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>Tue Feb 12 05:43:21 EST 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 membrane transport, 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

Dating back to Wigglesworth’s pioneering work in the 1930s (1), *Rhodnius prolixus* (Fig. 1) has served as a model organism for the study of cellular and physiological processes in insects, such as gametogenesis, the role of hemolymph proteins and lipids in oogenesis, and ion and water transport mechanisms. As an obligate blood-feeding hemipteran, this insect has adapted remarkably well to digesting and eliminating the potentially toxic by-products of blood digestion. *R. prolixus* is also a major vector of *Trypanosoma cruzi*, the parasitic protozan that causes Chagas disease in humans. This disease, commonly considered a disease of the poor, causes premature heart failure in humans and is responsible for high economic impact. The raw genomic reads have been deposited in the GenBank database (accession no. KP129556). The genomic assembly (RproC1) and supplementary material were deposited in the VectorBase repository, https://www.vectorbase.org. The raw genomic reads have been deposited in the Sequence Read Archive (BioSample accession nos. SAMN00993539 and SAMN00099354). See also Genbank Assembly accession no. GCA_000181055.2


This article is a PNAS Direct Submission.

Data deposition: The cloned cDNA sequences reported in this paper have been deposited in the GenBank database (accession no. KP129556). The genomic assembly (RproC1) and the consensus gene prediction (version 1.3) have been deposited in the VectorBase repository, https://www.vectorbase.org. The raw genomic reads have been deposited in the Sequence Read Archive (BioSample accession nos. SAMN00993539 and SAMN00099354). See also Genbank Assembly accession no. GCA_000181055.2

1. R.D.M. and R.J.V.-A. contributed equally to this work.
2. To whom correspondence should be addressed. Email: rdmesquita@iq.ufrj.br, edoslon@cck.gov, or pedro@biogmed.ufrj.br.
3. Present address: Centro de Pesquisa René Rachou, Minas Gerais 30190-002, Brazil.
4. E.M.D. and P.L.O. contributed equally to this work. This article contains supporting information online at www.pnas.org/lookup/suppl;doi/10.1073/pnas.1506226112/-/DCSupplemental.
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.

and social costs. Approximately 10,000 people die from the disease annually and 100 million people are at risk for infection (2).

Results and Discussion

Genome Landscape. We assembled 702.6 Mb (v3.0.1, designated rpDorsal vol. 112) 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. 1B), 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 a single superfamily (DTRP1-7), composed of seven new families, represents almost 3% of the genome (Fig. 1A and C and SI Appendix, Tables A1 and A2). Many of these sequences have full-length transposases and ORFs suggesting a recent period of increased transposition (8).

Horizontal gene transfers from Wolbachia were likely the origin of 27 genes (Tables D1–D3 in Dataset S1) located in some of the 85 Wolbachia-like genomic regions. These regions spanned 100–2,500 bp, which were smaller and more dispersed than those reported in Glossina morsitans (two segments of ∼500 Kb) (9). A previous transcriptome analysis showed that eight of these genes appear to be transcribed in the midgut (10). Codon use analysis of the horizontally transferred genes (HTG) revealed that, of the 61 amino acid codons, 35 codons showed a shift from a Wolbachia spp. use bias toward the Rhodnius codon frequency (SI Appendix, Fig. A2 and Table D4 in Dataset S1). This codon frequency shift suggests an ongoing adaptive process whereby the HTG coevolved with the insect tRNA profile for more efficient translation. The HTG also revealed two transposases (RPRC000742 and RRPC000770), one reverse transcriptase (RPRC000723), and DNA recombination/repair enzymes, including one Holliday junction helicase (RPRC004559), two DNA mismatch repair MutL proteins (RPRC010818 and RRPC011745), and one DNA polymerase I (RPRC006597). These findings suggest that extensive machinery to transpose, recombine, and repair the host DNA were likely instrumental for the success of Wolbachia gene transfer to R. prolixus.

