Report

The genome of the blind bee louse fly reveals deep convergences with its social host and illuminates Drosophila origins

Héloïse Bastide[1,*], Hélène Legout[1], Noé Dogbo[1], David Ogereau[1], Carolina Prediger[1], Julie Carcaud[1], Jonathan Filée[1], Lionel Garnery[1], Clément Gilbert[1], Frédéric Marion-Poll[1,2], Fabrice Requier[1], Jean-Christophe Sandoz[1], Amir Yassin[1]

[1]Université Paris-Saclay, CNRS, IRD, UMR Évolution, Génomes, Comportement et Écologie, 91198, Gif-sur-Yvette, France.

[2]Université Paris-Saclay, AgroParisTech, Paris, France

[*]**Correspondence:** Héloïse Bastide (heloise.bastide@universite-paris-saclay.fr)

Summary

The nests of social insects often harbor a rich fauna of intruders known as inquilines.[1] Social inquilines are usually closely-related to their host due to potential genetic predispositions,[2,3] but how phylogenetically distant non-social inquilines adapt to their hosts remains unclear. Here, we analyzed the genome of the wingless and blind bee louse fly *Braula coeca*, an inquiline kleptoparasite of the Western honey bee *Apis mellifera*.[4,5] Using large phylogenomic data, we confirmed recent accounts that the bee louse fly is an aberrant drosophilid,[6,7] and showed that it had likely evolved from a sap-breeder ancestor associated with honeydew and scale insects wax. Unlike many parasites, such as the human louse, the bee louse fly genome did not show significant erosion or strict reliance on an endosymbiont, likely due to a relatively recent age of inquilinism. However, a striking parallel evolution in a set of gene families was observed between the honey bee and the bee louse fly. 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 nest and a major diet of honey, pollen, and beeswax. Vision-related and odorant receptor genes also exhibited rapid losses. Only genes whose orthologs in the closely related *Drosophila melanogaster* respond to honey bee pheromones components or floral aroma were retained, whereas the losses included orthologous receptors responsive to the anti-ovarian honey bee queen pheromones. These results indicate that deep genomic convergences can underlie major morphological and neuroethological transitions during the evolution of inquilinism between non-social parasites and their social hosts.

Keywords: parasitism; morphological evolution; behavioral adaptations; gene family evolution; Drosophilidae phylogeny.

Results and Discussion

The bee louse fly Braula coeca *is an aberrant member of the Drosophilidae*

Among the several parasites and inquilines that are attracted by the rich resources and clean and protective shelter of the Western honey bee *Apis mellifera* nest, none has undergone as profound morphological changes 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 burrow tunnels in which they pupate without forming true puparia.[4,5] Following emergence, the adults attach to the body of worker bees, migrating from one individual to another until reaching the queen (Figure 1A). There, they move to the queen's head, stimulate regurgitation, and imbibe from her mouth honey and nectar.[4,5] The bee louse fly is considered an inquiline kleptoparasite with potential negative effects on honey bee colony health due to the galleries it makes in bee combs and the facilitation of transmitting serious pathogenic viruses to the bees.[8]

Ever since Réaumur's first description of the bee louse fly in 1740, and Nitzsch's creation of the genus *Braula* in 1818 (Document S1),[9,10] the positioning within the Diptera of the bee louse fly and affiliated species that were classified under the family Braulidae has been puzzling because of its aberrant morphology and unique adaptations to a social host. This family contains seven species belonging to the genera *Braula* and *Megabraula* that are all inquilines to honey bee species of the genus *Apis*. Recent phylogenetic analyses based on a transcriptome assembled from one adult fly and using 1,130 loci interestingly showed *Braula coeca*, the most widespread braulid, to constitute a basal lineage within the Drosophilidae that was sister to four genera of the subfamily Steganinae.[6,7] To reassess this hypothesis using a larger dataset, we sequenced the whole genome from a pooled sample of 15 unsexed *B. coeca* flies, all collected on Ouessant Island in Western France. We used a hybrid approach to assemble a draft genome using long-read Oxford Nanopore Technology (ONT) and short-read Illumina sequencing (see Methods). Benchmarking Universal Single-Copy Orthologs (BUSCO)[11] gave a score of 95.8% of the Dipteran conserved single-copy orthologs with 1.3% of duplicated genes. Merqury[12] estimated an assembly consensus quality value (QV) of 41, which exceeds the recommended threshold of QV40 for reference genome.[13] We assembled two genomes and one transcriptome of three additional steganine genera. We then built a supermatrix of 3,100 BUSCO genes (2,557,349 amino acids) that included 15 drosophilid species (representative members of the four main radiations in the family),[14] and 5 species

77 belonging to the superfamily Ephydroidea to which both the Drosophilidae and Braulidae
78 belong (Data S1). The maximum-likelihood phylogenetic analysis of this large dataset
79 reconfirmed the close-relatedness of *B. coeca* to the Drosophilidae. It further showed that
80 it is a full member of the subfamily Steganinae (Figure 1D). The taxonomic priority
81 principle should consider the family Drosophilidae, described in 1856,[15] a junior synonym
82 for the family Braulidae, described in 1853.[16] However, the asymmetric size and scientific
83 relevance of the two families argue against such a decision. We, therefore, opt for
84 synonymizing the Braulidae with the Drosophilidae, referring hereafter to *Braula* and
85 *Megabraula* as members of the subfamily Steganinae.
86
87 *Inquilinism in the bee louse fly likely evolved from sap breeders associated with scale insects*
88 To gain further insight into the history of the association between *Braula* and *Apis*,
89 we mapped the predominant ecological habitats of ephydroid families on the phylogeny.
90 The ancestral habitat of ephydroids was presumed to be rotting leaf molds.[17] From this,
91 multiple specializations took place, including the exploitation of aquatic molds (and
92 eventually algae) in the Ephydridae,[18] Mammal dung in Curtonotidae and Diastatidae,[18,19]
93 and fermenting vegetables and fruits, sap and fungi, with specialization mostly on yeasts
94 in the Drosophilidae[20] (Figure 1D). Bayesian reconstruction suggests the ancestral habitat
95 of the Drosophilidae to be tree sap breeding (Figure 1D) with fungus- and fruit-breeding
96 subsequently deriving and predominating in the genera *Leucophenga* and *Drosophila*,
97 respectively. Remarkably, the deepest branches in the Steganinae and the
98 Cryptochaetidae (the closest relative to the Drosophilidae) represent lineages whose
99 larvae are predatory of scale insects and mealy bugs, e.g., *Acletoxenus formosus* and *A.*
100 *indicus* on aleyrodoids, *Rhinoleucophenga brasiliensis* and *R. obesa* as well as
101 *Cryptochaetidum iceryae* and *C. grandicorne* on coccoids.[21] In those lineages, adults are
102 often seen to feed on the honeydew produced by the bugs, an abundant sugar-rich
103 substrate sucked from plants' sap, while larvae take shelter and develop in the waxy
104 secretions of these insects. This dependence on sugary substrate (honeydew) and
105 development in a waxy environment could have predisposed *Braula*'s inquilinism in bee
106 nests.
107
108 *The bee louse fly inquilinism is relatively recent*

109 To date *Braula* inquilinism, we inferred a fossil-calibrated phylogeny using 79
110 single-copy orthologs (63,192 amino acids) in 17 Acalyptrate dipteran and 25 Apocrite
111 Hymenopteran species (see Methods; Data S1; Figure 2A). Five non-ephydroid dipteran
112 species with Ref-Seq assemblies were included in this analysis to correct for tree
113 imbalance[22]. The divergence between *B. coeca* from its closest steganine relatives (1 in
114 Figure 2A: 44.9 [37.8-53.8] million years (myr) [95% confidence interval]) overlapped
115 with the origin of the Apidae (50.12 [42.9-64.5] myr) and with the transition from solitary
116 to subsocial (2: 40.25 [32.4-47.8] myr) and primitively social habits (3: 30.98 [26.3-36.0]
117 myr).[23] It is possible that the origin of the bee louse fly-apid interactions occurred at sap
118 breeding sites, when early subsocial apids started to gather resin and other plant
119 exudates, as well as scale insects honeydew, and stored them in their nests. As eusociality
120 evolved (4: 23.8 [18.9-28.2] myr), the proportion of resin to secreted wax diminished, and
121 some cells were also used to store nectar and honey for the brood.[24] A shift from the
122 putatively ancestral dependence on honey and wax produced by scale insects to those
123 produced by bees might have evolved by then. The transition to eusociality in the genus
124 *Apis* required an important division of labor that involved the evolution of pheromonal
125 control of the reproductive capacity of worker females by the queen and the evolution of
126 trophallaxis.[24] Adaptation of *Braula* to the queen pheromone compounds that have anti-
127 ovarian effects on a wide range of insects, including *Drosophila melanogaster*[25] and the
128 exploitation of trophallaxis[5] could not have evolved before the advancement of eusociality
129 (5: 17.1 [11.6-23.0] myr). The evolution of blindness and apterism should have
130 constrained the dispersal of the bee lice, relating their speciation history to that of their
131 hosts. Indeed, only seven bee louse fly species are known, of which five *Braula* species are
132 restricted to the Western honey bee *A. mellifera*, and two *Megabraula* species are
133 restricted to the giant honey bee *A. laboriosa* in the Himalayas.[26,27] The divergence
134 between these *Apis* species, and presumably between *Braula* and *Megabraula*, is
135 estimated at 6: 5.8 [2.8 -12.0] myr ago. Therefore, the evolution of the bee louse fly
136 inquilinism has likely taken place during the Mid- to Late Miocene period between 5.8 and
137 17.1 myr ago (Figure 2A). We cannot rule out even a more recent origin if the ancestor of
138 *Braula* or *Megabraula* has shifted from one *Apis* host to another, *i.e.* <5.8 myr ago.
139
140 *The bee louse fly inquilinism was accompanied by a reduction of gene content but not*
141 *genome size*

142 Loss of significant portions of genomic and gene contents is a characteristic of
143 obligate parasites specializing on specific hosts or inhabiting extreme environments. For
144 example, the human body louse, *Pediculus humanus*, has one of the smallest genomes and
145 the lowest numbers of genes in insects (108 megabases [Mb] and 10,773 protein-coding
146 genes).[28] For *B. coeca*, we obtained a final assembly size of 309.35 Mb shared by 2,477
147 contigs with an N50 of 347,211 bp. This N50 estimate is typical of hybrid genome
148 assemblies obtained using a pooled sample of wild-caught drosophilid flies from species
149 with large genome sizes (> 300 Mb).[29] No evidence for polyploidy or other endosymbiont
150 that could have biased the genome size estimate was detected (Figure S1). Genome size
151 prediction using k-mers distribution spectra predicted a genome of 308 Mb, concordant
152 with the assembly size (Figure S1). Such a genome size is significantly larger than the
153 remaining drosophilid species (Student's *t* one-sample test $P < 1.3 \times 10^{-4}$). Phylogenetic
154 analysis of genome size evolution indicates that the *B. coeca* genome likely retained the
155 size of the ancestral Steganinae, *i.e.* a stronger reduction occurred in the Drosophilinae
156 lineage containing *D. melanogaster* (Figure 2B; Figure S2; Data S1).
157 To determine the number of protein-coding genes, we used four rounds of Maker[30]
158 supported by the training of the gene finding and prediction tools SNAP[31] and Augustus[32].
159 The annotation, made on the repeat-masked genome, yielded 10,349 protein-coding
160 genes with an Annotation Edit Distance (AED) ≤ 0.5 for 96.4% of our gene models and a
161 Pfam domain found in 83.66% of the proteins. Using the same strategy, we annotated two
162 steganine genomes, namely *Phortica variegata* and *Leucophenga varia*. The annotation
163 yielded 11,067 (BUSCO score = 91%) and 13,160 (BUSCO score = 90.8%) protein-coding
164 genes, respectively. The annotation of the ephydrid *Ephydra gracilis* genome yielded
165 9,154 protein-coding genes (BUSCO score = 68.9%) (Figure S2). *Ephydra* is particular
166 among Ephydroidea in adapting to hypersaline waters and associated algal flora.[33] Given
167 the current low knowledge of ephydrid genetics, whether their low gene content is due to
168 their high specialization or an artifact of incomplete annotation, is hard to know.
169 Regardless, the bee louse fly has the lowest number of protein-coding-genes compared to
170 other drosophilids (Student's *t* one-sample test $P < 1.5 \times 10^{-5}$) despite having a total
171 genome size that is among the largest genomes in the family. Remarkably, a low gene
172 content is characteristic of bee genomes, compared to ants and wasps, with a remarkable
173 trend of gene reduction within the family Apidae during the evolution of the genus *Apis*
174 (Figure 2C; Figure S2; Data S1).

