RESEARCH ARTICLE

Sugarcane Giant Borer Transcriptome Analysis and Identification of Genes Related to Digestion

Fernando Campos de Assis Fonseca[1,2]*‡, **Alexandre Augusto Pereira Firmino**[1,3]‡, **Leonardo Lima Pepino de Macedo**[1,4], **Roberta Ramos Coelho**[1,2], **José Dijair Antonino de Sousa Júnior**[1,2], **Orzenil Bonfim Silva-Junior**[1,4], **Roberto Coiti Togawa**[1], **Georgios Joannis Pappas Jr**[2], **Luiz Avelar Brandão de Góis**[5], **Maria Cristina Mattar da Silva**[1], **Maria Fátima Grossi-de-Sá**[1,4]*

1 Embrapa Recursos Genéticos e Biotecnologia, Brasília, Distrito Federal, Brazil, **2** Universidade de Brasília, Brasília, Distrito Federal, Brazil, **3** Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil, **4** Universidade Católica de Brasília, Brasília, Distrito Federal, Brazil, **5** Usina triunfo, Maceió, Alagoas, Brazil

‡ These authors contributed equally to this work.
* fcafonseca@gmail.com (FCAF); fatima.grossi@embrapa.br (MFGS)

Abstract

Sugarcane is a widely cultivated plant that serves primarily as a source of sugar and ethanol. Its annual yield can be significantly reduced by the action of several insect pests including the sugarcane giant borer (*Telchin licus licus*), a lepidopteran that presents a long life cycle and which efforts to control it using pesticides have been inefficient. Although its economical relevance, only a few DNA sequences are available for this species in the GenBank. Pyrosequencing technology was used to investigate the transcriptome of several developmental stages of the insect. To maximize transcript diversity, a pool of total RNA was extracted from whole body insects and used to construct a normalized cDNA database. Sequencing produced over 650,000 reads, which were *de novo* assembled to generate a reference library of 23,824 contigs. After quality score and annotation, 43% of the contigs had at least one BLAST hit against the NCBI non-redundant database, and 40% showed similarities with the lepidopteran *Bombyx mori*. In a further analysis, we conducted a comparison with *Manduca sexta* midgut sequences to identify transcripts of genes involved in digestion. Of these transcripts, many presented an expansion or depletion in gene number, compared to *B. mori* genome. From the sugarcane giant borer (SGB) transcriptome, a number of aminopeptidase N (APN) cDNAs were characterized based on homology to those reported as Cry toxin receptors. This is the first report that provides a large-scale EST database for the species. Transcriptome analysis will certainly be useful to identify novel developmental genes, to better understand the insect's biology and to guide the development of new strategies for insect-pest control.

Citation: Fonseca FCdA, Firmino AAP, Macedo LLPd, Coelho RR, Antonino de Souza Júnior JD, Silva-Junior OB, et al. (2015) Sugarcane Giant Borer Transcriptome Analysis and Identification of Genes Related to Digestion. PLoS ONE 10(2): e0118231. doi:10.1371/journal.pone.0118231

Academic Editor: Erjun Ling, Institute of Plant Physiology and Ecology, CHINA

Received: September 8, 2014

Accepted: January 11, 2015

Published: February 23, 2015

Data Availability Statement: Pyrosequencing data were deposited in the Short Read Archive of the National Center for Biotechnology Information with accession number SRR1204999 (http://www.ncbi.nlm.nih.gov/). The DNA sequences submitted to the dBEST database under accession numbers JZ578314–JZ578318 (http://www.ncbi.nlm.nih.gov/dbEST/)

Funding: This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (http://www.capes.gov.br/), Empresa Brasileira de Pesquisa Agropecuária (http://www.embrapa.br/), Conselho Nacional de Desenvolvimento Científico e Tecnológico (http://www.cnpq.br/), and Fundação de Apoio à Pesquisa do Distrito Federal (http://www.fap.df.gov.br/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Luiz AvelarBrandao de Gois is affiliated with TRIUNFO AGROINDUSTRIAL LTDA - accordingly with current legislation, as a phytosanitary and environmental coordination, occupying the position of biologist. The authors declare that there is no interest from this company in patents or any other products that may be obtained from the present study described in the manuscript. This does not alter the authors' adherence to all PLOS ONE policies on sharing data and materials.

Introduction

Lepidopteran stem borers are economically important agricultural insect pests that cause severe damage to sugarcane crops worldwide. Because of its endophytic behavior, the use of chemical pesticides is mostly ineffective and requires manual labor, which is time consuming and increases the cost [1,2]. Historically, synthetic insecticides, which are usually aerially applied, were used against moth borers accomplishing moderated success. Furthermore, repetitious applications may contribute to environmental pollution and generate health problems [3,4]. Alternatives to chemical control in sugarcane crops include transgenes based on the endotoxins produced by the bacterium *Bacillus thuringiensis* (Bt), which is widely used for protection against lepidopterans in cereal crops such as rice and maize [5,6]. There are several lepidopteran insect pests that affect sugarcane worldwide that could be targeted by the expression of Bt toxins in transgenic plants [7]. However, to combat the numerous species of borers involved in the infestation of sugarcane crops, basic and applied research must be conducted to increase our knowledge of the biology and ecology, as well as management of the species.

The sugarcane giant borer *Telchin licus licus* (SGB) (Drury, 1773) (Lepidoptera: Castiniidae) has become a major insect pest in the sugarcane fields of Central and South America [8]. To control infestation, several methods were evaluated, including biological control, mechanical collection and identification of resistant plants, but none of these strategies turned out to be successful. Moreover, the insect spends most of its life cycle, -six to ten months-, feeding inside the plant, and is not easily affected by chemical pesticides [9]. Despite its economic importance, studies with this insect aiming pest control are still at an early stage. At present, no reports demonstrating the establishment of an insect rearing system for this insect is available, thereby hindering the study of its development. To date, a few published reports have discussed about the chemical composition of pheromones [10], body morphology [11], entomopathogenic activity of fungi [12] and *in vitro* evolution of Cry toxins [13]. Recently, 109 mitochondrial DNA sequences were used to determine the origin of invasive subspecies of *T*. *licus* in sugarcane producing regions in Brazil [14] and are now publicly available at the GenBank. The lack of information regarding the insect´s physiology and molecular biology obliged us to make use of data from model insects like *Bombyx mori*, whose genome has been fully sequenced [15] and Expressed Sequence Tags (ESTs) libraries from organisms of the same order [16], hampering cloning and characterization of genes and their expression analysis.

Large scale sequencing through next generation sequencing technologies has effectively increased the number and depth of genomic and transcriptomic data [17–19]. This approach is being applied in entomology for gene discovery [20], gene expression analysis of specific tissues and at different physiological conditions [21], to analyze single nucleotide polymorphisms (SNPs) [22], development studies of resistance to insecticides [23,24], understanding endosymbiosis interactions [25], and to determine specific target genes to be silenced through RNA interference [26]. With the increase of DNA sequence information, new opportunities for analysis are emerging. To date poplar leaf beetle's (*Chrysomela tremulae*) and tobacco hornworm's (*Manduca sexta*) midgut transcriptomes have been sequenced [27,28], making it possible to compare data from different organisms and identify new gene families. Likewise, transcriptome analysis can provide an assessment of gene families that can be used in the biotechnological industry as those involved in biomass conversion [29]. Recent reports have demonstrated the presence of genes related to cellulose degradation in some insect species [27], with the possibility of finding similar genes in different organisms.

The insect midgut is involved in several processes, including digestion, immunity and mechanical protection. For these reasons it became an important organ to study [30]. Searching for new genes may help to understand (i) the role of receptors that participate in toxin resistance [31], (ii) ecological traits [32], (iii) disease transmission [33] and (iv) identify which phenotypic alterations may be caused by silencing specific genes [34].

Depending on their feeding habits, insects can regulate gene expression to increase their adaptability to different diets [35]. In contrast to other lepidopteran caterpillars that feed primarily on leaves, the SGB presents an endophytic behavior, feeding within the steam of sugarcane plants, whose composition is richer in sucrose than starch. This characteristic can influence the differential expression of digestive enzymes, leading to a specific intestinal homogenate composition that guarantees the insect´s survival and development within the plant.

The insect digestive system is primarily composed of proteases, a major group of hydrolytic enzymes that participates on digestion [36] and proenzyme activation [37]. These enzymes are also related to developmental processes such as molting and metamorphosis, can act as a regulator of innate immune response [38–41] and are associated with Cry toxin activation and solubilization [42,43]. Several reports have proposed the digestive system as a target for pest population management [44–47], based on the fact that any alterations in the enzyme activity of the intestinal homogenate or the use of protease inhibitors (PIs) can modify the insects' metabolism and lead to a reduction of nutrient absorption, thus hindering its development [48,49]. Recently, a Kunitz-type protease inhibitor was characterized against insects' intestinal homogenate, including SGB larvae and a substantial decrease in midgut serine protease activity was observed, although they provided no information concerning the mortality of the insects [50]. In a variety of studies, protease inhibitors have been successfully evaluated, demonstrating that transgenic adoption strategy is an efficient method for insect pest control [51–53].

Among several classes of proteases that are targeted by insect management programs, aminopeptidase N (APN) stands out for its dual characteristic. Besides its importance in digestion of proteins and release of amino acids for cell metabolism, APNs have been studied, together with cadherin-like (BT-R) proteins and alkaline phosphatases (ALP) as putative Cry toxin receptors [54–57]. Silencing APN genes through RNAi demonstrated that the susceptibility of the insect to Cry toxins could be altered [58,59] and that, at least for a few species, the development of resistance to Bt-based pesticides involves the alteration of APN gene expression [60].

The present work describes the construction of an EST dataset of *T. licus licus*, a non-model organism, which attacks severely sugarcane crops and for which no information about its molecular biology and physiology is available. To identify a broad range of genes, a normalized whole-body library containing different life stages, including eggs, early-stage larvae, late-stage larvae, prepupae, pupae and adults (both male and female) was sequenced using the 454 sequencing system. This research focused on describing the transcripts involved in the expression of digestive enzymes and highlight potential candidates to be used for pest control. Over 24,000 contigs were obtained from which a selection is currently investigated to get insight into the insect's biology and to develop new strategies for a more effective pest management.

Materials and Methods

Insects and RNA extraction

The Chico Mendes Institute for Biodiversity Conservation (ICMBIO), through the Biodiversity Information System (SISBIO), authorizes the collection of specimens of Brazilian fauna for research purposes in public and private properties. We obtained a Permanent License for Fauna Collection, n° 34833-1, issued in 07/05/2012. This License authorized the Brazilian Agricultural Research Corporation (Embrapa) and its researchers to collect the specimens used in this study. No endangered or protected species were involved in this research.

Late-stage larvae, prepupae, pupae and adults were collected directly from infested sugarcane fields of Triunfo sugarcane mill, located in Maceió, Alagoas state (AL). Females were kept in entomological cages for oviposition. The eggs were collected and maintained at 28 ± 2°C, 70 ± 10% relative humidity and a photoperiod of 12:12 h (Light: Dark) until hatching. The larvae were individualized and reared with sugarcane pieces. A pool of insects of the same stage was frozen in liquid nitrogen prior to RNA isolation.

Total RNA was extracted separately from each insect stage, eggs, early-stage larvae, late-stage larvae, prepupae, pupae, and both male and female adults using Trizol Reagent (Invitrogen, Life Technologies), according to the manufacturer's protocol. RNA was treated with RNAse-free DNase I (Ambion, Life Technologies) at 37°C for 30 minutes, according to the manufacturer's protocol.

cDNA Library Normalization and Pyrosequencing

A pool of 30 μg of all insect stages total RNA was sent to Eurofins MWG Operon, in Huntsville, AL, USA (http://www.eurofinsdna.com/) to synthesize a cDNA library.

The RNA quality was assessed using an Agilent 2100 Bioanalyzer prior to cDNA library construction. Full-length, enriched cDNAs were generated using the SMART PCR cDNA synthesis kit (BD Clontech) following the manufacturer's protocol. The resulting double-stranded cDNAs were normalized using the Kamchatka crab duplex-specific nuclease method (Trimmer cDNA normalization kit, Evrogen) to prevent over-representation of the most common transcripts [61]. The normalized transcripts were submitted to a half-plate run using the 454 pyrosequencing, GS FLX Titanium technology, according to the protocols provided by the manufacturer (Roche 454 Life Sciences). Raw data from the sequencing run were submitted to the Sequence Read Archive repository of the National Center for Biotechnology Information (NCBI) under accession number SRR1204999.

Data Pre-Processing

Pre-processing of pyrosequencing reads for quality and adaptor trimming was first performed using the runAssembly function of Newbler version 2.5.3 using -cdna and -tr options. This latter option was used to output the trimmed reads. *Est2assembly* 1.03 platform was used on previously trimmed reads to prepare data for the assembler [62]. Contaminant sequences (prokaryotic, viral and mitochondrial sequences) were removed after BLAST analysis using locally prepared databases. Repetitive sequence identification and Poly A/T tail trimming was performed using RepeatMasker and standardizing options in the est2assembly preprocessed pipeline.

Assembly, Annotation and Gene Ontology (GO)

As there were no SGB data or DNA sequences of other phylogenetically related organisms, the contigs were *de novo* assembled using MIRA v3.3.0.1 [63]. The resulting contigs were submitted to the Transcriptome Shotgun Assembly repository of the National Center for Biotechnology Information (NCBI) under accession number SRR1204999. Unique sequences were determined by similarity searches using the BLASTx tool. Functional annotation by GO terms (http://www.geneontology.org), InterPro entries (InterProScan; http://www.ebi.ac.uk/Tools/pfa/iprscan/), enzyme classification codes (EC) and metabolic pathways (KEGG, Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/) were determined using the Blast2GO software suite v2.4.3 (http://www.blast2go.org) [64]. Sequences were submitted to the NCBI protein nr databank via BLASTx, with an e-value threshold of 1e^{-5}. False Discovery Rate (FDR) was used at probability level of 0.05%. GO terms were improved with the ANNEX tool [65], followed by GOSlim tool available at Blast2GO (goslim_generic.obo) [66]. Combined graphs were constructed at level 2, for Biological Process, Molecular Function and Cellular

Component categories. Enzymatic classification codes and KEGG metabolic pathways were generated by direct mapping of GO terms, with their respective ECs. InterPro searches were performed remotely from Blast2GO on InterProEBI server.

De novo Contig Assembly Quality Assessment

Contig sequences after assembly were analyzed for biases already known to be induced by the methods used to generate cDNA libraries accordingly to the NGS technologies requirements (i.e. fragmentation and synthesis) that could impact the quality and further functional analysis. To do this we carried a two-tier analysis of the sequencing coverage at base pair-level by: a) analyzing the alignments of the EST sequences on the reference genome of the related *B*. *mori* for all the annotated genomic loci representing protein-coding regions; b) analyzing the alignments of the EST sequences back on the 13,562 contigs that we did not found any functional information. For a) we use the GFF formatted file distributed along the *B*. *mori* genome (SilkDB 2.0) describing the genomic position of the exons for each one of the 16,823 annotated transcripts. In summary we use the GMAP program with the cross-species option to gather the read depth coverage at base pair-level from the alignment of the 61,775 ESTs on 1,009 putative protein-coding genomic loci with RPKM > 0.125. This cut-off value of RPKM was used accordingly as reported previously in a RNA-Seq study aimed to determine the minimum detectable level of expression below which expression can be considered as noise [67]. At this cut-off the mean depth of the coverage of data from the remaining aligned ESTs was 29.8x for the average sampled genes. For b) in the absence of reference gene structure we extracted, for each contig sequence, coverage data from the regions spanning 5% to 15% and 85% to 95% of the putative transcript, by length. These two coverage regions represent in our analysis the ends (3' UTR/ 5' UTR) of the transcript. The first and last 5% of the sequence, by length, was excluded to avoid artifacts from the assembly process. The remaining region, spanning 15% to 85% represent the middle of the cDNA strand. This approach rely largely in the processivity analysis for RNA-Seq coverage data described by Lahens and coworkers (2014) [68]. Both analysis for coverage of data were conducted using BEDTools program [69]. The geneBody_coverage module of the RSeQC program was used to infer if ESTs coverage along the reference models is uniform and if there is any bias towards the ends or the mid-range of the sequences [70].

Sequence Comparison with *M. sexta* and *B. mori* data

Contig sequences were BLASTed against 13,828 midgut DNA sequences of *M*. *sexta* that were obtained from InsectaCentral [71]. Sequences with e-value score above the cut-off ($1e^{-3}$) were selected. The BLAST results were organized by InterPro terms and description; and the most frequent results were listed. Digestive enzyme sequences were searched by annotation, sequence similarities and InterPro terms.

