Analysis of the testis transcriptome of the Chagas' disease vector *Rhodnius prolixus*.

Jovino César[1*]; Jose M. C. Ribeiro[2&*]; Marcos Horácio Pereira[2+]; Ricardo N. Araújo[2+]; Nelder F. Gontijo[2+]; Grasielle C. D. Pessoa[2+]; Mauricio V. Sant'Anna[2+]; Marcos H. F. Sorgine[3+]; David Majerowicz[4+]; Marcelo Medeiros[5+]; Antonio B. Carvalho[6+]; Gloria Braz[7+]; Rafael D. Mesquita[7+]; Pedro L. Oliveira[3+]; Leonardo B. Koerich[1+&]

[1]Departamento Parasitologia – UFMG – Brazil

[2]National Institute of Alergy and Infectious Diseases – NIH – US

[3]Instituto de Bioquímica Médica – UFRJ - Brazil

[4]Instituto de Farmácia – UFRJ - Brazil

[5]Instituto de Metrologia – RJ - Brazil

[6]Instituto de Biologia – UFRJ - Brazil

[7]Instituto de Química – UFRJ - Brazil

+Instituto Nacional de Ciência e Tecnologia – Entomologia Molecular

*These authors contributed equally to this work

[&]Corresponding authors: LBK:lbkoerich@ufmg.com; JMCR: jribeiro@niaid.nih.gov

Abstract

Rhodnius prolixus is amongst the most important vectors of *Trypanosoma cruzi* in the Americas, putting thousands of people at risk of contracting Chagas' Disease. This insect is also one of the most important models in insect physiology, especially regarding the blood-feeding process. However, studies on *R. prolixus* genetics lagged behind and our understanding on the regulation of gene expression is incipient. Transcriptomes have the power to study the expression of thousands of genes in a single experiment. A comprehensive *R. prolixus* transcriptome was performed in 2014, sequencing RNA from different tissues (anterior gut, midgut, posterior gut, rectum, ovaries, fat body, maphigian tubules and testes). However, on the occasion, only the gut transcriptome was deeply analysed. Here we evaluated the results of the testes transcriptome of *R. prolixus* with the objective to find and understand genes that could have an important role in male reproduction. We found that, from the 25,673 transcripts assembled in the whole transcriptome, 5,365 have a testis specific expression pattern. As expected, amongst the most abundant families of transcripts, are those related to spermatogenesis and male fertility, such as myosins, actins and dyneins. For our surprise, lipocalins, serine protease inhibitors (serpins) and lysozymes also were highly abundant in testes. The role of these classes of genes are well known in other tissues, such as salivary glands and gut, but very little is known on their role in testes (there are only a few hypothesis) and it would be interesting to further study the role of these genes on *R. prolixus* male fertility. Finally, as a reflection of the lack of knowledge on triatomine genetics, we found that almost half of transcript abundance in *R. prolixus* testes have no similarities to any other genes on reference databases. Our study shows that we still have a lot to know and to understand about reproduction in triatomine, especially in males. Besides the huge number of genes without described function (possibly novel genes), there are those in which the function is known for other tissues and we can only guess, at best, the role and importance of such genes for triatomine male fertility.

Introduction

Rhodnius prolixus (Hemiptera:Reduuviidae) is the main vector of Chagas' disease in Central and the northern part of South America. Chagas disease is amongst the most important parasitic infection in Latin America and nearly 6 million people are infected with *Trypanosoma cruzi* (the causative agent of Chagas' disease) in 21 countries, with ~40,000 new cases every year[1]. In addition, since the works of Sir Vicent Wigglesworth in the first half of the XXth century, *R. prolixus* is also a model in the study of insect physiology[2], especially in respect to hematophagy. However, as in many other hematophagous insects, most studies on *R. prolixus* focuses on female biology, while knowledge on male biology is very incipient.

In recent years, RNA sequencing techniques became cheaper and of easier access by the scientific community. Whole mRNA sequencing (or whole RNA sequencing in some cases) is now widely used as a tool to discover new genes, new transcript isoforms, to study genes differentially expressed in different conditions (or tissues and developmental stages), and to create a catalogue of gene candidates for further functional studies. As a tradition in vector biology studies, a few transcriptomes have catalogued expression of *R. prolixus* genes in the salivary glands[3], gut[4] and ovaries[5] in order to study hematophagy, vector-host interactions, vector-parasite interactions and female fertility. In ticks, blood-feeding also stimulates spermatogenesis, mating and production of male factors that trigger female reproduction[6]. However, no comprehensive investigation has addressed exclusively male biology. *R. prolixus* testes starts to develop in fifth instar nymphs and are fully functional in adults. A male adult has two testicles, and each testicle is composed of: two long testes; five short testes; a vas deferens duct; vesicular seminal duct; duct of accessory grands and; four accessory glands (three transparent and one opaque)[7]. Spermatogenesis occur in each of the seven testes arms and we still do not know if the different arms sizes have any specific role on male reproduction. On the other hand, many studies on *R. prolixus* accessory glands have been conducted since the beginning of the XX century, showing its importance on spermatophore production and

male fertility[8]. In insects, accessory glands are involved in[9]: *i)* facilitation of the insemination of the females; *ii)* sperm activation; *iii)* formation of mating plugs; *iv)* modification of female mating behaviour; *v)* sperm competition and; *vi)* egg maturation and oviposition. Still, little is known about key genes related to male fertility and reproduction biology in insects.

Transcriptomics and proteomics have been successfully applied to create a catalogue of testis specific genes in arthropods. The testes transcriptome of *Drosophila melanogaster* described ~8,500 genes and further analyses on gene expression revealed that 399 genes were upregulated in testes[10]. Most of the upregulated genes are typical components of sperm structure (*eg.* dyneins, mitochondria and outer dense fiber), sugar metabolism and peptidases. Among those, only one was previously known to cause infertility in males. In the leishmaniasis vector *Lutzomyia longipalpis* (Diptera:Psychodidae), the testes transcriptome suggested that many of the accessory gland proteins are involved in proteolysis (*e.g.* serine protease, metalloproteases and protease inhibitors), in immunity and redox metabolism[11]. In the tick *Dermacentor variabilis*, the testis transcriptome and spermatophore proteome showed the importance of serine/threonine kinase, metalloendoproteinases, ferritins, serine proteases, trypsin, cysteine proteases, serpins, a cystatin, GPCR and others, in the reproductive biology of these arthropods[12].

On the other hand, the physiology of *R. prolixus* male reproduction and its impact on female development are essentially a black box. Consequently, a new global approach using transcriptomics, proteomics, and quantitative gene expression is needed to understand male physiology and fertility of this insect. In 2010, a comprehensive RNA-seq from different tissues (anterior gut, midgut, posterior gut, rectum, malpighian tubules, ovaries, fat body and testes) of *R. prolixus* was performed, and the analysis of the gut transcriptome was published in 2014[4]. Now, in this study, we analysed the data from the testes transcriptome and found ~14,000 *R. prolixus* testes transcripts, with 5,365 testes specific transcripts. This is the first comprehensive genetic study on *R. prolixus* male reproductive

organs, and we expect to provide some insight on male specific fertility genes and candidates for further functional studies on this insect reproductive biology.

Results and Discussion

The original transcriptome[4] generated a total of 171,124 (454) reads that where assembled in 25,673 contigs. We found that 14,454 are expressed in testes, from which 5,365 contigs have a testes specific expression pattern, while 9,089 are transcribed in testes and in other tissues (Figure 1). We did not observe testes expression for 11,219 contigs. Transcript function was predicted by using Blast against the NR, SwissProt, COG, KOG, CDD and Pfam databases. A total of 7,691 transcripts did not show~~ed~~ any similarity with described proteins, possibly representing novel *R. prolixus* genes. From those transcripts showing blast hits, we classified them in seven major categories: housekeeping, secreted, immunity, transposable elements, viral, unknown conserved (proteins with unknown function but found in other organisms) and unknown (Supplementary table 1). Each transcript was then categorized in functional classes (*e.g.* protein synthesis, transcription machinery, signal transduction, etc) or protein families (*e*.g. lipocalins, serine proteases, lysozymes, etc.). Nearly 85% of the testes contigs belongs to unknown (6,395 contigs) and housekeeping (6,012 contigs) categories (Figure 2), with the relative abundance (RPKM) of transcripts following the same pattern. On the other hand, the abundance of unknown transcripts in other tissues is ~2.5-fold higher than the abundance of housekeeping transcripts. Another difference is observed in the secreted transcripts category. While the number of contigs in this category is almost the same for testes and other tissues (182 and 214 contigs respectively), the relative abundance of testes secreted transcripts is ~2-fold higher than in other tissues (Figure 2), suggesting an important role of testes as a secretion organ.

