages of the *grasshopper species Locusta migratoria manilensis* life cycle but the levels of *Hemocyanin* mRNA in the embryo were significantly higher than in adults (Yin et al., 2012).

Recently have been reported that the sex and body size do not seem to influence the copper content of the blood of scorpions (Padmanabhanaidu, 1966). Notably, the expression of Hemocyanin mRNA was affected by environmental or nutritional changes including the level of dietary copper and the absorption of the scorpion oxyhaemocyanin in the u.v. region or in the visible range of the spectrum (Padmanabhanaidu, 1966). Although, the preliminary studies on the physiological properties and structural of the Haemocyanin in scorpions has been confirmed the respiratory and also buffering function of Haemocyanin (Padmanabhanaidu, 1966; Jaenicke et al., 2012), none of them has been compared the expression of Homocyanin in adult and juvenile scorpions and this issue needs further investigation.

Papilin, an extracellular matrix glycoprotein, is other up-regulated protein of adult scorpions. It suggested to be influences cell rearrangements and may modulate metalloproteinases during organogenesis and development (Kramerova et al., 2000). With this explanation, simultaneous overexpression of Papilin with Homocyanin proteins may confirm overexpression of each other.

Other up-regulated mRNAs in juvenile scorpions are Vitellogenin, Transforming growth factor-beta-induced protein ig-h3 and Endoplasmic reticulum lectin 1. Vitellogenin (vtg) is a phospholipoglycoprotein involved in egg yolk formation and produced by most oviparous organisms. In the majority of insects, Vtg is synthesized extraovarially in the fat body (Valle, 1993; Roy-Zokan et al., 2015; Rosell and Coons, 1992; Hagedorn et al., 1973).

Previously, it was hypothesized that Vitellogenin is only expressed in females of social insects, but evidence has recently been obtained that Vtg is also expressed in males as well as in subsocial insects such as beetles (Roy-Zokan et al., 2015). In addition to its multiple social roles such as regulation of social behavior, Vtg has been shown to be involved in nutritional functions, antioxidant activities, immune defense, oocyte development, reproduction and longevity regulation (Sun and Zhang, 2015; Naderi et al., 2005; Nelson et al., 2007). Furthermore, Vtg production increases in male fish when exposed to exogenous estrogen, in mosquito when blood sucking (Hagedorn et al., 1973), in frog, yellowfin seabream and rooster under estrogen stimulation (Bergink et al., 1974; Wangh and Knowland, 1975; Naderi et al., 2005), in hymenoptera male and female during parental care (Roy-Zokan et al., 2015) and in honeybees during foraging behavior and primes bees for specialized foraging tasks (Nelson et al., 2007).

Evidence shows that estrogen in male plays a critical role in sexual function and it is essential for spermatogenesis. Moreover, the dynamic process of spermatogenesis in juvenile males is regulated by

estrogen (Naderi et al., 2005). Most importantly, the presence of Vtg in scorpions has been reported in this investigation, thus this study reports both the expression of this protein and evaluated the difference in its expression among adult and juvenile scorpions. The biological significance of DE analysis has been shown that the expression of this protein is significantly increased in juvenile scorpions. Increases in Vtg expression in juvenile males, may be due to exogenous estrogen stimulation (Naderi et al., 2005). In this study, we evaluated the some of the most important differentially expressed genes between adult and juvenile scorpions, to obtain as much information as possible, refer to Supplementary file 3 and 4.

Gene Ontology of DE mRNAs

Moreover, this study investigated dysregulated genes based on GO terms and KEGG pathways. Analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were used to enhance our knowledge of the biological functions of the DE mRNAs during development of *A. crassicauda*. Interestingly, the GO and KEGG analyses revealed that many dysregulated genes were related to metabolic, thermogenesis, cytoskeleton, estrogen signaling, GnRH signaling, growth hormone signaling and melanogenesis pathways. Therefore, in this study, altering the expression pattern of genes involved in these pathways suggested to be key developmental indicators in scorpion. Of these, some important genes *including*, Chs, Elovl, MYH, RDX, ACTN, VCL, PIP5K, PP1C, FGFR, GNAS, EGFR, CREB, CoA, PLCB, CALM, CACNA, PKA and CAMK genes were suggested key regulation factors of development in scorpion. These GO terms and KEGG pathways can provide new insight for the study of scorpion development. Several other studies were also undertaken to elucidate the role of this important mediators in reproductive processes of male. Leclerc et al. (2020) demonstrate that the Calmodulin (CALM) is a small conserved protein which known to be expressed *abundantly in the testis especially in the spermatozoa. This acidic protein mediates numerous intracellular Ca^{2+} dependent events include; sperm motility and fertilizing ability. In the present study, upregulation of calmodulin in adult scorpions further emphasize the importance of this Ca^{2+} mediator in sperm function.*

