



Multiple Roles of the Polycistronic Gene *Tarsal-less/Mille-Pattes/Polished-Rice* During Embryogenesis of the Kissing Bug *Rhodnius prolixus*

Vitória Tobias-Santos^{1,2}, Diego Guerra-Almeida¹, Flavia Mury¹, Lupis Ribeiro¹, Mateus Berni³, Helena Araujo^{3,4}, Carlos Logullo^{1,4}, Natália Martins Feitosa¹, Jackson de Souza-Menezes¹, Evenilton Pessoa Costa⁵ and Rodrigo Nunes-da-Fonseca^{1,4*}

¹ Instituto de Biodiversidade e Sustentabilidade (NUPEM), Universidade Federal do Rio de Janeiro, Campus Macaé, Rio de Janeiro, Brazil, ² Institut de Génomique Fonctionnelle de Lyon - CNRS UMR 5242, Lyon, France, ³ Laboratório de Biologia Molecular do Desenvolvimento, Instituto de Ciências Biomédicas, Universidade Federal do Rio de Janeiro, Cidade Universitária Ilha do Fundão, Rio de Janeiro, Brazil, ⁴ Instituto Nacional de Ciência e Tecnologia em Entomologia Molecular - INCT-EM. INCT-EM, Rio de Janeiro, Brazil, ⁵ Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Rio de Janeiro, Brazil

OPEN ACCESS

Edited by:

Elizabeth Jenness Duncan,
University of Leeds, United Kingdom

Reviewed by:

Leslie Pick,
University of Maryland, College Park,
United States
Miriam Rosenberg,
Hebrew University of Jerusalem, Israel

*Correspondence:

Rodrigo Nunes-da-Fonseca
mfonseca@macae.ufrj.br

Specialty section:

This article was submitted to
Evolutionary Developmental Biology,
a section of the journal
Frontiers in Ecology and Evolution

Received: 12 June 2019

Accepted: 20 September 2019

Published: 11 October 2019

Citation:

Tobias-Santos V, Guerra-Almeida D, Mury F, Ribeiro L, Berni M, Araujo H, Logullo C, Feitosa NM, de Souza-Menezes J, Pessoa Costa E and Nunes-da-Fonseca R (2019) Multiple Roles of the Polycistronic Gene *Tarsal-less/Mille-Pattes/Polished-Rice* During Embryogenesis of the Kissing Bug *Rhodnius prolixus*. *Front. Ecol. Evol.* 7:379. doi: 10.3389/fevo.2019.00379

Genes encoding small open-reading frames (smORFs) have been characterized as essential players of developmental processes. The smORF *Tarsal-less/mille-pattes/polished-rice* has been thoroughly investigated in holometabolous insects, such as the fruit fly *Drosophila melanogaster* and the red flour beetle *Tribolium castaneum*, while its function in hemimetabolous insects was only recently investigated. Thus, we analyzed the function of the *tal/pri/mlpt* ortholog in a hemimetabolous insect, the kissing bug *Rhodnius prolixus* (*Rp*). Sequence analysis shows that *Rp-mlpt* polycistronic mRNA encodes two small peptides (11–14 amino acids) containing a LDPTG(L/Q/T)Y motif. Interestingly, a new hemipteran-specific conserved peptide of ~80 amino acids was also identified by *in silico* analysis. *In silico* docking analysis supports the binding of the small *mlpt* peptides to the transcription factor Shavenbaby. *Rp-mlpt in situ* hybridization and knockdown via RNA interference showed a conserved role of *Rp-mlpt* during embryogenesis, with a major role in the regulation of thoracic vs. abdominal segmentation, leg development, and head formation. Altogether, our study shows that *mlpt* segmentation role is conserved in the common ancestor of Paraneoptera and suggests that polycistronic genes might generate order specific smORFs.

Keywords: smORF, peptides, evo-devo, arthropod development, hemiptera, evolution, MZT, hemimetabolous insect

INTRODUCTION

A large number of essential genes required for biological processes have been discovered by genetic screenings in model organisms such as the fruit fly *Drosophila melanogaster* (e.g., Nüsslein-Volhard and Wieschaus, 1980). While most loci important for developmental processes were identified in these original screenings, recent genetic and expression analyses in *D. melanogaster* and in the

beetle *Tribolium castaneum* showed that genes previously classified as putative non-coding RNAs encode functional small open reading frames (smORFs) or sORFs (Pueyo et al., 2016a; Chekulaeva and Rajewsky, 2018). smORFs, ORFs smaller than 100 amino acids, have been described as being fundamental for several developmental processes of insects (Savard et al., 2006; Galindo et al., 2007; Kondo et al., 2007), although gene prediction methods have, in general, discarded smORFs in genome-wide predictions (reviewed in Saghatelian and Couso, 2015). Comparative genomic analysis of Drosophilid species showed an unexpected conservation of smORF-containing genes, suggesting important biological roles for this new class of genes (Ladoukakis et al., 2011).

Later on, experimental analysis of conserved smORFs such as *sarcolamban* (*scl*), a conserved peptide involved in Ca^{2+} uptake at the sarco-endoplasmic reticulum (Magny et al., 2013), and *hemotin* (Pueyo et al., 2016b), a conserved phagocytosis regulator, provided further evidence that genes containing smORFs might constitute a reservoir of important players in metazoan genomes. In the past years, direct evidence of large-scale smORF translation has been obtained by ribosomal profiling of polysomal fractions in *Drosophila*, using the Poly Ribo-Seq technique (Aspden et al., 2014). This study was able to classify smORFs in two groups: “longer” smORFs of around 80 amino acids resembling canonical proteins, mostly containing transmembrane motifs, and shorter (“dwarf”) smORFs. These “dwarf” smORFs are generally shorter (around 20 amino-acid long), less conserved and mostly found in 5'-UTRs and non-coding RNAs.

While bioinformatic studies point to hundreds or thousands of genes containing putative smORFs, only a few functional studies have been performed. In insects, the function of the smORF founding member *mille-pattes* (*mlpt*) was only investigated in detail in holometabolous insects, such as the *D. melanogaster* and *T. castaneum* (Savard et al., 2006; Galindo et al., 2007; Kondo et al., 2007), and more recently in basally branching Diptera (Jiménez-Guri et al., 2018) and Lepidoptera (Cao et al., 2018). In *T. castaneum*, *mlpt* acts as a gap gene during the process of embryonic segmentation, regulating Hox genes and thoracic vs. abdominal specification; knockdown of *mlpt* leads to embryos with multiple legs, *mille-pattes* in French (Savard et al., 2006). In *D. melanogaster*, *tarsal-less* (*tal*) was identified in a spontaneous mutant with defects in the distal part of the legs, the tarsus (Galindo et al., 2007). Independently, in the same year, Polished Rice (*Pri*) peptides were shown to be required for the F-actin organization during trichome morphogenesis; embryos lacking *pri* show external cuticle defects, resembling *polished rice* (Kondo et al., 2007). More recently, Mlpt peptides were shown to be regulated by ecdysone, defining the onset of epidermal trichome development, through post-translational control of the Shavenbaby (Svb) transcription factor (Chanut-Delalande et al., 2014). Altogether, Mlpt peptides display context-specific roles and interactions within different developmental processes.

While in *D. melanogaster*, *mlpt* early embryonic gene expression is segmental displaying a pair-rule pattern, mutants

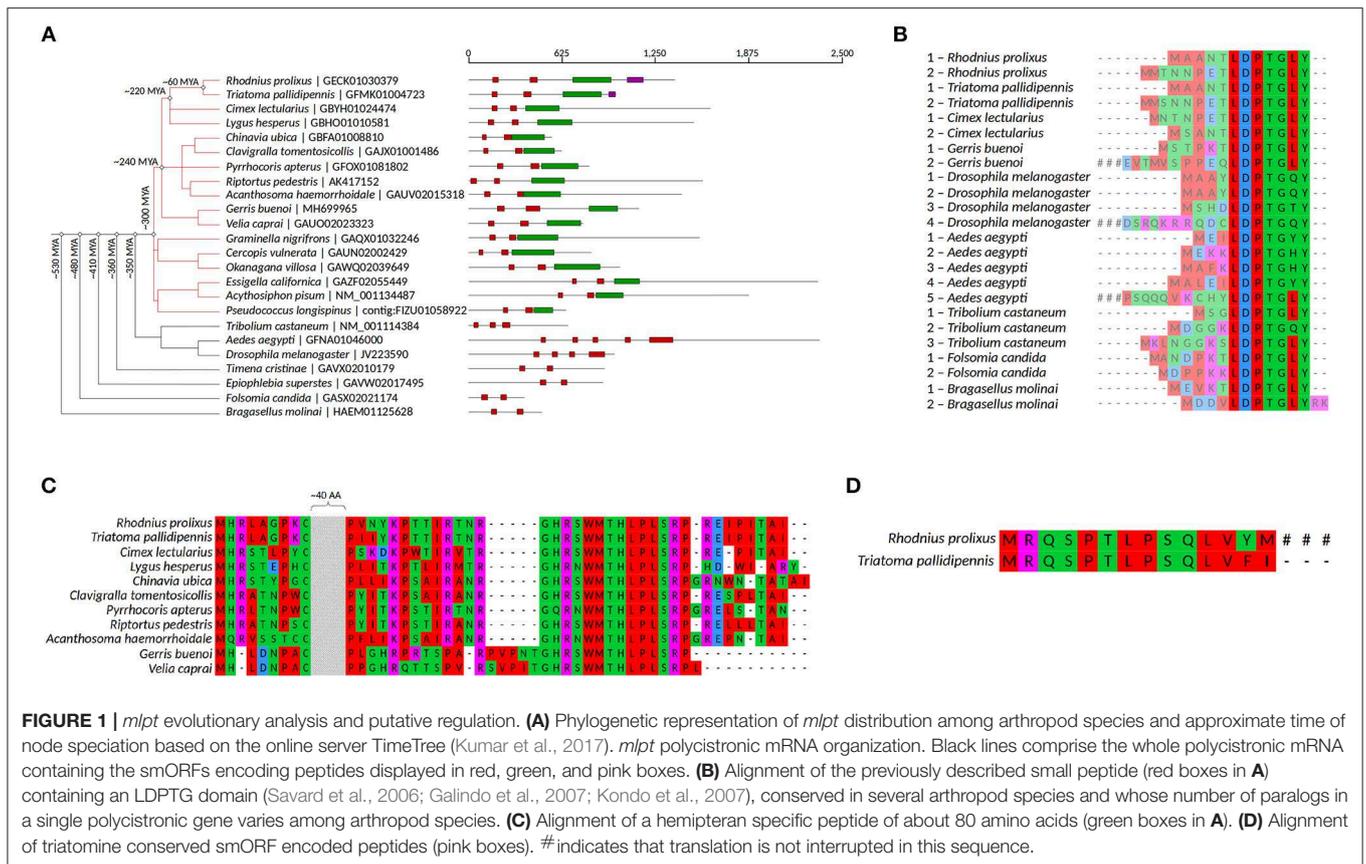
do not display segmentation or homeotic alterations as reported in *T. castaneum*. *D. melanogaster* embryonic mutant phenotypes include broken trachea, loss of cephalopharyngeal skeleton, abnormal posterior spiracles and lack of denticle belts, structures of late embryonic *mlpt* expression (Galindo et al., 2007).

