Comparative phylogenetic analysis of bacterial associates in Pyrrhocoroidea and evidence for ancient and persistent environmental symbiont reacquisition in Largidae (Hemiptera: Heteroptera).

Eric Robert Lucien Gordon[1], Quinn McFrederick[1], Christiane Weirauch[1]

[1] Department of Entomology, University of California, Riverside

Corresponding author: Eric R. L. Gordon, egord003@ucr.edu.

Abstract

The ancient insect order Hemiptera, one of the most well-studied insect lineages with respect to bacterial symbioses, still contains major branches which lack robust phylogenies and comprehensive characterization of associated bacterial symbionts. The Pyrrhocoroidea (Largidae [220 species]; Pyrrhocoridae [~300 species]) is a superfamily of the primarily-herbivorous hemipteran infraorder Pentatomomorpha, though relationships to related superfamilies are controversial. Studies on bacterial symbionts of this group have focused on members of Pyrrhocoridae, but recent examination of species of two genera of Largidae demonstrated divergent symbiotic complexes between these putative sister families. We surveyed bacterial diversity of this group using paired-end Illumina and targeted Sanger sequencing of bacterial 16S amplicons of 30 pyrrhocoroid taxa, including 17 species of Largidae, in order to determine the identity of bacterial associates and similarity of associated microbial communities among species. We also constructed the first comprehensive phylogeny of this superfamily (4,800 bp; 5 loci; 57 ingroup + 12 outgroup taxa) in order accurately trace the evolution of symbiotic complexes among Pentatomomorpha. We undertook multiple lines of investigation (*i.e.*, experimental rearing, FISH microscopy, phylogenetic and co-evolutionary analyses) to understand potential transmission routes of largid symbionts. We found a prevalent, specific association of Largidae with plant-beneficial-environmental clade *Burkholderia* housed in midgut tubules. As in other distantly-related Heteroptera, symbiotic bacteria seem to be acquired from the environment every generation. We review current understanding of symbiotic complexes within the Pentatomomorpha and discuss means to further investigations of the evolution and function of these symbioses.

Importance. Obligate symbioses with bacteria are common in insects, particularly for Hemiptera wherein varied forms of symbiosis occur, though knowledge of symbionts remains incomplete for major lineages. Thus, an accurate understanding of how these partnerships evolved and changed over millions of years is not yet achievable. We contribute to our

52 understanding of the evolution of symbiotic complexes in Hemiptera by characterizing bacterial
53 associates of Pyrrhocoroidea focusing on the family Largidae and by constructing a phylogeny
54 to establish evolutionary relationships of and within this group. Members of Largidae are
55 associated with specific symbiotic *Burkholderia* from a different clade than *Burkholderia*
56 symbionts in other Hemiptera and are members of the earliest-diverging superfamily of
57 *Burkholderia*-associated Hemiptera. Evidence suggests that species of Largidae reacquire
58 specific symbiotic bacteria every generation environmentally, a rare strategy for insects with
59 potentially volatile evolutionary ramifications, but one that has persisted in Largidae and other
60 related lineages since the Cretaceous.

61 **Introduction**

62 With over 82,000 species, Hemiptera is one of the most diverse lineages of animals and is
63 the most speciose order of insects which do not undergo complete metamorphosis (1). Their
64 success can be tied directly to bacterial symbionts, which presumably allowed the ancestor of
65 this lineage to exploit nutrient-limited diets of plant tissues such as xylem and phloem (2). Three
66 lineages (Sternorrhyncha, Auchenorrhyncha, Coleorrhyncha) are exclusively herbivorous and
67 nearly all constituent species remain associated with obligate intracellular symbionts that are
68 vertically transmitted to offspring before oviposition (3–5). These obligate partnerships with
69 bacteria may have helped to promote speciation through rapid development of potential genetic
70 incompatibilities of symbiont and hosts resulting in higher rates of hybrid mortality (6). Sources
71 of incompatibilities and other potentially deleterious mutations inevitably occur during population
72 bottlenecks of intracellular symbiont lineages during transmission between generations and can
73 eventually lead to compensatory mutations in host genomes (6). This comparatively rapid
74 evolution of both symbiont and host can effectively cripple the symbiosis over many generations
75 (7, 8). In many lineages, symbiont genomes become so reduced that former metabolic
76 capability has been supplanted by secondarily-acquired symbionts or partitioned by speciation
77 of one endosymbiont into two metabolically-distinct species (9–11).

Unlike other Hemiptera, the ancestor of the Heteroptera achieved independence from obligate symbionts at some point while transitioning from an herbivore to a predator (evolution of trophic strategies of Heteroptera summarized in [12]). However, two diverse clades of Heteroptera have secondarily re-evolved herbivory and together comprise more than 60% of heteropteran diversity, including many economically important pests of crops (13). In one of these radiations, Trichophora (Pentatomomorpha excluding Aradoidea), most members possess large populations of particular groups of extracellular symbiotic bacteria in blind tubules of the posterior midgut called caeca (14, 15). In the Pentatomoidea, these bacteria comprise various lineages of gammaproteobacteria which are primarily vertically transmitted via egg smearing, coprophagy or co-deposition in a jelly or capsule (16–19). Certain lineages of Lygaeoidea do not possess caeca and instead are associated with various bacteriome-inhabiting, vertically-transmitted gammaproteobacteria (20, 21), but all examined caeca-possessing members of the superfamilies Lygaeoidea (Berytidae, Blissidae in part, Rhyparochromidae, Pachygronthidae) and Coreoidea (Coreidae, Alydidae) are symbiotically partnered with members of *Burkholderia* (22). The mechanism of inoculation is from the environment (23–25) with young instar nymphs acquiring particular symbionts from soil (26), although some vertical transmission occurs in at least some lineages, such as Blissidae (27, 28). Caeca-possessing host species possess a symbiont-sorting organ at the junction of the third and fourth midgut section which blocks the passage of food and allows selective passage of particular bacteria (29).

In comparison to other hemipterans, herbivorous heteropterans tend to feed on relatively nutrient-rich parts of plants *e.g.* seeds, fruits or new buds. Despite this comparatively rich diet, hosts tend to face moderate to severe fitness deficits when deprived of beneficial symbionts (30–32). In Pentatomoidea, the function of symbionts is likely supplementation of amino acids as has been demonstrated for Urostylididae (19) and suspected for Pentatomidae and Plastapidae, where symbionts with reduced genomes retain capabilities to produce amino acids

104 (33, 34). In Parastrachiidae (Pentatomoidea), symbionts recycle uric acid during long periods of
105 diapause between availability of berries of their host plant (35). In Alydidae (Coreoidea), gene
106 knockout experiments have demonstrated some requirements for effective symbiotic
107 colonization, such as flagellar motility (29), and biosynthetic genes responsible for cell walls
108 (36), secondary messenger, c-di-GMP (37, 38), and a bacterial energy storage polymer (39).
109 These symbionts have been shown to enhance host innate immunity (40) or confer resistance
110 to fenitrothion (41) but the primary function of *Burkholderia* symbionts to hosts remains
111 unknown.

112 The Largidae together with Pyrrhocoridae, constitute the superfamily Pyrrhocoroidea which
113 contains ~520 extant species (42, 43). Members of Pyrrhocoridae are associated with a
114 characteristic microbiota primarily in the third section of the midgut which includes two obligate
115 actinobacterial species, *Coriobacterium glomerans* and *Gordonibacter* sp. passed on via
116 smearing of eggs (32, 44, 45). Only female Pyrrhocoridae are known to possess caeca though
117 they appear to be vestigial and do not contain bacteria (45). Elimination of symbionts by egg
118 sterilization results in host mortality and symbionts are thought to help their specialist hosts
119 subsist on seeds of Malvales through supplementation of B vitamins (46). In contrast to
120 Pyrrhocoridae, most largids appear to be generalist seed-feeding herbivores although some Old
121 World genera may be associated preferentially with seeds of Euphorbiaceae (13, 47, 48). Early
122 microscopic studies have demonstrated bacteria within caeca of members of Largidae
123 previously identified as *Bacillus* or *Lactobacillus (14, 49, 50)*. While conducting a survey of
124 bacterial diversity of Pyrrhocoridae, Sudarakaran et al. also discovered that representatives of
125 the largid genera, *Largus* and *Physopelta*, are associated with *Burkholderia (51)*. A focused
126 study on members of *Physopelta* found *Burkholderia* in midgut caeca from the plant-associated
127 beneficial and environmental (PBE) clade (52). All other known heteropteran *Burkholderia*
128 symbionts belong to a clade called the stinkbug-associated and beneficial environmental (SBE)

clade with the exception of one family, Blissidae, in which *Burkholderia* isolates may belong to the former two clades as well as a third *Burkholderia cepacia* complex (BCC) clade. (27, 28).

Alternate scenarios of the evolution of bacterial symbioses within this group of herbivorous true bugs may be invoked depending on the phylogenetic relationship of Pyrrhocoroidea with respect to related superfamilies. Evolutionary relationships of this superfamily have been controversial, with recent molecular phylogenies (listed in chronological order) placing Pyrrhocoroidea as sister to Coreoidea + Lygaeoidea (53), Alydidae (54), Coreoidea (55) or Lygaeoidea (56) and the most recent morphology-based cladistic analysis finding this superfamily as sister to Coreoidea (57). The monophyly of Largidae has also been questioned proposing that Pyrrhocoridae may have evolved from Largidae (58), or specifically, that the old World tribe Physopeltini may be more closely related to Pyrrhocoridae than to the New World Largini (43). Establishing the relationships of this superfamily will allow for a more accurate explanation of how different symbiotic complexes evolved among and within this group, and allow for testing concordance among host and symbiont phylogenies.