175

176 *Transposable elements (TEs) expanded in the bee louse fly with one element horizontally*
177 *transferred with the host*

178 The bee louse fly's large genome size and low gene content suggest an increase in
179 repetitive sequences. RepeatModeler and RepeatMasker analyses[34,35] indicated that
180 nearly 41.34% of the *B. coeca* genome consists of such sequences, compared to 22.05%
181 and 10.98% in *D. melanogaster* and *A. mellifera*, respectively (Figure 2D). Remarkably,
182 half of the bee louse fly repetitive sequences consisted of long interspersed nuclear
183 elements (LINEs) retrotransposons (14.94%). While LINEs are usually among the most
184 abundant transposable elements after LTRs within the Drosophilidae,[36] their values did
185 not exceed what was found in *B. coeca* (we found the highest percentage in *Leucophenga*
186 *varia* with 5.54%). It is at present unclear what factors influence the diversity of
187 transposable elements (TEs) landscapes among eukaryote species[37]. Nonetheless, this
188 difference means that whereas the bee louse fly has likely retained the ancestral large
189 genome size of the Drosophilidae, its TEs constitution has largely evolved.

190 Because host-parasite relationships have repeatedly been invoked as a factor that
191 may favor horizontal transfer of TEs,[38] we searched for evidence of such transfers
192 between *B. coeca* and *A. mellifera* (Document S2). We found one TE, a DNA transposon
193 *Famar1-like* element previously described in the earwig *Forficula auricularia*[39] that
194 belongs to the Tc1-mariner superfamily. This element showed a high similarity between
195 *B. coeca* and *A. mellifera* but was absent in all other drosophilid species for which a
196 genome is available in GenBank, which is highly suggestive of an acquisition through
197 horizontal transfer (Figure S3). Indeed, phylogenetic analysis of multiple copies of this TE
198 extracted from 37 widely divergent animal species (Figure 2E) supported a direct transfer
199 event between *B. coeca* and *A. mellifera*, although the directionality of the transfer cannot
200 be inferred since the elements from the two species form mutually-exclusive
201 monophyletic clades. Remarkably, all elements found in the genomes of four *A. mellifera*
202 subspecies, including *A. m. carnica*, *A. m. caucasia*, *A. m. mellifera* and *A. m. ligustica*,
203 formed an exclusively monophyletic clade. The transfer time between *B. coeca* and *A.*
204 *mellifera* has likely preceded the dispersion of this element among the subspecies or even
205 their differentiation 0.77 myr ago[40] if the element was ancestral in *A. mellifera*. On the
206 other hand, we did not find any trace of this element or any other related element in any
207 other *Apis* species, indicating that the maximal time of horizontal transfer likely does not

208 surpass 2.73 [0.70-8.9] myr ago, *i.e.* the time of divergence between *A. mellifera* and its
209 closest-relative *A. cerana* (Figure 2A). The tight ecological connection between the bee
210 louse fly and its host may have favored this transfer, as was suggested for blood- or sap-
211 sucking insects.[41,42]
212
213 *Gene families with excess losses show striking cross-order parallelism*
214 Despite their deep divergence, we tested whether parallel changes could explain
215 the reduction of protein-coding genes in both the honey bee and the bee louse fly. We used
216 OrthoFinder[43] to cluster orthologous proteins from the 25 hymenopteran and 17 dipteran
217 species. We identified 19,010 orthogroups. Of these, 935 showed significant size evolution
218 among the 42 species when analyzed using CAFE5[44] and after applying an error model
219 that accounted for misassemblies and misannotations. To classify those orthogroups into
220 functional categories, we extracted groups that contained *D. melanogaster* orthologs for
221 which a molecular function, *i.e.* a gene group, was assigned in the Flybase database,[45] e.g.,
222 the 60 odorant receptors of *D. melanogaster* clustered in 16 orthogroups. Of 1,078 gene
223 groups, 163 significantly deviated from the birth-death model estimated by CAFE5.
224 Thirty-nine gene groups had significant losses in the bee louse fly. Most of these
225 groups showed a striking parallelism with *Apis* in particular and hymenopterans in
226 general (Figure 3). The most significant groups were those involved in the chemical
227 detection of taste (gustatory receptors and divergent ionotropic receptors) and odors
228 (odorant receptors and odorant binding proteins). The remaining groups included those
229 involved in metabolism and/or immunity, such as C-type lectins (recognition), CLIP
230 serine proteases (signaling), serpin (signaling), Dorsal (signaling), lysozyme (other), as
231 well as in detoxification, such as cytochrome P450, GST-C, carboxylases, ABCC
232 transporters, acyl-co, and alcohol dehydrogenase. Indeed, bees have evolved a reduced
233 repertoire of immunity and detoxification genes, likely due to the evolution of social
234 behavior and their life in an overprotective and clean shelter, *i.e.* the nest.[46,47] Cytochrome
235 P450 genes are more expressed in foraging workers than in the castes that remain in the
236 nest (*i.e.* the queen and nurse workers).[48] This is also the case of alcohol dehydrogenase,[49]
237 an emblematic enzyme in *D. melanogaster* that was lost in *B. coeca*. The reduction of
238 peptidases in both the honey bee and the bee louse fly could also be due to the low protein
239 content of some of their food, *i.e.* nectar and honey. We also noted an underrepresentation
240 of chitin-binding domain proteins and chitinases in the bee louse fly and the honey bee.

241 Cuticles could act as barriers against environmental toxins, which may not be highly
242 encountered in the nest. Remarkably, *B. coeca* is unique among cyclorrhaphan dipterans
243 as its pupa, similarly to the honey bee's,[50] is contained in the unmodified cuticle of the
244 third instar larva, and no sclerotized puparium is formed.[4,5]
245
246 *Honey and wax feeding drove the loss of almost all bitter-tasting gustatory receptors*
247 The two most significantly evolving gene families in the bee louse fly, *i.e.* gustatory
248 receptors (GRs) and divergent ionotropic receptors (IR-DIVs), allow the detection of
249 soluble cues (Figure 3A-B). There are 60 GRs in *D. melanogaster*, of which 9 and 49
250 receptors respond primarily to sweet and bitter tastes, respectively, and 2 receptors
251 respond to carbon dioxide (CO_2).[51] The three categories clustered into 35 orthogroups
252 (Figure 4A), whose phylogenetic analysis indicates that the ancestral drosophilid
253 repertoire consisted of 6 sweet, 29 bitter, and 2 CO_2 GRs assuming functional conservation
254 of gustatory categories (Figure 4A). We identified 11 GRs in the bee louse fly with no
255 duplications using InsectOR[52] and manual curation. These GRs could be classified
256 according to their *D. melanogaster* orthologs into 2 sweet, 7 bitter, and 2 CO_2. That means
257 that the *D. melanogaster* lineage disproportionally evolved more bitter receptors from the
258 ancestral repertoire, whereas *B. coeca* disproportionally lost bitter receptors (Figure 4A).
259 InsectOR inferred the number of GRs in the steganine species *L. varia* and *P. variegata* to
260 be 21 and 26, respectively, further confirming that *B. coeca* has lost a significant portion
261 of the ancestral GR repertoire (Figure S4). Honey bees have only 11 GRs, of which 7 are
262 orthologous to sweet *Drosophila* GRs.[53] This is likely due to the bees' strong diet reliance
263 on sweet floral nectars and honey.[54] The loss of *B. coeca* bitter GRs and its retention of 2
264 ancestral sweet receptors is a strong convergence with its host.
265 Ionotropic receptors are another major class of chemoreceptors. They are divided
266 into antennal IRs, which are conserved across insects and are most likely involved in
267 olfaction, and divergent IRs (IR-DIVs), which evolve rapidly and are mostly involved in
268 the taste perception of carboxylic and amino acids. Only divergent IRs showed a
269 significant loss in *B. coeca* (Figure 3B). However, our knowledge about the function of the
270 42 *D. melanogaster* IR-DIVs is still limited.[55] We inferred the ancestral IR-DIV drosophilid
271 repertoire to contain 29 receptors, of which only 9 were retained in *B. coeca*. Remarkably,
272 whereas we found almost no direct orthologs between Diptera and Hymenoptera for IR-

273 DIVs (Figure 3B), bees are known to have few IRs in general[56] pointing to another possible
274 taste convergence between the bee louse fly and its host.
275
276 *One-fifth of ancestral odorant receptors was lost, including one receptor that is involved in*
277 *anti-ovarian response in* Drosophila melanogaster
278 Odorant receptors (ORs) are essential to detect volatile chemical cues from the
279 environment. This family has expanded in the honey bee to reach 170.[57] However, only 9
280 of the honey bee genes have orthologs with *D. melanogaster*, and phylogenetic analysis
281 indicates that this common OR repertoire has been gradually reduced during the
282 evolution of *Apis* (Figure 3C). The 60 ORs of *D. melanogaster* clustered within 16
283 orthogroups (Figure 4B). We inferred the ancestral drosophilid OR repertoire to contain
284 44 ORs, with at least one representative for each orthogroup (Figure 4B). We identified in
285 *B. coeca*, following InsectOR[52] and manual curation, 35 ORs in addition to *Orco*, *i.e.* one-
286 fifth of the ancestral repertoire was lost. The number of ORs was 50 and 51 in the two
287 closely related steganine species *L. varia* and *P. varia*, respectively. *Braula* ORs were direct
288 orthologs to 18 genes in *D. melanogaster* (Figure 4B). Judging from the response of these
289 orthologs to different volatiles in *D. melanogaster* as curated in the DoOR database[58] and
290 assuming potential conservation of function, the retained bee louse fly ORs may respond
291 to compounds produced by honey bee workers in a defense context (e.g., 1-hexanol,
292 farnesol, 2-heptanone),[59] and/or of floral, pollen and nectar aromas, such as
293 acetophenone and benzaldehyde, a major volatile of honey.[60,61] Two cases of
294 tetraplications were observed. One case involved three recent duplications of genes
295 orthologous to DmOr67b, a gene that is highly responsive in *D. melanogaster* to both
296 acetophenone and 1-hexanol. The second case involved three successive duplications of
297 a gene orthologous to DmOr74a, which responds in *D. melanogaster* larvae to 1-nonanol
298 and 1-heptanol, the latter being a major brood volatile,[10]/23/23 11:02:00 AM and 1-
299 hexanol, a component of the alarm pheromone.[62] Of these three duplications, two were
300 unique to *B. coeca* compared to its closely-related steganine species (Figure S5). Low
301 concentrations of isopentyl acetate, the main component of the alarm pheromone,
302 released by unstressed workers at the nest entrances attract the parasitic nest beetle
303 *Aethina tumida*,[63] suggesting that the detection of the host odors could be a common
304 strategy among phylogenetically distant inquilines and parasites of social insects.

