BMC Genomics

Open Access

The genome of the water strider *Gerris buenoi* reveals expansions of gene repertoires associated with adaptations to life on the water

David Armisén[1*], Rajendhran Rajakumar[2], Markus Friedrich[3], Joshua B. Benoit[4], Hugh M. Robertson[5], Kristen A. Panfilio[6,7], Seung-Joon Ahn[8,9], Monica F. Poelchau[10], Hsu Chao[11], Huyen Dinh[11], Harsha Vardhan Doddapaneni[11], Shannon Dugan[11], Richard A. Gibbs[11], Daniel S. T. Hughes[11], Yi Han[11], Sandra L. Lee[11], Shwetha C. Murali[12], Donna M. Muzny[11], Jiaxin Qu[11], Kim C. Worley[11], Monica Munoz-Torres[13], Ehab Abouheif[14], François Bonneton[1], Travis Chen[14], Li-Mei Chiang[10], Christopher P. Childers[10], Andrew G. Cridge[15], Antonin J. J. Crumière[1], Amelie Decaras[1], Elise M. Didion[4], Elizabeth J. Duncan[15,16], Elena N. Elpidina[17], Marie-Julie Favé[14], Cédric Finet[1], Chris G. C. Jacobs[18,19], Alys M. Cheatle Jarvela[20], Emily C. Jennings[4], Jeffery W. Jones[3], Maryna P. Lesoway[14,21,22], Mackenzie R. Lovegrove[15], Alexander Martynov[22], Brenda Oppert[23], Angelica Lillico-Ouachour[14], Arjuna Rajakumar[14], Peter Nagui Refki[1,24], Andrew J. Rosendale[4], Maria Emilia Santos[1], William Toubiana[1], Maurijn van der Zee[18], Iris M. Vargas Jentzsch[6], Aidamalia Vargas Lowman[1], Severine Viala[1], Stephen Richards[11*] and Abderrahman Khila[1*]

Abstract

Background: Having conquered water surfaces worldwide, the semi-aquatic bugs occupy ponds, streams, lakes, mangroves, and even open oceans. The diversity of this group has inspired a range of scientific studies from ecology and evolution to developmental genetics and hydrodynamics of fluid locomotion. However, the lack of a representative water strider genome hinders our ability to more thoroughly investigate the molecular mechanisms underlying the processes of adaptation and diversification within this group.

(Continued on next page)

* Correspondence: davidarmisen@gmail.com; stephenr@bcm.edu; abderrahman.khila@ens-lyon.fr
[1]Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5242, Ecole Normale Supérieure de Lyon 46, allée d'Italie, 69364 Lyon Cedex 07, France
[11]Human Genome Sequencing Center, Department of Human and Molecular Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
Full list of author information is available at the end of the article

(Continued from previous page)

Results: Here we report the sequencing and manual annotation of the *Gerris buenoi* (*G. buenoi*) genome; the first water strider genome to be sequenced thus far. The size of the *G. buenoi* genome is approximately 1,000 Mb, and this sequencing effort has recovered 20,949 predicted protein-coding genes. Manual annotation uncovered a number of local (tandem and proximal) gene duplications and expansions of gene families known for their importance in a variety of processes associated with morphological and physiological adaptations to a water surface lifestyle. These expansions may affect key processes associated with growth, vision, desiccation resistance, detoxification, olfaction and epigenetic regulation. Strikingly, the *G. buenoi* genome contains three insulin receptors, suggesting key changes in the rewiring and function of the insulin pathway. Other genomic changes affecting with opsin genes may be associated with wavelength sensitivity shifts in opsins, which is likely to be key in facilitating specific adaptations in vision for diverse water habitats.

Conclusions: Our findings suggest that local gene duplications might have played an important role during the evolution of water striders. Along with these findings, the sequencing of the *G. buenoi* genome now provides us the opportunity to pursue exciting research opportunities to further understand the genomic underpinnings of traits associated with the extreme body plan and life history of water striders.

Keywords: Water striders, Genome sequence, Water surface locomotion, Evolution, Adaptation,
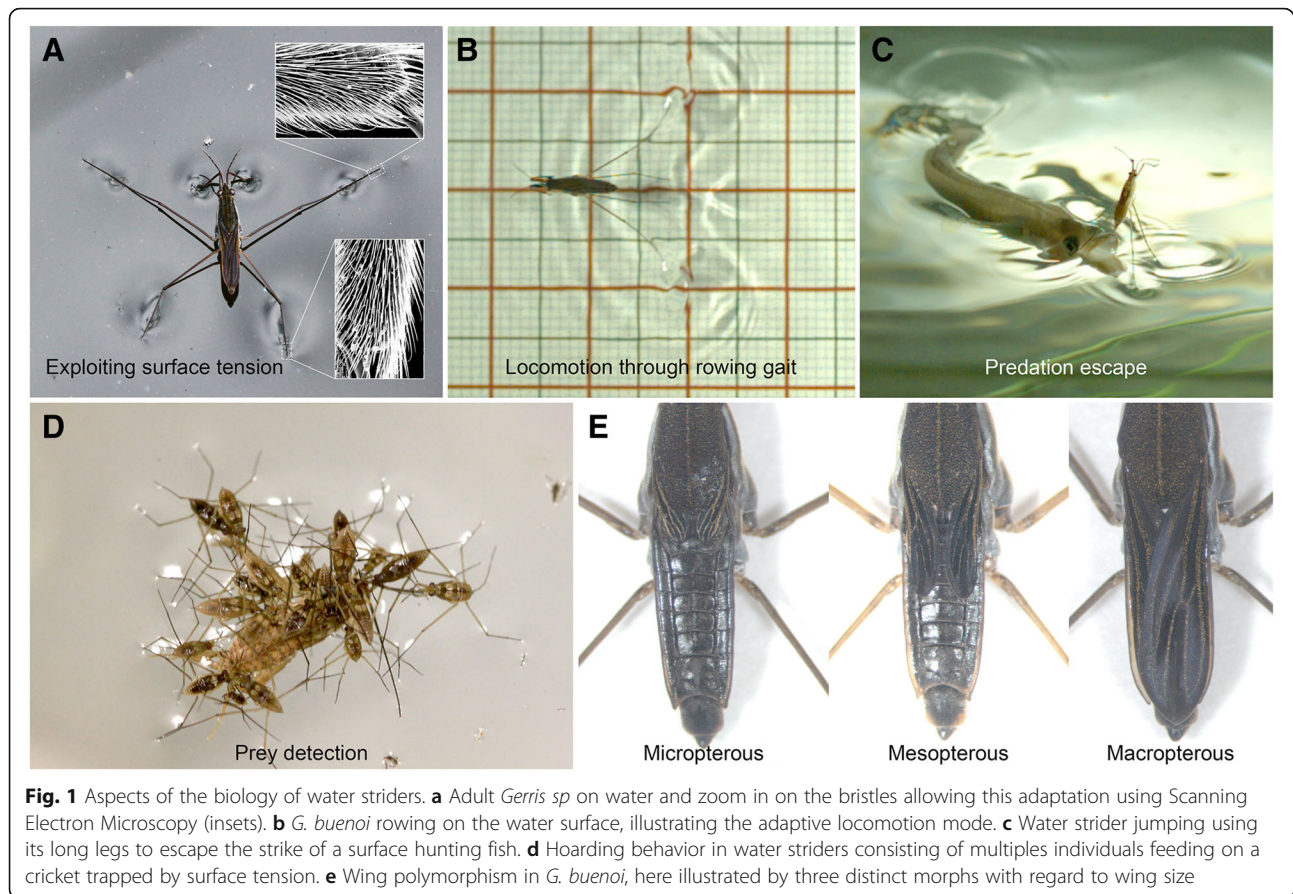
Background

The semi-aquatic bugs (Gerromorpha) are a monophyletic group of predatory heteropteran insects characterized by their ability to live at the water-air interface [1–4]. Over 200 million years ago, the ancestor of the Gerromorpha transitioned from terrestrial habitats to the water surface, leading to a radiation that has generated over 2,000 species classified into eight families [1]. Phylogenetic reconstructions suggest that the ancestral habitat of the Gerromorpha was either humid and terrestrial or marginally aquatic [1, 5, 6]. Water striders subsequently became true water surface dwellers and colonized a diverse array of niches, including streams, lakes, ponds, marshes, and the open ocean [1, 7, 8]. The invasion of these new habitats provided access to resources previously underutilized by insects and made the Gerromorpha the dominant group of insects at water surfaces [1]. This novel specialized life style makes the Gerromorpha an exquisite model system to study how new ecological opportunities can drive adaptation and species diversification [2, 9–11].

This shift in habitat exposed the Gerromorpha to new selective pressures compared to their terrestrial ancestors. The Gerromorpha face two primary challenges unique among insects: how to remain afloat and how to generate efficient thrust on the fluid substrate for locomotion [2, 3, 12]. Due to their specific arrangement and density, the bristles covering the legs of water striders are adapted to keep them afloat by acting as non-wetting structures, which exploit water surface tension by trapping air between the leg and water (Fig. 1a) [2, 3, 12, 13]. Furthermore, locomotion is made possible through evolutionary changes in the morphology and biomechanical adaptions associated with patterns of leg movement (Fig. 1b) [2, 3, 12, 13]. Two distinct modes of locomotion are employed by distinct species: an ancestral mode using a tripod gait with alternating leg movements, and a derived mode using a rowing gait through a simultaneous sculling motion of the pair of middle legs (Fig. 1b) [2, 12]. The rowing mode is characteristic of the Gerridae and some Veliidae and is associated with a derived body plan where the middle legs are the longest (Fig. 1a–b) [2, 12]. The evolutionary trajectory of this group is also thought to have been shaped by the novel predator-prey interactions (Fig. 1c and d) associated with their water surface life history. Following the invasion of water surfaces, other adaptations have emerged, including: (1) the adaption of their visual system to the surface-underwater environment; (2) the evolution of wing polymorphisms associated with dispersal strategies and habitat quality (Fig. 1e) [14], and changes in cuticle composition that optimized water exchange and homeostasis associated with living on water.