In the current study, we aim to identify symbionts across Pyrrhocoroidea with a focus on the family Largidae and determine the relations of these symbionts to other known heteropteran symbionts through an Illumina bacterial 16S amplicon survey and targeted full-length 16s rRNA gene sequencing. We also construct the first comprehensive phylogeny of Pyrrhocoroidea to accurately determine the evolutionary history of this group and how this branch on the tree of life relates to others with bacterial symbionts. We seek evidence to determine the method of transmission of Largidae symbionts to offspring through experimental rearing, fluorescence in situ hybridization (FISH) microscopy, culturing and investigation of patterns of symbiont and host phylogenies. We summarize our results and the current knowledge on evolution of bacterial symbiont complexes in pentatomomorphan Heteroptera.

Methods

Rearing of *Largus californicus*. Adult specimens were captured from Lytle Creek in San Bernardino National Forest, San Bernardino County, CA in June 2015 and were enclosed in plastic containers with soil substrate from UCR campus (not from field environment) and held at room temperature. Specimens were also provided with grapes, cabbage and water in a vial plugged with cotton. An egg batch (of approximately 70 eggs) was divided and approximately half of the eggs were washed with distilled water and DNA was extracted from washed eggs and wash with a QIAGEN DNeasy Blood and Tissue kit. The other half of the eggs were allowed to hatch, which they did after two weeks (on July 28th, 2015) and five first instar nymphs were pooled together and subjected to DNA extraction two days after hatching. The caeca-containing region of the midgut of an adult specimen was also dissected and DNA extracted. The presence of bacteria in DNA extracts was analyzed with PCR with universal and genus-specific bacterial 16S primers (Table S1).

Culturing. The caeca-containing region of the midgut of one adult specimen of *L. californicus* was dissected away from other gut tissue and macerated with an Eppendorf pestle for 2 minutes in 200 ul of PBS buffer (pH 7.4). An inoculating loop was used to spread the resulting cloudy homogenate liquid on a Luria-Bertani agar plate and incubated for three days at 37˚C. Several representatives of dominant colony morphotype were analyzed using colony PCR with universal bacterial primers (Table S2) and the resulting sequences were blasted against GenBank after cleaning and sequencing of PCR products to confirm the identity of the bacterial isolate.

Florescence in situ hybridization (FISH). Gut tissue from live specimens of *Largus californicus*, after anesthetization at -20˚C for three minutes, was dissected and stored separately in acetone as well as ~10 whole eggs from the unwashed half of the egg batch for whole mount microscopic preparations. We followed the protocol in (59), fixing tissue with Carnoy's solution overnight and staining gut tissue with DAPI for labelingl DNA, and two oligonucleotides probes: Cy-5-labeled universal bacterial probe (EUB-338; [60]) and Cy-3-

180 labeled *Burkholderia*-specific 16S oligonucleotide (Burk129; [22]) from biomers.net for Cy-5 and
181 Integrated DNA Technologies for Cy-3 (with HPLC purification) for staining of specific bacterial
182 symbionts. Confocal microscopy was conducted with a Leica TCS SP5 using 405, 543 and 655
183 nm lasers for visualization of DAPI, Cy-3 and Cy-5 respectively.

184 **Sampling and DNA extraction.** Individual specimens of all available of Pyrrhocoroidea
185 (from the worldwide-ethanol collection of Heteroptera of the Weirauch lab; details listed in Table
186 S1) along with two outgroup taxa (32 species total including 13 genera of Pyrrhocoroidea) were
187 surface sterilized with a 1% bleach solution for 2 minutes and rinsed with 100% ethanol before
188 removal of the abdomen from the thorax. Internal abdominal tissue was removed with sterile
189 forceps and the resulting material was homogenized with a bead beater for 3 minutes at 30 Hz
190 after addition of 100 µL of 0.1 mm glass beads and one 2.38 mm metal bead. Forceps were
191 washed with EtOH, flamed and sterilized with a 10% bleach solution before and after each
192 extraction. Each sample was incubated at 55˚ for 24 hours after addition of 10 µl of 800 U/ml
193 Proteinase K before proceeding with DNA extraction.

194 **Host phylogeny.** A total of ~4,800 bp of host DNA consisting of two mitochondrial protein-
195 encoding (COI and COII) and three ribosomal genes (16S, 18S, 28S) were amplified from DNA
196 extracts with primers listed in Table S2. PCR products were cleaned with Bioline Sureclean and
197 submitted to Macrogen for Sanger sequencing. Chromatographs were edited in Sequencher
198 v4.8 and aligned with MAFFT (E-INS-i strategy). A RAxML maximum likelihood phylogeny was
199 constructed after partitioning based on gene and codon position for protein encoding genes and
200 designation of the two included taxa of Pentatomoidea as an outgroup. Relevant sequences
201 from GenBank were included for a maximally comprehensive phylogeny of Pyrrhocoroidea (57
202 taxa and at least 20 genera) and all newly acquired sequences (accessions numbers
203 KX523359-KX523485) have been deposited on GenBank (Table S1). We noted that published
204 sequences from *Dindymus lanius* on GenBank clustered closely with *Antilochus* in contrast to
205 our own sequences from three species of *Dindymus* which included *Dindymus lanius* so we

206 excluded the former sequences from our analysis in case of possible misidentification. All
207 sampled representatives have been imaged with a Leica Microsystems imaging system and
208 databased using Arthropod Easy Capture (AEC) implemented in the Plant Bug PBI and these
209 data are publically available via ‘‘Heteroptera Species Pages”
210 (http://research.amnh.org/pbi/heteropteraspeciespage).
211 **Illumina 16S amplicons sequencing and analysis.** Gut tissue from a subset of 17 taxa
212 was subject to Illumina amplicon sequencing of a ~300 bp fragment of the bacterial 16S gene
213 using 799F and 1115R primers with barcodes for multiplexing, as in (61), intended to minimize
214 amplification of chloroplast DNA (62) and conducted in triplicate PCR (35 cycles annealing at
215 52˚) with 5 PRIME HotMasterMix. Triplicate PCR products were pooled and cleaned with
216 Ultraclean PCR cleanup kit (MoBio, Carlsbad, CA). Illumina adaptors were added to templates
217 via PCR on 1 µl of cleaned product with HPLC purified primers (15 cycles annealing at 58˚;
218 Table S2). Eighteen microliters of PCR product for each sample was normalized with a 96-well
219 SequalPrep™ Normalization Plate and 5 µl of each normalized sample was pooled and
220 assessed for quality with a 2100 Bioanalyser (Agilent, Santa Clara, CA) and the University of
221 California, Riverside Institute for Integrative Genome Biology.
222 Samples were multiplexed on an Illumina Miseq lane with a MiSeq Reagent Kit v3 with 2 x
223 300 paired-end sequencing. Paired-end read was assembled, trimmed (reads with ambiguous
224 bases or aberrant lengths removed) and demultiplexed using mothur v1.35.1 (63). Sequences
225 were aligned to the Silva v4 reference alignment and checked for chimeras with the UCHIME
226 algorithm using the most abundant sequences as a reference and assigned to OTUs at a 97%
227 identity level and OTUs were classified using a Bayesian classifier both implemented in mothur
228 (63). Jaccard and Bray Curtis community dissimilarity metrics along with weighted and
229 unweighted Unifrac distance matrices (64) were computed after rarefying dataset to 2,066 reads
230 per sample (the lowest number of reads in any sample after filtering) and clustered via PCoA

ordination plots visualized with Plotly (65). Raw data are available on the NCBI Sequence Read Archive (SRA) under accession number SRP078165.

For visualization of OTU abundance with a heat map, we removed all OTUs which together comprised <1% of the total dataset and any representatives of any OTU which constituted less than 0.5% of the total reads in a sample. Together, these represent slightly less than 2.5% of total reads after trimming. The most common representative of any OTU unclassified at a genus level was blasted against GenBank and manually curated to the lowest level possible based on highest scoring blastn to known organisms. Putative chimeric OTUs based on high blast hit identity to distantly related bacterial lineages were removed. OTUs with identical top blast hits were combined for visualization purposes. Read abundance was plotted on the log scale after division by 10 (equivalent to $\log_{10}(\text{reads}) - 1$) and plotted with the ggplot2 package in R.

Targeted full-length 16S PCR and phylogeny. Full length 16S sequences were retrieved from two transcriptomes of *Largus californicus* sequenced recently as a part of the Hemipteroid Tree of Life project. Based on these sequences, we modified existing primers to amplify full 16S sequences of *Burkholderia* with two sets of PCR with combination of PBE-specific primers and *Burkholderia*-specific primers (Primers in Table S2; newly acquired 16S sequences KX527603-KX527621 in Table S1). PCR products were cleaned with Bioline SureClean and sequences were processed as for host genes. A comprehensive set of 272 ribosomal 16S sequences for environmental and insect associated isolates of *Burkholderia* along with named *Burkholderia* species and 3 *Pandoraea* outgroups was downloaded from GenBank and a tree for the genus including full length 16S sequences of dominant *Burkholderia* obtained from Largidae samples via PCR and representative *Burkholderia* reads (300 bp) of less dominant OTUs from the Illumina dataset chosen from each sample in which they were present was constructed (Fig. S2). Taxa were pruned from this tree using Mesquite v3.04 (66) for easier visualization, retaining close relatives of newly sequenced bacterial 16S genes, representatives of other