RNAi Machinery and Target Genes. RNAi is a posttranscriptional gene-silencing mechanism triggered by double-stranded RNAs (dsRNAs) that degrade a target messenger RNA (mRNA) in a sequence-specific manner (11). The RNAi mechanism can also be triggered by microRNAs (miRNAs), which are small noncoding RNAs (12). We identified RNAi machinery (Table D5 in Dataset S1) as well as precursors and 87 mature miRNA sequences (SI Appendix, Figs. A3 and A4 and Tables D5–D9 in Dataset S1) that comprised a complete gene-silencing pathway for 804 potential target genes (SI Appendix, Fig. A3C and Tables D10 and D11 in Dataset S1). Some miRNA clusters were similar to those described for Caenorhabditis elegans (13), such as the cluster rpr-miR-71/2a-2/13a/13b/2a-1 (SI Appendix, Fig. A3B) that was identified in one of the introns of the phosphatase-four-like protein gene (RPRC004331). The miRNA-target gene pairs also showed conservation; for example, rpr-miR-124-3p targeting the ROCK1 protein kinase gene (RPRC007372-RA) (14) and rpr-miR-10-3p targeting the Huntington-like gene (RPRC000630-RA) (15).

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, Fax-associated protein with death domain (Fadd), death-related ced-3/Nedd2-like caspase (Dredd), and Caspar. A similar observation was reported for the Ae. aegypti, where 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, rprRelish (KP129556) (Fig. 2A). 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 rprRelish knockdown (SI Appendix, Fig. A6B), which reduced the levels of midgut defensin A (AA074624), whereas expression of lysozyme B (ABX11554) remained unchanged and lysozyme A expression (ABX11553) increased (Fig. 2D). These data suggest that rprRelish directly controls defensin A expression, but not the lysozymes. Silencing of rprRelish 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 rprRelish 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, rprRelish-silenced insects infected with T. cruzi (strain DM28C) 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 active evasion or tolerance mechanisms, because activation of these immune pathways did not control T. cruzi populations, as previously suggested (18, 19).

Mesquita et al.

PNAS | December 1, 2015 | vol. 112 | no. 48 | 14937
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 SI Appendix), but only two Sec proteomes were found: selenophosphate synthetase 2 (RPRC0109014), which is involved in Sec synthesis, and glutathione peroxidase (Gpx; RRPC011108), a major antioxidant protein. In all insect genomes sequenced to date, the known Gpx genes were nonselenium, cysteine-based enzymes. It was previously hypothesized that Sec was replaced by cysteine in an ancient common ancestor of all insects. The discovery of a Sec-based Gpx in an insect brings new insight into the evolution of this gene family, indicating that the Sec-Gpx genes are more widespread and conserved than previously thought (20). We also observed a recent Sec-to-cysteine conversion in one (RPRC014349) of the two R. prolixus thioredoxin reductases (Fig. 34). These results, together with the finding of a Sec-Gpx in R. prolixus, indicate that insects originally had a more extensive repertoire of selenoproteins, and that Sec loss and replacement by cysteine occurred independently in different species.

A substantial LSE of enzymes involved in drug and detoxification occurred, such as in carboxylesterases and cytochrome P450 (Tables D14 and D15 in Dataset S1) (see also ref. 21). These enzymes, together with glutathione S-transferases, can drive insecticide resistance. Their identification is particularly relevant for public health surveillance, as recent reports indicated that natural triatomine populations have responded with increasingly higher levels of insecticide resistance and could threaten the successful vector control initiatives that have occurred over the last two and half decades in Latin America (22, 23).

