The genome of the bee louse fly reveals deep convergences in the evolution of social inquilinism

Hélène Legout[1], David Ogereau[1], Jonathan Filée[1], Lionel Garnery[1], Clément Gilbert[1], Fabrice Requier[1], Amir Yassin[1], Héloïse Bastide[1]

[1]Laboratoire Évolution, Génomes, Comportement et Écologie, CNRS, IRD, Université Paris-Saclay – Institut Diversité, Écologie et Évolution (IDEEV), 12 route 128, 91190 Gif-sur-Yvette, France.

Corresponding author: Héloïse Bastide (heloise.bastide@universite-paris-saclay.fr)

Abstract

The nests of social insects often harbor a rich fauna of intruders, known as inquilines. Close relatedness between the host and the inquiline prevails due to potential genetic predispositions but how phylogenetically distant inquilines adapt to their hosts remains unclear. Here, we analyzed the genome of the wingless and blind bee louse fly *Braula coeca*, an inquiline of the Western honey bee *Apis mellifera*. We found that unlike many parasites, such as the human louse, the bee louse genome did not show significant erosion or strict reliance on an endosymbiont, likely due to a relatively recent age of inquilinism. However, a strikingly parallel evolution in a set of gene families was observed between the honey bee and the bee louse. Convergences included genes potentially involved in metabolism and immunity, and the loss of nearly all bitter-tasting gustatory receptors in agreement with life in a protective hive and a major diet of honey, pollens, and beeswax. Vision-related and odorant receptor genes also exhibited rapid losses. Only genes whose orthologs in the closely related *Drosophila melanogaster* respond to components of the honey bee alarm pheromones or floral aroma were retained, whereas the losses included orthologous receptors responsive to the anti-ovarian honey bee queen pheromone. These results establish a new model for the study of major morphological and neuroethological transitions and indicate that deep genetic convergences between phylogenetically distant organisms can underlie the evolution of social inquilinism.

Keywords: parasitism; morphological evolution; behavioral adaptations; gene family evolution; Drosophila.

Introduction

Inquilinism is a form of interspecific interactions wherein an organism, the inquiline, inhabits the "home" of another organism (Hegner 1926). Examples abound in nature (Moser 1964; Luczkovich *et al.* 1991; Sanver and Hawkins 2000; Kneitel and Miller 2002), and ancient cases are even present in the fossil record (Landman *et al.* 2014). The relationship between the inquiline and its host can range from commensalism to parasitism depending on the overlap in and availability of resources and space. Factors favoring the evolution of inquilinism greatly depend on the inquiline's capacity to hide its presence from the host. This is particularly true in the case of social inquilines, which constitute the most frequent case of inquilinism. wherein the efficient nest cleaning and care for the offspring performed by social organization are often fatal for the intruders. Therefore, phylogenetic relatedness with the host was suggested to play a major role in facilitating social inquilinism, since the inquiline would share some common genetic factors promoting camouflage, known as Emery's rule (Cini *et al.* 2019). Although recent phylogenetic studies have revised and sometimes rejected Emery's rule at the specific level (Huang and Dornhaus 2008; Lopez-Osorio *et al.* 2015; Romiguier *et al.* 2018; but see Savolainen and Vepsäläinen 2003; Degueldre *et al.* 2021), most cases of social inquilines remain related to the same genus (Jansen *et al.* 2010; Cardinal *et al.* 2010), tribe (Schrader *et al.* 2021), family (Ronquist 1994; Cardinal *et al.* 2010), or even order (Gilbert *et al.* 2012). The genetic basis underlying inquilines belonging to distinct orders or across wide phylogenetic distances is still less understood.

The hive of the Western honey bee *Apis mellifera* comprises several parasites and inquilines that are attracted by the hive's rich resources (honey, pollen, beeswax, and the bees themselves) and its clean and protective shelter (Winston 1987). These include several insects belonging to distinct orders such as wax moths, hive beetles, and endoparasitoid flies. None of these has endured as profound morphological changes that even its affiliation at the order-level was confounded, as the apterous and quasi-blind bee louse fly, *Braula coeca* (Figure 1A-C). The female lays eggs in honey (not brood) cells, and the hatched larvae eat pollen and wax, where they borrow tunnels in which they pupate without forming true puparia (Skaife 1922; Imms 1942). Following emergence, the adults attach to the body of worker bees, migrating from one individual to another until reaching the queen. There, they move to the head, stimulate regurgitation and imbibe from her mouth the honey and nectar (Skaife 1922; Imms 1942). The bee louse is considered an

inquiline kleptoparasite with potential negative effects on bee colony health due to galleries in bee combs and the facilitation of transmitting serious pathogenic viruses to the bees (Avalos *et al.* 2019).

The phylogenetic positioning within the Diptera of the family Braulidae, which contains the bee louse, has long been puzzling due to its modified morphology (Grimaldi and Underwood 1986). Interestingly, recent phylotranscriptomic and phylogenomic analyses show the bee louse to constitute a basal lineage within the Drosophilidae (Bayless *et al.* 2021; Winkler *et al.* 2022). This proximity to *Drosophila melanogaster*, the most investigated insect at the genetic, developmental, and neurobiological levels, makes the bee louse a unique model for the study of the genomic changes underlying major morphological and ecological shifts. We present here an annotated assembly of the bee louse *B. coeca* genome and compare the evolution of its genomic architecture and gene content with those of *D. melanogaster* (Adams *et al.* 2000), the honey bee *A. mellifera* (Weinstock *et al.* 2006), as well as to its homonym and morphologically-similar the ectoparasitic human louse *Pediculus humanus* (Kirkness *et al.* 2010). The comparisons revealed striking evidence of cross-order genomic parallelism and shed new light on the evolution of social inquilinism between phylogenetically distant organisms.

Results

Genome architecture

We sequenced whole genome from a pooled sample of 15 unsexed *Braula coeca* individuals, all collected from the same geographical location, the Island of Ouessant in France. We used a hybrid approach to assemble a draft genome using both long-read Oxford Nanopore Technology (ONT) and short-read Illumina sequencing (see Methods).

Size. The final assembly of the bee louse showed a size of 309 Mb, an N50 of 347227 bp and a GC content of 34.95%. BUSCO gave a score of 95.8% of the Dipteran conserved single-copy orthologs with 1.3% of duplicated genes. Analysis of heterozygous k-mers pairs distribution indicated that the genome is diploid (Supplementary Figure 1A) and genome size prediction using k-mers distribution spectra predict a genome size of 308 Mb, concordant with the assembly size (Supplementary Figure 1B). The bee louse genome is therefore larger than the genomes of the honey bee (227 Mb), *Drosophila melanogaster* (143 Mb), and the ectoparasitic human louse (108 Mb).

104 **Endosymbionts.** The taxonomic assignment of each contigs show that most of them
105 (96%) match with arthropods indicating no (or very few) DNA contamination
106 (Supplementary Figure 1C,D). Unlike in the human louse, no evidence for an obligate
107 endosymbiont was detected in the bee louse. Imms (1942) discussed the possibility that
108 the bee louse larvae have gut microbes that facilitate the digestion of the beeswax. We
109 cannot rule out this hypothesis, but our results show that such microbiota, if present,
110 likely do not persist in the adult stage.
111 **Gene content.** The annotation of the bee louse genome yielded 11,221 protein-coding
112 genes. This number is higher than in *A. mellifera* and *P. humanus* where 9,935 and 10,773
113 protein-coding genes are found, respectively, but lower than that of *D. melanogaster*
114 (13,968 protein-coding genes). The Annotation Edit Distance (AED), which measures the
115 congruence between gene annotation and its supporting evidence was ≤ 0.5 for 96.7% of
116 our gene models, indicating the near completeness of our annotation. Similarly, 79.79%
117 of the corresponding proteins had a Pfam domain, which is another indication of
118 annotation completeness since it varies between 57% and 75% in most eukaryotes.
119 **Synteny.** Orthologous genes were used to map *B. coeca* scaffolds to their corresponding
120 Muller's elements in *D. melanogaster*. The alignment showed strong consistency
121 indicating persistent synteny between the two lineages (Figure 1D). However, given the
122 short length of the scaffolds it was difficult to assess how much collinearity and
123 rearrangement events took place since their divergence. There is no karyotypic map of *B.*
124 *coeca* so it remains difficult at this stage to infer the chromosomal number from the
125 assembly alone.
126 **Mitochondrial genome.** The mitochondrial genome consisted of a single scaffold of 19
127 kb (Figure 1E) unlike in the human louse, which has multiple fragments inside the
128 mitochondria. Compared to conserved gene content and collinearity of the *D.*
129 *melanogaster* and *A. mellifera* mitogenomes (Crozier and Crozier 1993), the
130 mitochondrial gene content of the bee louse was incomplete, lacking the *ND1* gene,
131 included several duplications of the *ND2* and *ND4L* genes that are truncated and/or
132 display frameshifts and had a rearrangement between the *COX1* and *COX2* genes. The
133 nuclear genomes contained 165 mitochondrial DNA insertions (NUMTs) distributed on
134 88 contigs and totalizing 278 kb (~0.09% of the genome). The proportion of NUMTs in
135 the bee louse approaches that of the honey bee (0.08%, (Behura 2007)) but exceeds that
136 of the genus *Drosophila* (0.03%, (Rogers and Griffiths-Jones 2012).