On the other hand, protein kinase A (PKA) and calcium/calmodulin-dependent protein kinase (CaMK) also have been investigated to regulate cellular growth, differentiation, sperm maturation and function. Furthermore, they are known to play a major role in spermatogenesis. Both protein kinase A (PKA) and CaMKIV can phosphorylate CREM τ (cAMP-responsive element modulator- τ), which an essential transcription factor for spermatogenesis (Jenardhanan and Mathur, 2014; Wu et al., 2001). The results of present study can be further validated by a report by Ma et al (2015), who demonstrated that PP1C (serine/threonine-protein phosphatase pp1 catalytic subunit) involved in spermatogenesis and by Hasegawa et al (2012) who indicated that PIP5K1A (phosphatidylinositol-4 phosphate 5-Kinase) may play a role in the maintenance of sperm number and morphology during spermatogenesis.

Conclusion

The lack of information about juvenile scorpions venom, as well as the differences between their venom components with adult scorpions, highlights the need for such studies. In the present study, a rich source of known and novel venom proteins was identified using the RNA-seq technique and using various databases for the isolation and identification of protein coding RNAs. Comparison of venom components of juvenile and adult scorpions using gene expression analyses identified DE genes and pathways significantly associated with scorpion growth. By analyzing DE genes, we suggested the Chs, Elovl, MYH, RDX, ACTN, VCL, PIP5K, PP1C, FGFR, GNAS, EGFR, CREB, CoA, PLCB, CALM, CACNA, PKA and CAMK genes as key regulators of scorpion growth. Taking our results together, these findings broadened our knowledge about the differences between adult and juvenile scorpion's venom. Furthermore, this study provides new insights into the application of comparative transcriptome analysis to identify the special biomarkers.

Abbreviations

DE: Differentially expressed

GNAS: Guanine nucleotide-binding protein

PLCB: Phosphatidylinositol phospholipase C, beta

CALM: Calmodulin

CREB: Cyclic AMP-dependent transcription factor ATF-4

EGFR: Epidermal growth factor receptor

CoA: Nuclear receptor coactivator 1

PKA: Protein kinase A

CAMK: Calcium/calmodulin-dependent protein kinase (CaM kinase) II

CACNA: Voltage-dependent calcium channel L type alpha

RAC1: Ras-related c3 botulinum toxin substrate

PIP5K: Phosphatidylinositol-4 phosphate 5-Kinase

PP1C: Serine/threonine-protein phosphatase pp1 catalytic subunit

Chs: Chitin synthase

Elovl: Fatty acid elongase

FGFR: Fibroblast growth factor receptor

Vtg: Vitellogenin

GO: Gene ontology

KEGG: Kyoto Encyclopedia of Genes and Genomes

Declarations

Ethics approval and consent to participate

No animals were euthanized as part of this study and all sample collection methods and experimental procedures described herein were rigorously reviewed and approved by the Institutional Animal Care Committee of Razi Vaccine and Serum Research Institute (Permit number 960354-020-18-83-2) and AREEO protocols, which are in compliance with Iran guidelines for work with animals. This study also adheres to the ARRIVE Guidelines for reporting animal research.

Consent for publication

Not applicable.

Availability of data and materials

The mRNA and protein sequences reported in this paper are appearing in the NCBI under the accession numbers OM296278, OM296279, OM296280, OM296281, OM296282, OM296283, OM296284, OM296285, OM296286, OM296287, OM296288, OM296289, OM296290, OM296291, OM296292, OM296293. The data are included in the Electronic Supplementary Materials.

Competing interests

The authors declare that they have no conflict of interest.