To clarify whether the segmentation function of *mlpt* is ancestral but has been lost in the lineage giving rise to *D. melanogaster*, or whether it is a recently arisen specialization of *T. castaneum*, it is important to study *mlpt* function in other insect groups. Since hemipterans, as hemimetabolous, comprise the sister group of holometabolous insects (reviewed in Panfilio and Angelini, 2018), a functional characterization of *mlpt* in the emergent hemiptera model, the kissing bug *Rhodnius prolixus* would be important (Nunes-da-Fonseca et al., 2017). *R. prolixus* available genomic and transcriptomic resources (Medeiros et al., 2011; Ribeiro et al., 2014; Mesquita et al., 2015; Marchant et al., 2016; Brito et al., 2018), an established embryonic staging system (Berni et al., 2014) and the recent availability of *in situ* hybridization and RNA interference techniques are great advantages of this model system (Nunes-da-Fonseca et al., 2017). Recently, *mlpt* function has been investigated in several insects, including two hemipteran species, the water strider *Gerris buenoi* (Hemiptera, Gerridae) and the milkweed bug *Oncopeltus fasciatus* (Hemiptera, Lygaeidae) (Ray et al., 2019). This analysis provided evidence that Mlpt peptides, the ubiquitin-ligase Ubr3 and the transcription factor Shavenbaby (Svb), constitute an ancient developmental module required for early insect embryo patterning.

Here, we report interesting conserved and new functional aspects of the prototypic smORF *mlpt* gene of *R. prolixus*. First, sequence analysis identified a new hemiptera-specific peptide in the polycistronic mRNA of *mlpt*, a gene conserved throughout the Pancrustacean clade. Second, molecular docking analysis indicates that the small peptide containing a LDPTG(L/Q/T)Y consensus motif interacts with the N-terminus of the transcription factor Svb, as in *D. melanogaster*. Third, expression and functional analysis shows that the ortholog of *mlpt* acts during embryonic segmentation, regulating the transition between thoracic and abdominal identity, similarly to its role previously described in *T. castaneum*. Fourth, a conserved role in tarsal patterning was also observed. Overall, segmentation and tarsal patterning are conserved and ancestral roles of *mlpt* and our data provide evidence that generation of new peptides from smORFs might constitute an underestimated mechanism for the evolution of new genes.

RESULTS

The recent description of several biologically important genes encoding smORFs among metazoans opened new avenues for molecular biology and functional genomics research (Saghatelian and Couso, 2015; Pueyo et al., 2016a; Zanet et al., 2016). *mlpt* function has been largely investigated in holometabolous insects, while its function in hemimetabolous insects has only been



recently reported (Ray et al., 2019). In the current manuscript, we describe a bioinformatic and functional analysis of *mlpt* function in the kissing bug *R. prolixus*.

mlpt Polycistronic Gene and Peptide Distribution Among Arthropods

Previous BLAST searches for genes encoding *mlpt* peptides against different arthropod genomes and transcriptomes provided evidence that this gene containing smORFs is restricted to insects and crustaceans (Galindo et al., 2007). The increase in genome sequences in the past years (Tribolium Genome Sequencing Consortium et al., 2008; Schwager et al., 2017; Panfilio et al., 2019) allowed us to perform a complete search in the available arthropod genomes. Using a non-stringent BLAST approach (see methods for details), we identified *mlpt* orthologs in available insect genomes (Supplementary Table 1). These orthologs encode between two and several copies of a peptide of 11–32 amino acids, containing a LDPTG(L/Q/T)Y consensus motif (Figure 1A–red boxes, Figure 1B). These peptides have been previously shown to mediate the switch of the *svb* transcription factor from a repressor to an activator (Kondo et al., 2010) via the ubiquitin-conjugating complex, UbcD6-Ubr3, and proteasome recruitment (Zanet et al., 2015). Remarkably, hemiptera *mlpt* orthologs not only contain two smORFs encoding the LDPTG(L/Q/T)Y consensus motif (Figure 1A–red boxes), but

also contain a larger hemiptera-specific smORF of about 80 amino acids (Figure 1A–green boxes and Figure 1C). We named this new peptide smHemiptera due to its restricted phylogenetic distribution. Interestingly, a polycistronic transcript containing all the aforementioned smORFs was identified in a digestive tract *R. prolixus* transcriptome (Ribeiro et al., 2014 and our own observations). This peptide appears to be smaller and less conserved in basally branching hemipteran species with circa 60 amino acids in the genome of the pea aphid *Acyrtosiphon pisum* and *Pseudococcus longispinus* and larger than 80 amino acids in triatomines, such as *Rhodnius prolixus* and *Triatoma pallidipennis*. smHemiptera contains a conserved stretch of 14 amino acids (GHR(S/N)WMTHLPLSRP) in almost all derived hemipteran species with available transcriptomes (Figures 1A,C). The unique exception is the genome of the milkweed bug, *O. fasciatus* (Panfilio et al., 2019), which lacks a *mlpt* gene sequence in the current assembly. *Of-mlpt* transcript recently published by Ray et al. (2019) does not contain smHemiptera. Thus, either *O. fasciatus mlpt* gene lost smHemiptera or the published transcript is not complete. We favor the latter possibility since smHemiptera is also present in *mlpt* genes of all species analyzed, including the recently published *mlpt* sequence of water strider, *Gerris buenoi* (Hemiptera, Gerridae) (Ray et al., 2019). Lastly, another peptide of 14 amino acids was identified at the 3' region of two triatomine species (Figure 1A–pink boxes and Figure 1D).

Shavenbaby-*mlpts* Binding Affinity, Interaction Modes, and Hotspots Predictions

Previous work in *D. melanogaster* identified the 31 N-terminal residues of the Shavenbaby (Svb) transcription factor (TF) as a *mlpt*-dependent degradation signal, or degron (Zanet et al., 2015). In order to investigate the interaction modes between the peptides encoded by the polycistronic mRNA of *mlpt* and the Svb TF from *D. melanogaster* (Dmel-Svb) and *R. prolixus* (Rprol-Svb), 3D structures were predicted by an *ab initio* approach and protein-peptide docking assays were performed. First, exhaustive analyses about the secondary structure profile and disordered regions on Dmel-Svb and Rprol-Svb amino acid sequences were performed. Our results suggest that both Dmel-Svb and Rprol-Svb display mostly disordered and coiled regions (Supplementary Figures 1, 2), as previously suggested (Zanet et al., 2015). The disordered and coiled regions appear in large domains of both proteins, being interspersed by small alpha-helices and short beta-sheets (Figures 2A–R; Supplementary Figures 2A,B). According to DISOPRED3, around to C-terminal in both Svbs, the disorder probability decreased, indicating that the Zinc-finger double domain located in this region is more ordered. Smaller peptides containing LDPTG(L/Q/T)Y motifs, Dmel-pptd1 (Supplementary Figure 2C), Rprol-Mlpt1 (Supplementary Figure 2E) and Rprol-Mlpt2 (Supplementary Figure 2F) assumed disordered conformations. In contrast, Dmel-Mlpt4 (Supplementary Figure 2D), a peptide with 32 aa and unknown function (Galindo et al., 2007; Kondo et al., 2007), presented an ordered structure with two alpha-helices, intercalated by two small disordered regions. Moreover, DISOPRED3 program also predicted the protein-binding sites on the Dmel-Svb N-terminal ranging from Leu19 to Asn23 and in Rprol-Svb from Ser24 to Leu42 (Supplementary Figures 1A,B). Interestingly, our findings about Dmel-Svb-Mlpt interactions corroborates previous findings of Zanet et al. (2015), which showed that the region between the first 11 (N11) and 31 (N31) amino acids are required for Mlpt-mediated processing. Therefore, we further conducted docking assays with Svb and Mlpts, taking into account the data prediction and previously published biological data (Zanet et al., 2015).

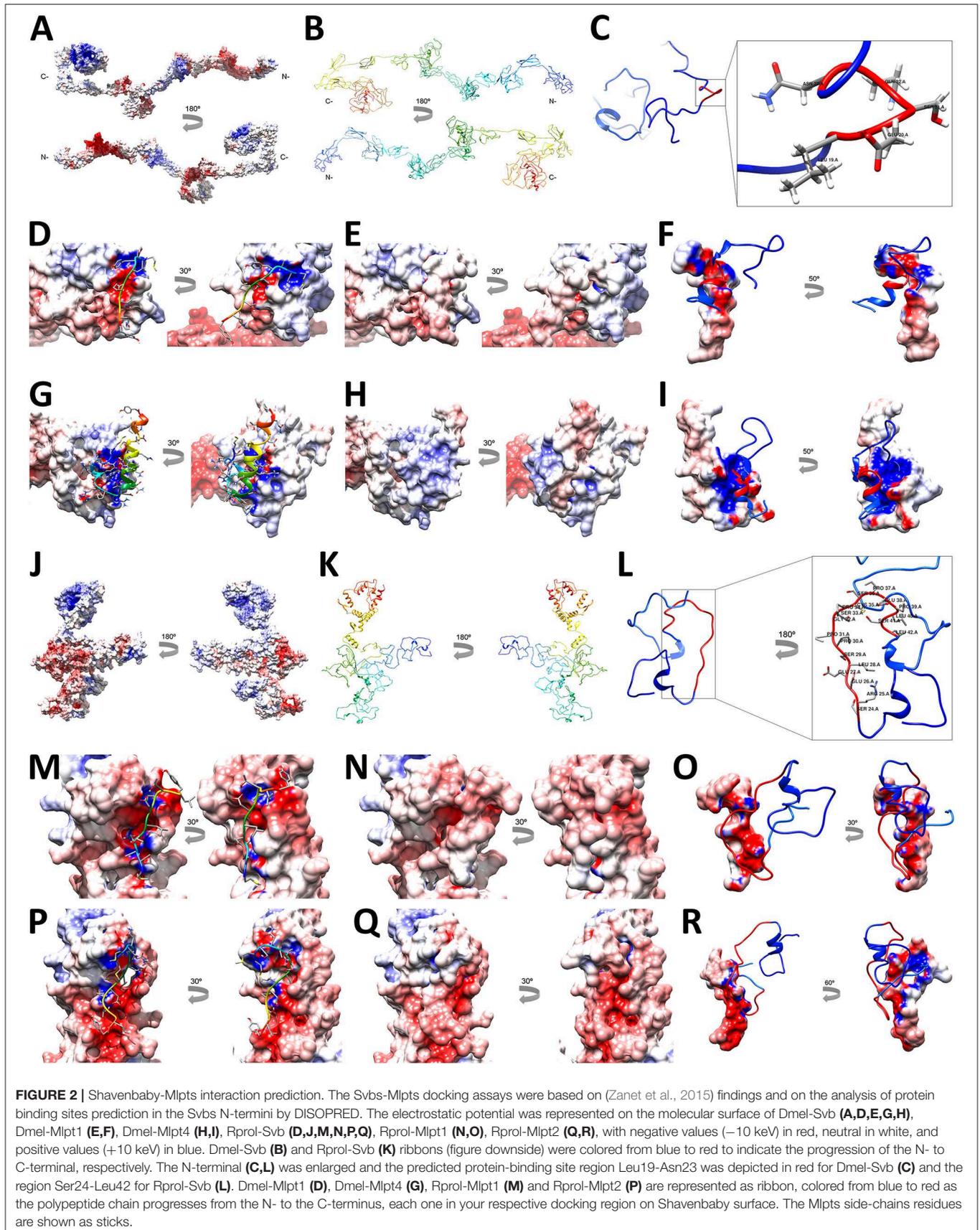
Docking assays focused in the predicted protein-binding sites on the Svb N-terminal. Our results suggest that at least half of LDPTG(L/Q/T)Y motifs residues from Mlpts peptides are involved with complex formation (Supplementary Table 2). The possible interactions at the complex interfaces (Supplementary Figure 3), the binding free energy contribution per-residue (Supplementary Table 2) as well as their theoretical interaction affinities (Supplementary Table 2), were predicted. The binding affinity measured by HawkDock program, using the MM/GBSA method, showed that Dmel-Svb_Dmel-Mlpt4 has the highest affinity, followed by Rprol-Svb_Rprol-Mlpt1, after by Rprol-Svb_Rprol-Mlpt2 and Dmel-Svb_Dmel-Mlpt1 (Supplementary Table 3). The dissociation constant (Kd) was measured by PRODIGY program and suggests a low affinity between them (Supplementary Table 3). The