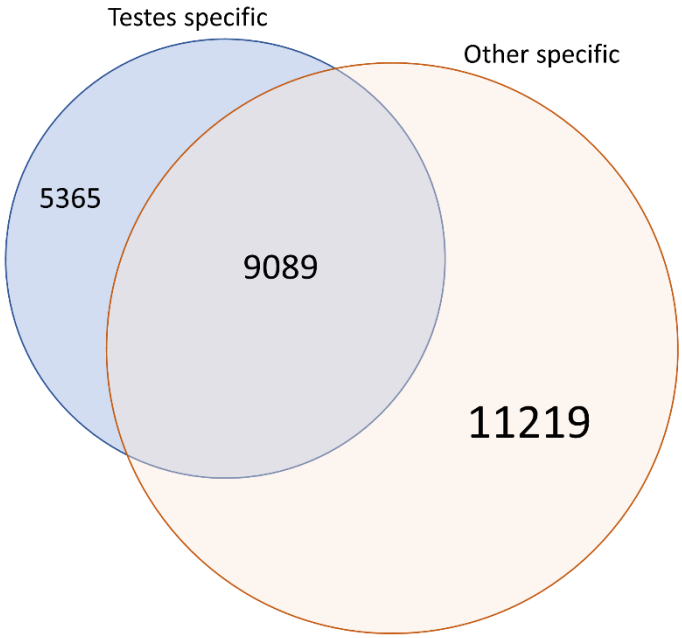

Figure 1. Distribution of specific and non-specific testes contigs in *Rhodnius prolixus*. Other specific includes transcripts found in gut, ovaries, fat body and Malpighian tubules.

Overview in housekeeping, secreted and immunity-related genes

Housekeeping: When we evaluate the relative abundance of transcripts by functional class in each major category, we observe differences in only a few groups (Figure 3). In the housekeeping category, we observe that there is a trade-off between protein synthesis, cytoskeletal and signal transduction transcripts (Figure 3 and Supplementary table 2). While the abundance of protein synthesis transcripts decreases ~1.5-fold in testes, the abundance of cytoskeletal and signal transduction transcripts increases ~2.5 and ~1.5-fold respectively. The increase in cytoskeletal transcripts should be expected due to genes involved spermatogenesis and the flagellar structure in spermatozoa, which involves actins, myosins, tubulins and dyneins. Actins and myosins have a key role in mitosis and meiosis and we found many actin and myosin transcripts overexpressed in testis. In fact, two of the most expressed genes in testes (Table 1) belongs to the myosin family; a myosin

class II transcript (1.5-fold higher in testes) representing ~2% of all testis's transcripts, and a stretchin-mlck which has expression ~24-fold higher in testes (representing 0.75% of all testes transcripts). Tubulins and dyneins are the main components flagella and we found 38 tubulins and 27 dyneins expressed in testes. Tubulins represents 7.3% of all testis's cytoskeletal transcripts with a RPKM ~3-fold higher in testes, while dyneins represent 4.1% of all testis's cytoskeletal transcripts with a RPKM ~20-fold higher in testes. We also observed discrete abundancies of kinesins (~5-fold increase in testes), corresponding to 2.9% of all testis's cytoskeletal transcripts. Discussion on myosin's, actins, tubulins, and dyneins are detailed further.

Secreted: Our data suggests that the expression profile of secreted proteins in the testes of *R. prolixus* is very similar to other tissues (Figure 3 and supplementary table 3), with lipocalins and serine protease inhibitors (serpins) as the most abundant families (corresponding to 71% of testes and 70% of other tissues secreted transcripts). In hematophagous insects, lipocalins and serpins are usually remembered as key salivary proteins to inhibit host hemostasis during blood ingestion. However, a few studies have demonstrated the role of lipocalins in sperm maturation and the role of serpins in spermatogenesis. We found one lipocalin and one serpin amongst the most expressed genes in testes (table 1) and an in-depth discussion on these families are detailed further in this manuscript.

Immunity: Immunity-related transcripts represents less than 2% of all transcripts expressed in testes and other tissues (Figure 2). Nonetheless, our data suggests striking differences in the expression profile of testes immunity related genes in comparison to other tissues (Figure 3). We observed that lysozymes correspond for 75% of all immunity related transcripts in testes, while these proteins correspond for 37% of all immunity related transcripts in other tissues (Figure 3 and supplementary table 4). Lysozymes are enzymes with antimicrobial activity that recognize peptidoglycans. Lysozymes have been found on seminal fluids of insects and mammals, but its role in this tissue remains to be elucidated. Discussion on *R. prolixus* testes lysozymes are detailed further.

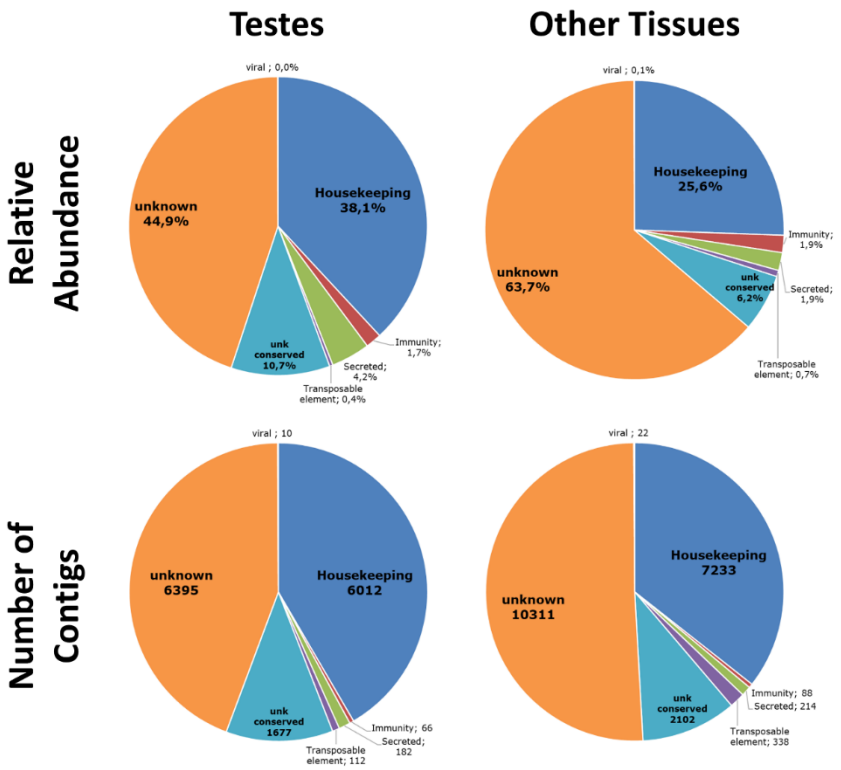

Figure 2. Relative abundance of transcripts and number of contigs by functional class in testes and other tissues of *Rhodnius prolixus.* Other tissues of *R. prolixus* are gut (anterior gut, midgut, posterior gut and rectum), ovaries, malpighian tubules and fat body. Abundance is based in reads per kilobase million (RPKM).

