Correction for Mesquita et al., Genome of Rhodnius prolixus, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Mesquita, Rafael D., et al. “Correction for Mesquita et al., Genome of Rhodnius Prolixus, an Insect Vector of Chagas Disease, Reveals Unique Adaptations to Hematophagy and Parasite Infection.” Proc Natl Acad Sci USA 113, no. 10 (February 29, 2016): E1415–E1416.</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.1600205113">http://dx.doi.org/10.1073/pnas.1600205113</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences (U.S.)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Fri Apr 12 03:39:56 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/110831">http://hdl.handle.net/1721.1/110831</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
**Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection**


**Rhodnius prolixus** not only has served as a model organism for the study of insect physiology, but also is a major vector of Chagas disease, an illness that affects approximately seven million people worldwide. We sequenced the genome of *R. prolixus*, generated assembled sequences covering 95% of the genome (~702 Mb), including 15,456 putative protein-coding genes, and completed comprehensive genomic analyses of this obligate blood-feeding insect. Although immune deficiency (IMD)-mediated immune responses were observed, *R. prolixus* putatively lacks key components of the IMD pathway, suggesting a reorganization of the canonical immune signaling network. Although both Toll and IMD effectors controlled intestinal microbiota, neither affected *Trypanosoma cruzi*, the causal agent of Chagas disease, implying the existence of evasion or tolerance mechanisms. *R. prolixus* has experienced an extensive loss of selenoprotein genes, with its repertoire reduced to only two proteins, one of which is a selenocysteine-based glutathione peroxidase, the first found in insects. The genome contained actively transcribed, horizontally transferred genes from *Wolbachia* sp., which showed evidence of codon use evolution toward the insect use pattern. Comparative protein analyses revealed many lineage-specific expansions and putative gene absences in *R. prolixus*, including tandem expansions of genes related to chemoreception, feeding, and digestion that possibly contributed to the evolution of a blood-feeding lifestyle. The genome assembly and these associated analyses provide critical information on the physiology and evolution of this important vector species and should be instrumental for the development of innovative disease control methods.

**Rhodnius prolixus** | genome | hematophagy | immunity | Chagas disease

**Data mining**


**Edited by Alberto Carlos Frasch, Universidad de San Martín and National Research Council (Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina), San Martín-C.P., Argentina, and approved October 6, 2015 (received for review June 3, 2015).**
Significance

*Rhodnius prolixus* is a major vector of Chagas disease, an illness caused by *Trypanosoma cruzi* which affects approximately 7 million people worldwide. This report describes the first genome sequence of a nondipteran insect vector of an important human parasitic disease. This insect has a gene repertoire substantially distinct from dipteran disease vectors, including immune signaling pathways that display major departures from the canonical network. Large gene expansions related to chemoreception, feeding, and digestion have facilitated triatomine adaptation to a blood-feeding lifestyle. This study provides information about the physiology and evolution of an important disease vector that will boost understanding of transmission of a life-threatening parasite and may lead to the development of innovative control methods.

Results and Discussion

**Genome Landscape.** We assembled 702.6 Mb (v3.0.1, designated RproC1) of the predicted 733-Mb genome size (3, 4), with mean depth coverage of 8x. The current assembly includes 27,872 scaffolds (SI Appendix, Fig. A1), with a measured GC content of 27.1% that was slightly lower than that of another hemipteran, the pea aphid, *Acyrthosiphon pisum* (29.2%), although comparable GC content was found in protein-coding regions (5). The consensus gene prediction (VectorBase 1.3) includes 15,456 protein-coding genes and 738 RNA genes (Figs. L4 and SI Appendix, Fig. A1). We also found 25 Y-chromosome linked scaffolds, including nine Y-linked genes.

Orthologous gene clustering analyses (6) classified *R. prolixus* genes with orthologs in other insects and outgroup species (Fig. L8), and highlighted 630 lineage-specific expansions (LSE) in many *R. prolixus* gene groups related to defense mechanisms, behavior, development, and physiology. A comparison of gene families based on protein domain annotations revealed a few putative lineage-specific reductions (LSR), where *R. prolixus* possessed substantially fewer genes. These LSEs and LSRs will be specifically discussed below in relation to their biological roles.

