

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

Main Manuscript for No Evidence That Haplodiploidy Favours the Evolution of Eusociality.

Rosa Bonifacii¹, Louis Bell-Roberts², Alan Grafen¹ and Stuart West¹.

¹ Department of Biology, University of Oxford, 11a Mansfield Road, Oxford, OX1 3SZ. UK.

² Georgina Mace Centre for the Living Planet, Imperial College London, Silwood Park Campus, Ascot, SL5 7PY

*Rosa Bonifacii.

Email: rosa.bonifacii@biology.ox.ac.uk

PNAS strongly encourages authors to supply an [ORCID identifier](#) for each author. Do not include ORCIDs in the manuscript file; individual authors must link their ORCID account to their PNAS account at www.pnascentral.org. For proper authentication, authors must provide their ORCID at submission and are not permitted to add ORCIDs on proofs.

Author Contributions: R.B., S.W., and L.B.R. conceived the project. R.B. curated the data and performed the analyses with advice from A.G. and L.B.R. L.B.R. edited the phylogenetic tree. S.W. and R.B. wrote the manuscript with comments and editing by L.B.R and A.G. S.W. acquired funding and supervised the project.

Competing Interest Statement: The authors declare no competing interests.

Classification: Biological Sciences - Evolution

Keywords: haplodiploidy, eusociality, social evolution, insects, phylogenetic pseudoreplication.

This PDF file includes:

Main Text

Figures 1 to 5

27 **Abstract**

28 The potential role of haplodiploid sex determination in promoting the evolution of altruism and
29 eusociality has been the subject of intense debate for over 50 years. Different theoretical models
30 have suggested that haplodiploidy influences relatedness in a way that either does, or does not,
31 make it easier for altruism to evolve. This debate over the 'haplodiploidy hypothesis' can only be
32 resolved with a decisive empirical test that controls for potential phylogenetic bias. Here, we
33 critically examine the current state of evidence for an adaptive link between haplodiploidy and
34 eusociality, applying phylogenetically informed methods to ensure that statistical tests reflect
35 independent evolutionary transitions. Using data from 5678 species, across all insect groups, we
36 find no evidence that haplodiploidy favours an increased rate of eusocial evolution in insects. We
37 show that this result is robust to: (a) different analytical approaches; (b) alternative ways of
38 defining both eusociality and haplodiploidy; and (c) noise in eusociality assignments for uncertain
39 clades. Our analyses suggest that previously reported associations between haplodiploidy and
40 eusociality are likely to have been artefacts, false positive results primarily driven by a high
41 transition rate to eusociality within the Hymenoptera. This high transition rate could be explained
42 by any factor associated with that group, such as parental care, monogamy or the possession of
43 a powerful sting.

44 **Significance Statement**

45 Social insects represent one of the highest levels of social organisation. In many species,
46 reproduction is monopolised by a single queen whose offspring are raised by sterile workers. For
47 over 50 years researchers have debated whether such extreme worker altruism is facilitated by
48 haplodiploid sex determination, whereby males develop from unfertilized (haploid) eggs and
49 females from fertilized (diploid) eggs. Haplodiploidy can cause female helpers to be more closely
50 related to their sisters than their own offspring, potentially favouring the altruistic rearing of
51 siblings. Despite extensive theoretical debate, there has been no decisive empirical test of this
52 "haplodiploidy hypothesis". Our analysis of 5678 species, across all insect groups, finds no
53 evidence that haplodiploidy has driven an increased rate of eusocial evolution.

54
55 **Main Text**

56
57 **Introduction**

58 The eusocial insects represent one of the pinnacles of social life (Figure 1; 1–4). A typical colony
59 of leaf-cutter ants consists of a single reproducing queen and up to 10 million workers. These
60 workers are the ultimate altruists - they are completely sterile and devote their entire adult lives to
61 rearing the offspring laid by the queen. The eusocial insects are dominated by the Hymenoptera,
62 the wasps, ants and bees. Although this order comprises only 6-8% of insect diversity, it is
63 thought to have produced more independent origins of eusociality than all other insects
64 combined.

65
66 Hamilton suggested that eusociality has arisen so frequently in the Hymenoptera because they
67 have haplodiploid sex determination, where males develop from unfertilized (haploid) eggs and
68 females from fertilized (diploid) eggs (5, 6). Haplodiploid sex determination leads to females being
69 more highly related to their sisters ($R=0.75$; assuming monandry) than they are their daughters
70 ($R=0.5$), which may make it easier for altruism to evolve. This 'haplodiploidy hypothesis' has
71 received support, or mixed support, from three empirical studies, which have reported that the
72 transition rate to eusociality is higher in haplodiploid lineages than in non-haplodiploid (7–9).

73

74 The support for the haplodiploidy hypothesis is, however, potentially misleading. The methods
75 used to test for correlated evolution assume that all evolutionary transitions to eusociality are
76 phylogenetically independent events (10). This can be a reasonable assumption if transitions are
77 spread across a phylogenetic tree. If this is not the case, and transitions are clumped, then
78 problems of nonindependence arise (11–13). As a worst-case scenario, imagine that n transitions
79 occurred in response variable y , within a clade of a tree where the explanatory variable x does
80 not change state. While this would contribute n data points to an analysis, they are not
81 independent, and are also correlated with anything shared by the members of that clade, not just
82 the state of x .

83
84 This worst-case scenario of ‘pseudoreplication within lineages’ is exactly what has happened in
85 the Hymenoptera. There has only been one evolutionary transition to haplodiploidy, at the base of
86 this group, and so the different transitions to eusociality within Hymenoptera do not represent
87 independent trials of the hypothesis that haplodiploidy promotes eusocial evolution - they
88 represent only one (Fig. 2; 11). The correlation between haplodiploidy and eusociality could
89 therefore be an artefact of eusociality being particularly promoted in the Hymenoptera, due to any
90 other factors that are shared across this group. For example, both parental care and monogamy
91 being common, or the possession of a powerful sting to facilitate group defence (14–22). This
92 potential problem is so great that the pattern in the Hymenoptera has previously been used as a
93 case study to illustrate the exact problem of pseudoreplication within lineages (11).

94
95 While this problem is significant, it is not insurmountable, and phylogenetically informed methods
96 exist that are able to account for this lack of independence within lineages. We used a dataset of
97 5,678 species spanning all major insect groups to test the haplodiploidy hypothesis while
98 explicitly controlling for pseudoreplication within lineages. To assess the robustness of our
99 conclusions, we used three different analytical approaches. First, we applied Read & Nee’s
100 (1995) method of sister-taxon comparison, which transforms the data into phylogenetically
101 independent comparisons, each representing a single transition in the explanatory variable (11).
102 Second, we used three phylogenetic regression approaches, which make different assumptions
103 in order to deal with the underlying problem of non-independence: (a) Grafen’s (1989)
104 phylogenetic regression (23); (b) Ives & Garland’s (2010) alternative phylogenetic logistic
105 regression (24); and (c) Adams et al.’s (2024) robust phylogenetic regression (25). Third, we used
106 Boyko and Beaulieu’s (2023) hidden-Markov models (HMM), which examine the possible role of
107 alternative drivers by explicitly introducing a ‘hidden state’ (26). This state may represent any
108 number of unobserved factors, biological or otherwise, that is correlated with the rate of eusocial
109 evolution across the phylogeny.

110 **Results**

111 We compiled a dataset of 5678 species, which could be placed on Chesters’ (2017) phylogeny of
112 insects (27), and for which data were available on both the genetics of sex determination and
113 social traits.

114
115 **Distribution of Haplodiploidy and Eusociality.** Independent origins of haplodiploidy and
116 eusociality were identified using 9aI state reconstructions under an “all rates different” model.
117 Overall, 21% of the species we examined were haplodiploid by our most inclusive definition.
118 Using maximum likelihood ancestral state reconstructions, we inferred 10 independent transitions
119 to haplodiploidy across the insect tree. We also tested the robustness of our analyses by
120 repeating them including only arrhenotokous species (excluding those species which have

121 paternal genome elimination). By this definition, 19.6% of species in our dataset are classified as
122 haplodiploid, comprising five independent origins.

123
124 The occurrence and number of transitions to eusocial behaviour depends on how it is defined. A
125 potential source of confusion is that 'eusociality' is used inconsistently across studies and
126 textbooks (See Methods) (4, 28–32). Consequently, to test the robustness of our conclusions we
127 repeated the analyses with three possible definitions, that are increasingly restrictive, and which
128 allow us to focus on specific traits. First, for our broadest categorisation, we used Michener and
129 Wilson's definition of eusociality which incorporates all group living species exhibiting both: (a)
130 cooperative brood care; and (b) reproductive division of labour where 'helpers' aid 'breeders'
131 (behavioural castes) (1, 33, 34). Using this definition, ancestral state reconstructions estimate that
132 eusociality has arisen 16 times (haplodiploid: 12 times, diploid: 4 times).

133
134 Second, we used Crespi & Yanega's definition of eusociality with castes, which includes only
135 those cooperatively breeding species where individuals become fixed into specialised castes prior
136 to reproductive maturity (30). By this definition, eusociality has arisen 19 times, 15 times in
137 Haplodiploid lineages and 4 times in diploids.

138
139 Third, for our most restrictive definition, we consider only species which have gone through a
140 major evolutionary transition to superorganismality (here characterised by morphological caste
141 differentiation and worker sterility) (3, 4, 35). Superorganismality was found to have arisen 6
142 times, 4 times in haplodiploid lineages and twice in diploid. We provide a graphical illustration of
143 the estimated phylogenetic distribution of haplodiploidy and eusociality (according to our broad
144 *Michener-Wilson* definition) in Fig. 3.

145
146 Nb. These transition counts reflect estimates inferred from the taxa sampled in our dataset. While
147 they provide valuable insight into evolutionary patterns, they should be interpreted as indicative
148 rather than exact measures of the "true" number of transitions. A table showing transition counts
149 for haplodiploid vs. diploid lineages, as well as for Hymenoptera vs. non-Hymenoptera, can be
150 found in Supplementary Table S1. For a comprehensive depiction of all inferred transitions across
151 the phylogeny, see Fig.S1.

152
153 **Sister-Taxon Comparison.** Read & Nee's (1995) method of sister-taxon comparison did not
154 support the hypothesis that haplodiploidy is associated with a higher rate of eusocial evolution in
155 haplodiploid insects. Examining Michener & Wilson's definition of eusociality, we transformed the
156 dataset into a ten independent contrasts between haplodiploid and non-haplodiploid clades, each
157 representing a single transition in haplodiploidy (Fig. 4). Four of these contrasts showed a
158 difference in the proportion of eusocial species between sister clades: three showed a higher
159 incidence of eusociality in the haplodiploid clade, and one showed a higher incidence in the
160 diploid clade (Binomial test: $p = 0.32$; 95% CI: 0.25 – 1). The conclusion of no significant influence
161 of haplodiploidy was robust to: (i) alternative analyses that included all contrasts (Wilcoxon
162 signed-rank test: $p = 0.24$, $Z = 0.93$); (ii) choice of contrast sister clade pairing (in 100 replicate
163 datasets using alternative plausible sister clades, none produced a significant result with $p < 0.05$,
164 Fig.S2) and; (iii) alternative definitions of eusociality and haplodiploidy (Supplementary Table S2).