Funding

Financial support for the all authors was provided by Razi Vaccine and Serum Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran (960354-020-18-83-2).

Author's contributions

Conceived and designed the experiments: F.S. Collect and identify the scorpion specimens: H.J. and F.S. Analyzed the data: F.S. wrote the manuscript: F.S. and H.J. All authors have reads and approved the manuscript.

Acknowledgements

We thank the Razi Vaccine and Serum Research Institute, of Iran for their financial support and our colleagues for many stimulating discussions.

Author information

Department of Venomous Animals and Anti-Venom Production, Razi Vaccine and Serum Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Ahvaz, Iran.

Salabi, F & Jafai, H

References

1. Abu-Zaid A, Inaya A, Farooqui A, Ahmed N, Bakheet, D, Dzimiri N, Nader, M (2012) Striatin is a novel regulator of calcium homeostasis in cardiomyocytes.
2. Ahmadi S, Knerr JM, Argemi L, Bordon KC, Pucca MB, Cerni FA, Arantes EC, Çalışkan F, Laustsen AH (2020) Scorpion venom: detriments and benefits. *Biomedicines*. 8(5):118.
3. Ahmed ST. Morphology And Histology Of Venom Gland Of Scorpio Maurus Kruglovi (Birula, 1910) (Scorpionidae: Scorpiones) (2015) Morphology And Histology Of Venom Gland Of Scorpio Maurus Kruglovi (Birula, 1910)(Scorpionidae: Scorpiones). *Zanco Journal of Pure and Applied Sciences*, 27(5), pp.59-62.
4. Al-Asmari AK, Kunnathodi F, Al Saadon K, Idris MM (2016) Elemental analysis of scorpion venoms. *Journal of venom research*, 7, p.16.
5. Bannenberg G, Serhan CN (2010) Specialized pro-resolving lipid mediators in the inflammatory response: An update. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*. 1;1801(12):1260-73.
6. Barth FG, Wadeuhl M (1975) Slit sense organs on the scorpion leg (*Androctonus australis* L., Buthidae). *Journal of morphology*, 145(2), pp.209-227.
7. Bergink EW, Wallace RA, Van de Berg JA, Bos ES, Gruber M, AB G (1974) Estrogen-induced synthesis of yolk proteins in roosters. *American Zoologist*, 14(4), pp.1177-1193.
8. Blake DJ, Weir A, Newey SE, Davies KE (2002) Function and genetics of dystrophin and dystrophin-related proteins in muscle. *Physiological reviews*.