While we are starting to uncover the developmental genetic and evolutionary processes underlying the adaptation of water striders to the requirements of water surface locomotion, predator-prey, and sexual interactions [2, 15–19], studies of these mechanisms at the genomic level are hampered by the lack of a representative genome. Here we report the genome of the water strider *G. buenoi*, the first sequenced member of the infraorder Gerromorpha. *G. buenoi* is part of the family Gerridae, and has been previously used as a model to study sexual selection and developmental genetics [15, 20–22]. Moreover, *G. buenoi* can easily breed in laboratory conditions and is closely related to several other *Gerris* species used as models for the study of water-walking hydrodynamics, salinity tolerance, and sexual conflict. With a particular focus on manual annotation and analyses of processes involved in phenotypic adaptations to life on water, our analysis of the *G. buenoi* genome suggests that the genomic basis of water surface invasion might be, at least in part, underpinned by clustered gene family expansions and tandem gene duplications.
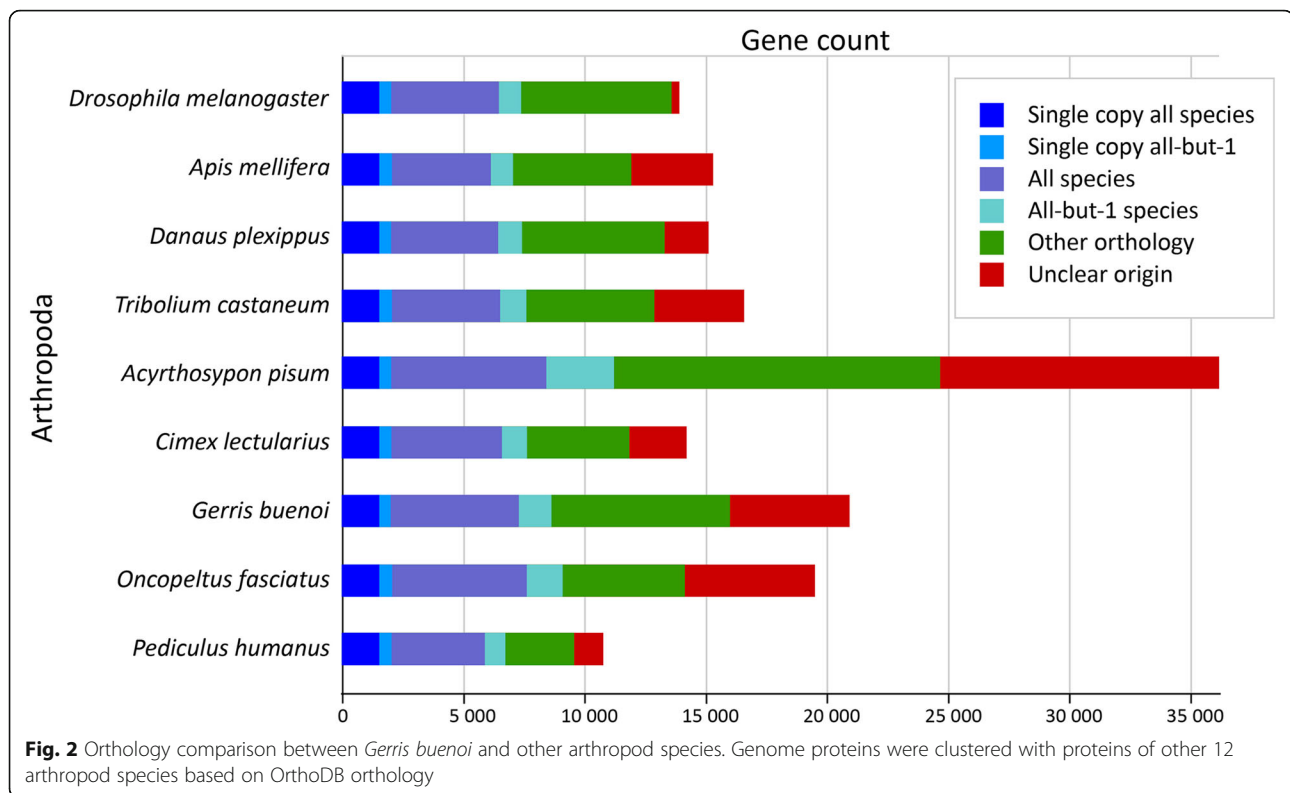
Fig. 1 Aspects of the biology of water striders. **a** Adult *Gerris sp* on water and zoom in on the bristles allowing this adaptation using Scanning Electron Microscopy (insets). **b** *G. buenoi* rowing on the water surface, illustrating the adaptive locomotion mode. **c** Water strider jumping using its long legs to escape the strike of a surface hunting fish. **d** Hoarding behavior in water striders consisting of multiples individuals feeding on a cricket trapped by surface tension. **e** Wing polymorphism in *G. buenoi*, here illustrated by three distinct morphs with regard to wing size

Results and discussion

General features of the *G. buenoi* genome

The draft assembly of *G. buenoi* genome comprises 1,000,194,699 bp (GC content: 32.46%) in 20,268 scaffolds and 304,909 contigs (N50 length is 344,118 and 3812 bp, respectively). The assembly recovers ~ 87% of the genome size estimated at ~ 1.15 GB based on k-mer analysis. The *G.buenoi* genome is organized into 18 autosomal chromosomes with a XX/X0 sex determination system [23]. The MAKER automatic annotation pipeline predicted 20,949 protein-coding genes, which is greater than the 16,398 isogroups previously annotated in the transcriptome of the closely related species *Limnoporus dissortis* (PRJNA289202) [18, 24], as well as the 14,220 genes in the bed bug *Cimex lectularius* genome [25] and the 19,616 genes in the genome of the milkweed bug *Oncopeltus fasciatus* [26]. The final *G. buenoi* official gene set (OGS) 1.0 includes 1,277 manually annotated genes, including 1,378 mRNAs and 15 pseudogenes, representing development, growth, immunity, cuticle formation as well as olfaction and detoxification pathways genes, amongst others (see Additional file 1). Using OrthoDB [27, 28], we found that ~ 75% of *G. buenoi* genes have at least one orthologue in other arthropod species (Fig. 2). We then used benchmarking sets of universal single-copy orthologs (BUSCOs) [29, 30] to assess the completeness of the assembly. A total of 85.4% of BUSCOs were found complete and 12.3% were fragmented.

In addition to BUSCOs, we used Hox and Iroquois Complex (Iro-C) gene clusters as indicators of draft genome quality and as an opportunity to assess synteny among species. The Hox cluster is conserved across the Bilateria [31], and the Iro-C is found throughout the Insecta [25, 32]. In *G. buenoi*, we were able to find and annotate gene models for all ten Hox genes (Additional file 1: Table S3). While linkage of the highly conserved central class genes *Sex combs reduced*, *fushi tarazu*, and *Antennapedia* occurred in the expected order and with the expected transcriptional orientation, the linked models of *proboscipedia* and *zerknüllt* (*zen*) occur in opposite transcriptional orientations (head-to-head, rather than both 3′ to 5′). Inversion of the divergent *zen* locus is not new in the Insecta [33], but was not observed in the hemipteran *C. lectularius*, in which the complete Hox cluster was fully assembled [25]. Future genomic data will help to determine whether such a microinversion within the Hox cluster is conserved within the

Fig. 2 Orthology comparison between *Gerris buenoi* and other arthropod species. Genome proteins were clustered with proteins of other 12 arthropod species based on OrthoDB orthology

hemipteran family Gerridae. Assembly limitations are also evident in our Hox cluster analysis. For example, the complete gene model for *labial* is present but split across scaffolds, while only partial gene models could be created for *Ultrabithorax* and *Abdominal-B*. Furthermore, while there are clear single-copy orthologues of members of the small Iroquois complex, *iroquois* and *mirror*, they are not linked in the current assembly (Additional file 1: Table S3). However, both genes are located near the ends of their scaffolds, and direct concatenation of the scaffolds (5′-Scaffold451–3′, 3′-Scaffold2206-5′) would correctly reconstruct this cluster: (1) with both genes in the 5′-to-3′ transcriptional orientation along the (+) DNA strand, (2) with no predicted intervening genes within the cluster, and (3) with a total cluster size of 308 Kb, which is fairly comparable with that of other recently sequenced hemipterans in which the Iro-C cluster linkage was recovered (391 Kb in *C. lectularius* [25] and 403 Kb in *O. fasciatus* [26]). Lastly, building on the automated BUSCO assessment for presence of expected genes, we examined genes associated with autophagy processes, which are highly conserved among insects, and all required genes are present within the genome (Additional file 2). Therefore, along with the Hox and Iroquois Complex (Iro-C) gene cluster analyses, the presence of a complete set of required autophagy genes suggest good gene representation and supports further analysis.

Adaptation to water surface locomotion

One of the most important morphological adaptations that enabled water striders to conquer water surfaces is the change in shape, density, and arrangement of the bristles that span the contact surface between their legs and the fluid substrate. These bristles, by trapping air, act as non-wetting structures, forming a cushion between the legs and the water surface (Fig. 1a) [2, 3, 12, 13]. QTL studies in flies uncovered dozens of candidate genes and regions linked to variation in bristle density and morphology [34]. In the *G. buenoi* genome we were able to annotate 90 out of 120 genes known to be involved in bristle development [34, 35] (Additional file 1: Table S4). Among these, we found a single duplication, the gene *Beadex* (*Bx*). A similar duplication found in *C. lectularius* and *H. halys* suggest that the *Bx* duplication may have predated the separation of these lineages and the radiation of Gerromorpha, although a broader phylogenetic sampling is needed to strengthen this conclusion. In *Drosophila*, *Bx* is involved in neural development by controlling the activation of *achaete-scute* complex genes [36] and mutants of *Bx* have extra sensory organs [36]. Based on this, it is reasonable to speculate that duplication of *Beadex* might have been exploited by water striders and subsequently linked to changes in bristle pattern and density. This possibility opens up new research avenues to further understand the adaptation of water striders to living on the water surface.

A new duplication in the Insulin Receptor gene family in the Gerromorpha

The insulin signaling pathway coordinates hormonal and nutritional signals in animals [37–39]. This facilitates the complex regulation of several fundamental molecular and cellular processes including transcription, translation, cell stress, autophagy, and physiological states, such as aging and starvation [39–42]. The action of insulin signaling is mediated through the Insulin Receptor (InR), a transmembrane receptor of the tyrosine kinase class [43]. While vertebrates possess one copy of the InR [44], arthropods generally possess either one or two copies, although the highly duplicated *Daphnia pulex* genome [45] contains four copies [46]. Interestingly, the *G. buenoi* genome contains three distinct InR copies. Further sequence examination using in-house transcriptome databases of multiple Gerromorpha species confirmed that this additional copy is common to all of them, indicating that it was present in the common ancestor of the group (Fig. 3). In addition, cloning of the three InR sequences using PCR indicates that these sequences originate from three distinct coding genes that are actively transcribed in this group of insects. Comparative protein sequence analysis revealed that the three InR copies possess all the characteristic domains found in InR in both vertebrates and invertebrates (Fig. 3a). Together, these results validate the presence of three InR copies in Gerromorpha, an exceptional situation amongst Arthropoda.

