Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection

Rafael D. Mesquita[a,b,1,2], Raquel J. Vionette-Amaral[c,1], Carl Lowenberger[d], Rolando Rivera-Pomar[e,f], Fernando A. Monteiro[b,g], Patrick Minx[h], John Spieth[h], A. Bernardo Carvalho[b,i], Francisco Panzera[j], Daniel Lawson[k], André Q. Torres[a,g], Jose M. C. Ribeiro[l], Marcos H. F. Sorgine[b,c], Robert M. Waterhouse[m,n,o,p], Michael J. Montague[h], Fernando Abad-Franch[q,3], Michele Alves-Bezerra[c], Laurence R. Amaral[r], Helena M. Araujo[b,s], Ricardo N. Araujo[b,t], L. Aravind[u], Georgia C. Atella[b,c], Patricia Azambuja[b,g], Mateus Berni[s], Paula R. Bittencourt-Cunha[c], Gloria R. C. Braz[a,b], Gustavo Calderón-Fernández[v], Claudia M. A. Carareto[w], Mikkel B. Christensen[k], Igor R. Costa[c], Samara G. Costa[g], Marilvia Dansa[x], Carlos R. O. Daumas-Filho[c], Iron F. De-Paula[c], Felipe A. Dias[b,c], George Dimopoulos[y], Scott J. Emrich[z], Natalia Esponda-Behrens[e], Patricia Fampa[aa], Rita D. Fernandez-Medina[bb], Rodrigo N. da Fonseca[b,cc], Marcio Fontenele[b,s], Catrina Fronick[h], Lucinda A. Fulton[h], Ana Caroline Gandara[b,c], Eloi S. Garcia[b,g], Fernando A. Genta[b,g], Gloria I. Giraldo-Calderón[dd], Bruno Gomes[b,g], Katia C. Gondim[b,c], Adriana Granzotto[w], Alessandra A. Guarneri[b,ee], Roderic Guigó[ff,gg], Myriam Harry[hh,ii], Daniel S. T. Hughes[k], Willy Jablonka[c], Emmanuelle Jacquin-Joly[jj], M. Patricia Juárez[v], Leonardo B. Koerich[b,i], José Manuel Latorre-Estivalis[b,ee], Andrés Lavore[e], Gena G. Lawrence[kk], Cristiano Lazoski[b,i], Claudio R. Lazzari[ll], Raphael R. Lopes[c], Marcelo G. Lorenzo[b,ee], Magda D. Lugon[x], David Majerowicz[c,mm], Paula L. Marcet[kk], Marco Mariotti[ff,gg], Hatisaburo Masuda[b,c], Karine Megy[k], Ana C. A. Melo[a,b], Fanis Missirlis[nn], Theo Mota[oo], Fernando G. Noriega[pp], Marcela Nouzova[pp], Rodrigo D. Nunes[b,c], Raquel L. L. Oliveira[a], Gilbert Oliveira-Silveira[c], Sheila Ons[e], Lucia Pagola[e], Gabriela O. Paiva-Silva[b,c], Agustina Pascual[e], Marcio G. Pavan[g], Nicolás Pedrini[v], Alexandre A. Peixoto[b,g], Marcos H. Pereira[b,t], Andrew Pike[y], Carla Polycarpo[b,c], Francisco Prosdocimi[c], Rodrigo Ribeiro-Rodrigues[qq], Hugh M. Robertson[rr], Ana Paula Salerno[ss], Didier Salmon[c], Didac Santesmasses[ff,gg], Renata Schama[b,g], Eloy S. Seabra-Junior[ss], Livia Silva-Cardoso[c], Mario A. C. Silva-Neto[b,c], Matheus Souza-Gomes[r], Marcos Sterkel[c], Mabel L. Taracena[c], Marta Tojo[tt], Zhijian Jake Tu[uu], Jose M. C. Tubio[vv], Raul Ursic-Bedoya[d], Thiago M. Venancio[b,x], Ana Beatriz Walter-Nuno[c], Derek Wilson[k], Wesley C. Warren[h], Richard K. Wilson[h], Erwin Huebner[ww], Ellen M. Dotson[kk,2,4], and Pedro L. Oliveira[b,c,2,4]

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

Dating back to Wigglesworth's pioneering work in the 1930s (1), *Rhodnius prolixus* (Fig. 1A) 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 protozoan 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