Digestive enzyme sequences were submitted to the online tool TRAPID [72] for comparative analysis of SGB transcripts and the *B*. *mori* genome. The OrthoMCL-DB proteome database was used as a reference. Protein family groups were identified and organized accordingly to the gene expansion or depletion information.

5’ RACE of APN1

Sequencing of aminopeptidase N1 was completed by Rapid Amplification of cDNA Ends (RACE). Neonate larvae of SGB were frozen in liquid nitrogen and total RNA was extracted using TRIZOL reagent (Invitrogen, Life Technologies). An aliquot was applied on 1% agarose gel and subjected to electrophoresis to analyze the integrity of the RNA sample. Quantification was performed using a Nanovue Plus Spectrophotometer (GE life sciences). After incubation with DNase I (Ambion, Life Technologies), 5 μg total RNA was used for cDNA synthesis using M-MLV reverse transcriptase (Invitrogen, Life Technologies) and gene specific primer 1. Addition of a 3' hydroxyl terminus tail was made using terminal transferase enzyme (New England BioLabs). All experiments were performed accordingly to manufacture's protocols. Amplification of DNA fragment was carried out in two rounds of polymerase chain reaction (PCR). The first round consisted of 5 μL of cDNA template, 1X PCR buffer, 2 mM $MgCl_2$, 0.2 mM dNTP mix, 0.2 μM oligo-dT_{30} adapter primer, 0.2 μM gene specific primer 2, 0.1 units of *Taq* DNA Polymerase (Ludwig Biotec) and deionized water to a final volume of 20 μL. The reaction conditions were: 94°C for 1', 30 cycles of 94°C for 45", 60°C for 45" and 72°C for 1', followed by a final step of 72°C for 5'. The second round of PCR consisted of 1μl (1:20) of the first round template, 1X PCR buffer, 2 mM $MgCl_2$, 0.2 mM dNTP mix, 0.2 μM anchor primer, 0.2 μM gene specific primer 3, 0.1 units of *Taq* DNA Polymerase (Ludwig Biotec) and deionized water to a final volume of 20 μL. The reaction conditions were the same as cited previously. Primer sequences are shown in S1 Table. The entire volume was applied on 1% agarose gel and subjected to electrophoresis. DNA fragments were collected from the gel under UV light and purified using the QIAquick Gel Extraction Kit (Qiagen), followed by ligation into PCR 2.1 vector (Invitrogen, Life Technologies). After transformation of OMNIMAX *E. coli* cells using Heat Shock method (Invitrogen, Life Technologies) and purification of plasmids by alkaline lysis [73], sequencing was carried out using M13 forward and reverse primers on a ABI377 sequencer (Applied Biosystems, Life Technologies). The DNA sequences were submitted to the dBEST database under accession numbers JZ578314—JZ578318.

Real-Time PCR Experiments

Transcript levels of serine proteases and APNs were determined in three different tissues (Anterior midgut, Posterior midgut and Carcass) by qPCR using a 7500 Fast Real-Time PCR System (Applied Biosystems, Life Technologies). Insects were dissected under a stereomicroscope (Zeiss Stemi SV6, Jena, Germany) and the tissues directly frozen in liquid nitrogen. A pool of tissues was used for total RNA extraction using Trizol Reagent (Invitrogen, Life Technologies). After DNase I treatment (Ambion, Life Technologies), 1 μg total RNA was used for cDNA synthesis using M-MLV reverse transcriptase (Invitrogen, Life Technologies) and oligo-dT_{30} adapter primer, accordingly to manufacture's protocol. Reaction mixtures contained 2 μL of cDNA (1:20 diluted), 2.5 μL of Fast SYBR Green Master Mix (Applied Biosystems), 0.2 μM of each primer (S2 Table) and double distilled H_2O to a final volume of 10 μL. PCR conditions were as follows: 95°C for 20 s, followed by 40 cycles of 95°C for 3 s and 60°C for 30 s. At the end of the program a melting curve for each primer pair (60–94°C read every 0.5°C) was acquired to ensure that only single products were amplified. The SGB glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and 18S ribosomal subunit (RPS18) were used for normalization of qPCR data (S2 Table). Raw data were treated using the online tool qPCR miner (http://www.miner.ewindup.info) [74] to determine primers efficiency and Cq values. The relative expression of each gene was calculated using the qBASE plus program (Biogazelle, Belgium). Three independent quantitative qPCR reactions were carried out per sample and two biological replicates were performed.

Protease and Aminopeptidase Sequence Analysis

The amino acid sequences of serine proteases and aminopeptidases were obtained by *in silico* translation using TrEMBL (http://www.expasy.ch/tools/dna.html) [75]. Manual screening was carried out to correct mis-assembled contigs and frameshifts, when necessary. Prediction of signal peptides, molecular weight, isoelectric point and glycosylation sites were predicted by using, respectively, SignaIP 4.1, Compute pI/MW, NetNGlyc 1.0 and NetOGlyc 4.0 online tools hosted at ExPaSy: SIB Bioinformatics Resource Portal (http://www.expasy.org/tools/). The GPI anchoring signal was predicted by using PredGPI online tool (http://gpcr2.biocomp.unibo.it/gpipe/index.htm) [76].

Sequence Alignment and Phylogenetic Analysis

All sequences were aligned using ClustalW2 [77] and edited by BioEdit software v.7 [78]. Phylogenetic analysis were conducted using MEGA v.5 where neighbor-joining trees were constructed with bootstrap of 10,000 replicates and evolutionary divergence calculated by p-distance method [79].

Results and Discussion

Sequence Analysis, De Novo Assembly and Annotation

A cDNA library synthesized from a mixture of insect total RNA from different life stages was normalized to reduce over-abundant transcripts. Sequencing was carried out on a half plate of the GS-FLX 454 pyrosequencer, resulting in 653,511 reads with 381,273,406 bp. After bioinformatics preprocessing, 362,412 high-quality reads were obtained and assembled into 23,824 contigs with average depth of coverage of 8.5 sequences per nucleotide position and length from 290 bp to 5,527 bp, with an average of 633 bp (Table 1). A similar pattern of coverage data was observed for other non-model insects that had their transcriptomes sequenced on the 454 Titanium: The non-normalized transcriptome of several life stages of the insect *Anopheles funestus* were pyrosequenced on a half plate of the 454 Titanium generating a 8.3 read per contig coverage [80]. *De novo* transcriptome assembly was performed for the apple maggot (*Rhagoletis pomonella*) and obtained 13.92 reads per contig coverage [32]. The same strategy depicted in this study was recently used by our group to generate an average coverage of 9.58 for the coleopteran *Anthonomus grandis* transcriptome [26]. Sequencing raw data for the

Table 1. Summary of the *Telchin licus licus* transcriptome.

Number of reads before pre-processing	653511
Number of bases before pre-processing	381273406
Average read length before pre-processing	583
Number of reads after pre-processing	362412
Number of bases after pre-processing	140286056
Average read length after pre-processing	387
Number of contigs	23824
Number of bases in contigs	15166298
Average contig length	633
Min. contig length	290
Max. contig length	5527
Average read coverage per contig	8.5
% contigs with at least 1 GO term	18
% contigs with an EC number	0.6
% contigs with at least 1 IPR	16
Contigs with at least 1 BLAST hit against NR	8708
Contigs with no BLAST hits	15116

doi:10.1371/journal.pone.0118231.t001

present study was deposited in the Short Read Archive of the National Center for Biotechnology Information with accession number SRR1204999.

The assembled sequences (23,824 contigs) were analyzed for similarities with known sequences against non-redundant protein database at NCBI using the BLASTx program of the BLAST suite of tools [81]. At the superior cut-off threshold for blast search set to $1e^{-5}$, 8,708 contigs (~37%) returned hits against this database. About 60% of the contigs did not show significant sequence similarity at protein-level, reinforcing the findings reported in other studies aiming to explore insect midgut transcriptome [28,82]. The remaining contig sequences (15,116) were inspected for the occurrence of ORFs and domain search querying Pfam-A to provide functional information using Transdecoder program [83]. This procedure resulted in the annotation of 582 additional contigs. Alignment of the contigs to the reference genome of *B. mori* [84] using the cross-species option in GMAP program [85] allowed to the recognition of another 605 contigs for which the most probable placement in the genome was coincident with annotated gene model in this related species. Additional search for sequence similarity was carried out using BLASTn program against the NCBI collection of nucleotides (nt), using a stringent e-value cut-off of $1e^{-10}$ to recognize sequences assembled into contigs in which content was typically associated with untranslated regions resulted in 367 matches. This disjoint conjunct of sequences based on similarities searches resulted in a total of 10,262 contigs (43%) putatively representing reliable set of genes for SGB.

Quality of the remaining ~13,000 contigs that did not fall into the previous attempts to identify their protein-coding potentialities were assessed to quality and accuracy accordingly to previously described in Materials and Methods. Fig. 1 summarizes our findings, suggesting that the sequencing of the 106,271 ESTs that generated the unannotated contigs occurred more frequently at the middle of the DNA strand towards to one of the ends (most probably the 3' end), and gave uneven profiling of the transcriptome assembly. This observed profile is very contrasting with the coverage data gathered from the ESTs that formed contig sequences that could be successfully placed on the genomic loci of a set of probable putative orthologous between SGB and the related specie *B. mori*. In this latter, the sequencing seems to occurs more frequently at the middle of the DNA strand and both ends are nearly equally represented. This analysis suggests that a bias occurred in the library prep and/or in the sequencing instrument that led to incomplete representation of this particular set of EST. Incomplete cDNA synthesis, inconsistent or poor fragmentation of the cDNA sample are known sources of biases that can be related to this observation [86]. These ESTs are almost certainly the product of poor quality sequencing for which the bioinformatics steps used previously the assembly process markedly reduce their coverage of data (mean depth of coverage 5.8x and contigs shorter than 1,200 bp). A similar discussion in a study of the transcriptome of midgut of *M. sexta* using 454 technology led to closely report in number of contigs that did not return informative similarity against known protein-coding sequences [28]. In this study the authors reported ~10,000 contigs in that condition. Accordingly to our analysis and similar findings in related literature we observe that we cannot offer an exhaustive picture of the SGB transcriptome. However we emphasize that sufficient caution was considered in our bioinformatics pipeline to establish a reliable set of sequences for downstream functional characterization at least at the same level of confidence observed in similar studies related to our sample.

A high number of contigs with unknown functions were also observed on previously sequenced insect transcriptomes [20,87,88]. However, taking into account the inaccuracies in the sequencing, we cannot completely rule out that the lack of annotation can also suggests an important number of species-specific genes, which may be useful in several studies, particularly using RNAi strategies [24]. The BLASTx hits distribution, accordingly to the adopted e-value of $1e^{-5}$, is shown in Fig. 2. To determine the coverage of our library, we grouped the contigs

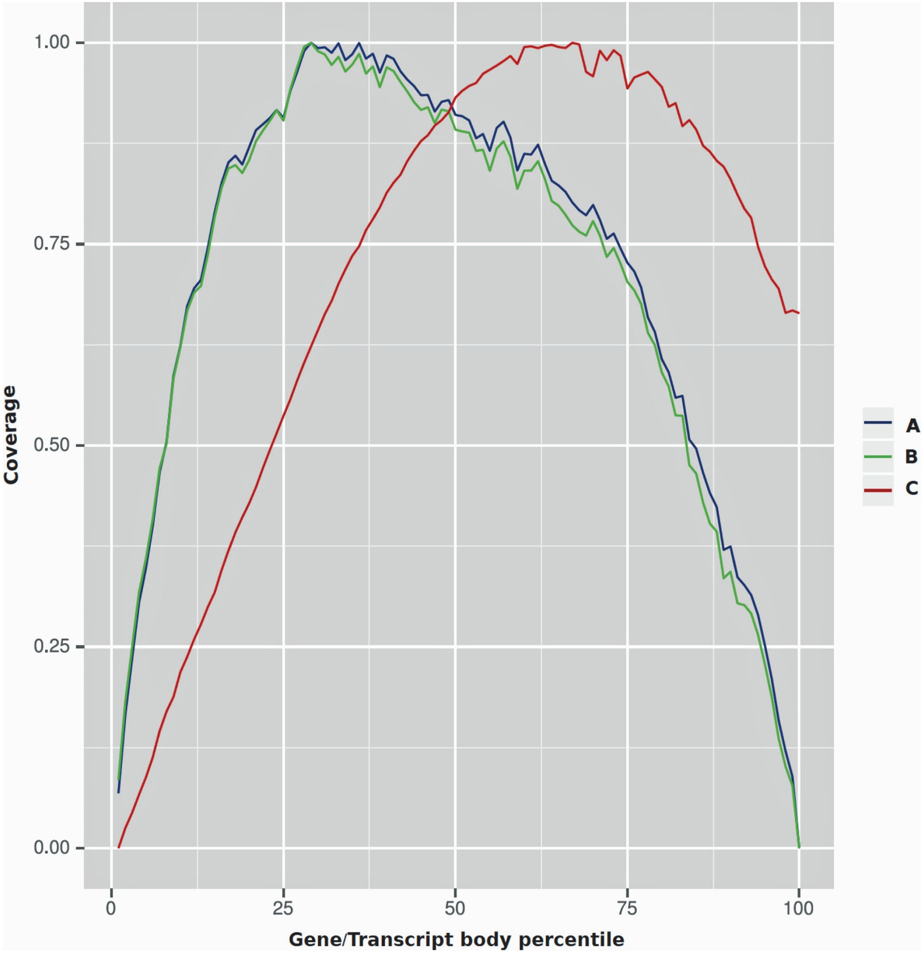


Fig 1. Coverage of data along the gene/transcript length suggests the presence of sequencing bias in the particular set of ESTs. Coverage signals were extracted from BAM files and 100 quantiles were obtained from each transcript in BED files. A) All genes for which ESTs produced alignments to the reference genome of the related specie *B*. *mori*. B) Same as before considering only genes for which the ESTs coverage of data was above 0.125 RPKM. C) Assembled contigs that did not return signals of protein-coding capacities for which ESTs produced alignments.

doi:10.1371/journal.pone.0118231.g001

accordingly to the most frequent species similarities. The highest percentage of sequence hits occurred with insect proteins, particularly Lepidoptera (30%), Hymenoptera (18%), Diptera (12%) and Coleoptera (7%). Though SGB is a lepidopteran, the high number of sequence similarities with dipterans and hymenopterans reflects mostly the influence of the large number of DNA sequences for these species in the GenBank. After comparing the top hits distribution, as expected, there was a higher percentage of similarity to protein sequences of lepidopterans (87%), especially *B. mori* (40%) and *Danaus plexippus* (33%) both with fully sequenced genomes (Fig. 3).

As the insects were collected directly from sugarcane fields and total RNA was isolated from whole body organisms, it was not possible to clean the content of the midgut, leaving the possibility for the presence of parasites and microorganisms, as well as plant tissues derived from the insect diet. Among our BLAST hits results, we observed a low number of contigs derived from species other than insects: *Branchiostoma floridae* (Lancelet or Amphioxus), *Hydra vulgaris* (Hydrozoan), *Saccoglossus kowalevskii* (Hemichordate), and *Picea sitchensis* (Seed plant).