137 **Transposable elements.** The large genome size of the bee louse despite its low gene
138 content compared to *D. melanogaster* suggest an increase in repetitive sequences.
139 RepeatModeler and RepeatMasker analyses indicated that nearly 41.41% of the *B. coeca*
140 genome consist of such sequences, compared to 22.15% and 11.14% in *D. melanogaster*
141 and *A. mellifera*, respectively (Supplementary Table 1). There is a far larger proportion of
142 long interspersed nuclear elements (LINEs) retrotransposons in *B. coeca* (15.18%)
143 compared to only 2.38% and none in *D. melanogaster* and *A. mellifera*, respectively. Long
144 terminal repeat (LTR) elements on the other hand were fewer in *B. coeca* (0.57%) and *A.*
145 *mellifera* (0.17%) than in *D. melanogaster* (6.99%).
146 Because host-parasite relationships have repeatedly be invoked as a factor that
147 may favor horizontal transfer of transposable elements (TE) (Gilbert *et al.* 2010; Ortiz *et*
148 *al.* 2015; Venner *et al.* 2017), we searched for evidence of such transfers between *B. coeca*
149 and *A. mellifera* (Supplementary Text 1). We found one TE, *Famar1-like* element,
150 previously described in the earwig *Forficula auricularia* (Barry *et al.* 2004a) that shows
151 high similarity between *B. coeca* and *A. mellifera* but was absent in *D. melanogaster*, highly
152 suggestive of acquisition through horizontal transfer (Supplementary Figure 2). However,
153 phylogenetic analysis of multiple copies of this TE extracted from 37 widely divergent
154 animal species (Supplementary Figure 3) shows that it was most likely acquired through
155 independent transfer events in *B. coeca* and *A. mellifera*, and that the tight ecological
156 connection between the two species are unlikely to have played a direct role in these
157 transfers (Supplementary Text 1).
158
159 *Gene family evolution*
160 **Families with excess losses.** We identified gene families that underwent reduction or
161 extension as compared to *D. melanogaster* using DAVID (Sherman *et al.* 2022).
162 Underrepresented families (27 families with False Discovery Rate (FDR) < 0.05) showed
163 a striking parallelism with the honey bee, with three of the most underrepresented
164 families being similar (given in Honey bee Genome Sequencing Consortium 2006;
165 Supplementary Table 2). These included InterPro-defined families such as Peptidase S1
166 (IPR001254), Ecdysteroid kinase-like (IPR004119) and Zinc finger AD-type (IPR012934).
167 Peptidases play a major role in detoxification and digestion of protein-rich diets.
168 Their reduction in both the honey bee and the bee louse could be due to the
169 overprotectiveness of their mutual habitat, the hive, and/or the low protein content of

170 their food, nectar and honey. The Ecdysteroid kinase-like (EcKL) family is also suspected
171 to include proteins involved in detoxification (Scanlan *et al.* 2020). We also noted a
172 functional affinity between two underrepresented families involved in the formation of
173 the exoskeleton: the Insect cuticle protein (IPR000618) in the honey bee and Chitin-
174 binding domain (IPR002557) in the bee louse. Cuticles could act as barriers against
175 environmental toxins, which may not be highly encountered in the hive. Remarkably, *B.*
176 *coeca* is unique among Cyclorrhaphan Diptera as its pupa, similarly to the honey bee's
177 (Winston 1987), is contained in the unmodified cuticle of the third instar larva, and no
178 sclerotized puparium is formed (Skaife 1922; Imms 1942).

179 Significant biological processes terms (FDR < 0.05) were mostly related to
180 proteolysis and lipid metabolism (e.g., fatty-acyl-co-A biosynthesis), whereas cellular
181 components terms were associated with the membranes and extracellular space in
182 agreement with a biased loss of genes potentially involved with metabolism,
183 detoxification and/or immunity (Supplementary Table 2).

184 **Families with excess gains.** We did not find a similar important overlap in
185 overrepresented gene families between the bee louse and the honey bee (5 families with
186 FDR < 0.05; Supplementary Table 3). Expanding gene families unique to the honey bee
187 included 7TM odorant receptor (IPR004117), Ankyrin (IPR002110), Yellow/royal jelly
188 protein (IPR003534) and LysR substrate-binding (IPR005119) (Weinstock *et al.* 2006).
189 We found one family, Pleckstrin homology domain (IPR001849), that was
190 overrepresented in both species. This domain is a part of several lipases that are known
191 to be involved in wax ingestion. Indeed, larvae of the bee louse mine the beeswax (Imms
192 1942), whereas honey bee workers use their salivary lipases while chewing the wax to
193 form the hive combs (Kurstjens *et al.* 1985).

194 Gene Ontology (GO) biological and cellular terms with significant enrichment at
195 FDR < 0.05 associated with morphological (e.g., imaginal disc-derived leg morphogenesis,
196 autophagy, and dorsal closure) and/or neurological (e.g., dendrite morphogenesis, axon
197 guidance, sensory perception of sound, and neuromuscular junction) developments
198 (Supplementary Table S3). The bee louse duplicated genes hence may play a role in the
199 evolution of the particular morphologies that helped adaptation to the phoretic lifestyle.

200 **Vision.** The species Latin name refers to the early assumption that the bee louse was blind
201 due to the reduction of the eye size and the loss of the ocelli. However, thin optic nerves
202 connect the brain to the rudimentary eyes, which lack ommatidia and pigments

203 (Müggenburg 1892), and the bee louse demonstrates negative phototaxis, indicating a
204 certain degree of light perception (Kaschef 1959). In agreement with reduced vision in
205 the bee louse, we found three out of the seven rhodopsin genes, which are responsible for
206 colored vision and the positive phototaxis of *D. melanogaster*. Two of the three genes, *Rh1*
207 and *Rh6*, are expressed in the ommatidia and are sensitive to light with long wavelengths
208 (Senthilan and Helfrich-Förster 2016), whereas the third one, *Rh7*, is expressed in the
209 brain and regulates light-dependent circadian entrainment (Ni *et al.* 2017). The role of
210 these opsins in light detection despite the absence of ommatidia is unclear. Remarkably,
211 *Rh1*, *Rh6* and *Rh7* are structurally required in mechanosensory bristles to control larval
212 locomotion (Zanini *et al.* 2018), and *Rh1* and *Rh6* also detect temperature (Leung and
213 Montell 2017), whereas *Rh1* and *Rh7* detect low concentrations of a bitter plant
214 component (Leung *et al.* 2020). Therefore, the retention of these rhodopsins in the bee
215 louse could mainly be due to their unconventional functions. On the other hand, the red-
216 sensitive rhodopsin *Rh2*, which is exclusively expressed in the ocelli and used for horizon
217 detection in *D. melanogaster* (Mishra *et al.* 2021), is among those which were lost in the
218 bee louse, in agreement with the loss of the ocelli in the bee louse.
219 **Olfaction.** Odorant receptors (ORs) are essential to detect volatile chemical signals from
220 the environment. In most *Drosophila* species, there are nearly 61 ORs, whereas in the
221 honey bee this family has expanded to reach 160 (Robertson and Wanner 2006). We
222 found 29 ORs in the bee louse in addition to the Orco ortholog. Of these four had no direct
223 orthologs in *D. melanogaster*, but the remaining 25 genes were orthologous to 21 genes
224 in *D. melanogaster* and 12 were duplicates specific to the bee louse lineage (Figure 2A;
225 Supplementary Figure 3). Judging from the response of those ORs to different volatiles in
226 *D. melanogaster* as curated in the DOOR database (Münch and Galizia 2016) and assuming
227 potential conservation of function, we can divide the bee louse ORs into three categories.
228 First, 11 ORs respond to different components of the honey bee workers alarm and
229 mandibular gland pheromones, e.g., 2-heptanol, propyl acetate, 2-heptanone, 1-hexanol,
230 butyl acetate, isopentyl acetate, etc. These included *Or13a* (x2 paralogues), *Or42a*,
231 *Or42b/Or59b*, *Or43a*, *Or47b*, *Or49a/Or85f*, *Or67b* (x4), *Or74a*, *Or85b/Or85c*, and *Or85e*
232 (x2). Second, 12 ORs respond to different floral, pollen and nectar aromas, such as
233 acetophenones, ethylguaiacol, geranyl acetate or fenchone, but most importantly to
234 benzaldehyde, a major volatile of honey (Machado *et al.* 2020; Starowicz *et al.* 2021).
235 These included *Or30a*, *Or45b* (x2), *Or59a*, *Or67b* (x4), *Or82a*, *Or85e* (x2) and *Or94a*. Note