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 RproC1) of the predicted 733-Mb genome size (3, 4), with mean depth coverage of 8×. 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. 1*A* 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. 1*B*), 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 (DTTRP1-7), composed of seven new families, represents almost 3% of the genome (Fig. 1 *A* 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, 38 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 RPRC000770), one reverse transcriptase (RPRC000723), and DNA

recombination/repair enzymes, including one Holliday junction helicase (RPRC004559), two DNA mismatch repair MutL proteins (RPRC010818 and RPRC011745), 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 D6–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/2c/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 (RPRC007732-RA) (14) and rpr-miR-10-3p targeting the Huntington-like gene (RPRC006430-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*, Fas-associated protein with death domain (*Fadd*), death-related ced-3/Nedd2-like caspase (*Dredd*), and *Caspar*. A similar observation was reported for the pea aphid, where a more extensive 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-κB/Rel homolog, *rpRelish* (KP129556) (Fig. 2*A*). 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. 2 *B* and *C*), thereby indicating an active IMD pathway in *R. prolixus*. This possibility was investigated by *rpRelish* knockdown (*SI Appendix*, Fig. A6B), which decreased expression of midgut *defensin A* (AAO74624), whereas expression of *lysozyme B* (ABX11554) remained unchanged and *lysozyme A* expression (ABX11553) increased (Fig. 2D). These data suggest that *rpRelish* directly controls *defensin A* expression, but not the lysozymes. Silencing of *rpRelish* also increased the population of the symbiotic bacteria *Rhodococcus rhodnii* (Fig. 2 *E–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 *rpRelish* or, alternatively, to a novel rewiring of the immune network (Fig. 2I). 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, *rpRelish*-silenced insects infected with *T. cruzi* (strain DM28C) did not change parasite loads after 7 (*SI Appendix*, Fig. A6C) or 14 days postinfection (Fig. 2H). Even more surprisingly, silencing the transcription factor *rpDorsal* of the Toll pathway (17) (*SI Appendix*, Fig. A6A), 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).
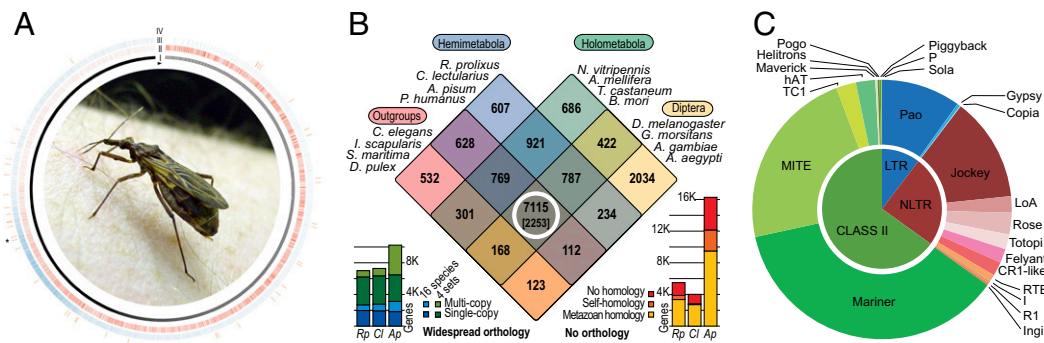
GENETICS

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 genic (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.

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 synthase 2 (RPRC009014), 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 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. 3*A*). 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).
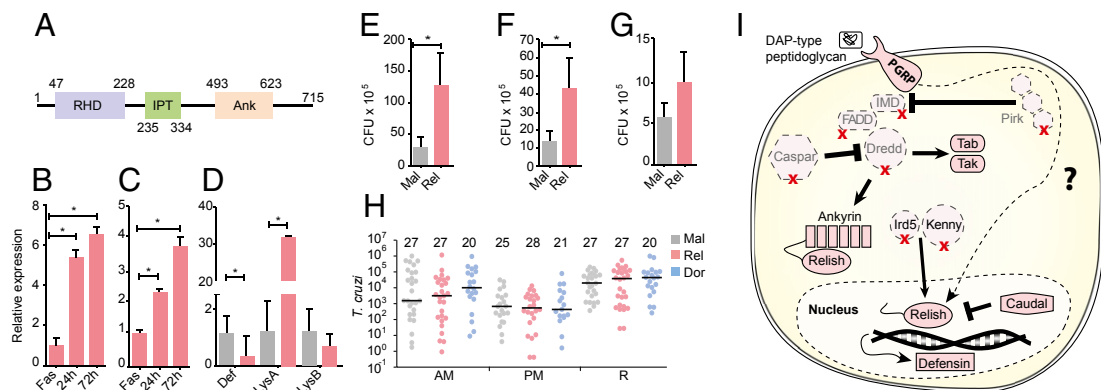