amount of bonds and contacts were measured by PDBsum (Supplementary Figure 3; Supplementary Table 3), and the results suggest that the interaction model Rprol-Svb_Rprol-Mlpt1 has the greatest number of non-bonded contacts, followed by Rprol-Svb_Rprol-Mlpt2. The Rprol-Svb_Rprol-Mlpt1 and Dmel-Svb_Dmel-Mlpt4 established the greatest number of hydrogens bonds but Dmel-Svb_Dmel-Mlpt4 has the greatest number of salt-bridges. Further analysis by *in silico* site-directed mutagenesis validated the contribution of each amino acid for peptide-TF binding (Supplementary Figure 4). Importantly, the predictions of binding affinity changes upon mutations indicate that globally at least 30% of residues at the interface are critical for protein complex stabilization. Analysis of binding affinity changes upon mutations corroborate previous biological findings of Zanet et al. (2015), since single mutations of Leucines (L24, L25, and L19) and Glutamic Acid (E20) at the N-terminal region of Dmel-Svb affect this region's interaction with Dmel-Mlpt1 (Supplementary Figure 4). These amino acids lie on the region previously experimentally demonstrated to be required for Dmel-Svb processing, mediated by Dmel-Mlpt1 (Zanet et al., 2015). To summarize, the proposed model was able to recover previously data of Dmel-Svb_Dmel-Mlpt biological interaction (Zanet et al., 2015) and to suggest the putative regions of interaction between Rprol-Svb and Rprol-Mlpts.

Rp-mlpt Spatial Expression Pattern and Relative Expression During *Rhodnius prolixus* Embryogenesis

mlpt embryonic expression was originally described in the beetle *T. castaneum* and has been characterized by a very dynamic patterning (Savard et al., 2006). In *D. melanogaster*, *mlpt* expression was correlated with tissue folding, acting as a connection between patterning and morphogenesis (Galindo et al., 2007). Early expression of *D. melanogaster mlpt* starts as seven blastodermal stripes and a cluster of cells in the anterior part of the embryo, although segmentation is not affected in *mlpt* mutants (Galindo et al., 2007). Later, after segmentation, *mlpt* is present in the trachea, posterior spiracles, pharynx, hindgut, and presumptive denticle belts. *D. melanogaster mlpt* mutants display reduced cuticular structures and ectopic expression of *mlpt* in the head induces extra skeleton components (Galindo et al., 2007). In hemipterans, *O. fasciatus* and *G. buenoi mlpt* is expressed in an anterior domain at early stages, during germ band elongation through recently added segments anterior to the growth zone, and later expression is observed in presumptive neurons in the central nervous system, as well as in the limb buds and mouth parts (Ray et al., 2019).

To investigate *mlpt* expression in *R. prolixus* during embryogenesis, *in situ* hybridization of several embryonic stages was performed. The original data of *Rp-mlpt* expression using a colorimetric substrate and DAPI stainings are provided in Supplementary Figure 5, while, for simplicity, overlay images of pseudo-fluorescence expression and DAPI stainings are provided in Figure 3. *Rp-mlpt* distinct expression can be observed at stage 1B, between 6 and 12 h after egg laying (AEL), in an anterior domain (Figure 3A), which corresponds to the presumptive head



eggs (Supplementary Figure 6A). Altogether, these RT-qPCR data show that *Rp-mlpt* expression is higher during zygotic segmentation stages than maternally deposited mRNAs.

Rp-mlpt Parental RNA Interference Is Efficient and Leads to a Series of Knockdown Phenotypes

Since *Rp-mlpt* gene displays a complex expression patterning during embryogenesis, we sought to analyze its function by parental RNA interference (pRNAi) in *R. prolixus*, as previously described (Berni et al., 2014). In this method, females are injected with double-stranded (dsRNA) synthesized against the gene of interest and phenotypic effects are evaluated in the offspring. *Rp-mlpt* dsRNA eggs showed a decrease up to 60–70% in *Rp-mlpt* expression when compared to its expression in the control dsRNA group, validating the knockdown (Supplementary Figure 6B). An increase in *Rp-mlpt* dsRNA injection concentration up to six micrograms per microliter (6 $\mu\text{g}/\mu\text{l}$) did not further increase knockdown efficiency (data not shown).

Upon egg collection and fixation, nuclear DAPI staining was performed and a plethora of embryonic mutant phenotypes were observed. A first phenotypic class of *Rp-mlpt* RNAi embryos comprises embryos with a wild-type number of segments, but with thoracic or gnathal segmental fusion. Embryos with this morphology were evident at stage 5 and 7 (Figures 4B,B',C,C'). A second phenotypic class was constituted by embryos with four pairs of legs in the *Rp-mlpt* RNAi instead of the three pairs observed in controls (Figures 4E,E'). Last, a third phenotypic class of *Rp-mlpt* RNAi embryos showed improper germ band elongation and lack of the posterior region at stage 3 (Figures 4A,A'), which is more evident at stage 9, when a lack of abdominal segmentation and fused legs were observed (Figures 4D,D').

These phenotypic classes can also be observed by dissection of embryos at very late stages when nuclear DAPI staining is not possible due to cuticle formation (Figures 5A–D). Thoracic red pigmentation can be used to compare wild-type and RNAi embryos. In the first *Rp-mlpt* RNAi phenotypic class (20 out of 45 embryos—45%), embryos with two pairs of thoracic segments can be identified, indicating improper thoracic segmentation (Figures 5B–B'''). In the second class (15 out of 45 embryos—33%), four thoracic segments are present (Figures 5C–C'''), while in the third class (10 out of 45 embryos—22%), the strongest knockdown phenotype, the late embryo displays impairment of germ band elongation and most abdominal segmentation is lost, while dorsal closure shows clear defects (Figures 5D–D'''). We noticed that injection of higher concentrations of *Rp-mlpt* dsRNA increased the frequency of the third phenotypic class and the observation of head defects that were not identified in any of the previously described classes (see below).

Interestingly, injection of *Rp-mlpt* dsRNA at lower concentrations (up to 1 $\mu\text{g}/\mu\text{L}$) into *R. prolixus* females lead to a small percentage of nymphs, which are able to leave the eggshell (10–15%). A few of these *Rp-mlpt* dsRNA

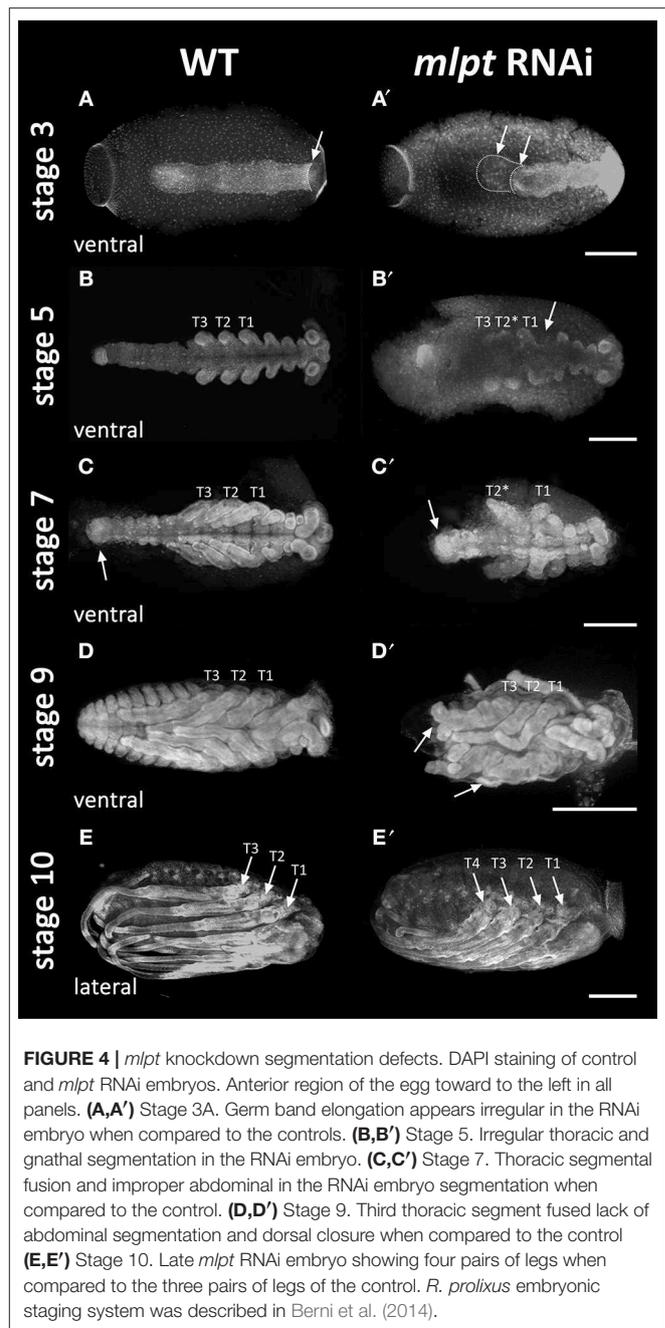


FIGURE 4 | *mlpt* knockdown segmentation defects. DAPI staining of control and *mlpt* RNAi embryos. Anterior region of the egg toward the left in all panels. (A,A') Stage 3A. Germ band elongation appears irregular in the RNAi embryo when compared to the controls. (B,B') Stage 5. Irregular thoracic and gnathal segmentation in the RNAi embryo. (C,C') Stage 7. Thoracic segmental fusion and improper abdominal in the RNAi embryo segmentation when compared to the control. (D,D') Stage 9. Third thoracic segment fused lack of abdominal segmentation and dorsal closure when compared to the control (E,E') Stage 10. Late *mlpt* RNAi embryo showing four pairs of legs when compared to the three pairs of legs of the control. *R. prolixus* embryonic staging system was described in Berni et al. (2014).

nymphs display unilateral segmental fusion of first and second thoracic segments leading to abnormal distal leg morphology (Figures 6A,B). Distal regions of legs in wild-type *R. prolixus* are composed of a stereotypic pattern of tibia, tarsus and a pair of claws (Figure 6A). dsRNA *Rp-mlpt* nymphs display improper cuticle darkening and different types of leg deviations from the wild-type patterning, including lack of the proper distinction between tarsus and claws (Figures 6B',B''). Altogether, our results show that *Rp-mlpt* is important for the distinction between thoracic and abdominal segments, for the process of germ band elongation and distal leg patterning.