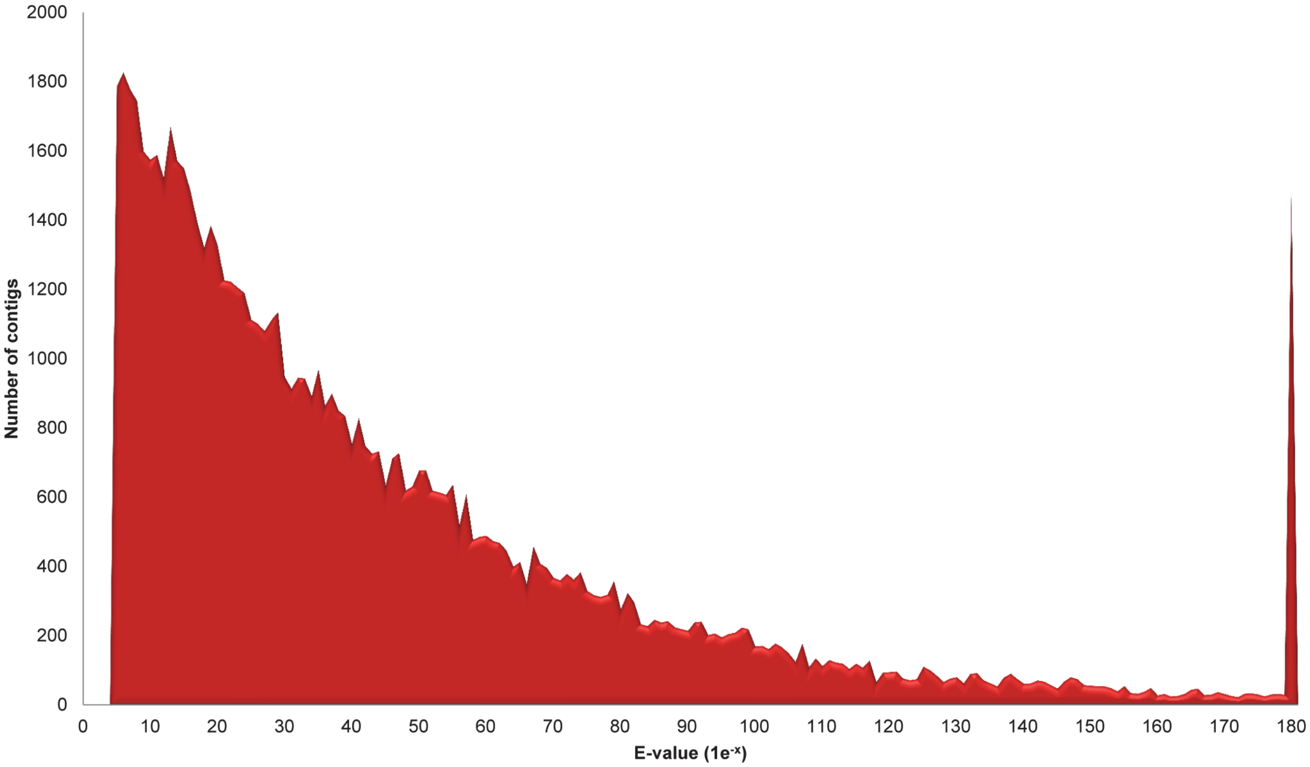


Fig 2. E-value distribution of the top BLASTx hits. Sequences with e-value equal to 0 are represented in the right peak. The cut-off used was 1e^{-5}.

doi:10.1371/journal.pone.0118231.g002

Although there is no known ecological relationship between the first three species and SGB, primarily because they have an aquatic lifestyle, a more detailed analysis of the contigs indicated that most of the sequences are associated with hypothetical proteins, possibly because the genome of those species have also been sequenced [89,90] and the lack of annotation is hampering the determination of protein function. The similarity of contigs with sequences of *P. sitchensis* could indicate contamination of our sample with plant tissues; however, all of the sequences were classified as unknown proteins. In fact, accordingly to the authors of the *P. sitchensis* sequencing project, there could be a contamination of their samples with insect cDNA since the plants were subjected to herbivory prior to RNA extraction and sequencing [91]. We searched our database for sequences of other plant species and found a few contigs with similarities to *Oryza sativa*, *Zea mays*, *Arabidopsis sp*. and *Vitis vinifera* but most of them code for transposon and retrotransposon proteins or proteins highly conserved between eukaryotes. No similarities were found after restricting the analysis of BLASTx (nr database) to *saccharum* sp sequences at the GenBank. Thus, there appears to be no significant influence of DNA contamination of SGB cDNA library.

Gene Ontology and Function Classification

Gene ontology analysis (GO) was performed to classify the functions of the predicted proteins (Fig. 4). We observed a dominance of Biological Process GO terms for metabolic (29%) and cellular (29%) processes (Fig. 4A). For Molecular Function, it was observed a high percentage of terms for catalytic activity (39%) and binding (38%) (Fig. 4B). For Cellular Component, a high percentage of GO terms were predicted for cell components (42%) (Fig. 4C). The same pattern of GO classification was observed for other insect transcriptomes and confirmed that

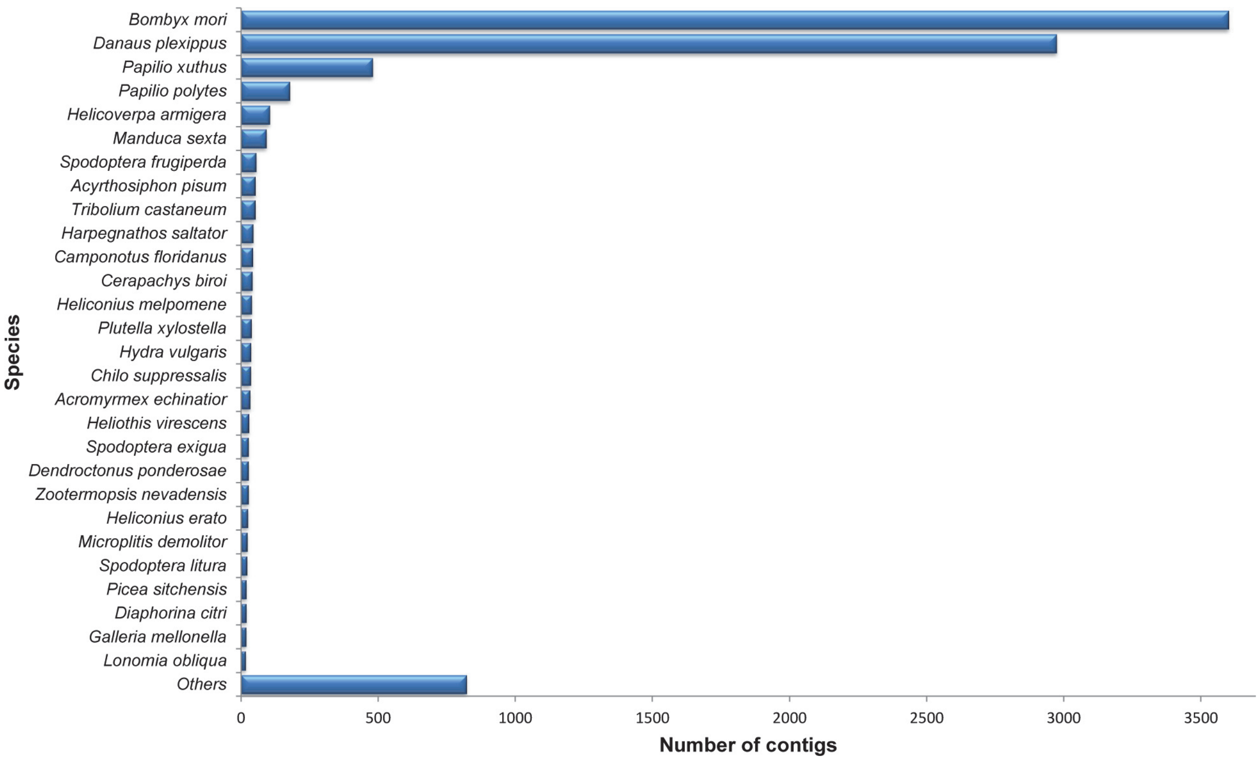


Fig 3. Species distribution of the top BLAST hits for each unique sequence. A higher similarity was observed with proteins from the lepidopteran *Bombyx mori*.

doi:10.1371/journal.pone.0118231.g003

our database is representative and consistent with other reported data [92–95]. The InterPro database was used to obtain a more detailed classification of predicted proteins. Of almost 24,000 contigs, 16% presented InterPro entries (Table 1). The top 25 InterPro hits are shown in Table 2. The most frequent identified proteins were insect cuticle proteins (292 entries), NAD (P)-binding domain (264 entries) and cytochrome P450 (252 entries). Several contigs encoding putative digestive enzymes were also observed: peptidase S1/S6; chymotrypsin/Hap (218 entries), Serine/cysteine peptidase; trypsin-like (142 entries), carboxypeptidases (86 entries), lipases (84 entries) and peptidase S1A; chymotrypsin (79 entries) were the most frequent.

To increase our knowledge of proteins possibly expressed specifically in the SGB midgut, the EST library was compared with midgut sequences of *M*. *sexta* obtained from the Insecta-Central database. Out of 13,828 *M*. *sexta* sequences, 6473 hits were achieved above the cut-off (1x10^{-3}); the most frequent InterPro hits are shown in S3 Table. Many of the contigs were classified as proteins for cell metabolism, digestion and detoxification. The presence of cuticle proteins among them is intriguing, although such characteristic has been observed in *Anopheles gambie* and *B*. *mori* and are most likely involved with immune defense response and midgut growth [96,97]. In addition, functional analyses of the BLAST hits were performed by grouping the contigs with a predicted function for digestive enzymes to estimate the number of unigenes and how many sequences from our library had no similarities to other known *M*. *sexta* genes (Table 3). Of the 120 contigs of serine protease transcripts identified in the SGB database, 96 presented BLAST hits against known midgut sequences, corresponding to 59 *M*. *sexta* unigenes. For aminopeptidase N, 16 contigs were found in our database and, interestingly, 18 contigs were obtained after cross-species similarities searches against *M*. *sexta* sequences using blat and gmap sequence alignment tools. The annotation of these two contigs could not be achieved

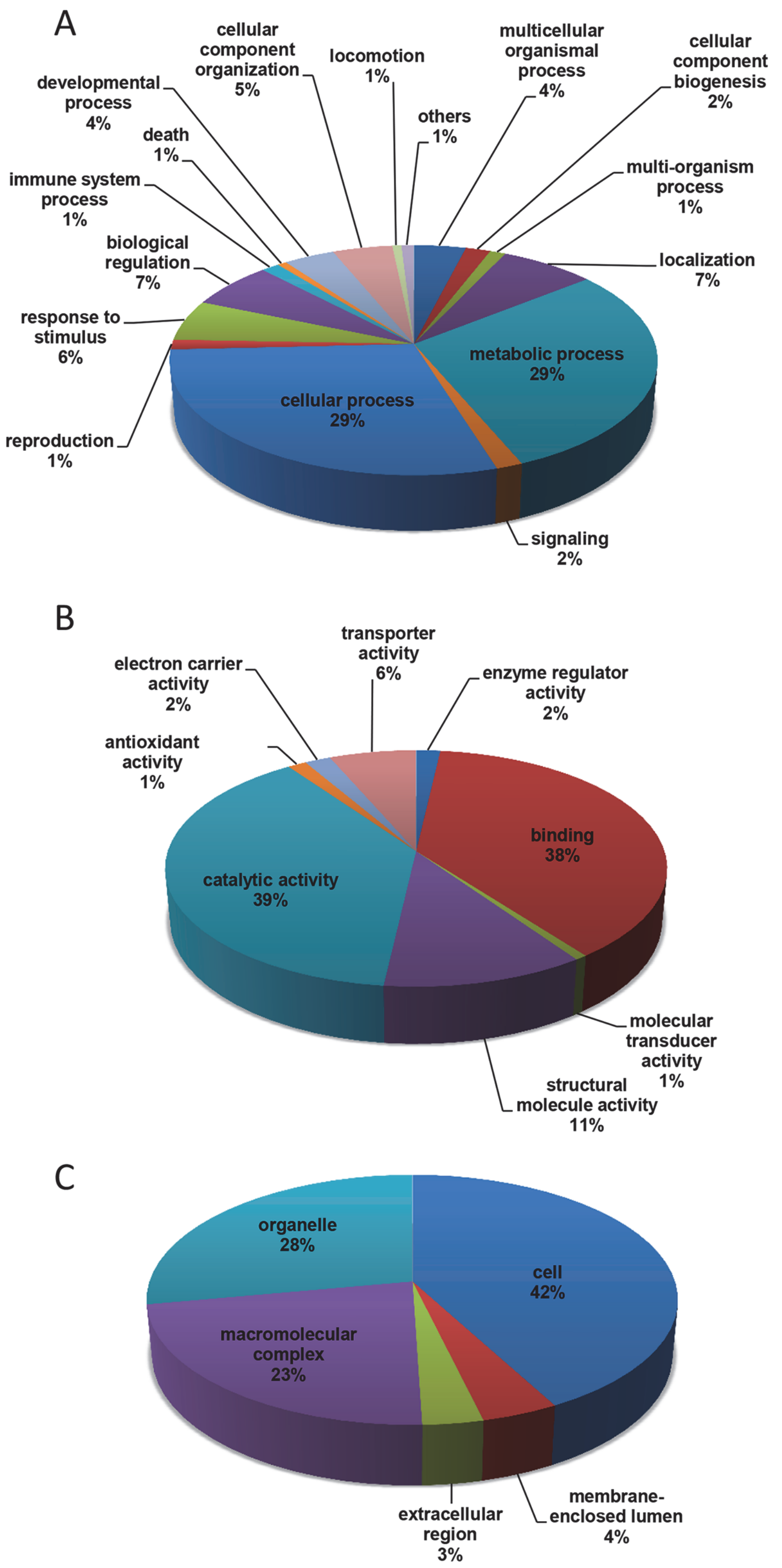


Fig 4. Gene Ontology (GO) assignments for SGB transcriptome. All contigs were classified on level 2 for **A**) Biological Process, **B**) Molecular Function and **C**) Cellular Component.

doi:10.1371/journal.pone.0118231.g004

Table 2. Summary of top 25 protein domains found in the *Telchin licus licus* transcriptome.

InterPro	Frequency	Description
IPR000618	292	Insect cuticle protein
IPR016040	264	NAD(P)-binding domain
IPR001128	252	Cytochrome P450
IPR001254	218	Peptidase S1/S6, chymotrypsin/Hap
IPR002198	209	Short-chain dehydrogenase/reductase SDR
IPR015880	206	Zinc finger, C2H2-like
IPR007087	162	Zinc finger, C2H2-type
IPR002347	153	Glucose/ribitol dehydrogenase
IPR001680	144	WD40 repeat
IPR019781	143	WD40 repeat, subgroup
IPR009003	142	Serine/cysteine peptidase, trypsin-like
IPR000215	137	Protease inhibitor I4, serpin
IPR002557	132	Chitin binding protein, peritrophin-A
IPR000504	126	RNA recognition motif, RNP-1
IPR001395	105	Aldo/keto reductase
IPR002018	92	Carboxylesterase, type B
IPR000834	86	Peptidase M14, carboxypeptidase A
IPR000734	84	Lipase
IPR001251	83	Cellular retinaldehyde-binding/triple function, C-terminal
IPR000217	81	Tubulin
IPR001314	79	Peptidase S1A, chymotrypsin
IPR016196	73	Major facilitator superfamily, general substrate transporter
IPR005055	71	Insect pheromone-binding protein A10/OS-D
IPR012677	71	Nucleotide-binding, alpha-beta plait
IPR011046	69	WD40 repeat-like-containing domain

doi:10.1371/journal.pone.0118231.t002

previously because the sequences are too short and have no conserved domains, hampering to identify protein family groups.

Additional functional, comparative and phylogenetic analysis of *de novo* transcriptome of SGB and *B. mori* genome was performed using the online tool TRAPID. Of all the groups

Table 3. Digestive enzymes found in the *Telchin licus licus* transcriptome and their correspondence with *Manduca sexta* midgut enzymes.

Classification	InterPro	Total number of contigs	Number of contigs x *M. sexta* (unigenes)
Serine protease	IPR009003/IPR001314	120	96 (59)
Cystein protease	IPR015643	19	3 (2)
Carboxypeptidase	IPR000834	31	33 (18)
Aminopeptidase	IPR001930	16	18 (11)
Dipeptidase	IPR005320/IPR001548	9	9 (3)
α-amylase	IPR006047	15	9 (4)
α-glucosidase	IPR000322	4	5 (4)
β-glucosidase	IPR001360	12	13 (13)
β-galactosidase	IPR001944	3	3 (2)
Trehalase	IPR001661	2	2 (1)
Lipase	IPR006693/IPR000734/IPR013818/IPR002331	45	39 (16)

doi:10.1371/journal.pone.0118231.t003

Table 4. Gene expansion/depletion analysis among *T. licus licus* and *B. mori* databases.