Table 1. Top ten expressed genes in *Rhodnius prolixus* testes

Name	Category	Family	RPKM		Relative Abundance	
			Testes	Other tissues	Testes	Other tissues
rp_asb-229	Housekeeping	Myosin Class II	13515.66	8852.31	1.51%	0.59%
rp_asb-316	unk conserved	unk conserved	13166.88	4800.01	1.47%	0.32%
rp_asb-165	Immunity	Lysozyme	11466.31	10011.70	1.28%	0.67%
rp_asb-47	Secreted	Lipocalin	10417.37	6401.19	1.16%	0.43%
rp_asb-1369	unknown	Unknown Product	9215.73	0.00	1.03%	0.00%
rp_asb-3088	unknown	Unknown Product	8660.44	8.46	0.97%	0.00%
rp_asb-4409	Housekeeping	Stretchin-Mlck	6769.18	274.04	0.75%	0.02%
rp_asb-433	Secreted	Serpin	6756.90	5829.12	0.75%	0.39%
rp_asb-3455	unknown	Unknown Product	6640.39	2.77	0.74%	0.00%
rp_asb-3046	unknown	Unknown Product	5300.36	8.48	0.59%	0.00%

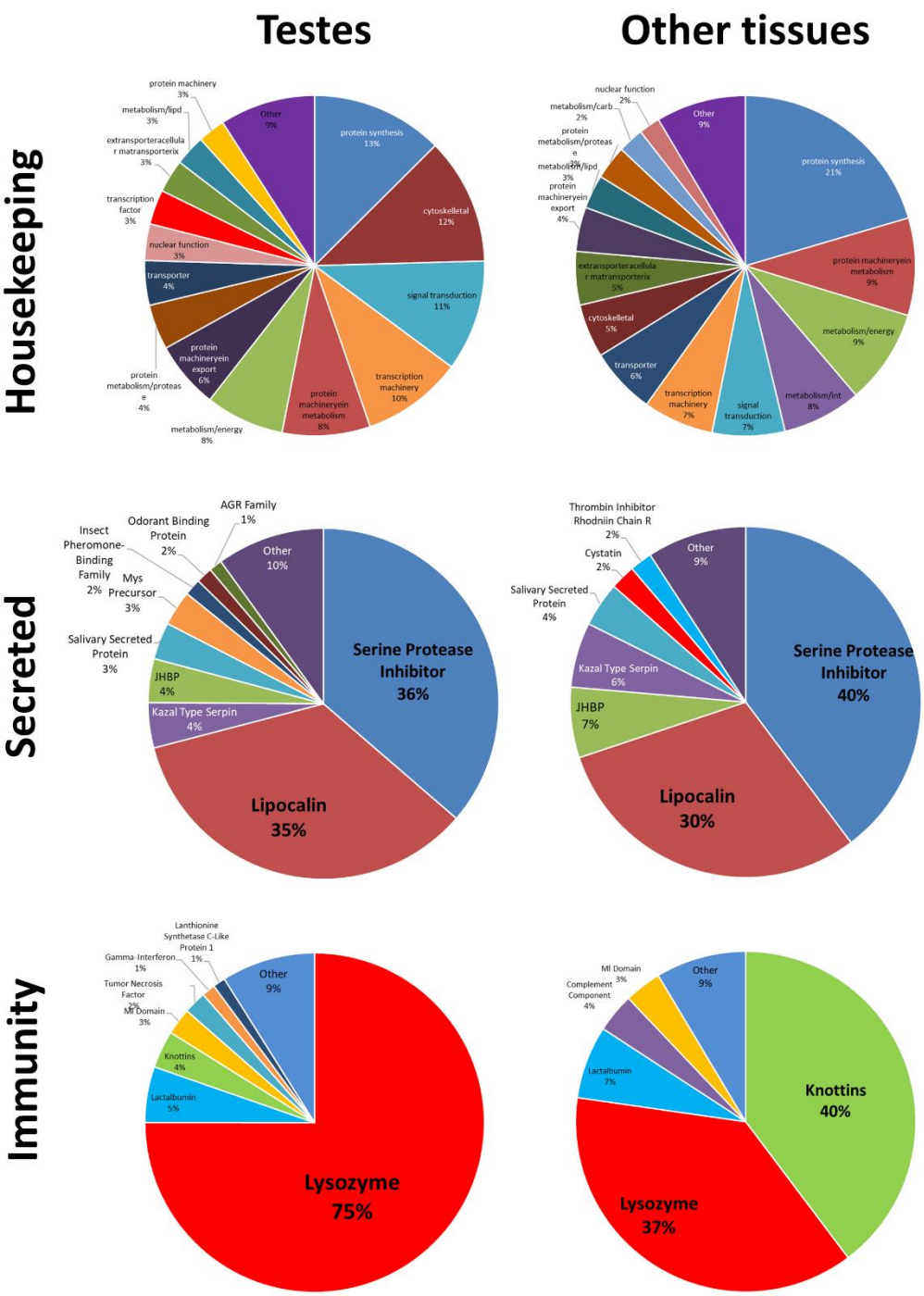

Figure 3. Relative abundance of transcripts by functional ontologies in three different major functional classes in testes and other tissues of *Rhodnius prolixus*. Other tissues of *R. prolixus* are gut, ovaries, malpighian tubules and fat body. Abundance is based in reads per kilobase million (RPKM).

Specific and differentially expressed genes in *Rhodnius prolixus* testes.

To have some insights on the regulation of gene expression of *R. prolixus* testes, we performed a heatmap analysis using the mean z-scores for each gene family (Figure 4). Although the mean z-score can hide some overexpressed genes, this analysis avoids the complexity of evaluating the expression pattern of tens of thousands of genes, providing a simple and fast identification of gene families that may have important roles in tissue organization, spermatogenesis, sperm maturation, fertility, etc. Heatmaps were separated in the three major categories: secreted (Figure 4 panel A), housekeeping (Figure 4 panel B), and immunity (Figure 4 panels C1 and C2). As expected, we could not observe clear differences in the expression profile of housekeeping gene families (Figure 4 panel B), even in cytoskeletal transcripts, which have a relative abundance in testes much higher than in other tissues. This result suggests that the abundance of cytoskeletal genes in testes is a result of a few overexpressed genes. In fact, from the 634 cytoskeletal genes found in the transcriptome, only 429 are expressed in testes, while 519 are expressed in other tissues (~20% more than testes). On the other hand, we observed clear differences in secreted and immunity categories. From the secreted gene families, our data suggests that the expression of mucins, anterior gradient (Agr), mys precursor and serine protease inhibitors of the serpin and Kazal families are generally upregulated in testes and down regulated in other tissues, while ubiquitin, triabin, apyrases, fringe-like proteins, salivary phospholipases, tryptophan-rich proteins, 47kda salivary proteins, inositol polyphosphates and platelet aggregation inhibitor proteins seems to be down regulated in testes. Interestingly, many of the downregulated gene families in testes are important salivary genes with a key role in inhibiting host hemostasis during blood feeding. In the immunity category (Figure 4 panel C1), we observed a general downregulation in lysozymes, tyrosine phosphatase, defensins, chemokines and spatzle-3 gene families. It is interesting to note the down regulation of lysozymes in *R. prolixus* testes, since transcripts of this family corresponds to 75% of all immunity transcripts. The reason is that only two lysozymes out of nine are transcribed in testes, and one of them is the third most abundant gene in

the testis's transcriptome (table 1). Since the number of immunity related transcripts in the whole transcriptome is relatively low (exactly 100 genes), we decided to analyse the z-scores individually in this category (Figure 4 panel C2). From those, 12 genes are specifically expressed in testes (Gamma-Interferon, Lanthionine Synthetase, Membrane Glycoprotein, 2 Phosphatases, Rhomboid, 4 Tumour Necrosis Factor genes and 2 unclassified genes) and 34 genes are not expressed in testes (mostly from the downregulated families). A more detailed discussion on testes specific genes are detailed further in this manuscript.

An interesting question regards the expression profile of uncharacterized genes (unknown and unknown conserved), since this group corresponds for more than seven thousand genes and for ~55% of transcript abundancy in testes. From the 2,533 unknown conserved genes in the whole transcriptome, 431 have a testes specific expression profile, while 856 are not expressed in testes (supplementary excel table). From those genes with expression in both tissues, at least one of them seems to be upregulated in testes, with a RPKM value 3-fold higher in testes than in other tissues (table 1). From the 13,874 genes with unknown function and not conserved in other organisms, 3,563 contigs are specifically expressed in testes and 7,479 are not transcribed in testes (supplementary excel table). Four of those genes are amongst the top ten transcribed products in testes (table 1). From the 5,364 testes specific transcripts in this transcriptome, ~75% (3,994 transcripts) have unknown function and most of them could be novel *R. prolixus* proteins. Many proteomes and transcriptomes of the male reproductive organs describe the diversity of seminal fluid proteins produced by the male accessory glands, and we suspect that many of the testes specific unknown proteins found here could belong to this class of proteins. Our reason to believe that is because such proteins are known to be species specific short peptides (many of them have a role in species compatibility), that evolve very fast. Indeed, the mean size of *R. prolixus* testes specific peptides with unknown function is 100.68 amino acids, while the mean size of peptides in the whole

transcriptome is 160.05 amino acids. Still, such claims should be answered in a further transcriptome aiming specifically to male accessory glands.