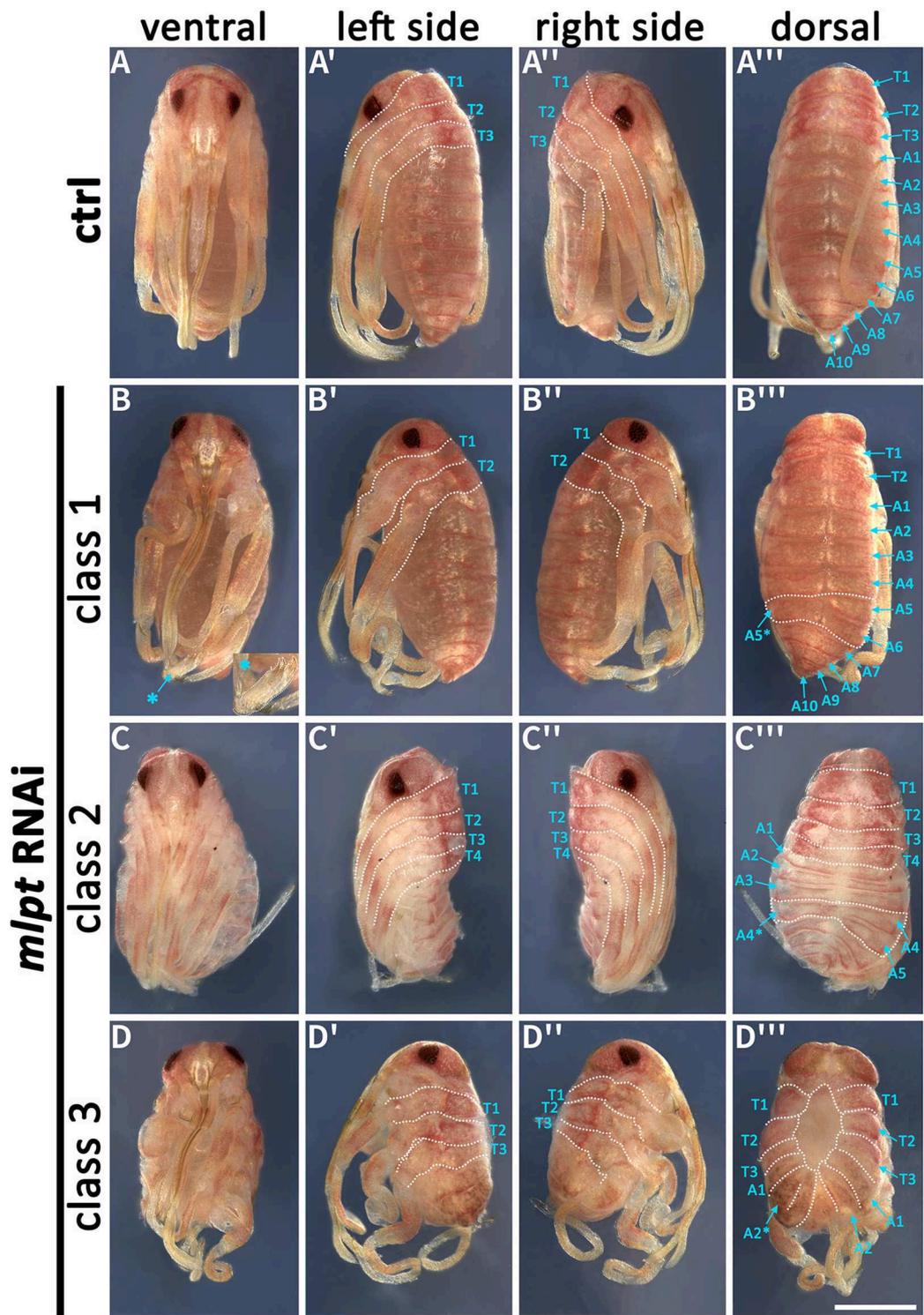
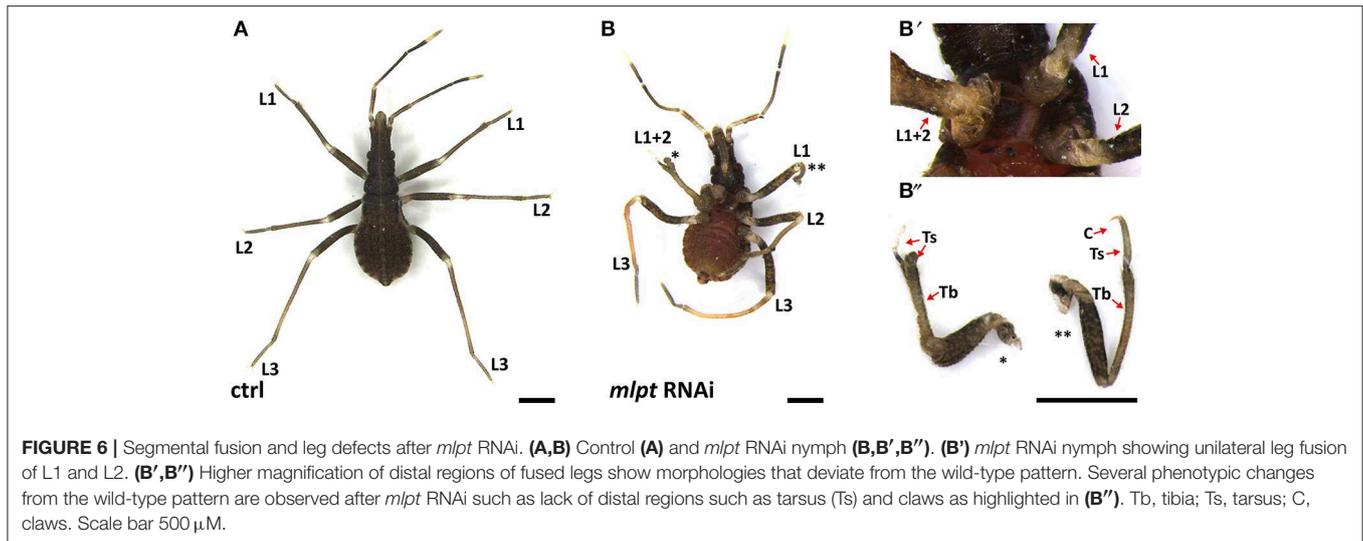


FIGURE 5 | Phenotypic classes of *m/pt* RNAi late embryos. *m/pt* knockdown embryos were classified in three phenotypical classes when compared to the control (**A-A'''**). (**B-B'''**) Class 1—Embryos displaying only two thoracic segments and distal leg duplications (asterisk) (**B-D**). Abdominal segmentation appears normal with the exception of some asymmetrical segmental divisions on left and right sides (asterisk) (20 out of 45 embryos). (**C-C'''**) Class 2—Embryos displaying four thoracic segments and irregular abdominal segmentation (15 out of 45 embryos). (**D-D'''**) Class 3—Shorter embryos only displaying thoracic segments and fewer abdominal segments than controls (10 out of 45 embryos).



Rp-mlpt dsRNA Injection Leads to Segmental Fusion and Impairment of Head Formation in the Strongest Phenotypes

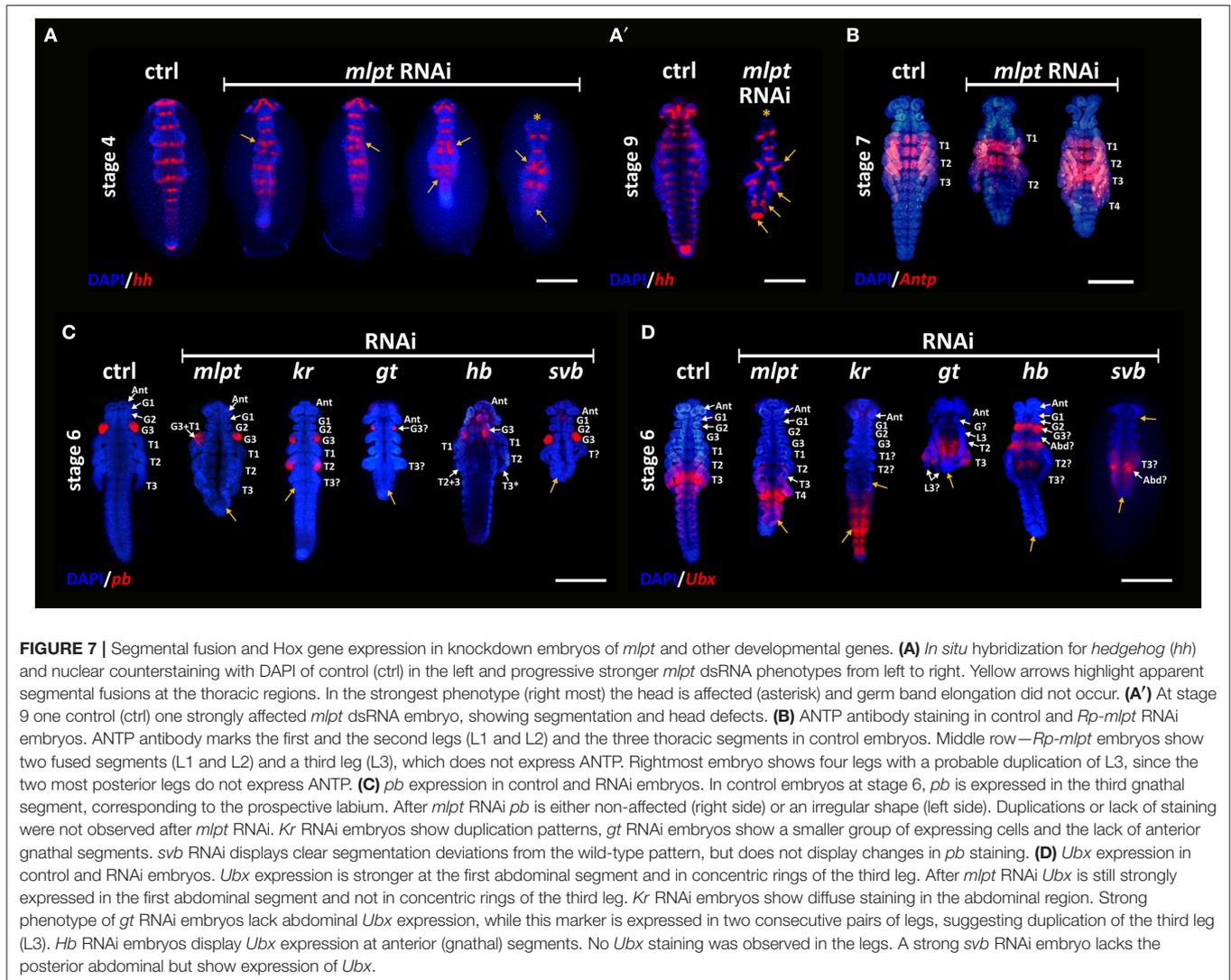
Reduction in *Rp-mlpt* expression affected segment formation with clear phenotypic effects in the thorax and abdomen, before posterior invagination and at later segmentation stages (Figures 4, 5). Since a reduction or increase in the number of thoracic segments was observed in late embryonic stages after *Rp-mlpt* RNAi, we evaluated the expression of the ortholog of the segment polarity gene *hedgehog* (*hh*) in control and *Rp-mlpt* RNAi embryos. Like in other insects, the posterior region of every segment expresses *Rp-hh* during embryonic segmentation, and thoracic and abdominal segments can be identified by their differential widths (Figure 7A). A phenotypic series of *Rp-mlpt* RNAi embryos shows that the distance between segmental stripes is reduced and that the distinction between thoracic and abdominal segments is less clear. Remarkably, in apparently stronger knockdown phenotypes, a lack of the anterior-most segments of the head was also observed (Figure 7A, asterisk). *hh* stainings of these strongest *mlpt* phenotypes at stage 9 confirms that the head is affected at later stages (Figure 7A'). Phenotypic defects in the head have only been observed upon injection of higher dsRNA concentrations (four to 6 μ g per microliter) and in a low frequency (\sim 10% of the embryos). This strongest phenotype cannot be attributed to any of the previously described three classes. Altogether, *Rp-hh* expression analysis shows that *Rp-mlpt* is essential for proper thoracic and abdominal segmental distinction and for head formation.