While this manuscript as under evaluation, an independent study reported the presence of a third InR gene in Blattodea [47]. To determine: (1) the origin of the three InR copies in the *G. buenoi* genome; and (2) whether the third copy in Gerromorpha and Blattodea share a common ancestor, we performed a phylogenetic reconstruction that included the sequences of eight Gerromorpha (three InR copies), four Blattodea (three InR copies), *Daphnia* (four copies) and an additional sample of 126 Arthropoda, all of which possess either one or two InR copies (see Additional files 3 and 4). The four InR duplicates of *Daphnia* were all lineage-specific and together formed a sister group to those found in insects. Within insects, this analysis clustered two InR copies into distinct InR1 and InR2 clusters (Fig. 3b). Furthermore, gerromorphan InR1 and InR2 copies clustered with bed bug and milkweed bug InR1 and InR2, respectively, while the Gerromorpha-restricted copy clustered alone (Fig. 3b; Additional file 1: Figure S1). These data suggest that the new InR copy, which we designated InR1-like, most likely originated from the InR1 gene in the common ancestor of the Gerromorpha. In contrast, the third InR copy in Blattodea clustered with InR2, suggesting an independent origin of novel InR copies in Gerromorpha, which we therefore would suggest be

designated InR2-like. A closer examination of the organization of the genomic locus of the InR1-like gene in *G. buenoi* revealed that this copy is intronless. This observation, together with the phylogenetic reconstruction, suggests that InR1-like is a retrocopy of InR1 that may have originated through RNA-based duplication [48]. In addition, our analysis suggests two independent losses of InR2. InR2 is lost among the parasitoid wasps yet retained in other wasps, and InR2 is also lost in the common ancestor of Diptera and Lepidoptera. Taken together, our current phylogenetic reconstruction demonstrates that: (1) InR was duplicated at the base of insects, generating InR1 and InR2; (2) InR1 was subsequently duplicated within the Gerromorpha, while InR2 was duplicated at the common ancestor of Blattodea; (3) InR2 was independently lost in the common ancestor of Lepidoptera and Diptera as well as among the parasitoid wasps, while other wasps have retained it.

In insects, the insulin signaling pathway has been implicated in the developmental regulation of complex nutrient-dependent growth phenotypes such as beetle horns and wing polyphenisms in plant hoppers, as well as morphological caste differentiation in social termites and bees [49–52]. In the particular case of wing polymorphism in *G. buenoi* [1, 14, 52], our analysis found no DNA methylation signature, as previously found in wing polyphenic ants and aphids [53–57], but rather an increased number of histone clusters and a unique duplication of the histone methyltransferase *grappa* (see Additional file 1: Supplementary Data). Taken together, it will be of interest to test the functional significance of the new InR copy in relation to wing polyphenism, as well as more generally how it may be potentially involved in appendage plasticity, either independent of, or alongside, epigenetic processes. Moreover, a comparative functional approach between the novel InR genes in Gerromorpha and Blattodea will shed light on the role independent insulin receptor duplications have played in functional convergence and/or diversification.

A lineage-specific expansion and possible sensitivity shifts of long wavelength sensitive opsins

Visual ecology at the air-water interface and the exceptionally specialized visual system of water striders has drawn considerable interest [58, 59]. Consisting of over 900 ommatidia, the prominent compound eyes of water striders are involved in prey localization, mating partner pursuit, predator evasion and dispersal by flight [60–62]. Realization of the first three tasks is associated with dorsal-ventral differences in the photoreceptor organization of the eye [63, 64], and polarized light-sensitivity [65] (see Additional file 1: Supplementary Data). Each water strider ommatidium contains six outer and two inner Recent work has produced evidence of at least two types of ommatidia,
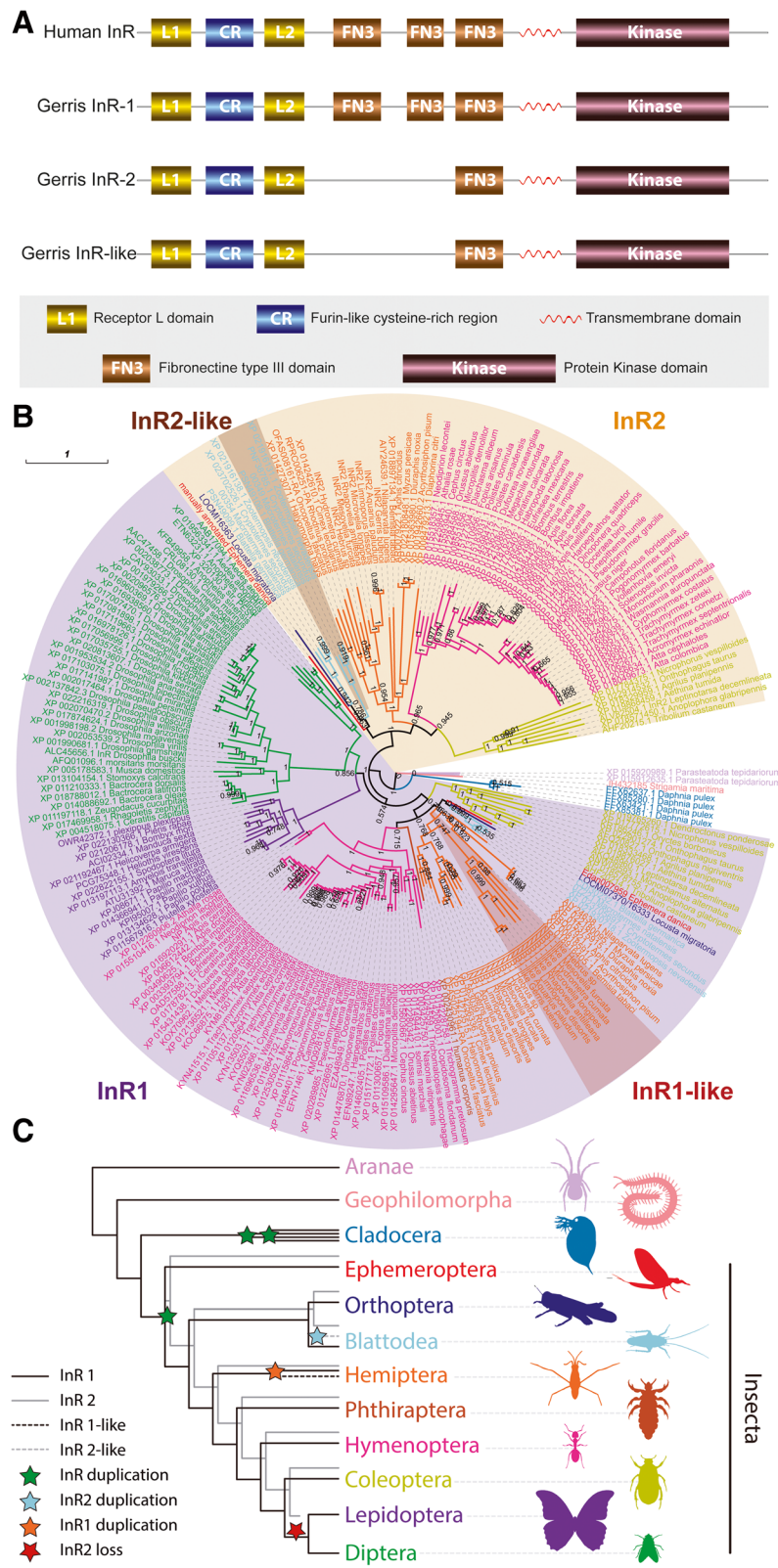
Fig. 3 Characterization of the three copies of the Insulin Receptor in *Gerris buenoi*. **a** Protein domain comparison between the three InRs of *G. buenoi* and the Human InR. **b** InR phylogenetic relationship amongst Insecta. Branch support numbers at branches. A non-circular version included in Additional file 1: Figure S1. **c** Simplified Arthropoda phylogeny based on [115] depicting InR duplications and loss events

with outer photoreceptors that are sensitive to either green (~ 530 nm) or blue (~ 470–490 nm) wavelengths [66], but the wavelength specificity of the two inner photoreceptors cells is still unknown. At the molecular level, the wavelength specificity of photoreceptor subtypes is mostly determined by the expression of paralogous opsins (light sensitive G-protein coupled receptor proteins), which differ in their wavelength absorption maxima. Interestingly, our genomic analysis of opsin diversity in *G. buenoi* uncovered 8 opsin homologs. Among these, we uncovered three arthropod non-retinal opsins (c-opsin, Arthropsin and Rh7 opsin) (see Additional file 1: Supplementary Data) in addition to five retinal opsins (Fig. 4a; Additional file 1: Figure S2). One of these five retinal opsins was identified as a member of the UV-sensitive opsin subfamily and the other four were identified as tandem, clustered members of the long wavelength sensitive (LWS) opsin subfamily (Fig. 4b).

Surprisingly, both genomic and transcriptomic searches in *G. buenoi* and other water strider species failed to detect sequence evidence of homologs of the otherwise deeply conserved blue-sensitive opsin subfamily (Fig. 4b; Additional file 1: Table S5) [67]. Although the apparent lack of blue opsin in *G. buenoi* was unexpected given the presence of blue sensitive photoreceptors [66], it was consistent with the lack of blue opsin sequence evidence in the available genomes and transcriptomes of other heteropteran species including *Halyomorpha halys*, *Oncopeltus fasciatus*, *Cimex lectularius*, and *Rhodnius prolixus*. Blue opsin, however, is present in other hemipteran clades, including Cicadomorpha (*Nephotettix cincticeps*) and Sternorrhyncha (*Pachypsylla venusta*) (Fig. 4b). Based on the currently available sample of hemipteran species, these data suggest that the blue-sensitive opsin subfamily was lost early in the last common ancestor of the Heteroptera (Fig. 4b and Additional file 1: Table S5). This raises the question of which compensatory events explain the presence of blue sensitive photoreceptors in water striders.

Studies in butterflies and beetles produced evidence of blue sensitivity shifts in both UV- and LWS-opsin homologs following gene duplication [68–70]. In butterflies, molecular evolutionary studies have implicated amino acid residue differences at four protein sequence sites in sensitivity shifts from green to blue: Ile17Met, Ala64Ser, Asn70Ser, and Ser137Ala [68, 69] (Fig. 4c; Additional file 1: Figure S2 and Supplementary Data). Based on sequence information from physiologically characterized LWS opsins in other insect orders and the degree of amino acid residue conservation at these sites in a sample of 114 LWS opsin homologs from 54 species representing 12 insect orders (Additional file 1: Supplementary Data and Additional file 5), we could identify *G. buenoi* LWS opsin 3 as a high confidence candidate for a blue-shifted paralog, followed by *G. buenoi* LWS opsin

1 and 2. Moreover, the *G. buenoi* LWS opsin 4 paralog matches all of the butterfly green-sensitive amino acid residue states, thus favoring this paralog as green-sensitive (Fig. 4). These conclusions are further backed by the fact that water striders lack ocelli, which implies that all four paralogs are expressed in photoreceptors of the compound eye. Overall, it is most likely that the differential expression of the highly diverged *G. buenoi* LWS opsin paralogs accounts for the presence of both blue- and green-sensitive peripheral photoreceptors in water striders. Moreover, given that the outer blue photoreceptors have been specifically implicated in the detection of contrast differences in water striders [66], it is tempting to speculate that the deployment of blue-shifted LWS opsins is a convergent characteristic of a fast-tracking visual system, similar to visual systems in dipteran species that also feature open rhabdomeres, neural superposition, and polarized light-sensitivity.