165
166 **Phylogenetic Regression.** With all three phylogenetic regression approaches, we reached the
167 same conclusion, that there was no evidence of eusociality evolving at a higher rate in
168 haplodiploid insects. Results for Michener and Wilson's definition of eusociality are as follows -
169 Grafen's (1989) Phylogenetic regression: $p = 0.94$, $F_{1,147} = 0.005$; Ives & Garland (2010) binomial

170 regression: $p = 0.90$, $Z = 0.13$); Adams et al. (2024) Robust regressions: $p = 0.94$, $Z = -0.07$. We
171 also reached the same conclusion when using alternative definitions of both haplodiploidy and
172 eusociality (Supplementary Table S3)

173

174 **Hidden Markov models (HMIMs).** Using Boyko and Beaulieu's (2023) method of hidden-Markov
175 models (HMM) (26), we also reached the same conclusion that there was no evidence that
176 haplodiploidy is associated with a higher rate of eusocial evolution in haplodiploid insects. The full
177 two-rate-class hidden Markov independent model was the best-supported model, with by far the
178 lowest AIC, especially versus the single rate "dependent" model (log-likelihood = -208.42, $K = 10$
179 parameters, ΔAIC of 51.12). This implies that when we allow the rates of trait-change to vary in
180 an unlinked way, there is no residual evidence that haplodiploidy itself predicts eusociality. What
181 looked like a correlation under simpler models is better explained by a so-called "hidden state",
182 which may represent any number of unobserved factors, biological or otherwise.

183

184 For these analyses we reached the same conclusion under five of the six possible combinations
185 of eusociality and haplodiploidy trait definitions. The only exception was the combination of Crespi
186 & Yanega's eusociality and the "strict" haplodiploidy definition, where the dependent HMM model
187 was better supported (Supplementary Table S4). A key caveat with the HMM method is that,
188 although it can mitigate potential bias by absorbing heterogeneity in transition rates, it does not
189 explicitly control for (eliminate) pseudoreplication within lineages. This issue of pseudoreplication
190 is likely to be more pronounced under the Crespi & Yanega definition because this categorisation
191 is harder to score reliably, leading to noise that can increase the number of apparent eusocial
192 transitions within Hymenoptera. To assess whether the support for the dependent model reflects
193 reliable biological signal rather than uncertainty in trait coding rather than, we examined the
194 possible influence of classification error to our results.

195 **Robustness to classification error.** To test whether our results were sensitive to potential
196 classification error, we re-ran all phylogenetic regressions and Hidden Markov models after
197 introducing 10% random classification noise in eusociality assignments within socially variable
198 clades. Across all replicate datasets and all six combinations of trait definitions, the three
199 phylogenetic regression approaches consistently recovered a non-significant effect of
200 haplodiploidy (all $p > 0.05$, Supplementary Tables S5- S7). We did not rerun the sister-taxon
201 analyses because these broad-scale comparisons will be unaffected by small amounts of
202 misclassification, and the results would be unchanged.

203 Model rankings under the Hidden Markov framework were similarly stable. Across all replicates,
204 the two-rate-class independent HMM model was most frequently favoured, though the strength of
205 support varied with trait definitions (Supplementary Table S8). In four of the six combinations, the
206 independent model dominated the replicates (proportion of "wins" = 0.68–1.00). Under the Crespi
207 and Yanega definition, results showed no consistent support for either dependent or independent
208 models (0.56 for both strict and inclusive). It appears that results for this trait definition are
209 unstable and highly sensitive to trait uncertainty. Nonetheless, even under substantial (10%)
210 reclassification noise, no trait combination reversed our main conclusion.

211 **Discussion**

212 We found no support for the hypothesis that haplodiploid sex determination has favoured the
213 evolution of eusociality. This conclusion was robust to five different methods of analysis, and

214 three possible definitions of eusociality (Fig. 3-5). Our conclusions differ from previous analyses
215 because we specifically used methods that are robust to the problem of pseudoreplication within
216 lineages. Previous analyses are likely to have been biased by the high transition rate to
217 eusociality within the Hymenoptera, which only represents a single unreplicated transition in
218 haplodiploidy (Fig. 2) (11–13).

219
220 We found only 10 independent transitions between haplodiploid and diploid clades across the
221 insects. This number of evolutionary transitions is not high, and so an important limitation of any
222 possible analysis is that the insect phylogeny has limited statistical power to test the role of
223 haplodiploidy. However, amongst these 10 transitions, six showed no difference in the prevalence
224 of eusociality, one showed a higher rate of eusocial evolution in the diploid clade and only three
225 showed a higher transition rate in the haplodiploid clade (Fig. 5). This is not a strong suggestive
226 pattern. Of course, it is always a possibility that there is an influence that it is too weak to detect.
227 For example, if we just focus on the transitions where there is a difference, and assume that the
228 3:1 ratio is accurate, then approximately 30 transitions would be required to have an 80% chance
229 of detecting a significant result. Nonetheless, and given all possible caveats, the empirical data
230 clearly do not provide significant support for the haplodiploidy hypothesis.

231
232 Recent support for the haplodiploidy hypothesis come from analyses based on Pagel's test for
233 correlated evolution (8, 9). Pagel's method implicitly assumes that all transitions to eusociality are
234 independent evolutionary events (7–9). This is a reasonable assumption when transitions are
235 distributed across a phylogeny, but not when they are clumped within groups where the
236 explanatory variable does not vary (11–13). Multiple simulations have clearly demonstrated that
237 these models are likely to incorrectly infer false-positive associations under such circumstances
238 (13, 36). This scenario of 'pseudoreplication within lineages' occurs in our data set because of the
239 multiple transitions to eusociality in the Hymenoptera, after a single transition to haplodiploidy.
240 Our conclusions differed from these previous studies because we used methods that specifically
241 deal with this pseudoreplication problem. Joshi & Wiens provided mixed support for the
242 haplodiploidy hypothesis because they used a combination of methods that both do, and don't,
243 lead to this pseudoreplication (8).

244
245 The previous false positive results are likely to be a result of the "Hymenoptera effect" - an
246 artefact specifically resulting from the high transition rate to eusociality within the Hymenoptera.
247 Indeed, our hidden-Markov model analysis suggests that the significant result found in previous
248 analyses is better explained by an unknown (hidden) variable which could represent any
249 combination of factors. While the high transition rate in the Hymenoptera is correlated with
250 haplodiploidy, it could equally be caused by any other factor shared across this group. For
251 example, both parental care and monogamy being common, or the possession of a powerful sting
252 to facilitate group defence (14–22). A period of extended parental care with an appreciable
253 possibility of parental mortality provides an efficiency benefit to cooperative brood care (15–17,
254 37–39). Monogamy leads to a high relatedness within family groups, increasing the indirect (kin
255 selected) benefit that can be gained from helping rear siblings (20, 22, 40). Examining the
256 influence of these factors across all insects is a key future step.

257
258 There has been considerable theoretical debate over whether haplodiploidy should favour
259 altruism and therefore facilitate the evolution of eusociality or superorganismality. While
260 haplodiploidy can make females more highly related to their sisters ($R=0.75$) than they are to their
261 daughters ($R=0.5$), it also causes females to be less related to their brothers ($R=0.25$) than they
262 are to their sons ($R=0.5$) (6, 41). Consequently, whether haplodiploidy facilitates altruism can

263 depend upon life history details, and especially the pattern of sex allocation (41). Numerous
264 theoretical studies have shown that certain patterns of sex allocation, such as sex ratio variation
265 between partially overlapping generations, can lead to haplodiploidy favouring altruism (41–47,
266 47–53). But it is not clear that those patterns occurred during the evolution of eusociality. It has
267 also been shown that certain circumstances may lead to haplodiploidy disfavouring the evolution
268 of altruism (47, 49, 54). Consequently, biological details matter in a way that prevents a clear
269 prediction being made for whether haplodiploidy alone is likely to favour the evolution of altruism.
270 Our analysis has shown that, at present, we have no evidence that haplodiploidy has consistently
271 led to a higher rate of evolution of eusociality. A lack of support for the haplodiploidy hypothesis
272 has also been found when examining which sex helps. The helping sex is explained by the
273 ecological benefit and form of helping (brood care or defence) and not by the method of sex
274 determination (55).

275
276 Our results also highlight the importance of careful consideration of the underlying structure of the
277 data when conducting comparative analyses. Simply accounting for phylogeny is often not
278 enough, we must carefully consider the power we have to test any given hypothesis by
279 considering the number of independent transitions, not only in the response variable, but also in
280 the explanatory variables of interest. Failure to do this can result in potentially false positive
281 results and erroneous conclusions.

282
283 To conclude, we have provided a decisive negative test of the haplodiploidy hypothesis. We
284 emphasise that this is not a problem for inclusive fitness theory, because inclusive fitness theory
285 does not necessarily predict that haplodiploidy should matter. Indeed, one of us has previously
286 suggested that in the biologically most plausible scenarios, the influence of haplodiploidy was
287 likely to have been negligible or even to hinder the evolution of eusociality (47, 49, 54). In
288 contrast, there is considerable support for the cases where inclusive fitness can make clear
289 predictions, such as how altruistic helping and eusociality can be favoured by the role of
290 monogamy in producing a high relatedness; or the role of selection pressures such as 'life
291 insurance' in species with extended parental care and 'fortress defence' providing ecological
292 benefits to cooperation (15–22, 37, 38, 56–58). Perhaps ironically, it appears that haplodiploidy
293 has been more important for explaining conflict rather than cooperation within social insect
294 species. There is clear empirical support for inclusive fitness predictions over how the relatedness
295 asymmetries generated by haplodiploidy can lead to conflict within haplodiploid social groups,
296 resulting in both 'worker policing' and 'split sex ratios' (59–62).

297 298 **Materials and Methods**

299 **Sex Determination.** We obtained data on sex determination mechanisms from the Tree of Sex
300 database which compiles information on reproductive systems across eukaryotic life
301 (<https://treeofsex.sanger.ac.uk/database>). To supplement this data set, we also obtained
302 information on known haplodiploid lineages from a published review of sex determination
303 mechanism in insects (63).

304 We initially categorised species as diploid or haplodiploid with an inclusive definition of
305 haplodiploidy which includes species with either: (a) arrhenotoky, where males develop from
306 unfertilized eggs and are haploid; or (b) paternal genome elimination, in which males inherit a full
307 diploid genome but later discard the paternal contribution in either somatic or germline cells.
308 Although arrhenotoky and parental genome elimination differ mechanistically, both systems result

309 in equivalent patterns of genetic inheritance and relatedness structures, leading to comparable
310 evolutionary consequences in the context of our study (64).

311 To test the robustness of our results, we repeated our analyses including only arrhenotokous
312 species as haplodiploid. Arrhenotoky is the predominant form of haplodiploidy in insects, and the
313 traditional focus of discussions on eusocial evolution. In addition, arrhenotoky may facilitate
314 adjustment of offspring sex ratios, which can help haplodiploidy facilitate the evolution of altruism.
315 We reached the same conclusions in all analyses.

