Evolutionary genomics of socially polymorphic populations of *Pogonomyrmex californicus*

Mohammed Errbii[1], Ulrich R. Ernst[1,§,£], Aparna Lajmi[2], Eyal Privman[2], Jürgen Gadau[1,‡], Lukas Schrader[1,‡]

[1]: Molecular Evolution and Sociobiology Group, Institute for Evolution & Biodiversity, University of Münster, Hüfferstr. 1, DE-48149 Münster, Germany

[2]: Department of Evolutionary and Environmental Biology, Institute of Evolution, University of Haifa, Haifa, Israel

[‡]: equal contribution

[§]: current address: Apicultural State Institute, University of Hohenheim, Erna-Hruschka-Weg 6, DE-70599 Stuttgart, Germany

[£]: Center for Biodiversity and Integrative Taxonomy (KomBioTa), University of Hohenheim, DE-70599 Stuttgart, Germany

*Corresponding authors: Lukas Schrader, Jürgen Gadau
Email: lukas.schrader@uni-muenster.de; gadauj@uni-muenster.de

Keywords: haplometrosis, pleometrosis, colony founding, supergene, social niche polymorphism, polygenic trait, *Pogonomyrmex*

Abstract

Social insects vary considerably in their social organization both between and within species. In the California harvester ant, *Pogonomyrmex californicus* (Buckley 1867), colonies are commonly founded and headed by a single queen (haplometrosis, primary monogyny). However, in some populations in California (USA), unrelated queens cooperate not only during founding (pleometrosis) but throughout the life of the colony (primary polygyny). The genetic architecture and evolutionary dynamics of this complex social niche polymorphism (haplometrosis vs pleometrosis) have remained unknown. Here, we provide a first analysis of its genomic basis and evolutionary history using population genomics comparing individuals from a haplometrotic population to those from a pleometrotic population. We discovered a recently evolved (< 200 k years), 8 Mb non-recombining region segregating with the observed social niche polymorphism. This region shares several characteristics with supergenes underlying social polymorphisms in other socially polymorphic ant species. However, we also find remarkable differences from previously described social supergenes. Particularly, four additional genomic regions not in linkage with the supergene show signatures of a selective sweep in the pleometrotic population. Within these regions, we find for example genes crucial for epigenetic regulation via histone modification (*chameau*) and DNA methylation (*Dnmt1*). These results suggest that social morph in this species is a polygenic trait involving a potential young supergene. Further studies targeting haplo- and pleometrotic individuals from a single population are however required to conclusively resolve whether these genetic differences underlie the alternative social phenotypes or have emerged through genetic drift.

Introduction

Colonies of social Hymenoptera are led by a single (monogyny) or multiple reproductive females (polygyny) [1–3]. Such variations of social structure are suggested to have an adaptive value in certain conditions [2,4], and are conceptualized as alternative “social niches”, i.e. “the set of social environments in which the focal individual has non-zero inclusive fitness” [5]. In ants, species are often considered to be either monogyne or polygyne, but more and more cases of intraspecific variability in social organization have been described within several ant genera. Examples include multiple *Formica* and *Myrmica* species [6], various *Solenopsis* species [7–9], *Messor pergandei* [10], *Pogonomyrmex californicus* [11], *Cataglyphis niger* [12] *Leptothorax acevorum* [13] and *L. longispinosus* [14].

Species showing intraspecific variation in social organization are ideal to study the genomic architecture and evolutionary dynamics underlying complex trait polymorphisms. Recent population genomic studies in *Solenopsis* fire ants [9,15] and *Formica* wood ants [16–18] have provided significant insights into the origin, evolution, and geographic distribution of social variants in these species. Importantly, these studies led to the discovery of convergently evolved supergenes, i.e. genomic regions containing clusters of linked loci [19], underlying social niche polymorphisms in these species. Suppression of recombination in these supergenes drives the evolution of two or more diverged haplotype groups, similar to the X and Y haplotypes of sex chromosomes. Such genomic architecture facilitates the evolution and maintenance of complex trait polymorphisms (e.g. Brelsford et al., 2020; Hohenlohe et al., 2010; Joron et al., 2006; Matschiner et al., 2021; Nadeau et al., 2012). However social polymorphism can involve mechanisms other than supergenes as shown recently for the big-headed ant *Pheidole pallidula*, where genetic and environmental factors are suspected to contribute to the expression of alternative social morphs [24].