Classification	Gene family	Transcript count	
		T. licus	*B. mori*
Serine protease	1133_BGIBMGA002205	2	1
	1133_OG5_215701	4	2
	1133_OG5_149528	2	1
	1133_OG5_136800	3	1
	1133_BGIBMGA008883	2	1
	1133_OG5_174722	2	1
	1133_OG5_158561	3	1
	1133_BGIBMGA004487	2	1
	1133_OG5_158544	2	1
	1133_OG5_142367	2	1
	1133_OG5_141149	14	5
	1133_BGIBMGA004425	2	1
	1133_OG5_142493	3	1
	1133_BGIBMGA005172	2	1
	1133_OG5_152309	3	6
	1133_OG5_163947	1	6
	1133_OG5_130858	7	15
Cystein protease	1138_OG5_127800	3	1
	1138_OG5_126607	6	2
Carboxypeptidase	1134_BGIBMGA004799	3	1
	1134_OG5_128876	4	1
	1134_OG5_127925	1	2
Aminopeptidase	1136_OG5_129538	2	1
Alpha-amylase	1119_OG5_128640	7	3
Beta-galactosidase	1142_OG5_128163	2	1
Lipase	1135_OG5_130874	3	1
	1135_BGIBMGA004157	2	1
	1135_BGIBMGA011895	3	1
	1135_OG5_135981	2	4
	1135_OG5_150673	1	10

doi:10.1371/journal.pone.0118231.t004

shown in Table 4, many presented gene expansion or depletion among the gene families identified. In the serine protease group, the most notorious change occurred in family 1133_OG5_141149, in which there was an expansion of 9 genes, and in family 1133_OG5_130858, with a depletion of 8 genes. For the aminopeptidase group, there was an expansion of 1 gene in family 1136_OG5_129538. In general, there was expansion/depletion in almost all protein groups, except dipeptidase, α-glucosidase, β-glucosidase and trehalase (Table 4). Although more than one transcript could be part of the same gene, which could lead to an overrepresentation of the total number of genes, both species showed a close sequence count, indicating minimal influence by sequence mis-assembly. Changes in the number of genes are frequently found among eukaryotes and can be accounted for different mechanisms such as gene duplication, transposition and gene loss via the accumulation of mutations, leading to divergence of proteins function and adaptation. Even among closely related species, a large number of gene gain and loss can be observed. Depending on the biological functions of these genes it is reasonable to expect that such characteristic may influence the genotypic and phenotypic alterations observed among species [98]. The differences observed for SGB may explain the adaptation of the insect to a sugarcane based diet.

Serine Proteases

Insects are adapted to almost all environments and can feed on a variety of nutrient sources such as grains, stems, cellulose, flesh and blood. Their diet serves as a primary source of fats, sugars and proteins that are digested by the action of several enzymes of the gut tract [36]. The major class of proteases found in the lepidopterans gut is the serine protease, particularly trypsin (EC 3.4.21.4) and chymotrypsin (EC 3.4.21.1); both characterized by the catalytic triad His 57, Asp 102, Ser 195 (bovine chymotrypsin numbering) [99]. Trypsin cleaves preferentially protein chains on the carboxyl side of basic amino acids, such as arginine and lysine. Conversely, chymotrypsin cleaves protein chains on the carboxyl side of aromatic amino acids, such as phenylalanine and tyrosine [36]. The specificity of serine proteases is determined by a (S1) binding pocket that interacts with the side chains of amino acids at the P1-P1' site of the substrate [100]. In trypsin-like proteases the main residues that form the S1 pocket are Asp 189, Gly 216 and Gly 226. Chymotrypsin-like proteases usually contain the residues Gly/Ser 189, Gly 216 and Gly 226 [101].

In the SGB database several contigs for serine proteases were identified by InterPro analyses. Searching for those contigs presenting complete protein coding sequences, five contigs for trypsin-like and four chymotrypsin-like proteases were found and later confirmed by BLAST search against NCBI nr database. According to the MEROPS peptidase database (http://merops.sanger.ac.uk), all the sequences were classified as belonging to the SA1 family and many preserved all the characteristics of digestive serine proteases.

The sequences have a 5' UTR and 3' UTR regions and an open reading frame (ORF) of approximately 800 bp. The trypsin-like transcripts coded for predicted proteins of 259–306 amino acids with predicted isoelectric points ranging from 5.48 to 8.70 and theoretical molecular weights of 26 to 33 kDa. All the genes presented higher transcript levels in the midgut tissues, compared to the carcass (Fig. 5 A-E). The same characteristic was observed with other insect serine proteases, indicating that the SGB genes could be part of the intestinal homogenate and participate on digestion of proteins [102]. The chymotrypsin-like transcripts coded for predicted proteins of 279–299 amino acids with isoelectric points ranging from 6.67 to 8.27 and molecular weights of 29 to 32.8 kDa (S4 Table). Likewise, the chymotrypsin genes presented higher expression in the midgut (Fig. 6 A-D). The characterization of several trypsin- and chymotrypsin-like *Ostrinia nubialis* cDNAs revealed that insect digestive proteases could possess similar characteristics [103].

Sequence alignment was carried out to identify conserved sites. The trypsin-like transcripts were named Tl-TRY1—Tl-TRY5 and contain an N-terminal signal peptide. A conserved IXGG (where X stands for any amino acid) propeptide-processing site was observed on all sequences but not in Tl-TRY2, suggesting that this molecule could remain in the organism as a zymogen. The catalytic triad was conserved on sequences 1, 3 and 4 as well as the binding pocket site. For sequence Tl-TRY2, His74 was replaced by Ser, Ser230 was replaced by Glu and the binding pocket contained substitutions in all residues. Sequence Tl-TRY5 conserved only

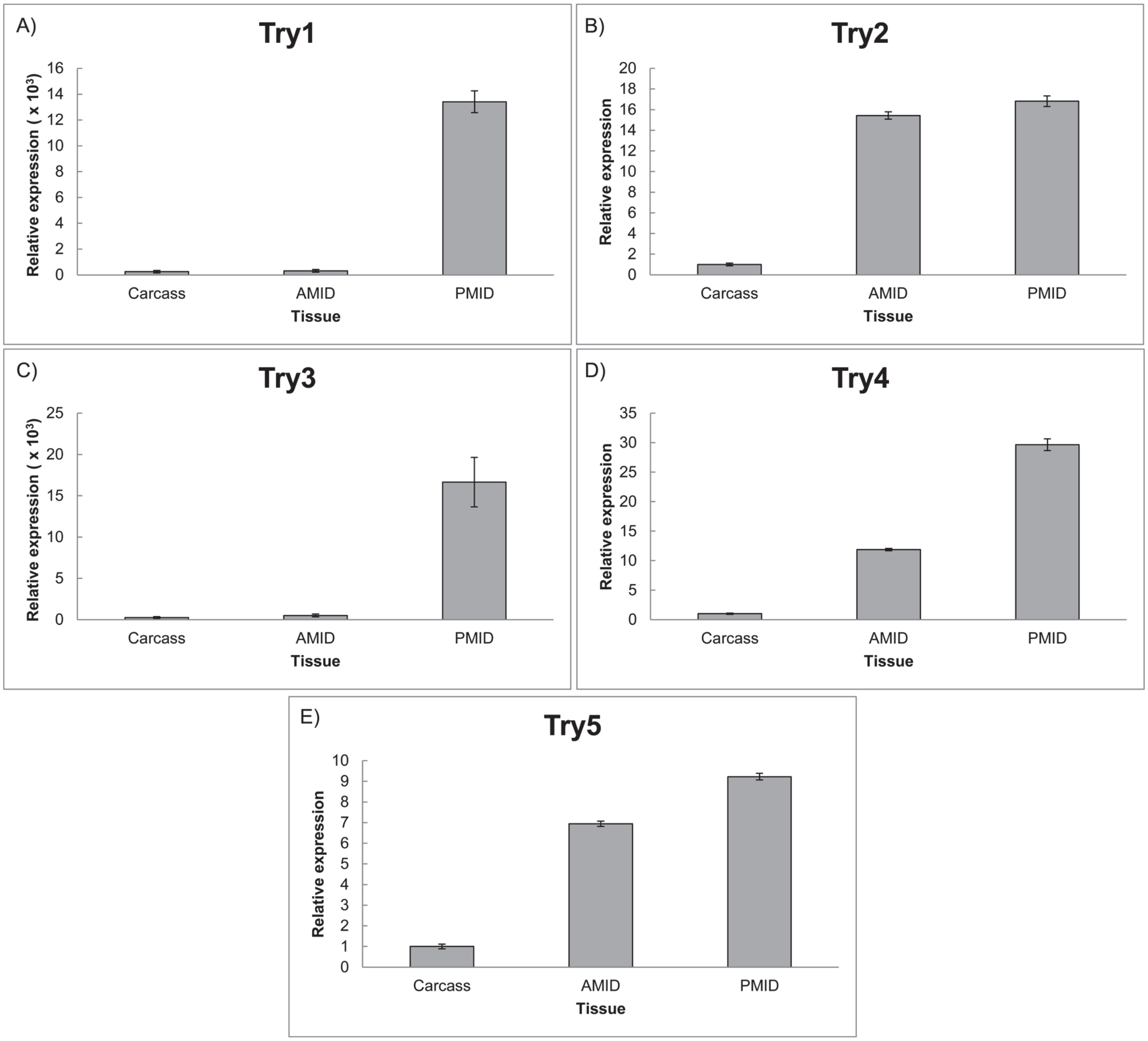


Fig 5. Real time PCR of trypsin-like genes in three different tissues of SGB. A) Trypsin 1, B) Trypsin 2, C) Trypsin 3, D) Trypsin 4 and E) Trypsin 5. AMID (Anterior midgut). PMID (Posterior midgut). Each bar represents the relative expression in comparison to the tissue that had the smaller expression value, arbitrarily designated as 1. Standard errors of technical triplicate are shown.

doi:10.1371/journal.pone.0118231.g005

Asp114, which is part of the catalytic site (Fig. 7). As a mechanism for avoiding the effects of PIs, insects change the expression of many proteases, leading to activate many digestive enzymes or to produce insensitive proteins with sequence variations that helps to prevent the binding of PIs [104,105], explaining the high number of trypsin genes in the genome and transcriptome databases. Another important characteristic observed in the insect physiology is the expression of proteases with mutations at the active site, which could indicate that these molecules are inactive [106]. Notwithstanding, such molecules have been identified in several insects, and many of them classified as serine protease homologs (SPH) that are possibly involved in innate immune response [107,108]. The amino acid substitutions observed on sequences 2, and 5 are easily found in serine protease sequences at the GenBank (S1 Fig. and S2

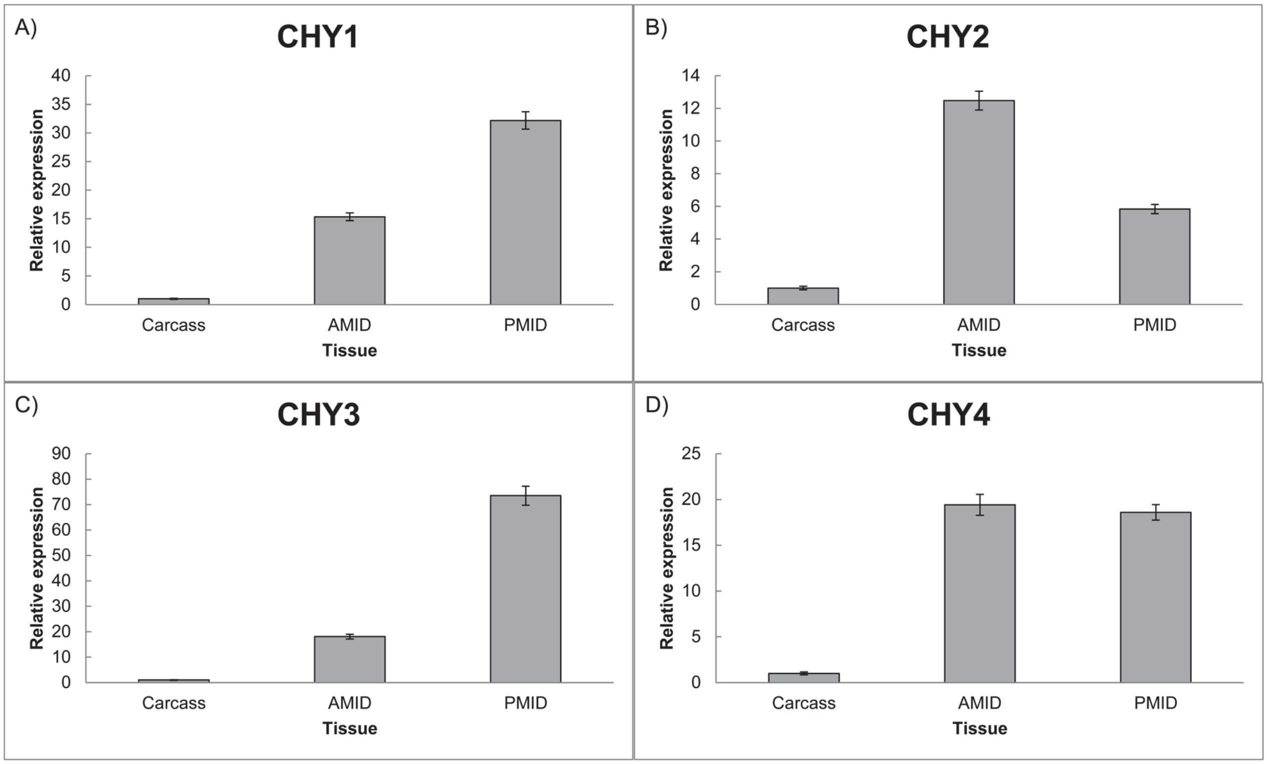


Fig 6. Real time PCR of chymotrypsin-like genes in three different tissues of SGB. A) Chymotrypsin 1, B) Chymotrypsin 2, C) Chymotrypsin 3 and D) Chymotrypsin 4. AMID (Anterior midgut). PMID (Posterior midgut). Each bar represents the relative expression in comparison to the tissue that had the smaller expression value, arbitrarily designated as 1. Standard errors of technical triplicate are shown.

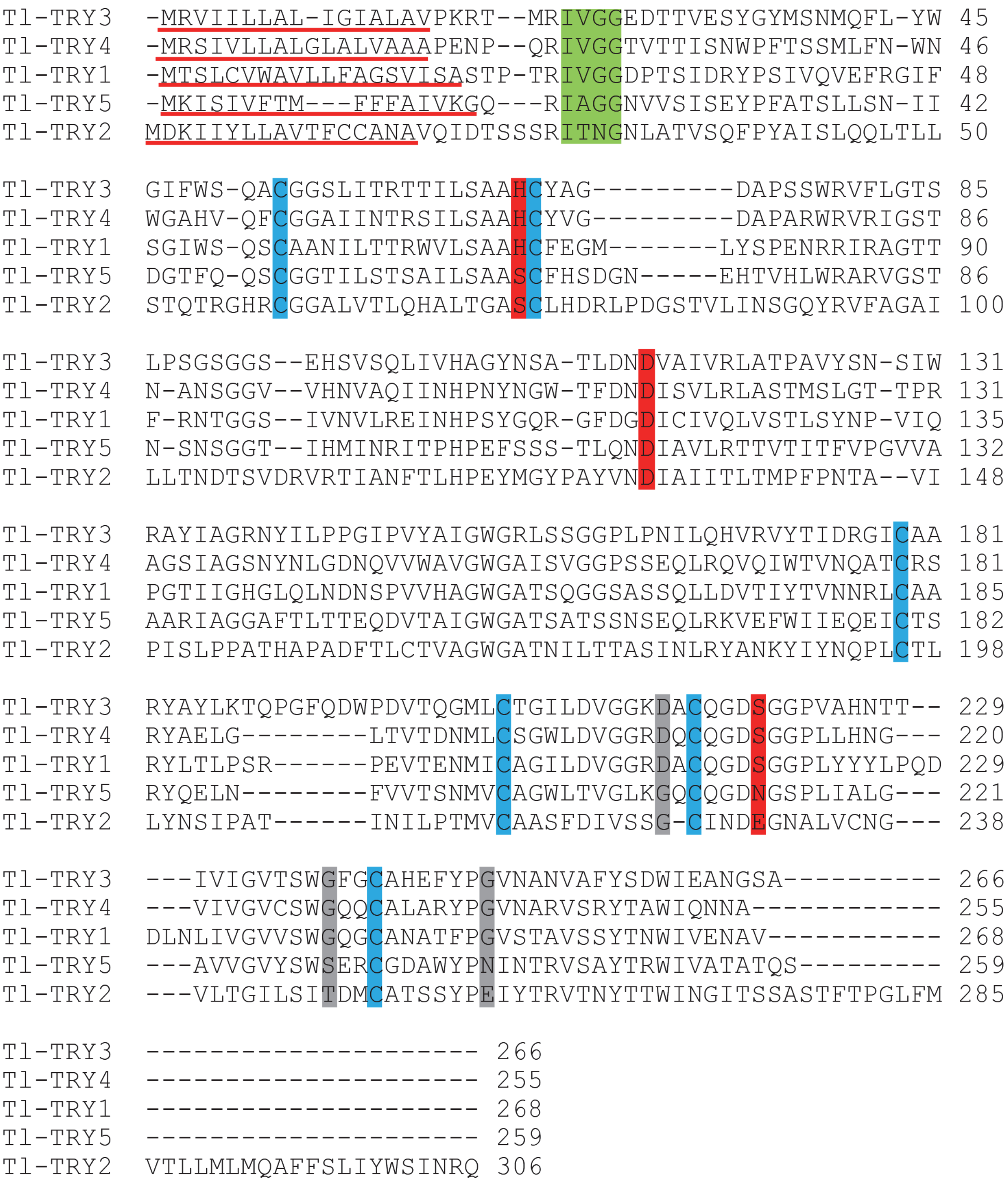
doi:10.1371/journal.pone.0118231.g006

Fig.) and most likely indicate that such mechanism could be conserved among different insect species.