236 that the multiple *B. coeca* paralogs orthologous to *D. melanogaster Or67b* and *Or85e* genes
237 belong to both categories. Third, one OR, *Or67d*, which is involved in sexual pheromone
238 11-cis-vaccenyl acetate perception (Ha and Smith 2006), is present in the bee louse. Three
239 among the ORs that were lost in the bee louse compared to other drosophilids, *Or49b*,
240 *Or56a* and *Or98a*, are responsible to *D. melanogaster* response to the anti-ovarian honey
241 bee queen mandibular pheromones 9-oxo-2-decenoic acid (9ODA) and 10-hydroxy-2-
242 decenoic acid (10HDA) (Galang *et al.* 2019). Their loss most likely protected the capacity
243 of the bee louse to reproduce in the hive.
244 **Taste.** Gustatory receptors (GRs) allow detecting soluble chemical signals. There are
245 nearly 68 GRs in *D. melanogaster*, that respond mostly to sweet, bitter and carbon dioxide
246 (CO_2) tastes (Weiss *et al.* 2011). Unlike their expanded OR family, the honey bee has only
247 10 GRs, of which 7 are orthologous to sweet *Drosophila* GRs (Robertson and Wanner
248 2006). This is likely due to the bees' food reliance on sweet floral nectars and honey. We
249 found 13 GRs in the bee louse, with no duplications (Figure 2B; Supplementary Figure 4).
250 These GRs could be classified according to their *D. melanogaster* orthology into three
251 categories. First, 6 GRs belong to the sweet class, namely *Gr43a*, *Gr61a* and *Gr64a,b,e,f*
252 which usually respond to sucrose, maltose and fructose and other major honey sugars.
253 Second, 5 GRs belong to the bitter class, namely *Gr33a*, *Gr57a*, *Gr66a*, *Gr93a* and *Gr94a*,
254 suggesting that nearly 50 mostly bitter tasting drosophilid GRs were lost in the bee louse.
255 Note that in *D. melanogaster*, *Gr33a* and *Gr66a*, are potentially involved in sexual
256 pheromones detection (Lacaille *et al.* 2007; Moon *et al.* 2009). Third, 2 GRs, *Gr21a* and
257 *Gr63a*, which detects CO_2 odor in *D. melanogaster* (Jones *et al.* 2007) are present in the
258 bee louse. Those GRs are absent in the honey bee despite its ability to perceive CO_2 (Stange
259 and Diesendorf 1973). High CO_2 concentrations are probably characteristic of largely
260 populated hives and induce fanning response in bees (Seeley 1974). The quasi-blind bee
261 louse may therefore use CO_2 concentrations to detect the bees in the dark hive
262 environment.
263
264 **Discussion**
265 That the enigmatic bee louse is indeed a drosophilid (Winkler *et al.* 2022), a lineage
266 within the most investigated insect family with more than 100 fully sequenced genomes
267 (Kim *et al.* 2021), is undoubtedly one of the most exciting discoveries in Dipteran
268 phylogeny. How could a fly with a typical drosophilid genome become partly ecologically

269 like a bee and partly morphologically like a louse? Our cross-order comparisons of the bee
270 louse to its homonyms shed significant light on the genomic basis of these spectacular
271 convergences.

272 The genome of the human louse is among the smallest sequenced insect genomes
273 (Kelley *et al.* 2014). Loss of significant portions of genomic and gene contents is a
274 characteristic of obligate parasites specializing on specific hosts or inhabiting extreme
275 environments. Indeed, the gene content of the bee louse genome approached that of the
276 human louse, but the ~309 Mb-long genome of the bee louse is longer than that of most
277 drosophilid species, even being slightly longer from the largest genome in this family
278 (~304 Mb-long; Kim *et al.* 2021). Besides, the human louse has an obligatory bacterial
279 endosymbiont that may compensate the loss of its genes (Kirkness *et al.* 2010) but no
280 evidence for such associations is present in the bee louse. These differences may mainly
281 indicate the relative recency of the shift to inquilinism in the bee louse compared to the
282 230 million years (myr) of specialization in true lice (Misof *et al.* 2014). *Megabraula*, the
283 closest relative to the genus *Braula*, is an inquiline of the giant honey bee *Apis laboriosa*,
284 whose divergence from the Western honey bee *A. mellifera* is around 23 myr ago
285 (Grimaldi and Underwood 1986; Cardinal *et al.* 2010). The crown age of the Drosophilidae
286 and the divergence time between *Apis* and its closest pollen-basket (corbiculate) bees are
287 estimated at 66-70 myr ago (Cardinal *et al.* 2010; Suvorov *et al.* 2021). This indicates that
288 association between braulids and *Apis* has arisen between 70 to 20 myr ago.

289 The origin of social inquilinism requires the evolution of multiple convergences
290 that can locate the host, deceive it to enter the colony, adapt to its social organization, and
291 remain undetected or not easily removed from the colony. These phenotypes, which are
292 often host-specific, evolve more easily when the inquiline and the host are
293 phylogenetically close (Cini *et al.* 2019). Despite the distance of the bee louse from honey
294 bees, gene family analysis provided strong evidence for convergent evolution mostly for
295 genes potentially involved in immunity, detoxification, metabolism and chemical
296 perception. Although *Braula* has lost nearly half of the typical drosophilids odorant
297 receptor repertoire, in contrast to the major expansion of this family in the honey bee, it
298 predominantly retained genes whose orthologs in *D. melanogaster* detect compounds of
299 the honey bee pheromones and/or floral aroma and honey odors. Low concentrations of
300 isopentyl acetate (IPA), the major component of the alarm pheromone, released by
301 unstressed workers at hive entries attract the parasitic hive beetle *Aethina tumida* (Torto

302 *et al.* 2007), suggesting that the detection of the host odors could be a common strategy
303 among phylogenetically distant social inquilines and parasites. We also noted that the bee
304 louse has multiple copies related to *Or67b* which responds to several components of bee
305 alarm pheromones, such as 1-hexanol, 2-heptanone, 1-butanol and 3-methyl-1-butanol,
306 as well as to benzaldehyde, the major honey volatile, in *D. melanogaster* (Münch and
307 Galizia 2016), suggesting a possible dual ancestral function that might have facilitated
308 association with the honey bees. The multiple copies related to *Or67b* are remarkable
309 since the copy number expansion of this gene was associated with the evolution of
310 herbivory and strong plant-association in the drosophilid genus *Scaptomyza* (Goldman-
311 Huertas *et al.* 2015; Matsunaga *et al.* 2022). Therefore, the bee louse might have evolved
312 from a flower-breeding or plant-associated lineage within the Drosophilidae, a lifestyle
313 that has recurrently evolved in this family (Yassin 2013).
314 Whereas major molecular convergences could exist between the inquiline and its
315 social host, divergent strategies to adapt to the eusocial lifestyle requirements are still
316 needed. The loss of the three odorant receptors, *Or56a*, *Or49b* and *Or98a*, that respond in
317 *D. melanogaster* to the honey bee queen's pheromones which "sterilize" the bee workers,
318 mainly 9-oxo-2-decenoic acid (9ODA) and possibly 9-hydroxy-2-decenoic acid (9HDA)
319 (Galang *et al.* 2019), is a notable example. The queen's pheromones elicit anti-ovarian
320 response in other insects including *D. melanogaster* mostly through the activation of the
321 neurons bearing these receptors. Therefore, a *sine qua non* condition for reproducing in a
322 beehive is to protect against the effects of those pheromones, with the loss of the
323 responding receptors being a preliminary and effective strategy. However, this raises the
324 question of how the bee lice recognize the queen, which they preferentially infest (Imms
325 1942). It is therefore possible that higher sensitivity to worker pheromones, as suggested
326 by the repertoire of retained odorant receptors (see above), may help to mostly
327 discriminate the workers hence facilitating the recognition of the queen, who has a
328 rudimentary sting gland. Another possibility is that strong CO_2 emission by the court
329 surrounding the queen may be an indicator of her location, as could be suggested from
330 the retention of the two CO_2-smelling gustatory receptors, *Gr21a* and *Gr63a*, whose
331 orthologs are absent in the honey bee. It also remains unclear how the bee lice sexually
332 communicate in the hive given the predominance of chemical camouflage; the
333 hydrocarbon profile of bee lice from different colonies mimicked that of the colony of
334 origin (Martin and Bayfield 2014). The retention of some chemosensory genes whose

335 orthologs are involved in sexual pheromone detection in *D. melanogaster*, such as *Or67d*,
336 *Gr33a* and *Gr66a* (Kurtovic *et al.* 2007; Lacaille *et al.* 2007; Moon *et al.* 2009) suggest that
337 pheromonal communications may still be present in the bee louse but perhaps at low
338 undetectable levels.

339 Small size, loss of wings and the evolution of strongly clinging legs are all
340 morphological adaptations that could prevent the honey bees getting rid of the bee lice.
341 All these adaptations are convergent with ectoparasitic true lice, and for some, such as
342 apterism, represent major recurrent changes that have responded to distinct pressures
343 throughout the history of insects (Roff 1990). We found intact most of the main wing
344 development genes whose mutations severely reduce the wing in *D. melanogaster*, such
345 as *wingless*, *apterous* or *vestigial*. This means that the major morphological changes more
346 likely resulted from regulatory changes of these core genes or modifications of other
347 genes. The regression of visual systems and their underlying genes in the bee louse
348 spending most of its life cycle in the bee hives is a common phenomenon in animals
349 inhabiting dark environments, such as fossorial mammals (Partha *et al.* 2017) and
350 cavefishes (Policarpo *et al.* 2021). With its genetic relatedness to *Drosophila* and
351 ecological association to *Apis*, two major laboratory models, and with new genomic tools
352 presented here, the bee louse *Braula coeca* is a promising model to address questions
353 related to deep convergences that are still difficult to approach in multiple highly
354 specializing animals.

355

356 **Materials and Methods**

357 *Sample collection and genomic library preparation*

358 Samples of *Braula coeca* were collected from honey bee colonies on the Island of Ouessant
359 in France and kindly provided to us by the *Association Conservatoire de l'Abeille Noire*
360 *Bretonne* (A.C.A.N.B.). Genomic DNA was extracted from 15 unsexed individuals
361 conserved in alcohol, using the Nucleobond AXG20 kit and buffer set IV from Macherey-
362 Nagel (ref. 740544 and 740604, https://www.mn-net.com, Düren, Germany).