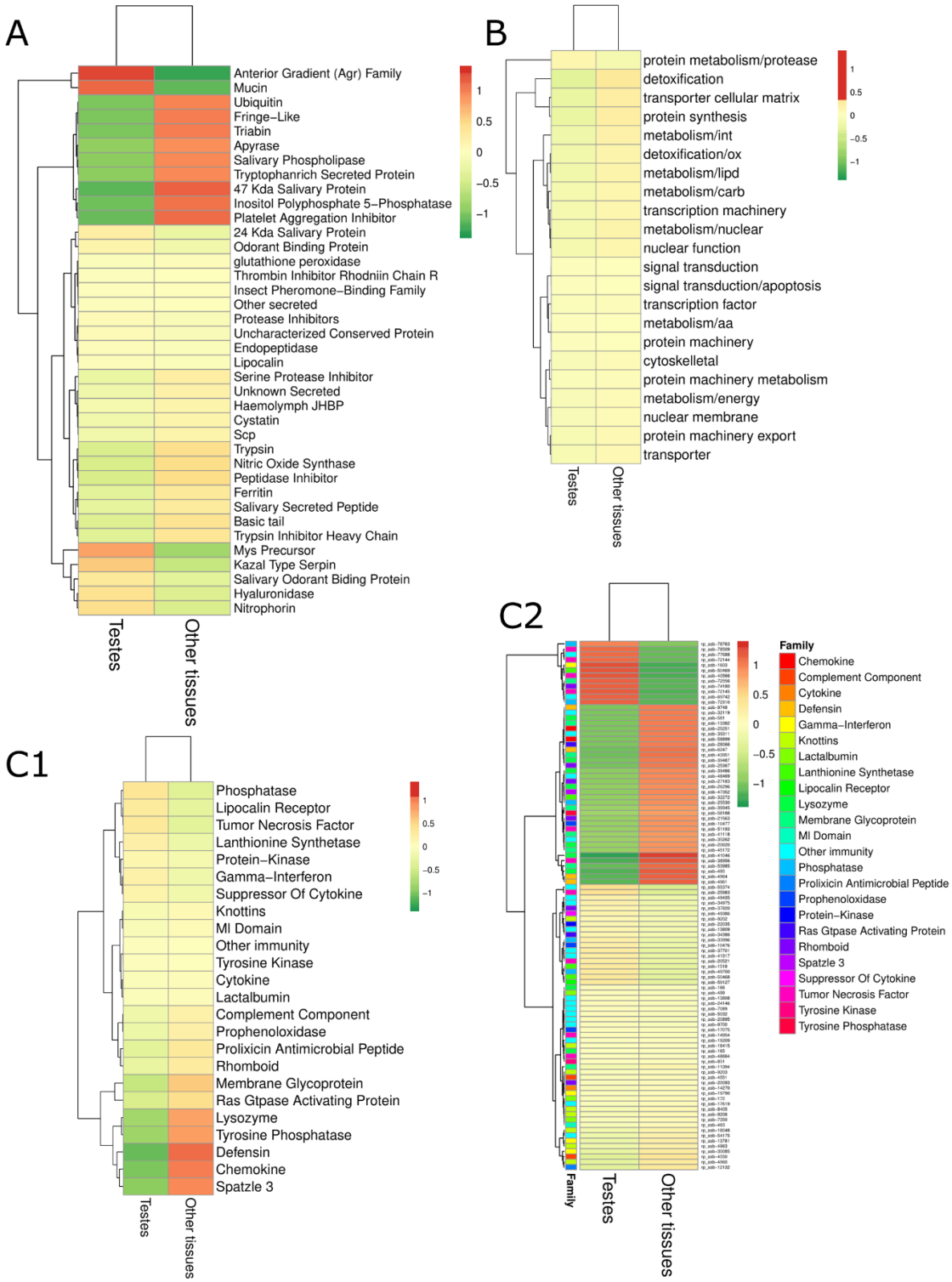

Figure 4. Heatmaps for each major functional category. Heatmaps show transcript abundancy (expression level) in each sample. **A** Housekeeping; **B** Secreted; **C1** Immunity; **C2** All immunity genes. In Bright Red are upregulated classes of genes (>2-fold standard deviation); Bright Green are downregulated classes of genes (< 2-fold standard deviation); Brownish green or red are classes with standard transcription. Classes were grouped by a pattern of transcription (eg. Genes that are upregulated in both tissues; genes that are differentially expressed, genes that are downregulated in both tissues; genes that has a standard regulation in both tissues; etc).

Male fertility factors and other flagellar components expressed in testes.

Flagella is found in many prokaryotic and eukaryotic cells and are a typical component of sperm. In eukaryotes, the flagellum main structure is the axoneme, composed mainly by an intricate combination of microtubules (tubulins) and dyneins[13]. While microtubules are responsible for support, dyneins act as motors, giving movement to the flagellum[14]. There are several isoforms of axonemal dynein heavy chains (α, β, γ, 1-α, 1-β, etc.) that associate to form the inner and outer arms of the axonemes. In *D. melanogaster*, the Y chromosome harbours three dynein heavy chains known as *kl-2*, *kl-3* and *kl-5*, and knockout of any of these genes causes male sterility due to lack of flagellum movement[15–17]. We found 45 tubulins and 30 dyneins in the *R. prolixus* transcriptome. From those, 19 tubulins and 15 dyneins presented testes specific expression and only six tubulins and three dyneins were not expressed in testes (Supplementary excel table). Tubulins are 3-fold more expressed in testes than in other tissues, while dyneins are 20-fold more expressed in testes. Evolutionary analysis of the dynein heavy chains (Figure 5) suggests that *R. prolixus* have two orthologs of *D. melanogaster kl-2* (which belongs to the dynein heavy chain 1-β family), and one of them has a testis specific expression pattern. However, none of them are amongst the most expressed dyneins in the transcriptome. The two most expressed dyneins are rp-asb-15029 and rp-asb-15030 (RPKM 397.39 and 80.10 respectively). These two dyneins are not testes specific, and both belongs to the dynein heavy chain 6 family. Interestingly, we did not find any orthologs for *kl-3* and *kl-5*, which are well conserved genes in eukaryotes, arising a question if these transcripts were not detected in our study or if such genes were lost on the *R. prolixus* genome. Still, there is no studies on the role of dyneins in the fertility of kissing bugs and further experiments using silencing (RNAi) or knockout (CRISPR/cas-9) approaches should be performed to ascertain such role.