305 Whereas major molecular convergences could exist between the inquiline and its
306 social host, divergent strategies to adapt to the eusocial lifestyle requirements are still
307 needed. In honey bees, colony cohesion is driven by the volatile queen's mandibular
308 pheromone (QMP), which "sterilizes" the bee workers.[64] This pheromone elicits an anti-
309 ovarian response in other insects, including *D. melanogaster*.[25] An RNA interference
310 (RNAi)-screen identified DmOr49b, DmOr56a, and DmOr98a to be potentially involved in
311 the detection of the QMP compounds and the suppression of fecundity.[25,65] A *sine qua non*
312 condition for a drosophilid to reproduce in a bee nest would, therefore, be to lose those
313 receptors or to modify their response or effect. We found that the bee louse fly does not
314 have an ortholog for DmOr98a, a receptor specific to the genus *Drosophila* (Figure S5).
315 The bee louse fly has a pseudogene, orthologous to DmOr49b, that was InsectOR
316 identified. Orthologs of this *D. melanogaster* receptor are present and complete in all
317 dipteran species, including *L. varia* and *P. variegata* (Figure S5). The bee louse fly had a
318 receptor that we called BcOr22, which was orthologous to DmOr56a (Figure 4B). This last
319 receptor is narrowly tuned in many *Drosophila* species to a single component, the mold
320 volatile geosmin, whose perception also inhibits oviposition in *D. melanogaster*,[66]
321 pointing to a possible conserved role in reproduction. Therefore, further functional
322 analyses of the response of candidate ORs to various QMP compounds are required in
323 both *D. melanogaster* and *B. coeca* to understand how modifications of these genes in *B.*
324 *coeca* might have facilitated the evolution of the bee louse fly inquilinism.
325
326 *Blindness and life in a dark nest were accompanied by the loss of multiple rhodopsins*
327 The species Latin name of the bee louse fly refers to the assumption that it was
328 blind due to the reduction of the eye size and the loss of the ocelli. In agreement with
329 reduced vision in the bee louse fly, we found only two out of the seven rhodopsin genes,
330 which are responsible for colored vision and positive phototaxis in *D. melanogaster* and
331 which were all present in the ancestral drosophilid repertoire. *D. melanogaster* orthologs
332 of the *Rh1* and *Rh6* genes are expressed in the ommatidia and are sensitive to light.[67] The
333 role of these opsins in light detection, despite the absence of ommatidia in the bee louse
334 fly is unclear. Remarkably, *Rh1* and *Rh6* are structurally required in mechanosensory
335 bristles to control larval locomotion.[68] They also detect temperature.[69] Therefore, the
336 retention of these rhodopsins in the bee louse fly could mainly be due to their
337 unconventional functions. On the other hand, the rhodopsin *Rh2*, which is exclusively

338 expressed in the ocelli and used for horizon detection in *D. melanogaster* [70], is among
339 those lost in the bee louse fly, in agreement with the loss of the ocelli. Regression of the
340 visual system and its underlying opsin genes is common in animals inhabiting dark
341 environments, such as fossorial mammals[71] and cavefishes,[72] representing a major
342 example of deep convergences.
343
344 *Apterism was not accompanied by the loss of major wing development genes*
345 Small size, loss of wings, and the evolution of strongly clinging legs are all
346 morphological changes that could prevent the honey bees from getting rid of the bee
347 lice.[73] All these potential adaptations are convergent with ectoparasitic true lice, and for
348 some, such as apterism, represent major recurrent changes that have responded to
349 distinct pressures throughout the history of insects.[74] We found intact most of the main
350 wing development genes whose mutations severely reduce the wing in *D. melanogaster*,
351 such as *wingless*, *apterous*, or *vestigial*. This means that the major morphological changes
352 more likely resulted from regulatory changes of these core genes or modifications of other
353 genes. Future developmental studies, specifically comparing the expression of wing and
354 leg morphogenic genes between the bee louse fly and *D. melanogaster*, will definitively
355 help shed light on the transcriptomic shifts underlying the major morphological changes
356 of the bee louse fly.
357
358 **Conclusion**
359 That the enigmatic bee louse fly is indeed a drosophilid, a lineage within the most
360 investigated insect family with more than 150 fully sequenced genomes, is undoubtedly
361 one of the most exciting discoveries in dipteran phylogeny. How could a fly with an
362 ancestral drosophilid genome become ecologically adapted to bees and morphologically
363 similar to lice? Our results show that a mosaic of deep convergences at the genomic level
364 underlies the relatively recent and dramatic changes of the bee louse fly to nest
365 inquilinism. This mosaicism involved deep convergences with the host, mostly in genes
366 likely involved in immunity, detoxification, and chemical perception, as well as
367 convergences with general features of fossorial animals in the visual systems. Future
368 developmental studies may elucidate whether general morphologies, such as apterism
369 and leg modifications, could also be shared between *Braula* and other ectoparasites. Due
370 to its genetic relatedness to *Drosophila* and ecological association to *Apis*, two major

laboratory models, the new genomic resources presented here can help establish the bee louse fly as a promising model to address questions related to deep convergences that are difficult to approach in multiple highly specializing animals.

Acknowledgments

We are very grateful to Marcus Stensmyr for insightful comments on an early draft of the manuscript, the *Association Conservatoire de l'Abeille Noire Bretonne* (A.C.A.N.B.) for help collecting *B. coeca* flies, Etienne Minaud for photographic images of *B. coeca* and Héloïse Muller for help with the genome annotation pipeline, Christian Cheminade and Michael Lang for help translating early literature on *Braula*. *Braula* genome sequencing was funded by a grant from Université Paris-Saclay (ADAPAR) to HB.

Author contributions

Conceptualization: H.B. and A.Y. ; Investigation: H.B., H.L., N.D., D.O., C.P., J.F., C.G., and A.Y.; Resources: H.L. and L.G.; Writing – Original draft: H.B., H.L., D.O., J.F., C.G., and A.Y.; Writing – Reviewing & Editing: H.B., J.C., C.G., F.M.P., F.R., J.C.S., and A.Y.; Visualization: H.B., J.F., C.G., and A.Y.; Supervision: H.B.; Funding Acquisition: H.B.

Declaration of interests

The authors declare no conflicts of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Grammarly (Grammarly Inc.) in order to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

399 **Figures titles and legends**

400

401 **Figure 1. The bee louse fly (*Braula coeca*) is an inquiline of the Western honey bee**
402 **(*Apis mellifera*) and has likely evolved from a sap-breeding drosophilid associated**
403 **with scale insects.**

404 (A) Tens of *B. coeca* adults preferentially attached to the honey bee queen (© Etienne
405 Minaud). Scale bar = 5 mm.

406 (B) Dorsal view of an adult showing the loss of the wings, halters, and scutum, mesonotum
407 reduction and the legs' robustness. Scale bar = 0.5 mm.

408 (C) Frontal view of an adult showing the reduction of the eyes and the loss of the ocelli.
409 Scale bar = 0.5 mm.

410 (D) Maximum likelihood phylogeny inferred from 3,100 conserved single-copy proteins
411 (2,557,349 amino acids) showing the position of *B. coeca* (orange) in the subfamily
412 Steganinae (light green) of the Drosophilidae (red). Outgroup species belong to the
413 superfamily Ephydroidea (light blue). All internal nodes had an ultra-fast bootstrap value
414 of 100% except * = 73%. Pie charts at internal nodes indicate the likelihood of ancestral
415 breeding niches inferred from the predominant niches of terminal taxa.

416

417 **Figure 2. Evolution of the bee louse fly inquilinism, its genome size, gene content,**
418 **and transposable elements in the bee louse fly with evidence for horizontal transfer**
419 **between the inquiline and its host.**

420 (A) Fossil-calibrated maximum-likelihood phylogeny inferred from 79 conserved single-
421 copy proteins (63,192 amino acids) demonstrating major stages in the evolution of the
422 inquiline and its social host. All internal nodes had an ultra-fast bootstrap value of 100
423 (except when given), with a blue interval indicating a 95% confidence level of divergence
424 time estimate inferred by MCMCTree. The red bar indicates the likely interval of the origin
425 of the bee louse fly-*Apis* association. Labels 1-6 refer to the major stages mentioned in the
426 text.

427 (B) Genome-size evolution. Red asterisk indicates the estimate for *B. coeca*.

428 (C) Gene content evolution. Red asterisk indicates the estimate for *B. coeca*.

429 (D) Proportions of transposable elements in the genomes of 42 dipteran and
430 hymenopteran species. DNAs = DNA transposons, LCs = low complexity elements, LINEs
431 = long interspersed nuclear elements, LTRs = long terminal repeats, SINEs = short

432 interspersed nuclear elements, sRNAs = small RNAs, SRs = single repeats, and Unclass. =
433 unclassified.
434 (E) Maximum-likelihood phylogeny of *Famar1*-like copies from 38 animal species. Filled
435 circles indicate ultrafast bootstrap values higher than 90%.
436
437 **Figure 3. Deep convergence in ecologically-relevant gene families evolution**
438 **between the bee louse fly and its hymenopteran host.**
439 A sample of significantly evolving gene families in *B. coeca* lineage as inferred by CAFE5
440 from functionally-grouped orthogroups defined by OrthoFinder (see text). For each panel,
441 the species tree from Figure 2 is given with branch colors reflecting inferred ancestral
442 family size reconstruction using fastanc in phylotools. CAFE5 *P*-values are noted for each
443 panel.
444 (A) GR = gustatory receptors.
445 (B) IR-DIV = divergent ionotropic receptors.
446 (C) OR = odorant receptors.
447 (D) OBP = odorant-binding-proteins.
448 (E) S1A-NPH = S1A non-peptidase homologs.
449 (F) CYP3 = Cytochrome P450 CYP3 clan.
450 (G) CYP4 = Cytochrome P450 CYP4 clan.
451 (H) GST-C = Cytosolic glutathione S-transferases.
452 (I) ECKL = Ecdysteroid kinase-like.
453 (J) CLECT = C-type lectin-like.
454 (K) CBD = Chitin binding domain-containing proteins.
455 (L) ADH = Alcohol dehydrogenases.
456
457 **Figure 4. Evolution of chemosensory receptors gene families in *Braula coeca* and**
458 ***Drosophila melanogaster***
459 (A) Maximum-likelihood phylogeny of gustatory receptors (GRs) with main taste
460 categories color code given in a frame.
461 (B) Maximum-likelihood phylogeny of odorant receptors (ORs) with main ligands for each
462 *D. melanogaster* receptor given in dark red. L = larva and A = adult expression.
463 For A and B, ultra-fast bootstrap values are given above nodes. Branches are colored
464 according to orthogroups defined by OrthoFinder for 42 dipteran and hymenopteran

465 species. Numbers in broken brackets before each orthogroup reflect the presumed
466 ancestral gene content inferred by phylotools.
467
468 **Methods**
469 *Lead contact*
470 Further information and requests for resources and reagents should be directed to and
471 will be fulfilled by the lead contact, Héloïse Bastide (heloise.bastide@universite-paris-
472 saclay.fr). Raw sequence data are deposited on NCBI Sequence Read Archive (SRA) under
473 Bioproject PRJNA1000103. Perl scripts are available on Github depository at
474 https://github.com/AmirYassinLab/Supermatrix and
475 https://github.com/AmirYassinLab/OG2GG. Commands for all programs are available on
476 Github depository at https://github.com/hbastide/Braula_genome_analysis. Genome
477 assemblies and all data associated to this study are deposited in Figshare and will be made
478 available upon publication.
479
480 *Sample collection and genomic library preparation*
481 Samples of *Braula coeca* were collected from honey bee colonies on the Island of Ouessant
482 in France and kindly provided to us by the *Association Conservatoire de l'Abeille Noire*
483 *Bretonne* (A.C.A.N.B.). Genomic DNA was extracted from 15 unsexed individuals
484 conserved in alcohol using the Nucleobond AXG20 kit and buffer set IV from Macherey-
485 Nagel (ref. 740544 and 740604, https://www.mn-net.com, Düren, Germany).
486
487 *Genome sequencing and assembly*
488 We used a hybrid approach to assemble a draft genome of *B. coeca* using both long-
489 read Oxford Nanopore Technology (ONT) and short-read Illumina sequencing.[75] Before
490 nanopore sequencing, a size selection was conducted on the DNA using the SRE XS from
491 Circulomics (https://www.circulomics.com/, Baltimore, Maryland, USA). The SQK-
492 LSK110 kit from ONT (https://nanoporetech.com/)[76] was then used to prepare the
493 samples for nanopore sequencing following the manufacturer's protocol. The library was
494 loaded and sequenced on an R9.4.1 flow cell (ref. FLO-Min106) for sequencing. Raw data
495 were basecalled using Guppy v5.0.11 and the "sup" algorithm. The ONT raw data size was
496 4.4 Gb in 1,399,323 reads (mean read length 3,146 kb, longest read of 123.3 Mb), with an
497 N50 of 4,677 kb. Phred scores ranged from 8 to 18, with a median of 13, as assessed by