9. Bottrall JL, Madaras F, Biven CD, Venning MG, Mirtschin PJ (2010) Proteolytic activity of Elapid and Viperid Snake venoms and its implication to digestion. *Journal of venom research*. 1:18.
10. Bougis PE, Rochat H, Smith LA (1989) Precursors of *Androctonus australis* scorpion neurotoxins. Structures of precursors, processing outcomes, and expression of a functional recombinant toxin II. *J Biol Chem*. 264:19259–65.
11. Castanón MJ, Walko G, Winter L, Wiche G (2013) Plectin–intermediate filament partnership in skin, skeletal muscle, and peripheral nerve. *Histochemistry and cell biology*, 140(1), pp.33-53.
12. Chen JV, Kao LR, Jana SC, Sivan-Loukianova E, Mendonça S, Cabrera OA, Singh P, Cabernard C, Eberl DF, Bettencourt-Dias M, Megraw TL (2015) Rootletin organizes the ciliary rootlet to achieve neuron sensory function in *Drosophila*. *Journal of Cell Biology*, 211(2), pp.435-453.
13. Cid-Urbe JI, Santibáñez-López CE, Meneses EP, Batista CV, Jiménez-Vargas JM, Ortiz E, Possani LD (2018) The diversity of venom components of the scorpion species *Paravaejovis schwenkmeyeri* (Scorpiones: Vaejovidae) revealed by transcriptome and proteome analyses. *Toxicon*, 151, pp.47-62.
14. Conti FJ, Monkley SJ, Wood MR, Critchley DR, Müller U (2009) Talin 1 and 2 are required for myoblast fusion, sarcomere assembly and the maintenance of myotendinous junctions. *Development*, 136(21), pp.3597-3606.
15. Crucitti P (1999) The scorpions of Anatolia: biogeographical patterns. *Biogeographica* 20, 81–94.
16. Daniel JG, Panizzi JR (2019) Spatiotemporal expression profile of embryonic and adult ankyrin repeat and EF-hand domain containing protein 1-encoding genes *ankef1a* and *ankef1b* in zebrafish. *Gene Expression Patterns*, 34, p.119069.
17. Dunlop J.A, (2010) Geological history and phylogeny of Chelicerata. *Arthropod Struct. Dev.* 39, 124e142.
18. Durbeej M, Henry MD, Ferletta M, Campbell KP, Ekblom P (1998) Distribution of dystroglycan in normal adult mouse tissues. *Journal of Histochemistry & Cytochemistry*, 46(4), 449-457.
19. e Oliveira UC, Candido DM, Dorce VA, Junqueira-de ID (2015) The transcriptome recipe for the venom cocktail of *Tityus bahiensis* scorpion. *Toxicon*, 95, pp.52-61.
20. Ekblom P (1996) Receptors for laminins during epithelial morphogenesis. *Curr Opin Cell Biol* 5:700–706.
21. El Refaey MM, Mohler PJ (2017) Ankyrins and spectrins in cardiovascular biology and disease. *Frontiers in Physiology*, 8, p.852.
22. Escalante T, Rucavado A, Gutiérrez JM (2009) Snake venom metalloproteinases. Biological roles and participation in the pathophysiology of envenomation. *Handbook of venoms and toxins of reptiles*, pp.115-138.
23. Fawcett DW, Porter KR (1954) A study of the fine structure of ciliated epithelia. *Journal of Morphology*, 94(2), pp.221-281.
24. Foelix RF, Müller-Vorholt G (1983b) The fine structure of scorpion sensory organs. II. Pecten sensilla. *Bulletin of the British Arachnological Society*, 6(2), pp.68-74.

25. Foelix RF, Schabronath J (1983a) The fine structure of scorpion sensory organs. I. Tarsal sensilla. *Bulletin of the British Arachnological Society*, 6, pp.53-67.
26. Ginkel LM, Wordeman L (2000) Expression and partial characterization of kinesin-related proteins in differentiating and adult skeletal muscle. *Molecular biology of the cell*, 11(12), 4143-4158.
27. Gomes ER, Cadot B (2017) Molecular motors and nuclear movements in muscle. *Communicative & Integrative Biology*, 10(3), p.e1319537.
28. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology*, 29(7), p.644.
29. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*. 8(8):1494-512.
30. Hagedorn HH, Fallon AM, Laufer H (1973) Vitellogenin synthesis by the fat body of the mosquito *Aedes aegypti*: Evidence for transcriptional control. *Developmental Biology*, 31(2), pp.285-294.
31. Han MK, Van Der Krogt GN, De Rooij J (2017) Zygotic vinculin is not essential for embryonic development in zebrafish. *PloS one*, 12(8), p.e0182278.
32. Hara Y, Campbell KP (2014) Dystroglycan: an extracellular matrix receptor that links to the cytoskeleton. *Glycoscience: Biology and Medicine*, pp.1245-1251.
33. Hasegawa H, Noguchi J, Yamashita M, Okada R, Sugimoto R, Furuya M, Unoki T, Funakoshi Y, Baba T, Kanaho Y (2012) Phosphatidylinositol 4-phosphate 5-kinase is indispensable for mouse spermatogenesis. *Biology of reproduction*, 86(5), pp.136-1.
34. Henry M.D, Campbell K.P, (1998) A role for dystroglycan in basement membrane assembly. *Cell*, 95(6), pp.859-870.
35. Hooper SL, Thuma JB (2005) Invertebrate muscles: muscle specific genes and proteins. *Physiological reviews*, 85(3), pp.1001-1060.
36. Hosseinzadeh S, Ramsey J, Mann M, Bennett L, Hunter WB, Shams-Bakhsh M, Hall DG, Heck M (2019) Color morphology of *Diaphorina citri* influences interactions with its bacterial endosymbionts and 'Candidatus *Liberibacter asiaticus*'. *PloS one*, 14(5), p.e0216599.
<https://doi.org/10.1371/journal.pone.0216599>
37. Huai D, Zhang Y, Zhang C, Cahoon EB, Zhou Y (2015) Combinatorial effects of fatty acid elongase enzymes on nervonic acid production in *Camelina sativa*. *PloS one*, 10(6), p.e0131755.
38. Jaenicke E, Pairet B, Hartmann H, Decker H (2012) Crystallization and Preliminary Analysis of Crystals of the 24-Meric Hemocyanin of the Emperor Scorpion (*Pandinus imperator*). *PLoS ONE* 7(3): e32548.
39. Jalali, A. and Rahim, F., 2014. Epidemiological review of scorpion envenomation in Iran. *Iranian journal of pharmaceutical research: IJPR*, 13(3), p.743.