**Transposable Elements and Horizontal Gene Transfers.** Approximately 1,400 transposable elements (TEs), with members from most of the known TE superfamilies, comprised 5.6% of the genome; this percentage was much lower relative to *Aedes aegypti* (~50%) and *Anopheles gambiae* (~20%) but similar to other anophelines (2–11%) (7). Nearly 70% of the TEs belonged to class II (mariner TEs) and highlighted 630 lineage-specific expansions (LSE) in many *R. prolixus* gene groups related to defense mechanisms, behavior, development, and physiology. A comparison of gene families based on protein domain annotations revealed a few putative lineage-specific reductions (LSR), where *R. prolixus* possessed substantially fewer genes. These LSEs and LSRs will be specifically discussed below in relation to their biological roles.

**Immune Pathways and Their Effects on Intestinal Microbiota and *T. cruzi.*** We identified most of the genes from canonical immune pathways, including the Toll, immune deficiency (IMD), and Jak/STAT pathways, and several immune effectors, including a substantial LSE of defensins (SI Appendix, Fig. A5 and Table D12 in Dataset S1). However, some canonical components of the IMD pathway were not detected: IMD, Fas-associated protein with death domain (*FADD*), death-related ced-3/Nedd2-like caspase (*Dredd*), and CaspAr. A similar observation was reported for the hindgut, where loss of the IMD pathway genes purportedly allowed the development of its obligate endosymbiont (16). In *R. prolixus*, however, some members of the IMD pathway were found, such as the peptidoglycan recognition proteins (PGRPs) and the NF-kB/Rel homolog, *rpr* (KP129556) (Fig. 2D). The expression of the latter was up-regulated in both the intestinal epithelium and fat body 24 and 72 h after a blood meal (Fig. 2B and C), thereby indicating an active IMD pathway in *R. prolixus*. This possibility was investigated by *rpr* knock-down (SI Appendix, Fig. A6D), which decreased the midgut defensin A (AAO74624), whereas expression of *lysozyme B* (ABX11554) remained unchanged and *lysozyme A* expression (ABX11553) increased (Fig. 2D). These data suggest that *rpr* directly controls defensin A expression, but not the lysozymes. Silencing of *rpr* also increased the population of the symbiotic bacteria *Rhodococcus rhodnii* (Fig. 2E–G), thus providing further support for an active IMD pathway despite the lack of several canonical proteins. This could be explained by either the existence of unknown alternative components linking the PGRP receptors to *rpr* or, alternatively, to a novel rewiring of the immune network (Fig. 2F). The role of the IMD pathway in the control of the gut microbiota led us to investigate its role in the control of *T. cruzi*. Unexpectedly, *rpr*-silenced insects infected with *T. cruzi* (strain DM25c) did not change parasite loads after 7 (SI Appendix, Fig. A6C) or 14 days post-infection (Fig. 2H). Even more surprisingly, silencing the transcription factor *rpDorsal* of the Toll pathway (17) (SI Appendix, Fig. A6D), did not change parasite levels (Figs. 2H and SI Appendix, Fig. A6C). These findings strongly indicate that either *T. cruzi* infection does not activate the insect’s immune system or that the parasite is not affected by antimicrobial peptides that are produced in response to parasitism or the ingestion of a blood meal. A plausible hypothesis for these data is that *T. cruzi* developed evasive or tolerance mechanisms, because activation of these immune pathways did not control *T. cruzi* populations, as previously suggested (18, 19).
Selenocysteine Machinery, Selenoproteins, and Detoxification Enzymes.

Selenocysteine (Sec) is a proteinogenic amino acid that is incorporated into proteins during translation by the reassignment of specific UGA codons. We identified enzymes involved in Sec synthesis and insertion (Table D13 in Dataset S1), but only two Sec proteins were found: selenophosphate synthetase 2 (RPRC0109014), which is involved in Sec synthesis, and glutathione peroxidase (GPx; RPRC011108), a major antioxidant protein. In all insect genomes sequenced to date, the known GPx genes were non-selenium, cysteine-based enzymes. It was previously hypothesized that Sec was replaced by cysteine in an ancient common ancestor of all insects. A recent discovery of a Sec-based GPx in an insect brings new insight into the evolution of this gene family, indicating that the Sec-GPx repertoire of selenoproteins, and that Sec loss and replacement by cysteine occurred independently in different species. This discovery is important because it suggests that insects may have developed an alternative antioxidant system to cope with oxidative stress. The identification of Sec-based GPx in an insect has implications for the understanding of insecticide resistance and could potentially be used to develop new insecticides.