363

364 *Genome sequencing and assembly*

365 We used a hybrid approach to assemble a draft genome of *B. coeca* using both long-
366 read Oxford Nanopore Technology (ONT) and short-read Illumina sequencing, as in Miller
367 et al. (2019). Before nanopore sequencing, a size selection was conducted on the DNA

368 using the SRE XS from Circulomics (https://www.circulomics.com/, Baltimore, Maryland,
369 USA). The SQK-LSK110 kit from Oxford Nanopore Technology (Lu *et al.* 2016;
370 https://nanoporetech.com/) was then used to prepare the samples for nanopore
371 sequencing following manufacturer's protocol. The library was loaded and sequenced on
372 a R9.4.1 flow cell (ref FLO-Min106) for sequencing. Raw data were basecalled using
373 Guppy v5.0.11 and the "sup" algorithm. Illumina paired-end sequencing was performed
374 by Novogene Company Limited (https://en.novogene.com, Cambridge, UK) on the same
375 DNA library.
376 We used MaSuRCA v4.0.3 (Zimin *et al.* 2017) to produce the hybrid assembly of
377 our genome, using the Cabog assembler. We obtained a final assembly size of 309,35Mb
378 in 2477 contigs, with a N50 of 347211 pb. The completeness of the assembly was
379 estimated to 95,8% with Busco v5.0 on the *diptera_odb10* dataset
380 (C:95.8%[S:94.6%,D:1.2%],F:0.7%,M:3.5%,n:3285).
381
382 *Estimation of genome size, endosymbionts detection and mitogenome assembly*
383 K-mers frequencies within short-read data were obtained with KMC 3 (Kokot *et al.*
384 2017). Genome size and ploidy were inferred using GenomeScope v2.0 with k-mer size =
385 21 and Smudgeplot (Ranallo-Benavidez *et al.* 2020). Contig taxonomy was performed
386 using Blobtools (Laetsch and Blaxter 2017) with Diamond as search engine (Buchfink *et*
387 *al.* 2015) against the UniProt database using a local copy of the NCBI TaxID file for the
388 taxonomic assignation of the best hit. Minimap2 (Li 2018) was used for read mapping.
389
390 *Genome annotation*
391 The *B. coeca* genome was annotated using Maker v2.31.10 (Cantarel *et al.* 2008),
392 following the protocol given in Muller *et al.* (2021), wherein multiple rounds of Maker
393 supported by the training of the SNAP v.2006-07-28 (Korf 2004) and Augustus v.3.3.3
394 (König *et al.* 2016) gene finding and prediction tools, were conducted. Transcriptome of
395 *B. coeca* (NCBI accession no. SRR2046564; 1KITE Consortium; (Bayless *et al.* 2021)) and
396 proteomes of five *Drosophila* species, namely *D. innubila* (Hill *et al.* 2019), *D. albomicans*
397 (Mai *et al.* 2020), *D. bipectinata* (Kim *et al.* 2021), *D. melanogaster* (Adams *et al.* 2000) and
398 *D. virilis* (Clark *et al.* 2007) were used to guide the annotation. Protein-Protein BLAST
399 2.9.0+ (Altschul *et al.* 1997) (-evalue 1e-6 -max_hsps 1 -max_target_seqs 1) was then used
400 to assess putative protein functions in *B. coeca* by comparing the protein sequences given

401 by Maker to the protein sequences from the annotated genome of *D. melanogaster*. The
402 colocalization of *B. coeca* and their *D. melanogaster* orthologs on their respective contigs
403 and chromosomal arms was analyzed to test for synteny.
404
405 *Transposable elements annotation and transfer*
406 Transposable elements were identified in each species following a two-step
407 protocol. First, we used RepeatModeler v 2.0.1 (Flynn *et al.* 2020) with default parameters
408 to generate a *de novo* library of repetitive regions. RepeatMasker v 4.0.9 (Flynn *et al.*
409 2020) was then run with the newly generated library and the options -a (create a .align
410 output file) and -s (slow search; more sensitive) to create a summary of the families of
411 transposable elements found in each reference genome along with the percentage of the
412 genome they represent. Horizontal transfer analyses protocols of the *Famar1*-like
413 element are given in Supplementary Text 1.
414
415 *Gene family evolution*
416 The Database for Annotation, Visualization and Integrated Discovery DAVID 2021
417 (Sherman *et al.* 2022) was used to test for gene ontology (GO) terms enrichments among
418 lists of *D. melanogaster* orthologs annotated on the *B. coeca* genome (see above). The
419 program was either fed with the list of *D. melanogaster* genes that were absent or
420 duplicated in the *B. coeca* genome to test for under- and overrepresented gene families,
421 respectively, compared to *D. melanogaster* genome, as denoted by the False Discovery
422 Rate (FDR) correction for multiple tests. Each analysis was conducted for biological
423 processes, molecular function, and cytological components.
424
425 *Chemosensory superfamilies evolution*
426 Protein sequences of the odorant (ORs) and gustatory (GRs) receptors of *A.*
427 *mellifera* and *D. melanogaster* were obtained from Robertson and Wanner (2006) to
428 which we added the annotated ORs and GRs protein sequences of *B. coeca*. Sequences
429 were aligned using Molecular Evolutionary Genetic Analysis (MEGA X) software package
430 (Kumar *et al.* 2018), which was also used to infer a maximum-likelihood phylogenetic tree
431 for each family. We used iToL v4 (Letunic and Bork 2019) to visualize the trees. *B. coeca*
432 protein sequences are given in Supplementary Datasets 2 and 3 for ORs and GRs,
433 respectively.

Acknowledgments

We are very grateful to Drs. Julie Carcaud, Frédéric Marion-Poll and Jean-Christophe Sandoz for insightful comments on an early draft of the manuscript. *Braula* genome sequencing was funded by a grant from Université Paris Saclay (ADAPAR) to HB.

Conflicts of interest

The authors declare no conflicts of interest.

References

Adams M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, *et al.*, 2000 The Genome Sequence of Drosophila melanogaster. Science 287: 2185–2195. https://doi.org/10.1126/science.287.5461.2185

Altschul S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 25: 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Avalos J., H. Rosero, G. Maldonado, and F. J. Reynaldi, 2019 Honey bee louse (Braula schmitzi) as a honey bee virus vector? Journal of Apicultural Research 58: 427–429. https://doi.org/10.1080/00218839.2019.1565726

Barry E. G., D. J. Witherspoon, and D. J. Lampe, 2004a A Bacterial Genetic Screen Identifies Functional Coding Sequences of the Insect *mariner* Transposable Element *Famar1* Amplified From the Genome of the Earwig, *Forficula auricularia*. Genetics 166: 823–833. https://doi.org/10.1534/genetics.166.2.823

Barry E. G., D. J. Witherspoon, and D. J. Lampe, 2004b A Bacterial Genetic Screen Identifies Functional Coding Sequences of the Insect mariner Transposable Element Famar1 Amplified From the Genome of the Earwig, Forficula auricularia. Genetics 166: 823–833. https://doi.org/10.1093/genetics/166.2.823

Bayless K. M., M. D. Trautwein, K. Meusemann, S. Shin, M. Petersen, *et al.*, 2021 Beyond Drosophila: resolving the rapid radiation of schizophoran flies with phylotranscriptomics. BMC Biol 19: 23. https://doi.org/10.1186/s12915-020-00944-8

Behura S. K., 2007 Analysis of Nuclear Copies of Mitochondrial Sequences in Honeybee (Apis mellifera) Genome. Molecular Biology and Evolution 24: 1492–1505. https://doi.org/10.1093/molbev/msm068

467 Buchfink B., C. Xie, and D. H. Huson, 2015 Fast and sensitive protein alignment using
468 DIAMOND. Nat Methods 12: 59–60. https://doi.org/10.1038/nmeth.3176
469 Cantarel B. L., I. Korf, S. M. C. Robb, G. Parra, E. Ross, *et al.*, 2008 MAKER: An easy-to-use
470 annotation pipeline designed for emerging model organism genomes. Genome Res. 18: 188–
471 196. https://doi.org/10.1101/gr.6743907
472 Cardinal S., J. Straka, and B. N. Danforth, 2010 Comprehensive phylogeny of apid bees reveals
473 the evolutionary origins and antiquity of cleptoparasitism. Proceedings of the National
474 Academy of Sciences 107: 16207–16211. https://doi.org/10.1073/pnas.1006299107
475 Cini A., S. Sumner, and R. Cervo, 2019 Inquiline social parasites as tools to unlock the secrets
476 of insect sociality. Philosophical Transactions of the Royal Society B: Biological Sciences 374:
477 20180193. https://doi.org/10.1098/rstb.2018.0193
478 Clark A. G., M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, *et al.*, 2007 Evolution of
479 genes and genomes on the Drosophila phylogeny. Nature 450: 203–218.
480 https://doi.org/10.1038/nature06341
481 Crozier R. H., and Y. C. Crozier, 1993 The mitochondrial genome of the honeybee Apis
482 mellifera: complete sequence and genome organization. Genetics 133: 97–117.
483 https://doi.org/10.1093/genetics/133.1.97
484 Degueldre F., P. Mardulyn, A. Kuhn, A. Pinel, C. Karaman, *et al.*, 2021 Evolutionary history
485 of inquiline social parasitism in Plagiolepis ants. Molecular Phylogenetics and Evolution 155:
486 107016. https://doi.org/10.1016/j.ympev.2020.107016
487 Emery C., 1909 Über den Ursprung der dulotischen, parasitischen und myrmekophilen
488 Ameisen. Biologisches Centralblatt 29: 352–362.
489 Flynn J. M., R. Hubley, C. Goubert, J. Rosen, A. G. Clark, *et al.*, 2020 RepeatModeler2 for
490 automated genomic discovery of transposable element families. Proceedings of the National
491 Academy of Sciences 117: 9451–9457. https://doi.org/10.1073/pnas.1921046117
492 Galang K. C., J. R. Croft, G. J. Thompson, and A. Percival-Smith, 2019 Analysis of the
493 Drosophila melanogaster anti-ovarian response to honey bee queen mandibular pheromone.
494 Insect Molecular Biology 28: 99–111. https://doi.org/10.1111/imb.12531
495 Gilbert C., S. Schaack, J. K. Pace, P. J. Brindley, and C. Feschotte, 2010 A role for host-parasite
496 interactions in the horizontal transfer of transposons across phyla. Nature 464: 1347–50.
497 https://doi.org/10.1038/nature08939
498 Gilbert J. D. J., L. A. Mound, and S. J. Simpson, 2012 Biology of a new species of socially
499 parasitic thrips (Thysanoptera: Phlaeothripidae) inside Dunatothrips nests, with evolutionary
500 implications for inquilinism in thrips. Biological Journal of the Linnean Society 107: 112–122.