40. Jenardhanan, P. and Mathur, P.P., 2014. Kinases as targets for chemical modulators: Structural aspects and their role in spermatogenesis. *Spermatogenesis*, 4(2), p.e979113.
41. Kaltsas D, Stathi I, Mylonas M (2008) The foraging activity of *Mesobuthus gibbosus* (Scorpiones: Buthidae) in central and south Aegean archipelago. *J Nat Hist* 42:513–527
42. Keil, T.A., 2012. Sensory cilia in arthropods. *Arthropod structure & development*, 41(6), pp.515-534.
43. Knowlton, E.D. and Gaffin, D.D., 2011. Functionally redundant peg sensilla on the scorpion pecten. *Journal of Comparative Physiology A*, 197(9), p.895.
44. Koch, L. The taxonomy, geographic distribution and evolutionary radiation of Australo-Papuan scorpions, (Western Australian Museum, 1977).
45. Koenig, M., Monaco, A.P. and Kunkel, L.M., 1988. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell*, 53(2), pp.219-228.
46. Kramerova, I.A., Kawaguchi, N., Fessler, L.I., Nelson, R.E., Chen, Y., Kramerov, A.A., Kusche-Gullberg, M., Kramer, J.M., Ackley, B.D., Sieron, A.L. and Prockop, D.J., 2000. Papilin in development; a pericellular protein with a homology to the ADAMTS metalloproteinases. *Development*, 127(24), pp.5475-5485.
47. Leclerc, P., Goupil, S., Rioux, J.F., Lavoie Ouellet, C., Clark, M.È., Ruiz, J. and Saindon, A.A., 2020. Study on the role of calmodulin in sperm function through the enrichment and identification of calmodulin binding proteins in bovine ejaculated spermatozoa. *Journal of Cellular Physiology*, 235(6), pp.5340-5352.
48. Li, B. and Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 12(1), p.323.
49. Ma, G.X., Zhou, R.Q., Song, Z.H., Zhu, H.H., Zhou, Z.Y. and Zeng, Y.Q., 2015. Molecular mechanism of serine/threonine protein phosphatase 1 (PP1 α –PP1 γ) in spermatogenesis of *Toxocara canis*. *Acta tropica*, 149, pp.148-154.
50. 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.
51. Mack, I., Hector, A., Ballbach, M., Kohlhäufel, J., Fuchs, K.J., Weber, A., Mall, M.A. and Hartl, D., 2015. The role of chitin, chitinases, and chitinase-like proteins in pediatric lung diseases. *Molecular and cellular pediatrics*, 2(1), pp.1-8.
52. Martin-Eauclaire, M.F. and Bougis, P.E., 2012. Potassium channels blockers from the venom of *Androctonus mauretanicus mauretanicus*. *Journal of toxicology*, 2012.
53. Matute-Blanch, C., Calvo-Barreiro, L., Carballo-Carbajal, I., Gonzalo, R., Sanchez, A., Vila, M., Montalban, X. and Comabella, M., 2020. Chitinase 3-like 1 is neurotoxic in primary cultured neurons. *Scientific Reports*, 10(1), pp.1-6.
54. Merzendorfer, H. and Zimoch, L., 2003. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *Journal of Experimental Biology*, 206(24), pp.4393-4412.