Fig. 2. Characterization of IMD pathway and control of T. cruzi replication. Although several proteins from the IMD pathway are missing from the R. prolixus genome, the pathway is still active. (A) The rpRelish protein contains a Rel homology domain (RHD), an Ig-like fold, Plexins, a transcription factor domain (IPT), and several ankyrin (Ank) domains. (B and C) In response to the growth of the native microbiota following a blood meal, rpRelish expression is increased at 24 and 72 h post-blood meal in both the midgut (B) and fat body (C). (D) The rpRelish expression is knocked down (pink) in the gut using RNAi (SI Appendix, Fig. A6B), the expression of defensin A (Def) is greatly reduced, and the expression of lysisyme A (LyzA) is increased (D). (E) Upon knockdown of rpRelish expression (pink), the bacteria load in the anterior midgut (AM) is increased, with the same trend observed in the rectum (R) (G). dsMal control was shown in gray. *P < 0.05. Upon knockdown of rpDorsal or rpRelish expressions (SI Appendix, Fig. A6B), T. cruzi levels are not altered in any section of the digestive tract 14 d after infection (H). dsMal control (gray), dsRelish (pink) and dsDorsal (blue). Median values are at the graphic top. (I) Representation of the IMD R. prolixus immune pathways depicting the activity of key IMD pathway components. In R. prolixus, although the Toll and Jak/STAT pathways are usually highly conserved, members of IMD pathway are missing (dashed lines shapes with red ‘*’), including the IMD, Fadd, and Dredd genes and the negative modulators Pirk (poor IMD response upon knock-in) and Caspar. However, the pathway receptors, PGRPs and a homolog of the transcription factor Relish (rpRelish) are present, suggesting that the activity of the pathway is exerted through a non-canonical mechanism.
**Signaling and Development.** We identified and manually curated most of the major metazoan cell signaling pathways related to development and metabolism (Table D16 in Dataset S1). Tyrosine kinases are a class of protein kinases found exclusively in metazoans that regulate functions related to multicellularity, including intercellular communication, growth, differentiation, adhesion, and cell death. The tyrosine kinase (TK) of *R. prolixus* contained only 17 genes (Table D17 in Dataset S1), which were confined to 13 subfamilies (10 genes encode receptor TKs and 7 encode soluble TKs). This represents the smallest TK described to date; for comparison, *A. gambiae* has 32 TKs and *H. sapiens* has 90 TK members distributed among 30 subfamilies (24, 25).

Using orthologous relations to other insects (26), we identified genes involved in oogenesis, anteposterior (AP) and dorso-ventral (DV) axes determination (Tables D18 and D19 in Dataset S1). Of note, the germ plasm genes *osk* and *valos* (vls) were absent from *R. prolixus*. The absence of *osk* likely reflects a gene loss in insect evolution, because it was present in the transcriptome of the basally branching cricket *Gryllus bimaculatus* (27), where it was incorporated during neural development. Among the embryonic AP genes, *Kruppel* (RPRC001002) and *giant* (RPRC001027) were functionally confirmed as gap genes (28, 29). We identified single-copy orthologs for the DV patterning genes, including most Toll and bone morphogenetic protein (BMP) pathway elements. The BMP type 1 receptor, however, was not found in *R. prolixus* or in previous analyses (30). Our functional analysis showed that one of the *tP Toll* genes (RPRC009262), similarly to that reported for hymenopterans (31, 32), performs both DV and AP embryonic patterning roles (30), pointing to a potentially novel role that this pathway element might display during embryogenesis. The role of hypoxia in the development of the tracheal system—originally reported in *R. prolixus* by Wigglesworth (33)—involves a set of remarkably conserved genes (Table D20 in Dataset S1) (34). The role of the corpora allata hormone, later named juvenile hormone (JH), in the control of insect metamorphosis and reproduction was also another of Wigglesworth’s landmarks in the history of biology using *R. prolixus* as the experimental model (35). Although eight different forms of JH have been identified in other insects, ironically, the identity of the JH forms present in triatomines are still unknown. Nevertheless, we identified *R. prolixus* genes coding for the JH biosynthetic pathway/signaling system (36) (Table D21 in Dataset S1).