Fig. 1. Rhodnius genome. (A) Genome overview. All scaffolds are represented in layer I and are organized clockwise from the longest to the shortest, starting at the arrowhead. The generic (layer II, red) and TEs (layer III, blue) showed opposite densities until the asterisk (*) and were similarly low from this point until the end (shorter scaffolds). The Wolbachia sp. insertions (layer IV, orange) were observed throughout the genome without a trend. The kernel picture illustrates an adult R. prolixus. (B) Gene clustering. The Venn diagram partitions 15,439 OrthoMCL gene clusters according to their species compositions for R. prolixus and three other Hemimetabola (blue), four Diptera (yellow), four other Holometabola (green), and four noninsect outgroup species (pink). The 6,993 R. prolixus genes show widespread orthology (white circle, and bars, Bottom Left): these are part of the 7,115 clusters that have representatives from each of the four species sets, of which a conserved core of 2,253 clusters have orthologs in all 16 species. The 5,498 R. prolixus genes show no confident orthology (bars, Bottom Right), but most of these are homologous (e-value < 1e-05) to genes from other animals or to genes in its own genome. (C) TE distribution. The inner chart represents the three main classes of TEs (LTRs, non-LTRs, and class II), and the outer shows the distribution of TE superfamilies within each class. The charts are based on the total base pairs occupied by TE-related sequences longer than 0.5 Kb, as shown in SI Appendix, Table A1.

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 (PTF), 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 postblood meal in both the midgut and fat body. (C) Fas: fasting. When 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 lysozyme A (LizA) is increased (D). Upon knockdown of rpRelish expression (pink), the bacteria load in the anterior midgut (AM) (E) and posterior midgut (PM) (F) 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 absence 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 “X”), 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 noncanonical 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 curated 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* 100 TK members distributed among 30 subfamilies (24, 25).

Using orthologous relations to other insects (26), we identified genes involved in oogenesis, anteroposterior (AP) and dorsoventral (DV) axes determination (Tables D18 and D19 in Dataset S1). Of note, the germ plasm genes *osk* and *valos* (vl) 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, *Knuppel* (RPRC000102) and *giant* (RPRC000127) 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 *rToll* 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* were driven in part by transposable elements, 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). 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 salvia* was previously studied for its capacity to interfere with host blood clotting, platelet aggre- gation, and vasoconstriction (41), and also for its immunomod- ulatory activity. Lipocalin genes (*n* = 51 genes, some previously described (10, 41)) which 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 hyposmotic 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, mechanosensation, 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. pism* (47), most urea cycle enzymes were not observed, with the exception of arginosuccinate synthetase.

---

**Fig. 3.** Thioether oxidoreductases and nitrophorins trees. (A) The tree shows the phylogenetic relationship of thioether oxidoreductase (TR) in nine Panoperaetra genomes. Colored balls indicate the amino acid aligned to the Sec position: green (U) selenocysteine, red (C) cysteine, and gray (V) 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. TRID, Triatoma dimidiata; TRIMA, Triatoma matagrossensis. Support values based on bootstrap are included at nodes.

Mesquita et al. PNAS  | December 1, 2015 | vol. 112 | no. 48 | 14939
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 automatic annotation and paralog clustering using AnoXcel (50). We performed orthology-based clustering using OrthoMCL (6) and gene family clusters 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 nonlong-terminal repeats (non-LTRs) with RPS-blast (52). Sec- ond, we identified LTRs using homology-based approaches (53, 54). The third method, used to discover MITEs, relied upon RepeatScout (56). We assessed the genomic coverage of all TE using BLAT (57). We identified genomic duplicated regions with GLASTN (52) of all scaffolds against each other with 5-Kb cut-off to avoid TE detection.

miRNA precursors and mature sequences were retrieved using Einverted (52) and filtered. Their target genes were predicted using miRanda (59). Selenoprotein genes were identified using Selenoprotocols (60) and Sebastian (61). We also identified SECS elements using SECSearch (62). 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 Ex- pander Tool (65).

Immunity genes were searched in expressed sequence tags, gene pre- diction, assembled genome, and genomic unassembled 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 in- terest. 7 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.

Acknowledgments.

We thank Charles B. Beard for the Rhodnius prolixus picture. Genome sequencing was funded by National Institutes of Health Grant HG003079; 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 Cientifico e Tecnológico, Fundación de Amargo 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 Consejo Nacional de Inves- tigaciones Científicas y Técnicas, Argentina, Agencia Nacional de Ciencia y Tecnología, and Fundación Bunge y Born; R.D.F.-M. was funded by Coordenación de Aperfeiçoamento de Pessoal de Nível Superior/FAPERJ, and R.M.W. was supported by Marie Curie PIF-GA-2011-303312.