Chymotrypsin-like transcripts were named Tl-CHY1—Tl-CHY4 and they preserved the N-terminal signal peptide. A conserved IXGG site is observed in all sequences except Tl-CHY3. The catalytic triad was conserved in all sequences. Substrate binding sites contained amino acid substitutions at different positions (Fig. 8). The differences in the binding pocket sites indicate that these enzymes have diverse activities because chymotrypsin sequences show more variations at the S1 pocket, which could increase protein flexibility and result in differential substrate recognition [41]. Although SGB serine protease activity has yet to be experimentally investigated, the identification and characterization of different genes will help researches to understand the enzymatic arsenal used by the insect to process the nutritive content of a sugarcane-based diet and to study protease involvement in the Cry toxin activation/deactivation mechanism. Such information will be important to guide the development of alternative strategies for pest control, such as genetic transformation of Elite sugarcane events by expressing Cry toxins or PIs, targeting specifically the activity of digestive enzymes [109,110].

Aminopeptidases

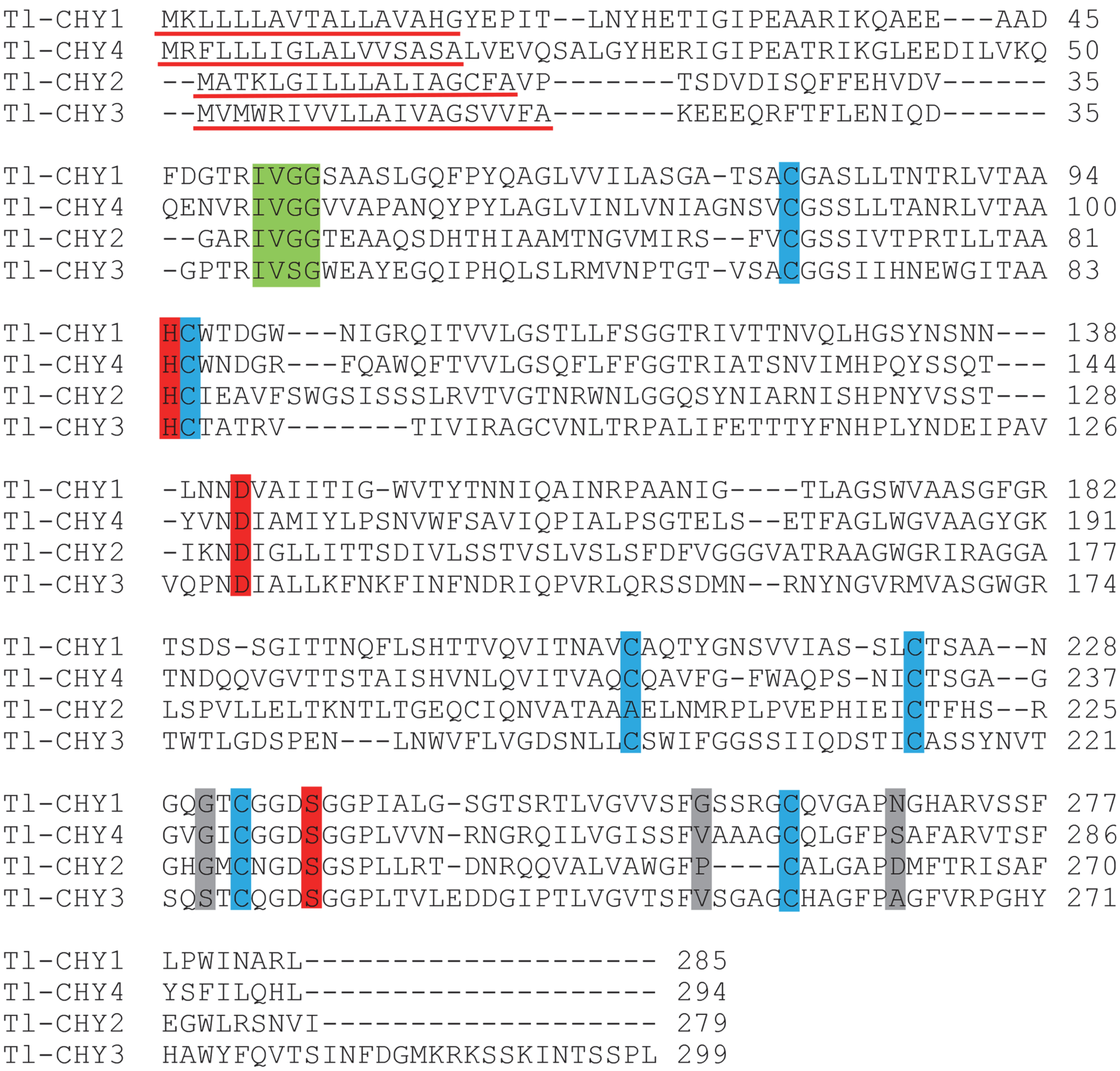
Midgut aminopeptidases N (APNs—EC 3.4.11.2) are enzymes that participate in the digestion of proteins by cleaving neutral amino acids from the N-terminus of polypeptides [36]. APNs are classified as belonging to the M1 family and have a zinc binding motif HEXXH (where X stands for any amino acid), followed by a conserved glutamic acid 24-amino acids downstream from the first histidine. While the histidines and the last glutamic acid form the zinc binding



```
Tl-TRY3   -MRVIILLAL-IGIALAVPKRT--MRIVGGEDTTVESYGYMSNMQFL-YW 45
Tl-TRY4   -MRSIVLLALGLALVAAAPENP--QRIVGGTVTTISNWPFTSSMLFN-WN 46
Tl-TRY1   -MTSLCVWAVLLFAGSVISASTP-TRIVGGDPTSIDRYPSIVQVEFRGIF 48
Tl-TRY5   -MKISIVFTM---FFFAIVKGQ---RIAGGNVVSISEYPFATSLLSN-II 42
Tl-TRY2   MDKIIYLLAVTFCCANAVQIDTSSSRITNGNLATVSQFPYAISLQQLTLL 50

Tl-TRY3   GIFWS-QACGGSLITRTTILSAAHCYAG---------DAPSSWRVFLGTS 85
Tl-TRY4   WGAHV-QFCGGAIINTRSILSAAHCYVG---------DAPARWRVRIGST 86
Tl-TRY1   SGIWS-QSCAANILTTRWVLSAAHCFEGM-------LYSPENRRIRAGTT 90
Tl-TRY5   DGTFQ-QSCGGTILSTSAILSAASCFHSDGN-----EHTVHLWRARVGST 86
Tl-TRY2   STQTRGHRCGGALVTLQHALTGASCLHDRLPDGSTVLINSGQYRVFAGAI 100

Tl-TRY3   LPSGSGGS--EHSVSQLIVHAGYNSA-TLDNDVAIVRLATPAVYSN-SIW 131
Tl-TRY4   N-ANSGGV--VHNVAQIINHPNYNGW-TFDNDISVLRLASTMSLGT-TPR 131
Tl-TRY1   F-RNTGGS--IVNVLREINHPSYGQR-GFDGDICIVQLVSTLSYNP-VIQ 135
Tl-TRY5   N-SNSGGT--IHMINRITPHPEFSSS-TLQNDIAVLRTTVTITFVPGVVA 132
Tl-TRY2   LLTNDTSVDRVRTIANFTLHPEYMGYPAYVNDIAIITLTMPFPNTA--VI 148

Tl-TRY3   RAYIAGRNYILPPGIPVYAIGWGRLSSGGPLPNILQHVRVYTIDRGICAA 181
Tl-TRY4   AGSIAGSNYNLGDNQVVWAVGWGAISVGGPSSEQLRQVQIWTVNQATCRS 181
Tl-TRY1   PGTIIGHGLQLNDNSPVVHAGWGATSQGGSASSQLLDVTIYTVNNRLCAA 185
Tl-TRY5   AARIAGGAFTLTTEQDVTAIGWGATSATSSNSEQLRKVEFWIIEQEICTS 182
Tl-TRY2   PISLPPATHAPADFTLCTVAGWGATNILTTASINLRYANKYIYNQPLCTL 198

Tl-TRY3   RYAYLKTQPGFQDWPDVTQGMLCTGILDVGGKDACQGDSGGPVAHNTT-- 229
Tl-TRY4   RYAELG--------LTVTDNMLCSGWLDVGGRDQCQGDSGGPLLHNG--- 220
Tl-TRY1   RYLTLPSR------PEVTENMICAGILDVGGRDACQGDSGGPLYYYLPQD 229
Tl-TRY5   RYQELN--------FVVTSNMVCAGWLTVGLKGQCQGDNGSPLIALG--- 221
Tl-TRY2   LYNSIPAT------INILPTMVCAASFDIVSSG-CINDEGNALVCNG--- 238

Tl-TRY3   ---IVIGVTSWGFGCAHEFYPGVNANVAFYSDWIEANGSA---------- 266
Tl-TRY4   ---VIVGVCSWGQQCALARYPGVNARVSRYTAWIQNNA------------ 255
Tl-TRY1   DLNLIVGVVSWGQGCANATFPGVSTAVSSYTNWIVENAV----------- 268
Tl-TRY5   ---AVVGVYSWSERCGDAWYPNINTRVSAYTRWIVATATQS--------- 259
Tl-TRY2   ---VLTGILSITDMCATSSYPEIYTRVTNYTTWINGITSSASTFTPGLFM 285

Tl-TRY3   --------------------- 266
Tl-TRY4   --------------------- 255
Tl-TRY1   --------------------- 268
Tl-TRY5   --------------------- 259
Tl-TRY2   VTLLMLMQAFFSLIYWSINRQ 306
```

Fig 7. ClustalW2 alignment of *T. licus licus* trypsin-like proteins. Many sequences preserved the characteristics of digestive serine proteases. The signal peptide is shown with a red underline. The cleavage site is shown in a green box. Red boxes indicate the active site residues. Grey boxes indicate the substrate binding region and cysteines that are possibly involved with disulfide bonds are shown in blue.

doi:10.1371/journal.pone.0118231.g007



```
Tl-CHY1  MKLLLLAVTALLAVAHGYEPIT--LNYHETIGIPEAARIKQAEE---AAD 45
Tl-CHY4  MRFLLLIGLALVVSASALVEVQSALGYHERIGIPEATRIKGLEEDILVKQ 50
Tl-CHY2  --MATKLGILLLALIAGCFAVP-------TSDVDISQFFEHVDV------ 35
Tl-CHY3  --MVMWRIVVLLAIVAGSVVFA-------KEEEQRFTFLENIQD------ 35

Tl-CHY1  FDGTRIVGGSAASLGQFPYQAGLVVILASGA-TSACGASLLTNTRLVTAA 94
Tl-CHY4  QENVRIVGGVVAPANQYPYLAGLVINLVNIAGNSVCGSSLLTANRLVTAA 100
Tl-CHY2  --GARIVGGTEAAQSDHTHIAAMTNGVMIRS--FVCGSSIVTPRTLLTAA 81
Tl-CHY3  -GPTRIVSGWEAYEGQIPHQLSLRMVNPTGT-VSACGGSIIHNEWGITAA 83

Tl-CHY1  HCWTDGW---NIGRQITVVLGSTLLFSGGTRIVTTNVQLHGSYNSNN--- 138
Tl-CHY4  HCWNDGR---FQAWQFTVVLGSQFLFFGGTRIATSNVIMHPQYSSQT--- 144
Tl-CHY2  HCIEAVFSWGSISSSLRVTVGTNRWNLGGQSYNIARNISHPNYVSST--- 128
Tl-CHY3  HCTATRV-------TIVIRAGCVNLTRPALIFETTTYFNHPLYNDEIPAV 126

Tl-CHY1  -LNNDVAIITIG-WVTYTNNIQAINRPAANIG----TLAGSWVAASGFGR 182
Tl-CHY4  -YVNDIAMIYLPSNVWFSAVIQPIALPSGTELS--ETFAGLWGVAAGYGK 191
Tl-CHY2  -IKNDIGLLITTSDIVLSSTVSLVSLSFDFVGGGVATRAAGWGRIRAGGA 177
Tl-CHY3  VQPNDIALLKFNKFINFNDRIQPVRLQRSSDMN--RNYNGVRMVASGWGR 174

Tl-CHY1  TSDS-SGITTNQFLSHTTVQVITNAVCAQTYGNSVVIAS-SLCTSAA--N 228
Tl-CHY4  TNDQQVGVTTSTAISHVNLQVITVAQCQAVFG-FWAQPS-NICTSGA--G 237
Tl-CHY2  LSPVLLELTKNTLTGEQCIQNVATAAAELNMRPLPVEPHIEICTFHS--R 225
Tl-CHY3  TWTLGDSPEN---LNWVFLVGDSNLLCSWIFGGSSIIQDSTICASSYNVT 221

Tl-CHY1  GQGTCGGDSGGPIALG-SGTSRTLVGVVSFGSSRGCQVGAPNGHARVSSF 277
Tl-CHY4  GVGICGGDSGGPLVVN-RNGRQILVGISSFVAAAGCQLGFPSAFARVTSF 286
Tl-CHY2  GHGMCNGDSGSPLLRT-DNRQQVALVAWGFP----CALGAPDMFTRISAF 270
Tl-CHY3  SQSTCQGDSGGPLTVLEDDGIPTLVGVTSFVSGAGCHAGFPAGFVRPGHY 271

Tl-CHY1  LPWINARL-------------------- 285
Tl-CHY4  YSFILQHL-------------------- 294
Tl-CHY2  EGWLRSNVI------------------- 279
Tl-CHY3  HAWYFQVTSINFDGMKRKSSKINTSSPL 299
```

Fig 8. ClustalW2 alignment of *T. licus licus* chymotrypsin-like proteins. Many sequences preserved the characteristics of digestive serine proteases. The signal peptide is shown with a red underline. The cleavage site is shown in green box. Red boxes indicate the active site residues. Grey boxes indicate the substrate binding region and cysteines that are possibly involved with disulfide bonds are shown in blue.

doi:10.1371/journal.pone.0118231.g008

region, the first glutamic acid participate in the catalytic process [111]. A highly conserved GAMEN motif is also found upstream from the zinc binding motif and is believed to be a part of the active site [112]. In insects, APNs have a cleavable N-terminal signal peptide that directs the protein to the outer surface of the cell membrane where they are attached through a glycosylphosphatidylinositol (GPI) anchor located at the C-terminus [113].