Fig. 2. Characterization of IMD pathway and control of *T. cruzi* replication. Although several proteins from the IMD pathway are missing from the *R. prolixus* genome, the pathway is still active. (*A*) The rpRelish protein contains a Rel homology domain (RHD), an Ig-like fold, Plexins, a transcription 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 postblood meal in both the midgut (*B*) and fat body (*C*). Fas: fasting. When *rpRelish* expression is knocked down (pink) in the gut using RNAi (*SI Appendix*, Fig. A6*B*), 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. A6), *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 kinome (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, anteroposterior (AP) and dorsoventral (DV) axes determination (Tables D18 and D19 in Dataset S1). Of note, the germ plasm genes *osk* and *valois* (*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* (RPRC000102) 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 I receptor *Sax*, however, was not found in *R. prolixus* or in previous analyses (30). Our functional analysis showed that one of the *rpToll* 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*, as well as *A. pisum*, were driven in part by transponsons, 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 vasoconstriction (41), and also for its immunomodulatory 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 proteinases found in most insects by lysosomal proteinases 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 argininosuccinate synthetase
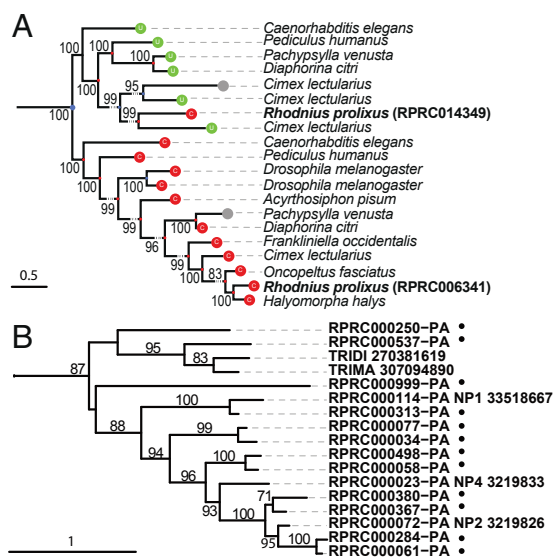
Fig. 3. Thioredoxin reductases and nitrophorins trees. (*A*) The tree shows the phylogenetic relationship of thioredoxin reductases (TR) in nine Paraneoptera genomes. Colored balls indicate the amino acid aligned to the Sec position: green (U) selenocysteine, red (C) cysteine, and gray (–) unknown/unaligned. 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.

(RPRC003927) (Table D35 in Dataset S1). The uricolysis pathway, which is present in mosquitoes (48), is putatively absent in *R. prolixus*. Absence of both pathways explains the lack of elimination of nitrogen as urea, also originally described by Wigglesworth, even before the discovery of the urea cycle (1). As is the case for most eukaryotes, the *R. prolixus* genome does not encode several pathways for essential amino acid synthesis (47) (Table D36 in Dataset S1), but the lack of almost all urea cycle enzymes (Table D35 in Dataset S1) also precludes the de novo synthesis of arginine. This is noteworthy because arginine is the precursor for nitric oxide, an important salivary vasodilator for blood-feeding insects. We postulate that adaptation to a diet rich in amino acids, such as blood, allowed for relaxation of the constraints on the arginine synthesis pathway.

This first analysis of the *R. prolixus* genome generated novel insights regarding multiple adaptive mechanisms that likely contributed to a strict blood-feeding lifestyle, and thus provided new working hypotheses that will drive future research in the biology of triatomines and Chagas disease. Some of the unique features of *R. prolixus*, including its numerous lineage-specific gene family expansions, the peculiar immune network, and a "silent" relationship between the insect host and the trypanosome parasite, also provide a new starting point for further understanding its evolutionary adaptations. In summary, our data illustrate the important role of *R. prolixus* as an additional study model to that of dipteran species in the vector biology arena and in the broader insect biology landscape as well. This work also reinforces the extraordinarily rich legacy of Wigglesworth, whose experimental studies on *R. prolixus* illuminated the basic concepts of insect physiology still used today. We hope the findings described here will allow the development of novel strategies to control this important Chagas disease vector.