256 insect-associated lineages and named species and the resulting dataset was realigned and
257 phylogeny reconstructed (Fig. 3).
258 A cophylogenetic analysis of *Burkholderia* and host phylogenies was performed with
259 TreeMap 3 (67) and Parafit (68). A maximally reconciled set branching pattern of host and
260 symbiont phylogenies was visualized with TreeMap 3. Patristic distances were calculated with
261 the cophenetic function in the ape package in R and the resulting phylogenetic relationships
262 were tested for correlation of bacterial and host phylogenies via the statistical test implemented
263 in Parafit with 10,000 permutations.
264 **Results**
265 **Gut morphology, rearing, culturing.** As has been previously described for congeners (Fig.
266 1B inset), the gut morphology of *Largus californicus* (adult pictured Fig. 1A) consists of five
267 morphologically distinct midgut sections (Fig. 1B). The third section is the largest and most
268 voluminous and is followed by a constricted region (Fig. 1G), homologous to the symbiont
269 sorting organs described in other Heteroptera (29). Caeca are relatively short and numerous in
270 the fifth midgut section consisting of two rows of tubes (Fig. 1B,C,F) and tend to be closely
271 associated with the distal part of the Malpighian tubules *in situ.* FISH microscopy highlights a
272 large density of bacteria in the caeca (Fig. 1D), which were stained with a genus-specific probe
273 for *Burkholderia* (Fig. 1E). Egg batches did not contain any prominent co-deposited substance
274 (Fig. 1H), however first instar nymphs did probe remains of hatched eggs (Fig. 1I, J).
275 Sequencing of PCR products amplified with general bacterial primers (Table S2) on DNA
276 extracts of the isolated caeca-containing region (Fig. 1C) produced a clean chromatograph
277 sequence matching *Burkholderia* (Table S1). DNA extracts of unwashed and washed eggs, first
278 instar nymphs and the wash from washed eggs produced no product when assayed with PCR
279 with *Burkholderia*-specific primers. Both sets of eggs and first instar nymphs did produce a band
280 when assayed with general bacterial primers which when sequenced produced a clean
281 chromatograph matching *Rickettsia* (Table S1). The *Burkholderia* symbiont was apparently

282 easily cultured on LB plates as all sequenced representatives of the dominant small yellow
283 colonies after plating of homogenized caeca were identical in sequence to each other and 100%
284 identical to the PBE-clade *Burkholderia australis,* isolated from sugarcane roots, JQ994113.1
285 [69]).

286 **Phylogenetic relationships within and among Pyrrhocoroidea.** Our phylogeny
287 reconstructed Pyrrhocoroidea as sister to Coreoidea + Lygaeoidea (Fig. 2; Fig S1) with a
288 bootstrap support value >85%. We recover, with moderate support, a monophyletic Largidae
289 with a sister group relationship between the two subfamilies Physopeltinae and Larginae. The
290 two predatory genera, *Antilochus* which are specialist predators of other Pyrrhocoridae (70), and
291 *Dindymus* which possess variable trophic strategies though some are specialist predators of
292 mollusks (71), do not appear to be most closely related to each other. In the unidentified female
293 pyrrhocorid specimen L54, which may belong to the genus *Ectatops* or *Saldoides*, we observed
294 caeca when dissecting gut tissue that were well-developed and comparable to those seen in
295 *Largus*.

296 **Bacterial associates of Pyrrhocoroidea.** After quality control, we recovered a total of
297 319,021 paired-end Illumina reads (average of 18,766 per sample). We found a highly prevalent
298 association of *Burkholderia* in gut extracts of all Largidae except for one specimen of
299 *Stenomacra tungurahuana* (L40). We also recovered a previously described Pyrrhocoridae-
300 associated *Clostridium* strain exclusively in each of our Pyrrhocoridae samples, although at a
301 below 0.5% level for a *Dysdercus* species from Australia L23 and *Dindymus lanius* L47 (OTUs
302 6, 10, 19, 23 in Table S3; File S1). The presence of *Gordonibacter sp.* (OTUs 11, 28) was also
303 observed exclusively in Pyrrhocoridae except for *Dindymus lanius* and only represented by two
304 reads out of nearly 20,000 in *Di. pulcher* (below 0.5% of sample reads for *Probergrothius* and
305 the *Dysdercus* species from Australia). We observed OTUs assigned to *Coriobacterium* only in
306 our two sampled *Dysdercus* species at a sub 0.5% level (OTU 64 in Table S3).

307 The principal coordinate analysis plot displayed in Fig. 2 is that of a subsampled distance
308 matrix based on abundance-weighted UniFrac distances (64) of OTUs and explains >45% of
309 the variance in the data with the two plotted axes. Principal coordinate analysis of bacterial
310 communities based on other distance measurements results in separate clusters of Largidae,
311 Pyrrhocoridae and Lygaeinae in all but Bray-Curtis community dissimilarity matrices in which
312 *Stenomacra tungurahuana* (L40) and *Stenomacra marginella* (L37) cluster separately from other
313 Largidae (Fig S2.) near the Lygaeinae and Pyrrhocoridae samples. Two *Dindymus* species
314 often cluster separately from most herbivorous Pyrrhocoridae sometimes also along with
315 *Dysdercus* sp. L23 (Fig. 2; Fig. S2).
316 **Phylogenetics of *Burkholderia* associates.** A phylogeny of *Burkholderia* symbionts based
317 on full length 16S sequences (when available) was not concordant with host phylogenies (Fig.
318 3; top). There was no evidence of codiversification (Parafit: p-value = 0.5129). Five species of
319 Largidae from geographically distant areas (Florida, Colombia, Argentina, Mexico and Costa
320 Rica) harbored the same or very similar strains of *Burkholderia* which closely matched 16S
321 sequences of PBE clade *Burkholderia*, including those of curated elite commercial inoculants
322 used in agriculture for nodulation of legume crops in Brazil, housed at the SEMIA *Rhizobium*
323 Culture Collection (e.g., 100% identical to *Burkholderia* sp. SEMIA 6385 / SEMIA 6382 isolated
324 from *Piptadenia gonoacantha* / *Mimosa caesalpiniifolia* roots, FJ025136.1 / AY904775.1 [72]).
325 A phylogeny constructed with all newly sequenced *Burkholderia* (orange and a subset of
326 those in red) and relatives present on GenBank shows that all Largidae-associated strains of
327 *Burkholderia* belong to the described PBE group (Fig. 3). Insect associates are closely related
328 to species that are known nodulating bacteria of legumes and particularly of Mimosoideae
329 (*Burkholderia diazotrophica*, *B. caribensis*, *B. phymatum*, *B. tuberum;* [73]) or other plant-
330 associated nitrogen fixing species (*B. tropica*, *B. heleia;* [74]). All other non-Largidae
331 heteropteran-associated *Burkholderia* fall within the SBE clade with the exception of a number
332 of symbiont strains isolated from Blissidae (marked in green on Fig. 3). A tree with more

comprehensive sampling of *Burkholderia* representatives including environmental isolates is shown in Fig. S3.

Discussion

Associated bacterial communities of Pyrrhocoroidea. We confirm an association of Largidae species from at least six genera with *Burkholderia* strains, specifically strains from plant-associated PBE clade, though there was also a single exception to this pattern, sample L40. Sample L40 (*Stenomacra tungurahuana*) instead possessed a high number of reads from *Rickettsia*. The *Dysdercus* specimen, L23, also had an unusual bacterial community profile with a high concentration of a species of Acetobacteraceae previously characterized from bees and also possessing a significant number reads from the intracellular pathogen, *Bartonella.* Both samples may be unrepresentative due to infection with another bacterium or could represent samples that suffered from decomposition or degradation of DNA after inadequate preservation in ethanol. Ethanol preservation has been shown to have an effect on sequenced bacterial diversity in other insects (75, 76). However, even these two samples cluster with other related members in most PCoA plots based on distance matrices except for that of Bray-Curtis distances. While L40 completely lacks any *Burkholderia* OTU, the *Burkholderia* strain present in sample L37 has less than 97% identity to 16S rRNA sequences from other largid-associated *Burkholderia* strains, thus it is not surprising that this similarity in genus composition would not be represented by phylogenetically-independent distance metrics. Members of two largid genera that we were not able to sample, *Macrocheraia grandis* and *Iphita limbata,* have been shown to possess caeca filled with rod-shaped bacteria potentially consistent with their identity as *Burkholderia (49, 50)*, though after culturing, isolates displayed differences with *Burkholderia,* such as being Gram-positive or spore-forming.

Although it was not our primary goal, we recovered a slightly more restricted distribution of some previously described Pyrrhocoridae symbionts than has been previously described. The most notable difference is the presence of *Coriobacterium* only in the *Dysdercus* species in our

359 sampled representatives. Previously, *Coriobacterium* has been shown to make up high
360 proportions of bacterial communities in members of the genera *Scantius*, *Pyrrhocoris* and
361 *Dysdercus* (only the latter sampled in the current study) but represented <5% of reads also in
362 *Antilochus*, *Probergothius* and *Dindymus*. Similarly, *Gordonibacter* has been shown to be
363 present at low levels in many genera of Pyrrhocoridae including *Dindymus,* whereas we
364 retrieved either extremely low or no numbers of *Gordonibacter* reads present in the
365 representatives of *Dindymus* we sampled. In both of these predatory representatives, we
366 instead found a microbiome dominated by a *Citrobacter*-type Enterobacteriaceae and a
367 *Lactococcus* strain both of which appear to be common insect gut inhabitants in Largidae and
368 perhaps other insects. The retention of any strict association of a predator with a bacterium,
369 which we observe with the Pyrrhocoridae-associated *Clostridium* species is surprising as the
370 evolution of a predatory life strategy in other Pentatomomorpha (such as the Asopinae
371 [Pentatomidae] and Geocorinae [Geocoridae]) seems to negate dependence on any particular
372 bacterium. This may reflect a relatively recent evolution of this predatory trophic strategy. The
373 observation of well-developed caeca in one female specimen (L54) of Pyrrhocoridae should be
374 investigated with fresh or acetone preserved specimens, if possible. Although vestigial caeca
375 have been noted in *Dysdercus*, *Antilochus*, *Probergothius* and *Pyrrhocoris (51)*, it may be
376 possible that at least some members of the well supported clade containing *Melamphaus* +
377 *Dermatinus* + *Ectatops* + *Euscopus* retain functional caeca.