The California harvester ant *Pogonomyrmex californicus* [25] occurs throughout the Southwestern USA and Northwestern Mexico [26] and varies in terms of monogyny and polygyny, with primary monogyny as the predominant social organization [27]. In contrast to *Solenopsis* and *Formica*

species, this variation is based on differences in colony founding strategy, i.e. haplometrosis (queens found a colony alone, “primary monogyny”) vs pleometrosis (multiple queens found a colony together, “primary polygyny”) [11,27–29], and not re-adoption of queens or colony budding [30,31]. Populations of *P. californicus* exhibit geographical variation in colony-founding strategies and are almost fixed for either the haplo- or the pleometrotic strategy [11,32]. Common garden experiments with foundresses belonging to these populations revealed that haplo- and pleometrotic behaviors are expressed independent of environmental conditions, consistent with a genetic polymorphism underlying these alternative social strategies [28,33], and representing adaptations to different ecological and geographic niches in this species.

In this study, we analyzed the genomic architecture, population dynamics, and evolutionary history of the social niche polymorphism in *P. californicus*, by comparing individuals from two populations in Southern California that are virtually fixed for either haplometrosis (H-population) or pleometrosis (P-population) [28,32]. We identify a non-recombining region of ca. 8 Mb that segregates with the studied populations, resembling previously described supergenes driving social polymorphisms in ants (Wang et al. 2013, Purcell et al. 2014). Additionally, we also find signatures of selective sweeps in the pleometrotic population outside the non-recombining region, suggesting that the social niche polymorphism in *P. californicus* is a polygenic trait involving not only a potential young supergene, but also additional unlinked modifier loci.

100 **Results**

101 *Pogonomyrmex californicus genome assembly*

102 To improve the available fragmented genome assembly, we combined minION long-read
103 sequencing and 10X sequencing [34] of pleometrotic individuals to produce a highly contiguous
104 reference genome assembly for *P. californicus*. The assembly spans 252.3 Mb across 199
105 scaffolds (N50 = 10.4 Mb, largest scaffold 20.75 Mb). We recovered 98.5% complete BUSCOs
106 (S:97.8%,D:0.7%,F:1.1%,M:0.4%,n:4415), suggesting a nearly complete genome assembly.
107 Genome annotation identified 22.79% repetitive sequences and 15,899 protein coding genes,
108 resembling repeat and gene contents reported for other published ant genomes [35] (Dataset S1).

109 *The haplometrotic and pleometrotic populations are genetically distinct*

110 For our population genomic analyses, we performed whole genome sequencing of 35 founding
111 queens (19 from the P-population and 16 from the H-population). Principal Component Analyses
112 (PCA) using 314,756 SNPs clearly separated queens of the two populations (Figures 1A and S1).
113 Further, the PCA suggested higher genetic diversity in the H-population compared to the P-
114 population. One founding queen (P99P) collected from the P-population was positioned between
115 both populations in the PCA (Figure 1A), likely representing an F1 hybrid between a haplo- and a
116 pleometrotic individual.

117 A detailed analysis of coancestry showed that queens share more haplotypes in within-population
118 pairs than in between-population pairs (Figure 1B). Consistent with ongoing admixture between the
119 two populations, we found significant between-population haplotype-sharing in some individuals
120 (e.g. P99P, H33G and H46G) corroborated by ADMIXTURE analyses (Figures 1B and S2).