Several lepidopteran APNs have been experimentally shown to bind different classes of Bt δ-endotoxins [31]. According to the most recent pore formation model, Cry1A toxins, once activated in the midgut of a susceptible insect, participate in a series of binding events with

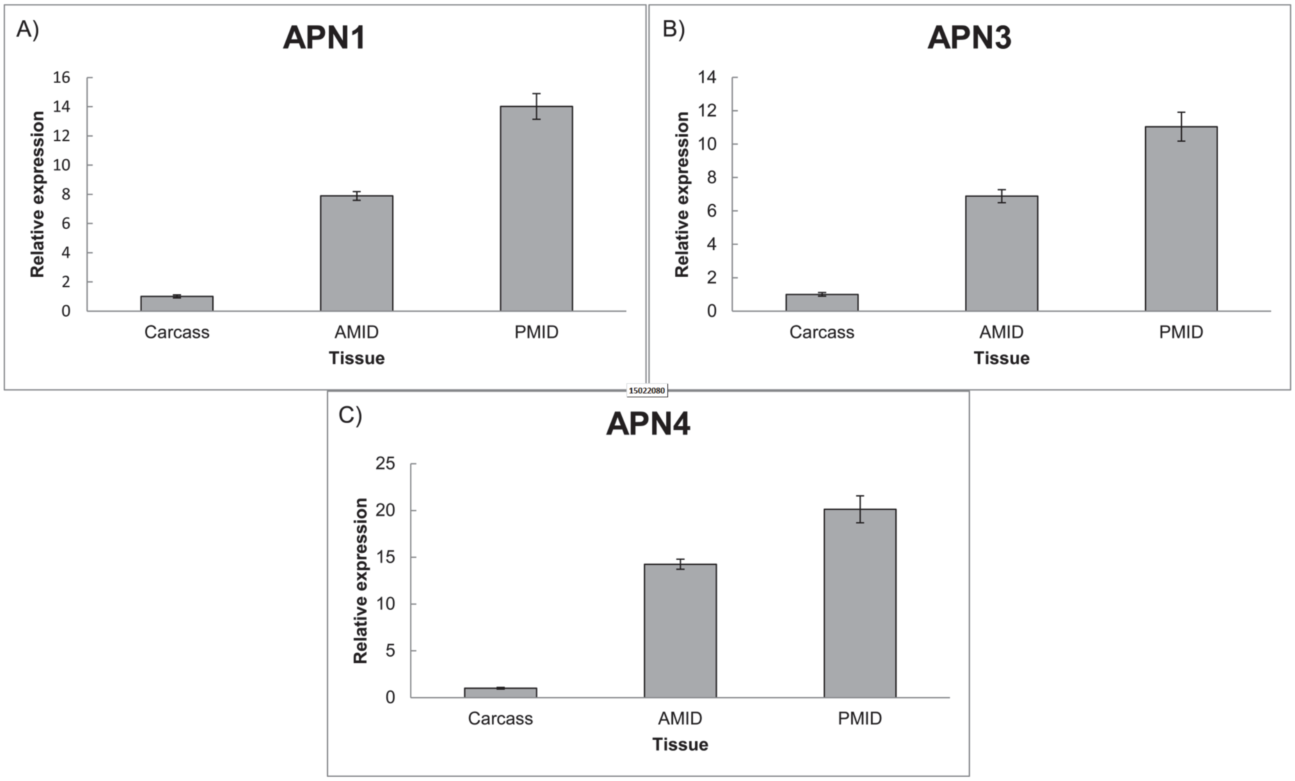


Fig 9. Real time PCR of APN genes in three different tissues of SGB. A) Aminopeptidase 1, B) Aminopeptidase 2 and C) Aminopeptidase 3. AMID (Anterior midgut). PMID (Posterior midgut). Each bar represents the relative expression in comparison to the tissue that had the smaller expression value, arbitrarily designated as 1. Standard errors of technical triplicate are shown.

doi:10.1371/journal.pone.0118231.g009

protein receptors present in the intestinal epithelium. In *M*. *sexta*, where the mode of action is better characterized, the first interaction consists of weak binding of a monomeric toxin to an APN or ALP receptor, allowing its recognition by Cadherin receptors. The protein-protein interactions induce an oligomer formation of Cry molecules, which binds again to the APN and is introduced into the plasma membrane forming a pore that causes osmotic lysis [114]. Other lepidopteran APNs were shown to participate in the development of resistance to insecticides [115] suggesting that such molecules are important targets for insect population management.

In the SGB database, a total of 18 APN contigs were identified after a similarity search against *M*. *sexta* cDNA sequences. Two contigs are actually two complete protein coding sequences and a third contig lacked the N-terminal region that was solved by 5'-RACE (S3 Fig.). Real time PCR experiments showed that APN transcript levels were higher in midgut tissues than the carcass (Fig. 9 A-C).The protein sequences were named TlAPN1, TlAPN3 and TlAPN4, according to the phylogenetic relationships of lepidopteran APNs. All sequences coded for proteins of approximately 1000 amino acids. A glycosylphosphatidylinositol anchoring signal, as well as an N-terminal signal peptide of 20 amino acids were predicted for all proteins. Sequence alignment confirmed that the HEXXH$(X)_{18}$E and GAMEN motifs are also conserved (S4 Fig.). The percentage of identity among the APN proteins ranged from 34% to 44% (S5 Table).

Glycosylation sites were predicted for different amino acid residues over the entire protein sequences. Four N-glycosylations and four O-glycosylations were predicted for TlAPN1, whereas six N-glycosylations and 18 O-glycosylations where predicted for TlAPN3. Moreover, six N-glycosylations and two O-glycosylations were predicted for TlAPN4 (S6 Table). It has been shown in previous reports that Cry1Ac toxins are capable of interacting with N-acetyl-galactosamine (GalNAc). In fact, competitive assays demonstrated that GalNAc binding to the toxin impairs the interaction with the plasma membrane receptors [116], which could reflect on a reduced activity. According to Sangadala and coworkers (2001) [117], there is a proportion of 4% carbohydrate in GPI cleaved *Manduca sexta* aminopeptidase 1 (MsAPN1), which presents a molar ratio of approximately 6: 10: 7: 3 for, respectively, GalNAc/GlcNAc/Man/Fuc. Glycosylation prediction sites indicate the presence of four possible N-glycosylations and 13 O-glycosylations in MsAPN1 [118]. A more detailed analysis of the N-linked oligosaccharides of MsAPN1 through mass spectrometry has revealed that 3 of the 4 N-glycosylations are occupied with highly fucosylated N-glycans ($Hex_3HexNAc_3Fuc_3$), while the remaining site is occupied by a paucimannosidic N-glycan ($Man_3GlcNAc_2$) [119]. No GalNAc was observed in any of the major N-glycans on MsAPN1, suggesting that 5 or 6 of the O-glycosylation sites are occupied with simple (GalNAc-peptide) type O-glycans [118,119]. It is believed that the presence of GalNAc in the C-terminus and its proximity to the plasma membrane could increase the affinity and specificity of binding to Cry1Ac toxin [120].

To better characterize SGB APNs, a phylogenetic analysis was performed to discern evolutionary relationships among representative APNs of several lepidopteran species. The sequences were fully distributed among eight phylogenetic classes, unlike the five classes previously identified when there were not much APN sequences at the GenBank [121]. Protein sequences of SGB were grouped in classes, 1, 3 and 4 (Fig. 10). Many of these proteins were reported as possible participants in the Cry mechanism of action, either by direct interaction with the toxin or by changing its expression levels in response to Bt infection. More than one APN for the same species was grouped among the eight classes, which could indicate that the protein belongs to a group of duplicated APN genes [122] or was derived from different tissues rather than just the midgut [123]. Class 1 presented proteins that were previously reported as Cry toxin receptors [31] as well as proteins identified in recent publications. Three *Helicoverpa armigera* proteins were included in the analysis. Two of them (accession numbers: ACC68682 and ACC68683) are associated with the accumulation of mutations that led to the development of resistance to Cry1Ac [124] and the third sequence (accession number: AAN75693) represents a recombinantly expressed protein that interacts differently with Cry1A toxins [125]. This group also included the *Ostrinia nubialis* sequence (accession number: AEO12690), which gene expression was substantially decreased in Cry-resistant insects [126] and *Diatraea saccharalis* APN1, in which silencing trough RNAi increased insect survival to Cry1Ab [59]. Class 2 included a *D. saccharalis* APN2 that was involved with survival of the insect to Cry1Ab survival. Class 3 presented a *D. saccharalis* APN3 as well as an *O. nubialis* APN (AEO12696), both of which were involved with Cry1Ab resistance. Class 4 added sequences that had not been observed in previously reported phylogenetic trees, including *Achaea Janata* APN, which binds to Cry1A toxins [127]. Class 5 was composed of eight proteins, but only one is known to bind to Cry toxins [128]. Class 6 was a group first observed by Crava and coworkers (2010) and included APNs of two insect species [129]. In this work, the analysis extended the number of APN proteins to six species, including a *Trichoplusia ni* APN6 that was significantly up-regulated in greenhouse-evolved resistant insects [60]. Classes 7 and 8 are composed of a few insect species but lacked any reported Cry toxin binding information. The features observed between SGB and other lepidopteran APNs, associated with the susceptibility of the insect to Cry toxins, indicate that some of these proteins could act as toxin receptors, raising the possibility of using such information to develop Cry toxins with increased activity towards specific targets.

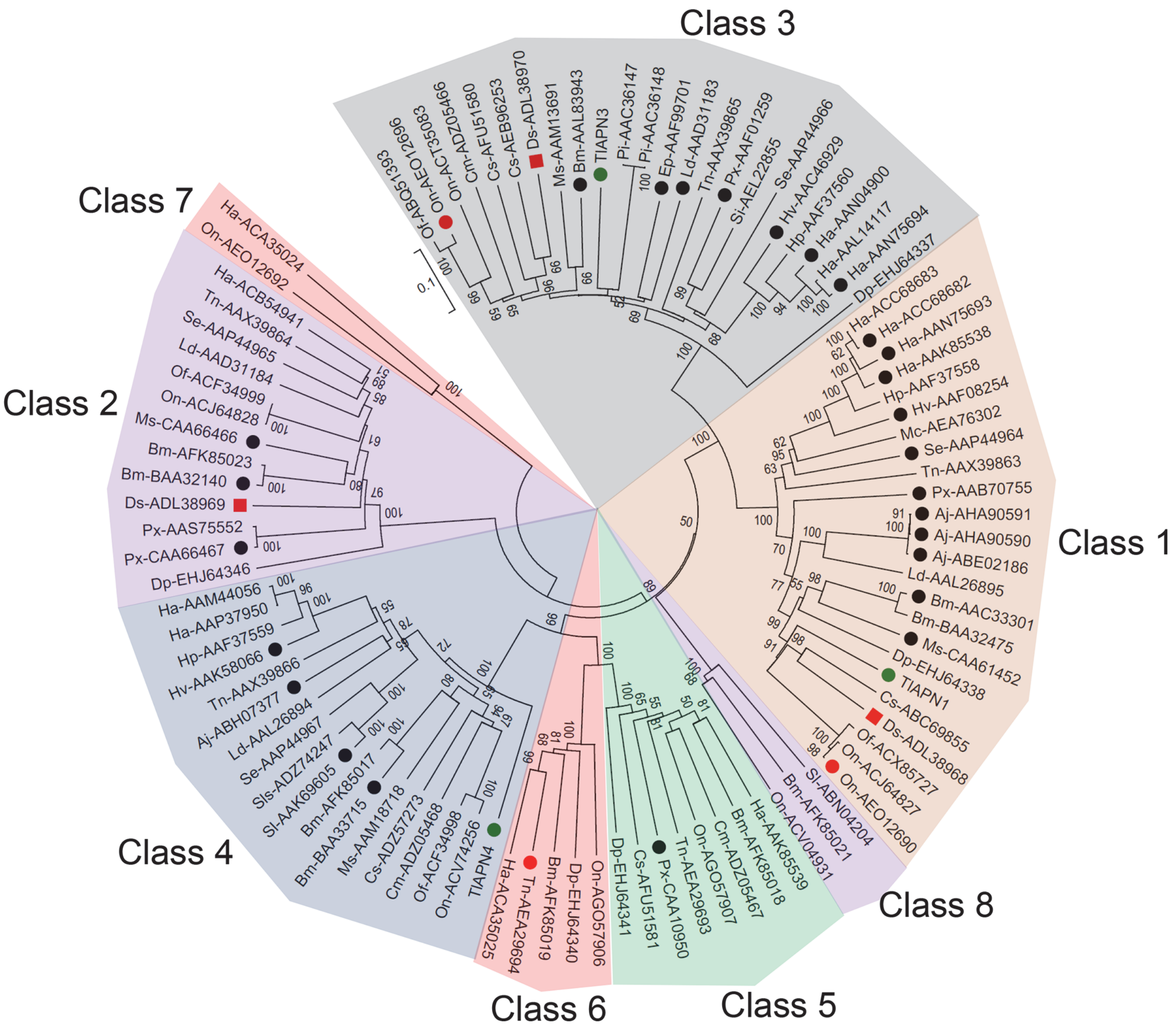


Fig 10. Phylogenetic analysis of representative lepidopteran APNs. The proteins of *T. licus licus* are shown in green. Black spots indicate the APNs that have been reported as putative Cry toxin receptors. Red spots indicate the APNs which gene expression changed after Bt infection. Red squares indicate the APNs which silencing induced resistance of the insect to Cry1Ab. GenBank accession number is shown for each protein. Species abbreviations: Ha, *Helicoverpa armigera*; Hp, *Helicoverpa punctigera*; Hv, *Heliothis virescens*; Px, *Plutella xylostella*; Se, *Spodoptera exigua*; Tn, *Trichoplusia n*i; Bm, *Bombyx mori*; Ms, *Manduca sexta*; Ld, *Lymantria dispar*; Ep, *Ephiphyas postvittana*; Pi, *Plodia interpunctella*; Sl, *Spodoptera litura*; Cs, *Chilo suppressalis*; Sls, *Spodoptera litoralis*; Cm, *Cnaphalocrocis medinalis*; Ds, *Diatraea saccharalis*; Si, *Sesamia inferens*; Dp, *Danaus plexippus*; Os, *Ostrinia nubilalis*; Mc, *Mamestra configurata*; Aj, *Achaea janata*; Of, *Ostrinia furnacalis*.

doi:10.1371/journal.pone.0118231.g010

Conclusions

In this work we report on the construction of a database of sugarcane giant borer (*T. licus licus*) by pyrosequencing of whole body mRNAs from several life stages. Before the data here presented, only 60 mitochondrial DNA sequences were deposited at the GenBank for this subspecies. As such, we believe our reported data will contribute significantly to advance future research on this organism. We have started characterizing many serine proteases and aminopeptidases, aiming specifically to study midgut enzymes that could be used in pest management assays. Next, we will characterize the activity of the serine proteases and determine the pattern of expression of these genes during the insect life cycle. To unravel the role of Cry toxin susceptibility we identified putative receptors that will be tested for their ability to bind to several Cry toxins. In addition, we intend to identify and validate genes through RNAi, a study which also will contribute to the better understanding of the insect's developmental biology.

Supporting Information

S1 Fig. ClustalW2 alignment of SGB trypsin-like Tl-TRY2 contig. BLASTp results show that sequence variation on the substrate binding site is frequently found at the GenBank. The cleavage site is shown in green. Red boxes indicate the active site. Gray boxes indicate the substrate binding region and blue boxes show the cysteins that are most likely involved with disulfide bonds. Abbreviations: Mc, *Mamestra configurata*; Ha, *Helicoverpa armigera*; Dp, *Danaus plexippus*. GenBank accession numbers are indicated.
(TIFF)

S2 Fig. ClustalW2 alignment of SGB trypsin-like Tl-TRY5 contig. BLASTp results show that sequence variation on the substrate binding site is frequently found at the GenBank. The cleavage site is shown in green. Red boxes indicate the active site. Gray boxes indicate the substrate binding region and blue boxes show the cysteins that are most likely involved with disulfide bonds. Abbreviations: On, *Ostrinia nubialis*; Cs, *Chilo suppressalis*; Ms, *Manduca sexta* Of, *Ostrinia furnacalis*; Dp, *Danaus plexippus*. GenBank accession numbers are indicated.
(TIFF)

S3 Fig. Agarose gel showing the 750 bp DNA fragment that was sequenced to complete the TlAPN1 N-terminus region. 1) 1Kb DNA Plus ladder (Invitrogen Life Sciences). 2)
PCR product.
(TIF)

S4 Fig. ClustalW2 alignment of *Telchin licus licus* APNs. Signal peptide is indicated with a red underline. Blue and green boxes show the active site of the protein. Red boxes indicate the predicted GPI anchoring site.
(TIF)

S1 Table. Primer sequences for amplification of TlAPN1 N-terminus.
(DOCX)

S2 Table. Primer sequences for qPCR of protease contigs.
(DOCX)

S3 Table. Top protein domains found in *Telchin licus licus* transcriptome after BLASTx against *Manduca sexta* midgut proteins.
(DOCX)

S4 Table. Theoretical physico-chemical parameters of serine proteases identified in SGB transcriptome.
(DOCX)

S5 Table. Amino acid sequence identity among SGB APNs.
(DOCX)

S6 Table. Glycosylation sites predicted for SGB APNS.
(DOCX)

Author Contributions

Conceived and designed the experiments: FCAF AAPF GJPJ MFGS. Performed the experiments: FCAF AAPF LLPM. Analyzed the data: FCAF AAPF LLPM RRC JDASJ RCT OBSJ GJPJ. Contributed reagents/materials/analysis tools: RCT OBSJ GJPJ LABG MCMS MFGS. Wrote the paper: FCAF OBSJ MCMS MFGS.