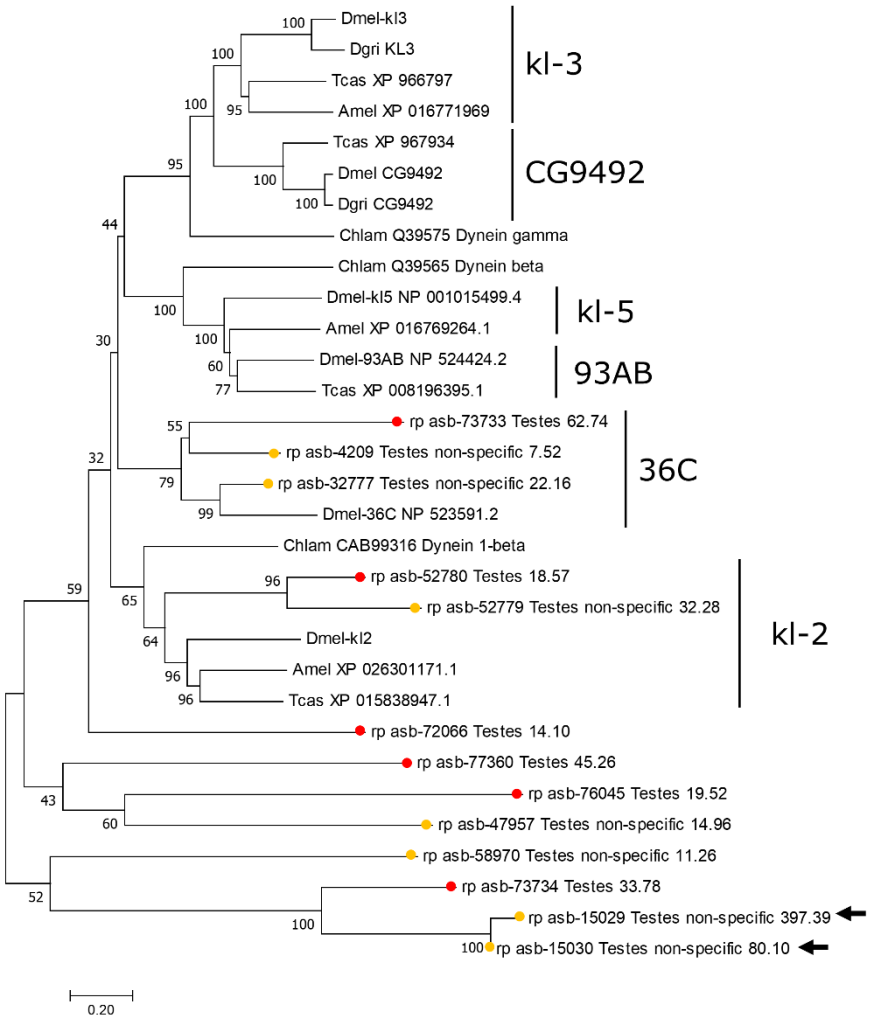

Figure 5. Evolutionary relationships of *R. prolixus* dynein heavy chain proteins. The evolutionary history of the *R. prolixus* dyneins was inferred using the Neighbor-Joining method. Red dots show transcripts with testes specific expression pattern and yellow dots shows non-specifc expression pattern. RPKM values of each transcript are shown in each branch and arrows indicate transcript with the highest RPKM values. Accession numbers of non-*Rhodnius* proteins are shown in each branch after species abbreviation. Species abbreviations are: Dmel = *Drosophila melanogaster*; Dgri= *D. grimshawii*; Tcas=*Tribolium castaneum*; Amel=*Apis mellifera*; Chlam=*Chlamydomonas* sp. Evolutionary analyses were conducted in MEGA.

Spermatogenesis related genes.

Spermatogenesis is an interesting phenomenon present in eukaryote and is a predominant part of sexual reproduction. The process begins with mitotic divisions of spermatogonia stem cells to create spermatocyte cells[18]. Spermatocytes suffer meiotic division to form haploid cells that will maturate

(during spermiogenesis) into spermatozoa. Despite its importance, mechanisms underlying spermatogenesis (specially the meiotic process) remains largely obscure. There are two characteristics in kissing bugs spermatogenesis that turns this process even more intriguing. First is the fact that kissing bugs chromosomes are holokinetic (they do not have a centromere) and processes of segregation of holokinetic chromosomes during cell division remains unclear[19,20]. The second characteristic is that, in kissing bugs, meiosis of sex chromosomes occurs in an inverse order, in which sister chromatids segregate in the first stage of meiosis and homologs segregate in the second stage[21]. We identified three family of genes that are well known for their role in spermatogenesis: myosin, actin and tumour necrosis factors (TNF).

Myosin are a superfamily of actin-dependent molecular motor proteins known for their roles in muscle contraction in eukaryotes. These proteins also have a key role in spermatogenesis, such as acrosome biogenesis, spindle assembly and positioning, karyokinesis and spermatid individualization[22]. We found 114 expressed transcripts from the myosin family, from which 19 presented a testes specific expression pattern (Supplementary excel table). The most abundant transcript in testes is the myosin rp-asb-70879 (table 1), and it has non-specific expression pattern, being also very abundant in gut tissues[23] (Supplementary excel table). However, the second and third most expressed myosin in testes are testes-specific and it would be very interesting to further evaluate their role in spermatogenesis using silencing (RNAi) techniques. On the other hand, only one actin (rp-asb-71413) from the 44 identified actins is specifically expressed in testes. Actins are critical for correct nuclear positioning, germinal vesicle breakdown, spindle migration, spindle rotation and chromosome segregation in spermatogenesis[22]. Hence, the question on the role of this testes specific actin remains open.

The specific role of TNFs in spermatogenesis of insects is still unknown. In vertebrates, TNFs are cytokines involved in the regulation of immune cells. It can induce fever, apoptotic cell death, inflammation, inhibition of tumorigenesis, etc. Many studies have demonstrated that disruption of

TNF expression in testes can affect spermatogenesis, imputing these molecules as essential for maintaining germ cell homeostasis and functional spermatogenesis in testis[24,25]. In mammals, testes specific TNF-alpha has an anti-hormonal role and play interactions between Steroli and germ cells (Sertoli cells are thought to play a central role in male-specific cell interactions, including those that occur during spermatogenesis)[26]. In Drosophila, testis contains two types of stem cells, germline stem cells (GSCs) and cyst stem cells (CySCs)[27] and it remains to be seen if *R. prolixus* has the same population of germline stem cells. From the ten *R. prolixus* TNFs identified in the transcriptome, we found that two TNFs (rp_asb-25983 and rp_asb-20521) are differentially expressed in testes, and four (rp_asb-40566, rp_asb-72145, rp_asb-72144, rp_asb-78509) that have a testis specific expression pattern (Figure 4, C2). Although rp_asb-25983 is also expressed in the anterior midgut (AM), it is ~20-fold more expressed in testes than in AM (supplementary excel table). Such genes could be targets to understand the role of TNF in insects and its interaction, if any, with GSCs and CySCs.

Secreted lipocalins

The lipocalins are a family of proteins which transport small hydrophobic molecules such as steroids, bilins, retinoids, and lipids[28]. They share limited regions of sequence homology and a common tertiary structure architecture[29]. Lipocalins have been associated with many biological processes, among them immune response, pheromone transport, biological prostaglandin synthesis, retinoid binding, and cancer cell interactions[28]. Ticks and kissing bugs evolved salivary lipocalins that act as efficient scavengers of biogenic amines[30–32]. Lipocalins represented 35% of all secreted transcripts in testes, and from the 19 lipocalins identified in *R. prolixus* transcriptome, we found that eight (42%) are specifically expressed in testes, while only five are not expressed in testes. The lipocalin rp_asb-47 is the fourth most abundant transcript in testes, and it is nearly 2-fold more abundant in testes than in other tissues. Interestingly, none of the testes specific lipocalins are highly expressed (RPKM values bellow 50). Phylogenetic analysis comparing the transcriptome lipocalins with lipocalins from

GenBank suggests that a few testes specific lipocalins could have general roles in platelet aggregation inhibition, more specifically: pallipidin (rp_asb-73726, rp_asb-73728, rp_asb-73729), triabin/procalin (rp_asb-15401) and platelet aggregation inhibitors (rp_asb-34425). One could wonder why inhibitors of platelet aggregation would be specifically expressed in *R. prolixus* testes. First, in hematophagous insects, lipocalins have been mostly studied in the light of the blood feeding process, which creates a bias in our knowledge. Secondly, it is important to note that salivary glands (SG) were not included in this transcriptome (the reason was that SG had already been extensively studied before). Hence, it is possible that some of the testes specific lipocalins are also expressed in SG and that they role in testes is not analogous to their role in SG. Lipocalins have been detected in the semen of other organisms. However, their role in male fertility is incipient. The best known testes specific lipocalins are the prostaglandin D-synthase lipocalins of mammals[33], where it is hypothesised that lipocalins have a role as carriers of hormones and retinoids to the developing germ cells in the seminiferous tubules and the maturing spermatozoa[34]. Interestingly, all three lipocalins that showed higher expression in testis (rp_asb-47, rp_asb-61810, rp_asb-7470) do~~es~~ not correlate with lipocalins with known function. Although these genes are not specifically expressed in testes, it is possible that they might have a central role in sperm maturation with a role not yet known to insect lipocalins, turning them important targets for further functional studies.