Expansion of cuticle gene repertoires

Desiccation resistance is essential to the colonization of terrestrial habitats by arthropods [71]. However, contrary to most insects, the Gerromorpha spend their entire life cycle in contact with water and exhibit poor desiccation resistance [1]. Cuticle proteins and aquaporins are essential for desiccation resistance through regulation of water loss and rehydration [72–75]. Unexpectedly in the *G. buenoi* genome, most members of cuticular and aquaporin protein families are present in similar numbers compared to other hemipterans (Additional file 1: Table S6 and Figure S3; Additional files 6 and 7). We identified 155 putative cuticle proteins belonging to five cuticular families: CPR (identified by Rebers and Riddiford Consensus region), CPAP1 and CPAP3 (Cuticular Proteins of Low-Complexity with Alanine residues), CPF (identified by a conserved region of about 44 amino acids), and TWDL (Tweedle) [76, 77] (Additional file 1: Table S6). Interestingly, almost half of them are arranged in clusters, indicative of local duplication events (Additional file 1: Table S7). Moreover, while most insect species, including other hemipterans, have only three TWDL genes, we found that the TWDL family in *G. buenoi* has been expanded to ten genes (Additional file 1: Figure S4). This expansion of the TWDL family is similar to that observed in some Diptera that possess *Drosophila*-specific and mosquito-specific TWDL expansions [77, 78]. Mutations in the *Drosophila TwdlD* are known to alter body shape [78]. Given the high diversification in body sizes and shapes in association with various aquatic habitats in the Gerromorpha in general [1, 2] and the Gerridae in particular [79, 80], it is possible that the expansion of the TWDL gene family is linked to this diversification. Therefore, a functional analysis of TWDL genes and comparative analysis with other hemipterans will provide
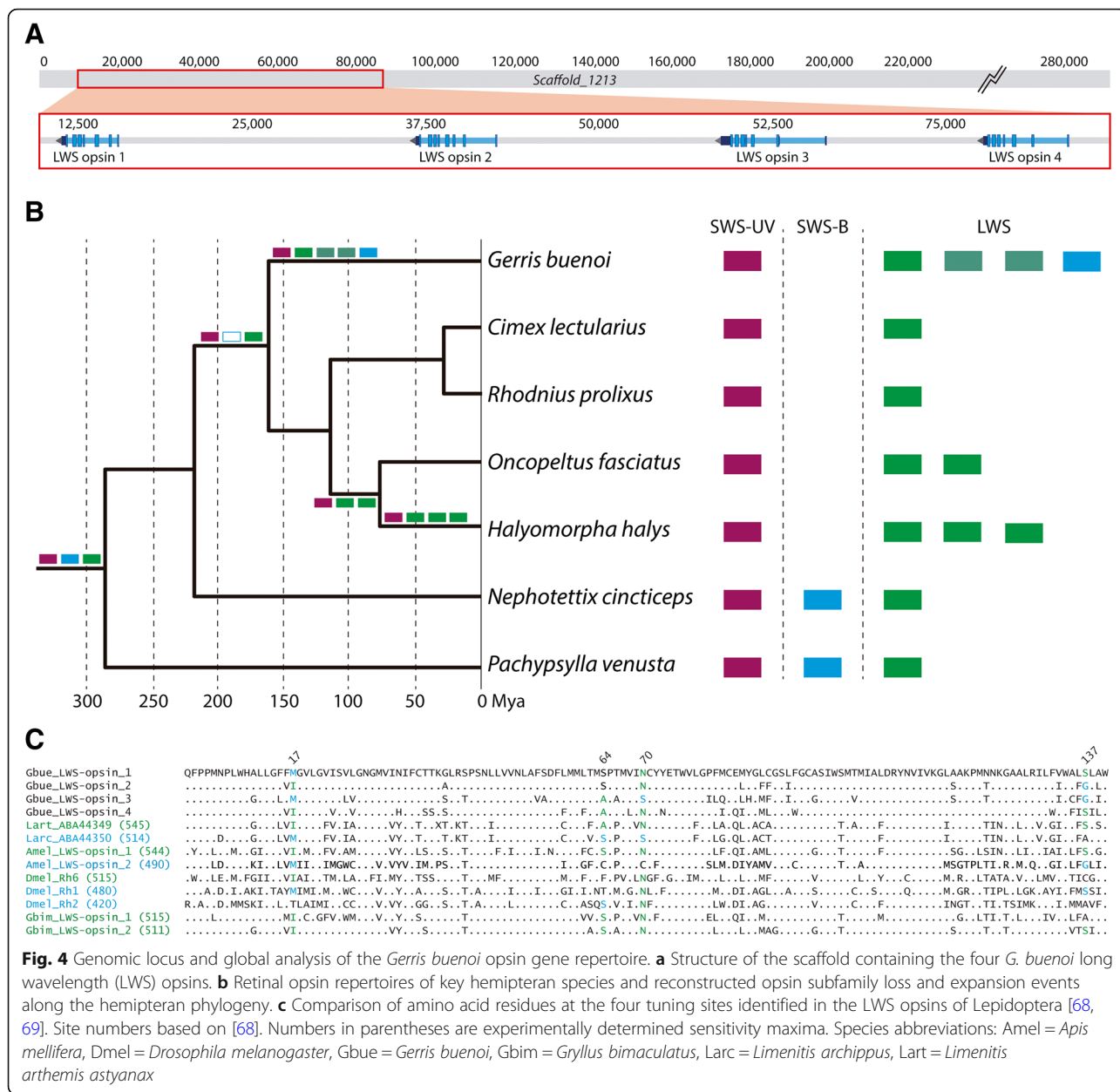
Fig. 4 Genomic locus and global analysis of the *Gerris buenoi* opsin gene repertoire. **a** Structure of the scaffold containing the four *G. buenoi* long wavelength (LWS) opsins. **b** Retinal opsin repertoires of key hemipteran species and reconstructed opsin subfamily loss and expansion events along the hemipteran phylogeny. **c** Comparison of amino acid residues at the four tuning sites identified in the LWS opsins of Lepidoptera [68, 69]. Site numbers based on [68]. Numbers in parentheses are experimentally determined sensitivity maxima. Species abbreviations: Amel = *Apis mellifera*, Dmel = *Drosophila melanogaster*, Gbue = *Gerris buenoi*, Gbim = *Gryllus bimaculatus*, Larc = *Limenitis archippus*, Lart = *Limenitis arthemis astyanax*

important insights into the evolutionary origins and functional significance of TWDL expansion in *G. buenoi*.

Prey detection in water surface environments

Unlike many closely related species that feed on plant sap or animal blood, *G. buenoi* feeds on various arthropods trapped by surface tension (Fig. 1d), thus making their diet highly variable. Chemoreceptors play a crucial role for prey detection and selection, in addition to vibrational and visual signals. We annotated the three families of chemoreceptors that mediate most of the sensitivity and specificity of chemoperception in insects: odorant receptors (ORs; Additional file 1: Figure S5A and

Additional file 8), gustatory receptors (GRs; Additional file 1: Figure S5B and Additional file 8) and ionotropic receptors (IRs; Additional file 1: Figure S5C and Additional file 8) (e.g. [81, 82]). Interestingly, we found an increase in the number of chemosensory genes in *G. buenoi* (Additional file 1: Table S8). First, the OR family is expanded, with a total of 155 OR proteins. This expansion is the result of lineage-specific "blooms" of particular gene subfamilies, including expansions of the 4, 8, 9, 13, 13, 16, 18, and 44 subfamilies (Additional file 1: Figure S5A and Supplementary Data). Second, the GR family is also fairly large (Additional file 1: Figure S5B), but the expansions here are primarily the result of extensive alternative splicing, such that 60 genes encode

135 GR proteins (Additional file 1: Table S8). These GRs include six genes encoding proteins related to the carbon dioxide receptors of flies, three related to sugar receptors, and one related to the fructose receptor (Additional file 1: Figure S5B). The remaining GRs include several highly divergent proteins, as well as four blooms, the largest of which comprises 80 proteins (Additional file 1: Figure S5B and Supplementary Data). By analogy with *D. melanogaster*, most of these proteins are likely to be "bitter" receptors, although some might be involved in perception of cuticular hydrocarbons and other molecules. Finally, the IR family is expanded to 45 proteins. In contrast with the OR/GR families, where the only orthologs across four heteropterans (*Rhodnius prolixus*, *Cimex lectularius*, *Oncopeltus fasciatus* and *Gerris buenoi*) and *Drosophila* are the single OrCo and fructose receptors, the IR family has single orthologs in each species. This is not restricted to only the highly conserved co-receptors (IR8a, 25a, and 76b) but also includes receptors implicated in sensing amino acids, temperature, and humidity (Ir21a, 40a, 68a, and 93a). As is common in other insects the amine-sensing IR41a lineage is expanded to four genes, while the acid-sensing IR75 lineage is highly expanded to 24 genes, and like the other heteropterans there are nine more highly divergent IRs (Additional file 1: Figure S5C and Supplementary Data).

We hypothesize that the high number of ORs may be linked to prey detection mediated by odor molecules at the air-water interface, although functional analysis will be needed to test this. As *G. buenoi* are faced with prey that have fallen on the water surface, and therefore individuals exhibit more of a scavenger strategy as compared to a hunter strategy, this expansion of ORs may enhance their ability to evaluate palatability. As toxic molecules are often perceived as bitter, the GR expansion might provide a complex bitter taste system to detect and even discriminate between molecules of different toxicities [83]. Finally, expansion of the IR family could be linked with prey detection as well as pheromone detection of water-soluble hydrophilic acids and amines, many of which are common chemosensory signals for aquatic species [84, 85].

Detoxification pathways

Water striders can be exposed to various toxic compounds found in the water, including those generated by pesticides, insecticides, and from other human activities as well as those found in their prey. Insect cytochrome P450 (CYP) proteins play a role in metabolic detoxification of xenobiotics including insecticides [86, 87]. They are also known to be responsible for the synthesis and degradation of endogenous molecules, such as ecdysteroids [88] and juvenile hormone [89]. The insect CYPs, one of the oldest and largest gene families in insects, underwent a high degree of diversification after multiple instances of

gene duplication, which may have enhanced a species' adaptive range [90]. In addition to CYP proteins we have also surveyed the presence of UDP-glycosyltransferase (UGT) genes in *G. buenoi*. UGTs are important for xenobiotic detoxification and the regulation of endobiotics in insects [91]. UGTs catalyse the conjugation of a range of small hydrophobic compounds to produce water-soluble glycosides that can be easily excreted in a number of insects [92, 93].