Materials and Methods

All candidate colonies were genotyped to assure *R. prolixus* identity before extracting genetic material from ovaries or testis. Whole-genome shotgun sequencing using Sanger and 454 technologies produced 8× 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 counts based on conserved protein domain annotations from InterProScan (51) for 16 species.

We used three different methods for identifying TEs. First, we identified class II and nonlong-terminal repeats (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 Findmite (55) and a repeat library produced with RepeatScout (56). We assessed the genomic coverage of all TE using BLAT (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 Einverted (58) and filtered. Their target genes were predicted using miRanda (59).

Selenoprotein genes were identified using Selenoprofiles (60) and Seblastian (61). We also identified SECIS elements using SECISearch3 (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, assembled genome, and genomic unassembled raw reads to assure reliability of the absences; 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.

[a]Departamento de Bioquímica, Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-909, Brazil; [b]Instituto Nacional de Ciência e Tecnologia em Entomologia Molecular, Rio de Janeiro 21941-591, Brazil; [c]Programa de Biologia Molecular e Biotecnologia, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-591, Brazil; [d]Biological Sciences, Simon Fraser University, Burnaby, BC, Canada V5A 1S6; [e]Centro Regional de Estudios Genomicos, Universidad Nacional de La Plata, La Plata 1900, Argentina; [f]Centro de Bioinvestigaciones, Universidad Nacional del Noroeste de Buenos Aires, 2700 Pergamino, Argentina; [g]Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro 21040-900, Brazil; [h]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO 63108; [i]Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-590, Brazil; [j]Sección Genética Evolutiva, Facultad de Ciencias, Universidad de la República, Montevideo 11400, Uruguay; [k]European Molecular Biology Laboratory, European Bioinformatics Institute, Welcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom; [l]Section of Vector Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD 20852; [m]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 1211, Switzerland; [n]Swiss Institute of Bioinformatics, Geneva 1211, Switzerland; [o]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; [p]The Broad Institute of MIT and Harvard, Cambridge, MA 02142; [q]Grupo de Pesquisa em Ecologia de Doenças Transmissíveis na Amazônia, Instituto Leônidas e Maria Deane, Fundação Oswaldo Cruz, Amazonas 69057-070, Brazil; [r]Laboratório de Bioinformática e Análises Moleculares, Instituto de Genética e Bioquímica, Faculdade de Computação, Universidade Federal de Uberlândia, Minas Gerais 38700-002, Brazil; [s]Instituto de Ciências Biomédicas, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-591, Brazil; [t]Departamento de Parasitologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Minas Gerais 31270-901, Brazil; [u]National Library of Medicine, National Center for Biotechnology Information, National Institutes of Health, Rockville, MD 20894; [v]Instituto de Investigaciones Bioquímicas de La Plata, Facultad de Ciencias Médicas, Universidad Nacional de La Plata-CONICET, La Plata 1900, Argentina; [w]Departamento de Biologia, Universidade Estadual Paulista, São José do Rio Preto, São Paulo 15054-000, Brazil; [x]Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Rio de Janeiro 28013-602, Brazil; [y]Department of Molecular Microbiology and Immunology, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD 21205-2179; [z]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556; [aa]Departamento de Biologia Animal, Instituto de Ciências Biológicas e da Saúde, Universidade Federal Rural do Rio de Janeiro, Rio de Janeiro 23897-000, Brazil; [bb]Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, Rio de Janeiro 21040-360, Brazil; [cc]Núcleo de Pesquisas Ecológicas de Macaé, Universidade Federal do Rio de Janeiro, Rio de Janeiro 27910-970, Brazil; [dd]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556; [ee]Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Minas Gerais 30190-002, Brazil; [ff]Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Barcelona 08003, Catalonia, Spain; [gg]Universitat Pompeu Fabra, Barcelona 08003, Catalonia, Spain; [hh]Laboratoire d'Evolution, Génome et Spéciation, Centre National de la Recherche Scientifique UPR9034/UR 072, Institut de Recherche pour le Développement, 91198 Gif sur Yvette, France; [ii]Université Paris-Sud, 91400 Orsay, France; [jj]French National Institute For Agricultural Research, Institute of Ecology and Environmental Sciences of Paris, 78000 Versailles, France; [kk]Division of