Among developmental genes, marked LSEs have occurred in several transcription factor (TF) families. Recent independent expansions of the Pipsqueak TF family in *R. prolixus* and *A. pisum* were driven in part by transposons, with the potential reuse of DNA-binding domains of transposases as new TFs. Given the developmental roles of Pipsqueak TFs in *Drosophila*, these LSEs might have played some role in the morphological and behavioral diversification of hemipterans. Other predicted TF expansions were also found as resulting from proliferation of transposable elements (37). Other expansions related to development were found in cuticle genes, including those that transcribe cuticular proteins and sclerotization enzymes (SI Appendix, Figs. A7–A10 and Tables D22 and D23 in Dataset S1).

**Hematophagy-Related Genes.** All developmental stages of *R. prolixus*, from first instar nymphs to adults of both sexes, feed exclusively on blood. Chemoreception is essential to host finding (38), and several LSEs that we identified included odorant and gustatory receptors as well as odorant and chemosensory binding proteins (SI Appendix, Figs. A11–A14 and Tables D24–D27 in Dataset S1). Some of the recent LSEs, such as OR58-87 (SI Appendix, Fig. A11) (named “recent Rhodnius expansion”) were found in tandem arrays (Table D24 in Dataset S1). As in other blood-feeding insects, detection of carbon dioxide contributes to host finding in triatomines (39). Nevertheless, the carbon dioxide receptor subfamily was absent in *R. prolixus*, as previously reported in the body louse and the pea aphid, despite being highly conserved in holometabolous insects (40). The sugar receptor subfamily was also missing in *R. prolixus*, but this may represent a more recent loss as it was found in the pea aphid (SI Appendix, Fig. A12). *Rhodnius* saliva was previously studied for its capacity to interfere with host blood clotting, platelet aggregation, and vasocostriction (41), and also for its immunomodulatory activity. Lipocalin genes (*r = 51* genes, some previously described (10, 41)] that transcribe the most abundant salivary proteins, formed large LSEs, and occurred as tandem clusters (SI Appendix, Fig. A15 and Table D28 in Dataset S1). We discovered 12 previously unidentified members of the nitrophorin clade (Fig. 3B), which encode lipocalins that carry nitric oxide in *R. prolixus* saliva. We also found LSEs occurring in tandem clusters for aspartic peptidases (cathepsin-D like aspartic proteases), probably reflecting a replacement of digestive serine proteases found in most insects by lysosomal proteases for blood digestion in hemipterans (Table D29 in Dataset S1). Among the putative LSRs (Table D30 in Dataset S1), a reduced set of only seven amiloride-sensitive sodium channels (SI Appendix, Fig. A16) may also be linked to hematophagy because amiloride-sensitive sodium channels are involved in the malpighian tubules filtration. This paucity could have evolved to limit sodium transport for the production of hypoosmotic urine and can be related to the unusually weak response of *R. prolixus* sodium channels to amiloride (42).

Moreover, we manually annotated other gene families related to hematophagy but without displaying signals of LSE. These families included the following pathways/functions: iron/heme binding, transport and metabolism (43, 44) (Table D31 in Dataset S1); behavioral control of the blood-seeking habits (45), including locomotory and visual activities, memory formation, temperature and humidity detection, mechanoreception, and circadian clock control (Tables D32 and D33 in Dataset S1); energy metabolism pathways (glycolysis, gluconeogenesis, and pentose-phosphate pathways (Table D34 in Dataset S1)); and regulatory peptides and receptors, as well as 37 neuropeptide precursors or hormone genes (46).

As reported for *A. pisum* (47), most urea cycle enzymes were not observed, with the exception of arginosuccinate synthetase.