498 PycoQC.[77] Illumina paired-end sequencing was performed by Novogene Company Limited
499 (https://en.novogene.com, Cambridge, UK) on the same DNA sample. The Illumina
500 sequencing produced 119,719,537 paired 150 bp reads. Phred scores averaged 36 per read
501 as analyzed by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We
502 used MaSuRCA v4.0.3[78] to produce the hybrid assembly of our genome using the Cabog
503 assembler. We obtained a final assembly size of 309,35Mb in 2477 contigs, with a N50 of
504 347,227 bp. The completeness of the assembly was estimated to 95.8% with BUSCO v5.0.0
505 on the *diptera_odb10* dataset (C:95.8%[S:94.5%,D:1.3%],F:0.7%,M:3.5%,n:3285), and to
506 93.6% using Merqury.
507
508 *Estimation of genome size and endosymbionts detection*
509 K-mers frequencies within short-read data were obtained with KMC 3.[79] Genome
510 size and ploidy were inferred using GenomeScope v2.0 with k-mer size = 21 and
511 Smudgeplot.[80] Contig taxonomy was performed using Blobtools[81] with Diamond as search
512 engine[82] against the UniProt database using a local copy of the NCBI TaxID file for the
513 taxonomic assignation of the best hit. Minimap2[83] was used for read mapping.
514
515 *Genome annotation*
516 The *B. coeca* genome was annotated using Maker v2.31.10,[30] following Muller et
517 al.'s[84] protocol, wherein multiple rounds of Maker supported by the training of the SNAP
518 v.2006-07-28[31] and Augustus v.3.3.3[32] gene finding and prediction tools, were conducted.
519 RepeatModeler v2.0.1 was first used to identify the repeat-enriched regions that were
520 masked by RepeatMasker v4.0.9 as implemented in Maker. Proteomes of five *Drosophila*
521 species, namely *D. innubila*, *D. albomicans*, *D. bipectinata*, *D. melanogaster*, and *D. virilis*
522 were obtained from NCBI and used to guide the annotation. Protein-Protein BLAST
523 2.9.0+[85] (-evalue 1e-6 -max_hsps 1 -max_target_seqs 1) was then used to assess putative
524 protein functions in *B. coeca* by comparing the protein sequences given by Maker to the
525 protein sequences from the annotated genome of *D. melanogaster*. The completeness of
526 genome annotation was assessed using BUSCO at each round and the round with the
527 highest score was retained.
528
529 *Phylogenomic analysis of the Ephydroidea*

530 Besides our *B. coeca* assembly, we obtained from NCBI repository genome
531 assemblies for 12 species, transcriptome shotgun assemblies (TSA) for four species, and
532 sequence read runs (SRR) for three species (Data S1). Paired-end DNA raw data of two
533 species, namely *Rhinoleucophenga* cf. *bivisualis* and *Cacoxenus indagator* were assembled
534 using MaSuRCA with default parameters. The transcriptome of *Acletoxenus* sp. was
535 assembled using Trinity software package[86] on the Galaxy Europe website[87] following
536 standard protocol[88]. BUSCO v.5.0 was used to assess the completeness of those assemblies
537 and to extract single-copy BUSCO genes for all species. Protein sequences of 3,100 single
538 and complete BUSCO genes were aligned using MAFFT[89] and concatenated into a single
539 supermatrix (2,557,349 amino acids). A maximum-likelihood (ML) phylogeny was then
540 inferred for the supermatrix using IqTREE 2[90] with 1,000 ultrafast bootstrap iterations[91]
541 and using the JTT+R substitution model inferred by ModelFinder[92] implemented by
542 IqTree.
543
544 *Reconstruction of the ancestral ecological niches*
545 For each of the 20 ephydroid species we obtained a predominant ecological niche
546 from the taxonomic literature.[17,20,21,93] Eight predominant niches were coded as discrete
547 traits, and the Multistate program of the BayesTraits v.4 package[94] was used under the
548 Reverse Jump MCMC model with 1,010,000 chain iterations and a burnin sample of
549 10,000.
550
551 *Phylogenomic analysis of Diptera and Hymenoptera*
552 The second phylogenomic analysis involved, besides *B. coeca*, 25 Hymenopteran
553 and 16 Dipteran species for which an assembly can be downloaded from the NCBI Genome
554 repository (Data S1). Protein sequences for all species but three, namely *Leucophenga*
555 *varia*, *Phortica variegata*, and *Ephydra gracilis*, were obtained from NCBI. For these three
556 species, we used the same four-round annotation procedure that we used for *B. coeca* to
557 identify protein-coding genes and translate their sequences. We used BUSCO to assess the
558 completeness of all annotated and downloaded genomes and their corresponding
559 assemblies. OrthoFinder[43] was used to generate protein sequences of protein-coding-
560 genes of the 42 species and to cluster these sequences into orthogroups. Only the longest
561 isoform (*i.e.* the primary transcript) was used for genes with multiple isoforms. 79
562 orthogroups contained a single copy ortholog from each species, and their protein

563 sequences were aligned using MAFFT and concatenated into a single supermatrix (63,192
564 amino acids). A maximum-likelihood (ML) phylogeny was then inferred for the
565 supermatrix using IqTREE 2[90] with 1,000 ultrafast bootstrap iterations[91] and the JTT+R
566 substitution model inferred by ModelFinder[92] implemented by IqTree.
567 MCMCTree[95] was used to date the inferred ML trees based on recently published
568 fossil-calibrated phylogenies. First, two time points were obtained for the 42-species
569 phylogeny. These included the divergence between ants and bees between 90-120 myr
570 ago[96] and between *Scaptodrosophila* and *Drosophila* between 50-56 myr ago,[97] with a
571 maximum root age for the ancestor of Hymenoptera and Diptera at 344 myr ago.[96]
572
573 *Genome size and gene content evolution*
574 Genome size and gene content (number of OrthoFinder generated protein-coding-
575 genes after retaining the longest isoform for genes with multiple transcripts) inferred for
576 each of the 42 Dipteran and Hymenopteran genomes were mapped on the phylogenetic
577 tree, and values at the ancestral nodes were inferred and visualized using the fastAnc
578 command in the R package Phylotools v0.2.2.[98]
579
580 *Transposable elements annotation and transfer*
581 Transposable elements were identified in *B. coeca*, *D. melanogaster*, and *A.*
582 *mellifera* following a two-step protocol. First, we used RepeatModeler v2.0.1[34] with
583 default parameters to generate a *de novo* library of repetitive regions. RepeatMasker v
584 4.0.9[34] was then run with the newly generated library and the options -a (create a .align
585 output file) and -s (slow search; more sensitive) to create a summary of the families of
586 transposable elements found in each reference genome along with the percentage of the
587 genome they represent. Horizontal transfer analysis protocols of the *Famar1*-like element
588 are given in Document S2.
589
590 *Gene family evolution*
591 We used CAFE v. 5[44] to model and infer gene family evolution. We conducted
592 CAFÉ5 using an error model on the 19,011 orthgroups generated by OrthoFinder and
593 using the time-calibrated phylogenetic tree of the 42 Dipteran and Hymenopteran species.
594 We also grouped OrthoFinder orthogroups into functional gene groups using the
595 customized perl script OG2GG.pl (https://github.com/AmirYassinLab/OG2GG). The

596 script extracts and groups orthogroups containing *D. melanogaster* orthologs that were
597 assigned to large functional gene groups in Flybase[45]
598 ("gene_group_data_fb_2023_02.tsv"). CAFE5 was then run on the gene groups' gene
599 counts using a single birth rate and the error model to correct for possible assembly and
600 annotation errors.
601
602 *Chemosensory superfamilies evolution*
603 To curate *B. coeca* gustatory receptors (GR) and odorant receptors (OR) genes, we
604 queried *D. melanogaster* GRs and ORs protein sequences on *B. coeca*, *L. varia* and *P.*
605 *variegata* assemblies using Exonerate ver. 2.2[99] with option –maxintron 2000 and -p
606 pam250. The output, along with the assembly, were fed to InsectOR[52] with option 7tm_7
607 and 7tm_6 activated for GR and OR analyses, respectively. From the output files, we
608 extracted 300-500 amino acids-long complete sequences with 7tm_7 or 7tm_6 motif
609 detected and with start codon present and no internal stop codon, *i.e.* pseudogenes
610 excluded. Protein sequences were analyzed using Miniprot[100] to produce an exhaustive
611 gff file with exact genomic coordinates for each species. Protein sequences were aligned
612 using MAFFT and a maximum-likelihood phylogenetic tree for each family using IqTREE
613 2 with the same options as the phylogenomic analysis. The literature was reviewed to
614 classify GRs into bitter, sweet, and CO_2 categories[101] and identify volatile ligands eliciting
615 the strongest response in odorant neurons in *D. melanogaster*.[58] Because CAFE5 inferred
616 ancestral counts for orthogroups with significant deviation only, we estimated and
617 visualized ancestral counts for each orthogroups of these two families using FastAnc
618 command in the R package Phylotools.
619
620 **List of supplementary materials**
621
622 **Data S1.** List of all genome assemblies and annotations downloaded or generated in this
623 study with data on genome size and gene content.
624
625 **Figure S1. Features of *Braula coeca* genome assembly**
626 (A) Smudgeplot comparing the sum of heterozygous kmer pair coverages (A+B) to their
627 relative coverage of the minor kmer (B/A+B). The hottest smudge corresponds to
628 expected diploid kmer pairs (AB).

629 (B) Kmer profile spectrum estimating the length of the genome at 307,538,118 bp
630 generated by GenomeScope2.
631 (C) Square-binned blob plot showing the distribution of assembly scaffolds on GC
632 proportion and coverage axes. Squares within each bin are colored according to
633 taxonomic annotation and scaled according to total span.
634 (D) ReadCovPlots generated by Blobtools visualising the proportion of reads of *B. coeca*
635 that are unmapped or mapped and showing the percentage of mapped reads by
636 taxonomic group, as barcharts.
637
638 **Figure S2. Evolution of genome size and gene content and completeness of assembly**
639 **and annotation of 42 dipteran and hymenopteran genomes.**
640 (A) Genome-size evolution and genome assembly completeness inferred by BUSCO across
641 42 dipteran and hymenopteran species.
642 (B) Gene content evolution and genome annotation completeness inferred by BUSCO
643 across 42 dipteran and hymenopteran species.
644 Branch colors in A and B reflect inferred ancestral reconstructions using fastanc in
645 phylotools. CS = complete and single-copy BUSCOs; CD = complete and duplicated
646 BUSCOs; F = fragmented BUSCOs; M = missing BUSCOs.
647
648 **Figure S3. Horizontal transposon transfer between *Braula coeca* and *Apis mellifera*.**
649 Comparison of Famar1-like transposon synonymous distance (dS) and orthologous gene
650 synonymous distances between *Braula coeca* and *Apis mellifera*. The red line indicates the
651 0.5% quantile (=1.76) of the distribution of orthologous gene dS calculated over 2,179
652 codon alignments. The distribution is bimodal, with genes having highly saturated dS
653 values showing a peak centered on 9.99 and highly conserved genes showing less
654 saturated dS values showing another peak around 2.5. The *Famar1*-like dS (green line, =
655 0.12) was calculated over the transposase open reading frame of one copy of the element
656 extracted from the *A. mellifera* genome and another copy from the *B. coeca* genome.
657
658 **Figure S4.** Maximum-likelihood phylogeny of gustatory receptors (GRs) from *D.*
659 *melanogaster*, *B. coeca*, *L. varia* and *P. variegata*.
660

Figure S5. Maximum-likelihood phylogeny of odorant receptors (ORs) from *D. melanogaster*, *B. coeca*, *L. varia* and *P. variegata*.