References

1. White WH, Viator RP, Dufrene EO, Dalley CD, Richard EP Jr, et al. (2008) Re-evaluation of sugarcane borer (Lepidoptera: Crambidae) bioeconomics in Louisiana. Crop Protection 27: 1256–1261.
2. Oliveira CM, Auad AM, Mendes SM, Frizzas MR (2014) Crop losses and the economic impact of insect pests on Brazilian agriculture. Crop Protection 56: 50–54.
3. Pimentel D (2009) Environmental and Economic Costs of the Application of Pesticides Primarily in the United States. In: Peshin R, Dhawan A, editors. Integrated Pest Management: Innovation-Development Process: Springer Netherlands. pp. 89–111.
4. Chrisman JdR, Koifman S, de Novaes Sarcinelli P, Moreira JC, Koifman RJ, et al. (2009) Pesticide sales and adult male cancer mortality in Brazil. International Journal of Hygiene and Environmental Health 212: 310–321. doi: 10.1016/j.ijheh.2008.07.006 PMID: 18838335
5. Nayak P, Basu D, Das S, Basu A, Ghosh D, et al. (1997) Transgenic elite indica rice plants expressing CryIAc ∂-endotoxin of *Bacillus thuringiensis* are resistant against yellow stem borer (*Scirpophaga incertulas*). Proceedings of the National Academy of Sciences 94: 2111–2116. PMID: 9122157
6. Koziel MG, Beland GL, Bowman C, Carozzi NB, Crenshaw R, et al. (1993) Field Performance of Elite Transgenic Maize Plants Expressing an Insecticidal Protein Derived from *Bacillus thuringiensis*. Nature Biotechnology 11: 194–200.
7. Srikanth J, Subramonian N, Premachandran MN (2011) Advances in Transgenic Research for Insect Resistance in Sugarcane. Tropical Plant Biology 4: 52–61.
8. Mendonça AF, Viveiros AJA, Sampaio FF (1996) A broca gigante da cana-de-açúcar, *Castnia licus* Drury, 1770 (Lep.: *Castniidae*). In: Mendonça AF, editor. Pragas da cana-de-açúcar Insetos, Cia. Maceió. pp. 133–167.
9. Pinto AS, Garcia JF, de Oliveira HN (2006) Manejo das Principais Pragas da Cana-de-açúcar. In: Segato SV, Pinto, A. S., Jendiroba, E., Nóbrega, J. C. M., editor. Atualização em Produção de Cana-de-açúcar. 1 ed. Piracicaba, SP: Piracicaba: CP 2. pp. 415p.
10. Rebouças LMC, Caraciolo MdSB, Sant'Ana AEG, Pickett JA, Wadhams LJ, et al. (1999) Composição química da glândula abdominal da fêmea da mariposa *Castnia licus* (Drury) (Lepidoptera:Castniidae): possíveis feromônios e precursores. Química Nova 22: 645–648. doi: 10.1016/j.ecoenv.2015.01.026 PMID: 25638565
11. Moraes SS, Duarte M (2009) Morfologia externa comparada das três espécies do complexo *Telchin licus* (Drury) (Lepidoptera, Castniidae) com uma sinonímia. Revista Brasileira de Entomologia 53: 245–265.
12. Rohde C, Alves LFA, Neves PMOJ, Alves SB, Silva ERLd, et al. (2006) Seleção de isolados de *Beauveria bassiana* (Bals.) Vuill. e *Metarhizium anisopliae* (Metsch.) Sorok. contra o cascudinho *Alphitobius diaperinus* (Panzer) (Coleoptera: Tenebrionidae). Neotropical Entomology 35: 231–240. PMID: 17348135
13. Craveiro KIC, Gomes Junior JE, Silva MCM, Macedo LLP, Lucena WA, et al. (2010) Variant Cry1Ia toxins generated by DNA shuffling are active against sugarcane giant borer. Journal of Biotechnology 145: 215–221. doi: 10.1016/j.jbiotec.2009.11.011 PMID: 19931577
14. Silva-Brandão KL, Almeida LC, Moraes SS, Cônsoli FL (2013) Using population genetic methods to identify the origin of an invasive population and to diagnose cryptic subspecies of *Telchin licus* (Lepidoptera: Castniidae). Bulletin of Entomological Research 103: 89–97. doi: 10.1017/S0007485312000430 PMID: 22971459
15. Consortium TISG (2008) The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. Insect Biochemistry and Molecular Biology 38: 1036–1045. doi: 10.1016/j.ibmb.2008.11.004 PMID: 19121390
16. Negre V, Hotelier T, Volkoff AN, Gimenez S, Cousserans F, et al. (2006) SPODOBASE: an EST database for the lepidopteran crop pest *Spodoptera*. BMC Bioinformatics 7: 1–10. PMID: 16393334
17. Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics 92: 255–264. doi: 10.1016/j.ygeno.2008.07.001 PMID: 18703132

18. Rothberg JM, Leamon JH (2008) The development and impact of 454 sequencing. Nature Biotechnology 26: 1117–1124. doi: 10.1038/nbt1485 PMID: 18846085

19. Bass C, Hebsgaard MB, Hughes J (2012) Genomic resources for the brown planthopper, *Nilaparvata lugens*: Transcriptome pyrosequencing and microarray design. Insect Science 19: 1–12.

20. Hahn DA, Ragland GJ, Shoemaker DD, Denlinger DL (2009) Gene discovery using massively parallel pyrosequencing to develop ESTs for the flesh fly *Sarcophaga crassipalpis*. Bmc Genomics 10: 1–9. doi: 10.1186/1471-2164-10-1 PMID: 19121221

21. Mittapalli O, Bai X, Mamidala P, Rajarapu SP, Bonello P, et al. (2010) Tissue-Specific Transcriptomics of the Exotic Invasive Insect Pest Emerald Ash Borer (*Agrilus planipennis*). Plos One 5: e13708. doi: 10.1371/journal.pone.0013708 PMID: 21060843

22. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, et al. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology 17: 1636–1647. doi: 10.1111/j.1365-294X.2008.03666.x PMID: 18266620

23. Peng X, Zha W, He R, Lu T, Zhu L, et al. (2011) Pyrosequencing the midgut transcriptome of the brown planthopper, *Nilaparvata lugens*. Insect Molecular Biology 20: 745–762. doi: 10.1111/j.1365-2583.2011.01104.x PMID: 21919985

24. Agunbiade TA, Sun W, Coates BS, Djouaka R, Tamò M, et al. (2013) Development of Reference Transcriptomes for the Major Field Insect Pests of Cowpea: A Toolbox for Insect Pest Management Approaches in West Africa. Plos One 8: e79929. doi: 10.1371/journal.pone.0079929 PMID: 24278221

25. Bai X, Zhang W, Orantes L, Jun T-H, Mittapalli O, et al. (2010) Combining Next-Generation Sequencing Strategies for Rapid Molecular Resource Development from an Invasive Aphid Species, *Aphis glycines*. Plos One 5: e11370. doi: 10.1371/journal.pone.0011370 PMID: 20614011

26. Firmino AAP, Fonseca FCdA, de Macedo LLP, Coelho RR, Antonino de Souza JD Jr, et al. (2013) Transcriptome Analysis in Cotton Boll Weevil (*Anthonomus grandis*) and RNA Interference in Insect Pests. Plos One 8: e85079. doi: 10.1371/journal.pone.0085079 PMID: 24386449

27. Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D, et al. (2009) Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera. Insect Biochemistry and Molecular Biology 39: 403–413. doi: 10.1016/j.ibmb.2009.04.001 PMID: 19364528

28. Pauchet Y, Wilkinson P, Vogel H, Nelson DR, Reynolds SE, et al. (2010) Pyrosequencing the *Manduca sexta* larval midgut transcriptome: messages for digestion, detoxification and defence. Insect Molecular Biology 19: 61–75. doi: 10.1111/j.1365-2583.2009.00936.x PMID: 19909380

29. Sun J-Z, Scharf ME (2010) Exploring and integrating cellulolytic systems of insects to advance biofuel technology. Insect Science 17: 163–165.

30. Zhang S, Xu Y, Fu Q, Jia L, Xiang Z, et al. (2011) Proteomic Analysis of Larval Midgut from the Silkworm (*Bombyx mori*). Comparative and Functional Genomics 2011: 1–13.

31. Pigott CR, Ellar DJ (2007) Role of receptors in *Bacillus thuringiensis* crystal toxin activity. Microbiology and Molecular Biology Reviews 71: 255–281. PMID: 17554045

32. Schwarz D, Robertson HM, Feder JL, Varala K, Hudson ME, et al. (2009) Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot *Rhagoletis pomonella*. Bmc Genomics 10: 1–14. doi: 10.1186/1471-2164-10-1 PMID: 19121221

33. Zhang F, Guo H, Zheng H, Zhou T, Zhou Y, et al. (2010) Massively parallel pyrosequencing-based transcriptome analyses of small brown planthopper (*Laodelphax striatellus*), a vector insect transmitting rice stripe virus (RSV). Bmc Genomics 11: 1–13. doi: 10.1186/1471-2164-11-1 PMID: 20044946

34. Broehan G, Arakane Y, Beeman RW, Kramer KJ, Muthukrishnan S, et al. (2010) Chymotrypsin-like peptidases from *Tribolium castaneum*: a role in molting revealed by RNA interference. Insect Biochemistry and Molecular Biology 40: 274–283. doi: 10.1016/j.ibmb.2009.10.009 PMID: 19897036

35. Chikate YR, Tamhane VA, Joshi RS, Gupta VS, Giri AP (2013) Differential protease activity augments polyphagy in *Helicoverpa armigera*. Insect Molecular Biology 22: 258–272. doi: 10.1111/imb.12018 PMID: 23432026

36. Terra WR, Ferreira C (1994) Insect digestive enzymes: properties, compartmentalization and function. Comparative Biochemistry and Physiology 109B: 1–62.

37. Khan AR, James MNG (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. Protein Science 7: 815–836. PMID: 9568890

38. Kanost MR, Gorman MJ (2008) 4—Phenoloxidases in Insect Immunity. Insect Immunology. San Diego: Academic Press. pp. 69–96.

39. Law JH, Dunn PE, Kramer KJ (2006) Insect Proteases and Peptidases. Advances in Enzymology and Related Areas of Molecular Biology: John Wiley & Sons, Inc. pp. 389–425.

40. Neurath H (1984) Evolution of proteolytic enzymes. Science 224: 350–357. PMID: 6369538

41. Srinivasan A, Giri AP, Gupta VS (2006) Structural and functional diversities in lepidopteran serine proteases. Cellular & Molecular Biology Letters 11: 132–154. doi: 10.1016/j.bios.2015.01.052 PMID: 25638815

42. Diaz-Mendoza M, Farinos GP, Castanera P, Hernandez-Crespo P, Ortego F (2007) Proteolytic processing of native Cry1Ab toxin by midgut extracts and purified trypsins from the Mediterranean corn borer *Sesamia nonagrioides*. Journal of Insect Physiology 53: 428–435. PMID: 17336999

43. Gill SS, Cowles EA, Pietrantonio PV (1992) The mode of action of *Bacillus thuringiensis* endotoxins. Annual Review of Entomology 37: 615–636. PMID: 1311541

44. Christou P, Capell T, Kohli A, Gatehouse JA, Gatehouse AM (2006) Recent developments and future prospects in insect pest control in transgenic crops. Trends in Plant Science 11: 302–308. PMID: 16690346

45. Hakim RS, Baldwin K, Smagghe G (2010) Regulation of midgut growth, development, and metamorphosis. Annual Review of Entomology 55: 593–608. doi: 10.1146/annurev-ento-112408-085450 PMID: 19775239

46. Rodriguez-Cabrera L, Trujillo-Bacallao D, Borras-Hidalgo O, Wright DJ, Ayra-Pardo C (2010) RNAi-mediated knockdown of a *Spodoptera frugiperda* trypsin-like serine-protease gene reduces susceptibility to a *Bacillus thuringiensis* Cry1Ca1 protoxin. Environmental Microbiology 12: 2894–2903. doi: 10.1111/j.1462-2920.2010.02259.x PMID: 20545748

47. Whyard S, Singh AD, Wong S (2009) Ingested double-stranded RNAs can act as species-specific insecticides. Insect Biochemistry and Molecular Biology 39: 824–832. doi: 10.1016/j.ibmb.2009.09.007 PMID: 19815067

48. Franco OL, dos Santos RC, Batista JA, Mendes AC, de Araujo MA, et al. (2003) Effects of black-eyed pea trypsin/chymotrypsin inhibitor on proteolytic activity and on development of *Anthonomus grandis*. Phytochemistry 63: 343–349. PMID: 12737983

49. Carlini CR, Grossi-de-Sa MF (2002) Plant toxic proteins with insecticidal properties. A review on their potentialities as bioinsecticides. Toxicon 40: 1515–1539. PMID: 12419503

50. Rufino FPS, Pedroso VMA, Araujo JN, França AFJ, Rabêlo LMA, et al. (2013) Inhibitory effects of a Kunitz-type inhibitor from *Pithecellobium dumosum* (Benth) seeds against insect-pests' digestive proteinases. Plant Physiology and Biochemistry 63: 70–76. doi: 10.1016/j.plaphy.2012.11.013 PMID: 23238511

51. Gatehouse AM, Norton E, Davison GM, Babbe SM, Newell CA, et al. (1999) Digestive proteolytic activity in larvae of tomato moth, *Lacanobia oleracea*; effects of plant protease inhibitors *in vitro* and *in vivo*. Journal of Insect Physiology 45: 545–558. PMID: 12770339

52. Srinivasan T, Kumar KR, Kirti PB (2009) Constitutive expression of a trypsin protease inhibitor confers multiple stress tolerance in transgenic tobacco. Plant & Cell Physiology 50: 541–553. doi: 10.1016/j.bios.2015.01.054 PMID: 25638814

53. Quilis J, López-García B, Meynard D, Guiderdoni E, San Segundo B (2014) Inducible expression of a fusion gene encoding two proteinase inhibitors leads to insect and pathogen resistance in transgenic rice. Plant Biotechnology Journal 12: 367–377. doi: 10.1111/pbi.12143 PMID: 24237606

54. Yang ZX, Wu QJ, Wang SL, Chang XL, Wang JH, et al. (2012) Expression of cadherin, aminopeptidase N and alkaline phosphatase genes in Cry1Ac-susceptible and Cry1Ac-resistant strains of *Plutella xylostella* (L.). Journal of Applied Entomology 136: 539–548.

55. Knight PJ, Knowles BH, Ellar DJ (1995) Molecular cloning of an insect aminopeptidase N that serves as a receptor for *Bacillus thuringiensis* CryIA(c) toxin. The Journal of biological chemistry 270: 17765–17770. PMID: 7629076

56. Park Y, Kim Y (2013) RNA interference of cadherin gene expression in *Spodoptera exigua* reveals its significance as a specific Bt target. Journal of Invertebrate Pathology 114: 285–291. doi: 10.1016/j.jip.2013.09.006 PMID: 24055650

57. Arenas I, Bravo A, Soberon M, Gomez I (2010) Role of alkaline phosphatase from *Manduca sexta* in the mechanism of action of *Bacillus thuringiensis* Cry1Ab toxin. The Journal of Biological Chemistry 285: 12497–12503. doi: 10.1074/jbc.M109.085266 PMID: 20177063

58. Rajagopal R, Sivakumar S, Agrawal N, Malhotra P, Bhatnagar RK (2002) Silencing of midgut aminopeptidase N of *Spodoptera litura* by double-stranded RNA establishes its role as *Bacillus thuringiensis* toxin receptor. Journal of Biological Chemistry 277: 46849–46851. PMID: 12377776

59. Yang Y, Zhu YC, Ottea J, Husseneder C, Leonard BR, et al. (2010) Molecular characterization and RNA interference of three midgut aminopeptidase N isozymes from *Bacillus thuringiensis*-susceptible and -resistant strains of sugarcane borer, *Diatraea saccharalis*. Insect Biochemistry and Molecular Biology 40: 592–603. doi: 10.1016/j.ibmb.2010.05.006 PMID: 20685334

60. Tiewsiri K, Wang P (2011) Differential alteration of two aminopeptidases N associated with resistance to *Bacillus thuringiensis* toxin Cry1Ac in cabbage looper. Proceedings of the National Academy of Sciences of the United States of America 108: 14037–14042. doi: 10.1073/pnas.1102555108 PMID: 21844358

61. Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, et al. (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Research 32: 1–8. PMID: 14704337

62. Papanicolaou A, Stierli R, ffrench-Constant R, Heckel D (2009) Next generation transcriptomes for next generation genomes using est2assembly. BMC Bioinformatics 10: 1–16. doi: 10.1186/1471-2105-10-1 PMID: 19118496

63. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WE, et al. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Research 14: 1147–1159. PMID: 15140833

64. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676. PMID: 16081474

65. Myhre S, Tveit H, Mollestad T, Lægreid A (2006) Additional Gene Ontology structure for improved biological reasoning. Bioinformatics 22: 2020–2027. PMID: 16787968

66. Conesa A, Gotz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics 2008: 1–12.

67. Hackett N, Butler M, Shaykhiev R, Salit J, Omberg L, et al. (2012) RNA-Seq quantification of the human small airway epithelium transcriptome. Bmc Genomics 13: 82. doi: 10.1186/1471-2164-13-82 PMID: 22375630

68. Lahens N, Kavakli I, Zhang R, Hayer K, Black M, et al. (2014) IVT-seq reveals extreme bias in RNA sequencing. Genome Biology 15: R86. doi: 10.1186/gb-2014-15-6-r86 PMID: 24981968

69. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. doi: 10.1093/bioinformatics/btq033 PMID: 20110278

70. Wang L, Wang S, Li W (2012) RSeQC: quality control of RNA-seq experiments. Bioinformatics 28: 2184–2185. doi: 10.1093/bioinformatics/bts356 PMID: 22743226

71. Papanicolaou A, Gebauer-Jung S, Blaxter ML, Owen McMillan W, Jiggins CD (2008) ButterflyBase: a platform for lepidopteran genomics. Nucleic Acids Research 36: 582–587.