https://doi.org/10.1111/j.1095-8312.2012.01928.x

Goldman-Huertas B., R. F. Mitchell, R. T. Lapoint, C. P. Faucher, J. G. Hildebrand, *et al.*, 2015 Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. PNAS 112: 3026–3031. https://doi.org/10.1073/pnas.1424656112

Grimaldi D., and B. A. Underwood, 1986 Megabraula, a new genus for two new species of Braulidae (Diptera), and a discussion of braulid evolution. Systematic Entomology 11: 427–438. https://doi.org/10.1111/j.1365-3113.1986.tb00534.x

Ha T. S., and D. P. Smith, 2006 A Pheromone Receptor Mediates 11-cis-Vaccenyl Acetate-Induced Responses in Drosophila. Journal of Neuroscience 26: 8727–8733. https://doi.org/10.1523/JNEUROSCI.0876-06.2006

Hegner R. W., 1926 The Biology of Host-Parasite Relationships Among Protozoa Living in Man. The Quarterly Review of Biology 1: 393–418.

Hill T., B. S. Koseva, and R. L. Unckless, 2019 The Genome of Drosophila innubila Reveals Lineage-Specific Patterns of Selection in Immune Genes. Molecular Biology and Evolution 36: 1405–1417. https://doi.org/10.1093/molbev/msz059

Huang M. H., and A. Dornhaus, 2008 A meta-analysis of ant social parasitism: host characteristics of different parasitism types and a test of Emery's rule. Ecological Entomology 33: 589–596. https://doi.org/10.1111/j.1365-2311.2008.01005.x

Imms A. D., 1942 On Braula coeca Nitsch and its affinities. Parasitology 34: 88–100. https://doi.org/10.1017/S0031182000016012

Jansen G., R. Savolainen, and K. Vepsäläinen, 2010 Phylogeny, divergence-time estimation, biogeography and social parasite–host relationships of the Holarctic ant genus Myrmica (Hymenoptera: Formicidae). Molecular Phylogenetics and Evolution 56: 294–304. https://doi.org/10.1016/j.ympev.2010.01.029

Jones W. D., P. Cayirlioglu, I. Grunwald Kadow, and L. B. Vosshall, 2007 Two chemosensory receptors together mediate carbon dioxide detection in Drosophila. Nature 445: 86–90. https://doi.org/10.1038/nature05466

Kaschef A.-H., 1959 The sensory physiology and behaviour of the Honeybee LouseBraula coeca Nitzcsh (Diptera, Braulidae). Ins. Soc 6: 313–342. https://doi.org/10.1007/BF02225778

Kelley J. L., J. T. Peyton, A.-S. Fiston-Lavier, N. M. Teets, M.-C. Yee, *et al.*, 2014 Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. Nat Commun 5: 4611. https://doi.org/10.1038/ncomms5611

Kim B. Y., J. R. Wang, D. E. Miller, O. Barmina, E. Delaney, *et al.*, 2021 Highly contiguous assemblies of 101 drosophilid genomes, (G. Coop, P. J. Wittkopp, and T. B. Sackton, Eds.).

535 eLife 10: e66405. https://doi.org/10.7554/eLife.66405
536 Kirkness E. F., B. J. Haas, W. Sun, H. R. Braig, M. A. Perotti, *et al.*, 2010 Genome sequences
537 of the human body louse and its primary endosymbiont provide insights into the permanent
538 parasitic lifestyle. Proceedings of the National Academy of Sciences 107: 12168–12173.
539 https://doi.org/10.1073/pnas.1003379107
540 Kneitel J. M., and T. E. Miller, 2002 Resource and Top-Predator Regulation in the Pitcher Plant
541 (sarracenia Purpurea) Inquiline Community. Ecology 83: 680–688.
542 https://doi.org/10.1890/0012-9658(2002)083[0680:RATPRI]2.0.CO;2
543 Kokot M., M. Długosz, and S. Deorowicz, 2017 KMC 3: counting and manipulating k-mer
544 statistics. Bioinformatics 33: 2759–2761. https://doi.org/10.1093/bioinformatics/btx304
545 König S., L. W. Romoth, L. Gerischer, and M. Stanke, 2016 Simultaneous gene finding in
546 multiple genomes. Bioinformatics 32: 3388–3395.
547 https://doi.org/10.1093/bioinformatics/btw494
548 Korf I., 2004 Gene finding in novel genomes. BMC Bioinformatics 5: 59.
549 https://doi.org/10.1186/1471-2105-5-59
550 Kumar S., G. Stecher, M. Li, C. Knyaz, and K. Tamura, 2018 MEGA X: Molecular
551 Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution
552 35: 1547–1549. https://doi.org/10.1093/molbev/msy096
553 Kurstjens S. P., H. R. Hepburn, F. R. L. Schoening, and B. C. Davidson, 1985 The conversion
554 of wax scales into comb wax by African honeybees. J Comp Physiol B 156: 95–102.
555 https://doi.org/10.1007/BF00692930
556 Kurtovic A., A. Widmer, and B. J. Dickson, 2007 A single class of olfactory neurons mediates
557 behavioural responses to a Drosophila sex pheromone. Nature 446: 542–546.
558 https://doi.org/10.1038/nature05672
559 Lacaille F., M. Hiroi, R. Twele, T. Inoshita, D. Umemoto, *et al.*, 2007 An Inhibitory Sex
560 Pheromone Tastes Bitter for Drosophila Males. PLOS ONE 2: e661.
561 https://doi.org/10.1371/journal.pone.0000661
562 Laetsch D. R., and M. L. Blaxter, 2017 BlobTools: Interrogation of genome assemblies
563 Landman N. H., R. H. B. Lraaije, S. M. Klofak, N. L. Larson, G. A. Bishop, *et al.*, 2014
564 Inquilinism of a Baculite by a Dynomenid Crab from the Upper Cretaceous of South Dakota.
565 novi 2014: 1–16. https://doi.org/10.1206/3818.1
566 Letunic I., and P. Bork, 2019 Interactive Tree Of Life (iTOL) v4: recent updates and new
567 developments. Nucleic Acids Research 47: W256–W259. https://doi.org/10.1093/nar/gkz239
568 Leung N. Y., and C. Montell, 2017 Unconventional Roles of Opsins. Annu. Rev. Cell Dev.

569 Biol. 33: 241–264. https://doi.org/10.1146/annurev-cellbio-100616-060432
570 Leung N. Y., D. P. Thakur, A. S. Gurav, S. H. Kim, A. Di Pizio, *et al.*, 2020 Functions of
571 Opsins in Drosophila Taste. Current Biology 30: 1367-1379.e6.
572 https://doi.org/10.1016/j.cub.2020.01.068
573 Li H., 2018 Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094–
574 3100. https://doi.org/10.1093/bioinformatics/bty191
575 Lopez-Osorio F., A. Perrard, K. M. Pickett, J. M. Carpenter, and I. Agnarsson, 2015
576 Phylogenetic tests reject Emery's rule in the evolution of social parasitism in yellowjackets and
577 hornets (Hymenoptera: Vespidae, Vespinae). Royal Society Open Science 2: 150159.
578 https://doi.org/10.1098/rsos.150159
579 Lu H., F. Giordano, and Z. Ning, 2016 Oxford Nanopore MinION Sequencing and Genome
580 Assembly. Genomics, Proteomics & Bioinformatics 14: 265–279.
581 https://doi.org/10.1016/j.gpb.2016.05.004
582 Luczkovich J. J., G. M. Watters, and B. L. Olla, 1991 Seasonal Variation in Usage of a Common
583 Shelter Resource by Juvenile Inquiline Snailfish (Liparis inquilinus) and Red Hake (Urophycis
584 chuss). Copeia 1991: 1104–1109. https://doi.org/10.2307/1446107
585 Machado A. M., M. G. Miguel, M. Vilas-Boas, and A. C. Figueiredo, 2020 Honey Volatiles as
586 a Fingerprint for Botanical Origin—A Review on their Occurrence on Monofloral Honeys.
587 Molecules 25: 374. https://doi.org/10.3390/molecules25020374
588 Mai D., M. J. Nalley, and D. Bachtrog, 2020 Patterns of Genomic Differentiation in the
589 Drosophila nasuta Species Complex. Molecular Biology and Evolution 37: 208–220.
590 https://doi.org/10.1093/molbev/msz215
591 Martin S. J., and J. Bayfield, 2014 Is the bee louse Braula coeca (Diptera) using chemical
592 camouflage to survive within honeybee colonies? Chemoecology 24: 165–169.
593 https://doi.org/10.1007/s00049-014-0158-1
594 Matsunaga T., C. E. Reisenman, B. Goldman-Huertas, P. Brand, K. Miao, *et al.*, 2022 Evolution
595 of Olfactory Receptors Tuned to Mustard Oils in Herbivorous Drosophilidae. Molecular
596 Biology and Evolution 39: msab362. https://doi.org/10.1093/molbev/msab362
597 Mishra A. K., C. Fritsch, R. Voutev, R. S. Mann, and S. G. Sprecher, 2021 Homothorax controls
598 a binary Rhodopsin switch in Drosophila ocelli. PLOS Genetics 17: e1009460.
599 https://doi.org/10.1371/journal.pgen.1009460
600 Misof B., S. Liu, K. Meusemann, R. S. Peters, A. Donath, *et al.*, 2014 Phylogenomics resolves
601 the timing and pattern of insect evolution. Science 346: 763–767.
602 https://doi.org/10.1126/science.1257570