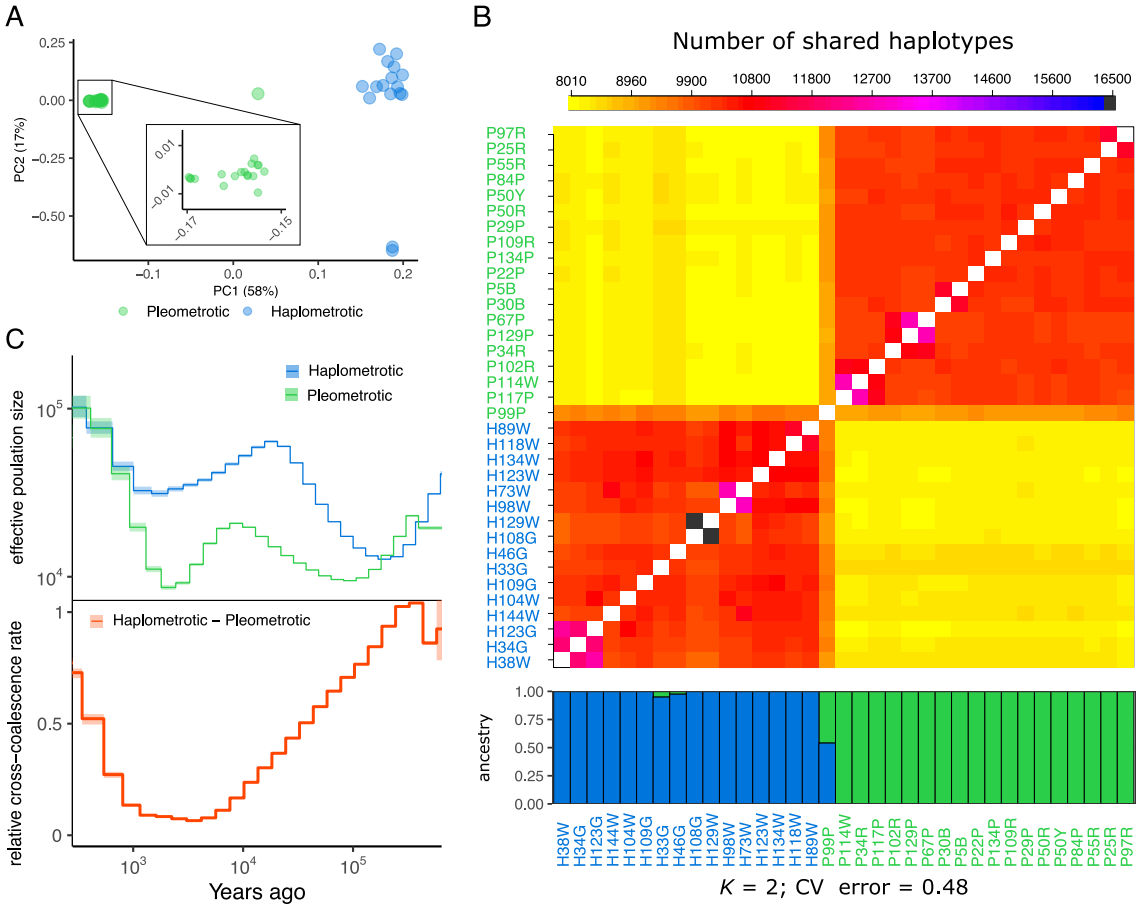

Figure 1. Population structure and demographic history in *P. californicus* populations. (A) Principal Component Analysis based on genome-wide SNP data of queens of the pleometrotic and haplometrotic populations (a scatterplot matrix showing four principal components can be found in Figure S1). (B) fineSTRUCTURE coancestry heatmap showing the number of shared haplotypes between donor (columns) and recipient (rows) queen pairs. ADMIXTURE plot below the heatmap shows patterns of ancestry among populations of *P. californicus* at k=2 (additional k values can be found in Figure S2). Note that one queen (P99P) is an apparent hybrid between the two populations. (C) Demographic history in *P. californicus* computed using MSMC2 with eight phased haplotypes (i.e. four queens), representing the pleometrotic and haplometrotic populations. Effective population size (N_e) is displayed in the top panel and the relative cross-coalescence rate (rCCR) in the lower panel. Lines and shaded areas are means and 95% confidence intervals, respectively.