Parasitic Diseases and Malaria, Entomology Branch, Centers for Disease Control and Prevention, Atlanta, GA 30329; ^{ll}Institut de Recherche sur la Biologie de l'Insecte, UMR7261, Centre National de la Recherche Scientifique, Université François Rabelais, 37200 Tours, France; ^{mm}Departamento de Biotecnologia Farmacêutica, Faculdade de Farmácia, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-902, Brazil; ⁿⁿDepartment of Physiology, Biophysics and Neuroscience, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Zacatenco, Mexico City 03760, Mexico; ^{oo}Departamento de Fisiologia e Biofísica, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Minas Gerais 31270-901, Brazil; ^{pp}Department of Biological Sciences, Florida International University, Miami, FL 11200; ^{qq}Núcleo de Doenças Infecciosas, Universidade Federal do Espirito Santo, Espirito Santo 29043-900, Brazil; ^{rr}Department of Entomology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; ^{ss}Instituto Federal de Educação Ciência e Tecnologia do Rio de Janeiro, Rio de Janeiro 20270-021, Brazil; ^{tt}Department of Physiology, School of Medicine–Center for Resesarch in Molecular Medicine and Chronic Diseases, Instituto de Investigaciones Sanitarias, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain; ^{uu}Department of Biochemistry, Virginia Polytechnic Institute, Blacksburg, VA 24061; ^{vv}Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, United Kingdom; and ^{ww}Department of Biological Sciences, University of Manitoba, Winnipeg, MB, Canada R3T 2N2