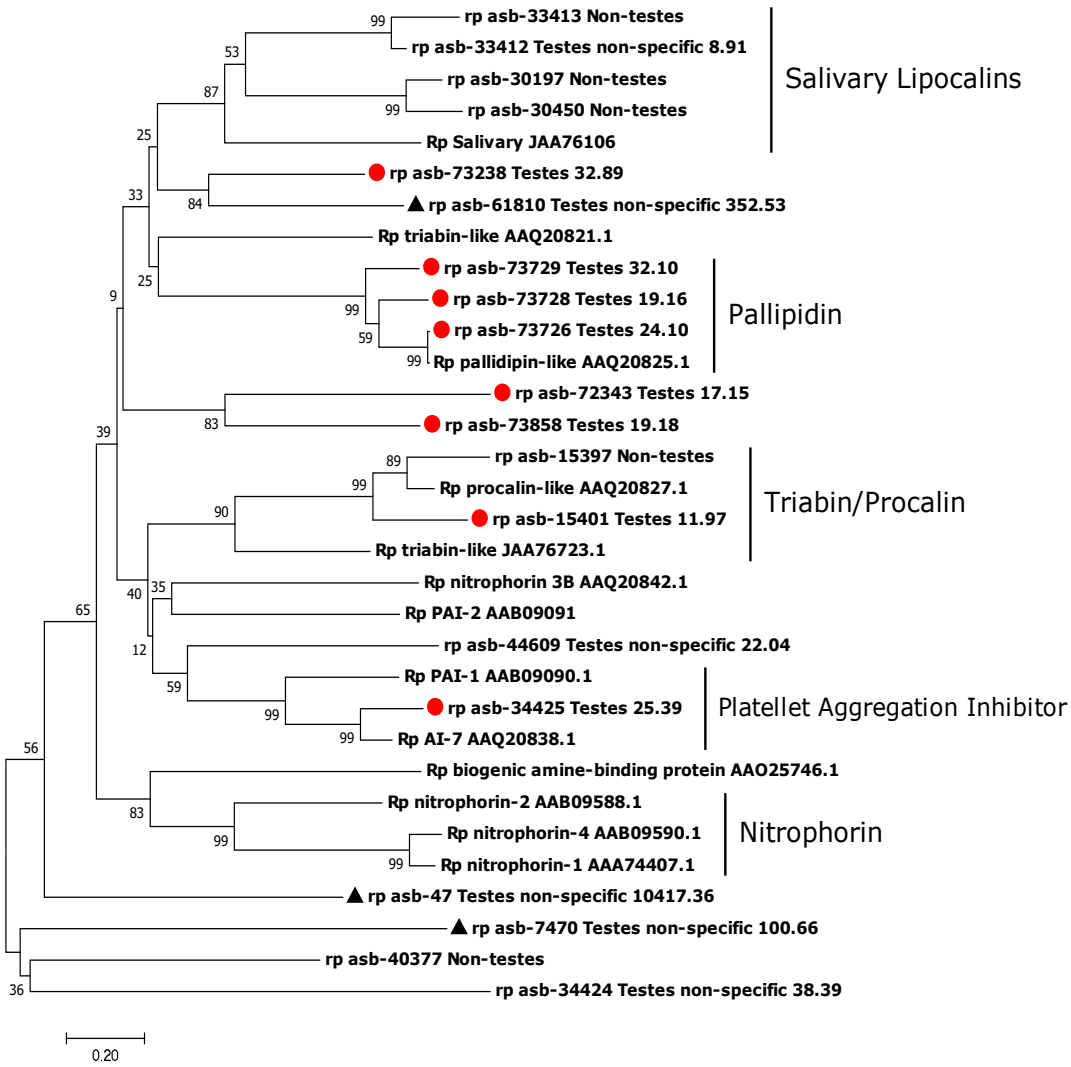

Figure 6. Evolutionary relationships of *R. prolixus* lipocalins. The evolutionary history of the *R. prolixus* lipocalins was inferred using the Neighbor-Joining method. Red dots show transcripts with testes specific expression pattern, black triangles show transcripts overexpressed in testes. RPKM values of each transcript are shown in each branch and arrows indicate transcripts with the highest RPKM values. Accession numbers of non-*Rhodnius* proteins are shown in each branch after species abbreviation. Evolutionary analyses were conducted in MEGA.

Secreted serine protease inhibitors

Serpins transcripts constitute ~36% of all secreted proteins in testes. Serpins are a superfamily of with similar structures that were first identified for their protease inhibition activity and are found in all kingdoms of life. In the *R. prolixus* transcriptome we found 36 serpins, but none of them

presented a testis specific expression pattern. Twenty-three of them seems to be highly, and equally, expressed in testes and gut (mean RPKM of 611.4 in testes and 498.7 in gut). Serpins have been detected in turkey testis, epididymis, ductus deferens, spermatozoa surface and in seminal plasma using electrophoretic methods[35]. From the 36 serpins found, transcript rp_asb-21159 caught our attention due to its differential expression in relation to other tissues (~25-fold higher). However, when we looked up for known domains in rp_asb-21159 sequence, we found that it contains a well-known 7tm-odorant binding domain, suggesting a miss annotation in the transcriptome. Hence, although serpins are highly abundant in testes, we could not identify specific contigs or patterns that differentiates from other tissues, and the question on the role of serpins in male reproduction remains open ended. In addition to serpins, Kazal-family serine protease inhibitors are found in seminal plasma, known as acrosin inhibitors. A study showed that serine protease inhibitor Kazal-type 2 (SPINK) is required for maintaining normal spermatogenesis and potentially regulates serine protease-mediated apoptosis in male germ cells[36].

Lysozymes

Lysozymes are antimicrobial enzymes produced by animals that forms part of the innate immune system. These proteins have a specific role in the hydrolysis of N-acetyl-D-glucosamine residues in peptidoglycan, causing lysis of bacteria. Lysozyme is abundant in secretions including tears, saliva, human milk, and mucus. *R. prolixus* transcriptome suggests that Lysozymes may have an important role in testes, representing 75% of all immunity related transcripts. Recently, testis specific lysozyme-like genes (Lyzl) belonging to the c-type lysozyme family have been described in sperm proteome of humans and mouse. However, their role in male fertility and mammal reproduction remains unclear. In the bed bug *Cimex lectularius*, the antimicrobial activity of the seminal fluid is attributed to Lyzl, and it is hypothesized that these genes could have a role in helping the female to protect themselves from bacteria introduced during copula. From the nine Lysozyme transcripts identified in *R. prolixus* transcriptome, only two of them are expressed in testes (both are expressed

in other tissues, mainly gut). Interestingly, evolutionary analysis (Figure 7) suggests that both testes lysozymes are closely related, probably from a recent duplication, and are orthologous to the lysozyme duplications *Lys* of *Drosophila melanogaster* (Figure 7, clade I).

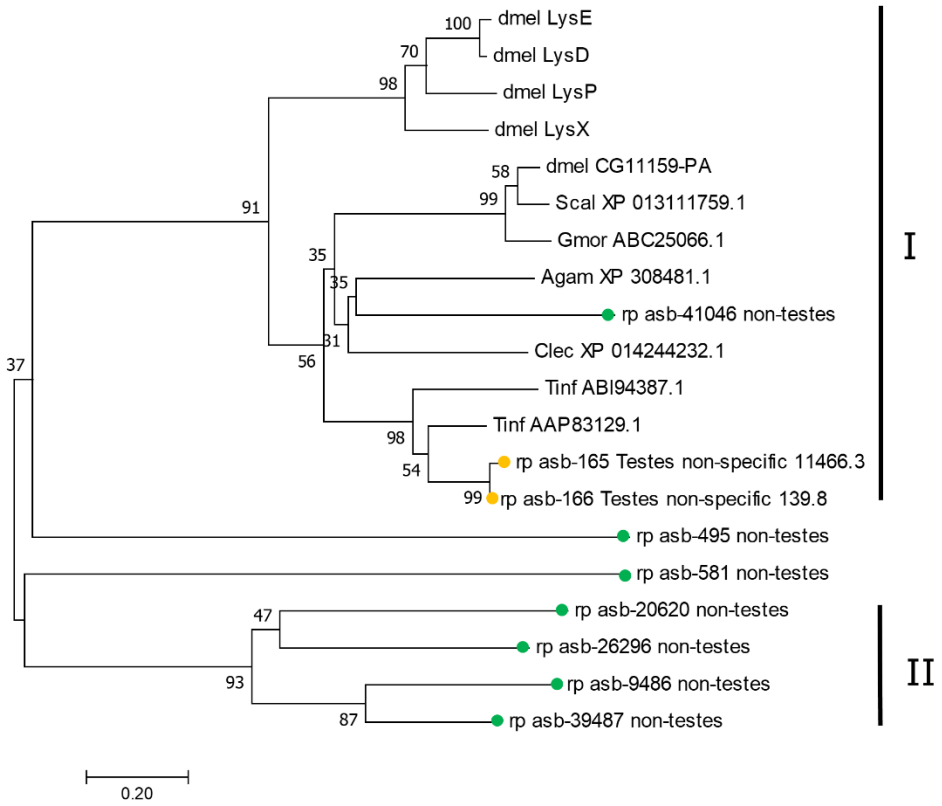

Figure 7. Evolutionary relationships of *R. prolixus* lysozymes. The evolutionary history of the *R. prolixus* lysozymes was inferred using the Neighbor-Joining method. Red dots show transcripts with testes specific expression pattern, yellow dots show non-specific expression pattern and green dots shows transcripts not expressed in testes. RPKM values of each transcript are shown in each branch and arrows indicate transcripts with the highest RPKM values. Accession numbers of non-*Rhodnius* proteins are shown in each branch after species abbreviation. Species abbreviations are: Dmel = *Drosophila melanogaster*; Scal= *Stomoxys calcitrans*; Tinf=*Triatoma infestans*; Agam=*Anopheles gambiae*; Clec=*Cimex lectularius.* Evolutionary analyses were conducted in MEGA.