We annotated and analyzed a total of 103 CYP genes (Additional file 1: Table S9, Additional files 3 and 9) and 28 putative UGT genes, including several partial sequences due to genomic gaps (Additional file 1: Table S10). Ten more CYP fragments were found, but they were not included in this analysis due to their short lengths (<250 aa). This is the largest number of CYP genes among the hemipteran and other species' genomes in which CYPomes were annotated: *O. fasciatus* (58 CYPs), *R. prolixus* (88 CYPs) and *N. lugens* (68 CYPs) [26, 94, 95], *D. melanogaster* (85), *A. mellifera* (45), and *B. mori* (86) (Additional file 1: Table S9). Indeed, the *G. buenoi* CYP protein family size is only exceeded by that of *T. castaneum* (131 proteins). CYP genes fall into one of the four distinct subfamilies: Clan 2 (6 genes), Clan mito (62 genes), Clan 3 (25 genes) and Clan 4 (10 genes) (Fig. 5; see Additional file 1: Supplementary Data). Similarly, the number of UGT genes is also higher than that of *O. fasciatus* (1) [26], *C. lectularius* (7) [25], *D. melanogaster* (11), *A. mellifera* (6) and *B. mori* (14) [96], and identical to *T. castaneum* (28) [96].

Interestingly, both CYP and UGT gene family expansions seem to be closely linked with tandem duplication events. In the particular case of *G. buenoi* CYPs, the Clan 2 and Clan mito have undergone relatively little gene expansion (Fig. 5a and b). However, an exceptional gene expansion is observed in the mitochondrial Clan of the *G. buenoi* CYPs, where seven CYP302Bs form a lineage-specific cluster (Fig. 5b). The Clan 3 and Clan 4 are highly expanded in insects such as *T. castaneum*, *B. mori*, *R. prolixus*, and *N. lugens*, as well as in *G. buenoi*, of which 45% (28/62 CYP genes) might have been generated by tandem gene duplications (Fig. 5c and d). On the other hand, ten UGT genes are clustered on Scaffold1549, suggesting gene duplication events may have produced this large gene cluster (Additional file 1: Figure S6). In addition, multiple UGT genes are linked within Scaffold1323, Scaffold3228, and Scaffold2126. A consensus Maximum-likelihood tree (Additional file 1: Figure S7) based on the conserved C-terminal half of the deduced amino acid sequences from *G. buenoi* UGTs supports the conclusion that genes clustered within the genome derive from recent tandem duplications.
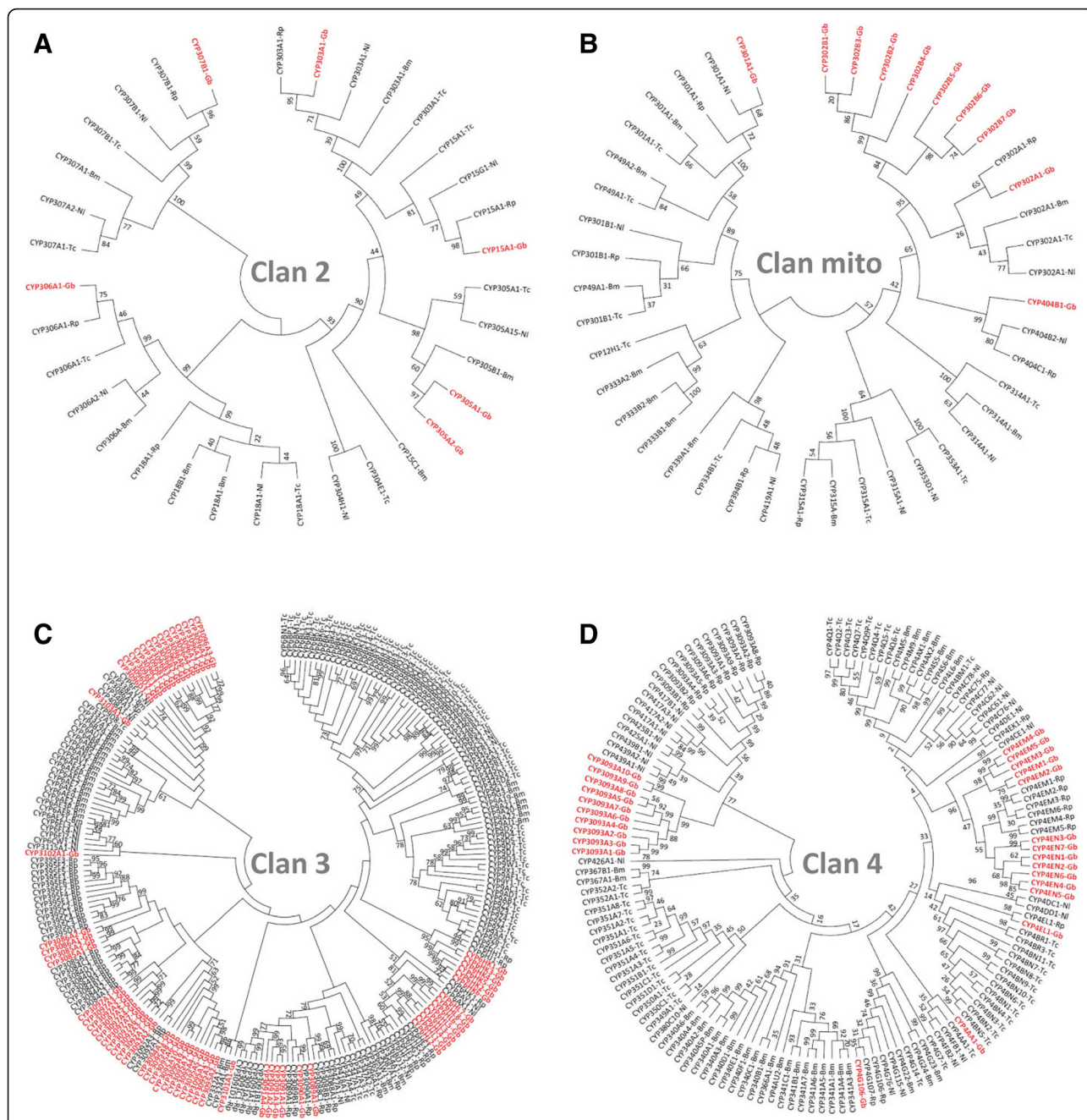
Armisén *et al. BMC Genomics*   (2018) 19:832

Page 10 of 16

Fig. 5 Phylogenetic analysis of four different Clans of the cytochrome P450s of *Gerris buenoi* with other insect species. **a** Clan 2, **b** Clan mitochondria, **c** Clan 3, and **d** Clan 4. The *G. buenoi* sequences are indicated in red and bold

Overall, our phylogenetic analysis revealed the conservation of CYPs and UGTs across insects, and the possibility for expansions via lineage-specific gene duplication. We hypothesize that this expansion may have been important in order to diversify the xenobiotic detoxification range and the regulation of endobiotics during the terrestrial-to-water surface transition.

Conclusions

The sequencing of the *G. buenoi* genome provides a unique opportunity to understand the molecular mechanisms underlying initial adaptations to water surface life and the subsequent diversification that followed. In particular, gene duplication is known to drive the evolution of adaptations and evolutionary innovations in a variety of lineages including water striders [80, 97–99]. The *G.*

buenoi genome revealed a number of clustered duplications in genes that can be linked to processes associated with the specialized life style of water striders. Some are shared with closely related Hemiptera, for example, the duplicated factor *Beadex* is an activator of the *Achaete/Scute* complex known to play an important role in bristle development. Other genes and gene family duplications are particularly rare, such as that found with the insulin receptors, which are known in other insects to be involved in a range of rocesses including wing development, growth, as well as a number of life history traits including reproduction [49, 52, 100]. The functional significance of the duplication of the histone methyltransferase *grappa* and histone cluster duplications remains unknown, yet opens up new avenues for investigation into the relationship between epigenetics and phenotypic plasticity. Expansions in the cuticle protein families involved in desiccation resistance or genes repertoires involved in xenobiotic detoxification and endobiotic regulation pathways may have played an important role during water surface specialization [78, 101]. Furthermore, the expansion of the opsin gene family and possible light sensitivity shifts are also likely associated with particularities of polarized light detection within the aquatic environment in which *G. buenoi* specializes. The impact of these duplications on the ability of water striders to function efficiently in water surface habitats remains to be experimentally tested. *G. buenoi*, which is now emerging as a tractable experimental model, offers a range of experimental tools to test these hypotheses. More generally, the *G. buenoi* genome provides a good opportunity to further understand the molecular and developmental genetic basis underlying adaptive radiations and diversification upon the conquest of new ecological habitats.

Methods

Animal collection and rearing

Adult *G. buenoi* individuals were collected from a pond in Toronto, Ontario, Canada. *G. buenoi* were kept in aquaria at 25 °C with a 14-h light/10-h dark cycle and fed on live crickets. Pieces of floating Styrofoam were regularly supplied to female water striders to lay eggs. The colony was inbred following a sib-sib mating protocol for six generations prior to DNA/RNA extraction.

DNA and total RNA extraction

Genomic DNA was isolated from adults using Qiagen Genome Tip 20 (Qiagen Inc., Valencia CA). The 180 and 500 bp paired-end libraries as well as the 3 kb mate-pair library were made from eight adult males. The 8 kb mate-pair library was made from eight adult females. Total RNA was isolated from 39 embryos, three first instar nymphs, one second instar nymph, one third instar nymph, one fourth instar nymph, one fifth instar nymph,

one adult male and one adult female. RNA was extracted using a Trizol protocol (Invitrogen).

Genome sequencing and assembly

Genomic DNA was sequenced using HiSeq2500 Illumina technology. 180 and 500 bp paired-end and 3 and 10 kb mate-pair libraries were constructed and 100 bp reads were sequenced. Estimated coverage was 28.6×, 7.3×, 21×, 17×, 72.9× respectively for each library. Sequenced reads were assembled in draft assembly using ALLPATHS-LG [102] and automatically annotated using custom MAKER2 annotation pipeline [103]. (More details can be found in Additional file 1: Supplementary Data). Expected genome size was calculated counting from Kmer based methods and using Jellyfish 2.2.3 and perl scripts from [104].

Community curation of the *G. buenoi* genome

International groups within the i5k initiative have collaborated on manual curation of *G. buenoi* automatic annotation. These curators selected genes or gene families based on their own research interests and manually curated MAKER-predicted gene set GBUE_v0.5.3 at the i5k Workspace@NAL [105] resulting in the non-redundant Official Gene Set OGSv1.0 [106].

Assessing genome assembly and annotation completeness with BUSCOs

Genome assembly completeness was assessed using BUSCO [29]. The Arthropoda gene set of 2675 single copy genes was used to test *G. buenoi* predicted genes.

Orthology analyses

OrthoDB8 (http://orthodb.org/) was used to find orthologues of *G. buenoi* (OGS 1.0) on 76 arthropod species. Proteins on each species were categorised using custom Perl scripts according to the number of hits on other eight arthropod species: *Drosophila melanogaster*, *Danaus plexippus*, *Tribolium castaneum*, *Apis mellifera*, *Acyrthosiphon pisum*, *Cimex lectularius*, *Pediculus humanus* and *Daphnia pulex*.