Comparison of Knockdown Phenotypes of *mlpt* and Other Developmental Genes

Previous functional analysis of *mlpt* in the *T. castaneum* and *D. melanogaster* showed that this gene is required for early patterning in the beetle but not in fruit flies. In beetles, *mlpt* acts as a gap gene, while its role in fruit flies is restricted to later (post-segmental) embryonic stages (Savard et al., 2006; Galindo et al.,

2007; Kondo et al., 2007, 2010; Ray et al., 2019). Recent analysis of *mlpt* function in two other hemipteran species and in a wasp showed a requirement for early patterning processes, suggesting that this is the ancestral function of the gene (Ray et al., 2019). From our phenotypic analysis, it remains open whether the knockdown of *Rp-mlpt* leads to similar morphological changes as the knockdown of classical gap genes, the transcription factors *Krüppel* (*Kr*), *giant* (*gt*), and *hunchback* (*Hb*). We performed parental RNAi against the orthologs of these three gap genes and analyzed the expression of Hox genes. *Krüppel* (*Kr*) and *giant* (*gt*) RNAi phenotypes have been published elsewhere (Lavore et al., 2012, 2014). Expression of the Hox gene *proboscipedia* (*pb*) in the third gnathal segment is comparable in control and *mlpt* RNAi embryos, although the posterior region in the knockdown has been clearly affected (Figure 7C). In contrast, *pb* expression is duplicated after *Rp-kr* RNAi, largely diminished after *Rp-gt* RNAi and slightly affected upon *Rp-Hb* RNAi (Figure 7C).

The expression of the Hox gene *Ultrabithorax* (*Ubx*) was also evaluated in control and in knockdowns of *mlpt* and other gap genes. In *R. prolixus* control embryos, *Ubx* is strongly expressed in the first abdominal segment and in concentric rings around the third leg (L3) (Figure 7C). As previously described, the effect of *mlpt* RNAi can generate at least three classes of embryonic phenotypes. Analysis of an embryo showing four pairs of legs (class 2 phenotype) showed that only the posterior-most leg expresses *Ubx*, suggesting that this leg corresponds to the L3 and that a duplication of L1 or L2 must have occurred. Interestingly, the expression of *Ubx* in the L3 does not occur as a concentric ring as in the control, but rather in one side of the legs, suggesting that leg patterning was also affected (Figure 7C). In addition, a slight expansion of *Ubx* toward posterior segments was observed in *mlpt* RNAi embryos (Figure 7C). *Ubx* expression was also analyzed upon knockdown of other gap genes, such as *Kr*, *gt*, and *hunchback* (*Hb*) pRNAi. *Kr* RNAi embryos show only two pairs of legs and lack of *Ubx* staining in the legs, suggesting that the third leg (L3) is absent. In addition, a large expansion of *Ubx* expression toward the posterior region in *Kr* RNAi embryos



was observed. *Rp-gt* RNAi embryos lack gnathal and abdominal segments and show a duplication of L3. *Rp-Hb* RNAi embryos show a large expansion of the posterior fates toward the anterior region, also pointing to a major role of this gap gene in *R. prolixus* (Figure 7D). Altogether, these results suggest that the segmentation role *Rp-mlpt* role in *R. prolixus* is rather limited and mainly concentrated at the transition between the thoracic and abdominal region and that classical gap genes, such as *kr*, *gt*, *hb*, display a broader role in the segmental cascade of the hemiptera *R. prolixus* (Figure 7). Further support for the hypothesis that *mlpt* function is mainly required for thoracic vs. abdominal distinction is provided by the immunohistochemical analysis using a monoclonal antibody against the homeobox transcription factor ANTP. ANTP demarcates the thoracic segments (T1-T3) and the leg one and two (L1 and L2) in the controls (Figure 7B). As in controls, ANTP is only expressed in L1 and L2 in *mlpt* dsRNA embryos with four legs (L1–L4), suggesting that L4 in these embryos might be a duplication of L3 (Ray et al., 2019).

Lastly, morphological analysis of the knockdown of the *mlpt* interaction partner, the transcription factor *svb* (Kondo et al.,

2010; Ray et al., 2019) was performed. As recently demonstrated by Ray et al. (2019) in two hemipteran species, *mlpt*, and *svb*, RNAi phenotypes appear remarkably similar. As observed for *mlpt* RNAi, *svb* RNAi did not show any alteration on the expression pattern of *pb* or *Ubx* either, although germ band elongation and the distinction between thoracic vs. abdominal segmentation was clearly affected (Figure 7C).

DISCUSSION

Nucleotide and Peptide Sequence Analysis Provides Insights Into the Evolution of the *mlpt* Gene Function

In the past years, genes encoding smORFs have been characterized as new and important players of an unknown part of animal genomes (Saghatelian and Couso, 2015; Zanet et al., 2015; Pueyo et al., 2016a). *mlpt* is the most widely studied polycistronic gene encoding smORFs in insects. Previous studies and our own sequence analysis presented here has shown that

mlpt is not present in chelicerate and myriapod genomes, being restricted to Pancrustacea.

Previous analyses identified specific roles for *mlpt* short peptides containing LDPTG(L/Q/T)Y motifs in the insect *D. melanogaster*, while the function of a second weakly conserved predicted peptide (ORF-B in Galindo et al., 2007), remains open since it is not translated in *D. melanogaster* cell culture nor functionally required in rescue experiments (Galindo et al., 2007; Kondo et al., 2007). Mechanistically, small peptides from *mlpt* containing LDPTG(L/Q/T)Y motifs are essential for a selective proteasome-mediated N-terminal processing and activation of the transcription factor *svb* (Kondo et al., 2010; Zanet et al., 2015). The new hemipteran specific peptide identified here (smHemiptera) is ~35% longer in *R. prolixus* than in the most basally branching species analyzed, *Pseudococcus longispinus* (Figures 1A,C). It has been recently proposed that smORFs could evolve to major ORFs via an “elongation” pattern (Couso and Patraquim, 2017). Although smHemiptera is not a major ORF, size distribution of the smHemiptera peptide in hemiptera phylogeny corroborates this hypothesis. Future studies should address the function of smHemiptera. It is unlikely that a predicted smORF peptide with a stretch of 14 amino acids with 100% identity over 240 million years would be non-functional. It must be considered that the pRNAi technique used in our study leads to the knockdown of the mature transcript and presumably of all predicted peptides encoded by the hemipteran *mlpt* gene. Cas9/CRISPR editing technology has been established in other non-model arthropod species (Gilles et al., 2015) and its establishment in *R. prolixus* might help to unveil the specific function of each small peptide present in *Rp-mlpt* transcript at different developmental stages. Recent studies demonstrated that *mlpt* is required for Svb activation in adult tissues, particularly in nephric and intestinal *D. melanogaster* stem cells (Al Hayek et al., 2019). Since *mlpt* is also expressed in *R. prolixus* digestive tract (Ribeiro et al., 2014 and our own observations), it will also be interesting to address the functional role of this gene containing smORFs in other developmental contexts.

To the best of our knowledge, this is the first study to provide a theoretical interaction model for the Svb-Mlpt complexes for *D. melanogaster* (Dmel) and *R. prolixus* (Rprol). Our predictions were able to capture previously published data for *D. melanogaster* (Zanet et al., 2015), even in the absence of experimentally validated 3D structures from *D. melanogaster* Svb. Secondary structure and fold for both Svb and Mlpts, and protein-binding sites on the N-terminal of Svbs were modeled and agree with (Zanet et al., 2015). Both Svbs (Dmel and Rprol) display around 80–90% conserved disordered regions, but the Zinc-finger double domain near the C-terminal were predicted to be more ordered (Supplementary Figure 1C). The results of the docking assays and binding affinity changes upon mutation on key residues at the interaction interface offer an interesting possibility to propose a theoretical interaction model for Svb-Mlpt of both species (Supplementary Figures 3, 4). Leucines (L24, L25, and L19) and Glutamic Acid (E20) at the N-terminal region of Dmel-Svb affects its interaction with Dmel-Mlpt1, amino acids

previously identified in the region of interaction between Dmel-Svb and Dmel_Mlpt1 (Zanet et al., 2015). Additionally, our results suggest that both organisms use at least half of LDPTG(L/Q/T)Y motif residues to anchor on Shavenbaby N-terminal region. All these *in silico* data might be useful for future NMR approaches to test these structural predictions and to test the effects of mutations in smORF peptides using Cas9/CRISPR technology.