135 *Gene flow between the H- and P-populations is a recent phenomenon*

136 We used MSMC2 [36] to reconstruct the demographic history of the P- and H-populations (Figure
137 1C). Both populations showed a similar trajectory with a zigzag-like pattern in effective population
138 size (N_e) between approximately 500,000 to 400 years ago, indicative of two separate bottlenecks:
139 an ancient one about 400,000 until about 150,000 years ago and a recent one between 15,000 to
140 roughly 2,000 years ago. Following the first bottleneck, N_e consistently remained smaller in the
141 pleometrotic population until between 2,000 and 400 years ago, when N_e started to increase again
142 in both populations, reaching similarly high estimates in recent times.

143 We also inferred how these populations separated over time by estimating relative cross-
144 coalescence (rCCR) [37] and found that the two populations started diverging during the first
145 bottleneck at ~300,000 years ago (Figure1C). Around 5,000 years ago, both populations were
146 almost completely isolated (rCCR ≈ 0). However, until 400 years ago, rCCR increased rapidly,
147 suggesting a growing genetic exchange, coinciding with population expansions in both populations.
148 This predicted recent gene flow is corroborated by evidence of recent genetic admixture (Figures
149 1A and 1B).

150 *Genetic diversity is significantly lower in the pleometrotic population*

151 The P-population is genetically less variable than the H-population based on genome-wide
152 nucleotide diversity (π) and heterozygosity (Figures 2A and 2B). Both estimates were significantly
153 lower in the P- (median π = 4.71e-4, median het = 0.167) compared to the H-population (median π
154 = 7.33e-4, median het = 0.274) (Figures 2A and 2B; π: Wilcoxon test: W = 3497100, p < 2.2e-16;
155 het: Wilcoxon test: W = 288, p = 4.507e-07). Accordingly, inbreeding was highly prevalent in the P-
156 population (median = 0.341) but absent in the H-population (median = -0.083) (Figure 2C; Wilcoxon
157 test: W = 16, p = 4. 507e-07). Such genome-wide reductions of diversity in the pleometrotic
158 population could be explained by a population bottleneck (founder effect), concordant with an
159 overall positive Tajima's *D* in both populations (Figure S3). This pattern is consistent with the
160 repeated colonization and extinction of ephemeral habitats by *P. californicus*.

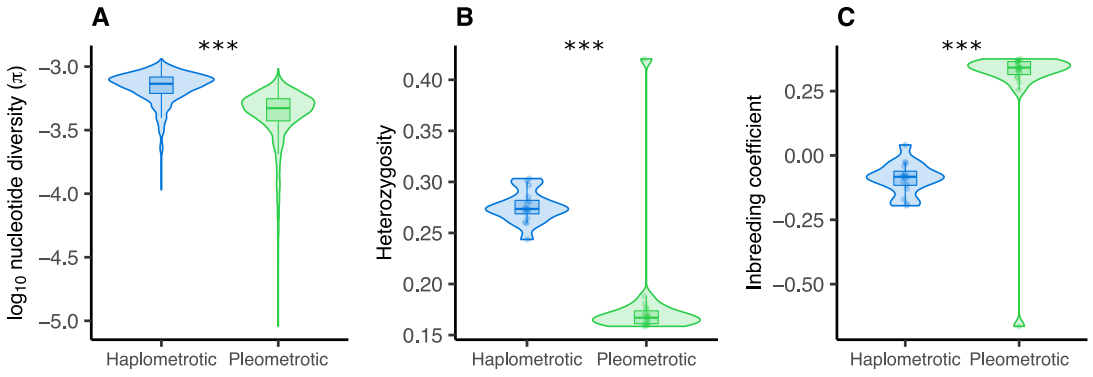

Figure 2. Higher Inbreeding and lower heterozygosity in the pleometrotic population. (A) Nucleotide diversity (π), (B) heterozygosity and (C) inbreeding coefficient in the haplometrotic (blue) and pleometrotic (green) populations. Heterozygosity and inbreeding coefficient were calculated on a per individual basis, while π estimates were calculated from 100-kb non-overlapping windows across the genome and averaged for each population. ***: $P < 0.0001$.