316 **Social Behaviour.** We used three approaches to systematically gather data on social behaviour
317 across all insect taxa. First, we conducted comprehensive literature searches in Web of Science
318 and Google Scholar with combinations of search terms (*eusociality*, *solitary*, *subsocial*,
319 *communal*, *quasisocial*, *semisocial*, *primitively eusocial*, *advanced eusocial*), and taxon-specific
320 terms. We focused on taxa known, or suspected, to exhibit social behaviour: Vespidae (vespid
321 wasps), Crabronidae (aphid wasps), Halictidae (sweat bees), Apidae (apid bees), Formicidae
322 (ants), Phlaeothripidae (gall-thrips), Curculionidae (bark and ambrosia beetles), Aphididae
323 (aphids), Termitoidae (termites) and Colletidae (*Amphylaeus morosus*). Second, we extracted
324 relevant data from key monographs, reviews, and existing datasets (1, 33, 66–72). Third, we
325 conducted forward and backward citation searches on all relevant papers found via the methods
326 above to identify additional sources of data.

327

328 **Eusociality.** Eusociality is defined inconsistently across studies and textbooks (30, 32). One
329 potential problem is that eusociality is used as a broad-brush category, lumping multiple
330 characteristics in way that draws attention away from distinct traits (32). However, the term is
331 deeply embedded and still widely used in the broader literature. Consequently, and to test the
332 robustness of our conclusions we repeated our analyses with three increasingly restrictive
333 definitions of eusociality.

334

335 First, for our broadest categorisation of eusociality, we used Wilson and Michener's definition.
336 Under this definition, a species was categorised as eusocial if exhibits both: (a) cooperative brood
337 care - where group members provide care for offspring that are not their own, such as
338 provisioning and defence; and (b) reproductive division of labour - where social groups have one
339 or more reproductively dominant individuals, evidenced by either consistent behavioural
340 differences (e.g., breeders vs. helpers) or by differences in ovarian development/ mating status.
341 This represents all species exhibiting reproductive altruism, and includes 'primitively eusocial'
342 species in which all group members retain reproductive totipotency (1, 73).

343

344 For our second definition, we use Crespi & Yanega's more restrictive categorisation: *eusociality*
345 *with castes* (see also: 18–20, 71). This definition includes only those cooperatively breeding
346 species where individuals become fixed into specialised castes prior to reproductive maturity
347 (preimaginal castes). Castes are evidenced by individuals having distinct bimodal size
348 differences, or non-allometric polymorphism that are correlated with reproductive specialisation.
349 This categorisation includes both 'facultatively' and 'obligately' eusocial species in which
350 members of at least one caste have irreversibly lost behavioural totipotency (30).

351

352 Third, for our most restrictive definition, we consider only species which have gone through a
353 major evolutionary transition to *superorganismality*. In these species castes are mutually
354 dependent upon each other and exhibit distinct non-allometric morphological differentiation. The
355 breeder (queen) cannot reproduce without the help of the helpers (workers), and the workers can

356 only pass genes to the next generation by helping related breeders. This definition aligns with
357 Crespi & Yanega's definition of 'obligately eusocial' species (30), a major evolutionary transition
358 in individuality (35) and Boomsma's discussion of superorganisms (32). For all definitions, we
359 cross-validated data across multiple sources to ensure accuracy and consistency. The highest
360 level of social behaviour reported for a species was used for analyses.

361

362 Within socially variable clades, species were scored under a definition only when supported by a
363 species-specific source; otherwise, they were conservatively scored "unknown". Entire lineages
364 were scored only when peer-reviewed sources report universal absence or presence of
365 eusociality. The full data is provided in Dataset S1.

366

367 **Phylogenetic Tree.** We used a pruned version of the Chesters (2017) species-level phylogeny
368 (27), retaining only those taxa where complete trait data could be obtained or inferred. However,
369 as the Chesters tree does not include two social lineages for which trait data are available
370 (*Kladothrips* and *Allodapini*), we supplemented the phylogeny with additional taxa from two
371 published sources. For the thrips, we incorporated species from the McLeish et al. (2007)
372 phylogeny (75). Specifically, we replaced *Kladothrips* species in the Chesters backbone with
373 those from McLeish et al. (2007), anchoring the *Kladothrips* clade to its closest relative
374 *Bactrothrips* as identified in the Chesters tree. To incorporate the allodapine bees, we used the
375 same approach with the Sless & Rehan (2023) phylogeny of carpenter bees (76). This integration
376 ensured that both eusocial thrips and allodapine bees were represented in the combined
377 phylogeny, while maintaining consistency with the Chesters topology.

378

379 **Divergence Time Estimation.** We estimated divergence times using two approaches, reflecting
380 alternative assumptions regarding evolutionary change.

- 381 (1) Time-calibrated tree: Applying a penalized likelihood method implemented in *treePL* (77).
382 The undated maximum-likelihood tree was matched to a dated family-level reference
383 phylogeny (78) using *congruify* in *geiger*, which identifies shared nodes between trees
384 and transfers corresponding secondary calibrations. These calibrations were applied in
385 *treePL* and cross-validation was used to choose the optimal smoothing parameter (λ).
386 The resulting ultrametric tree was scaled to absolute time (millions of years).
387 (2) Grafen-scaled tree: We also used Grafen's ρ , which rescales branch lengths according to
388 the number of descendant tips rather than elapsed time. This provides an alternative
389 representation of evolutionary change that does not assume a constant-rate (Brownian
390 motion) of evolutionary change.

391 Results presented are those obtained using the time-calibrated tree; corresponding analyses
392 based on the Grafen-scaled tree can be found in Supplementary Tables S3 & S4(b).

393 **Ancestral State Reconstructions.** To estimate transitions between ploidy and eusociality states
394 across the phylogeny, we used maximum likelihood ancestral state reconstruction via the *fitMk()*
395 function in the R package *phytools* (v 2.4-4). For each combination of traits (i.e., each of the three
396 eusociality definitions paired with each of the two haplodiploidy classifications), we applied the
397 "ARD" (all rates different) model. The most likely state at each node was inferred using marginal
398 likelihoods from the fitted model, and transitions were identified by comparing the most likely
399 states of parent and descendant nodes along each branch of the phylogeny.

400 **Sister-Taxon Comparisons.** Read & Nee's (1995) method transforms the data into a set of
401 independent contrasts, each one representing a single transition in the explanatory variable,
402 thereby removing the bias that arises from multiple lineages sharing a common evolutionary
403 origin (11). Importantly, the maximum number of datapoints generated for this analysis is
404 inherently limited to the number of independent transitions in Haplodiploidy. We implemented this
405 approach by identifying a set of independent contrasts, where each contrast consisted of a
406 haplodiploid clade (single or multi-species) and a paired non-haplodiploid sister clade. These
407 contrasts formed the basis for subsequent phylogenetically informed trait comparisons.

408 Each haplodiploid clade was paired with a phylogenetically adjacent non-haplodiploid sister
409 clade. In cases where the parent node had more than two "children" nodes, several valid sister
410 clades could exist. Because all candidate sister clades share the same parent node, they are
411 equally related to the focal haplodiploid clade in evolutionary terms. To reduce potential biases
412 introduced by comparing clades of very different sizes, we selected the clade containing the
413 number of species most similar to the focal haplodiploid clade. If multiple candidate sister clades
414 were equally close in size, one was selected at random. To ensure that contrasts were
415 phylogenetically independent, candidate sister clades were only accepted if no part of their
416 lineage (i.e., tips or internal nodes) overlapped with any previously used contrast.

417 After identifying informative contrasts between haplodiploid and non-haplodiploid sister clades,
418 we then calculated the proportion of eusocial species found within each clade. Following Read
419 and Nee (1995), contrasts were considered informative if the two clades differed in their
420 proportion of eusocial species. We used two methods to carry out statistical tests, (1) focusing on
421 just informative contrasts (binomial test); and (2) also including contrasts with no differences
422 (Wilcoxon signed-rank test with 10,000 permutations).

423 To assess the robustness of our results to the specific choice of sister clades, we implemented a
424 simulation-based approach using repeated randomization. For each of 100 replicate analyses, we
425 iteratively identified valid focal clades and then selected a sister clade at random from among the
426 10 most plausible candidates. These top candidates were defined based on (a) minimal size
427 difference from the focal clade and (b) minimal phylogenetic distance (i.e., fewest nodes levels up
428 in the tree). As above, at each iteration, we ensured that no tips or nodes were reused across
429 contrasts. Each replicate generated a unique set of contrasts, and we repeated the downstream
430 statistical analysis to examine the consistency of results across different plausible sets of
431 contrasts. This approach provides a nonparametric measure of the stability of our findings and
432 helps mitigate concerns about the subjectivity or uniqueness of any single clade pairing.

433 **Phylogenetic Regressions.** To test for an association between haplodiploidy and eusociality,
434 while explicitly accounting for shared ancestry, we applied three alternative phylogenetic
435 regression approaches.

436 First, we used Grafen's (1989) phylogenetic regression as implemented in the phyregB package
437 (80), which limits the degrees of freedom to the number of independent evolutionary events rather
438 than the number of species. Grafen's method directly transforms our branch lengths according to
439 the number of descendant tips (diversification) rather than elapsed time. This p -adjusted
440 phylogeny forms the basis for test statistics, yielding a robust assessment of the association
441 between haplodiploidy and eusociality (13, 81). Grafen & Ridley (1996) have previously

442 demonstrated the validity of using Grafen's phylogenetic regression for discrete data such as
443 'eusociality' (81).

444 Second, we used implemented phylogenetic logistic regression (phyloglm; (24). Models were fit
445 with the R package phylolm v2.6.2 using the function phyloglm (method="logistic_MPLE"). We
446 ran 1000 parametric bootstrap replicates to obtain 95% confidence intervals for both regression
447 coefficients and the phylogenetic-signal parameter (α), which controls the decay of covariance
448 with tree distance.

449 Third, we used robust phylogenetic regression (ROBRT), which has been explicitly validated to
450 remain reliable under scenarios of unreplicated evolution (25). We implemented the L1 (least-
451 absolute-deviations) estimator within a PGLS framework, retaining Pagel's λ to model
452 phylogenetic covariance but replacing the usual squared-residual with absolute deviations. By
453 leveraging these three approaches - each with different assumptions about branch-scaling,
454 residual structure, and robustness to outliers - we were able to test the robustness of our
455 conclusions.

456 **Hidden-Markov Models.** One final method that has been proposed to resolve the problem of
457 unreplicated evolution are hidden-Markov models (25). HMMs allow us to introduce a so-called
458 "hidden state," which can represent any number of unobserved factors, biological or otherwise.
459 Based on the presence or absence of this hidden state, changes between observed states are
460 allowed to vary. By explicitly accounting for background rate heterogeneity not tied to the focal
461 trait, HMIM models provide a more realistic framework for evaluating potential trait correlations in
462 the presence of singular evolutionary events (i.e., unreplicated bursts).

463 Following Boyko and Beaulieu's (2023) methodology, we fit four alternative discrete-trait models
464 using the corHMM package; 1) a dependent model based on Pagel's (1995) correlated evolution
465 model, where the transition rate in one trait depends on the state of the other (10); 2) an
466 independent model where each trait does not depend to the state of the other; 3) HMIM (Hidden
467 Markov Independent Model), a character-independent model that allows rates of evolution to vary
468 across the tree according to an unobserved (hidden) rate class, thereby capturing rate
469 heterogeneity not attributable to the traits themselves, and 4) HMM (Hidden Markov Dependent
470 Model), a character-dependent model that allows rates of evolution to vary across the tree
471 according to two rate classes, in these models transitions rates are attributable to haplodiploidy.
472 Each model was fit using maximum likelihood, and model performance was compared using AIC
473 scores and Akaike weights.