.

Final insights on the *Rhodnius prolixus* testes transcriptome.

Although *R. prolixus* is a model for studies on insect physiology, the knowledge in the genetics of this species has lagged behind. In such cases, transcriptome studies have the power of creating a catalogue of candidate genes to be further investigated to better understand insect genetics. The results found on *R prolixus* transcriptome is a reflection on our lack of knowledge in triatomine genetics, in which the function of more than 60% of the annotated transcripts remains unknown. In triatomines, most studies aim to better understand the hematophagy process and, although the *R. prolixus* transcriptome was generated from a multitude of tissues, the paper describing *R. prolixus* transcripts focused on the physiology of gut. Therefore, in our study we evaluated the transcripts from *R. prolixus* testes, and this is the first study focused on *R. prolixus* male reproduction in the genomic era. As expected, we do not have a clue on the function of ~50% of the testis's transcripts; and five of the top ten most expressed transcripts in testes belongs to this category. However, even the reproduction role for transcripts containing well described domains remains unclear. And this is the case for genes well studied in *R. prolixus* physiology, such as lipocalins, serpins and lysozymes. These three families of genes have important roles in the blood-feeding process and, surprisingly, these were the most abundant families of secreted and immunity related transcripts. Interestingly, mosquitoes have an expanded salivary family, named the D7 family[37–39], containing one or two odorant-binding motifs that inhibit hemostasis by binding biogenic amines and leukotrienes, a function similar to the salivary lipocalins of Rhodnius. A study of the accessory glands of the mosquito *Aedes aegypti* identified the presence of one typical highly expressed salivary D77 protein[40]. It is possible that such genes have a role in testes very similar to their counterparts expressed in the salivary glands or gut, but it is also very possible that lipocalins, serpins and lysozymes specifically expressed in testes have distinct roles that we still do not understand. On the other hand, the testes transcriptome generated a catalogue of transcripts with well described roles on spermatogenesis and insect fertility. We identified myosins, dyneins, actins, tumour necrosis factors

and other genes that could disrupt spermatogenesis and cause male sterility. Such genes should be targets on functional studies to better understand triatomine male biology, a field of study that is mostly neglected by the scientific community.

Material and Methods

Insects

Insects used for transcriptome were *Rhodnius prolixus* from a colony kept at UFRJ (Rio de Janeiro), fed with rabbit blood, and raised at 28°C and 70% relative humidity. Adult females (five from each condition) receiving their second blood meal after the imaginal molt were dissected before feeding, twelve hours after blood meal (ABM), twenty-four hours ABM, two days ABM, and five days ABM. A group of males (blood fed, five days ABM) was dissected to obtain testes. Organs (anterior midgut, posterior midgut, fat body, ovary, Malpighian tubules, and testes) were dissected, homogenized in TriZol, and processed as described below.

RNA extraction, library preparation, and sequencing

Organs were homogenized in TriZol, and total RNA was isolated, followed by mRNA purification using the Micro-Fasttrack 2.0 kit from Invitrogen (San Diego, CA, USA) according to manufacturer's instructions. Libraries were constructed using the Smart cDNA Library Construction kit from Clontech (Palo Alto, CA, USA) and normalized using the Trimmer cDNA Normalization kit from Evrogen (Moscow, Russia). The libraries were sequenced on a 454 genome sequencer FLX Titanium machine (Roche, Roche 454 Life Sciences, Branford, CT, USA).

Bioinformatics

Detailed bioinformatic analysis of our pipeline can be found in our previous publication. Pyrosequencing reads were removed from vector and primer sequences by running VecScreen. The resulting assemblies plus the clean pyrosequenced data were joined by an iterative BLAST and cap3 assembler. This assembler tracks all reads used for each contig, allowing deconvolution of the number of reads used from each library for tissue expression comparisons using traditional reads per kilobase million (RPKM). Non-testes RPKM is the sum of reads from all libraries except for testes library.

Coding sequences were extracted using an automated pipeline based on similarities to known proteins or by obtaining CDS from the larger open reading frame of the contigs containing a signal peptide. A non-redundant set of the coding and their protein sequences was mapped into a hyperlinked Excel spreadsheet, which is presented as Supplementary Excel Table. Signal peptide, transmembrane domains, cleavage sites, and mucin-type glycosylation were determined with software from the Centre for Biological Sequence Analysis (Technical University of Denmark, Lyngby, Denmark). To assign coding sequences as being of bacterial, viral, or invertebrate origins, the top blastp scores of the deducted proteins against each database were compared. If the ratio between the top two scores was larger than 1.25 and the e value of the blastp against pathogen or vertebrate was smaller than 1e-15, then the CDS was assigned to the top-scoring organism group.

Raw sequences were deposited on the Sequence Read Archive (SRA) from the NCBI under bioproject accession PRJNA191820. The individual run files received accession numbers SRR206936, SRR206937, SRR206938, SRR206946, SRR206947, SRR206948, SRR206952, SRR206983, and SRR206984. A total of 2,475 coding sequences and their translations were submitted to the Transcriptome Shotgun Assembly (TSA) project deposited at DDBJ/EMBL/GenBank under the accessions GAHY01000001-2475.

Differential expression analysis

For differential expression analysis, RPKM values were transformed into z-scores, using data from all libraries to calculate the mean Log. For heatmap analysis, we calculated the mean z-score for each protein class based on their putative role (*e.g.* protein synthesis machinery, amino-acid metabolism, detoxification, etc). Heatmap of protein classes was constructed using Heatmapper[41]. Since we were interested only in having an insight in contig abundancy, we did not perform in depth differential expression analysis using common approaches (*e.g.* DEGseq or edgeR). Thus, we do not present and discuss any DE contigs in terms of statistical significance.

Evolutionary analysis

Protein sequences from other organisms were obtained at NCBI and aligned with Muscle[42]. Evolutionary analyses were conducted in MEGA7[43]. The evolutionary history was inferred using the Neighbor-Joining method (10.000 replicates; pairwise deletion)[44]. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All accession numbers are shown in the respective figures.

References

1. Organization, W. H. Chagas disease in Latin America: an epidemiological update based on 2010 estimates. *Wkly. Epidemiol. Rec.* **6,** 33–43 (2015).

2. Beament, J. Wigglesworth's contribution to insect biology and the state of the art. *Insect Biol. Futur.* 1–17 (1980). doi:10.1016/B978-0-12-454340-9.50005-7

3. Ribeiro, J. M. . *et al.* Exploring the sialome of the blood-sucking bug Rhodnius prolixus. *Insect Biochem. Mol. Biol.* **34,** 61–79 (2004).

4. Ribeiro, M. C. *et al.* An Insight into the Transcriptome of the Digestive Tract of the Bloodsucking Bug , Rhodnius prolixus. **8,** (2014).

5. Medeiros, M. *et al.* Transcriptome and gene expression profile of ovarian follicle tissue of the triatomine bug Rhodnius prolixus. *Insect Biochem. Mol. Biol.* **41,** 823–831 (2011).