55. Mirshamsi, O.; Sari, A.; Hosseini, S. History of study and checklist of the scorpion fauna (Arachnida: Scorpiones) of Iran. *Prog. Biol. Sci.* 2011, 1, 16–28.
56. Mohamed Abd El-Aziz, T., Soares, A.G. and Stockand, J.D., 2019. Snake venoms in drug discovery: valuable therapeutic tools for life saving. *Toxins*, 11(10), p.564.
57. Nader, M., Alotaibi, S., Alsomale, E., Bakheet, D. and Dzimiri, N., 2017. P3495Striatin regulates the dynamics of the Caveolin-3/Calmodulin complex and modulates cardiomyocyte spontaneous contraction rate. *European Heart Journal*, 38(suppl_1).
58. Naderi, M., Safahieh, A., Madiseh, S.D., Zolgharnein, H. and Ghatrami, E.R., 2015. Induction of vitellogenin synthesis in juvenile male yellowfin seabream (*Acanthopagrus latus*) exposed to 4-nonylphenol and 17 β -estradiol. *Toxicology and Industrial Health*, 31(3), pp.209-220.
59. Nakamura, M. T., Yudell, B. E. & Loor, J. J. Regulation of energy metabolism by long-chain fatty acids. *Prog. Lipid Res.* 53, 124–144 (2014).
60. Nelson, C.M., Ihle, K.E., Fondrk, M.K., Page Jr, R.E. and Amdam, G.V., 2007. The gene vitellogenin has multiple coordinating effects on social organization. *PLoS Biol*, 5(3), p.e62.
61. Oukkache N, Rosso JP, Alami M, Ghalim N, Saile R, Hassar M, Bougis PE, Martin-Eauclaire MF: New analysis of the toxic compounds from the *Androctonus mauretanicus mauretanicus* scorpion venom. *Toxicon*. 2008, 51 (5): 835-852. 10.1016/j.toxicon.2007.12.012.
62. Padmanabhanaidu, B., 1966. Physiological properties of the blood and haemocyanin of the scorpion *Heterometrus pulvipes*. *Comparative Biochemistry and Physiology*, 17(1), pp.167-181.
63. Paixão-Cavalcante, D., Van den Berg, C.W., de Freitas Fernandes-Pedrosa, M., de Andrade, R.M.G. and Tambourgi, D.V., 2006. Role of matrix metalloproteinases in HaCaT keratinocytes apoptosis induced by *Loxosceles* venom sphingomyelinase D. *Journal of investigative dermatology*, 126(1), pp.61-68.
64. Powers, K., Schappacher-Tilp, G., Jinha, A., Leonard, T., Nishikawa, K. and Herzog, W., 2014. Titin force is enhanced in actively stretched skeletal muscle. *Journal of Experimental Biology*, 217(20), pp.3629-3636.
65. Pratt, M.B., Titlow, J.S., Davis, I., Barker, A.R., Dawe, H.R., Raff, J.W. and Roque, H., 2016. *Drosophila* sensory cilia lacking MKS proteins exhibit striking defects in development but only subtle defects in adults. *Journal of cell science*, 129(20), pp.3732-3743.
66. Quintero-Hernández, V., Jiménez-Vargas, J.M., Gurrola, G.B., Valdivia, H.H. and Possani, L.D., 2013. Scorpion venom components that affect ion-channels function. *Toxicon*, 76, pp.328-342.
67. Quintero-Hernández, V., Ramírez-Carretero, S., Romero-Gutiérrez, M.T., Valdez-Velázquez, L.L., Becerril, B., Possani, L.D. and Ortiz, E., 2015. Transcriptome analysis of scorpion species belonging to the *Vaejovis* genus. *PloS one*, 10(2), p.e0117188.
68. Reynolds, S. E. and Samuels, R. I. (1996). Physiology and biochemistry of insect moulting fluid. *Adv. Insect Physiol.* 26,157 -232. Google Scholar
69. Rivera-Perez, C., Nouzova, M., Clifton, M.E., Garcia, E.M., LeBlanc, E. and Noriega, F.G., 2013. Aldehyde dehydrogenase 3 converts farnesal into farnesoic acid in the corpora allata of mosquitoes. *Insect biochemistry and molecular biology*, 43(8), pp.675-682.