378 **Transmission method of *Burkholderia*.** Although it has not yet been shown
379 experimentally, all available evidence supports environmental reacquisition of *Burkholderia* by
380 new generations of species of Largidae. The lack of evidence of *Burkholderia* in or on eggs or in
381 lab-reared first instar nymphs as well as the lack of any sort of cophylogenetic signal suggest
382 horizontal transmission. Booth (77) noted that nearly all first instar nymphs of *Largus*
383 *californicus* reared in a laboratory setting died before molting, though eggs reared in the field
384 and field-caught first instars brought to the lab were both viable to adulthood. This suggests that

385 the first instar stage is when symbionts are acquired from the environment unlike in Alydidae,
386 where symbionts are acquired during the second instar (26). Also suggestive of a horizontal
387 transmission strategy is that in all other known cases of Heteroptera associated with
388 *Burkholderia,* the symbiosis is overwhelmingly acquired from the environment (although with up
389 to 30% vertical transmission in Blissidae; [(27)]).
390 Environmentally-acquired obligate symbiosis with bacteria is common in some
391 microhabitats such as in soil or the ocean where it occurs in partnerships between nitrogen-
392 fixing bacteria and legumes or bioluminescent bacteria and squids (78). However, this method
393 of transmission for obligate symbionts is very rare in insects and association with *Burkholderia*
394 in Heteroptera may be one of the oldest stably maintained symbiosis in insects, as the node
395 containing all *Burkholderia*-associated members has been dated to the early Cretaceous era
396 ~130 mya (54). Such associations have the potential to evolve towards pathogenicity of
397 symbionts as horizontal acquisition can select for pathogenic or cheating phenotypes (79) and
398 other similar systems show policing by hosts to punish symbionts which do not participate (80).
399 The gut crypts of Largidae and other heteropterans may provide a much-needed mechanism to
400 police symbionts as individuals crypts could be modulated individually, though this has not yet
401 been shown.
402 **Evolutionary transitions of symbiont complexes.** The exclusive association of Largidae
403 with PBE clade *Burkholderia* unlike all other *Burkholderia*-associated Heteroptera is unlikely to
404 be incidental. The Pyrrhocoroidea have now been shown to be one of two lineages of
405 *Burkholderia*-associated Heteroptera, with the Largidae the earliest diverging single family
406 which still retains such a symbiosis. It is likely that the ancestor of all *Burkholderia*-associated
407 Heteroptera was also associated with *Burkholderia* but it is not as clear which clade of
408 *Burkholderia* or if there was any specificity at all. Perhaps this ancestor had the same lack of
409 specificity towards *Burkholderia* that members of Blissidae seem to display when associating
410 with different strains of *Burkholderia* spanning the three described clades. As lineages diverged,

411 different mechanisms of specificity may have evolved that selected for particular subclades of
412 *Burkholderia* with increased evolutionary benefits for that lineage depending on its biology.
413 Many members of the PBE clade of *Burkholderia* are rhizobia that form nodules on legumes and
414 these symbiotic properties tend to be tied to symbiotic genomic islands that contain genes for
415 nodulation and nitrogen fixation that are easily horizontally transferred (81). Other PBE
416 *Burkholderia* are also nitrogen-fixing and associated with plants but their genomic architecture is
417 not yet known (74, 82). Nitrogen fixation may be the role of these symbiotic bacteria to their
418 largid hosts as has been suggested of other insect symbionts from lineages of bacteria that
419 undergo nitrogen-fixing symbioses with plants, such as in herbivorous ants (83). It is possible
420 that an increased efficiency of nitrogen fixation (e.g., of elite commercial inoculants) is the
421 mechanism that selects for similar strains in geographically disparate Largidae. These bacteria
422 can presumably live within either of two symbiotic systems or as a free-living bacterium within
423 the environment and must face different selective pressures in each of these states.
424 Strict vertical transmission of symbionts, especially intracellular symbionts, can lead to rapid
425 degradation of symbiont genomes and may be non-beneficial in some respects, trapping hosts
426 and symbiont in “an evolutionary rabbit hole” (6). However, the ancestor of Heteroptera
427 achieved independence from intracellular symbionts, but after re-evolving herbivory, some
428 members developed new associations with extracellular bacteria. In the Pentatomoidea, these
429 symbionts are primarily vertically transmitted, although there is accruing evidence suggesting
430 that vertically transmitted symbionts can be supplanted by environmental bacteria, thus
431 replacing any possibly degraded symbiont genome (84). For *Burkholderia*-associated members,
432 symbionts are primarily acquired from the environment every generation thus avoiding any
433 transmission bottleneck that could lead to genome degradation (22). However, within the
434 Pyrrhocoridae, there has been a transition to vertically transmitted extracellular symbionts, and
435 in some lineages of Lygaeoidea, several reversions to vertically-transmitted intracellular
436 symbioses, the ancestral condition of non-heteropteran Hemiptera (Fig. 4). It has been

437 suggested or implied that these transitions to vertically transmitted symbionts has been driven
438 by evolutionary pressures tied to host plant specificity (20, 51). But host plant specialists are
439 also widespread in species that retain environmentally-acquired symbionts (13, 85). Perhaps,
440 there are more specific evolutionary pressures that lead to this more stable but eventually
441 degenerate form of symbiosis.

442 There remains outstanding questions with respect to the mechanism and evolutionary
443 benefit of specificity to different clades of *Burkholderia* in different heteropterans. Generalist vs.
444 specialist evolutionary pressures may play a role. Many families remain to be characterized
445 (Fig. 4) especially using modern molecular methods and other families demonstrate such a
446 diversity of symbiotic complexes in known members that many more members should be
447 examined. Particularly, the identity of any symbiont present in members of the two small families
448 sister to all other Coreoidea, Hyocephalidae and Stenocephalidae (the former, associated with
449 *Acacia* and *Eucalyptus* seeds, and the latter, specialists on seeds of Euphorbiaceae [42]), may
450 contribute to our understanding of how the evolution of strict associations with clade-specific
451 *Burkholderia* arose. While the genome of a SBE clade *Burkholderia* isolated from Alydidae has
452 been published (86), it has so far not provided much insight into the probable function of this
453 symbiont. Unlike strict symbionts, these bacteria have large genomes and thus it is more difficult
454 to discern function from the retention of gene sequences alone. The sequencing of PBE clade
455 *Burkholderia* symbionts from Largidae, which in at least one case, is easily cultured, may
456 provide some insight into shared genes between these symbiotic strains as well as differences
457 among them.

458 **Acknowledgments**

459 We would like to thank Kaleigh Russell for assistance in culturing bacteria, Alex Knyshov
460 for guidance with FISH microscopy and Paul Masonick for help with collecting *Largus*
461 *californicus*.

462 **Funding information**

This study was funded by the Dr. Mir S. Mulla and Lelia Mulla Endowed Scholarship Fund and the National Science Foundation Graduate Research Fellowship awarded to E. R. L. Gordon and initial complement funds awarded to to Q. S. McFrederick by University of California, Riverside.

Supplementary data:

Table S1. Specimen information and accession numbers.

Table S2. PCR primers and conditions

Table S3. OTU count table

File S1. Representative OTU fasta file

Figure S1. Phylogeny of Pyrrhocoroidea with all bootstrap supports

Figure S2. PCoA plots based on other distance metrics.

Figure S3. Untrimmed *Burkholderia* phylogeny

References

1. **Cryan JR**, **Urban JM**. 2012. Higher-level phylogeny of the insect order Hemiptera: is Auchenorrhyncha really paraphyletic? Syst Entomol **37**:7–21.

2. **Buchner P**. 1965. Endosymbiosis of animals with plant microorganisms. Interscience Publishers, New York, N.Y.

3. **Moran NA**, **Tran P**, **Gerardo NM**. 2005. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. Appl Environ Microbiol **71**:8802–8810.

4. **Moran NA**, **McCutcheon JP**, **Nakabachi A**. 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet **42**:165–190.

5. **Kuechler SM**, **Gibbs G**, **Burckhardt D**, **Dettner K**, **Hartung V**. 2013. Diversity of bacterial endosymbionts and bacteria-host co-evolution in Gondwanan relict moss bugs (Hemiptera: Coleorrhyncha: Peloridiidae). Environ Microbiol **15**:2031–2042.

6. **Bennett GM**, **Moran NA**. 2015. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. Proc Natl Acad Sci U S A **112**:10169–10176.

7. **Wilson ACC**, **Duncan RP**. 2015. Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. Proc Natl Acad Sci U S A **112**:10255–10261.

8. **Bennett GM**, **McCutcheon JP**, **McDonald BR**, **Moran NA**. 2016. Lineage-specific patterns of genome deterioration in obligate symbionts of sharpshooter leafhoppers. Genome Biol Evol **8**:296–301.

9. **McCutcheon JP**, **Moran NA**. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol Evol **2**:708–718.

10. **Koga R**, **Moran NA**. 2014. Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. ISME J **8**:1237–1246.

11. **Van Leuven JT**, **Meister RC**, **Simon C**, **McCutcheon JP**. 2014. Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. Cell **158**:1270–1280.

12. **Walker AA**, **Weirauch C**, **Fry BG**, **King GF**. 2016. Venoms of heteropteran insects: a treasure trove of diverse pharmacological toolkits. Toxins **8**:43.

13. **Schaefer CW**, **Panizzi AR**. 2000. Heteroptera of economic importance. CRC Press.

14. **Glasgow H**. 1914. The gastric caeca and the caecal bacteria of the Heteroptera. Biol Bull.

15. **Miyamoto S**. 1961. Comparative morphology of alimentary organs of Heteroptera, with the phylogenetic consideration. Sieboldia **2**:197–259.

16. **Hosokawa T**, **Kikuchi Y**, **Meng XY**, **Fukatsu T**. 2005. The making of symbiont capsule in the plataspid stinkbug *Megacopta punctatissima*. FEMS Microbiol Ecol **54**:471–477.