474 **Error-robustness analysis.** To evaluate whether our findings were sensitive to potential
475 misclassification of eusociality, we performed a structured sensitivity analysis across all three
476 definitions of eusociality. For each definition, we generated 25 replicate datasets in which 10% of
477 species were randomly reclassified within clades known to exhibit variability or uncertain
478 classification (for clade details, by eusociality trait definition, see Supporting text). Flips were
479 confined to these socially variable lineages and applied in both directions (0→1 and 1→0). Each
480 replicate dataset was then re-analysed using both phylogenetic regression and corHMM

481 Nb. Read and Nee sister-clade analyses were not repeated, as these comparisons (based on
482 proportions of eusocial vs non-eusocial species in matched clades) are expected to be largely

483 unaffected by small amounts of misclassification. Instead, separate sensitivity checks were
484 performed for these analyses by varying the choice of sister taxa (see above).

485 **Acknowledgments and Funding Sources.**

486
487 We thank Dr. Laurence Belcher for helpful comments on the manuscript, and Dr. Anna Dewar for
488 advice on analysis and valuable discussions. This research was supported by funding from the
489 European Research Council (ERC).

490

491 **References**

- 492 1. E. O. Wilson, *The insect societies*. (Cambridge, Massachusetts, USA, Harvard University
493 Press [Distributed by Oxford University Press, 1971).
- 494 2. B. Holldobler, E. O. Wilson, *The Superorganism – The Beauty, Elegance and Strangeness of*
495 *Insect Societies* (W. W. Norton & Company, 2009).
- 496 3. A. F. G. Bourke, *Principles of Social Evolution* (Oxford University Press, 2011).
- 497 4. J. J. Boomsma, *Domains and Major Transitions of Social Evolution* (Oxford University Press,
498 2022).
- 499 5. W. D. Hamilton, The genetical evolution of social behaviour. II. *J. Theor. Biol.* **7**, 17–52
500 (1964).
- 501 6. W. Hamilton, Altruism and Related Phenomena, Mainly in Social Insects. *Annu. Rev. Ecol.*
502 *Syst.* 193–232 (1972).
- 503 7. R. H. Crozier, Advanced eusociality, kin selection and male haploidy. *Aust. J. Entomol.* **47**,
504 2–8 (2008).
- 505 8. C. H. Joshi, J. J. Wiens, Does haplodiploidy help drive the evolution of insect eusociality?
506 *Front. Ecol. Evol.* **11** (2023).
- 507 9. S. Suresh, T. A. Linksvayer, Eusociality evolved at a much higher rate in haplodiploid
508 insects. [Preprint] (2025). Available at:
509 <http://biorxiv.org/lookup/doi/10.1101/2025.01.14.633090> [Accessed 22 January 2025].
- 510 10. M. Pagel, Detecting correlated evolution on phylogenies: a general method for the
511 comparative analysis of discrete characters. *Proc. R. Soc. Lond. B Biol. Sci.* **255**, 37–45
512 (1994).
- 513 11. A. F. Read, S. Nee, Inference from binary comparative data. *J. Theor. Biol.* **173**, 99–108
514 (1995).
- 515 12. A. Grafen, M. Ridley, A New Model for Discrete Character Evolution. *J. Theor. Biol.* **184**, 7–
516 14 (1997).
- 517 13. W. P. Maddison, R. G. FitzJohn, The Unsolved Challenge to Phylogenetic Correlation Tests
518 for Categorical Characters. *Syst. Biol.* **64**, 127–136 (2015).

- 519 14. J. W. Stubblefield, E. L. Charnov, Some conceptual issues in the origin of eusociality.
520 *Heredity* **57**, 181–187 (1986).
- 521 15. D. C. Queller, The evolution of eusociality: Reproductive head starts of workers. *Proc. Natl.*
522 *Acad. Sci.* **86**, 3224–3226 (1989).
- 523 16. D. C. Queller, Extended parental care and the origin of eusociality. *Proc. R. Soc. Lond. B*
524 *Biol. Sci.* **256**, 105–111 (1994).
- 525 17. R. Gadagkar, Evolution of Eusociality: The Advantage of Assured Fitness Returns. *Philos.*
526 *Trans. Biol. Sci.* **329**, 17–25 (1990).
- 527 18. R. Gadagkar, Demographic Predisposition to the Evolution of Eusociality: A Hierarchy of
528 Models. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10993–10997 (1991).
- 529 19. J. J. Boomsma, Kin Selection versus Sexual Selection: Why the Ends Do Not Meet. *Curr.*
530 *Biol.* **17**, R673–R683 (2007).
- 531 20. J. J. Boomsma, Lifetime monogamy and the evolution of eusociality. *Philos. Trans. R. Soc. B*
532 *Biol. Sci.* **364**, 3191–3207 (2009).
- 533 21. J. J. Boomsma, Beyond promiscuity: mate-choice commitments in social breeding. *Philos.*
534 *Trans. R. Soc. B Biol. Sci.* **368**, 20120050 (2013).
- 535 22. W. O. H. Hughes, B. P. Oldroyd, M. Beekman, F. L. W. Ratnieks, Ancestral Monogamy
536 Shows Kin Selection Is Key to the Evolution of Eusociality. *Science* **320**, 1213–1216 (2008).
- 537 23. A. Grafen, The phylogenetic regression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **326**, 119–
538 157 (1989).
- 539 24. A. R. Ives, T. Garland Jr., Phylogenetic Logistic Regression for Binary Dependent Variables.
540 *Syst. Biol.* **59**, 9–26 (2010).
- 541 25. R. Adams, Z. Cain, R. Assis, M. DeGiorgio, Robust Phylogenetic Regression. *Syst. Biol.* **73**,
542 140–157 (2024).
- 543 26. J. D. Boyko, J. M. Beaulieu, Reducing the Biases in False Correlations Between Discrete
544 Characters. *Syst. Biol.* **72**, 476–488 (2023).
- 545 27. D. Chesters, Construction of a Species-Level Tree of Life for the Insects and Utility in
546 Taxonomic Profiling. *Syst. Biol.* **66**, 426–439 (2017).
- 547 28. R. Gadagkar, Why the Definition of Eusociality Is Not Helpful to Understand Its Evolution
548 and What Should We Do about It. *Oikos* **70**, 485–488 (1994).
- 549 29. P. W. Sherman, E. A. Lacey, H. K. Reeve, L. Keller, The eusociality continuum. *Behav. Ecol.*
550 **6**, 102–108 (1995).
- 551 30. B. J. Crespi, D. Yanega, The definition of eusociality. *Behav. Ecol.* **6**, 109–115 (1995).

- 552 31. R. M. Dew, S. M. Tierney, M. P. Schwarz, Social evolution and casteless societies: needs for
553 new terminology and a new evolutionary focus. *Insectes Sociaux* **63**, 5–14 (2016).
- 554 32. J. J. Boomsma, R. Gawne, Superorganismality and caste differentiation as points of no
555 return: how the major evolutionary transitions were lost in translation. *Biol. Rev.* **93**, 28–54
556 (2018).
- 557 33. C. D. Michener, *The social behavior of the bees: a comparative study* (Harvard University
558 Press, 1974).
- 559 34. S. Batra, Nests and Social Behavior of Halictine bees of India (Hymenoptera: Halictidae).
560 *Indian J. Entomol.* **28**, 375–393 (1966).
- 561 35. S. A. West, R. M. Fisher, A. Gardner, E. T. Kiers, Major evolutionary transitions in
562 individuality. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10112–10119 (2015).
- 563 36. J. D. Gardner, C. L. Organ, Evolutionary Sample Size and Consilience in Phylogenetic
564 Comparative Analysis. *Syst. Biol.* **70**, 1061–1075 (2021).
- 565 37. G. Shreeves, M. A. Cant, A. Bolton, J. Field, Insurance-based advantages for subordinate
566 co-foundresses in a temperate paper wasp. *Proc. R. Soc. Lond. B Biol. Sci.* **270**, 1617–1622
567 (2003).
- 568 38. J. Field, G. Shreeves, S. Sumner, M. Casiraghi, Insurance-based advantage to helpers in a
569 tropical hover wasp. *Nature* **404**, 869–871 (2000).
- 570 39. R. L. Bonifacii, J. Field, Extended parental care in the mass provisioning silk wasp,
571 *Microstigmus rosae*. *Behav. Ecol. Sociobiol.* **78**, 20 (2024).
- 572 40. C. K. Cornwallis, S. A. West, K. E. Davis, A. S. Griffin, Promiscuity and the evolutionary
573 transition to complex societies. *Nature* **466**, 969–972 (2010).
- 574 41. R. L. Trivers, H. Hare, Haplodiploidy and the Evolution of the Social Insect. *Science* **191**,
575 249–263 (1976).
- 576 42. R. Craig, Parental Manipulation, Kin Selection, and the Evolution of Altruism. *Evolution* **33**,
577 319–334 (1979).
- 578 43. J. Seger, Partial bivoltinism may cause alternating sex-ratio biases that favour eusociality.
579 *Nature* **301**, 59–62 (1983).
- 580 44. A. Grafen, Split sex ratios and the evolutionary origins of eusociality. *J. Theor. Biol.* **122**,
581 95–121 (1986).
- 582 45. L. Fromhage, H. Kokko, Monogamy and haplodiploidy act in synergy to promote the
583 evolution of eusociality. *Nat. Commun.* **2**, 397 (2011).
- 584 46. R. A. Johnstone, M. A. Cant, J. Field, Sex-biased dispersal, haplodiploidy and the evolution
585 of helping in social insects. *Proc. R. Soc. B Biol. Sci.* **279**, 787–793 (2011).

- 586 47. A. Gardner, J. Alpedrinha, S. A. West, Haplodiploidy and the Evolution of Eusociality: Split
587 Sex Ratios. *Am. Nat.* **179**, 240–256 (2012).
- 588 48. A. Gardner, L. Ross, Haplodiploidy, Sex-Ratio Adjustment, and Eusociality. *Am. Nat.* **181**,
589 E60–E67 (2013).
- 590 49. J. Alpedrinha, A. Gardner, S. A. West, Haplodiploidy and the Evolution of Eusociality:
591 Worker Revolution. *Am. Nat.* **184**, 303–317 (2014).
- 592 50. A. E. Quiñones, I. Pen, A unified model of Hymenopteran preadaptations that trigger the
593 evolutionary transition to eusociality. *Nat. Commun.* **8**, 15920 (2017).
- 594 51. P. Rautiala, H. Helanterä, M. Puurtinen, Extended haplodiploidy hypothesis. *Evol. Lett.* **3**,
595 263–270 (2019).
- 596 52. P. Kennedy, A. N. Radford, Sibling quality and the haplodiploidy hypothesis. *Biol. Lett.* **16**,
597 20190764 (2020).
- 598 53. E. Rees-Baylis, I. Pen, J. J. Kreider, Maternal manipulation of offspring size can trigger the
599 evolution of eusociality in promiscuous species. *Proc. Natl. Acad. Sci. U. S. A.* **121**,
600 e2402179121 (2024).
- 601 54. J. Alpedrinha, S. A. West, A. Gardner, Haplodiploidy and the Evolution of Eusociality:
602 Worker Reproduction. *Am. Nat.* **182**, 421–438 (2013).
- 603 55. L. Ross, A. Gardner, N. Hardy, S. A. West, Ecology, Not the Genetics of Sex Determination,
604 Determines Who Helps in Eusocial Populations. *Curr. Biol.* **23**, 2383–2387 (2013).
- 605 56. W. A. Foster, Experimental evidence for effective and altruistic colony defence against
606 natural predators by soldiers of the gall-forming aphid *Pemphigus spyrothecae*
607 (Hemiptera : Pemphigidae). *Behav. Ecol. Sociobiol.* **27**, 421–430 (1990).
- 608 57. J. Field, S. Brace, Pre-social benefits of extended parental care. *Nature* **428**, 650–652
609 (2004).
- 610 58. E. R. Lucas, J. Field, Assured fitness returns in a social wasp with no worker caste. *Proc. R.
611 Soc. B Biol. Sci.* **278**, 2991–2995 (2011).
- 612 59. F. L. W. Ratnieks, K. R. Foster, T. Wenseleers, Conflict resolution in insect societies. *Annu.
613 Rev. Entomol.* **51**, 581–608 (2006).
- 614 60. T. Wenseleers, F. L. W. Ratnieks, Comparative Analysis of Worker Reproduction and
615 Policing in Eusocial Hymenoptera Supports Relatedness Theory. *Am. Nat.* **168**, E163–E179
616 (2006).
- 617 61. J. Meunier, S. A. West, M. Chapuisat, Split sex ratios in the social Hymenoptera: a meta-
618 analysis. *Behav. Ecol.* **19**, 382–390 (2008).
- 619 62. S. A. West, *Sex Allocation* (Princeton Univ. Press, 2009).