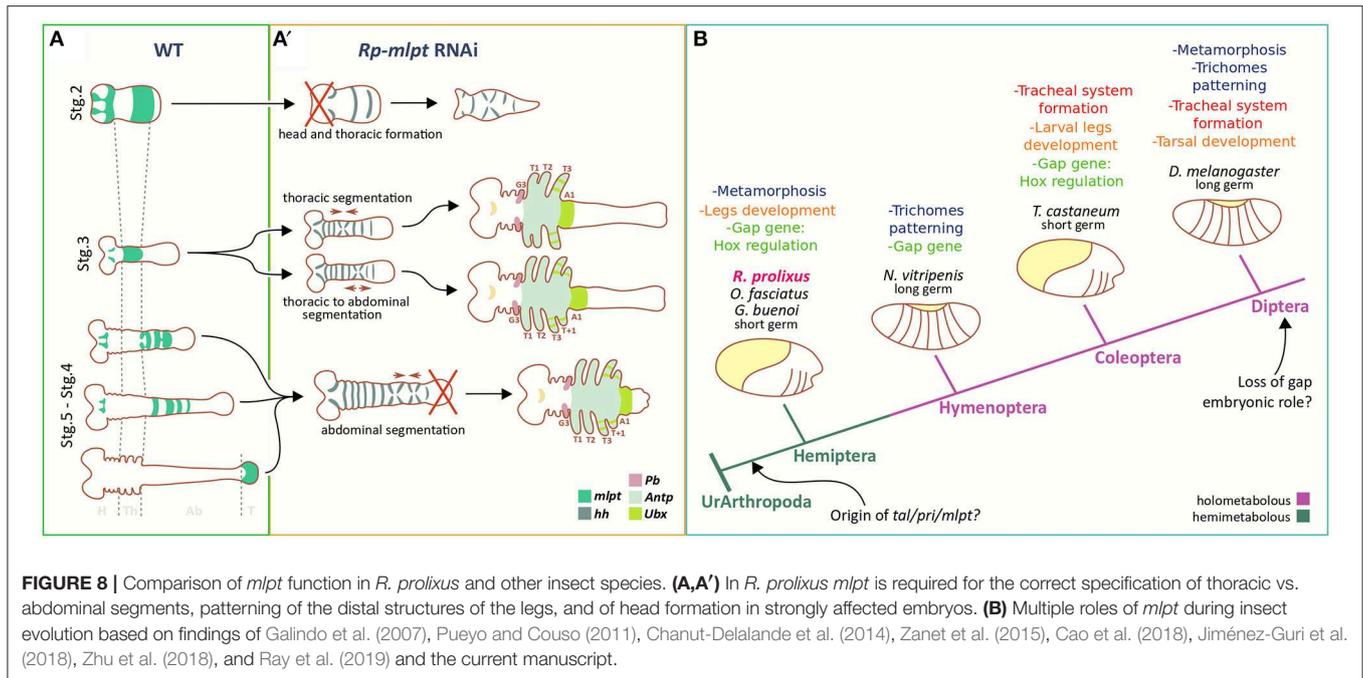
Conservation of *mlpt* Expression Among Insects

In situ hybridization analysis of the *Rp-mlpt* gene during embryonic development shows a complex pattern of spatial expression, similarly to previously reported expression patterns in beetles, fruit flies, and other holometabolous insects (Savard et al., 2006; Galindo et al., 2007; Cao et al., 2018; Jiménez-Guri et al., 2018; Zhu et al., 2018; Ray et al., 2019). Most differences in expression have been observed in early developmental stages. In *D. melanogaster*, the expression occurs in seven blastodermal stripes, a pair-rule pattern, while in most other insect species, including the basally branching Diptera midge *Clogmia albipunctata*, the expression appears as a gap domain (Jiménez-Guri et al., 2018). Thus, as observed for *T. castaneum* (Savard et al., 2006) and more recently in the hemiptera *O. fasciatus* and *G. buenoi* (Ray et al., 2019), *Rp-mlpt* is first expressed in an anterior domain overlapping with the head and shortly after expression in a posterior domain overlapping with the prospective thoracic segments (Figure 3). Later on, during germ band elongation and dorsal closure, the dynamic expression of *Rp-mlpt* is observed in several tissues including the tips of the legs, the head and the antenna (Figure 3). While expression in the head and in the legs have been reported in other species (Ray et al., 2019), expression in the antennae at later stages has, to the best of our knowledge, only been reported here and recently in another hemipteran species *O. fasciatus* (Ray et al., 2019).

Comparison of *Rp-mlpt* expression by RT-PCR between fertilized and non-fertilized eggs during the first hours of development (0–6 h AEL) and 36–48 h AEL shows 10 times higher expression at later stages when compared to freshly laid or non-fertilized eggs (0–6 h AEL) (Supplementary Figure 6). Although the developmental time of the process of maternal-zygotic transition in *R. prolixus* is unknown, these results suggest that large changes in *mlpt* transcriptional levels occurs zygotically, at later stages (36–48 h AEL). It is possible that the translated Mlpt peptides might post-transcriptionally switch Svb from a repressor to an activator at 36–48 h AEL, as previously reported for *D. melanogaster* (Kondo et al., 2010; Zanet et al., 2015).

mlpt Functions Not as a Classical Gap Gene in *R. prolixus* but Rather Participates in Thoracic vs. Abdominal Segmental Identity

Three phenotypic classes were observed in *mlpt* knockdown embryos, mainly affecting segment formation (Figures 4, 5). Comparison of the morphological defects of *mlpt* knockdown



embryos with the defects of transcription factor gap gene knockdowns, such as *Kr*, *Hb*, and *Gt* (Figure 7), showed clear differences (Figure 7). *mlpt* knockdown embryos display localized phenotypic changes mainly at the transition between thoracic and abdominal segments, while knockdown of the aforementioned transcription factors shows larger effects and changes in molecular marker expression (Figure 7). Our results can be reconciled with several recently published studies about *mlpt* in other insect orders (Figures 8A,B). In *T. castaneum* *mlpt* knockdown leads to extensive homeotic transformations and embryos show up to six leg pairs (Savard et al., 2006). Recent data in two other hemipterans showed that knockdown of three genes *mlpt*, *Ubr3*, and *svb* shows similar phenotypic effects as we describe here for *R. prolixus* such as posterior truncation, with the fusion/loss of thoracic segments, shortened legs and head appendages (Ray et al., 2019). While in *R. prolixus*, *mlpt* RNAi embryos with four pairs of legs were observed, *O. fasciatus* and *G. buenoi* knockdown embryos show thoracic segmental fusion, and an increase in the number of thoracic segments was not reported. Since ectopic segment formation after *mlpt* RNAi occurs in a low frequency in *R. prolixus*, it is possible that it also occurs at low frequency in *O. fasciatus* and *G. buenoi* *mlpt* knockdown. It is intriguing that, in these three species, *mlpt* is mainly required when germ band elongation and abdominal segmentation takes place and future studies using embryonic live imaging, as recently described for other insect species (Benton, 2018), might help to understand the role of *mlpt* during germ band elongation and thoracic vs. abdominal segment specification.

Finally, several distal duplications of leg segments and leg malformations have been observed after *Rp-mlpt* RNAi (Figure 6). While several leg defects might be attributed to thoracic segment fusion, at least some examples of distal duplications are likely generated by local effects of *Rp-mlpt* in the

legs. We did not detect multiple rings of *Rp-mlpt* expression in *R. prolixus*; however, such rings were observed in *Periplaneta* and also in other hemipteran species (Chesebro, 2013). In summary, our data is compatible with a model that *mlpt* is required during early stages of *R. prolixus* development for the distinction between thoracic and abdominal segmentation, during germ band elongation and for head formation (Figure 8A). Our data is important for defining the multiple roles of *mlpt* in *R. prolixus* embryogenesis, particularly during early patterning. Since recent analysis in other two hemipteran species, *O. fasciatus*, and *G. buenoi*, demonstrated a similar role of *mlpt* and its interaction partners for early developmental processes (Ray et al., 2019), one can infer that, in the common ancestor of Paraneoptera, *mlpt* already displayed multiple roles (Figure 8B). Most of these roles of *mlpt* have been described in Diptera and Coleoptera, and it remains open if all of them are conserved in hemimetabolous insects, as well as its ancestral function in insects (Figure 8B). Lastly, our data contribute to the study of *mlpt* evolution, providing the interesting finding of a new Hemipteran conserved peptide whose function should be addressed by future functional studies.

MATERIALS AND METHODS

Bioinformatic Analyses

Mlpt peptide sequences from *D. melanogaster* and *T. castaneum* were used for BLAST searches (Altschul et al., 1997) against available arthropod genomes and transcriptomes using relaxed parameters to maximize the chances to obtain genes encoding smORFs. The species used for the smORF identification are depicted in Supplementary Table 1. *Rp-mlpt/tal/pri* was identified in *R. prolixus* genome and transcriptomes (Medeiros et al., 2011; Ribeiro et al., 2014; Mesquita et al., 2015; Marchant

et al., 2016). BLAST results and domain architectures were manually annotated.

Structural Modeling and Shavenbaby-mlpts Docking Assays

The 3D models of Dmel-Svb (1351 aa; Uniprot: P51521-1) and Rprol-Svb (834 aa; VectorBase: GECK01059228) were predicted via *ab initio* modeling using the I-Tasser program (Yang et al., 2015). The 3D structures of Mlpts (Dmel-pptd1 and 4) were predicted via CABS-dock (Kurcinski et al., 2015) and I-Tasser, respectively; Rprol-pptd1 and 2, both via CABS-dock. The global and local stereochemical quality of all predicted models were performed by MolProbity (Chen et al., 2010), Verify 3D (Eisenberg et al., 1997), ProSA (Wiederstein and Sippl, 2007), VoroMQA (Olechnovic and Venclovas, 2017), ProQ3D (Uziela et al., 2017), Qprob (Cao and Cheng, 2016), DeepQA (Cao et al., 2016), and SVMQA (Manavalan and Lee, 2017). The most suited structural models were refined by ModRefiner (Xu and Zhang, 2011), 3Drefine (Bhattacharya et al., 2016), GalaxyRefine (Heo et al., 2013), and FG-MD (Zhang et al., 2011). The protein-peptide docking assays were directed to the Shavenbaby N-terminal region (first 31 aa) inspired by the findings of Zanet et al. (2015). The secondary structures, disordered regions and peptide-binding sites predictions on Shavenbabys were further analyzed by PSIPRED (Jones, 1999), Porter (Mirabello and Pollastri, 2013), SPOT-Disorder (Hanson et al., 2017), and DISOPRED3 (Jones and Cozzetto, 2015), respectively. The dockings were performed by HADDOCK (van Zundert et al., 2016). The best interaction complexes were selected after exhausting analyses with DockScore (Malhotra et al., 2015), PPCheck (Sukhwai and Sowdhamini, 2015), and CCharPPI (Moal et al., 2015) programs. After selection of the best complex, further refinements were performed via GalaxyRefineComplex (Heo et al., 2016). The refinements were completed using the “Energy minimization of side chains” function on PDB_Hydro webserver (Azuara et al., 2006). The binding free energy prediction (ΔG) using MM/GBSA method was carried out by HawkDock (Weng et al., 2019), dissociation constant (K_d) prediction was carried out by PRODIGY (Xue et al., 2016), and the analysis of hotspot amino acids located in the interaction interface were performed by HotRegion (Cukuroglu et al., 2012) and ANCHOR (Meireles et al., 2010), respectively. Calculations on the change in the binding energy of the protein-protein complex upon mutations in the residues on interaction interface were performed by BindProfX (Xiong et al., 2017). The electrostatic calculations were carried out by PDB2PQR and APBS (Dolinsky et al., 2004; Unni et al., 2011). Interatomic contacts and types of interactions maintained at Svbs-MLTPs interfaces were analyzed by PDBsum (Laskowski et al., 2018). The images of Svbs-MLTPs complexes were obtained by UCSF Chimera 1.13.1 (Pettersen et al., 2004).