168 *The social polymorphism in P. californicus shows characteristics of a polygenic trait involving a*
169 *supergene*
170 Genome scans uncovered two scaffolds (22 and 23) that showed a very distinct pattern from the
171 rest of the genome. Both scaffolds had high levels of genetic differentiation between the two social
172 morphs, reduced nucleotide diversity, negative Tajima's *D*, as well as very high extended haplotype
173 scores (*xpEHH)* in the pleometrotic population (Figure S4). Further analyses revealed that SNPs
174 were in high linkage disequilibrium (LD), with little to no LD decay in these scaffolds (Figure S5).

175 Linkage analysis on 108 haploid male offspring from a single queen belonging to the monogynous
176 H-population using 2,980 ddRAD markers (double-digest Restriction-site Associated DNA
177 sequencing) with no missing data put all markers in 29 Linkage groups (data not shown). The non-
178 recombining region was found in linkage group 14 (339 RAD markers, 255 cM over 26 Mb; Figure
179 3A) and included scaffolds 22 and 23 as well as three other small scaffolds. This ~8 Mb region
180 contains 69 mapped markers in complete linkage (Figure S6). Hence, this is a non-recombining
181 region co-segregating with the P- and H-populations, spanning across five complete physical
182 scaffolds (22, 23, 38, 42, 48), and parts of scaffolds 6 and 14 (Table S2). Visual inspection of 182
183 markers from both males and queens showed that distinct haplotypes extended along the entire
184 region, indicating absence of recombination (Figure 3A) across at least ~8 Mb, consistent with
185 patterns described for the supergenes in *Solenopsis* and *Formica* ants of non-recombining regions
186 of 12.7 and 11 Mb, respectively [15,17]. Dense population genomic data using 5,363 SNPs
187 genotyped across all 35 queens further confirmed this finding (Figure S7).