17. **Prado SS**, **Rubinoff D**, **Almeida RPP**. 2006. Vertical transmission of a pentatomid caeca-associated symbiont. Ann Entomol Soc Am **99**:577–585.

18. **Hosokawa T**, **Hironaka M**, **Mukai H**, **Inadomi K**, **Suzuki N**, **Fukatsu T**. 2012. Mothers never miss the moment: a fine-tuned mechanism for vertical symbiont transmission in a subsocial insect. Anim Behav **83**:293–300.

19. **Kaiwa N**, **Hosokawa T**, **Nikoh N**, **Tanahashi M**, **Moriyama M**, **Meng X-Y**, **Maeda T**,

Yamaguchi K, **Shigenobu S**, **Ito M**, **Fukatsu T**. 2014. Symbiont-supplemented maternal investment underpinning host's ecological adaptation. Curr Biol **24**:2465–2470.

20. **Kuechler SM**, **Renz P**, **Dettner K**, **Kehl S**. 2012. Diversity of symbiotic organs and bacterial endosymbionts of lygaeoid bugs of the families Blissidae and Lygaeidae (Hemiptera: Heteroptera: Lygaeoidea). Appl Environ Microbiol **78**:2648–2659.

21. **Matsuura Y**, **Kikuchi Y**, **Hosokawa T**, **Koga R**, **Meng X-Y**, **Kamagata Y**, **Nikoh N**, **Fukatsu T**. 2012. Evolution of symbiotic organs and endosymbionts in lygaeid stinkbugs. ISME J **6**:397–409.

22. **Kikuchi Y**, **Hosokawa T**, **Fukatsu T**. 2011. An ancient but promiscuous host-symbiont association between *Burkholderia* gut symbionts and their heteropteran hosts. ISME J **5**:446–460.

23. **Olivier-Espejel S**, **Sabree ZL**, **Noge K**, **Becerra JX**. 2011. Gut microbiota in nymph and adults of the giant mesquite bug (*Thasus neocalifornicus*) (Heteroptera: Coreidae) is dominated by *Burkholderia* acquired de novo every generation. Environ Entomol **40**:1102–1110.

24. **Kikuchi Y**, **Yumoto I**. 2013. Efficient colonization of the bean bug *Riptortus pedestris* by an environmentally transmitted *Burkholderia* symbiont. Appl Environ Microbiol **79**:2088–2091.

25. **Garcia JR**, **Laughton AM**, **Malik Z**, **Parker BJ**, **Trincot C**, **Chiang SSL**, **Chung E**, **Gerardo NM**. 2014. Partner associations across sympatric broad-headed bug species and their environmentally acquired bacterial symbionts. Mol Ecol **23**:1333–1347.

26. **Kikuchi Y**, **Hosokawa T**, **Fukatsu T**. 2011. Specific developmental window for establishment of an insect-microbe gut symbiosis. Appl Environ Microbiol **77**:4075–4081.

27. **Itoh H**, **Aita M**, **Nagayama A**, **Meng X-Y**, **Kamagata Y**, **Navarro R**, **Hori T**, **Ohgiya S**, **Kikuchi Y**. 2014. Evidence of environmental and vertical transmission of *Burkholderia* symbionts in the oriental chinch bug, *Cavelerius saccharivorus* (Heteroptera: Blissidae). Appl Environ Microbiol **80**:5974–5983.

28. **Boucias DG**, **Garcia-Maruniak A**, **Cherry R**, **Lu H**, **Maruniak JE**, **Lietze V-U**. 2012. Detection and characterization of bacterial symbionts in the Heteropteran, *Blissus insularis*. FEMS Microbiol Ecol **82**:629–641.

29. **Ohbayashi T**, **Takeshita K**, **Kitagawa W**, **Nikoh N**, **Koga R**, **Meng X-Y**, **Tago K**, **Hori T**, **Hayatsu M**, **Asano K**, **Kamagata Y**, **Lee BL**, **Fukatsu T**, **Kikuchi Y**. 2015. Insect's intestinal organ for symbiont sorting. Proc Natl Acad Sci U S A **112**:E5179–88.

30. **Kikuchi Y**, **Hosokawa T**, **Fukatsu T**. 2007. Insect-microbe mutualism without vertical transmission: a stinkbug acquires a beneficial gut symbiont from the environment every generation. Appl Environ Microbiol **73**:4308–4316.

31. **Kikuchi Y**, **Hosokawa T**, **Nikoh N**, **Fukatsu T**. 2011. Gut symbiotic bacteria in the cabbage bugs *Eurydema rugosa* and *Eurydema dominulus* (Heteroptera: Pentatomidae). Appl Entomol Zool **47**:1–8.

32. **Salem H**, **Kreutzer E**, **Sudakaran S**, **Kaltenpoth M**. 2013. Actinobacteria as essential symbionts in firebugs and cotton stainers (Hemiptera, Pyrrhocoridae). Environ Microbiol **15**:1956–1968.

33. **Nikoh N**, **Hosokawa T**, **Oshima K**, **Hattori M**, **Fukatsu T**. 2011. Reductive evolution of bacterial genome in insect gut environment. Genome Biol Evol **3**:702–714.

34. **Kenyon LJ**, **Meulia T**, **Sabree ZL**. 2015. Habitat Visualization and Genomic Analysis of "Candidatus Pantoea carbekii," the Primary Symbiont of the Brown Marmorated Stink Bug.

578 Genome Biol Evol **7**:620–635.

579 35. **Kashima T**, **Nakamura T**, **Tojo S**. 2006. Uric acid recycling in the shield bug, *Parastrachia*
580 *japonensis* (Hemiptera: Parastrachiidae), during diapause. J Insect Physiol **52**:816–825.

581 36. **Kim JK**, **Lee HJ**, **Kikuchi Y**, **Kitagawa W**, **Nikoh N**, **Fukatsu T**, **Lee BL**. 2013. Bacterial
582 cell wall synthesis gene uppP is required for *Burkholderia* colonization of the Stinkbug Gut.
583 Appl Environ Microbiol **79**:4879–4886.

584 37. **Kim JK**, **Jang HA**, **Won YJ**, **Kikuchi Y**, **Han SH**, **Kim C-H**, **Nikoh N**, **Fukatsu T**, **Lee BL**.
585 2014. Purine biosynthesis-deficient *Burkholderia* mutants are incapable of symbiotic
586 accommodation in the stinkbug. ISME J **8**:552–563.

587 38. **Kim JK**, **Kwon JY**, **Kim SK**, **Han SH**, **Won YJ**, **Lee JH**, **Kim C-H**, **Fukatsu T**, **Lee BL**.
588 2014. Purine biosynthesis, biofilm formation, and persistence of an insect-microbe gut
589 symbiosis. Appl Environ Microbiol **80**:4374–4382.

590 39. **Kim JK**, **Won YJ**, **Nikoh N**, **Nakayama H**, **Han SH**, **Kikuchi Y**, **Rhee YH**, **Park HY**, **Kwon**
591 **JY**, **Kurokawa K**, **Dohmae N**, **Fukatsu T**, **Lee BL**. 2013. Polyester synthesis genes
592 associated with stress resistance are involved in an insect-bacterium symbiosis. Proc Natl
593 Acad Sci U S A **110**:E2381–9.

594 40. **Kim JK**, **Lee JB**, **Huh YR**, **Jang HA**, **Kim C-H**, **Yoo JW**, **Lee BL**. 2015. *Burkholderia* gut
595 symbionts enhance the innate immunity of host *Riptortus pedestris*. Dev Comp Immunol
596 **53**:265–269.

597 41. **Kikuchi Y**, **Hayatsu M**, **Hosokawa T**, **Nagayama A**, **Tago K**, **Fukatsu T**. 2012. Symbiont-
598 mediated insecticide resistance. Proc Natl Acad Sci U S A **109**:8618–8622.

599 42. **Schuh RT**, **Slater JA**. 1995. True bugs of the world. Comstock, Ithaca, New York.

43. **Stehlík JL**. 2013. Review and reclassification of the Old World genus *Physopelta* (Hemiptera: Heteroptera: Largidae). Acta Entomologica Musei Natioalis Pragae **53**.

44. **Kaltenpoth M**, **Winter SA**, **Kleinhammer A**. 2009. Localization and transmission route of *Coriobacterium glomerans*, the endosymbiont of pyrrhocorid bugs. FEMS Microbiol Ecol **69**:373–383.

45. **Sudakaran S**, **Salem H**, **Kost C**, **Kaltenpoth M**. 2012. Geographical and ecological stability of the symbiotic mid-gut microbiota in European firebugs, *Pyrrhocoris apterus* (Hemiptera, Pyrrhocoridae). Mol Ecol **21**:6134–6151.

46. **Salem H**, **Bauer E**, **Strauss AS**, **Vogel H**, **Marz M**, **Kaltenpoth M**. 2014. Vitamin supplementation by gut symbionts ensures metabolic homeostasis in an insect host. Proc Biol Sci **281**:20141838.

47. **Dhiman SC**, **Bhardwaj MMH**. 2008. Host and pest relationship, host specificity and orientation towards food of *Physopelta schlanbuschii* (Heteroptera: Pyrrhocoroidea: Largidae). Annals of Plant Protection Sciences **16**:373–376.

48. **Dhiman SC**, **Gujral K**. 2002. Biology of *Iphita limbata* Stal, a pest of forest tree *Trewia nudiflora* Linn. Indian For **128**:54–64.

49. **Chattopadhyay AK**, **Choudhuri DK**. 1981. Studies on the endocellular procaryotes in *Lohita grandis* (Gray) (Hemiptera: Pyrrhocoridae): II. cultivation of presumptive endocellular procaryotes. Appl Entomol Zool **16**:162–164.