- 620 63. H. Blackmon, L. Ross, D. Bachtrog, Sex Determination, Sex Chromosomes, and Karyotype
621 Evolution in Insects. *J. Hered.* **108**, 78–93 (2017).
- 622 64. L. Ross, A. J. Mongue, C. N. Hodson, T. Schwander, Asymmetric Inheritance: The Diversity
623 and Evolution of Non-Mendelian Reproductive Strategies. *Annu. Rev. Ecol. Evol. Syst.* **53**,
624 1–23 (2022).
- 625 65. L. R. Hearn, O. K. Davies, M. P. Schwarz, Extreme reproductive skew at the dawn of
626 sociality is consistent with inclusive fitness theory but problematic for routes to
627 eusociality. *Proc. R. Soc. B Biol. Sci.* **289**, 20220652 (2022).
- 628 66. J. da Silva, Life History and the Transitions to Eusociality in the Hymenoptera. *Front. Ecol.*
629 *Evol.* **9** (2021).
- 630 67. C. D. Michener, *The Bees of the World*. 2nd. Ed Johns Hopkins Baltim. (2007).
- 631 68. R. M. Bohart, A. S. Menke, *Sphecid Wasps of the World. A generic revision*. Berkeley (Univ
632 of California Press, 1976).
- 633 69. C. K. Starr, Ed., *Encyclopedia of Social Insects* (Springer International Publishing, 2021).
- 634 70. J. C. Choe, B. J. Crespi, *The Evolution of Social Behaviour in Insects and Arachnids*
635 (Cambridge University Press, 1997).
- 636 71. J. H. Hunt, *The Evolution of Social Wasps* (Oxford University Press, 2007).
- 637 72. K. G. Ross, R. W. Matthews, Eds., *The Social Biology of Wasps* (Cornell Univ. Press, 1991).
- 638 73. C. D. Michener, Comparative Social Behavior of Bees. *Annu. Rev. Entomol.* **14**, 299–342
639 (1969).
- 640 74. N. B. Davies, J. R. Krebs, S. A. West, *An Introduction to Behavioural Ecology* (John Wiley &
641 Sons, 2012).
- 642 75. M. J. McLeish, T. W. Chapman, M. P. Schwarz, Host-driven diversification of gall-inducing
643 Acacia thrips and the aridification of Australia. *BMC Biol.* **5**, 3 (2007).
- 644 76. T. Sless, S. Rehan, Phylogeny of the carpenter bees (Apidae: Xylocopinae) highlights
645 repeated evolution of sociality. *Biol. Lett.* **19**, 20230252 (2023).
- 646 77. S. A. Smith, B. C. O’Meara, treePL: divergence time estimation using penalized likelihood
647 for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
- 648 78. J. L. Rainford, M. Hofreiter, D. B. Nicholson, P. J. Mayhew, Phylogenetic Distribution of
649 Extant Richness Suggests Metamorphosis Is a Key Innovation Driving Diversification in
650 Insects. *PLOS ONE* **9**, e109085 (2014).
- 651 79. M. W. Pennell, *et al.*, geiger v2.0: an expanded suite of methods for fitting
652 macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**, 2216–2218 (2014).

- 653 80. Alan Grafen, phyreg: The Phylogenetic Regression of Grafen (1989).
654 <https://doi.org/10.32614/CRAN.package.phyreg>. Deposited 8 February 2014.
- 655 81. A. Grafen, M. Ridley, Statistical Tests for Discrete Cross-species Data. *J. Theor. Biol.* **183**,
656 255–267 (1996).
- 657

658 **Figures Legends**

659

660 **Figure 1.** Eusociality has independently evolved in multiple arthropod lineages: (a) ants; (b) bees;
661 (c) wasps; (d) thrips; (e) beetles (f) aphids; (g) termites. The top row (a-d) represents taxa with
662 haplodiploid sex determination systems, the bottom row (e-g) comprises taxa with diploid
663 systems.

664

665 Image credits: Ants: *MD Jerry*, via Unsplash, Bees: *Shelby Cohron*, via Unsplash. Wasps (*Vespa*
666 *crabro*): *Aconcagua*, CC BY-SA 3.0, via Wikimedia Commons. Thrips (*Kladothrips*): *W. W.*
667 *Froggatt*, 1907, public domain, via Biodiversity Heritage Library. Ambrosia beetles: *Peter H. W.*
668 *Biedermann*, used with permission. Aphids (*Pemphigus spirothecae*): *Beentree*, CC BY-SA 3.0,
669 via Wikimedia Commons. Termites (*Syntermes* sp.): *Philipp Hoenle*, CC0 1.0, via Wikimedia
670 Commons.

671

672 **Figure 2.** Cladogram of Hymenoptera with a non-Hymenopteran outgroup. Read & Nee (1995)
673 used this example to illustrate the potential problem of pseudoreplication within lineages. Red
674 branches indicate lineages in which eusociality has evolved, and black branches represent non-
675 eusocial lineages. This part of the tree involves a single origin of haplodiploidy in ancestral
676 Hymenoptera (blue arrow), after which eusociality has evolved multiple times (red arrows). These
677 transitions to eusociality could equally be explained by haplodiploidy or any other feature(s) of
678 Hymenoptera.

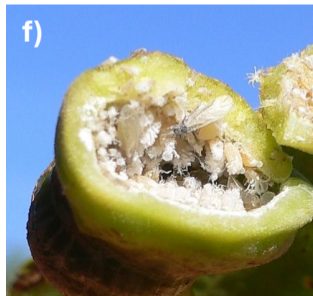
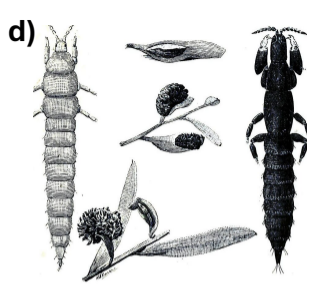
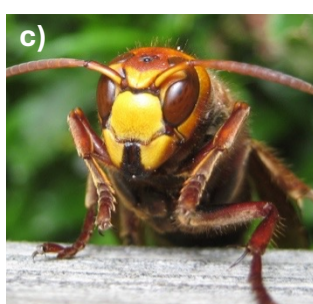
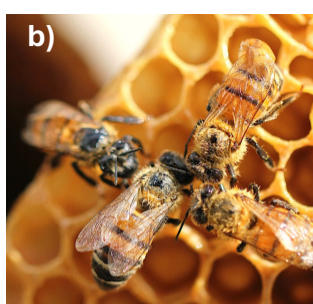
679

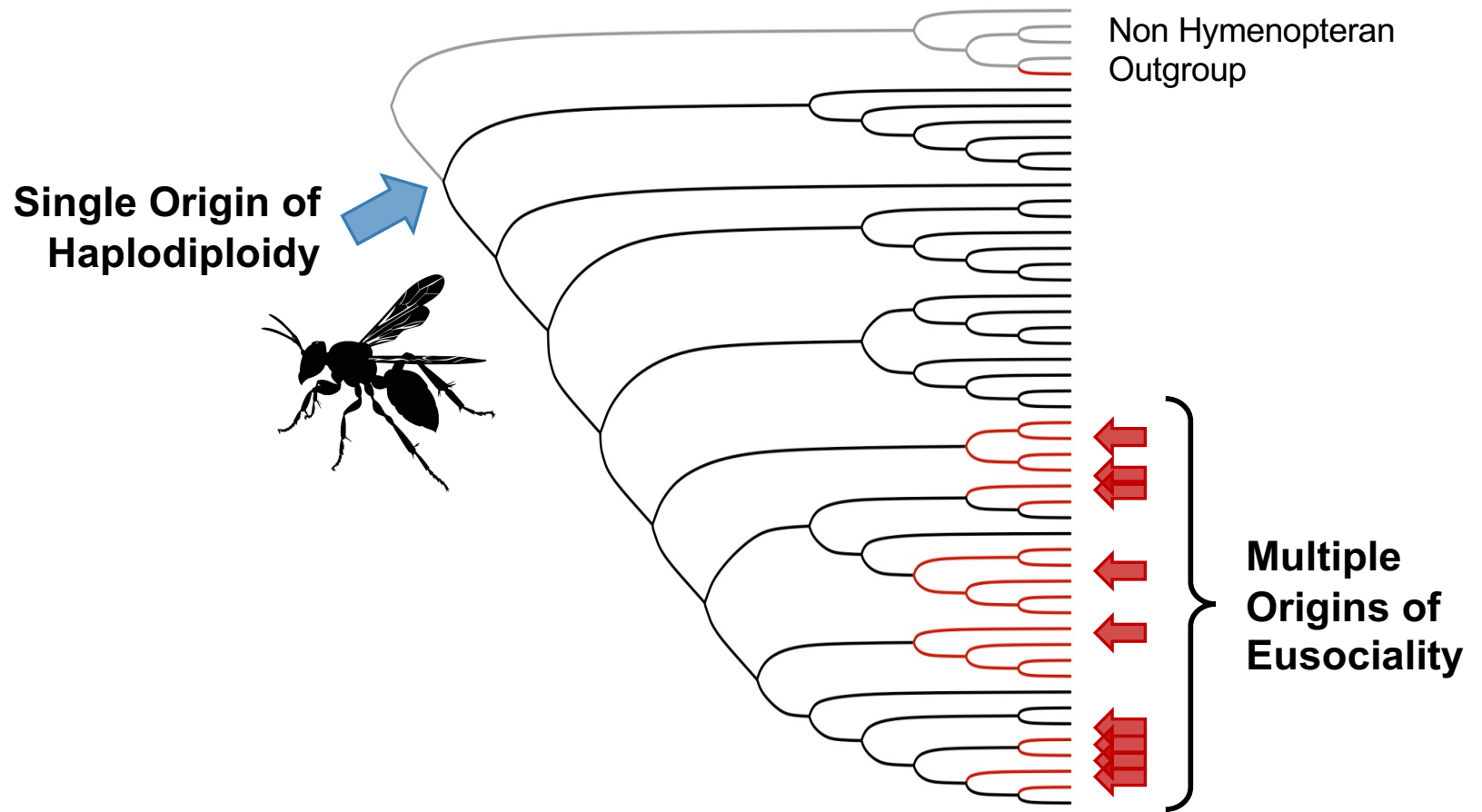
680 **Figure 3.** Schematic cladogram showing the evolutionary history of haplodiploidy and eusociality.
681 Haplodiploid lineages are indicated by dashed branches. Red circles denote lineages in which
682 eusociality has evolved (sensu Michener-Wilson), with numbers indicating the estimated number
683 of independent transitions represented by our sampled taxa. Asterisks(*) mark lineages that have
684 undergone transitions to superorganismality (sensu Boomsma-Gawne). These transitions are
685 estimated to have occurred once in termites, once in ambrosia beetles, and four times in the
686 bees, wasps and ants. Silhouettes from phylopic.org