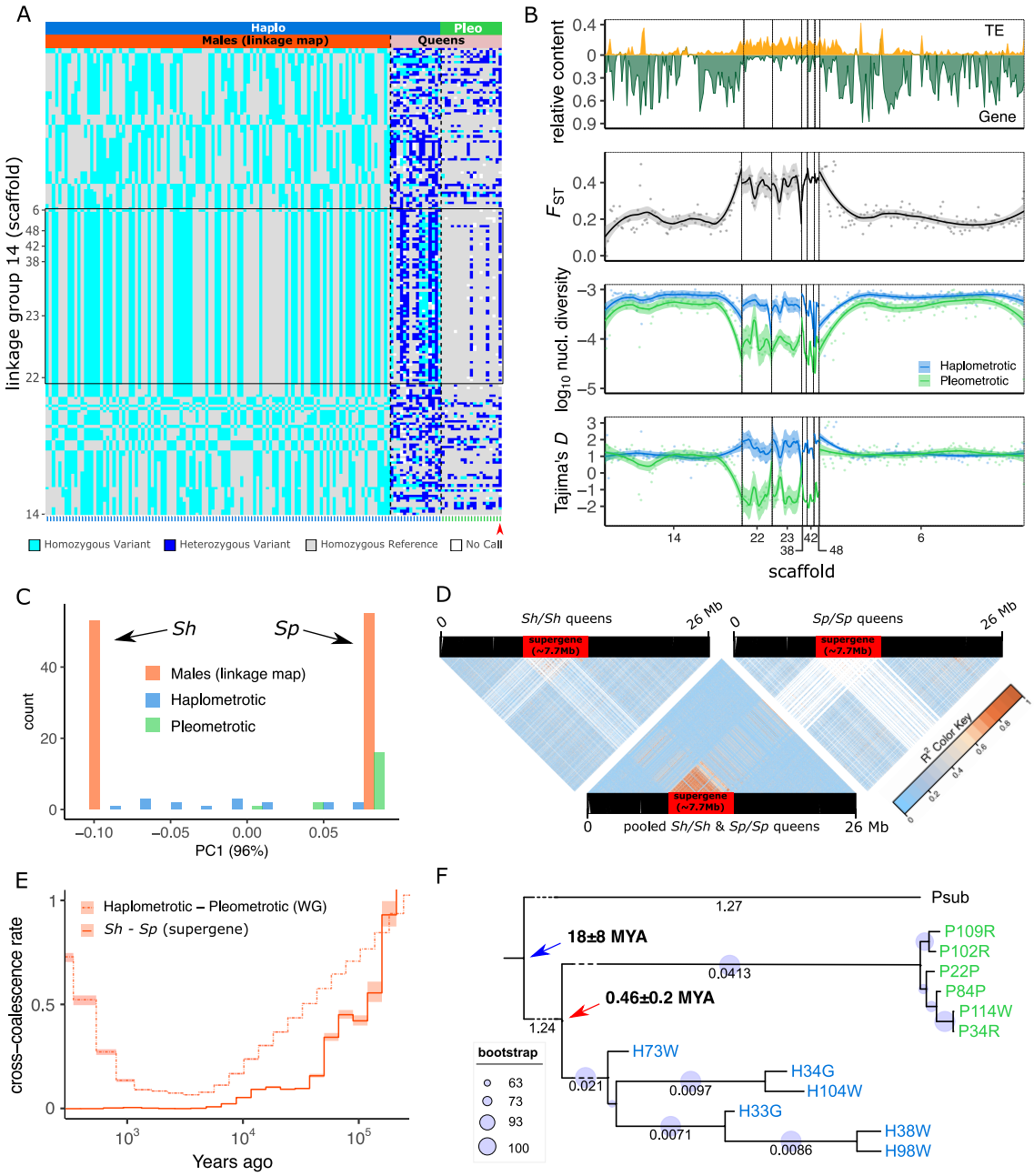

Figure 3. Genomic architecture and recombination rates at linkage group (LG) 14 harboring a putative supergene. (A) Genotypes for 108 sons from a single monogynous queen (haplometrotic population) based on RADseq markers and for the 35 queen samples from the pleometrotic and haplometrotic populations based on sites corresponding to RADseq markers from the whole genome resequencing data. Each row shows a SNP marker (182 in total) distributed across seven scaffolds on LG14. Columns (green ticks at the bottom for pleometrotic and blue ticks for haplometrotic) represent individual samples grouped by population (pleometrotic and haplometrotic) and dataset (males and queens). The black rectangle highlights the non-recombining putative supergene region, spanning scaffolds 22, 23, 38, 42, 48 and parts of scaffolds 6 and 14. Red arrowhead indicates the hybrid pleometrotic queen (P99P). (B) Sliding window analyses of LG14, showing (from top to bottom) the relative TE and gene content, genetic differentiation between populations (F_{ST}), nucleotide diversity (π) and Tajima's D estimates in the

pleometrotic (green) and haplometrotic (blue) populations. (C) Principal Component Analysis (PC1; 96%) based on 69 markers spanning the supergene region showing that the 108 male haplotypes are either Sh (negative PC1 scores) or Sp (positive PC1 scores). Diploid queens on the other hand are either homozygous for the Sp allele (most of pleometrotic queens, green bars) or heterozygous Sh/Sp (most of the haplometrotic queens, blue bars). (D) Linkage disequilibrium (LD) estimates (r^2) across LG 14 for Sh homozygous queens (n=6), pooled queens homozygous for Sh (n=6) and Sp (n=6), and Sp homozygous queens only (n=6). The red rectangle shows the region of no recombination between the Sh and Sp haplotypes. SNPs are ordered according to physical position on the chromosome after lift over. Note that invariant marker sites are displayed in white, dominating particularly the homozygous Sh/Sh and Sp/Sp LD plots. (E) Demographic history of the supergene computed using MSMC2 with queens homozygous for the Sh (n=4) and Sp (n=4) alleles. The solid line shows the relative cross-coalescence rate (rCCR) at the supergene, while the dash-dotted line gives the genome-wide rCCR estimates. Lines and shaded areas are means and 95% confidence intervals, respectively. (F) Maximum-likelihood (ML) tree for the supergene region using 33,424 SNPs and *P. subnitidus* as outgroup species. Small branch length values at the tips of the tree were omitted for ease of visualization (Newick tree can be found in Dataset S6). The blue arrow indicates the upper bond estimate for the speciation of *P. subnitidus* and *P. californicus* 18 MYA (95% HPD ±8). The red arrow indicates the age estimate for the formation of the Sh–Sp haplotype groups 0.46 (95% HPD ±0.2) MYA.