Document S1 – Translation of early works on the bee louse fly.

Document S2 – Analysis of horizontal transfer of transposable elements between *B. coeca* and *A. mellifera*.

References

1. Cini, A., Sumner, S., and Cervo, R. (2019). Inquiline social parasites as tools to unlock the secrets of insect sociality. Philosophical Transactions of the Royal Society B: Biological Sciences *374*, 20180193. 10.1098/rstb.2018.0193.
2. Rabeling, C., Schultz, T.R., Pierce, N.E., and Bacci, M. (2014). A Social Parasite Evolved Reproductive Isolation from Its Fungus-Growing Ant Host in Sympatry. Current Biology *24*, 2047–2052. 10.1016/j.cub.2014.07.048.
3. Borowiec, M.L., Cover, S.P., and Rabeling, C. (2021). The evolution of social parasitism in Formica ants revealed by a global phylogeny. Proceedings of the National Academy of Sciences *118*, e2026029118. 10.1073/pnas.2026029118.
4. Skaife, S.H. (1922). On *Braula Coeca* , Nitzsch, a Dipterous parasite of the honey bee. Transactions of the Royal Society of South Africa *10*, 41–48. 10.1080/00359192209519263.
5. Imms, A.D. (1942). On Braula coeca Nitsch and its affinities. Parasitology *34*, 88–100. 10.1017/S0031182000016012.
6. Bayless, K.M., Trautwein, M.D., Meusemann, K., Shin, S., Petersen, M., Donath, A., Podsiadlowski, L., Mayer, C., Niehuis, O., Peters, R.S., et al. (2021). Beyond Drosophila: resolving the rapid radiation of schizophoran flies with phylotranscriptomics. BMC Biol *19*, 23. 10.1186/s12915-020-00944-8.
7. Winkler, I.S., Kirk-Spriggs, A.H., Bayless, K.M., Soghigian, J., Meier, R., Pape, T., Yeates, D.K., Carvalho, A.B., Copeland, R.S., and Wiegmann, B.M. (2022). Phylogenetic resolution of the fly superfamily Ephydroidea–Molecular systematics of the enigmatic and diverse relatives of Drosophilidae. PLOS ONE *17*, e0274292. 10.1371/journal.pone.0274292.
8. Avalos, J., Rosero, H., Maldonado, G., and Reynaldi, F.J. (2019). Honey bee louse (Braula schmitzi) as a honey bee virus vector? Journal of Apicultural Research *58*, 427–429. 10.1080/00218839.2019.1565726.
9. de Réaumur, R.-A.F. (1740). Mémoires pour servir à l'histoire des insectes... Vol. 5 (de l'Imprimerie royale).
10. Nitzsch, C. (1818). Die Familien und Gattungen der Thierinsekten (insecta epizoica); als Prodromus einer Naturgeschichte derselben. In.
11. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. 10.1093/bioinformatics/btv351.
12. Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biology *21*, 245. 10.1186/s13059-020-02134-9.
13. Koren, S., Phillippy, A.M., Simpson, J.T., Loman, N.J., and Loose, M. (2019). Reply to 'Errors in long-read assemblies can critically affect protein prediction.' Nat Biotechnol *37*,

127–128. 10.1038/s41587-018-0005-y.
14. Yassin, A. (2013). Phylogenetic classification of the Drosophilidae Rondani (Diptera): the role of morphology in the postgenomic era. Systematic Entomology *38*, 349–364.
15. Rondani, C. (1856). Dipterologiae Italicae prodromus: Genera Italica Ordinis Dipterorum (Stocchi).
16. Egger, S. (1853). Himmlische Waffenrüstung für die Jugend bestehend aus den heiligen Sakramenten der Busse, des Altars und der Firmung: Ein praktischer Unterricht (Schmid).
17. Throckmorton, L. (1975). The phylogeny, ecology, and geography of *Drosophila*. In.
18. Keiper, J.B., Walton, W.E., and Foote, B.A. (2002). Biology and ecology of higher Diptera from freshwater wetlands. Annu. Rev. Entomol. *47*, 207–232. 10.1146/annurev.ento.47.091201.145159.
19. Pollock, J.N. (2002). Observations on the biology and anatomy of Curtonotidae (Diptera: Schizophora). Journal of Natural History *36*, 1725–1745. 10.1080/00222930110061869.
20. Markow, T.A., and O'Grady, P.M. (2005). Evolutionary genetics of reproductive behavior in *Drosophila*: connecting the dots. Annu. Rev. Genet. *39*, 263–291. 10.1146/annurev.genet.39.073003.112454.
21. Ashburner, M. (1981). Entomophagous and other bizarre *Drosophilidae*. In.
22. Duchêne, D., Duchêne, S., and Ho, S.Y.W. (2014). Tree imbalance causes a bias in phylogenetic estimation of evolutionary timescales using heterochronous sequences. Molecular Ecology Resources *15*, 785–794. 10.1111/1755-0998.12352.
23. Shell, W.A., Steffen, M.A., Pare, H.K., Seetharam, A.S., Severin, A.J., Toth, A.L., and Rehan, S.M. (2021). Sociality sculpts similar patterns of molecular evolution in two independently evolved lineages of eusocial bees. Commun Biol *4*, 253. 10.1038/s42003-021-01770-6.
24. Noll, F.B. (2002). Behavioral phylogeny of corbiculate Apidae (Hymenoptera; Apinae), with special reference to social behavior. Cladistics *18*, 137–153. 10.1111/j.1096-0031.2002.tb00146.x.
25. Galang, K.C., Croft, J.R., Thompson, G.J., and Percival-Smith, A. (2019). Analysis of the Drosophila melanogaster anti-ovarian response to honey bee queen mandibular pheromone. Insect Molecular Biology *28*, 99–111. 10.1111/imb.12531.
26. Grimaldi, D., and Underwood, B.A. (1986). Megabraula, a new genus for two new species of Braulidae (Diptera), and a discussion of braulid evolution. System Entomol *11*, 427–438. 10.1111/j.1365-3113.1986.tb00534.x.
27. Dobson, J.R. (1999). A "bee-louse" *Braula schmitzi* örösi-pál (Diptera: Braulidae) new to the British Isles, and the status of *Braula* spp. in England and Wales.
28. Kelley, J.L., Peyton, J.T., Fiston-Lavier, A.-S., Teets, N.M., Yee, M.-C., Johnston, J.S., Bustamante, C.D., Lee, R.E., and Denlinger, D.L. (2014). Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. Nat Commun *5*, 4611. 10.1038/ncomms5611.
29. Kim, B.Y., Wang, J.R., Miller, D.E., Barmina, O., Delaney, E., Thompson, A., Comeault, A.A., Peede, D., D'Agostino, E.R., Pelaez, J., et al. (2021). Highly contiguous assemblies of 101 drosophilid genomes. eLife *10*, e66405. 10.7554/eLife.66405.
30. Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A.S., and Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. *18*, 188–196. 10.1101/gr.6743907.
31. Korf, I. (2004). Gene finding in novel genomes. BMC Bioinformatics *5*, 59. 10.1186/1471-2105-5-59.

32. König, S., Romoth, L.W., Gerischer, L., and Stanke, M. (2016). Simultaneous gene finding in multiple genomes. Bioinformatics *32*, 3388–3395. 10.1093/bioinformatics/btw494.
33. van Breugel, F., and Dickinson, M.H. (2017). Superhydrophobic diving flies (*Ephydra hians*) and the hypersaline waters of Mono Lake. Proceedings of the National Academy of Sciences *114*, 13483–13488. 10.1073/pnas.1714874114.
34. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences *117*, 9451–9457. 10.1073/pnas.1921046117.
35. Smit, A.F.A., Hubley, R., and Green, P. (2015). RepeatMasker Open-4.0. 2013–2015 http://www.repeatmasker.org/.
36. Mérel, V., Boulesteix, M., Fablet, M., and Vieira, C. (2020). Transposable elements in *Drosophila*. Mobile DNA *11*, 23. 10.1186/s13100-020-00213-z.
37. Gilbert, C., Peccoud, J., and Cordaux, R. (2021). Transposable elements and the evolution of insects. Annu. Rev. Entomol. *66*, 355–372. 10.1146/annurev-ento-070720-074650.
38. Venner, S., Miele, V., Terzian, C., Biémont, C., Daubin, V., Feschotte, C., and Pontier, D. (2017). Ecological networks to unravel the routes to horizontal transposon transfers. PLOS Biology *15*, e2001536. 10.1371/journal.pbio.2001536.
39. Barry, E.G., Witherspoon, D.J., and Lampe, D.J. (2004). 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. 10.1534/genetics.166.2.823.
40. Carr, S.M. (2023). Multiple mitogenomes indicate things fall apart with Out of Africa or Asia hypotheses for the phylogeographic evolution of honey bees (*Apis mellifera*). Sci Rep *13*, 9386. 10.1038/s41598-023-35937-4.
41. Gilbert, C., Schaack, S., Pace II, J.K., Brindley, P.J., and Feschotte, C. (2010). A role for host–parasite interactions in the horizontal transfer of transposons across phyla. Nature *464*, 1347–1350. 10.1038/nature08939.
42. Gilbert, C., and Maumus, F. (2022). Multiple Horizontal Acquisitions of Plant Genes in the Whitefly Bemisia tabaci. Genome Biology and Evolution *14*, evac141. 10.1093/gbe/evac141.
43. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol *20*, 1–14. 10.1186/s13059-019-1832-y.
44. Mendes, F.K., Vanderpool, D., Fulton, B., and Hahn, M.W. (2020). CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics *36*, 5516–5518. 10.1093/bioinformatics/btaa1022.
45. Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J., Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V., et al. (2019). FlyBase 2.0: the next generation. Nucleic Acids Research *47*, D759–D765. 10.1093/nar/gky1003.
46. Evans, J.D., Aronstein, K., Chen, Y.P., and Hetru, C. (2006). Immune pathways and defence mechanisms in honey bees *Apis mellifera*. Insect Molecular Biology.
47. Berenbaum, M.R., and Johnson, R.M. (2015). Xenobiotic detoxification pathways in honey bees. Current Opinion in Insect Science *10*, 51–58. 10.1016/j.cois.2015.03.005.
48. Chan, Q.W.T., Chan, M.Y., Logan, M., Fang, Y., Higo, H., and Foster, L.J. (2013). Honey bee protein atlas at organ-level resolution. Genome Res. *23*, 1951–1960. 10.1101/gr.155994.113.
49. Miler, K., Stec, D., Kamińska, A., Pardyak, L., and Kuszewska, K. (2021). Alcohol intoxication resistance and alcohol dehydrogenase levels differ between the honeybee castes. Apidologie *52*, 230–241. 10.1007/s13592-020-00812-y.