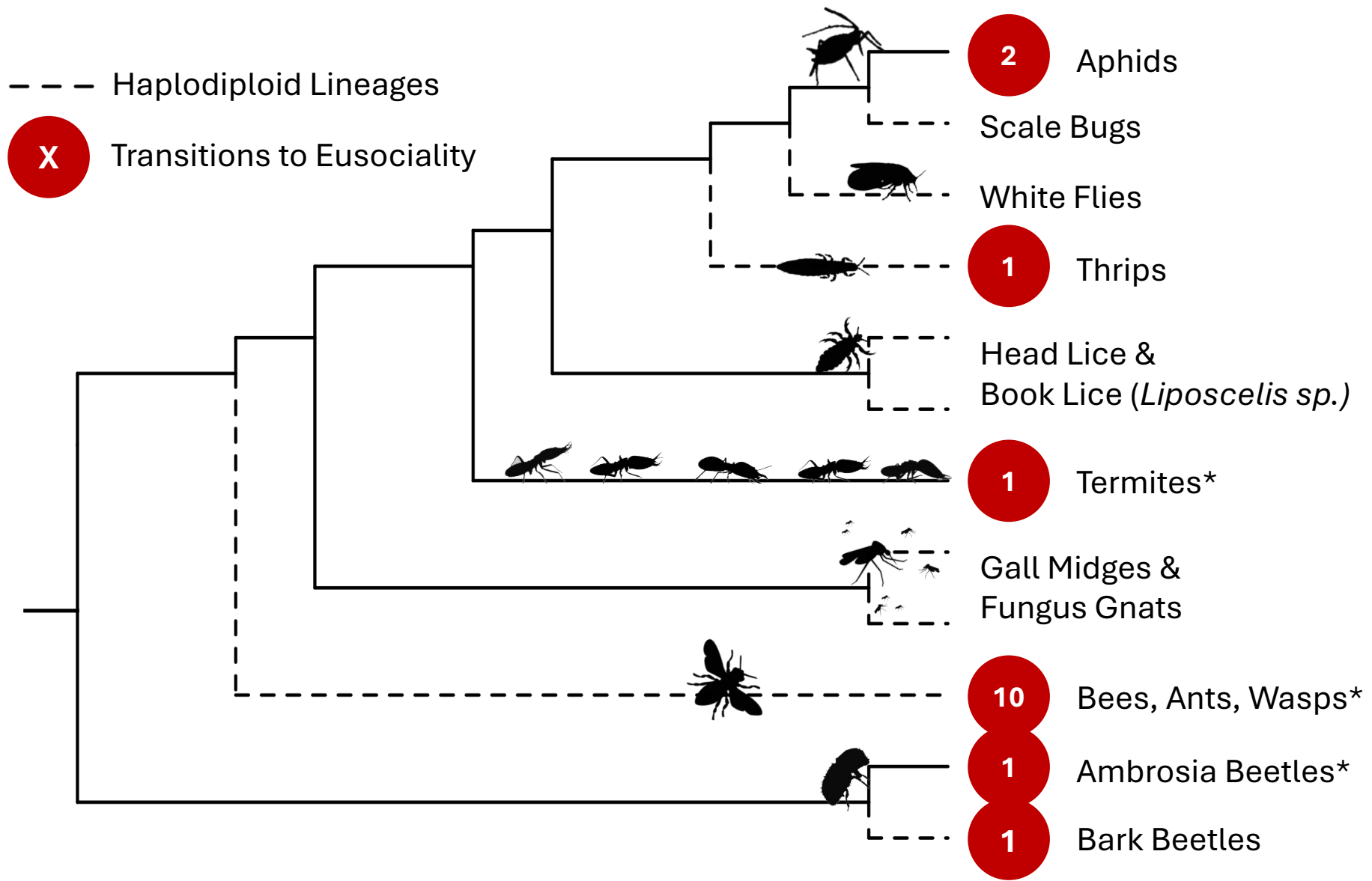
687 **Figure 4.** Cladogram showing the ten independent contrasts between haplodiploid and non-
688 haplodiploid clades, used for our sister-taxon comparisons. Focal and sister clades are indicated
689 by matched clade highlighting, with darker shaded clades representing the focal haplodiploid
690 clade (**F**) and lighter shaded clade representing the diploid contrast clades (**C**). Nodes
691 corresponding to the most recent common ancestors of each sister-taxon contrast are marked
692 with coloured circles; The four informative contrasts are indicated with an asterisk (*). Silhouettes
693 sourced from phylopic.org.

694 **Figure 5.** Difference in eusociality between focal (haplodiploid) and contrast (diploid) clades.
695 Each point shows the difference in the proportion of eusocial species for one contrast; positive
696 values indicate a higher proportion in the haplodiploid clade. Error bars represent 95% confidence
697 intervals for proportions based on clade size. Colours indicate which clade had the higher
698 proportion: green = more in haplodiploid, orange = more in diploid, and grey for no difference.

699

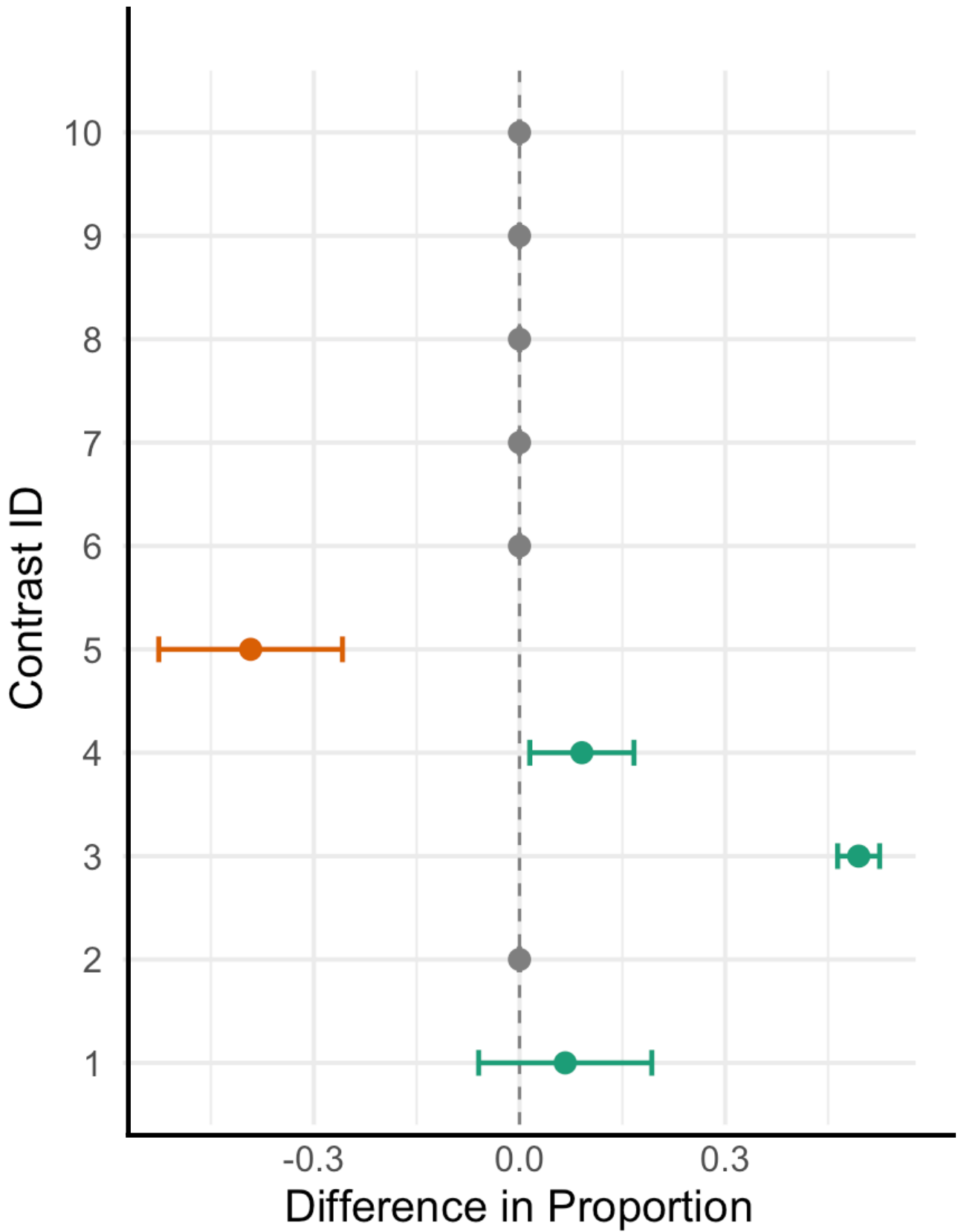






More in diploid

More in haplodiploid



Supporting Information for No Evidence That Haplodiploidy Favours the Evolution of Eusociality.

Rosa Bonifacii¹, Louis Bell-Roberts², Alan Grafen¹ and Stuart West¹.