50. **Chattopadhyay AK**, **Choudhuri DK**. 1984. Studies on the mycetomal procaryotes in *Iphita limbata* (Stal) (Pyrrhocoridae: Hemiptera: Insecta). Proc Indian Natl Sci Acad.

51. **Sudakaran S**, **Retz F**, **Kikuchi Y**, **Kost C**, **Kaltenpoth M**. 2015. Evolutionary transition in

622 symbiotic syndromes enabled diversification of phytophagous insects on an imbalanced
623 diet. ISME J **9**:2587–2604.

624 52. **Takeshita K**, **Matsuura Y**, **Itoh H**, **Navarro R**, **Hori T**, **Sone T**, **Kamagata Y**, **Mergaert P**,
625 **Kikuchi Y**. 2015. *Burkholderia* of plant-beneficial group are symbiotically associated with
626 bordered plant bugs (Heteroptera: Pyrrhocoroidea: Largidae). Microbes Environ **30**:321–
627 329.

628 53. **Hua J**, **Li M**, **Dong P**, **Cui Y**, **Xie Q**, **Bu W**. 2008. Comparative and phylogenomic studies
629 on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera).
630 BMC Genomics **9**:610.

631 54. **Li M**, **Tian Y**, **Zhao Y**, **Bu W**. 2012. Higher level phylogeny and the first divergence time
632 estimation of Heteroptera (Insecta: Hemiptera) based on multiple genes. PLoS One
633 **7**:e32152.

634 55. **Yuan M-L**, **Zhang Q-L**, **Guo Z-L**, **Wang J**, **Shen Y-Y**. 2015. Comparative mitogenomic
635 analysis of the superfamily Pentatomoidea (Insecta: Hemiptera: Heteroptera) and
636 phylogenetic implications. BMC Genomics **16**:460.

637 56. **Wang Y-H**, **Cui Y**, **Rédei D**, **Baňař P**, **Xie Q**, **Štys P**, **Damgaard J**, **Chen P-P**, **Yi W-B**,
638 **Wang Y**, **Dang K**, **Li C-R**, **Bu W-J**. 2015. Phylogenetic divergences of the true bugs
639 (Insecta: Hemiptera: Heteroptera), with emphasis on the aquatic lineages: the last piece of
640 the aquatic insect jigsaw originated in the Late Permian/Early Triassic. Cladistics.

641 57. **Henry TJ**. 1997. Phylogenetic analysis of family groups within the infraorder
642 Pentatomomorpha (Hemiptera: Heteroptera), with emphasis on the Lygaeoidea. Ann
643 Entomol Soc Am **90**:275–301.

644 58. **Schaefer CW**. 1964. The morphology and higher classification of the coreoidea

(Hemiptera-Heteroptera): Parts I and II. Ann Entomol Soc Am **57**:670–684.

59. **Koga R**, **Tsuchida T**, **Fukatsu T**. 2009. Quenching autofluorescence of insect tissues for in situ detection of endosymbionts. Appl Entomol Zool **44**:281–291.

60. **Amann RI**, **Binder BJ**, **Olson RJ**, **Chisholm SW**, **Devereux R**, **Stahl DA**. 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol **56**:1919–1925.

61. **Kembel SW**, **O'Connor TK**, **Arnold HK**, **Hubbell SP**, **Wright SJ**, **Green JL**. 2014. Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. Proc Natl Acad Sci U S A **111**:13715–13720.

62. **McFrederick QS**, **Rehan SM**. 2016. Characterization of pollen and bacterial community composition in brood provisions of a small carpenter bee. Mol Ecol **25**:2302–2311.

63. **Schloss PD**, **Westcott SL**, **Ryabin T**, **Hall JR**, **Hartmann M**, **Hollister EB**, **Lesniewski RA**, **Oakley BB**, **Parks DH**, **Robinson CJ**, **Sahl JW**, **Stres B**, **Thallinger GG**, **Van Horn DJ**, **Weber CF**. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol **75**:7537–7541.

64. **Lozupone CA**, **Hamady M**, **Kelley ST**, **Knight R**. 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol **73**:1576–1585.

65. **Inc. PT**. 2015. Collaborative data science. Plotly Technologies Inc., Montréal, QC.

66. **Maddison W**, **Maddison D**. 2015. Mesquite: a modular system for evolutionary analysis. Version 3.04. Computer program and documentation distributed by the author.

667 67. **Charleston M**. 1995. TreeMap 3.

668 68. **Legendre P**, **Desdevises Y**, **Bazin E**. 2002. A statistical test for host-parasite coevolution.
669 Syst Biol **51**:217–234.

670 69. **Paungfoo-Lonhienne C**, **Lonhienne TGA**, **Yeoh YK**, **Webb RI**, **Lakshmanan P**, **Chan**
671 **CX**, **Lim P-E**, **Ragan MA**, **Schmidt S**, **Hugenholtz P**. 2014. A new species of *Burkholderia*
672 isolated from sugarcane roots promotes plant growth. Microb Biotechnol **7**:142–154.

673 70. **Ahmad I**, **Schaefer CW**. 1987. Food plant and feeding biology of the Pyrrhocoroidea
674 (Hemiptera). Phytophaga **1**:75–92.

675 71. **Jackson RR**, **Barrion A**. 2004. Heteropteran predation on terrestrial gastropods, p. 483–
676 496. *In* Barker, GM (ed.), Natural enemies of terrestrial molluscs. CABI Publishing.

677 72. **Binde DR**, **Menna P**, **Bangel EV**, **Barcellos FG**, **Hungria M**. 2009. rep-PCR fingerprinting
678 and taxonomy based on the sequencing of the 16S rRNA gene of 54 elite commercial
679 rhizobial strains. Appl Microbiol Biotechnol **83**:897–908.

680 73. **Peix A**, **Ramírez-Bahena MH**, **Velázquez E**, **Bedmar EJ**. 2015. Bacterial associations with
681 legumes. CRC Crit Rev Plant Sci **34**:17–42.

682 74. **Aizawa T**, **Ve NB**, **Nakajima M**, **Sunairi M**. 2010. *Burkholderia heleia* sp. nov., a nitrogen-
683 fixing bacterium isolated from an aquatic plant, Eleocharis dulcis, that grows in highly acidic
684 swamps in actual acid sulfate soil areas of Vietnam. Int J Syst Evol Microbiol **60**:1152–
685 1157.

686 75. **Sanders JG**, **Powell S**, **Kronauer DJC**, **Vasconcelos HL**, **Frederickson ME**, **Pierce NE**.
687 2014. Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes.
688 Mol Ecol **23**:1268–1283.

689 76. **Hammer TJ**, **Dickerson JC**, **Fierer N**. 2015. Evidence-based recommendations on storing
690 and handling specimens for analyses of insect microbiota. PeerJ **3**:e1190.

691 77. **Booth CL**. 1990. Biology of *Largus californicus* (Hemiptera: Largidae). Southwest Nat
692 **35**:15–22.

693 78. **Bright M**, **Bulgheresi S**. 2010. A complex journey: transmission of microbial symbionts.
694 Nat Rev Microbiol **8**:218–230.

695 79. **Sachs JL**, **Wilcox TP**. 2006. A shift to parasitism in the jellyfish symbiont *Symbiodinium*
696 *microadriaticum*. Proc Biol Sci **273**:425–429.

697 80. **Toby Kiers E**, **Rousseau RA**, **West SA**, **Ford Denison R**. 2003. Host sanctions and the
698 legume–rhizobium mutualism. Nature **425**:78–81.

699 81. **Moulin L**, **Klonowska A**, **Caroline B**, **Booth K**, **Vriezen JAC**, **Melkonian R**, **James EK**,
700 **Young JPW**, **Bena G**, **Hauser L**, **Land M**, **Kyrpides N**, **Bruce D**, **Chain P**, **Copeland A**,
701 **Pitluck S**, **Woyke T**, **Lizotte-Waniewski M**, **Bristow J**, **Riley M**. 2014. Complete genome
702 sequence of *Burkholderia phymatum* STM815(T), a broad host range and efficient nitrogen-
703 fixing symbiont of *Mimosa* species. Stand Genomic Sci **9**:763–774.

704 82. **Reis VM**, **Estrada-de los Santos P**, **Tenorio-Salgado S**, **Vogel J**, **Stoffels M**, **Guyon S**,
705 **Mavingui P**, **Baldani VLD**, **Schmid M**, **Baldani JI**, **Balandreau J**, **Hartmann A**,
706 **Caballero-Mellado J**. 2004. *Burkholderia tropica* sp. nov., a novel nitrogen-fixing, plant-
707 associated bacterium. Int J Syst Evol Microbiol **54**:2155–2162.

708 83. **Russell JA**, **Moreau CS**, **Goldman-Huertas B**, **Fujiwara M**, **Lohman DJ**, **Pierce NE**.
709 2009. Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. Proc
710 Natl Acad Sci U S A **106**:21236–21241.

711 84. **Hosokawa T**, **Ishii Y**, **Nikoh N**, **Fujie M**, **Satoh N**, **Fukatsu T**. 2016. Obligate bacterial
712 mutualists evolving from environmental bacteria in natural insect populations. Nature
713 Microbiology **1**:15011.

714 85. **Fernandes JAM**, **Mitchell PL**, **Livermore L**, **Nikunlassi M**. 2015. Leaf-footed bugs
715 (Coreidae), p. 549–605. *In* Panizzi, AR, Grazia, J (eds.), True bugs (Heteroptera) of the
716 Neotropics. Springer Netherlands.

717 86. **Shibata TF**, **Maeda T**, **Nikoh N**, **Yamaguchi K**, **Oshima K**, **Hattori M**, **Nishiyama T**,
718 **Hasebe M**, **Fukatsu T**, **Kikuchi Y**, **Shigenobu S**. 2013. Complete genome sequence of
719 *Burkholderia* sp. strain RPE64, bacterial symbiont of the bean bug *Riptortus pedestris*.
720 Genome Announc **1**.