806 50. Winston, M.L. (1987). The Biology of the Honey Bee (Harvard University Press).
807 51. Weiss, L.A., Dahanukar, A., Kwon, J.Y., Banerjee, D., and Carlson, J.R. (2011). The
808 Molecular and Cellular Basis of Bitter Taste in Drosophila. Neuron *69*, 258–272.
809 10.1016/j.neuron.2011.01.001.
810 52. Karpe, S.D., Tiwari, V., and Ramanathan, S. (2021). InsectOR—Webserver for
811 sensitive identification of insect olfactory receptor genes from non-model genomes. PLoS
812 ONE *16*, e0245324. 10.1371/journal.pone.0245324.
813 53. Sadd, B.M., Barribeau, S.M., Bloch, G., de Graaf, D.C., Dearden, P., Elsik, C.G.,
814 Gadau, J., Grimmelikhuijzen, C.J., Hasselmann, M., Lozier, J.D., et al. (2015). The genomes
815 of two key bumblebee species with primitive eusocial organization. Genome Biol *16*, 1–32.
816 10.1186/s13059-015-0623-3.
817 54. Robertson, H.M., and Wanner, K.W. (2006). The chemoreceptor superfamily in the
818 honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family.
819 Genome Res. *16*, 1395–1403. 10.1101/gr.5057506.
820 55. Ni, L. (2021). The structure and function of ionotropic receptors in *Drosophila*. Front.
821 Mol. Neurosci. *13*, 638839. 10.3389/fnmol.2020.638839.
822 56. Park, D., Jung, J.W., Choi, B.-S., Jayakodi, M., Lee, J., Lim, J., Yu, Y., Choi, Y.-S.,
823 Lee, M.-L., Park, Y., et al. (2015). Uncovering the novel characteristics of Asian honey bee,
824 *Apis cerana*, by whole genome sequencing. BMC Genomics *16*, 1. 10.1186/1471-2164-16-1.
825 57. Karpe, S.D., Jain, R., Brockmann, A., and Sowdhamini, R. (2016). Identification of
826 Complete Repertoire of Apis florea Odorant Receptors Reveals Complex Orthologous
827 Relationships with Apis mellifera. Genome Biology and Evolution *8*, 2879–2895.
828 10.1093/gbe/evw202.
829 58. Münch, D., and Galizia, C.G. (2016). DoOR 2.0 - Comprehensive mapping of
830 *Drosophila melanogaster* odorant responses. Sci Rep *6*, 21841. 10.1038/srep21841.
831 59. Wang, Z., and Tan, K. (2019). Honey Bee Alarm Pheromone Mediates
832 Communication in Plant–Pollinator–Predator Interactions. Insects *10*, 366.
833 10.3390/insects10100366.
834 60. Machado, A.M., Miguel, M.G., Vilas-Boas, M., and Figueiredo, A.C. (2020). Honey
835 Volatiles as a Fingerprint for Botanical Origin—A Review on their Occurrence on Monofloral
836 Honeys. Molecules *25*, 374. 10.3390/molecules25020374.
837 61. Starowicz, M., Hanus, P., Lamparski, G., and Sawicki, T. (2021). Characterizing the
838 Volatile and Sensory Profiles, and Sugar Content of Beeswax, Beebread, Bee Pollen, and
839 Honey. Molecules *26*, 3410. 10.3390/molecules26113410.
840 62. Collins, A.M., and Blum, M.S. (1983). Alarm responses caused by newly identified
841 compounds derived from the honeybee sting. J Chem Ecol *9*, 57–65. 10.1007/BF00987770.
842 63. Torto, B., Boucias, D.G., Arbogast, R.T., Tumlinson, J.H., and Teal, P.E.A. (2007).
843 Multitrophic interaction facilitates parasite–host relationship between an invasive beetle and
844 the honey bee. Proceedings of the National Academy of Sciences *104*, 8374–8378.
845 10.1073/pnas.0702813104.
846 64. Van Oystaeyen, A., Oliveira, R.C., Holman, L., van Zweden, J.S., Romero, C., Oi,
847 C.A., d'Ettorre, P., Khalesi, M., Billen, J., Wäckers, F., et al. (2014). Conserved Class of
848 Queen Pheromones Stops Social Insect Workers from Reproducing. Science *343*, 287–290.
849 10.1126/science.1244899.
850 65. Camiletti, A.L., Percival-Smith, A., Croft, J.R., and Thompson, G.J. (2016). A novel
851 screen for genes associated with pheromone-induced sterility. Sci Rep *6*, 36041.
852 10.1038/srep36041.
853 66. Stensmyr, M.C., Dweck, H.K.M., Farhan, A., Ibba, I., Strutz, A., Mukunda, L., Linz,
854 J., Grabe, V., Steck, K., Lavista-Llanos, S., et al. (2012). A Conserved Dedicated Olfactory
855 Circuit for Detecting Harmful Microbes in Drosophila. Cell *151*, 1345–1357.

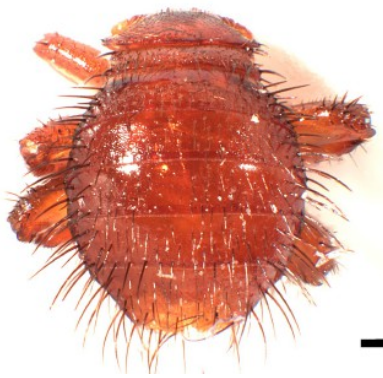
856 10.1016/j.cell.2012.09.046.
857 67. Senthilan, P.R., and Helfrich-Förster, C. (2016). Rhodopsin 7–The unusual Rhodopsin
858 in Drosophila. PeerJ *4*, e2427. 10.7717/peerj.2427.
859 68. Zanini, D., Giraldo, D., Warren, B., Katana, R., Andrés, M., Reddy, S., Pauls, S.,
860 Schwedhelm-Domeyer, N., Geurten, B.R.H., and Göpfert, M.C. (2018). Proprioceptive Opsin
861 Functions in Drosophila Larval Locomotion. Neuron *98*, 67-74.e4.
862 10.1016/j.neuron.2018.02.028.
863 69. Leung, N.Y., and Montell, C. (2017). Unconventional Roles of Opsins. Annu. Rev.
864 Cell Dev. Biol. *33*, 241–264. 10.1146/annurev-cellbio-100616-060432.
865 70. Mishra, A.K., Fritsch, C., Voutev, R., Mann, R.S., and Sprecher, S.G. (2021).
866 Homothorax controls a binary Rhodopsin switch in Drosophila ocelli. PLOS Genetics *17*,
867 e1009460. 10.1371/journal.pgen.1009460.
868 71. Partha, R., Chauhan, B.K., Ferreira, Z., Robinson, J.D., Lathrop, K., Nischal, K.K.,
869 Chikina, M., and Clark, N.L. (2017). Subterranean mammals show convergent regression in
870 ocular genes and enhancers, along with adaptation to tunneling. eLife *6*, e25884.
871 10.7554/eLife.25884.
872 72. Policarpo, M., Fumey, J., Lafargeas, P., Naquin, D., Thermes, C., Naville, M.,
873 Dechaud, C., Volff, J.-N., Cabau, C., Klopp, C., et al. (2021). Contrasting Gene Decay in
874 Subterranean Vertebrates: Insights from Cavefishes and Fossorial Mammals. Molecular
875 Biology and Evolution *38*, 589–605. 10.1093/molbev/msaa249.
876 73. Büscher, T.H., Petersen, D.S., Bijma, N.N., Bäumler, F., Pirk, C.W.W., Büsse, S.,
877 Heepe, L., and Gorb, S.N. (2022). The exceptional attachment ability of the ectoparasitic bee
878 louse Braula coeca (Diptera, Braulidae) on the honeybee. Physiological Entomology *47*, 83–
879 95. 10.1111/phen.12378.
880 74. Roff, D.A. (1990). The Evolution of Flightlessness in Insects. Ecological Monographs
881 *60*, 389–421. 10.2307/1943013.
882 75. Miller, D.E., Staber, C., Zeitlinger, J., and Hawley, R.S. (2018). Highly Contiguous
883 Genome Assemblies of 15 Drosophila Species Generated Using Nanopore Sequencing. G3:
884 Genes, Genomes, Genetics *8*, 3131–3141. 10.1534/g3.118.200160.
885 76. Lu, H., Giordano, F., and Ning, Z. (2016). Oxford Nanopore MinION Sequencing and
886 Genome Assembly. Genomics, Proteomics & Bioinformatics *14*, 265–279.
887 10.1016/j.gpb.2016.05.004.
888 77. Leger, A., and Leonardi, T. (2019). pycoQC, interactive quality control for Oxford
889 Nanopore Sequencing. JOSS *4*, 1236. 10.21105/joss.01236.
890 78. Zimin, A.V., Puiu, D., Luo, M.-C., Zhu, T., Koren, S., Marçais, G., Yorke, J.A.,
891 Dvořák, J., and Salzberg, S.L. (2017). Hybrid assembly of the large and highly repetitive
892 genome of Aegilops tauschii, a progenitor of bread wheat, with the MaSuRCA mega-reads
893 algorithm. Genome Res. *27*, 787–792. 10.1101/gr.213405.116.
894 79. Kokot, M., Długosz, M., and Deorowicz, S. (2017). KMC 3: counting and
895 manipulating k-mer statistics. Bioinformatics *33*, 2759–2761. 10.1093/bioinformatics/btx304.
896 80. Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and
897 Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun *11*, 1432.
898 10.1038/s41467-020-14998-3.
899 81. Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: Interrogation of genome
900 assemblies. 10.12688/f1000research.12232.1.
901 82. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment
902 using DIAMOND. Nat Methods *12*, 59–60. 10.1038/nmeth.3176.
903 83. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics
904 *34*, 3094–3100. 10.1093/bioinformatics/bty191.
905 84. Muller, H., Ogereau, D., Da Lage, J.-L., Capdevielle, C., Pollet, N., Fortuna, T.,

906 Jeannette, R., Kaiser, L., and Gilbert, C. (2021). Draft nuclear genome and complete
907 mitogenome of the Mediterranean corn borer, Sesamia nonagrioides, a major pest of maize.
908 G3 Genes|Genomes|Genetics *11*, jkab155. 10.1093/g3journal/jkab155.
909 85. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and
910 Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database
911 search programs. Nucl Acids Res *25*, 3389–3402. 10.1093/nar/25.17.3389.
912 86. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I.,
913 Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Trinity: reconstructing a
914 full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol *29*, 644–652.
915 10.1038/nbt.1883.
916 87. Boekel, J., Chilton, J.M., Cooke, I.R., Horvatovich, P.L., Jagtap, P.D., Käll, L., Lehtiö,
917 J., Lukasse, P., Moerland, P.D., and Griffin, T.J. (2015). Multi-omic data analysis using
918 Galaxy. Nat Biotechnol *33*, 137–139. 10.1038/nbt.3134.
919 88. Hiltemann, S., Rasche, H., Gladman, S., Hotz, H.-R., Larivière, D., Blankenberg, D.,
920 Jagtap, P.D., Wollmann, T., Bretaudeau, A., Goué, N., et al. (2023). Galaxy Training: A
921 powerful framework for teaching! PLOS Computational Biology *19*, e1010752.
922 10.1371/journal.pcbi.1010752.
923 89. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software
924 version 7: improvements in performance and usability. Molecular Biology and Evolution *30*,
925 772–780. 10.1093/molbev/mst010.
926 90. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von
927 Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for
928 Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution *37*, 1530–1534.
929 10.1093/molbev/msaa015.
930 91. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018).
931 UFBoot2: improving the ultrafast bootstrap approximation. Molecular Biology and Evolution
932 *35*, 518–522. 10.1093/molbev/msx281.
933 92. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S.
934 (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods
935 *14*, 587–589. 10.1038/nmeth.4285.
936 93. Okada, T. (1962). Bleeding sap preference of the Drosophilid flies. Japanese journal of
937 applied entomology and zoology *6*, 216–229. 10.1303/jjaez.6.216.
938 94. Pagel, M., and Meade, A. Bayesian analysis of correlated evolution of discrete
939 characters by reversible-jump markov chain monte carlo.
940 95. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Molecular
941 Biology and Evolution *24*, 1586–1591. 10.1093/molbev/msm088.
942 96. Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen,
943 P.B., Ware, J., Flouri, T., Beutel, R.G., et al. (2014). Phylogenomics resolves the timing and
944 pattern of insect evolution. Science *346*, 763–767. 10.1126/science.1257570.
945 97. Suvorov, A., Kim, B.Y., Wang, J., Armstrong, E.E., Peede, D., D'Agostino, E.R.R.,
946 Price, D.K., Waddell, P.J., Lang, M., Courtier-Orgogozo, V., et al. (2022). Widespread
947 introgression across a phylogeny of 155 *Drosophila* genomes. Current Biology *32*, 111-
948 123.e5. 10.1016/j.cub.2021.10.052.
949 98. Zhang, J., Pei, N., Mi, X., and Zhang, M.J. (2017). Package 'phylotools'. dimension
950 12. URL: https://github. com/helixcn/phylotools.
951 99. Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for
952 biological sequence comparison. BMC Bioinformatics *6*, 31. 10.1186/1471-2105-6-31.
953 100. Li, H. (2023). Protein-to-genome alignment with miniprot. Bioinformatics *39*,
954 btad014. 10.1093/bioinformatics/btad014.
955 101. Montell, C. (2009). A taste of the *Drosophila* gustatory receptors. Current Opinion in