Insect Rearing, Fixation and Dissection

Insect rearing was performed as described by Souza-Ferreira et al. (2014). Approximately 1 week after blood-feeding, eggs were collected daily and fixed in different stages of development. For fixation, up to 100 eggs were briefly washed with distilled

water to remove the debris and then transferred to a 1.5 mL microtube containing 1 mL of distilled water. This microtube was maintained for 90 s in boiling water, and after this period, the water was replaced by 1 mL of paraformaldehyde 12% (PBS) and fixed for 1 h (6–8°C). Then, the embryos were incubated with 1 mL of paraformaldehyde 4% containing 0.1% of Tween 20 under agitation (200 rpm) for 1 h at room temperature. The eggs were then washed repeatedly with PBST (PBS 1X, Tween 20 0.1%). For long-term storage, embryos were gradually transferred to ethanol 100% and then stored at –20°C. For the dissections, two forceps (Dumont No. 5) were used. The egg is held with the help of one of them, while the other is used to apply pressure at the chorionic rim to remove the operculum. The chorionic rim is held and the shell is opened transversely, leading to the embryonic release. For later stages, when segmentation is complete, at least stage 6, the yolk is easily removed, either with the help of a thin forceps or a glass needle, without damaging the embryo.

In situ Hybridization, Nuclear, and Antibody Staining

Embryos stored in ethanol 100% at –20°C were gradually transferred to PBST at room temperature. *In situ* hybridization and probe synthesis were performed as described by (Sachs et al., 2015), including the proteinase K treatment. DAPI (4',6-Diamidino-2'-phenylindole dihydrochloride, SIGMA) staining was performed as in Berni et al. (2014). Antibody staining followed the protocol described in Santos et al. (2013). All the images were acquired with the stereoscope Leica M205, and processed and analyzed with the software Leica Application Suite Advanced Fluorescence Version 0.4 (LAS AF v4—Leica Microsystems). Images were assembled and the *in situ* hybridization staining (NBT/BCIP) was converted to a false fluorescence as described in Benton et al. (2016).

RNA Interference and Real-Time PCR

The RNA interference (RNAi) was performed similarly to Berni et al. (2014) using a non-related dsRNA as a control (neomycin dsRNA-dsneo). For *Rp-mlpt*, two non-overlapping PCR fragments containing T7 promoter initiation sites at both ends were used as templates for dsRNA synthesis using Ambion T7 Megascript Kit (Cat. No. AM1334). The amount and integrity of the dsRNA samples were measured by spectrophotometry and agarose gel electrophoresis, respectively. For each silenced gene, between 6 and 12 μ g of dsRNA were injected. The quantification of RNAi efficiency and comparison of gene expression after silencing was measured through real-time PCR (RT-PCR) as performed in Berni et al. (2014) using the gene Elongation factor-1 (Efl) as endogenous reference gene. Gene bank or Vector base accession numbers of the genes analyzed by *in situ* hybridization or RT-PCR are provided in the **Supplementary Table 4**. For RT-PCR analysis, RNA was extracted in biological triplicates from eggs at appropriate hours after egg lay (AEL). Fertilized eggs (0–6 and 36–48 h AEL) and non-fertilized eggs obtained from female virgins (0–6 h AEL) were used for the analysis.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI.

AUTHOR CONTRIBUTIONS

VT-S, DG-A, FM, LR, MB, HA, CL, NF, JS-M, EP, and RN: experimental data generation and interpretation. HA, CL, and RN: funding acquisition. VT-S and RN: manuscript draft.

FUNDING

RN was supported by CNPq (307952/2017-7 and 431354/2016-2) and FAPERJ (E-26/210-150/2016 and E-26/203.298/2016). VT-S and DG-A were master students of PPG-PRODBIO-Macaé

REFERENCES

- Al Hayek, S., Alsawadi, A., Kambris, Z., Boquete, J.-P., Bohère, J., Ronsin, B., et al. (2019). Shavenbaby protein isoforms orchestrate the self-renewal versus differentiation of Drosophila intestinal stem cells. *bioRxiv*. doi: 10.1101/627554
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Aspden, J. L., Eyre-Walker, Y. C., Phillips, R. J., Amin, U., Mumtaz, M. A., Brocard, M., et al. (2014). Extensive translation of small open reading frames revealed by Poly-Ribo-Seq. *eLife* 3:e03528. doi: 10.7554/eLife.03528
- Auman, T., Vreede, B. M. I., Weiss, A., Hester, S. D., Williams, T. A., Nagy, L. M., et al. (2017). Dynamics of growth zone patterning in the milkweed bug *Oncopeltus fasciatus*. *Development* 144, 1896–1905. doi: 10.1242/dev.142091
- Azuara, C., Lindahl, E., Koehl, P., Orland, H., and Delarue, M. (2006). PDB_Hydro: incorporating dipolar solvents with variable density in the Poisson-Boltzmann treatment of macromolecule electrostatics. *Nucleic Acids Res.* 34, W38–42. doi: 10.1093/nar/gkl072
- Benton, M. A. (2018). A revised understanding of Tribolium morphogenesis further reconciles short and long germ development. *PLoS Biol.* 16:e2005093. doi: 10.1371/journal.pbio.2005093
- Benton, M. A., Pechmann, M., Frey, N., Stappert, D., Conrads, K. H., Chen, Y. T., et al. (2016). Toll genes have an ancestral role in axis elongation. *Curr. Biol.* 26, 1609–1615. doi: 10.1016/j.cub.2016.04.055
- Berni, M., Fontenele, M. R., Tobias-Santos, V., Caceres-Rodrigues, A., Mury, F. B., Vionette-do-Amaral, R., et al. (2014). Toll signals regulate dorsal-ventral patterning and anterior-posterior placement of the embryo in the hemipteran *Rhodnius prolixus*. *Evodevo* 5:38. doi: 10.1186/2041-9139-5-38
- Bhattacharya, D., Nowotny, J., Cao, R., and Cheng, J. (2016). 3Drefine: an interactive web server for efficient protein structure refinement. *Nucleic Acids Res.* 44, W406–W409. doi: 10.1093/nar/gkw336
- Brito, T., Julio, A., Berni, M., de Castro Poncio, L., Bernardes, E. S., Araujo, H., et al. (2018). Transcriptomic and functional analyses of the piRNA pathway in the chagas disease vector *Rhodnius prolixus*. *PLoS Negl Trop Dis* 12:e0006760. doi: 10.1371/journal.pntd.0006760
- Cao, G., Gong, Y., Hu, X., Zhu, M., Liang, Z., Huang, L., et al. (2018). Identification of tarsal-less peptides from the silkworm *Bombyx mori*. *Appl. Microbiol. Biotechnol.* 102, 1809–1822. doi: 10.1007/s00253-017-8708-4
- Cao, R., Bhattacharya, D., Hou, J., and Cheng, J. (2016). DeepQA: improving the estimation of single protein model quality with deep belief networks. *BMC Bioinformatics* 17:495. doi: 10.1186/s12859-016-1405-y
- Cao, R., and Cheng, J. (2016). Protein single-model quality assessment by feature-based probability density functions. *Sci. Rep.* 6:23990. doi: 10.1038/srep23990
- Chanut-Delalande, H., Hashimoto, Y., Pelissier-Monier, A., Spokony, R., Dib, A., Kondo, T., et al. (2014). Pri peptides are mediators of ecdysone for the temporal control of development. *Nat. Cell Biol.* 16, 1035–1044. doi: 10.1038/ncb3052
- (CAPES scholarships), MB a Ph.D. student from PPG-PCM-ICB-UFRJ and LR was a postdoc with CAPES scholarship from the National Institute of Molecular Entomology/CNPq (INCT-EM).

ACKNOWLEDGMENTS

The authors thank Roland Zimm for helpful suggestions in the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fevo.2019.00379/full#supplementary-material>

- Chekulaeva, M., and Rajewsky, N. (2018). Roles of long noncoding RNAs and circular RNAs in translation. *Cold Spring Harb. Perspect. Biol.* 11:a032680. doi: 10.1101/cshperspect.a032680
- Chen, V. B., Arendall, W. B. III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol. Crystallogr.* 66, 12–21. doi: 10.1107/S0907444909042073
- Chesebro, J. (2013). *Mechanisms of segmentation in the American cockroach, Periplaneta americana* (Doctoral thesis, PhD). Sussex: University of Sussex.
- Couso, J. P., and Patraquim, P. (2017). Classification and function of small open reading frames. *Nat. Rev. Mol. Cell Biol.* 18, 575–589. doi: 10.1038/nrm.2017.58
- Cukuroglu, E., Gursoy, A., and Keskin, O. (2012). HotRegion: a database of predicted hot spot clusters. *Nucleic Acids Res.* 40, D829–D833. doi: 10.1093/nar/gkr929
- Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32, W665–W667. doi: 10.1093/nar/gkh381
- Eisenberg, D., Lüthy, R., and Bowie, J. U. (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 277, 396–404. doi: 10.1016/S0076-6879(97)77022-8
- Galindo, M. I., Pueyo, J. I., Fouix, S., Bishop, S. A., and Couso, J. P. (2007). Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* 5:e106. doi: 10.1371/journal.pbio.0050106
- Gilles, A. F., Schinko, J. B., and Averof, M. (2015). Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle *Tribolium castaneum*. *Development* 142, 2832–2839. doi: 10.1242/dev.125054
- Hanson, J., Yang, Y., Paliwal, K., and Zhou, Y. (2017). Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics* 33, 685–692. doi: 10.1093/bioinformatics/btw678
- Heo, L., Lee, H., and Seok, C. (2016). GalaxyRefineComplex: refinement of protein-protein complex model structures driven by interface repacking. *Sci. Rep.* 6:32153. doi: 10.1038/srep32153
- Heo, L., Park, H., and Seok, C. (2013). GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* 41, W384–W388. doi: 10.1093/nar/gkt458
- Jiménez-Guri, E., Wotton, K. R., and Jaeger, J. (2018). tarsal-less is expressed as a gap gene but has no gap gene phenotype in the moth midge *Clogmia albipunctata*. *R. Soc. Open Sci.* 5:180458. doi: 10.1098/rsos.180458
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202. doi: 10.1006/jmbi.1999.3091
- Jones, D. T., and Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* 31, 857–63. doi: 10.1093/bioinformatics/btu744

- Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S., and Kageyama, Y. (2007). Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat. Cell Biol.* 9, 660–665. doi: 10.1038/ncb1595
- Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., et al. (2010). Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science* 329, 336–339. doi: 10.1126/science.1188158
- Kumar, S., Stecher, G., Suleski, M., and Heddes, S. B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34, 1812–1819. doi: 10.1093/molbev/msx116
- Kurcinski, M., Jamroz, M., Blaszczak, M., Kolinski, A., and Kmiecik, S. (2015). CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res.* 43, W419–W424. doi: 10.1093/nar/gkv456
- Ladoukakis, E., Pereira, V., Magny, E. G., Eyre-Walker, A., and Couso, J. P. (2011). Hundreds of putatively functional small open reading frames in *Drosophila*. *Genome Biol.* 12:R118. doi: 10.1186/gb-2011-12-11-r118
- Laskowski, R. A., Jablonska, J., Pravda, L., Varekova, R. S., and Thornton, J. M. (2018). PDBsum: structural summaries of PDB entries. *Protein Sci.* 27, 129–134. doi: 10.1002/pro.3289
- Lavore, A., Esponda-Behrens, N., Pagola, L., and Rivera-Pomar, R. (2014). The gap gene *Kruppel* of *Rhodnius prolixus* is required for segmentation and for repression of the homeotic gene *sex comb-reduced*. *Dev. Biol.* 387, 121–129. doi: 10.1016/j.ydbio.2013.12.030
- Lavore, A., Pagola, L., Esponda-Behrens, N., and Rivera-Pomar, R. (2012). The gap gene giant of *Rhodnius prolixus* is maternally expressed and required for proper head and abdomen formation. *Dev. Biol.* 361, 147–155. doi: 10.1016/j.ydbio.2011.06.038
- Magny, E. G., Pueyo, J. I., Pearl, F. M., Cespedes, M. A., Niven, J. E., Bishop, S. A., et al. (2013). Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* 341, 1116–1120. doi: 10.1126/science.1238802
- Malhotra, S., Mathew, O. K., and Sowdhamini, R. (2015). DOCKSCORE: a webserver for ranking protein-protein docked poses. *BMC Bioinformatics* 16:127. doi: 10.1186/s12859-015-0572-6
- Manavalan, B., and Lee, J. (2017). SVMQA: support-vector-machine-based protein single-model quality assessment. *Bioinformatics* 33, 2496–2503. doi: 10.1093/bioinformatics/btx222
- Marchant, A., Mougél, F., Mendonça, V., Quartier, M., Jacquín-Joly, E., da Rosa, J. A., et al. (2016). Comparing de novo and reference-based transcriptome assembly strategies by applying them to the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.* 69, 25–33. doi: 10.1016/j.ibmb.2015.05.009
- Medeiros, M. N., Logullo, R., Ramos, I. B., Sorgine, M. H., Paiva-Silva, G. O., Mesquita, R. D., et al. (2011). Transcriptome and gene expression profile of ovarian follicle tissue of the triatomine bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.* 41, 823–831. doi: 10.1016/j.ibmb.2011.06.004
- Meireles, L. M., Dömling, A. S., and Camacho, C. J. (2010). ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery. *Nucleic Acids Res.* 38, W407–W411. doi: 10.1093/nar/gkq502
- Mesquita, R. D., Vionette-Amaral, R. J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F. A., Minx, P., et al. (2015). Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14936–14941. doi: 10.1073/pnas.1506226112
- Mirabello, C., and Pollastri, G. (2013). Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics* 29, 2056–2058. doi: 10.1093/bioinformatics/btt344
- Moal, I. H., Jiménez-García, B., and Fernández-Recio, J. (2015). CCharPPI web server: computational characterization of protein-protein interactions from structure. *Bioinformatics* 31, 123–125. doi: 10.1093/bioinformatics/btu594
- Nunes-da-Fonseca, R., Berni, M., Tobias-Santos, V., Pane, A., and Araujo, H. M. (2017). *Rhodnius prolixus*: from classical physiology to modern developmental biology. *Genesis* 55, 1–11. doi: 10.1002/dvg.22995
- Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287, 795–801. doi: 10.1038/287795a0
- Olechnov, K., and Venclovas, C. (2017). VoroMQA: assessment of protein structure quality using interatomic contact areas. *Proteins* 85, 1131–1145. doi: 10.1002/prot.25278
- Paese, C. L. B., Schoenauer, A., Leite, D. J., Russell, S., and McGregor, A. P. (2018). A SoxB gene acts as an anterior gap gene and regulates posterior segment addition in a spider. *eLife* 7:e37567. doi: 10.7554/eLife.37567
- Panfili, K. A. (2008). Extraembryonic development in insects and the acrobatics of blastokinesis. *Dev. Biol.* 313, 471–91. doi: 10.1016/j.ydbio.2007.11.004
- Panfili, K. A., and Angelini, D. R. (2018). By land, air, and sea: hemipteran diversity through the genomic lens. *Curr. Opin. Insect Sci.* 25, 106–115. doi: 10.1016/j.cois.2017.12.005
- Panfili, K. A., Vargas Jentzsch, I. M., Benoit, J. B., Erezylmaz, D., Suzuki, Y., Colella, S., et al. (2019). Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Genome Biol.* 20:64. doi: 10.1186/s13059-019-1660-0
- Petersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084
- Pueyo, J. I., and Couso, J. P. (2011). Tarsal-less peptides control Notch signalling through the Shavenbaby transcription factor. *Dev. Biol.* 355, 183–193. doi: 10.1016/j.ydbio.2011.03.033
- Pueyo, J. I., Magny, E. G., and Couso, J. P. (2016a). New peptides under the s(ORF)ace of the Genome. *Trends Biochem. Sci.* 41, 665–678. doi: 10.1016/j.tibs.2016.05.003
- Pueyo, J. I., Magny, E. G., Sampson, C. J., Amin, U., Evans, I. R., Bishop, S. A., et al. (2016b). Hemotin, a regulator of phagocytosis encoded by a small ORF and conserved across metazoans. *PLoS Biol.* 14:e1002395. doi: 10.1371/journal.pbio.1002395
- Ray, S., Rosenberg, M. I., Chanut-Delalande, H., Decaras, A., Schwertner, B., Toubiana, W., et al. (2019). The mlpt/Ubr3/Svb module comprises an ancient developmental switch for embryonic patterning. *eLife* 8:e39748. doi: 10.7554/eLife.39748
- Ribeiro, J. M., Genta, F. A., Sorgine, M. H., Logullo, R., Mesquita, R. D., Paiva-Silva, G. O., et al. (2014). An insight into the transcriptome of the digestive tract of the bloodsucking bug, *Rhodnius prolixus*. *PLoS Negl. Trop. Dis.* 8:e2594. doi: 10.1371/journal.pntd.0002594
- Sachs, L., Chen, Y. T., Drechsler, A., Lynch, J. A., Panfili, K. A., Lässig, M., et al. (2015). Dynamic BMP signaling polarized by Toll patterns the dorsoventral axis in a hemimetabolous insect. *eLife* 4:e05502. doi: 10.7554/eLife.05502
- Saghatelian, A., and Couso, J. P. (2015). Discovery and characterization of smORF-encoded bioactive polypeptides. *Nat. Chem. Biol.* 11, 909–916. doi: 10.1038/nchembio.1964
- Santos, V. T., Ribeiro, L., Fraga, A., de Barros, C. M., Campos, E., Moraes, J., et al. (2013). The embryogenesis of the tick *Rhipicephalus* (Boophilus) microplus: the establishment of a new chelicerate model system. *Genesis* 51, 803–818. doi: 10.1002/dvg.22717
- Savard, J., Marques-Souza, H., Aranda, M., and Tautz, D. (2006). A segmentation gene in tribolium produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell* 126, 559–569. doi: 10.1016/j.cell.2006.05.053
- Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., et al. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 15:62. doi: 10.1186/s12915-017-0399-x
- Souza-Ferreira, P. S., Moreira, M. F., Atella, G. C., Oliveira-Carvalho, A. L., Eizemberg, R., Majerowicz, D., et al. (2014). Molecular characterization of *Rhodnius prolixus*' embryonic cuticle. *Insect Biochem. Mol. Biol.* 51, 89–100. doi: 10.1016/j.ibmb.2013.12.005
- Sukhwant, A., and Sowdhamini, R. (2015). PPCheck: a webserver for the quantitative analysis of protein-protein interfaces and prediction of residue hotspots. *Bioinform Biol. Insights* 9, 141–151. doi: 10.4137/BBI.S25928
- Tribolium Genome Sequencing Consortium, Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S. J., Denell, R., et al. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452, 949–955. doi: 10.1038/nature06784
- Unni, S., Huang, Y., Hanson, R. M., Tobias, M., Krishnan, S., Li, W. W., et al. (2011). Web servers and services for electrostatics calculations with APBS and PDB2PQR. *J. Comput. Chem.* 32, 1488–1491. doi: 10.1002/jcc.21720
- Uziela, K., Menéndez Hurtado, D., Shu, N., Wallner, B., and Elofsson, A. (2017). ProQ3D: improved model quality assessments using deep learning. *Bioinformatics* 33, 1578–1580. doi: 10.1093/bioinformatics/btw819
- van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastriitis, P. L., Karaca, E., et al. (2016). The HADDOCK2.2 web server: user-friendly

- integrative modeling of biomolecular complexes. *J. Mol. Biol.* 428, 720–725. doi: 10.1016/j.jmb.2015.09.014
- Weng, G., Wang, E., Wang, Z., Liu, H., Zhu, F., Li, D., et al. (2019). HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res.* 47, W322–W330. doi: 10.1093/nar/gkz397
- Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, W407–W410. doi: 10.1093/nar/gkm290
- Xiong, P., Zhang, C., Zheng, W., and Zhang, Y. (2017). BindProfX: assessing mutation-induced binding affinity change by protein interface profiles with pseudo-counts. *J. Mol. Biol.* 429, 426–434. doi: 10.1016/j.jmb.2016.11.022
- Xu, D., and Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys. J.* 101, 2525–2534. doi: 10.1016/j.bpj.2011.10.024
- Xue, L. C., Rodrigues, J. P., Kastiris, P. L., Bonvin, A. M., and Vangone, A. (2016). PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* 32, 3676–3678. doi: 10.1093/bioinformatics/btw514
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* 12, 7–8. doi: 10.1038/nmeth.3213
- Zanet, J., Benrabah, E., Li, T., Pélissier-Monier, A., Chanut-Delalande, H., Ronsin, B., et al. (2015). Pri sORF peptides induce selective proteasome-mediated protein processing. *Science* 349, 1356–1358. doi: 10.1126/science.aac5677
- Zanet, J., Chanut-Delalande, H., Plaza, S., and Payre, F. (2016). Small peptides as newcomers in the control of *Drosophila* development. *Curr. Top. Dev. Biol.* 117, 199–219. doi: 10.1016/bs.ctdb.2015.11.004
- Zhang, J., Liang, Y., and Zhang, Y. (2011). Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure* 19, 1784–1795. doi: 10.1016/j.str.2011.09.022
- Zhu, M., Hu, X., Cao, G., Xue, R., and Gong, C. (2018). Functions and impact of tal-like genes in animals with regard to applied aspects. *Appl Microbiol Biotechnol.* 102, 6841–6845. doi: 10.1007/s00253-018-9159-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Tobias-Santos, Guerra-Almeida, Murry, Ribeiro, Berni, Araujo, Logullo, Feitosa, de Souza-Menezes, Pessoa Costa and Nunes-da-Fonseca. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.