603 Moon S. J., Y. Lee, Y. Jiao, and C. Montell, 2009 A Drosophila Gustatory Receptor Essential
604 for Aversive Taste and Inhibiting Male-to-Male Courtship. Current Biology 19: 1623–1627.
605 https://doi.org/10.1016/j.cub.2009.07.061
606 Moser J. C., 1964 Inquiline Roach Responds to Trail-Marking Substance of Leaf-Cutting Ants.
607 Science 143: 1048–1049. https://doi.org/10.1126/science.143.3610.1048
608 Müggenburg F. H., 1892 *Der Rüssel der Diptera pupipara*. Nicolai.
609 Muller H., D. Ogereau, J.-L. Da Lage, C. Capdevielle, N. Pollet, *et al.*, 2021 Draft nuclear
610 genome and complete mitogenome of the Mediterranean corn borer, Sesamia nonagrioides, a
611 major pest of maize. G3 Genes|Genomes|Genetics 11: jkab155.
612 https://doi.org/10.1093/g3journal/jkab155
613 Münch D., and C. G. Galizia, 2016 DoOR 2.0 - Comprehensive Mapping of *Drosophila*
614 *melanogaster* Odorant Responses. Scientific Reports 6: 21841.
615 https://doi.org/10.1038/srep21841
616 Ni J. D., L. S. Baik, T. C. Holmes, and C. Montell, 2017 A rhodopsin in the brain functions in
617 circadian photoentrainment in Drosophila. Nature 545: 340–344.
618 https://doi.org/10.1038/nature22325
619 Ortiz M. F., G. L. Wallau, D. A. Graichen, and E. L. Loreto, 2015 An evaluation of the
620 ecological relationship between Drosophila species and their parasitoid wasps as an opportunity
621 for horizontal transposon transfer. Molecular Genetics and Genomics 290: 67–78.
622 https://doi.org/10.1007/s00438-014-0900-y
623 Partha R., B. K. Chauhan, Z. Ferreira, J. D. Robinson, K. Lathrop, *et al.*, 2017 Subterranean
624 mammals show convergent regression in ocular genes and enhancers, along with adaptation to
625 tunneling, (D. T. Odom, Ed.). eLife 6: e25884. https://doi.org/10.7554/eLife.25884
626 Policarpo M., J. Fumey, P. Lafargeas, D. Naquin, C. Thermes, *et al.*, 2021 Contrasting Gene
627 Decay in Subterranean Vertebrates: Insights from Cavefishes and Fossorial Mammals.
628 Molecular Biology and Evolution 38: 589–605. https://doi.org/10.1093/molbev/msaa249
629 Ranallo-Benavidez T. R., K. S. Jaron, and M. C. Schatz, 2020 GenomeScope 2.0 and
630 Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun 11: 1432.
631 https://doi.org/10.1038/s41467-020-14998-3
632 Robertson H. M., and K. W. Wanner, 2006 The chemoreceptor superfamily in the honey bee,
633 Apis mellifera: Expansion of the odorant, but not gustatory, receptor family. Genome Res. 16:
634 1395–1403. https://doi.org/10.1101/gr.5057506
635 Roff D. A., 1990 The Evolution of Flightlessness in Insects. Ecological Monographs 60: 389–
636 421. https://doi.org/10.2307/1943013

637 Rogers H. H., and S. Griffiths-Jones, 2012 Mitochondrial Pseudogenes in the Nuclear Genomes
638 of Drosophila. PLOS ONE 7: e32593. https://doi.org/10.1371/journal.pone.0032593
639 Romiguier J., J. Rolland, C. Morandin, and L. Keller, 2018 Phylogenomics of palearctic
640 Formica species suggests a single origin of temporary parasitism and gives insights to the
641 evolutionary pathway toward slave-making behaviour. BMC Evolutionary Biology 18: 40.
642 https://doi.org/10.1186/s12862-018-1159-4
643 Ronquist F., 1994 Evolution of Parasitism Among Closely Related Species: Phylogenetic
644 Relationships and the Origin of Inquilinism in Gall Wasps (hymenoptera, Cynipidae).
645 Evolution 48: 241–266. https://doi.org/10.1111/j.1558-5646.1994.tb01310.x
646 Sanver D., and B. A. Hawkins, 2000 Galls as habitats: the inquiline communities of insect galls.
647 Basic and Applied Ecology 1: 3–11. https://doi.org/10.1078/1439-1791-00001
648 Savolainen R., and K. Vepsäläinen, 2003 Sympatric speciation through intraspecific social
649 parasitism. Proceedings of the National Academy of Sciences 100: 7169–7174.
650 https://doi.org/10.1073/pnas.1036825100
651 Scanlan J. L., R. S. Gledhill-Smith, P. Battlay, and C. Robin, 2020 Genomic and transcriptomic
652 analyses in Drosophila suggest that the ecdysteroid kinase-like (EcKL) gene family encodes
653 the 'detoxification-by-phosphorylation' enzymes of insects. Insect Biochemistry and Molecular
654 Biology 123: 103429. https://doi.org/10.1016/j.ibmb.2020.103429
655 Schrader L., H. Pan, M. Bollazzi, M. Schiøtt, F. J. Larabee, *et al.*, 2021 Relaxed selection
656 underlies genome erosion in socially parasitic ant species. Nat Commun 12: 2918.
657 https://doi.org/10.1038/s41467-021-23178-w
658 Seeley T. D., 1974 Atmospheric carbon dioxide regulation in honey-bee (Apis mellifera)
659 colonies. Journal of Insect Physiology 20: 2301–2305. https://doi.org/10.1016/0022-
660 1910(74)90052-3
661 Senthilan P. R., and C. Helfrich-Förster, 2016 Rhodopsin 7–The unusual Rhodopsin in
662 Drosophila. PeerJ 4: e2427. https://doi.org/10.7717/peerj.2427
663 Sherman B. T., M. Hao, J. Qiu, X. Jiao, M. W. Baseler, *et al.*, 2022 DAVID: a web server for
664 functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic
665 Acids Research 50: W216–W221. https://doi.org/10.1093/nar/gkac194
666 Skaife S. H., 1922 On *Braula Coeca* , Nitzsch, a Dipterous parasite of the honey bee.
667 Transactions of the Royal Society of South Africa 10: 41–48.
668 https://doi.org/10.1080/00359192209519263
669 Stange G., and M. Diesendorf, 1973 The response of the honeybee antennal CO2-receptors to
670 N2O and Xe. J. Comp. Physiol. 86: 139–158. https://doi.org/10.1007/BF00702534