Insulin receptors phylogeny

Sequences were retrieved from 'nr' database by sequence similarity using BLASTp with search restricted to Insecta (taxid:50557). Each *G. buenoi* InR sequence was individually blasted and best 250 hits were recovered. A total of 304 unique id sequences were retrieved. Additionally, we recovered the genes annotated by Kremer et al. [47] as well as *Caenorhabditis elegans* insulin receptor homolog AAC47715.1 as outgroup. We performed a preliminary analysis aligning the sequences with Clustal Omega [107–109] and building a simple phylogeny using MrBayes

[110] (one chain, 100,000 generations). Based on that preliminary phylogeny, we selected a single isoform for each InR gene (Additional file 4). Final InR phylogeny tree was estimated aligning the sequences with MAFFT [111] using E-INS-i iterative method and using MrBayes (four chains, for 1,000,000 generations). Final phylogeny include InR sequences from (copy number in parenthesis):

Acromyrmex echinatior (2), *Acyrthosiphon pisum* (2), *Aedes aegypti* (1), *Aedes albopictus* (1), *Aethina tumida* (2), *Agrilus planipennis* (2), *Amyelois transitella* (1), *Anopheles darlingi* (1), *Anopheles gambiae* (1), *Anopheles sinensis* (1), *Anoplophora glabripennis* (2), *Aphis citricidus* (2), *Apis cerana* (2), *Apis dorsata* (2), *Apis florea* (1), *Apis mellifera* (2), *Aquarius paludum* (3), *Athalia rosae* (2), *Atta cephalotes* (2), *Atta colombica* (2), *Bactrocera dorsalis* (1), *Bactrocera latifrons* (1), *Bactrocera oleae* (1), *Bemisia tabaci* (2), *Blattella germanica* (3), *Bombus impatiens* (2), *Bombus terrestris* (2), *Bombyx mori* (1), *Caenorhabditis elegans* (1), *Camponotus floridanus* (2), *Cephus cinctus* (2), *Ceratina calcarata* (2), *Ceratitis capitata* (1), *Ceratosolen solmsi marchali* (1), *Cimex lectularius* (2), *Clunio marinus* (1), *Copidosoma floridanum* (1), *Cryptotermes secundus* (3), *Cyphomyrmex costatus* (2), *Danaus plexippus* (1), *Daphnia pulex* (4), *Dendroctonus ponderosae* (1), *Diachasma alloeum* (2), *Diaphorina citri* (1), *Dinoponera quadriceps* (2), *Diuraphis noxia* (2), *Drosophila ananassae* (1), *Drosophila arizonae* (1), *Drosophila biarmipes* (1), *Drosophila bipectinata* (1), *Drosophila busckii* (1), *Drosophila elegans* (1), *Drosophila erecta* (1), *Drosophila eugracilis* (1), *Drosophila ficusphila* (1), *Drosophila grimshawi* (1), *Drosophila kikkawai* (1), *Drosophila melanogaster* (1), *Drosophila miranda* (1), *Drosophila mojavensis* (1), *Drosophila obscura* (1), *Drosophila persimilis* (1), *Drosophila pseudoobscura* (1), *Drosophila rhopaloa* (1), *Drosophila sechellia* (1), *Drosophila serrata* (1), *Drosophila simulans* (1), *Drosophila suzukii* (1), *Drosophila takahashii* (1), *Drosophila virilis* (1), *Drosophila willistoni* (1), *Drosophila yakuba* (1), *Dufourea novaeangliae* (2), *Ephemera danica* (2), *Eufriesea mexicana* (2), *Fopius arisanus* (2), *Gerris buenoi* (3), *Glossina morsitans morsitans* (1), *Habropoda laboriosa* (2), *Halyomorpha halys* (2), *Harpegnathos saltator* (2), *Hebrus sp* (3), *Helicoverpa armigera* (1), *Heliothis virescens* (1), *Hydrometra cumata* (3), *Lasius niger* (2), *Leptinotarsa decemlineata* (2), *Limnoporus dissortis* (3), *Linepithema humile* (2), *Locusta migratoria* (2), *Macrotermes natalensis* (3), *Manduca sexta* (1), *Maruca vitrata* (1), *Megachile rotundata* (2), *Melipona quadrifasciata* (1), *Mesovelia furcata* (3), *Microplitis demolitor* (2), *Microvelia longipes* (3), *Monochamus alternatus* (1), *Monomorium pharaonis* (2), *Musca domestica* (1), *Myzus persicae* (2), *Nasonia vitripennis* (1), *Neodiprion lecontei* (2), *Nicrophorus vespilloides* (2), *Nilaparvata lugens* (2), *Oncopeltus fasciatus* (2), *Onthophagus nigriventris* (1), *Onthophagus taurus* (2), *Ooceraea biroi* (2), *Orussus abietinus* (2), *Oryctes borbonicus* (1), *Papilio machaon* (2), *Papilio polytes* (1), *Papilio xuthus* (1), *Parasteatoda tepidariorum* (2), *Pediculus humanus corporis* (1), *Pieris rapae* (1), *Plutella xylostella* (1), *Pogonomyrmex barbatus* (2), *Polistes canadensis* (2), *Polistes dominula* (2), *Pseudomyrmex gracilis* (2), *Rhagoletis zephyria* (1), *Rhagovelia antilleana* (3), *Rhodnius prolixus* (2), *Solenopsis invicta* (2), *Spodoptera litura* (1), *Stomoxys calcitrans* (1), *Strigamia maritima* (1), *Trachymyrmex cornetzi* (2), *Trachymyrmex septentrionalis* (2), *Trachymyrmex zeteki* (2), *Tribolium castaneum* (2), *Trichogramma pretiosum* (1), *Trichomalopsis sarcophagae* (1), *Vollenhovia emeryi* (2), *Wasmannia auropunctata* (2), *Zeugodacus cucurbitae* (1), and *Zootermopsis nevadensis* (3).

Cytochrome P450 proteins phylogeny

CYPs phylogenetic analysis was performed using Maximum-Likelihood method and the trees were generated by MEGA 6. The phylogenetic trees were generated by MEGA 6 with Maximum-Likelihood method using the amino acid sequences from *Gerris buenoi* (Gb), *Rhodnius prolixus* (Rp), *Nilaparvata lugens* (Nl), *Bombyx mori* (Bm) and *Tribolium castaneum* (Tc). All nodes have significant bootstrap support based on 1000 replicates.

Additional files

Additional file 1: Supplementary Online Information. Additional files 10 and 11. (ZIP 18100 kb)

Additional file 2: List of annotated autophagy genes in *Gerris buenoi*. (XLSX 12 kb)

Additional file 3: Accession number and gene name of 225 arthropod genes used to reconstruct Insulin receptors phylogeny (highlighted). (XLSX 91 kb)

Additional file 4: Nucleotide sequences of Insulin receptors genes annotated in Gerromorpha (*Gerris buenoi, Aquarius paludum, Limnoporus dissortis, Rhagovelia antilleana, Microvelia longipes, Mesovelia furcata, Hydrometra cumata* and *Hebrus sp*). (DOCX 40kb)

Additional file 5: SNPs summary at position 17, 64, 70 and 137 of 114 LWS opsin homologs from 54 species representing 12 insect orders. (XLSX 26 kb)

Additional file 6: List of annotated aquaporin genes in *Gerris buenoi* including gene names and closest identity. (XLSX 58 kb)

Additional file 7: List of annotated aquaporin genes in *Gerris buenoi* including gene names and closest identity. (XLSX 15 kb)

Additional file 8: Protein sequences of annotated chemoreceptors in *Gerris buenoi* : 155 Or, 135 Gr and 45 IR genes. (DOCX 83 kb)

Additional file 9: List of cytochrome P450 (CYP) genes annotated from the *Gerris buenoi* genome. (XLSX 14 kb)

Additional file 10: *Gerris buenoi* serosins nucleotide and protein sequences. (DOCX 12 kb)

Additional file 11: List of antioxidant enzyme genes identified in the *Gerris buenoi* genome. (XLSX 13 kb)

Acknowledgements

Funding

Availability of data and materials

The datasets generated and/or analysed during the current study are available in GenBank BioProject PRJNA203045. Genome assembly can found at [113], gene annotation can be found at [114], and official gene set can be found at [106].

Image rights

The authors acquired all images in this manuscript.

Authors' contributions

DA, AK, and SR conceived, managed, and coordinated the project; SD-P, DMM, RAG and SR managed the i5k project; SLL led the submissions team; HVD, HC led the library team; YH, HD led sequencing team; JQ, SCM, DSTH, SR and KCW assembled the genome; DSTH and SR did the automated annotation; DA, RR, MF, JBB, HMR, KAP, S-JA, MM-T, EA, FB, TC, CPC, AGC, AJJC, AD, EMD, ED, ENE, M-JF, CGCJ, AJ, ECJ, JWJ, MPL, ML, AM, BO, AL-O, AR, PNR, AJR, MES, WT, MvdZ, IMVJ, AVL and SV contributed to manual curation; LMC, CPC, MMT and MFP performed curation quality control and generated the OGS; EA, M-JF contributed to statistical analyses of wing development and polyphenism genes; IMVJ and KAP performed associated phylogenetic and synteny analyses of Wnt and homeodomain transcription factor gene clusters; DA performed InR phylogeny and InR analyses; RR contributed with Insulin/TOR signaling genes analyses and InR analyses; MF perfomed opsins analyses; JBB analyzed aquaporins, cuticle genes and antioxidant proteins; HMR carried out chemoreceptors analyses; AJ, KAP and IMVJ perfomed homeobox gene and Wnt signaling pathway analyses; S-JA analyzed UGTs detoxification genes; EA conducted wing polyphenism analyses; FB performed nuclear receptors and bHLH-PAS proteins analyses; AGC carried out early developmental genes analyses; ED conducted histone genes and histone modification machinery; ENE, AM and BO performed cysteine peptidase analyses; CF contributed to analysis of Beadex duplication;; CGCJ and MvdZ performed immune genes analyses; RR, MF, JBB, HMR, KAP, S-JA, EA, FB, AGC, ED, ENE, CF, AJ, BO, AM, AL-O, MvdZ and IMVJ contributed to report writing; DA and AK wrote the manuscript; DA, AK, RR, FB, MF, JBB, HMR, KAP and SR edited the manuscript; DA and AK organized the Additional file 1. All authors approved the final manuscript.

Ethics approval

The bugs were collected in road side ditches near Toronto and no permission was required.

Consent for publication

Not applicable.