72. Van Bel M, Proost S, Van Neste C, Deforce D, Van de Peer Y, et al. (2013) TRAPID: an efficient online tool for the functional and comparative analysis of de novo RNA-Seq transcriptomes. Genome Biology 14: R134. doi: 10.1186/gb-2013-14-12-r134 PMID: 24330842

73. Ahn SC, Baek BS, Oh T, Song CS, Chatterjee B (2000) Rapid mini-scale plasmid isolation for DNA sequencing and restriction mapping. BioTechniques 29: 466–468. PMID: 10997259

74. Zhao S, Fernald RD (2005) Comprehensive Algorithm for Quantitative Real-Time Polymerase Chain Reaction. Journal of Computational Biology 12: 1047–1064. PMID: 16241897

75. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, et al. (2012) ExPASy: SIB bioinformatics resource portal. Nucleic Acids Research 40: 597–603.

76. Pierleoni A, Martelli P, Casadio R (2008) PredGPI: a GPI-anchor predictor. BMC Bioinformatics 9: 1–11. doi: 10.1186/1471-2105-9-1 PMID: 18173834

77. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673–4680. PMID: 7984417

78. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98.

79. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution: 2731–2739.

80. Gregory R, Darby AC, Irving H, Coulibaly MB, Hughes M, et al. (2011) A de novo expression profiling of *Anopheles funestus*, malaria vector in Africa, using 454 pyrosequencing. Plos One 6: e17418. doi: 10.1371/journal.pone.0017418 PMID: 21364769

81. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410. PMID: 2231712

82. Ma W, Zhang Z, Peng C, Wang X, Li F, et al. (2012) Exploring the Midgut Transcriptome and Brush Border Membrane Vesicle Proteome of the Rice Stem Borer, *Chilo suppressalis* (Walker). Plos One 7: e38151. doi: 10.1371/journal.pone.0038151 PMID: 22666467

83. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8: 1494–1512. doi: 10.1038/nprot.2013.084 PMID: 23845962

84. group Ba, Xia Q, Zhou Z, Lu C, Cheng D, et al. (2004) A Draft Sequence for the Genome of the Domesticated Silkworm (*Bombyx mori*). Science 306: 1937–1940. PMID: 15591204

85. Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21: 1859–1875. PMID: 15728110

86. Bainbridge M, Warren R, Hirst M, Romanuik T, Zeng T, et al. (2006) Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. Bmc Genomics 7: 246. PMID: 17010196

87. Bai X, Mamidala P, Rajarapu SP, Jones SC, Mittapalli O (2011) Transcriptomics of the Bed Bug (*Cimex lectularius*). Plos One 6: e16336. doi: 10.1371/journal.pone.0016336 PMID: 21283830

88. Karatolos N, Pauchet Y, Wilkinson P, Chauhan R, Denholm I, et al. (2011) Pyrosequencing the transcriptome of the greenhouse whitefly, *Trialeurodes vaporariorum* reveals multiple transcripts encoding insecticide targets and detoxifying enzymes. Bmc Genomics 12: 1–14. doi: 10.1186/1471-2164-12-350 PMID: 21733171

89. Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, et al. (2010) The dynamic genome of Hydra. Nature 464: 592–596. doi: 10.1038/nature08830 PMID: 20228792

90. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, et al. (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453: 1064–1071. doi: 10.1038/nature06967 PMID: 18563158

91. Ralph S, Chun H, Kolosova N, Cooper D, Oddy C, et al. (2008) A conifer genomics resource of 200,000 spruce (Picea spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). Bmc Genomics 9: 484. doi: 10.1186/1471-2164-9-484 PMID: 18854048

92. Ewen-Campen B, Shaner N, Panfilio KA, Suzuki Y, Roth S, et al. (2011) The maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*. Bmc Genomics 12: 1–22. doi: 10.1186/1471-2164-12-350 PMID: 21733171

93. Hull JJ, Geib SM, Fabrick JA, Brent CS (2013) Sequencing and de novo assembly of the western tarnished plant bug (*Lygus hesperus*) transcriptome. Plos One 8: e55105. doi: 10.1371/journal.pone.0055105 PMID: 23357950

94. Magalhaes LC, van Kretschmar JB, Donohue KV, Roe RM (2013) Pyrosequencing of the adult tarnished plant bug, *Lygus lineolaris*, and characterization of messages important in metabolism and development. Entomologia Experimentalis et Applicata 146: 364–378.

95. Nirmala X, Schetelig MF, Yu F, Handler AM (2013) An EST database of the Caribbean fruit fly, *Anastrepha suspensa* (Diptera: Tephritidae). Gene 517: 212–217. doi: 10.1016/j.gene.2012.12.012 PMID: 23296060

96. Warr E, Aguilar R, Dong Y, Mahairaki V, Dimopoulos G (2007) Spatial and sex-specific dissection of the *Anopheles gambiae* midgut transcriptome. Bmc Genomics 8: 1–11. PMID: 17199895

97. Xu Q, Lu A, Xiao G, Yang B, Zhang J, et al. (2012) Transcriptional profiling of midgut immunity response and degeneration in the wandering silkworm, *Bombyx mori*. Plos One 7: e43769. doi: 10.1371/journal.pone.0043769 PMID: 22937093

98. Hahn MW, Han MV, Han S-G (2007) Gene Family Evolution across 12 *Drosophila* Genomes. PLoS Genetics 3: e197. PMID: 17997610

99. Klein B, Le Moullac G, Sellos D, Van Wormhoudt A (1996) Molecular cloning and sequencing of trypsin cDNAs from *Penaeus vannamei* (Crustacea, Decapoda): use in assessing gene expression during the moult cycle. The International Journal of Biochemistry & Cell Biology 28: 551–563. doi: 10.1016/j.biocel.2015.01.003 PMID: 25595463

100. Perona JJ, Craik CS (1995) Structural basis of substrate specificity in the serine proteases. Protein Science 4: 337–360. PMID: 7795518

101. Perona JJ, Craik CS (1997) Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. The Journal of Biological Chemistry 272: 29987–29990. PMID: 9374470

102. Yao J, Buschman LL, Oppert B, Khajuria C, Zhu KY (2012) Characterization of cDNAs Encoding Serine Proteases and Their Transcriptional Responses to Cry1Ab Protoxin in the Gut of *Ostrinia nubilalis* Larvae. Plos One 7: e44090. doi: 10.1371/journal.pone.0044090 PMID: 22952884

103. Coates BS, Hellmich RL, Lewis LC (2006) Sequence variation in trypsin- and chymotrypsin-like cDNAs from the midgut of *Ostrinia nubilalis*: methods for allelic differentiation of candidate *Bacillus thuringiensis* resistance genes. Insect Molecular Biology 15: 13–24. PMID: 16469064

104. Bolter C, Jongsma MA (1997) The adaptation of insects to plant protease inhibitors. Journal of Insect Physiology 43: 885–895. PMID: 12770458

105. Brioschi D, Nadalini LD, Bengtson MH, Sogayar MC, Moura DS, et al. (2007) General up regulation of *Spodoptera frugiperda* trypsins and chymotrypsins allows its adaptation to soybean proteinase inhibitor. Insect Biochemistry and Molecular Biology 37: 1283–1290. PMID: 17967347

106. Bown DP, Wilkinson HS, Gatehouse JA (1997) Differentially regulated inhibitor-sensitive and insensitive protease genes from the phytophagous insect pest, *Helicoverpa armigera*, are members of complex multigene families. Insect Biochemistry and Molecular Biology 27: 625–638. PMID: 9404008

107. Zou Z, Lopez DL, Kanost MR, Evans JD, Jiang H (2006) Comparative analysis of serine protease-related genes in the honey bee genome: possible involvement in embryonic development and innate immunity. Insect Molecular Biology 15: 603–614. PMID: 17069636

108. Zhao P, Wang G-H, Dong Z-M, Duan J, Xu P-Z, et al. (2010) Genome-wide identification and expression analysis of serine proteases and homologs in the silkworm *Bombyx mori*. Bmc Genomics 11: 405. doi: 10.1186/1471-2164-11-405 PMID: 20576138

109. Weng LX, Deng HH, Xu JL, Li Q, Zhang YQ, et al. (2011) Transgenic sugarcane plants expressing high levels of modified cry1Ac provide effective control against stem borers in field trials. Transgenic Research 20: 759–772. doi: 10.1007/s11248-010-9456-8 PMID: 21046242

110. Christy LA, Arvinth S, Saravanakumar M, Kanchana M, Mukunthan N, et al. (2009) Engineering sugarcane cultivars with bovine pancreatic trypsin inhibitor (aprotinin) gene for protection against top borer (*Scirpophaga excerptalis* Walker). Plant Cell Reports 28: 175–184. doi: 10.1007/s00299-008-0628-4 PMID: 18985354

111. Hooper NM (1994) Families of zinc metalloproteases. FEBS Letters 354: 1–6. PMID: 7957888

112. Laustsen PG, Vang S, Kristensen T (2001) Mutational analysis of the active site of human insulin-regulated aminopeptidase. European Journal of Biochemistry 268: 98–104. PMID: 11121108

113. Lu YJ, Adang MJ (1996) Conversion of *Bacillus thuringiensis* cryIAc-binding aminopeptidase to a soluble form by endogenous phosphatidylinositol phospholipase C. Insect Biochemistry and Molecular Biology 26: 33–40.

114. Pardo-Lopez L, Soberon M, Bravo A (2013) *Bacillus thuringiensis* insecticidal three-domain Cry toxins: mode of action, insect resistance and consequences for crop protection. FEMS Microbiology Reviews 37: 3–22. doi: 10.1111/j.1574-6976.2012.00341.x PMID: 22540421

115. Sivakumar S, Rajagopal R, Venkatesh GR, Srivastava A, Bhatnagar RK (2007) Knockdown of aminopeptidase-N from *Helicoverpa armigera* larvae and in transfected Sf21 cells by RNA interference reveals its functional interaction with *Bacillus thuringiensis* insecticidal protein Cry1Ac. The Journal of Biological Chemistry 282: 7312–7319. PMID: 17213205

116. Masson L, Lu YJ, Mazza A, Brousseau R, Adang MJ (1995) The CryIA(c) receptor purified from *Manduca sexta* displays multiple specificities. The Journal of biological chemistry 270: 20309–20315. PMID: 7657602

117. Sangadala S, Azadi P, Carlson R, Adang MJ (2001) Carbohydrate analyses of *Manduca sexta* aminopeptidase N, co-purifying neutral lipids and their functional interactions with *Bacillus thuringiensis* Cry1Ac toxin. Insect biochemistry and molecular biology 32: 97–107. PMID: 11719073

118. Knight PJK, Carroll J, Ellar DJ (2004) Analysis of glycan structures on the 120 kDa aminopeptidase N of *Manduca sexta* and their interactions with *Bacillus thuringiensis* Cry1Ac toxin. Insect Biochemistry and Molecular Biology 34: 101–112. PMID: 14976987

119. Stephens E, Sugars J, Maslen SL, Williams DH, Packman LC, et al. (2004) The N-linked oligosaccharides of aminopeptidase N from *Manduca sexta*: site localization and identification of novel N-glycan structures. European journal of biochemistry / FEBS 271: 4241–4258. PMID: 15511230

120. Pardo-Lopez L, Gomez I, Rausell C, Sanchez J, Soberon M, et al. (2006) Structural changes of the Cry1Ac oligomeric pre-pore from *Bacillus thuringiensis* induced by N-acetylgalactosamine facilitates toxin membrane insertion. Biochemistry 45: 10329–10336. PMID: 16922508

121. Herrero S, Gechev T, Bakker PL, Moar WJ, de Maagd RA (2005) *Bacillus thuringiensis* Cry1Ca-resistant *Spodoptera exigua* lacks expression of one of four Aminopeptidase N genes. Bmc Genomics 6: 96. PMID: 15978131

122. Angelucci C, Barrett-Wilt GA, Hunt DF, Akhurst RJ, East PD, et al. (2008) Diversity of aminopeptidases, derived from four lepidopteran gene duplications, and polycalins expressed in the midgut of *Helicoverpa armigera*: Identification of proteins binding the δ-endotoxin, Cry1Ac of *Bacillus thuringiensis*. Insect Biochemistry and Molecular Biology 38: 685–696. doi: 10.1016/j.ibmb.2008.03.010 PMID: 18549954

123. Ningshen TJ, Chaitanya RK, Hari PP, Vimala Devi PS, Dutta-Gupta A (2013) Characterization and regulation of *Bacillus thuringiensis* Cry toxin binding aminopeptidases N (APNs) from non-gut visceral

tissues, Malpighian tubule and salivary gland: Comparison with midgut-specific APN in the moth *Achaea janata*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 166: 194–202. doi: 10.1016/j.cbpb.2013.09.005 PMID: 24045122

124. Zhang S, Cheng H, Gao Y, Wang G, Liang G, et al. (2009) Mutation of an aminopeptidase N gene is associated with *Helicoverpa armigera* resistance to *Bacillus thuringiensis* Cry1Ac toxin. Insect Biochemistry and Molecular Biology 39: 421–429. doi: 10.1016/j.ibmb.2009.04.003 PMID: 19376227

125. Rajagopal R, Agrawal N, Selvapandiyan A, Sivakumar S, Ahmad S, et al. (2003) Recombinantly expressed isoenzymic aminopeptidases from *Helicoverpa armigera* (American cotton bollworm) midgut display differential interaction with closely related *Bacillus thuringiensis* insecticidal proteins. Biochemical Journal 370: 971–978. PMID: 12441000

126. Coates BS, Sumerford DV, Siegfried BD, Hellmich RL, Abel CA (2013) Unlinked genetic loci control the reduced transcription of aminopeptidase N 1 and 3 in the European corn borer and determine tolerance to *Bacillus thuringiensis* Cry1Ab toxin. Insect Biochemistry and Molecular Biology 43: 1152–1160. doi: 10.1016/j.ibmb.2013.09.003 PMID: 24121099

127. Budatha M, Meur G, Dutta-gupta A (2007) A novel aminopeptidase in the fat body of the moth *Achaea janata* as a receptor for *Bacillus thuringiensis* Cry toxins and its comparison with midgut aminopeptidase. Biochemical Journal 405: 287–297. PMID: 17402938

128. Nakanishi K, Yaoi K, Nagino Y, Hara H, Kitami M, et al. (2002) Aminopeptidase N isoforms from the midgut of *Bombyx mori* and *Plutella xylostella*—their classification and the factors that determine their binding specificity to *Bacillus thuringiensis* Cry1A toxin. FEBS Letters 519: 215–220. PMID: 12023048

129. Crava CM, Bel Y, Lee SF, Manachini B, Heckel DG, et al. (2010) Study of the aminopeptidase N gene family in the lepidopterans *Ostrinia nubilalis* (Hübner) and *Bombyx mori* (L.): Sequences, mapping and expression. Insect Biochemistry and Molecular Biology 40: 506–515. doi: 10.1016/j.ibmb.2010.04.010 PMID: 20420910