Rosa Bonifacii
Email: rosa.bonifacii@biology.ox.ac.uk

This PDF file includes:

Supporting text
Figure S1 & S2
Tables S1 to S8

Other supporting materials for this manuscript include the following:

Dataset S1

22 **Supporting Information Text**

23 **Socially Variable / Uncertain Clades**

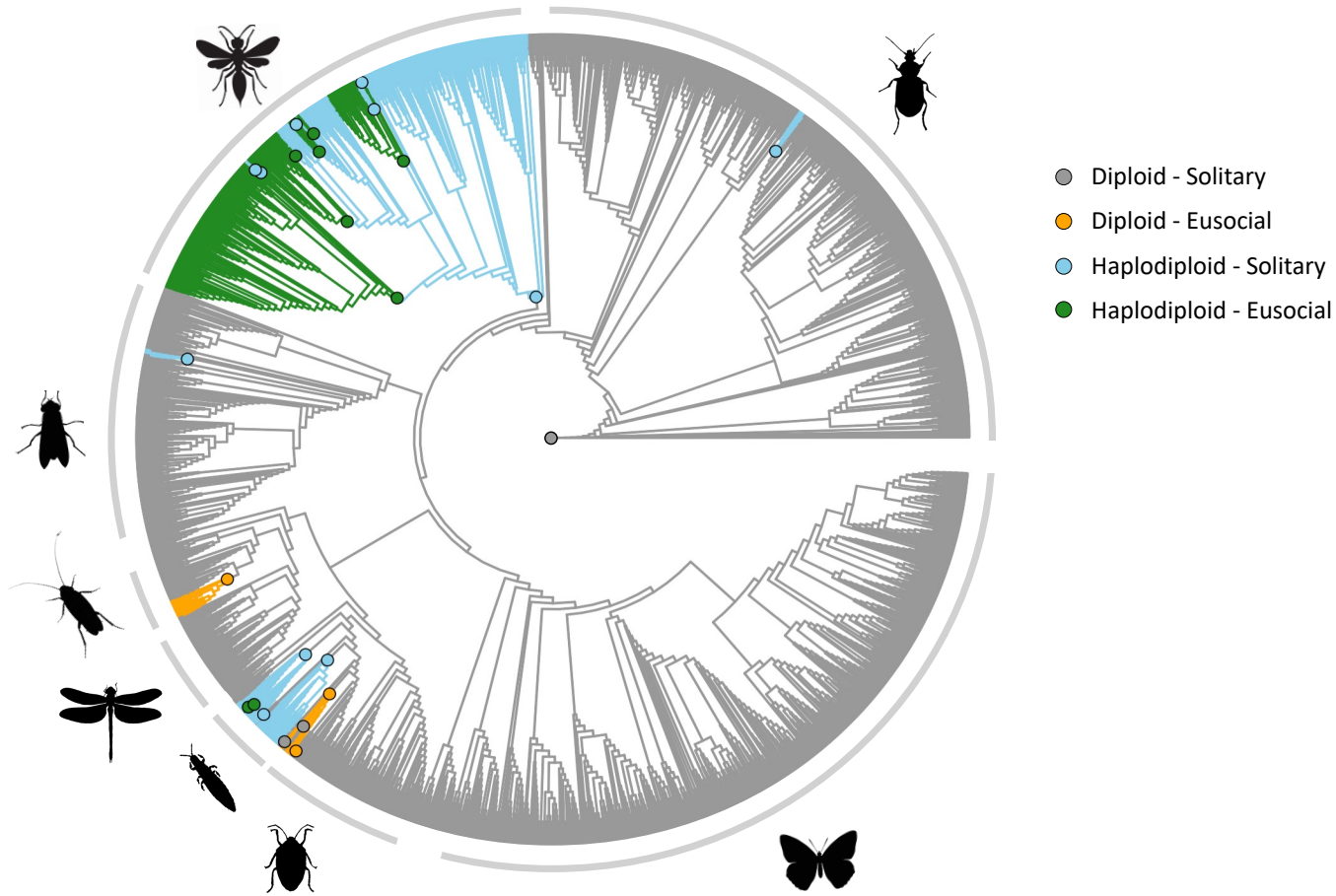
24 The clades listed below show social variation or uncertain eusociality status, and were therefore
25 used as the groups within which species were randomly reclassified for our error-robustness
26 analyses.

27 Michener-Wilson: *allodapini*, *xylocopini*, *ceratinini*, *euglossini*, *stenogastrinae*, *augochlorini*,
28 *halictini*, *kladothrips*, *hormaphidinae*, *eriosomatinae*, *xyleborini*, *platypodini*.

29
30 Boomsma- Gawne: *ponerinae*, *apinae*, *platypodinae*, *myrmeciinae*.

31
32 Crespi - Yanega: *allodapini*, *halictinae*, *kladothrips*, *epiponini*, *polistini*, *ropalidiini*, *hormaphidinae*,
33 *eriosomatinae*, *platypodinae*, *encyrtinae*.

34



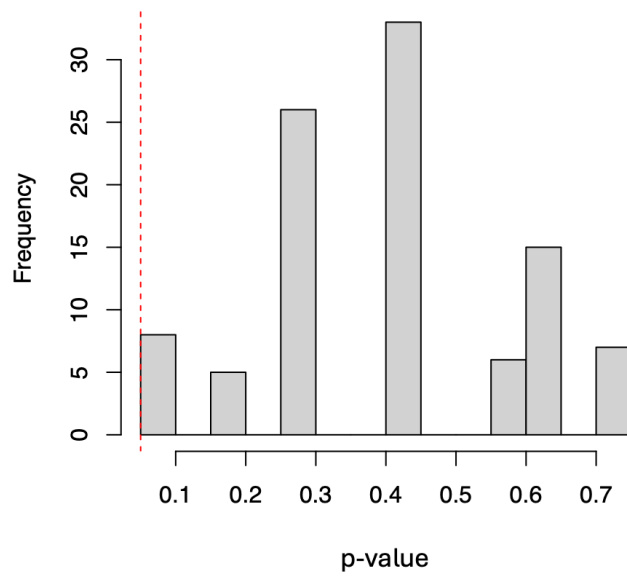
36

37 **Fig. S1.** Ancestral state reconstruction of haplodiploidy (inclusive definition) and eusociality
38 (Michener-Wilson definition) across the insect phylogeny. The tree shows inferred ancestral
39 states at each branch based on maximum likelihood estimation under an ARD (all-rates-different)
40 model. Branches are coloured according to the most likely state. Coloured circles indicate internal
41 nodes where state transitions are inferred to have occurred. Positions of major insect orders are
42 indicated. Silhouettes sourced from phylopic.org

43

44

P-values across replicate contrast sets



45

46 **Fig. S2.** Read & Nee's (1995) method of sister-taxon comparison did not support the hypothesis
47 that haplodiploidy has favoured a higher rate of eusocial evolution. Of 100 replicate datasets
48 using alternative plausible sister clades, none produced a significant result, $p < 0.05$. Results
49 presented use the Michener-Wilson definition of eusociality and the inclusive definition of
50 haplodiploidy.

51

52

53 **Table S1.** Numbers of inferred origins of eusociality under three alternative trait definitions,
 54 partitioned by (a) Hymenoptera vs. non-Hymenoptera and (b) haplodiploid vs. diploid lineages.
 55 These two partitions illustrate the problem of pseudoreplication: most eusocial origins in
 56 haplodiploid lineages occur within Hymenoptera, which itself represents only a single origin of
 57 haplodiploidy.
 58

Eusocial Definition	Number of Inferred Origins of Eusociality				Total
	Hymenoptera	Non-Hymenoptera	Haplodiploid	Diploid	
Michener-Wilson	9	7	12	4	16
Crespi-Yanega	12	7	15	4	19
Boomsma-Gawne	4	2	4	2	6

59
 60 Nb. These transition counts reflect estimates inferred from the taxa sampled in our dataset. While
 61 they provide valuable insight into evolutionary patterns, they should be interpreted as indicative
 62 rather than exact measures of the “true” number of transitions.
 63

64 **Table S2.** Results of Sister-Taxon Contrasts Analyses Using Alternative Trait Definitions. When
 65 using Crespi & Yanega's (1995) definition of eusociality, the number of informative contrasts
 66 drops to just three, two with greater eusociality in the haplodiploid clade, and one where
 67 eusociality is greater in the diploid clade. When examining superorganismality, we are left with
 68 just a single informative contrast.

Eusocial definition	Haplodiploid definition	Number contrasts	Number informative	Number successes	Binomial result	Wilcoxon result	Binomial # reps $p < 0.05$	Wilcoxon #reps $p < 0.05$
Michener-Wilson	inclusive	10	4	3	$p = 0.32$	$Z = 0.93,$ $p = 0.24$	0	0
Crespi-Yanega	inclusive	10	3	2	$p = 0.5$	$Z = 0.57,$ $p = 0.37$	0	0
Boomsma-Gawne	inclusive	10	1	1	$p = 0.5$	$Z = 1,$ $p = 0.5$	0	0
Michener-Wilson	strict	5	3	3	$p = 0.13$	$Z = 1.70,$ $p = 0.13$	0	0
Crespi-Yanega	strict	5	2	2	$p = 0.25$	$Z = 1.41,$ $p = 0.25$	0	0
Boomsma-Gawne	strict	5	1	1	$p = 0.5$	$Z = 1,$ $p = 0.49$	0	0

69

70 **Table S3.** Results of Phylogenetic Regression Analyses Examining the Association Between
 71 Haplodiploidy and Eusociality Under Alternative Trait Definitions. Results are shown for both the a
 72 time-calibrated phylogeny and Grafen's ρ -scaled phylogeny.

Eusocial definition	Haplodiploid definition	PhyregB results		logistic_MPLE		ROBRT	
		Time-scaled tree	Grafen's ρ scaled tree	Time-scaled tree	Grafen's ρ scaled tree	Time-scaled tree	Grafen's ρ scaled tree
Michener-Wilson	inclusive	NA	F(1,147) = 0.005, p = 0.94	z = 0.13, p = 0.90	z = 0.16, p = 0.87	z = -0.07, p = 0.94	z = 0.03, p = 0.98
Crespi - Yanega	inclusive	NA	F(1,202) = 0.006, p = 0.94	z = 0.21, p = 0.83	z = 0.45, p = 0.65	z = -0.06, p = 0.95	z = -0.02, p = 0.98
Boomsma-Gawne	inclusive	NA	F(1,124) = 0.001, p = 0.97	z = 0.001, p = 0.99	z = -0.06, p = 0.95	z = -0.08, p = 0.93	z = -0.02, p = 0.99
Michener-Wilson	strict	NA	F(1,147) = 0.016, p = 0.90	z = 0.12, p = 0.91	z = 0.14, p = 0.89	z = 1.47, p = 0.14	z = 1.04, p = 0.30
Crespi - Yanega	strict	NA	F(1,202) = 0.001, p = 0.97	z = 0.20, p = 0.84	z = 0.40, p = 0.69	z = -0.08, p = 0.94	z = -0.04, p = 0.97
Boomsma-Gawne	strict	NA	F(1,124) = 0.002, p = 0.97	z = 0.00, p = 0.99	z = 0.04, p = 0.97	z = -0.01, p = 0.91	z = -0.02, p = 0.98

73

74 **Table S4 (a).** Results of Hidden Markov Model analyses testing the association between
 75 haplodiploidy and eusociality under alternative trait definitions, using a **time-calibrated**
 76 **phylogeny**. Columns show the Hidden Markov Model variant (*Model*), the number of rate
 77 categories specified (*RateCat*), the log-likelihood of the fitted model (*logLik*), the Akaike
 78 Information Criterion (*AIC*), the difference in AIC relative to the best-fitting model (*deltaAIC*), and
 79 the relative model support expressed as Akaike weight (*Akaike weight*).