671 Starowicz M., P. Hanus, G. Lamparski, and T. Sawicki, 2021 Characterizing the Volatile and
672 Sensory Profiles, and Sugar Content of Beeswax, Beebread, Bee Pollen, and Honey. Molecules
673 26: 3410. https://doi.org/10.3390/molecules26113410
674 Suvorov A., B. Y. Kim, J. Wang, E. E. Armstrong, D. Peede, *et al.*, 2021 Widespread
675 introgression across a phylogeny of 155 Drosophila genomes. Current Biology 0.
676 https://doi.org/10.1016/j.cub.2021.10.052
677 Torto B., D. G. Boucias, R. T. Arbogast, J. H. Tumlinson, and P. E. A. Teal, 2007 Multitrophic
678 interaction facilitates parasite–host relationship between an invasive beetle and the honey bee.
679 Proceedings of the National Academy of Sciences 104: 8374–8378.
680 https://doi.org/10.1073/pnas.0702813104
681 Venner S., V. Miele, C. Terzian, C. Biémont, V. Daubin, *et al.*, 2017 Ecological networks to
682 unravel the routes to horizontal transposon transfers. PLOS Biology 15: e2001536.
683 https://doi.org/10.1371/journal.pbio.2001536
684 Weinstock G. M., G. E. Robinson, R. A. Gibbs, G. M. Weinstock, G. M. Weinstock, *et al.*,
685 2006 Insights into social insects from the genome of the honeybee Apis mellifera. Nature 443:
686 931–949. https://doi.org/10.1038/nature05260
687 Weiss L. A., A. Dahanukar, J. Y. Kwon, D. Banerjee, and J. R. Carlson, 2011 The Molecular
688 and Cellular Basis of Bitter Taste in Drosophila. Neuron 69: 258–272.
689 https://doi.org/10.1016/j.neuron.2011.01.001
690 Winkler I. S., A. H. Kirk-Spriggs, K. M. Bayless, J. Soghigian, R. Meier, *et al.*, 2022
691 Phylogenetic resolution of the fly superfamily Ephydroidea–Molecular systematics of the
692 enigmatic and diverse relatives of Drosophilidae. PLOS ONE 17: e0274292.
693 https://doi.org/10.1371/journal.pone.0274292
694 Winston M. L., 1987 *The Biology of the Honey Bee*. Harvard University Press.
695 Yassin A., 2013 Phylogenetic classification of the Drosophilidae Rondani (Diptera): the role of
696 morphology in the postgenomic era. Systematic Entomology 38: 349–364.
697 Zanini D., D. Giraldo, B. Warren, R. Katana, M. Andrés, *et al.*, 2018 Proprioceptive Opsin
698 Functions in Drosophila Larval Locomotion. Neuron 98: 67-74.e4.
699 https://doi.org/10.1016/j.neuron.2018.02.028
700 Zhang H.-H., J. Peccoud, M.-R.-X. Xu, X.-G. Zhang, and C. Gilbert, 2020 Horizontal transfer
701 and evolution of transposable elements in vertebrates. Nat Commun 11: 1362.
702 https://doi.org/10.1038/s41467-020-15149-4
703 Zimin A. V., D. Puiu, M.-C. Luo, T. Zhu, S. Koren, *et al.*, 2017 Hybrid assembly of the large
704 and highly repetitive genome of Aegilops tauschii, a progenitor of bread wheat, with the

MaSuRCA mega-reads algorithm. Genome Res. 27: 787–792. https://doi.org/10.1101/gr.213405.116

708 **Legends of figures**

709

710 **Figure 1** – The bee louse fly (*Braula coeca*) morphology and genome. A) Two adults
711 attached to a honey bee worker preserved in alcohol. B) Dorsal view of an adult showing
712 the loss of the wings, halters and scutum, the reduction of the mesonotum and the
713 robustness of the legs. C) Frontal view of an adult showing the reduction of the eyes and
714 the loss of the ocelli. Scale bars = 1 mm in A and 0.5 mm in B and C. D) Orthologous genes
715 colocalization in *D. melanogaster* Muller's elements and *B. coeca* contigs demonstrating
716 the predominant conservation of synteny. E) The mitochondrial genome of *B. coeca*. Blue
717 and violet ribbons indicate the protein-encoding genes in sens and anti-sens respectively,
718 red and yellow ribbons indicate the rRNA and tRNA genes in sens and anti-sens
719 respectively. Names with asterix indicate the presence of inactivated paraloguous genes.

720

721 **Figure 2** – Maximum-likelihood phylogenetic tree for A) odorant receptors (ORs) and B)
722 gustatory receptors (GRs) of *A. mellifera* (green), *B. coeca* (red) and *D. melanogaster*
723 (blue). Genes commented in the text are labelled. For full labels see Supplementary
724 Figures 4 and 5. Sequences for *B. coeca* are given in Supplementary Datasets 2 and 3 for
725 ORs and GRs, respectively. Sequences for *A. mellifera* and *D. melanogaster* are from
726 Robertson and Wanner (2006).

727

728

729 **List of supplementary materials**

730

731 **Supplementary Dataset 1** – Sequence of all *Famar* and *Famar1*-like copies used to
732 reconstruct the phylogeny shown in Supplementary Figure 3. The sequence names
733 contain the name of the species, the Genbank accession number of the contig from which
734 the copy was extracted, the start and end position of the copy in the contig, as well as a
735 final number that allows making a correspondence with leaves in the tree shown in
736 Supplementary Figure 3.

737

738 **Supplementary Dataset 2** – Sequence of odorant receptors (ORs) of *Braula coeca* used
739 to reconstruct the phylogeny shown in Supplementary Figure 4.

740

741 **Supplementary Dataset 3** – Sequence of gustatory receptors (GRs) of *Braula coeca* used
742 to reconstruct the phylogeny shown in Supplementary Figure 5.

743

744 **Supplementary Figure 1** – Genomic features of *Braula coeca.* A) Ploidy estimation using
745 the coverage and the distribution of heterozygous k-mers pairs. B) Genome size
746 estimation using the k-mer profile spectrum. C) Taxonomic assignation of the contigs
747 according to their GC% and their coverages. D) Proportion of short-reads that mapped
748 onto the genome assembly (left) and onto the different contigs according to their
749 taxonomic assignments (right).

750

751 **Supplementary Figure 2** – Comparison of Famar1-like synonymous distance and
752 orthologous gene synonymous distances between *Braula coeca* and *Apis mellifera*. To
753 calculate the distribution of gene or gene fragment synonymous distances (dS) between
754 *B. coeca* and *A. mellifera* we selected best reciprocal blastp hits between single copy
755 BUSCO genes retrieved from the two species and calculated dS for each of them using the
756 approach described in Zhang *et al.* (2020). The red line indicates the 0.5% quantile of this
757 distribution (=1.76). The distribution is bimodal, with genes having highly saturated dS
758 values showing a peak centered on 9.99 and genes showing less saturated dS values
759 showing another peak around 2.5. We verified that genes showing less saturated values
760 correspond to highly genes that evolve under strong purifying selection and are thus
761 highly conserved between Hymenoptera and Diptera. The *Famar1*-like dS (green line, =

762 0.12) was calculated over the transposase open reading from of one copy of the element
763 extracted from the *A. mellifera* genome and another copy extracted from the *B. coeca*
764 genome.
765
766 **Supplementary Figure 3** – Phylogeny of *Famar1*-like copies from 38 animal species. The
767 ten *Famar1*-like copies showing the highest nucleotide identity to the *Famar* element
768 initially described in the earwig (Barry *et al.* 2004b) were retrieved using online blastn
769 (see Supplementary Text 1) and extracted from the genome of 37 animal species. Filled
770 circles indicate bootstrap value higher than 70%, with the diameter of the circle
771 proportional to each individual value. Name of the copies are composed of the name of
772 the species from which they were extracted and a unique number that allows making a
773 correspondence with sequences provided in Supplementary Dataset 1.
774
775 **Supplementary Figure 4** – Labelled maximum-likelihood phylogeny of odorant
776 receptors (ORs) of *A. mellifera* (green), *B. coeca* (blue), and *D. melanogaster* (red) show in
777 Figure 2A.
778
779 **Supplementary Figure 5** – Labelled maximum-likelihood phylogeny of gustatory
780 receptors (GRs) of *A. mellifera* (green), *B. coeca* (blue), and *D. melanogaster* (red) show in
781 Figure 2B.
782
783 **Supplementary Table 1** – Proportions of transposable elements in the genomes of *B.*
784 *coeca*, *D. melanogaster* and *A. mellifera*.
785
786 **Supplementary Table 2** – Gene Ontology (GO) enrichment for *D. melanogaster* genes that
787 are absent from *B. coeca* genome as inferred using DAVID 2021.
788
789 **Supplementary Table 3** – Gene Ontology (GO) enrichment for *D. melanogaster* genes that
790 have additional copies in *B. coeca* genome as inferred using DAVID 2021.
791
792 **Supplementary Text 1** – Analysis of horizontal transfer of transposable elements
793 between *B. coeca* and *A. mellifera*.
794