1. Wigglesworth VB (1931) The physiology of excretion in a blood-sucking insect, *Rhodnius prolixus* (Hemiptera, Reduviidae) I. Composition of the urine. *J Exp Biol* 8(4):411–427.
2. WHO (2015) *Chagas Disease Fact Sheet No 340* (WHO, Geneva).
3. Panzera F, et al. (2010) Cytogenetics and genome evolution in the subfamily Triatominae (Hemiptera, Reduviidae). *Cytogenet Genome Res* 128(1-3):77–87.
4. DeSalle R, Gregory TR, Johnston JS (2005) Preparation of samples for comparative studies of arthropod chromosomes: Visualization, in situ hybridization, and genome size estimation. *Methods Enzymol* 395:460–488.
5. Anonymous; International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol* 8(2):e1000313.
6. Fischer S, et al. (2011) Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr Protoc Bioinformatics* Chapter 6, Unit 6.12:11–19.
7. Neafsey DE, et al. (2015) Mosquito genomics. Highly evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes. *Science* 347(6217):1258522.
8. Fernandez-Medina RD, Granzotto A, Ribeiro JM, Carareto CM (2015) Transposition burst of mariner-like elements in the sequenced genome of *Rhodnius prolixus*. *Insect Biochem Mol Biol*, 10.1016/j.ibmb.2015.09.003.
9. Brelsfoard C, et al. (2014) Presence of extensive *Wolbachia* symbiont insertions discovered in the genome of its host *Glossina morsitans morsitans*. *PLoS Negl Trop Dis* 8(4):e2728.
10. Ribeiro JMC, et al. (2014) An insight into the transcriptome of the digestive tract of the bloodsucking bug, *Rhodnius prolixus*. *PLoS Negl Trop Dis* 8(1):e2594.
11. Fire A, et al. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391(6669):806–811.
12. Bartel DP (2004) MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* 116(2):281–297.
13. Lim LP, et al. (2003) The microRNAs of *Caenorhabditis elegans*. *Genes Dev* 17(8):991–1008.
14. An L, Liu Y, Wu A, Guan Y (2013) microRNA-124 inhibits migration and invasion by down-regulating ROCK1 in glioma. *PLoS One* 8(7):e69478.
15. Hoss AG, et al. (2014) MicroRNAs located in the Hox gene clusters are implicated in Huntington's disease pathogenesis. *PLoS Genet* 10(2):e1004188.
16. Gerardo NM, et al. (2010) Immunity and other defenses in pea aphids, *Acyrthosiphon pisum*. *Genome Biol* 11(2):R21.
17. Ursic-Bedoya R, Buchhop J, Lowenberger C (2009) Cloning and characterization of Dorsal homologues in the hemipteran *Rhodnius prolixus*. *Insect Mol Biol* 18(5):681–689.
18. Eichler S, Schaub GA (2002) Development of symbionts in triatomine bugs and the effects of infections with trypanosomatids. *Exp Parasitol* 100(1):17–27.
19. Castro DP, et al. (2012) *Trypanosoma cruzi* immune response modulation decreases microbiota in *Rhodnius prolixus* gut and is crucial for parasite survival and development. *PLoS One* 7(5):e36591.
20. Dias FA, et al. (2015) Identification of a selenium-dependent glutathione peroxidase in the blood-sucking insect *Rhodnius prolixus*. *Insect Biochem Mol Biol*, 10.1016/j.ibmb.2015.08.007.
21. Schama R, et al. (2015) *Rhodnius prolixus* supergene families of enzymes potentially associated with insecticide resistance. *Insect Biochem Mol Biol*, 10.1016/j.ibmb.2015.06.005.
22. Capriotti N, Mougabure-Cueto G, Rivera-Pomar R, Ons S (2014) L925I mutation in the Para-type sodium channel is associated with pyrethroid resistance in *Triatoma infestans* from the Gran Chaco region. *PLoS Negl Trop Dis* 8(1):e2659.
23. Germano MD, et al. (2010) New findings of insecticide resistance in *Triatoma infestans* (Heteroptera: Reduviidae) from the Gran Chaco. *J Med Entomol* 47(6):1077–1081.
24. Shiu S-H, Li W-H (2004) Origins, lineage-specific expansions, and multiple losses of tyrosine kinases in eukaryotes. *Mol Biol Evol* 21(5):828–840.
25. Blume-Jensen P, Hunter T (2001) Oncogenic kinase signalling. *Nature* 411(6835):355–365.
26. Lynch JA, El-Sherif E, Brown SJ (2012) Comparisons of the embryonic development of *Drosophila*, *Nasonia*, and *Tribolium*. *Wiley Interdiscip Rev Dev Biol* 1(1):16–39.
27. Ewen-Campen B, Srouji JR, Schwager EE, Extavour CG (2012) *Oskar* predates the evolution of germ plasm in insects. *Curr Biol* 22(23):2278–2283.
28. Lavore A, Esponda-Behrens N, Pagola L, Rivera-Pomar R (2014) The gap gene *Krüppel* of *Rhodnius prolixus* is required for segmentation and for repression of the homeotic gene *sex comb-reduced*. *Dev Biol* 387(1):121–129.
29. Lavore A, Pagola L, Esponda-Behrens N, Rivera-Pomar R (2012) The gap gene giant of *Rhodnius prolixus* is maternally expressed and required for proper head and abdomen formation. *Dev Biol* 361(1):147–155.
30. Berni M, et al. (2014) Toll signals regulate dorsal-ventral patterning and anterior-posterior placement of the embryo in the hemipteran *Rhodnius prolixus*. *Evodevo* 5(1):38.
31. Özüak O, Buchta T, Roth S, Lynch JA (2014) Dorsoventral polarity of the *Nasonia* embryo primarily relies on a BMP gradient formed without input from Toll. *Curr Biol* 24(20):2393–2398.