721 87. **Tian X**, **Xie Q**, **Li M**, **Gao C**, **Cui Y**, **Xi L**, **Bu W**. 2011. Phylogeny of pentatomomorphan
722 bugs (Hemiptera-Heteroptera: Pentatomomorpha) based on six Hox gene fragments.
723 Zootaxa **2888**:57–68.

724 88. **Grazia J**, **Schuh RT**, **Wheeler WC**. 2008. Phylogenetic relationships of family groups in
725 Pentatomoidea based on morphology and DNA sequences (Insecta: Heteroptera).
726 Cladistics **24**:932–976.

727 89. **Péricart J**. 1998. Hémiptères lygaeidae euro-méditerranéens. Fédération Française des
728 Sociétés de Sciences Naturelles, Paris, France.

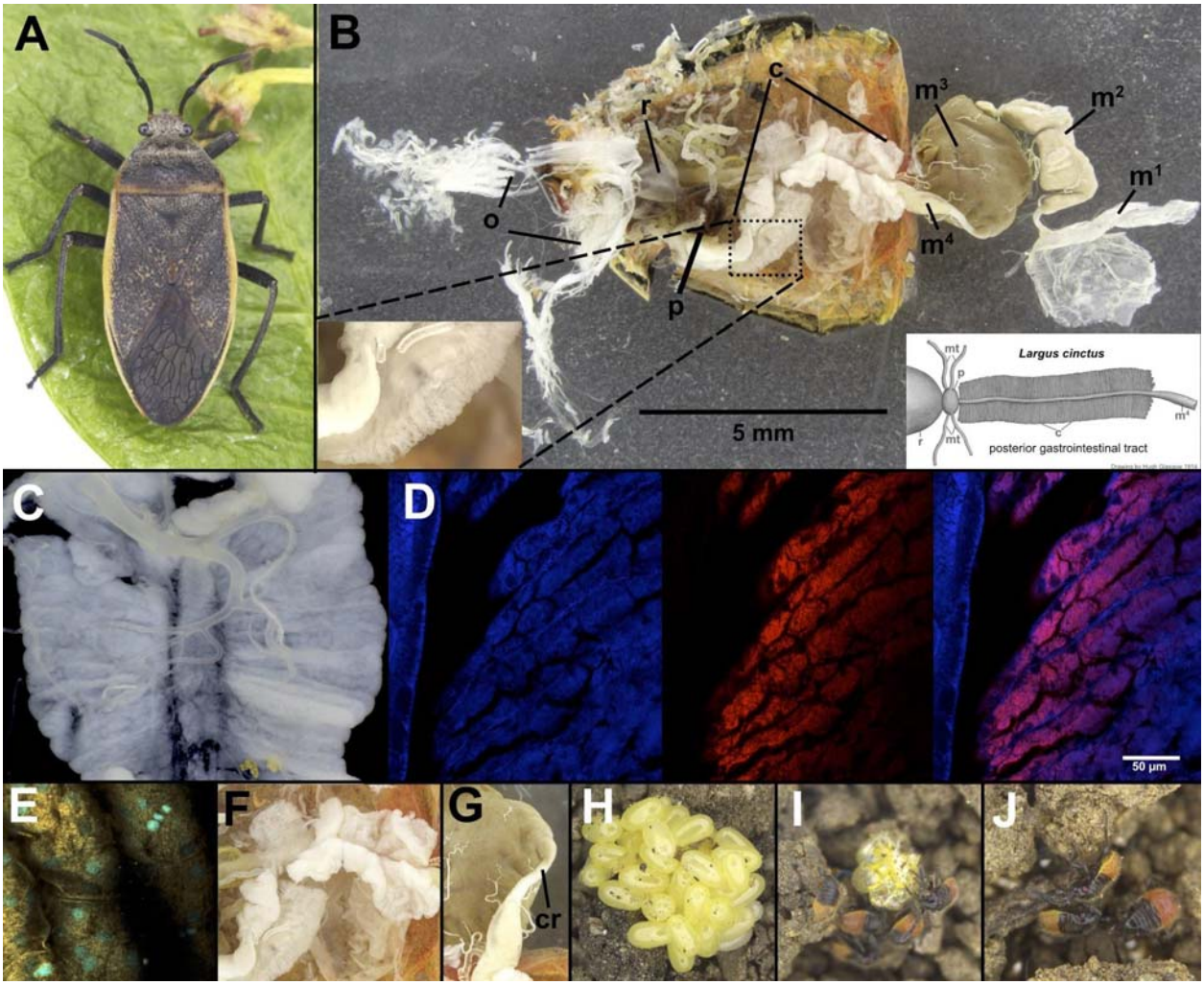
729
730
731
732
733

734

735

736

737 **Figures**



738

739 **Figure 1**. **A.** *Largus californicus* adult female. **B.** Gut morphology of *L. californicus*; **Inset:** Magnified
740 posterior midgut caeca (left) and drawing of *Largus cinctus* posterior gastrointestinal tract (modified from
741 Glasgow 1914; right). **C.** Caeca dissected away from other parts of gastrointestinal tract. **D.** FISH
742 micrographs of caeca stained with DAPI (blue) and Cy-5 universal bacterial probe (red) for 16S; Left: 405
743 nm laser; Middle: 655 nm laser; Right: Merged. **E.** Merged FISH micrograph of caecal tissue stained with
744 DAPI-only (blue) or with DAPI and Cy-3 *Burkholderia*-specific probe (orange). **F.** Magnified view of caeca-
745 containing section of midgut. Anterior of gut oriented towards right. **G.** Close up view of constricted region

between fourth and fifth midgut region. Anterior of gut oriented towards top. **H.** Egg batch after removal of about half of eggs. **I.** 1st instar nymph of *L. californicus* probing egg batch with labium. **J.** 1st instar nymphs of *L. californicus.*

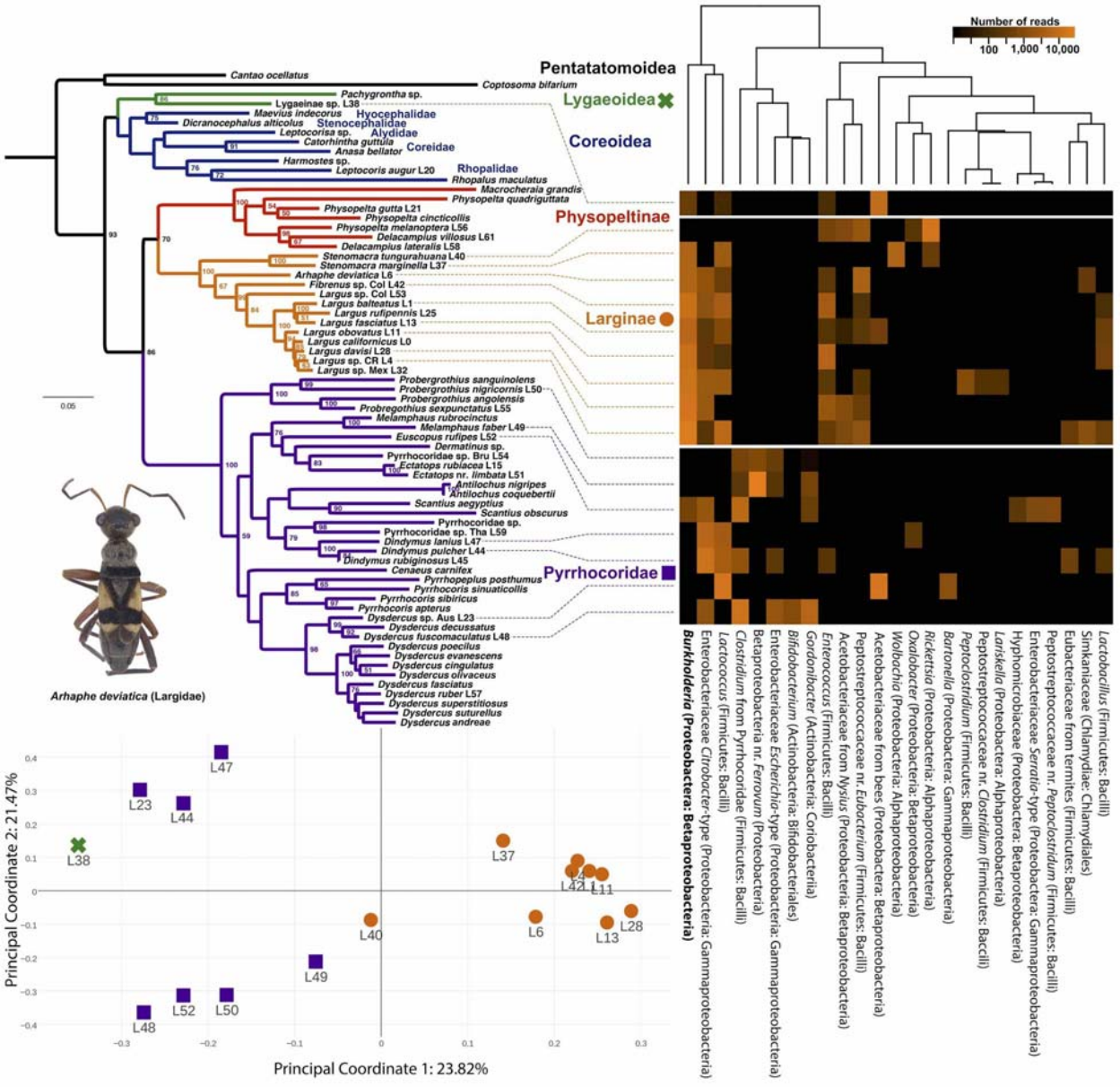
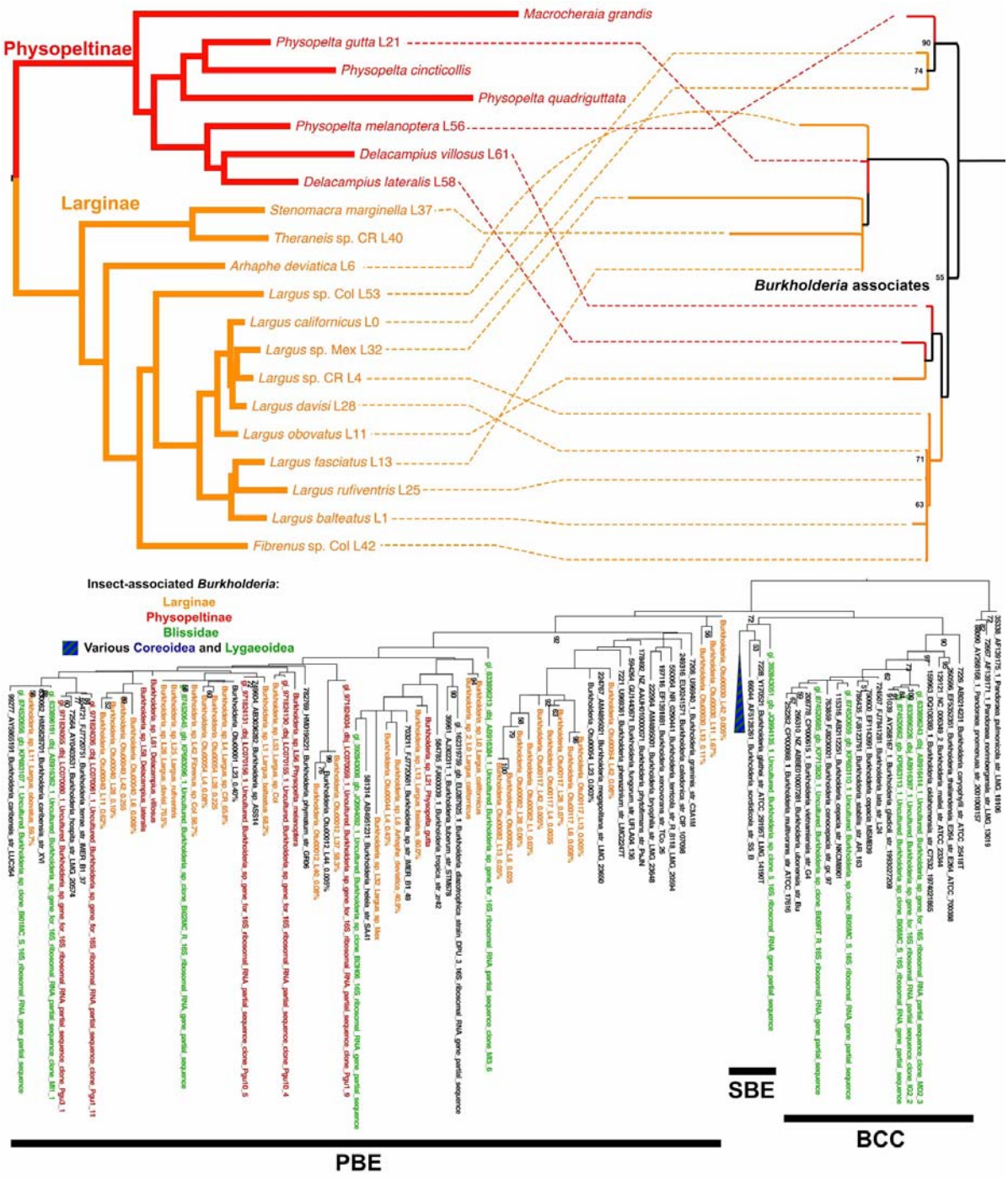