Figure 1

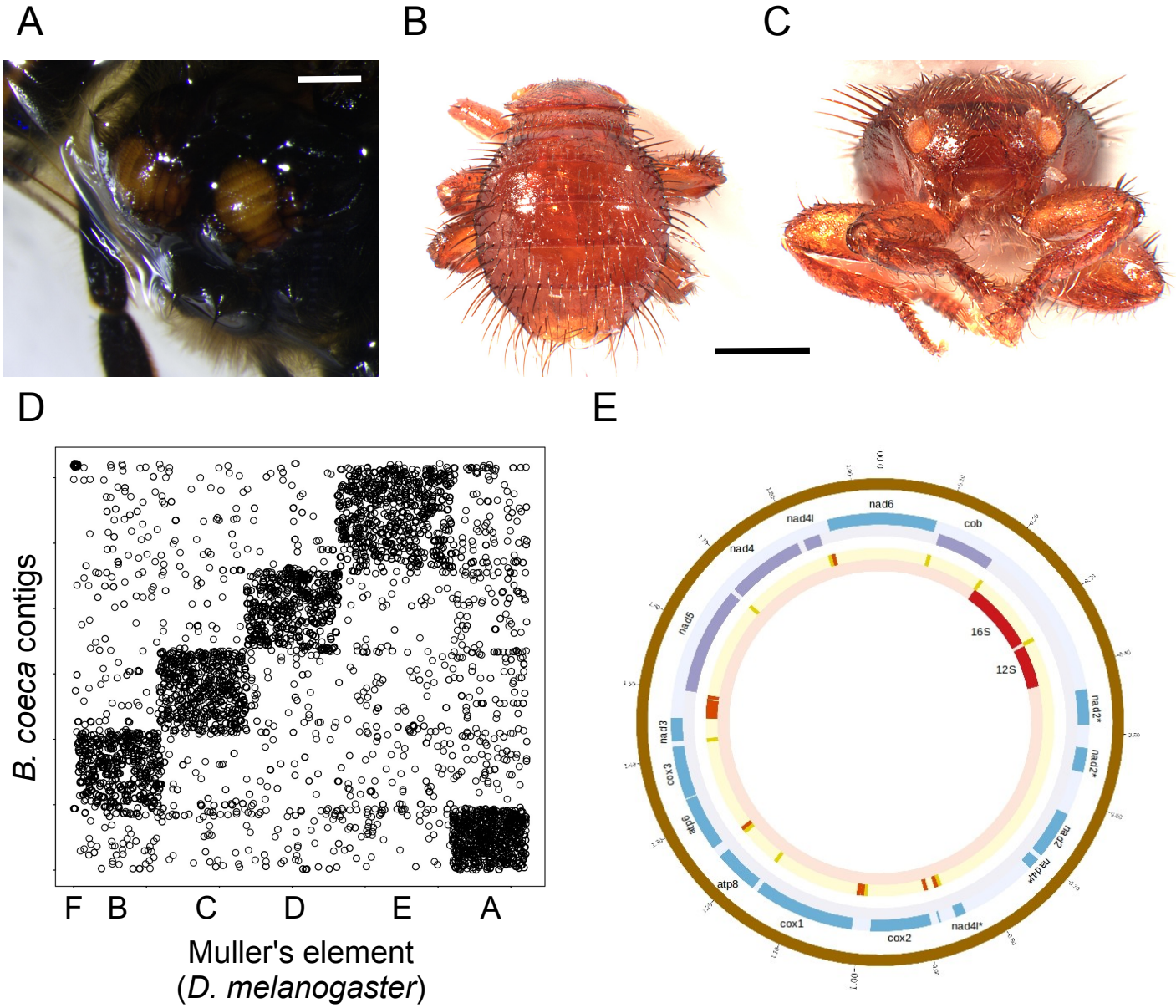
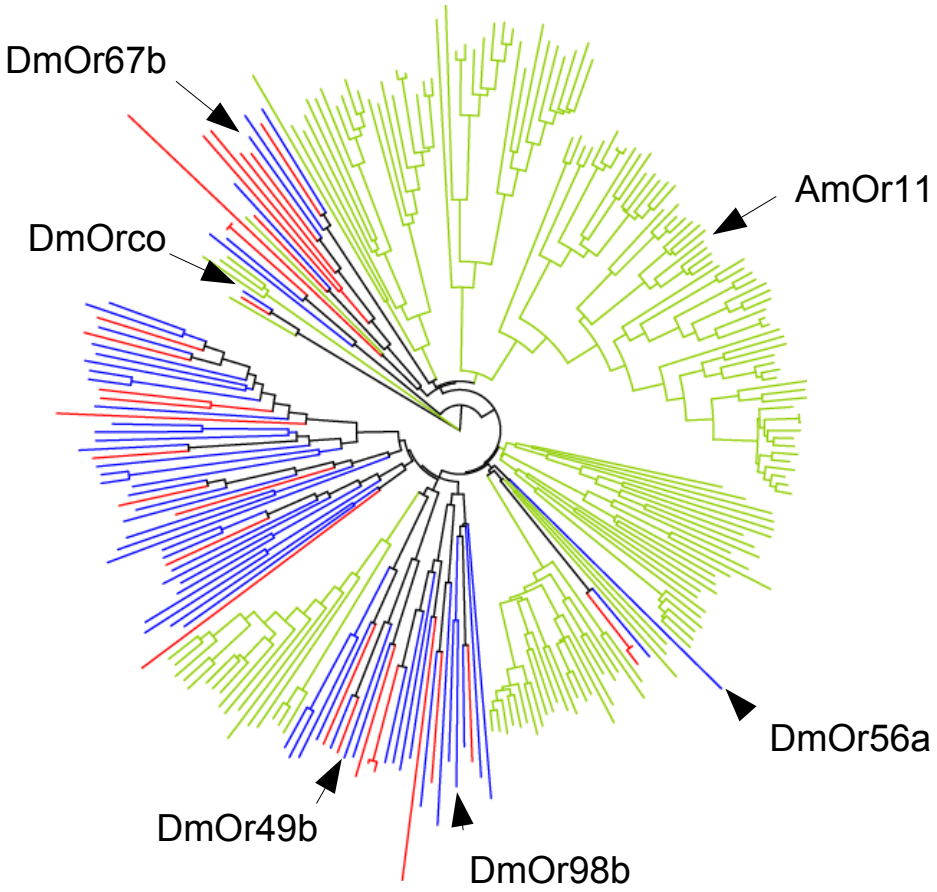
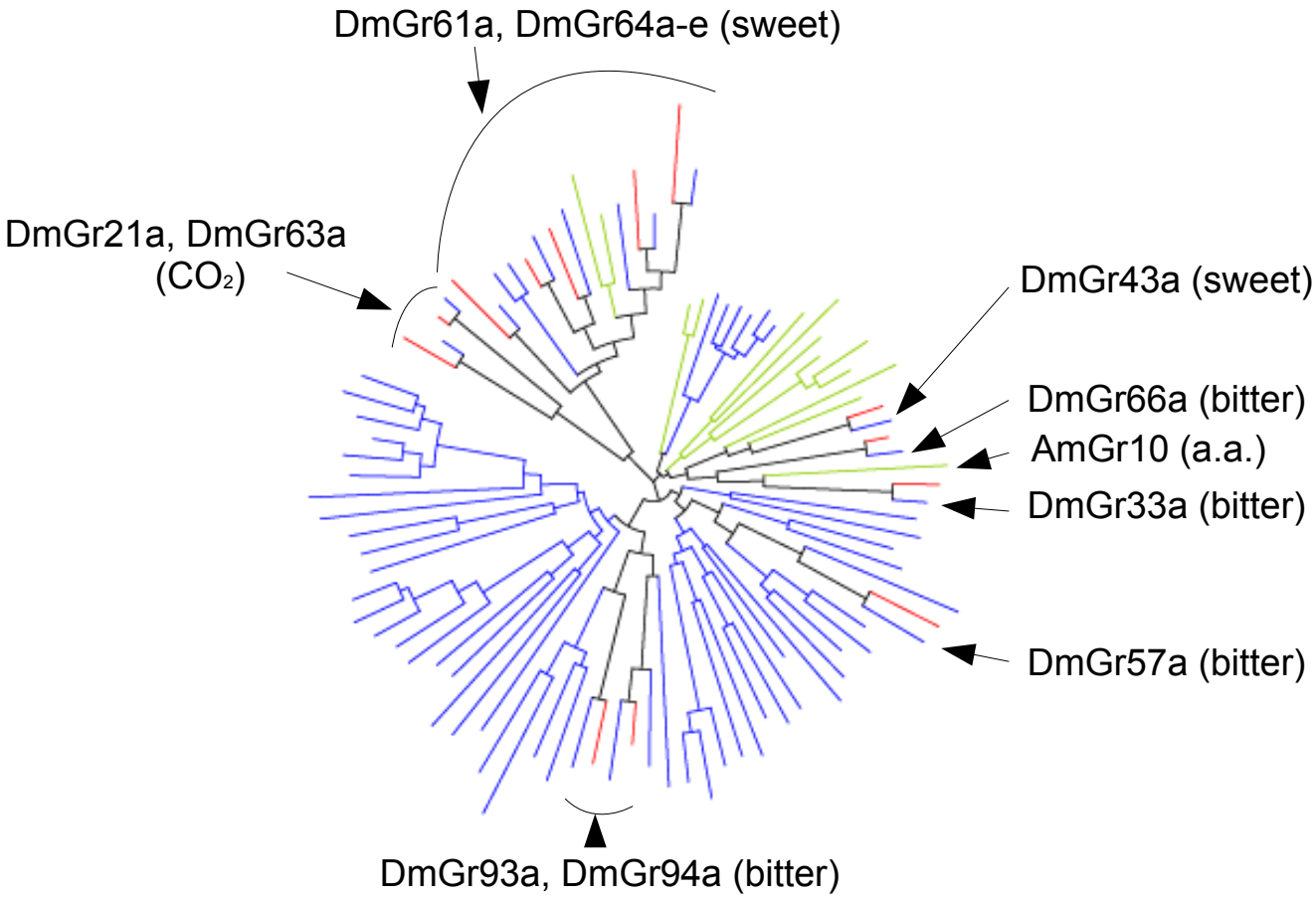

A
B
C
D
E
B. coeca contigs
F
B
C
D
E
A
Muller's element
(D. melanogaster)
nad6
cob
nad4l
nad4
nad5
16S
12S
nad2*
nad2*
nad2
nad4l*
nad4l*
nad3
cox3
atp6
atp8
cox1
cox2
nad4l*

Figure 2

A


DmOr67b
AmOr11
DmOrco
DmOr56a
DmOr49b
DmOr98b

B


DmGr61a, DmGr64a-e (sweet)
DmGr21a, DmGr63a
(CO_2)
DmGr43a (sweet)
DmGr66a (bitter)
AmGr10 (a.a.)
DmGr33a (bitter)
DmGr57a (bitter)
DmGr93a, DmGr94a (bitter)