956 Neurobiology *19*, 345–353. 10.1016/j.conb.2009.07.001.
957

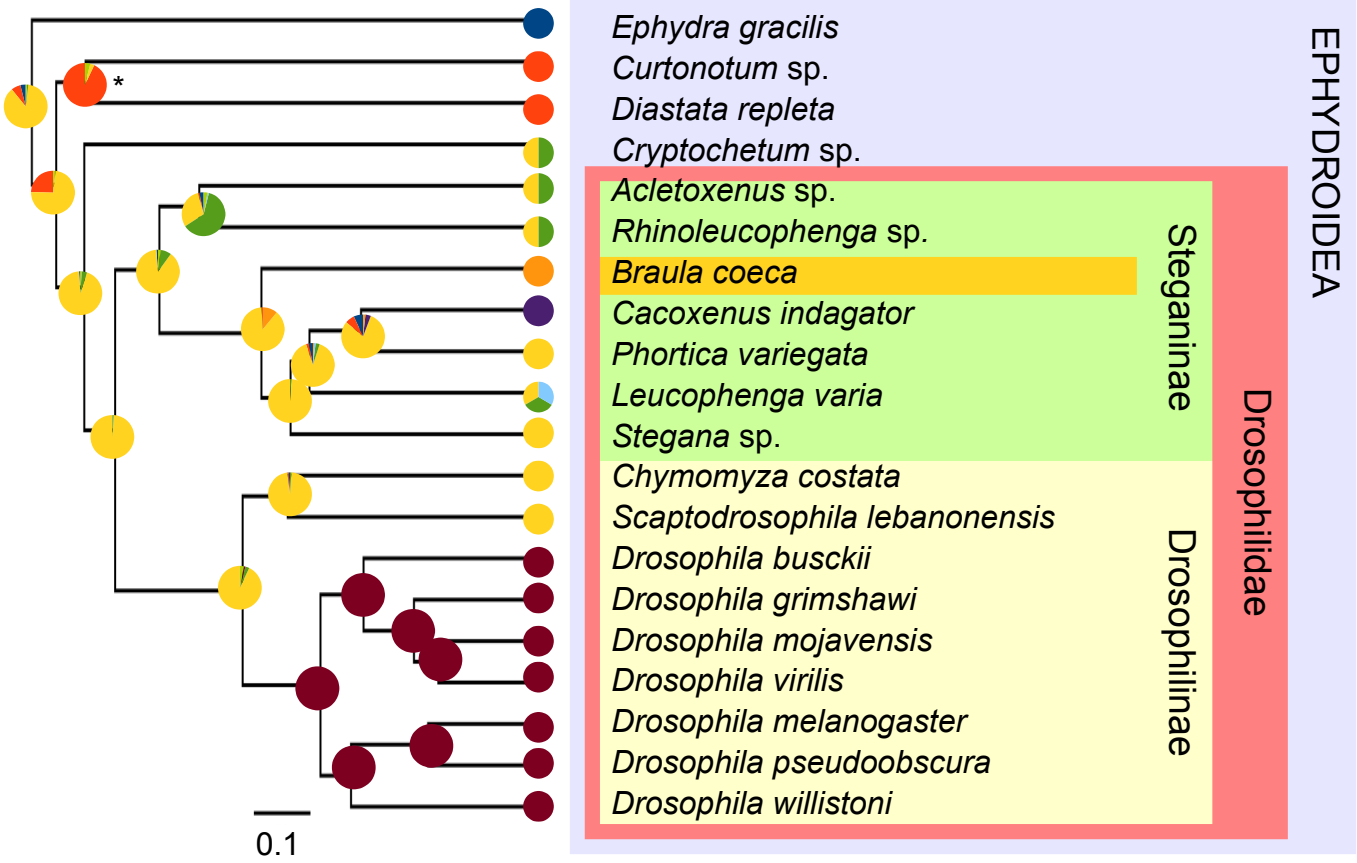
A

B



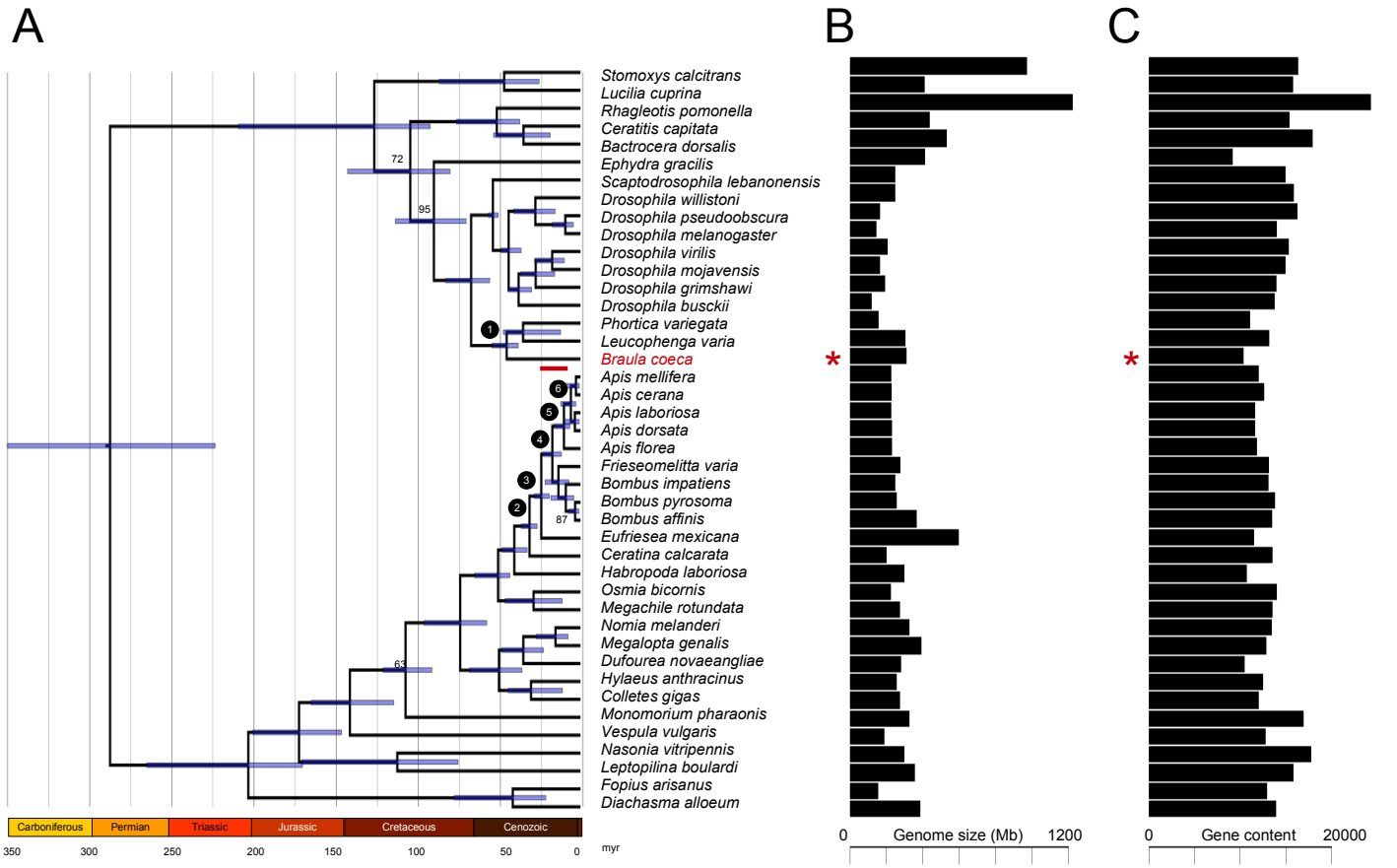
C

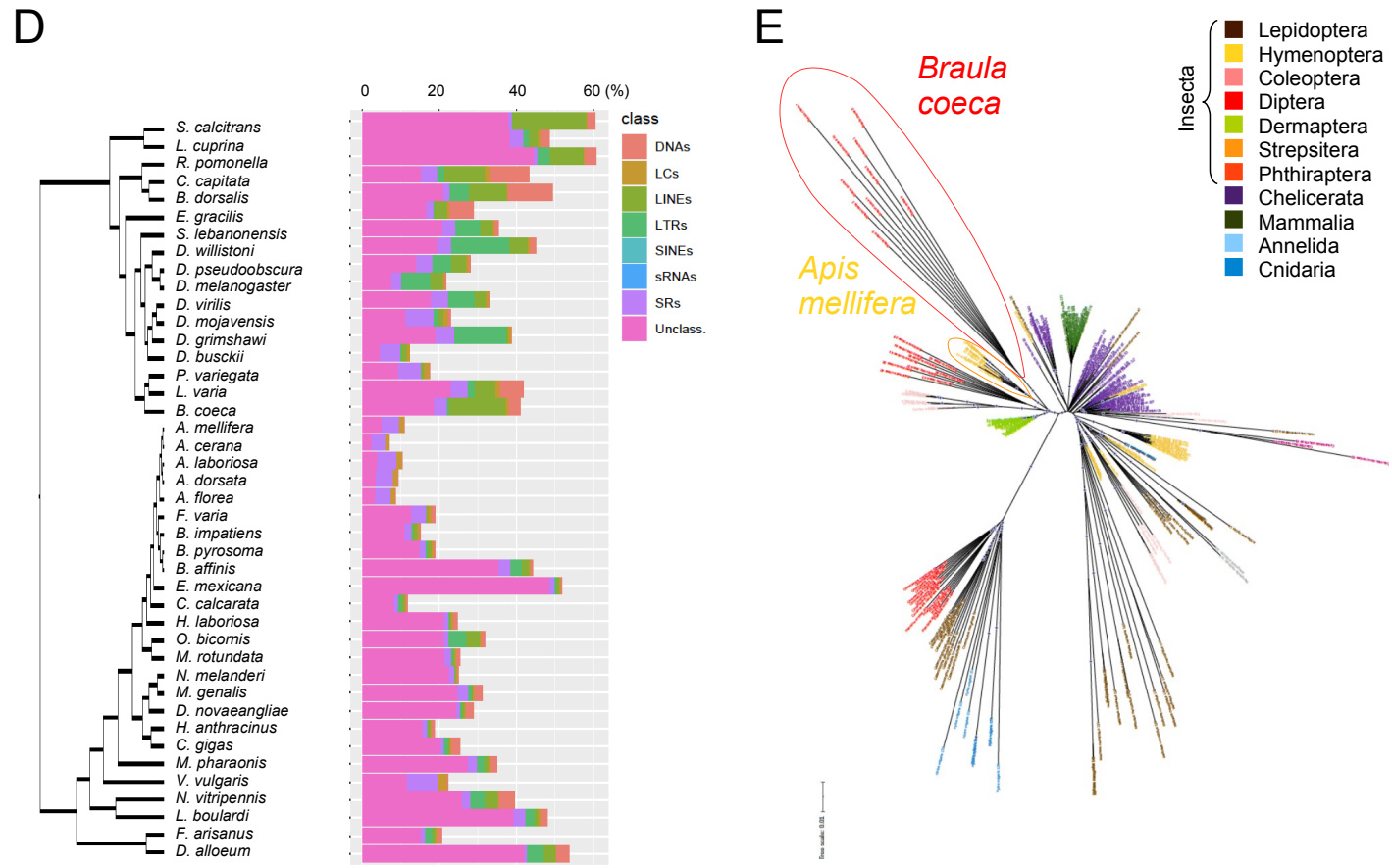
D

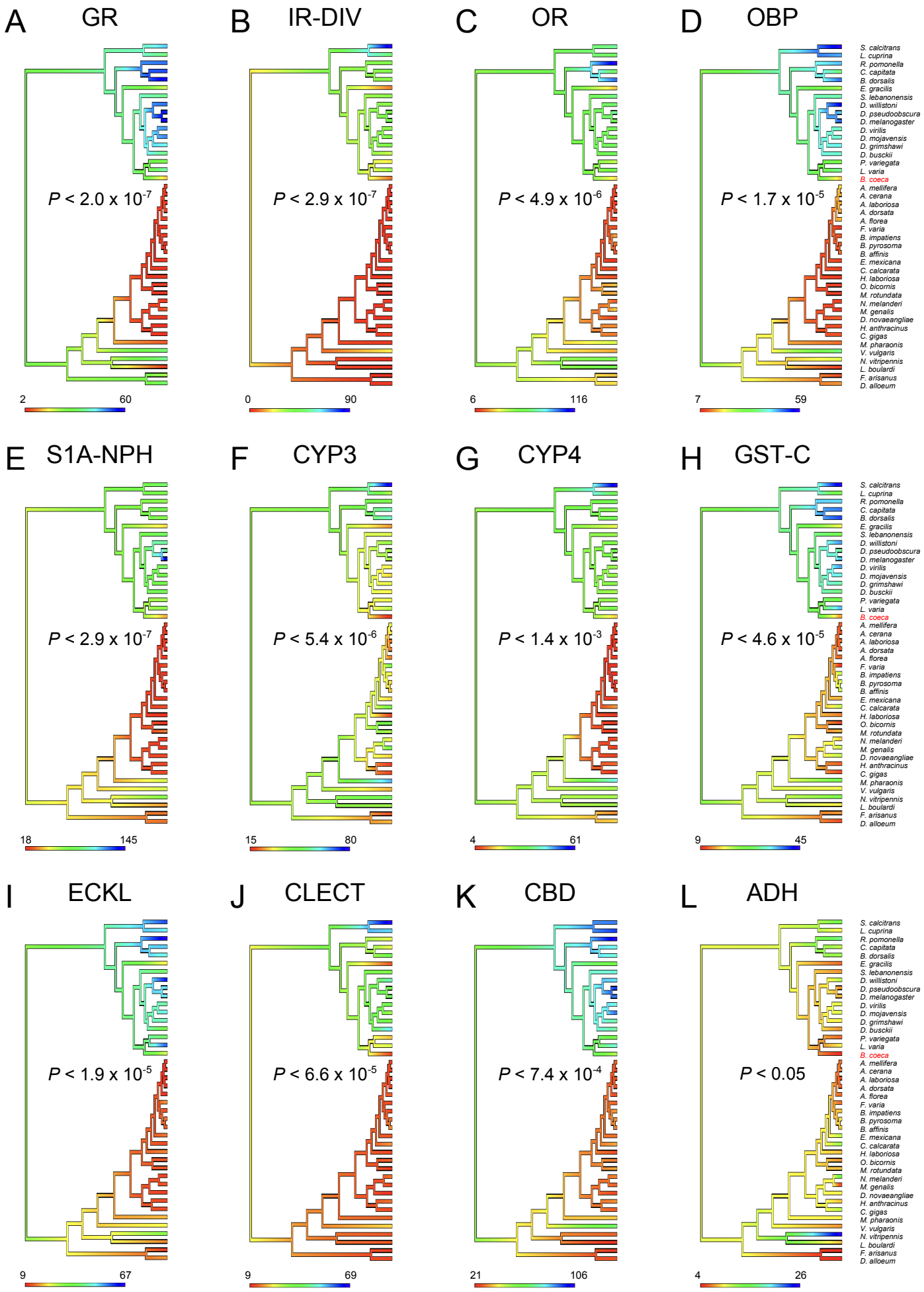


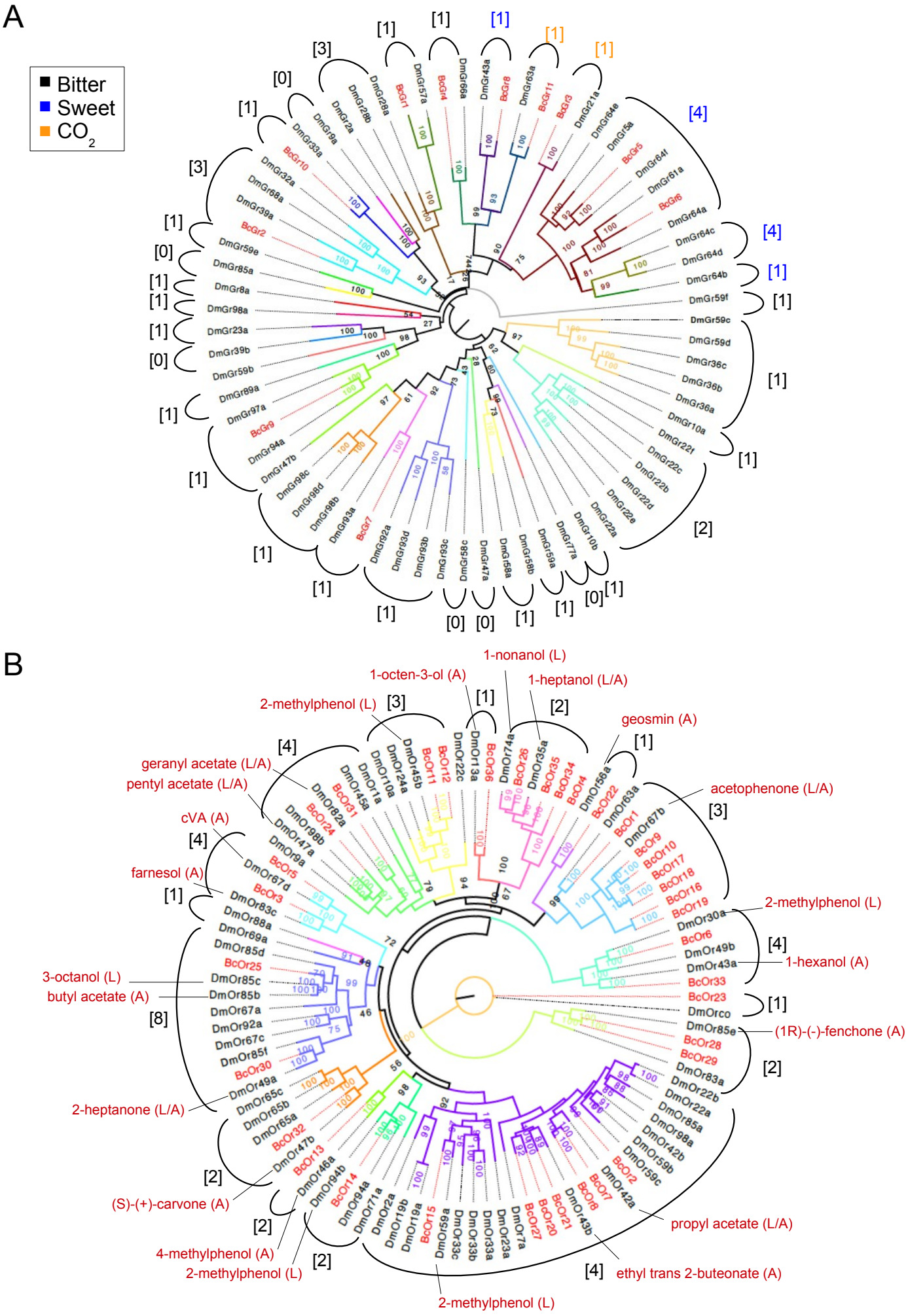
Predominant niche

- Algae + marsh plants
- Dung
- Plant sap
- Scale insects
- Fruits
- Fungi
- Solitary bee inquilinism
- Social bee inquilinism


A
B
C
Stomoxys calcitrans
Lucilia cuprina
Rhagleotis pomonella
Ceratitis capitata
Bactrocera dorsalis
Ephydra gracilis
Scaptodrosophila lebanonensis
Drosophila willistoni
Drosophila pseudoobscura
Drosophila melanogaster
Drosophila virilis
Drosophila mojavensis
Drosophila grimshawi
Drosophila busckii
Phortica variegata
Leucophenga varia
Braula coeca
Apis mellifera
Apis cerana
Apis laboriosa
Apis dorsata
Apis florea
Frieseomelitta varia
Bombus impatiens
Bombus pyrosoma
Bombus affinis
Eufriesea mexicana
Ceratina calcarata
Habropoda laboriosa
Osmia bicornis
Megachile rotundata
Nomia melanderi
Megalopta genalis
Dufourea novaeangliae
Hylaeus anthracinus
Colletes gigas
Monomorium pharaonis
Vespula vulgaris
Nasonia vitripennis
Leptopilina boulardi
Fopius arisanus
Diachasma alloeum
72
95
87
63
Carboniferous
Permian
Triassic
Jurassic
Cretaceous
Cenozoic
350
300
250
200
150
100
50
0
myr
0
Genome size (Mb)
1200
0
Gene content
20000


D
E
S. calcitrans
L. cuprina
R. pomonella
C. capitata
B. dorsalis
E. gracilis
S. lebanonensis
D. willistoni
D. pseudoobscura
D. melanogaster
D. virilis
D. mojavensis
D. grimshawi
D. busckii
P. variegata
L. varia
B. coeca
A. mellifera
A. cerana
A. laboriosa
A. dorsata
A. florea
F. varia
B. impatiens
B. pyrosoma
B. affinis