Figure 2. Maximum-likelihood phylogeny of Pyrrhocoroidea with bootstraps >50% displayed on branches and number of nucleotide changes indicated with scale bar indicating branch length corresponding to a mean of .05 nucleotide substitutions per site. Pentatomoidea was constrained as the outgroup. Displayed to the right is a heat map of bacterial OTUs from specimens indicated on the phylogeny with brightness corresponding to the $\log_{10}$ of number of Illumina reads divided by 10 or the

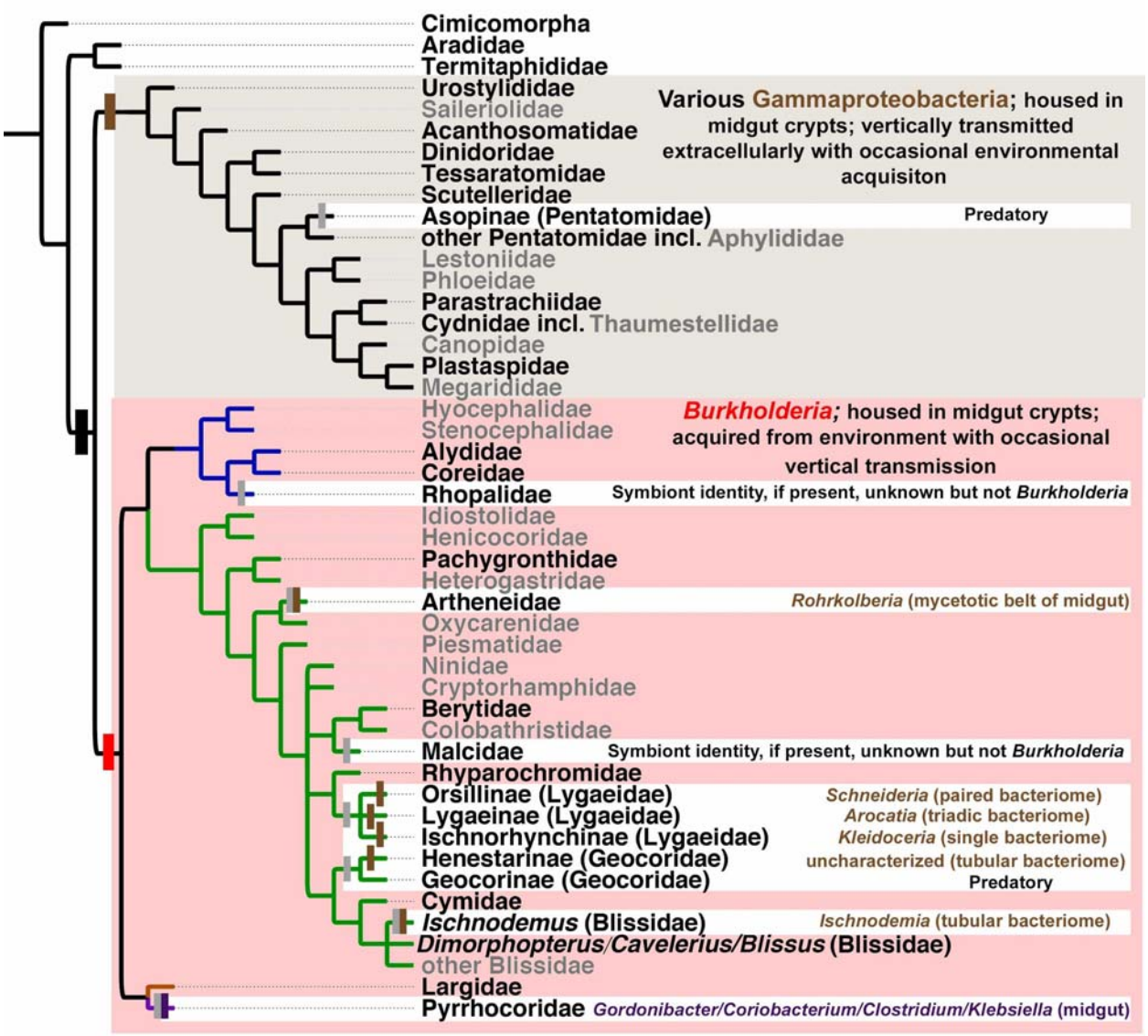
755 $\log_{10}$ of reads after subtracting 1. The dendrogram on top of the heatmap represents clustering of OTUs
756 based on their shared presence in samples. A principal coordinate analysis (PCoA) plot of abundance-
757 weighted Unifrac distances is displayed on the bottom left.
758
759
760
761
762



763

764 **Figure 3.** Bootstrap support >50% are displayed on bacterial phylogenies. Top. Maximally concordant
765 host (left) and symbiont (right) maximum likelihood phylogenies as produced by TreeMap3 with linkages
766 shown. Bottom. A phylogeny of newly obtained *Burkholderia* sequences with others from Genbank.
767 Sequences are colored by host as in Figure 1. OTUs from the Illumina dataset (as well as full length 16S
768 sequences from dominant *Burkholderia* strains) are followed by the percentage of reads they made up

769 out of the total for that sample. One node consisting only of lygaeoid and coreoid associated *Burkholderia*
770 in the SBE clade was collapsed for easier visualization.
771



772

773 **Figure 4.** The large black mark indicates the evolution of caeca. Grey marks indicate a loss of these
774 caeca (vestigial in Pyrrhocoridae). Brown marks indicate the evolution of association with
775 gammaproteobacteria, red with *Burkholderia*, and purple with a consortium of bacteria including two
776 Actinobacteria. Taxa in grey have not yet been examined. The phylogeny is based on relationships
777 recovered in this paper (for Coreoidea) and well-supported relationships from Bayesian and likelihood
778 analyses based on six Hox gene fragments (87). Families not represented in that analysis or with not

779 well-supported relationships were placed using successive weighting parsimony analyses based on
780 morphology as in (57) for Lygaeoidea and (88) for Pentatomoidea. References for symbiotic associates
781 are as in text (20–22, 32). Only one member of Lygaeinae (*Arocatus longiceps*) has so far been
782 demonstrated to be associated with a symbiont (20). Oxycarenidae are purported to possess an unpaired
783 bacteriome but the identity of the bacterial inhabitant is unknown (89).