70. Robinson, M.D., McCarthy, D.J. and Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139-140.
71. Rokyta, D.R. and Ward, M.J., 2017. Venom-gland transcriptomics and venom proteomics of the black-back scorpion (*Hadrurus spadix*) reveal detectability challenges and an unexplored realm of animal toxin diversity. *Toxicon*, 128, pp.23-37.
72. Rosell, R. and Coons, L.B., 1992. The role of the fat body, midgut and ovary in vitellogenin production and vitellogenesis in the female tick, *Dermacentor variabilis*. *International journal for parasitology*, 22(3), pp.341-349.
73. Roy-Zokan, E.M., Cunningham, C.B., Hebb, L.E., McKinney, E.C. and Moore, A.J., 2015. Vitellogenin and vitellogenin receptor gene expression is associated with male and female parenting in a subsocial insect. *Proceedings of the Royal Society B: Biological Sciences*, 282(1809), p.20150787.
74. Schuster, M., Treitschke, S., Kilaru, S., Molloy, J., Harmer, N.J. and Steinberg, G., 2012. Myosin5, kinesin1 and myosin17 cooperate in secretion of fungal chitin synthase. *The EMBO journal*, 31(1), pp.214-227.
75. Singiser, R.H. and McCann, R.O., 2006. Evidence that talin alternative splice variants from *Ciona intestinalis* have different roles in cell adhesion. *BMC cell biology*, 7(1), p.40.
76. Snake venoms and its implication to digestion. *J. Venom. Res.* 2010, 1, 18–28.
77. Stanley-Samuelson, D. W. & Dadd, R. H. Long-chain polyunsaturated fatty acids: Patterns of occurrence in insects. *Insect Biochem.* 13, 549–558 (1983).
78. Styczynska-Soczka, K. and Jarman, A.P., 2015. The *Drosophila* homologue of Rootletin is required for mechanosensory function and ciliary rootlet formation in chordotonal sensory neurons. *Cilia*, 4(1), p.9.
79. Sun, C. and Zhang, S., 2015. Immune-relevant and antioxidant activities of vitellogenin and yolk proteins in fish. *Nutrients*, 7(10), pp.8818-8829.
80. Tonino, P., Kiss, B., Strom, J., Methawasin, M., Smith, J.E., Kolb, J., Labeit, S. and Granzier, H., 2017. The giant protein titin regulates the length of the striated muscle thick filament. *Nature communications*, 8(1), pp.1-11.
81. Tsujioka M, Machesky LM, Cole SL, Yahata K, Inouye K: A unique talin homologue with a villin headpiece-like domain is required for multicellular morphogenesis in *Dictyostelium*. *Curr Biol.* 1999, 9 (7): 389-392. 10.1016/S0960-9822(99)80169-9.
82. Tsujioka M, Yoshida K, Inouye K: Talin B is required for force transmission in morphogenesis of *Dictyostelium*. *Embo J.* 2004, 23 (11): 2216-2225. 10.1038/sj.emboj.7600238.
83. Valdez-Velázquez, L.L., Quintero-Hernández, V., Romero-Gutiérrez, M.T., Coronas, F.I. and Possani, L.D., 2013. Mass fingerprinting of the venom and transcriptome of venom gland of scorpion *Centruroides tecomanus*. *PloS one*, 8(6), p.e66486.
84. Valle, D., 1993. Vitellogenesis in insects and other groups: a review. *Mem. Inst. Oswaldo Cruz*, pp.1-26.