32. Wilson MJ, Kenny NJ, Dearden PK (2014) Components of the dorsal-ventral pathway also contribute to anterior-posterior patterning in honeybee embryos (*Apis mellifera*). *Evodevo* 5(1):11.
33. Wigglesworth VB (1954) Growth and regeneration in the tracheal system of an insect, *Rhodnius prolixus* (Hemiptera). *Q J Microsc Sci* 95:115–137.
34. Lavore A, et al. (2015) Comparative analysis of zygotic developmental genes in *Rhodnius prolixus* genome shows conserved features on the tracheal developmental pathway. *Insect Biochem Mol Biol* 64:32–43.
35. Wigglesworth VB (1934) The physiology of ecdysis in *Rhodnius prolixus* (Hemiptera). II. Factors controlling molting and metamorphosis. *Q J Microsc Sci* 2(77):191–222.
36. Noriega FG (2014) Juvenile hormone biosynthesis in insects: What is new, what do we know, and what questions remain? *International Scholarly Research Notices*, 10.1155/2014/967361.
37. Vidal NM, Grazziotin AL, Iyer LM, Aravind L, Venancio TM (2015) Transcription factors, chromatin proteins and the diversification of Hemiptera. *Insect Biochem Mol Biol*, 10.1016/j.ibmb.2015.07.001.
38. Carey AF, Carlson JR (2011) Insect olfaction from model systems to disease control. *Proc Natl Acad Sci USA* 108(32):12987–12995.
39. Botto-Mahan C, Cattan PE, Canals M (2002) Field tests of carbon dioxide and conspecifics as baits for *Mepraia spinolai*, vector of Chagas disease. *Acta Trop* 82(3):377–380.
40. Robertson HM, Kent LB (2009) Evolution of the gene lineage encoding the carbon dioxide receptor in insects. *J Insect Sci* 9:19.
41. Ribeiro JMC, Francischetti IMB (2003) Role of arthropod saliva in blood feeding: Sialome and post-sialome perspectives. *Annu Rev Entomol* 48:73–88.
42. Paluzzi JP, Yeung C, O'Donnell MJ (2013) Investigations of the signaling cascade involved in diuretic hormone stimulation of Malpighian tubule fluid secretion in *Rhodnius prolixus*. *J Insect Physiol* 59(12):1179–1185.
43. Dansa-Petretski M, Ribeiro JMC, Atella GC, Masuda H, Oliveira PL (1995) Antioxidant role of *Rhodnius prolixus* heme-binding protein. Protection against heme-induced lipid peroxidation. *J Biol Chem* 270(18):10893–10896.
44. Oliveira PL, et al. (1995) A heme-binding protein from hemolymph and oocytes of the blood-sucking insect, *Rhodnius prolixus*. Isolation and characterization. *J Biol Chem* 270(18):10897–10901.
45. Guerenstein PG, Lazzari CR (2009) Host-seeking: How triatomines acquire and make use of information to find blood. *Acta Trop* 110(2-3):148–158.
46. Ons S, Richter F, Urlaub H, Pomar RR (2009) The neuropeptidome of *Rhodnius prolixus* brain. *Proteomics* 9(3):788–792.
47. Guedes RLM, et al. (2011) Amino acids biosynthesis and nitrogen assimilation pathways: A great genomic deletion during eukaryotes evolution. *BMC Genomics* 12(Suppl 4):S2.
48. Scaraffia PY, et al. (2008) Discovery of an alternate metabolic pathway for urea synthesis in adult *Aedes aegypti* mosquitoes. *Proc Natl Acad Sci USA* 105(2):518–523.
49. Miller JR, et al. (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24(24):2818–2824.
50. Ribeiro JM, Topalis P, Louis C (2004) AnoXcel: An *Anopheles gambiae* protein database. *Insect Mol Biol* 13(5):449–457.
51. Jones P, et al. (2014) InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.
52. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
53. Tubío JMC, Naveira H, Costas J (2005) Structural and evolutionary analyses of the Ty3/gypsy group of LTR retrotransposons in the genome of *Anopheles gambiae*. *Mol Biol Evol* 22(1):29–39.
54. Tubío JMC, et al. (2011) Evolutionary dynamics of the Ty3/gypsy LTR retrotransposons in the genome of *Anopheles gambiae*. *PLoS One* 6(1):e16328.
55. Tu Z (2001) Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. *Proc Natl Acad Sci USA* 98(4):1699–1704.
56. Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21(Suppl 1):i351–i358.
57. Kent WJ (2002) BLAT—The BLAST-like alignment tool. *Genome Res* 12(4):656–664.
58. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 16(6):276–277.
59. John B, et al. (2004) Human microRNA targets. *PLoS Biol* 2(11):e363.
60. Mariotti M, Guigó R (2010) Selenoprofiles: Profile-based scanning of eukaryotic genome sequences for selenoprotein genes. *Bioinformatics* 26(21):2656–2663.
61. Mariotti M, Lobanov AV, Guigo R, Gladyshev VN (2013) SECISearch3 and Seblastian: New tools for prediction of SECIS elements and selenoproteins. *Nucleic Acids Res* 41(15):e149.
62. Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25(5):955–964.
63. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302(1):205–217.
64. Huerta-Cepas J, et al. (2011) PhylomeDB v3.0: An expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic Acids Res* 39(Database issue, D1):D556–D560.
65. Ulitsky I, et al. (2010) Expander: From expression microarrays to networks and functions. *Nat Protoc* 5(2):303–322.

GENETICS