Competing interests

Elizabeth Duncan and Stephen Richards are an Associate Editors at BMC Genomics.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5242, Ecole Normale Supérieure de Lyon 46, allée d'Italie, 69364 Lyon Cedex 07, France. [2]Department of Molecular Genetics & Microbiology and UF Genetics Institute, University of Florida, 2033 Mowry Road, Gainesville, FL 32610-3610, USA. [3]Department of Biological Sciences, Wayne State University, Detroit, MI 48202, USA. [4]Department of Biological Sciences, McMicken College of Arts and Sciences, University of Cincinnati, 318 College Drive, Cincinnati, OH 45221-0006, USA. [5]Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [6]Institute for Zoology: Developmental Biology, University of Cologne, Zülpicher Str. 47b, 50674 Cologne, Germany. [7]School of Life Sciences, University of Warwick, Gibbet Hill Campus, Coventry CV4 7AL, UK. [8]USDA-ARS Horticultural Crops Research Unit, 3420 NW Orchard Avenue, Corvallis, OR 97330, USA. [9]Department of Crop and Soil Science, Oregon State University, 3050 SW Campus Way, Corvallis, OR 97331, USA. [10]USDA Agricultural Research Service, National Agricultural Library, Beltsville, MD 20705, USA. [11]Human Genome Sequencing Center, Department of Human and Molecular Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. [12]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA. [13]Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [14]Department of Biology, McGill University, 1205 Avenue Docteur Penfield Avenue, Montréal, Québec H3A 1B1, Canada. [15]Laboratory for Evolution and Development, Department of Biochemistry, University of Otago, P.O. Box 56, Dunedin, New Zealand. [16]School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK. [17]A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow 119991, Russia. [18]Institute of Biology, Leiden University, Sylviusweg 72, 2333 BE Leiden, Netherlands. [19]Max Planck Institute for Chemical Ecology, Hans-Knöll Strasse 8, 07745 Jena, Germany. [20]Department of Entomology, University of Maryland, College Park, MD 20742, USA. [21]Smithsonian Tropical Research Institute, Apartado Postal 0843-03092, Balboa Ancon, Panama City, Panama. [22]Center of Life Sciences, Skolkovo Institute of Science and Technology, Skolkovo 143025, Russia. [23]USDA ARS Center for Grain and Animal Health Research, 1515 College Ave., Manhattan KS-66502, USA. [24]Department of Evolutionary Genetics, Max-Planck-Institut für Evolutionsbiologie, August-Thienemann-Straße 2, 24306 Plön, Germany.

References

1. Andersen NM. The semiaquatic bugs, vol. 3. Klampenborg: Scandinavian Science Pres Ltd; 1982.
2. Crumiere AJ, Santos ME, Semon M, Armisen D, Moreira FF, Khila A. Diversity in morphology and Locomotory behavior is associated with niche expansion in the semi-aquatic bugs. Curr Biol. 2016;26(24):3336–42.
3. Hu DL, Bush JW. The hydrodynamics of water-walking arthropods. J Fluid Mech. 2010;644:29.
4. Polhemus JT. Surface wave communication in water striders; field observations of unreported taxa (Heteroptera: Gerridae, Veliidae). J N Y Entomol Soc. 1990;98(3):383–84.
5. Damgaard J. What do we know about the phylogeny of the semi-aquatic bugs (Hemiptera: Heteroptera: Gerromorpha)? Entomol Americana. 2012; 118(1):81–98.
6. Damgaard J, Andersen NM, Meier R. Combining molecular and morphological analyses of water strider phylogeny (Hemiptera–Heteroptera, Gerromorpha): effects of alignment and taxon sampling. Syst Entomol. 2005; 30(2):289–309.

Armisén *et al. BMC Genomics*   (2018) 19:832

Page 14 of 16

7. Andersen NM. The evolution of marine insects: phylogenetic, ecological and geographical aspects of species diversity in marine water striders. Ecography. 1999;22(1):98–111.

8. Ikawa T, Okabe H, Cheng L. Skaters of the seas – comparative ecology of nearshore and pelagic Halobates species (Hemiptera: Gerridae), with special reference to Japanese species. Mar Biol Res. 2012;8(10):915–36.

9. Santos ME, Berger CS, Refki PN, Khila A. Integrating evo-devo with ecology for a better understanding of phenotypic evolution. Brief Funct Genomics. 2015;14(6):384–95.

10. Schluter D. The ecology of adaptive radiation. New York: Oxford University Press; 2000.

11. Simpson GG. The major features of evolution. New York City: Columbia University Press; 1953.

12. Andersen NM. A comparative study of locomotion on the water surface in semiaquatic bugs (Insecta, Hemiptera, Gerromorpha). Vidensk Mcddr dansk naturh Foren. 1976;139:337–96.

13. Gao X, Jiang L. Biophysics: water-repellent legs of water striders. Nature. 2004;432(7013):36.

14. Fairbairn DJ, King E. Why do Californian striders fly? J Evol Biol. 2009; 22(1):36–49.

15. Khila A, Abouheif E, Rowe L. Evolution of a novel appendage ground plan in water striders is driven by changes in the Hox gene Ultrabithorax. PLoS Genet. 2009;5(7):e1000583.

16. Khila A, Abouheif E, Rowe L. Function, developmental genetics, and fitness consequences of a sexually antagonistic trait. Science. 2012;336(6081):585–9.

17. Khila A, Abouheif E, Rowe L. Comparative functional analyses of ultrabithorax reveal multiple steps and paths to diversification of legs in the adaptive radiation of semi-aquatic insects. Evolution. 2014;68(8):2159–70.

18. Refki PN, Armisen D, Crumiere AJ, Viala S, Khila A. Emergence of tissue sensitivity to Hox protein levels underlies the evolution of an adaptive morphological trait. Dev Biol. 2014;392(2):441–53.

19. Refki PN, Khila A. Key patterning genes contribute to leg elongation in water striders. Evodevo. 2015;6:14.

20. Arnqvist G, Rowe L. Antagonistic coevolution between the sexes in a group of insects. Nature. 2002;415(6873):787–9.

21. Arnqvist G, Rowe L. Correlated evolution of male and female morphologles in water striders. Evolution. 2002;56(5):936–47.

22. Devost E, Turgeon J. The combined effects of pre- and post-copulatory processes are masking sexual conflict over mating rate in Gerris buenoi. J Evol Biol. 2016;29(1):167–77.

23. Calabrese DM, Tallerico P. Chromosome study in males of Nearctic species of Gerris and Limnoporus Hemiptera Heteroptera Gerridae. Proc Entomol Soc Wash. 1982;84(3):4.

24. Armisen D, Refki PN, Crumiere AJ, Viala S, Toubiana W, Khila A. Predator strike shapes antipredator phenotype through new genetic interactions in water striders. Nat Commun. 2015;6:8153.

25. Benoit JB, Adelman ZN, Reinhardt K, Dolan A, Poelchau M, Jennings EC, Szuter EM, Hagan RW, Gujar H, Shukla JN, et al. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. Nat Commun. 2016;7:10165.

26. Panfilio KA, Vargas Jentzsch IM, Benoit JB, Erezyilmaz D, Suzuki Y, Colella S, Robertson HM, Poelchau MF, Waterhouse RM, Ioannidis P, et al. Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. bioRxiv. 2017:201731. https://doi.org/10.1101/201731.

27. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simao FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. 2017; 45(D1):D744–9.

28. Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simao FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids Res. 2015;43(Database issue):D250–6.

29. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.

30. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2018; 35(3):543–48.

31. Krumlauf R. Evolution of the vertebrate Hox homeobox genes. BioEssays. 1992;14(4):245–52.

32. Cavodeassi F, Modolell J, Gomez-Skarmeta JL. The Iroquois family of genes: from body building to neural patterning. Development. 2001; 128(15):2847–55.

33. Negre B, Ruiz A. HOM-C evolution in Drosophila: is there a need for Hox gene clustering? Trends Genet. 2007;23(2):55–9.

34. Dilda CL, Mackay TF. The genetic architecture of Drosophila sensory bristle number. Genetics. 2002;162(4):1655–74.

35. Norga KK, Gurganus MC, Dilda CL, Yamamoto A, Lyman RF, Patel PH, Rubin GM, Hoskins RA, Mackay TF, Bellen HJ. Quantitative analysis of bristle number in Drosophila mutants identifies genes involved in neural development. Curr Biol. 2003;13(16):1388–96.

36. Asmar J, Biryukova I, Heitzler P. Drosophila dLMO-PA isoform acts as an early activator of achaete/scute proneural expression. Dev Biol. 2008;316(2):487–97.

37. Baker KD, Thummel CS. Diabetic larvae and obese flies-emerging studies of metabolism in Drosophila. Cell Metab. 2007;6(4):257–66.

38. Edgar BA. How flies get their size: genetics meets physiology. Nat Rev Cancer. 2006;7(12):907–16.

39. Martin DE, Hall MN. The expanding TOR signaling network. Curr Opin Cell Biol. 2005;17(2):158–66.

40. Junger MA, Rintelen F, Stocker H, Wasserman JD, Vegh M, Radimerski T, Greenberg ME, Hafen E. The Drosophila forkhead transcription factor FOXO mediates the reduction in cell number associated with reduced insulin signaling. J Biol. 2003;2(3):20.

41. Puig O, Tjian R. Transcriptional feedback control of insulin receptor by dFOXO/FOXO1. Genes Dev. 2005;19(20):2435–46.

42. Wang MC, Bohmann D, Jasper H. JNK extends life span and limits growth by antagonizing cellular and organism-wide responses to insulin signaling. Cell. 2005;121(1):115–25.

43. Ward CW, Lawrence MC. Ligand-induced activation of the insulin receptor: a multi-step process involving structural changes in both the ligand and the receptor. BioEssays. 2009;31(4):422–34.

44. Hernandez-Sanchez C, Mansilla A, de Pablo F, Zardoya R. Evolution of the insulin receptor family and receptor isoform expression in vertebrates. Mol Biol Evol. 2008;25(6):1043–53.

45. Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A, Arnold GJ, Basu MK, et al. The ecoresponsive genome of Daphnia pulex. Science. 2011;331(6017):555–61.

46. Boucher P, Ditlecadet D, Dube C, Dufresne F. Unusual duplication of the insulin-like receptor in the crustacean Daphnia pulex. BMC Evol Biol. 2010;10:305.

47. Kremer LPM, Korb J, Bornberg-Bauer E. Reconstructed evolution of insulin receptors in insects reveals duplications in early insects and cockroaches. J Exp Zool B Mol Dev Evol. 2018; 330(5):305–11.

48. Kaessmann H, Vinckenbosch N, Long M. RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet. 2009;10(1):19–31.

49. Emlen DJ, Szafran Q, Corley LS, Dworkin I. Insulin signaling and limb-patterning: candidate pathways for the origin and evolutionary diversification of beetle 'horns'. Heredity (Edinb). 2006;97(3):179–91.

50. Hattori A, Sugime Y, Sasa C, Miyakawa H, Ishikawa Y, Miyazaki S, Okada Y, Cornette R, Lavine LC, Emlen DJ, et al. Soldier morphogenesis in the damp-wood termite is regulated by the insulin signaling pathway. J Exp Z B Mol Dev Evol. 2013;320(5):295–306.

51. Patel A, Fondrk MK, Kaftanoglu O, Emore C, Hunt G, Frederick K, Amdam GV. The making of a queen: TOR pathway is a key player in diphenic caste development. PLoS One. 2007;2(6):e509.

52. Xu HJ, Xue J, Lu B, Zhang XC, Zhuo JC, He SF, Ma XF, Jiang YQ, Fan HW, Xu JY, et al. Two insulin receptors determine alternative wing morphs in planthoppers. Nature. 2015;519(7544):464–7.