85. Vetter, I., Davis, J.L., Rash, L.D., Anangi, R., Mobli, M., Alewood, P.F., Lewis, R.J. and King, G.F., 2011. Venomics: a new paradigm for natural products-based drug discovery. *Amino acids*, 40(1), pp.15-28.
86. Wang, X., Ma, S., Mao, M., Li, C., Shen, X., Xu, S. and Yang, J., 2019. RNA-sequencing and bioinformatics analysis of long noncoding RNAs and mRNAs in the prefrontal cortex of mice following repeated social defeat stress. *BioMed research international*, 2019.
87. Wanhg, L.J. and Knowland, J., 1975. Synthesis of vitellogenin in cultures of male and female frog liver regulated by estradiol treatment in vitro. *Proceedings of the National Academy of Sciences*, 72(8), pp.3172-3175.
88. Ward, M.J., Ellsworth, S.A. and Rokyta, D.R., 2018. Venom-gland transcriptomics and venom proteomics of the Hentz striped scorpion (*Centruroides hentzi*; Buthidae) reveal high toxin diversity in a harmless member of a lethal family. *Toxicon*, 142, pp.14-29.
89. Wen, J., Egan, A.N., Dikow, R.B. and Zimmer, E.A., 2015. Utility of transcriptome sequencing for phylogenetic inference and character evolution. *Next-generation sequencing in plant systematics*.
90. Wiche G, Krepler R, Artlieb U, Pytela R, Denk H. Occurrence and immunolocalization of plectin in tissues. *J Cell Biol*. 1983;97(3):887–901.
91. Wu, J.Y., Ribar, T.J. and Means, A.R., 2001. Spermatogenesis and the Regulation of Ca²⁺-Calmodulin-Dependent Protein Kinase IV Localization Are Not Dependent on Calspermin. *Molecular and cellular biology*, 21(17), pp.6066-6070.
92. Yagmur, E.A.; Moradi, M.; Larti, M.; Lashkari, S. First record of *Androctonus robustus* Kovač & Ahmed, 2013. (Scorpiones: Buthidae) for Iran. *Zool. Middle East*. 2016, 62, 370–372.
93. Yang, J., Liu, X., Yue, G., Adamian, M., Bulgakov, O. and Li, T., 2002. Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet. *The Journal of cell biology*, 159(3), pp.431-440.
94. Yin, H., Guan, N., Dong, L., Yue, Q., Yin, X. and Zhang, D., 2012. Initial analysis of the hemocyanin subunit type 1 (Hc1 gene) from *Locusta migratoria manilensis*. *Molecular biology reports*, 39(3), pp.3305-3310.
95. Zhang, L., Shi, W., Zeng, X.C., Ge, F., Yang, M., Nie, Y., Bao, A., Wu, S. and Guoji, E., 2015. Unique diversity of the venom peptides from the scorpion *Androctonus bicolor* revealed by transcriptomic and proteomic analysis. *Journal of proteomics*, 128, pp.231-250.
96. Zhang, X.Y. and Zhang, P.Y., 2016. Scorpion venoms in gastric cancer. *Oncology letters*, 12(5), pp.3683-3686.
97. Zheng, T., Li, H., Han, N., Wang, S., Price, J.H., Wang, M. and Zhang, D., 2017. Functional characterization of two elongases of very long-chain fatty acid from *Tenebrio molitor* L. (Coleoptera: Tenebrionidae). *Scientific Reports*, 7(1), pp.1-10.

Figures

Figure 1

Overview of experimental and computational analyzing workflow. Briefly, the workflow involved stepwise filters based on annotations, gene expression, Differential expression analysis, and GO term and KEGG pathway enrichment analysis.

Figure 2

Characterization of *A. crassicauda* mRNA. (A) Length distribution of mRNA transcripts of adult and juvenile (B) Expression levels of mRNAs transcripts of adult and juvenile.

Figure 3

The profiles of differentially expressed mRNAs in adult and juvenile groups. (A) Up-regulated and down-regulated mRNAs (B) Differentially expressed mRNAs were analyzed using volcano plot. Red and blank points refer to those differentially and not differentially expressed genes. The scatter plot is a visualization method used for showing differentially expressed. (C) Crosswise comparison of the gene abundance in the two samples adult and juvenile. Colors from purple to yellow represent the gene expression abundance from poor to rich, respectively.

Figure 4

Classification of *up and down-regulated mRNAs* into venom components. **A.** Relative distribution of the up-regulated genes identified in the venom gland transcriptome of *A. crassicauda*. **B.** Relative distribution of the down-regulated genes identified in the venom gland transcriptome of *A. crassicauda*.

Figure 5

Gene Ontology (GO) analysis of the differentially expressed mRNAs from our dataset. (A) GO analysis of up-regulated mRNAs. (B) GO analysis of down-regulated mRNAs.

Figure 6

KEGG analyses of differentially expressed genes between the adult and juvenile groups. (A) The KEGG enrichment pathways for up-regulated mRNAs. (B) The KEGG enrichment pathways for down-regulated mRNAs. *P*-value cut-off was set at 0.01.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfiles.docx](#)
- [Supplementaryfile7.xlsx](#)
- [Supplementaryfile8.xlsx](#)