**Fig. 3.** Thioredoxin reductases and nitrophorins trees. (A) The tree shows the phylogenetic relationship of thioredoxin reductases (TR) in nine Paraneoptera genomes. Colored balls indicate the amino acid aligned to the Sec position: green (U) selenocysteine, red (C) cysteine, and gray (−) unknown/unaligned. One of the two Cys-TR in *R. prolixus* (RPRC014349) clusters with the Sec-TR proteins, implying a Sec to Cys conversion. (B) Nitrophorins branch of lipocalins tree exemplify a *Rhodnius* enriched cluster. Sequences clustered by OrthoMCL on group 691 together with unclustered sequences similar to lipocalins were aligned and a tree constructed (SI Appendix, Fig. A15). Black dots mark new *Rhodnius* sequences. TRIDI, *Triatoma dimidiata*; TRIMA, *Triatoma matogrossensis*. Support values based on bootstrap are included at nodes.

Mesquita et al. PNAS | December 1, 2015 | vol. 112 | no. 48 | 14939
Backgrounds by Ex-Wolbachia. Absence of both pathways explains the lack of genome coverage of all TE using BLAST (57). We identified genomic duplicated regions using BLASTN (52) of all scaffolds against each other with 5-kb cut-off to avoid TE detection.

miRNA precursors and mature sequences were retrieved using Enviest (58) and filtered. Their target genes were predicted using miRanda (59).

Selenoprotein genes were identified using Selenoproteins (60) and Sebastian (61). We also identified SECS elements using SECSearch (61). We used TRNAScan-SE (62) to detect tRNA-Sec. Selenoproteins multiple sequence alignments used T-Coffee (63) and then performed phylogenetic analyses by reconstructing trees with maximum likelihood using the best-fitting evolutionary model (64).

Wolbachia similar regions were searched using BLASTN (52) to compare the Rhodnius genome with 16 Wolbachia genomes. The gene models present in these regions were selected and grouped with those having a Wolbachia protein as best hit in a BLASTP (52). This gene group had their codon-use compared with Rhodnius and Wolbachia backgrounds by ExPander Tool (65).

Immunity genes were searched in expressed sequence tags, gene prediction, assembles genome, and genomic unsequenced raw reads to assure reliability of the absence; all searches used BLAST (52) and homologous genes from closely related taxa. The rpRelish gene was manually assembled using contigs of a previously published transcriptome (10). For confirmation, rpRelish was cloned and sequenced. Insects were immunologically challenged through blood meal as it increases endosymbiont (R. rhodnii) population in the gut. Total RNA was extracted using the TRIzol reagent (Invitrogen) following the manufacturer’s instructions. Real-time PCR (quantitative RT-PCR) was used to assess the transcript abundance and silencing efficiency of the genes of interest. T7 Megascript kit (Ambion) was used to generate and purify dsRNA from PCR-amplified genes following the manufacturer’s instructions. The dsRNA in sterile water was introduced into the thorax of adult females by injection. The microbiota of R. prolixus digestive tract was analyzed by counting colony forming units after plating homogenates of the different sections of the midgut.

Materials and Methods

All candidate colonies were genotyped to assure R. prolixus identity before extracting genetic material from ovaries or testes. Whole-genome shotgun sequencing using Sanger and 454 technologies produced 8x genome coverage. After end-sequencing a BAC library, we assembled all data by implementing the default parameters in CABOG (49).

We predicted gene sequences using ab initio and similarity-based approaches that were merged to construct the final set of gene models. We performed orthology-based clustering using OrthoMCL (6) and gene family based on conserved protein domain annotations from InterProScan (51) for 16 species.

We used three different methods for identifying TEs. First, we identified class I and non-long-terminal-sequences (non-LTRs) with RPS-blast (52). Second, we identified LTRs using homology-based approaches (53, 54). The third method, used to discover MITEs, relied upon Repeatmasker (55) and a repeat library produced with RepeatScout (56). We assessed the genomic coverage of all TE using BLAST (57). We identified genomic duplicated regions using BLASTN (52) of all scaffolds against each other with 5-kb cut-off to avoid TE detection.

ACKNOWLEDGMENTS. We thank Charles B. Beard for the Rhodnius prolixus picture. Genome sequencing was funded by National Institutes of Health Grant N01-HG-100307; VectorBase was supported by National Institutes of Health National Institute of Allergy and Infectious Diseases Grant HHSN272200900639C; Brazilian groups were funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Fundação de Amparo à Pesquisa do Estado de São Paulo, and Fundação de Amparo à Pesquisa do Estado de Minas Gerais; R.R.-P. was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/FAPERJ; and R.M.W. was supported by Marie Curie PIOP-GA-2011-303312.