53. Suen G, Teiling C, Li L, Holt C, Abouheif E, Bornberg-Bauer E, Bouffard P, Caldera EJ, Cash E, Cavanaugh A, et al. The genome sequence of the leaf-cutter ant Atta cephalotes reveals insights into its obligate symbiotic lifestyle. PLoS Genet. 2011;7(2):e1002007.

54. Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, Holt C, Hu H, Abouheif E, Benton R, et al. Draft genome of the red harvester ant Pogonomyrmex barbatus. Proc Natl Acad Sci. 2011;108(14):5667–72.

55. Smith CD, Zimin A, Holt C, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG, et al. Draft genome of the globally widespread and invasive argentine ant (Linepithema humile). Proc Natl Acad Sci. 2011; 108(14):5673–8.

56. Brisson JA. Aphid wing dimorphisms: linking environmental and genetic control of trait variation. Philos Trans R Soc Lond Ser B Biol Sci. 2010; 365(1540):605–16.

57. Elango N, Hunt BG, Goodisman MA, Yi SV. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, Apis mellifera. Proc Natl Acad Sci. 2009;106(27):11206–11.

58. Fischer C, Mahner M, Wachmann E. The rhabdom structure in the ommatidia of the Heteroptera (Insecta), and its phylogenetic significance. Zoomorphology. 2000;120(1):1–13.

59. Schneider L, Langer H. Die Struktur des Rhabdoms im "Doppelauge" des Wasserläufers Gerris lacustris. Z Zellforsch Mikrosk Anat. 1969;4:538–59.

60. Meyer HW. Visuelle Schlüsselreize für die Auslösung der Beutefanghandlung beim Bachwasserläufer Velia capraí (Hemiptera, Heteroptera). Z Vgl Physiol. 1971;72(3):260–97.

61. Rowe L. The costs of mating and mate choice in water striders. Anim Behav. 1994;48:1049–56.

62. Spence JR, Anderson N. Biology of water striders: interactions between systematics and ecology. Annu Rev Entomol. 1994;39(1):101–28.

63. Dahmen H. Eye specialisation in waterstriders: an adaptation to life in a flat world. J Comp Physiol A. 1991;169(5):623–32.

64. Wolburg-Buchholz K. The organization of the lamina ganglionaris of the hemipteran insects, Notonecta glauca, Corixa punctata and *Gerris lacustris*. Cell Tissue Res. 1979;197(1):39–59.

65. Bohn H, Täuber U. Beziehungen zwischen der Wirkung polarisierten Lichtes auf das Elektroretinogramm und der Ultrastruktur des Auges von Gerris lacustris L. Z Vgl Physiol. 1971;72(1):32–53.

66. Frolov R, Weckström M. Developmental changes in biophysical properties of photoreceptors in the common water strider (Gerris lacustris): better performance at higher cost. J Neurophysiol. 2014;112(4):913–22.

67. Henze MJ, Oakley TH. The dynamic evolutionary history of Pancrustacean eyes and opsins. Integr Comp Biol. 2015;55(5):830–42.

68. Frentiu FD, Bernard GD, Cuevas CI, Sison-Mangus MP, Prudic KL, Briscoe AD. Adaptive evolution of color vision as seen through the eyes of butterflies. Proc Natl Acad Sci. 2007;104 Suppl 1:8634–40.

69. Frentiu FD, Bernard GD, Sison-Mangus MP, Brower AV, Briscoe AD. Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies. Mol Biol Evol. 2007;24(9):2016–28.

70. Sharkey CR, Fujimoto MS, Lord NP, Shin S, McKenna DD, Suvorov A, Martin GJ, Bybee SM. Overcoming the loss of blue sensitivity through opsin duplication in the largest animal group, beetles. Reports. 2017;7(1):8.

71. Cloudsley-Thompson JL. Evolution and adaptation of terrestrial arthropods. Springer, Berlin, Heidelberg; 2012.

72. Benoit JB, Hansen IA, Attardo GM, Michalkova V, Mireji PO, Bargul JL, Drake LL, Masiga DK, Aksoy S. Aquaporins are critical for provision of water during lactation and intrauterine progeny hydration to maintain tsetse fly reproductive success. PLoS Negl Trop Dis. 2014;8(4):e2517.

73. Benoit JB, Hansen IA, Szuter EM, Drake LL, Burnett DL, Attardo GM. Emerging roles of aquaporins in relation to the physiology of blood-feeding arthropods. J Comp Physiol B. 2014;184(7):811–25.

74. Campbell EM, Ball A, Hoppler S, Bowman AS. Invertebrate aquaporins: a review. J Comp Physiol B. 2008;178(8):935–55.

75. Wigglesworth VB. Transpiration through the cuticle of insects. J Exp Biol. 1945;21(3–4):97–114.

76. Ioannidou ZS, Theodoropoulou MC, Papandreou NC, Willis JH, Hamodrakas SJ. CutProtFam-Pred: detection and classification of putative structural cuticular proteins from sequence alone, based on profile hidden Markov models. Insect Biochem Mol Biol. 2014;52:51–9.

77. Willis JH. Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. Insect Biochem Mol Biol. 2010;40(3):189–204.

78. Guan X, Middlebrooks BW, Alexander S, Wasserman SA. Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in Drosophila. Proc Natl Acad Sci. 2006;103(45):16794–9.

79. Tseng M, Rowe L. Sexual dimorphsim and allometry in the giant water strider *Gigantometra gigas*. Can J Zool. 1999;77(6):923–9.

80. Santos ME, Le Bouquin A, Crumiere AJJ, Khila A. Taxon-restricted genes at the origin of a novel trait allowing access to a new environment. Science. 2017;358(6361):386–90.

81. Benton R. Multigene family evolution: perspectives from insect chemoreceptors. Trends Ecol Evol. 2015;30(10):590–600.

82. Joseph RM, Carlson JR. Drosophila chemoreceptors: a molecular Interface between the chemical world and the brain. Trends Genet. 2015;31(12):683–95.

83. Weiss LA, Dahanukar A, Kwon JY, Banerjee D, Carlson JR. The molecular and cellular basis of bitter taste in Drosophila. Neuron. 2011;69(2):258–72.

84. Derby CD, Sorensen PW. Neural processing, perception, and behavioral responses to natural chemical stimuli by fish and crustaceans. J Chem Ecol. 2008;34(7):898–914.

85. Silbering AF, Rytz R, Grosjean Y, Abuin L, Ramdya P, Jefferis GS, Benton R. Complementary function and integrated wiring of the evolutionarily distinct Drosophila olfactory subsystems. J Neurosci. 2011;31(38):13357–75.

86. Ffrench-Constant RH, Daborn PJ, Le Goff G. The genetics and genomics of insecticide resistance. Trends Genet. 2004;20(3):163–70.

87. Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol. 1999;29(9):757–77.

88. Rewitz KF, O'Connor MB, Gilbert LI. Molecular evolution of the insect Halloween family of cytochrome P450s: phylogeny, gene organization and functional conservation. Insect Biochem Mol Biol. 2007;37(8):741–53.

89. Helvig C, Koener JF, Unnithan GC, Feyereisen R. CYP15A1, the cytochrome P450 that catalyzes epoxidation of methyl farnesoate to juvenile hormone III in cockroach corpora allata. Proc Natl Acad Sci. 2004;101(12):4024–9.

90. Good RT, Gramzow L, Battlay P, Sztal T, Batterham P, Robin C. The molecular evolution of cytochrome P450 genes within and between drosophila species. Genome Biol Evol. 2014;6(5):1118–34.

91. Ahn SJ, Vogel H, Heckel DG. Comparative analysis of the UDP-glycosyltransferase multigene family in insects. Insect Biochem Mol Biol. 2012;42(2):133–47.

92. Ahmad SA, Hopkins TL. Phenol β-glucosyltransferases in six species of insects: properties and tissue localization. Comp Biochem Physiol B Comp Biochem. 1993;104(3):515–9.

93. Morello A, Repetto Y. UDP-glucosyltransferase activity of housefly microsomal fraction. Biochem J. 1979;177(3):809–12.

94. Lao SH, Huang XH, Huang HJ, Liu CW, Zhang CX, Bao YY. Genomic and transcriptomic insights into the cytochrome P450 monooxygenase gene repertoire in the rice pest brown planthopper, *Nilaparvata lugens*. Genomics. 2015;106(5):301–9.

95. Schama R, Pedrini N, Juarez MP, Nelson DR, Torres AQ, Valle D, Mesquita RD. Rhodnius prolixus supergene families of enzymes potentially associated with insecticide resistance. Insect Biochem Mol Biol. 2016;69:91–104.

96. Hughes AL. Origin of Ecdysosteroid UDP-glycosyltransferases of Baculoviruses through horizontal gene transfer from Lepidoptera. Coevolution. 2013;1(1):1–7.

97. Pebusque MJ, Coulier F, Birnbaum D, Pontarotti P. Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. Mol Biol Evol. 1998;15(9):1145–59.

98. Seoighe C, Gehring C. Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. Trends Genet. 2004;20(10):461–4.

99. Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet. 2004;38:615–43.

100. Sang M, Li C, Wu W, Li B. Identification and evolution of two insulin receptor genes involved in Tribolium castaneum development and reproduction. Gene. 2016;585(2):196–204.

101. Cornman RS, Willis JH. Annotation and analysis of low-complexity protein families of Anopheles gambiae that are associated with cuticle. Insect Mol Biol. 2009;18(5):607–22.

102. Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB. ALLPATHS: de novo assembly of whole-genome shotgun microreads. Genome Res. 2008;18(5):810–20.

103. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12:491.

104. Estimate Genome Size perl scripts. https://github.com/josephryan/estimate_genome_size.pl. Accessed 21 Mar 2016.

105. Poelchau M, Childers C, Moore G, Tsavatapalli V, Evans J, Lee CY, Lin H, Lin JW, Hackett K. The i5k workspace@NAL--enabling genomic data access, visualization and curation of arthropod genomes. Nucleic Acids Res. 2015; 43(Database issue):D714–9.

106. *Gerris buenoi* Official Gene Set OGSv1.0. https://data.nal.usda.gov/dataset/gerris-buenoi-official-gene-set-v10. Accessed 21 Aug 2018.

107. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, Lopez R. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. 2015;43(W1):W580–4.

108. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. Analysis tool web services from the EMBL-EBI. Nucleic Acids Res. 2013;41(Web Server issue):W597–600.

109. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. Mol Syst Biol. 2011;7:539.

110. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.

111. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.

112. Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, Moran NA. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. Mol Biol Evol. 2014;31(4):857–71.

113. *Gerris buenoi* Genome Assembly. https://data.nal.usda.gov/dataset/gerris-buenoi-genome-assembly-10. Accessed 21 Aug 2018.

114. *Gerris buenoi* Genome Annotation. https://data.nal.usda.gov/dataset/gerris-buenoi-genome-annotations-v053. Accessed 21 Aug 2018.

115. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. Phylogenomics resolves the timing and pattern of insect evolution. Science. 2014;346(6210):763–7.