E. mexicana
C. calcarata
H. laboriosa
O. bicornis
M. rotundata
N. melanderi
M. genalis
D. novaeangliae
H. anthracinus
C. gigas
M. pharaonis
V. vulgaris
N. vitripennis
L. boulardi
F. arisanus
D. alloeum
0
20
40
60 (%)
class
DNAs
LCs
LINEs
LTRs
SINEs
sRNAs
SRs
Unclass.
Braula coeca
Apis mellifera
Insecta
Lepidoptera
Hymenoptera
Coleoptera
Diptera
Dermaptera
Strepsitera
Phthiraptera
Chelicerata
Mammalia
Annelida
Cnidaria


A
GR
P < 2.0 x 10-7
2
60
B
IR-DIV
P < 2.9 x 10-7
0
90
C
OR
P < 4.9 x 10-6
6
116
D
OBP
P < 1.7 x 10-5
7
59
E
S1A-NPH
P < 2.9 x 10-7
18
145
F
CYP3
P < 5.4 x 10-6
15
80
G
CYP4
P < 1.4 x 10-3
4
61
H
GST-C
P < 4.6 x 10-5
9
45
I
ECKL
P < 1.9 x 10-5
9
67
J
CLECT
P < 6.6 x 10-5
9
69
K
CBD
P < 7.4 x 10-4
21
106
L
ADH
P < 0.05
4
26
S. calcitrans
L. cuprina
R. pomonella
C. capitata
B. dorsalis
E. gracilis
S. lebanonensis
D. willistoni
D. pseudoobscura
D. melanogaster
D. virilis
D. mojavensis
D. grimshawi
D. busckii
P. variegata
L. varia
B. coeca
A. mellifera
A. cerana
A. laboriosa
A. dorsata
A. florea
F. varia
B. impatiens
B. pyrosoma
B. affinis
E. mexicana
C. calcarata
H. laboriosa
O. bicornis
M. rotundata
N. melanderi
M. genalis
D. novaeangliae
H. anthracinus
C. gigas
M. pharaonis
V. vulgaris
N. vitripennis
L. boulardi
F. arisanus
D. alloeum


A
Bitter
Sweet
CO2
B
1-nonanol (L)
1-octen-3-ol (A)
1-heptanol (L/A)
2-methylphenol (L)
geosmin (A)
geranyl acetate (L/A)
pentyl acetate (L/A)
acetophenone (L/A)
cVA (A)
farnesol (A)
2-methylphenol (L)
1-hexanol (A)
3-octanol (L)
butyl acetate (A)
(1R)-(-)-fenchone (A)
2-heptanone (L/A)
(S)-(+)-carvone (A)
propyl acetate (L/A)
4-methylphenol (A)
2-methylphenol (L)
ethyl trans 2-buteonate (A)
2-methylphenol (L)