80

Eusocial definition	Haplodiploid definition	Model	RateCat	logLik	AIC	deltaAIC	Akaike weight
Michener-Wilson	inclusive	Dependent	1	-234.97	485.95	51.12	0
Michener-Wilson	inclusive	Independent	1	-250.57	509.13	74.30	0
Michener-Wilson	inclusive	HMIM Independent	2	-208.42	434.83	0	0.803
Michener-Wilson	inclusive	HMM Dependent	2	-201.82	437.64	2.81	0.197
Crespi - Yanega	inclusive	Dependent	1	-266.02	548.04	42.78	0
Crespi - Yanega	inclusive	Independent	1	-283.78	575.57	70.31	0
Crespi - Yanega	inclusive	HMIM Independent	2	-243.63	505.26	0	0.863
Crespi - Yanega	inclusive	HMM Dependent	2	-237.47	508.95	3.68	0.137
Boomsma-Gawne	inclusive	Dependent	1	-165.86	347.72	3.05	0.135
Boomsma-Gawne	inclusive	Independent	1	-169.29	346.57	1.90	0.241
Boomsma-Gawne	inclusive	HMIM Independent	2	-163.34	344.68	0	0.622
Boomsma-Gawne	inclusive	HMM Dependent	2	-161.01	356.02	11.35	0.002
Michener-Wilson	strict	Dependent	1	-205.11	426.22	44.25	0
Michener-Wilson	strict	Independent	1	-221.87	451.74	69.77	0
Michener-Wilson	strict	HMIM Independent	2	-181.98	381.96	0	0.527
Michener-Wilson	strict	HMM Dependent	2	-174.09	382.18	0.22	0.473
Crespi - Yanega	strict	Dependent	1	-236.13	488.26	39.41	0
Crespi - Yanega	strict	Independent	1	-255.08	518.17	69.32	0
Crespi - Yanega	strict	HMIM Independent	2	-220.89	459.77	10.92	0.004
Crespi - Yanega	strict	HMM Dependent	2	-207.43	448.85	0	0.996
Boomsma-Gawne	strict	Dependent	1	-136.90	289.80	0.85	0.256
Boomsma-Gawne	strict	Independent	1	-140.59	289.18	0.23	0.350
Boomsma-Gawne	strict	HMIM Independent	2	-135.47	288.95	0	0.393
Boomsma-Gawne	strict	HMM Dependent	2	-133.15	304.50	14.9	0.000

81 **Table S4 (b).** Results of Hidden Markov Model analyses testing the association between
 82 haplodiploidy and eusociality under alternative trait definitions, using a **Grafen's p-scaled**
 83 **phylogeny**. Columns show the Hidden Markov Model variant (*Model*), the number of rate
 84 categories specified (*RateCat*), the log-likelihood of the fitted model (*logLik*), the Akaike
 85 Information Criterion (*AIC*), the difference in AIC relative to the best-fitting model (*deltaAIC*), and
 86 the relative model support expressed as Akaike weight (*Akaike weight*).

87

Eusocial def.	Haplodiploid def.	Model	RateCat	logLik	AIC	deltaAIC	Akaike weight
Michener-Wilson	inclusive	Dependent	1	-240.22	496.44	56.89	0.00
Michener-Wilson	inclusive	Independent	1	-252.09	512.19	72.63	0.00
Michener-Wilson	inclusive	HMIM	2	-210.78	439.55	0.00	1.00
Michener-Wilson	inclusive	HMM Dependent	2	-208.62	451.24	11.68	0.00
Crespi - Yanega	inclusive	Dependent	1	-304.96	625.93	59.01	0.00
Crespi - Yanega	inclusive	Independent	1	-325.43	658.86	91.95	0.00
Crespi - Yanega	inclusive	HMIM Independent	2	-274.46	566.92	0.00	0.89
Crespi - Yanega	inclusive	HMM Dependent	2	-268.58	571.16	4.24	0.11
Boomsma-Gawne	inclusive	Dependent	1	-179.58	375.16	17.19	0.00
Boomsma-Gawne	inclusive	Independent	1	-183.62	375.23	17.26	0.00
Boomsma-Gawne	inclusive	HMIM Independent	2	-169.99	357.98	0.00	0.84
Boomsma-Gawne	inclusive	HMM Dependent	2	-163.65	361.30	3.33	0.16
Michener-Wilson	strict	Dependent	1	-197.99	411.98	36.10	0.00
Michener-Wilson	strict	Independent	1	-210.51	429.02	53.14	0.00
Michener-Wilson	strict	HMIM Independent	2	-178.94	375.88	0.00	0.73
Michener-Wilson	strict	HMM Dependent	2	-171.95	377.89	2.02	0.27
Crespi - Yanega	strict	Dependent	1	-262.24	540.48	42.47	0.00
Crespi - Yanega	strict	Independent	1	-283.85	575.70	77.69	0.00
Crespi - Yanega	strict	HMIM Independent	2	-242.70	503.39	5.39	0.06
Crespi - Yanega	strict	HMM Dependent	2	-232.00	498.01	0.00	0.94
Boomsma-Gawne	strict	Dependent	1	-137.81	291.62	2.04	0.20
Boomsma-Gawne	strict	Independent	1	-142.03	292.07	2.48	0.16

Boomsma-Gawne	strict	HMIM Independent	2	-135.79	289.59	0.00	0.55
Boomsma-Gawne	strict	HMIM Dependent	2	-129.57	293.13	3.54	0.09

88

89

90 **Table S5.** Summary statistics from “phyregB” phylogenetic regression analyses across 25
 91 replicate datasets, used to assess the robustness of results to behavioural misclassification. In
 92 each replicate, 10% of eusociality states were randomly reassigned within socially variable
 93 clades. Columns show the number of replicates (*n_reps*), mean and standard deviation of the
 94 estimated haplodiploidy effect (*mean_beta*, *sd_beta*), minimum and maximum p-values (*min_p*,
 95 *max_p*) and the proportion of significant results (*prop_sig_0.05*).

Eusocial definition	Haplodiploid definition	n_reps	mean_beta	sd_beta	min_p	max_p	prop_sig_0.05
Michener-Wilson	inclusive	25	0.019	0.015	0.508	0.984	0
Crespi - Yanega	inclusive	25	-0.005	0.002	0.877	0.989	0
Boomsma-Gawne	inclusive	25	0.002	0.001	0.917	0.965	0
Michener-Wilson	strict	25	0.063	0.036	0.430	0.905	0
Crespi - Yanega	strict	25	0.006	0.002	0.933	0.992	0
Boomsma-Gawne	strict	25	0.005	0.001	0.910	0.957	0

96

97

98 **Table S6.** Summary statistics from phylogenetic logistic regression analyses across 25 replicate
 99 datasets used to assess robustness of results to behavioural misclassification. In each replicate,
 100 10% of eusociality states were randomly reassigned within socially variable clades. Columns
 101 show the number of replicates (*n_reps*), mean and standard deviation of the estimated
 102 haplodiploidy effect (*mean_beta*, *sd_beta*), minimum and maximum p-values (*min_p*, *max_p*) and
 103 the proportion of significant results (*prop_sig_0.05*).

Eusocial definition	Haplodiploid definition	n_reps	mean_beta	sd_beta	min_p	max_p	prop_sig_0.05
Michener-Wilson	inclusive	25	0.129	0.069	0.557	0.968	0
Crespi - Yanega	inclusive	25	0.250	0.064	0.550	0.829	0
Boomsma-Gawne	inclusive	25	0.050	0.068	0.911	0.995	0
Michener-Wilson	strict	25	0.142	0.083	0.599	0.995	0
Crespi - Yanega	strict	25	0.280	0.069	0.580	0.833	0
Boomsma-Gawne	strict	25	0.034	0.057	0.917	0.997	0

104

105

106 **Table S7.** Summary statistics from ROBRT phylogenetic regression analyses across 25 replicate
 107 datasets used to assess the robustness to behavioural misclassification. In each replicate, 10% of
 108 eusociality states were randomly reassigned within socially variable clades. Columns show the
 109 number of replicates (*n_reps*), mean and standard deviation of the estimated haplodiploidy effect
 110 (*mean_beta*, *sd_beta*), minimum and maximum p-values (*min_p*, *max_p*) and the proportion of
 111 significant results (*prop_sig_0.05*).

Eusocial definition	Haplodiploid definition	n_reps	mean_beta	sd_beta	min_p	max_p	prop_sig_0.05
Michener-Wilson	inclusive	25	-0.056	0.002	0.73	0.80	0
Crespi - Yanega	inclusive	25	-0.050	0	0.83	0.86	0
Boomsma-Gawne	inclusive	25	-0.056	0.002	0.73	0.80	0
Michener-Wilson	strict	25	-0.054	0.002	0.74	0.81	0
Crespi - Yanega	strict	25	-0.050	0	0.83	0.86	0
Boomsma-Gawne	strict	25	-0.043	0.002	0.74	0.80	0

112

113

114 **Table S8.** Summary statistics from the hidden-Markov model (HMM) analyses across the 25
 115 “noisy” datasets used to assess robustness to behavioural-state misclassification. In each
 116 replicate dataset, 10% of eusociality states were randomly reassigned within socially variable
 117 clades to mimic potential observation error. For each *RateCat* model, we report the mean Akaike
 118 Information Criterion (*mean_AIC*), the mean Akaike weight (*mean_weight*), and the number of
 119 replicates in which the model achieved the lowest AIC (No “wins”).

Eusocial definition	Haplodiploid definition	Model	RateCat	mean_AIC	mean_weight	No “wins”
Michener-Wilson	inclusive	Dependent	1	571.405	0.000	0
Michener-Wilson	inclusive	Independent	1	596.386	0.000	0
Michener-Wilson	inclusive	HMIM Independent	2	496.012	0.990	25
Michener-Wilson	inclusive	HMM Dependent	2	505.705	0.010	0
Crespi - Yanega	inclusive	Dependent	1	674.689	0.000	0
Crespi - Yanega	inclusive	Independent	1	719.700	0.000	0
Crespi - Yanega	inclusive	HMIM Independent	2	606.087	0.613	14
Crespi - Yanega	inclusive	HMM Dependent	2	607.933	0.387	11
Boomsma-Gawne	inclusive	Dependent	1	448.752	0.000	0
Boomsma-Gawne	inclusive	Independent	1	450.613	0.000	0
Boomsma-Gawne	inclusive	HMIM Independent	2	430.985	0.941	24
Boomsma-Gawne	inclusive	HMM Dependent	2	441.748	0.058	1
Michener-Wilson	strict	Dependent	1	486.543	0.000	0
Michener-Wilson	strict	Independent	1	513.219	0.000	0
Michener-Wilson	strict	HMIM Independent	2	431.960	0.672	17
Michener-Wilson	strict	HMM Dependent	2	434.505	0.328	8
Crespi - Yanega	strict	Dependent	1	588.585	0.000	0
Crespi - Yanega	strict	Independent	1	636.533	0.000	0
Crespi - Yanega	strict	HMIM Independent	2	538.870	0.465	14
Crespi - Yanega	strict	HMM Dependent	2	539.730	0.535	11
Boomsma-Gawne	strict	Dependent	1	365.164	0.037	1

Boomsma-Gawne	strict	Independent	1	367.446	0.009	0
Boomsma-Gawne	strict	HMIM Independent	2	355.734	0.786	22
Boomsma-Gawne	strict	HMM Dependent	2	360.264	0.168	2

120

121

122 **Dataset S1** (separate file). Trait data used in all comparative analyses. This dataset contains the
123 full set of trait data used across all analyses in the study. Each row represents a species included
124 in the modified phylogeny. Columns are: *species_name*: Taxon name matching tip labels in the
125 phylogenetic tree; *haplodiploidy_inclusive*: Broad definition of haplodiploidy, coded as 0 (absent)
126 or 1 (present); *haplodiploidy_strict*: Strict definition of haplodiploidy, coded as 0 or 1;
127 *haplodiploidy_ref*: Primary literature source(s) supporting the haplodiploidy classification for each
128 species. *eusocial_mw*: Eusociality classification based on Michener Wilson criteria, coded as 0 or
129 1; *eusocial_cy*: Classification based on Crespi & Yanega criteria, coded as 0 or 1;
130 *superorganism_bg*: Binary coding for superorganism-level social structure (per Boomsma &
131 Gawne), coded as 0 or 1; *eusocial_ref*: Primary literature source(s) used to assign
132 eusociality/superorganism classifications. Missing data are indicated as NA.
133