6. Sonenshine, D. E. *et al.* First transcriptome of the testis-vas deferens-male accessory gland and proteome of the spermatophore from dermacentor variabilis (acari: Ixodidae). *PLoS One* (2011). doi:10.1371/journal.pone.0024711

7. Wigglesworth, V. B. The function of the Corpus allatum in the growth and reproduction of Rhodnius prolixus (Hemiptera). *Q. J. Microsc. Sci.* **79,** 91–121 (1936).

8. Davey, K. G. Spermatophore production in Rhodnius prolixus. *Q. J. Microsc. Sci.* **100,** 221–230 (1959).

9. Leopold, R. A. The role of male accessory glands in insect reproduction. *Annu. Rev. Entomol* **21,** 199–221 (1976).

10. Mikhaylova, L. M., Nguyen, K. & Nurminsky, D. I. Analysis of the Drosophila melanogaster testes transcriptome reveals coordinate regulation of paralogous genes. *Genetics* **179,** 305–315 (2008).

11. Azevedo, R. V. D. M. *et al.* The transcriptome of Lutzomyia longipalpis (Diptera: Psychodidae) male reproductive organs. *PLoS One* **7,** e34495 (2012).

12. Sonenshine, D. E. *et al.* First transcriptome of the testis-vas deferens-male accessory gland and proteome of the spermatophore from Dermacentor variabilis (Acari: Ixodidae). *PLoS One* **6,** (2011).

13. Gibbons, I. R. Cilia and flagella of eukaryotes. *J. Cell Biol.* **91,** 107s–124s (1981).

14. Witman, G. B. & Minervini, N. Dynein arm conformation and mechanochemical transduction in the eukaryotic flagellum. *Symp. Soc. Exp. Biol.* **35,** 203–23 (1982).

15. Brosseau, G. Genetic analysis of the male fertility factors on the Y chromosome of Drosophila melanogaster. 257–274 (1960).

16. Carvalho, A. B., Lazzaro, B. P. & Clark, A. G. Y chromosomal fertility factors kl-2 and kl-3 of Drosophila melanogaster encode dynein heavy chain polypeptides. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 13239–44 (2000).

17. Gepner, J. & Hays, T. S. A fertility region on the Y chromosome of Drosophila melanogaster encodes a dynein microtubule motor. *Proc. Natl. Acad. Sci.* **90,** 11132–11136 (1993).

18. de Kretser, D. M., Loveland, K. L., Meinhardt, A., Simorangkir, D. & Wreford, N. Spermatogenesis. *Hum. Reprod.* **13,** 1–8 (1998).

19. Schreiber, G., Carvalho, H. C. & Espinola, H. N. Genetics of Triatoma and Other Vector Species. in *Genetics of Insect Vectors of Diseases* (eds. Wright, J. W. & Pal, R.) 441–456 (Elsevier Inc., 1967).

20. Schreiber, G. Vinte e cinco anos de citogenética de Triatominae. *Rev. Soc. Bras. Med. Trop.* **8,** 123–127 (1974).

21. Solari, A. J. Autosomal synaptonemal complexes and sex chromosomes without axes in Triatoma infestans (Reduviidae; Hemiptera). *Chromosoma* **72,** 225–240 (1979).

22. Sun, X., Kovacs, T., Hu, Y.-J. & Yang, W.-X. The role of actin and myosin during spermatogenesis. *Mol. Biol. Rep.* **38,** 3993–4001 (2011).

23. Ribeiro, J. M. C. *et al.* An Insight into the Transcriptome of the Digestive Tract of the Bloodsucking Bug, Rhodnius prolixus. *PLoS Negl. Trop. Dis.* **8,** 27 (2014).

24. Yao, P.-L., Lin, Y.-C. & Richburg, J. H. TNF Alpha-Mediated Disruption of Spermatogenesis

in Response to Sertoli Cell Injury in Rodents Is Partially Regulated by MMP21. *Biol. Reprod.* **80,** 581–589 (2009).

25. De, S. K. Expression of tumor necrosis factor-alpha in mouse spermatogenic cells. *Endocrinology* **133,** 389–396 (1993).

26. Benahmed, M. Role of tumor necrosis factor in the male gonad. *Contracept. Fertil. Sex.* **25,** 569–71

27. Fuller, M. T. & Spradling, A. C. Male and Female Drosophila Germline Stem Cells: Two Versions of Immortality. *Science (80-.).* **316,** 402–404 (2007).

28. Grzyb, J., Latowski, D. & Strzałka, K. Lipocalins – a family portrait. *J. Plant Physiol.* **163,** 895–915 (2006).

29. Flower, D. R. Experimentally determined lipocalin structures. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology* **1482,** 46–56 (2000).

30. Mans, B. J., Ribeiro, J. M. C. & Andersen, J. F. Structure, function, and evolution of biogenic amine-binding proteins in soft ticks. *J. Biol. Chem.* **283,** 18721–33 (2008).

31. Mans, B. J. & Ribeiro, J. M. C. Function, mechanism and evolution of the moubatin-clade of soft tick lipocalins. *Insect Biochem. Mol. Biol.* **38,** 841–852 (2008).

32. Assumpc, T. C. F. *et al.* An insight into the sialome of the blood-sucking bug Triatoma infestans, a vector of Chagas' disease. *Insect Biochem. Mol. Biol.* **38,** 213–232 (2008).

33. Leone, M. G., Haq, H. A. & Saso, L. Lipocalin type prostaglandin D-synthase: which role in male fertility? *Contraception* **65,** 293–295 (2002).

34. Watanabe, H. *et al.* Lipocalin 2 binds to membrane phosphatidylethanolamine to induce lipid raft movement in a PKA-dependent manner and modulates sperm maturation. *Development* **141,** 2157–2164 (2014).

35. Kotłowska, M., Kowalski, R., Glogowski, J., Jankowski, J. & Ciereszko, A. Gelatinases and serine proteinase inhibitors of seminal plasma and the reproductive tract of turkey (Meleagris gallopavo). *Theriogenology* **63,** 1667–1681 (2005).

36. Dietz, J. A. *et al.* Spink2 Modulates Apoptotic Susceptibility and Is a Candidate Gene in the Rgcs1 QTL That Affects Retinal Ganglion Cell Death after Optic Nerve Damage. *PLoS One* **9,** e93564 (2014).

37. Alvarenga, P. H. *et al.* The Function and Three-Dimensional Structure of a Thromboxane A2/Cysteinyl Leukotriene-Binding Protein from the Saliva of a Mosquito Vector of the Malaria Parasite. *PLoS Biol.* **8,** e1000547 (2010).

38. Calvo, E., Mans, B. J., Ribeiro, J. M. C. & Andersen, J. F. Multifunctionality and mechanism of ligand binding in a mosquito antiinflammatory protein. *Proc. Natl. Acad. Sci.* **106,** 3728–3733 (2009).

39. Mans, B. J., Calvo, E., Ribeiro, J. M. C. & Andersen, J. F. The Crystal Structure of D7r4, a Salivary Biogenic Amine-binding Protein from the Malaria Mosquito Anopheles gambiae. *J. Biol. Chem.* **282,** 36626–36633 (2007).

40. Sirot, L. K. *et al.* Towards a Semen Proteome of the Dengue Vector Mosquito: Protein Identification and Potential Functions. *PLoS Negl. Trop. Dis.* **5,** e989 (2011).

41. Babicki, S. *et al.* Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* **44,** W147–W153 (2016).

42. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).

43. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33,** 1870–1874 (2016).

44. Saitou, N. & Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425 (1987).

Acknowledgments

This research is part of the Brazilian research program Institutos Nacionais de Ciência e Tecnologia – Entomologia Molecular.

Author contributions

We declare that this work was done by the authors named in this article and all liabilities pertaining to claims relating to the content of this article will be borne by the authors. All authors wrote, reviewed and approved the manuscript including figures and tables. P.L.O., L.B.K. and J.M.R. conceived the study. L.B.K., P.L.O., D.M., M.M. and M.H.S. prepared samples for sequencing. JMR and RDM. assembled and annotated the transcriptome. J.C., R.N.A., M.R.V.S., N.F.G., M.H.P., G.D.P. and L.B.K contributed to testes transcriptome analysis. J.C. conducted abundance, heat-map and evolutionary analyses. J.C. and LBK edited the manuscript.

Additional information

Supplementary information

Supplementary dataset

Competing interests: The authors declare no competing interests.

Data availability: All sequences and annotation tables are freely available in the GenBank or as the Supplementary dataset.