221 Similar to the pattern described for the supergenes in *Solenopsis* and *Formica*, transposable
222 element (TE) content in the *P. californicus* non-recombining region was significantly increased
223 (Wilcoxon test: W = 14448, p < 2.2e-16), whereas exon content was significantly reduced compared
224 to the rest of the same linkage group (Figures 3B and S8A; Wilcoxon test: W = 1602.5, p < 2.2e-
225 16). This is consistent with the prediction that TEs accumulate in regions of low recombination
226 [38,39]. Further, genetic differentiation as measured by F_{ST} between the H- and P-population was
227 significantly increased in the supergene compared to the rest of the linkage group (Figures 3B and
228 S8B; Wilcoxon test: W = 12261, p < 2.2e-16). Nucleotide diversity and Tajima's *D* in this region
229 were significantly reduced in the pleometrotic population (Figures 3B, S8C and S8D; Kruskal–
230 Wallis rank sum test $_{\text{nucleotide diversity}}$, χ^2 = 254.77, df = 3, p < 2.2e-16; Kruskal–Wallis rank sum test
231 $_{\text{Tajima's } D}$, χ^2 = 177.81, df = 3, p < 2.2e-16; results of pairwise Wilcoxon rank sum *post hoc* tests are
232 displayed in the figure). Finally, 156 expressed transcripts from a previously published RNA-seq
233 dataset [40] mapped to the supergene. 13 of those transcripts (8.3 %) were reported to be
234 significantly associated with aggression depending on the social context of the queens (no
235 significant enrichment, Dataset S2).

236 PCA on 69 markers along this region showed that PC1, which explains 96% of the variation,
237 separates the male dataset of our mapping population into two distinct clusters, which we labeled
238 as the Sh and Sp haplotype groups, indicating that these males are indeed the progeny of a
239 heterozygous Sh/Sp mother queen (Figures 3C and S9). Superimposing SNP data from H- and P-
240 queens to the PCA revealed that all but two pleometrotic queens (besides the hybrid queen) were
241 homozygous for the Sp allele, while most haplometrotic queens were either heterozygous (Sh/Sp)
242 or homozygous for the Sh allele (Figures 3C and S9). Further support for the suppression of
243 recombination between Sh and Sp alleles came from LD analyses across queens. The region
244 showed stronger LD in a mixed pool of Sh/Sh and Sp/Sp queens, as well as heterozygous Sh/Sp,
245 in comparison to separate pools of Sh/Sh or Sp/Sp queens (Figures 3D and S10), indicating
246 recombination in homozygous but not in heterozygous queens. The observed LD in homozygous
247 Sh and Sp queens may be attributed to genomic rearrangements similar to *Formica* [16].
248 Suppression of recombination in heterozygotes is a key feature of supergenes [9,15,19,41]. Taken

249 together, these data indicate the presence of an ~8 Mb non-recombining supergene with two major
250 haplotypes (Sh and Sp) segregating with the two populations of *P. californicus*. Unlike other social
251 supergene systems (e.g. in *Solenopsis*) [7], there were no deviations from Hardy-Weinberg
252 equilibrium (HWE) regarding the frequency of Sh/Sh, Sh/Sp and Sp/Sp genotypes at the supergene
253 locus in either population (pleometrotic: χ^2 (df = 1, N = 19) = 0.1396, p = 0.713; haplometrotic: χ^2
254 (df = 1, N = 16) = 0. 0.0711, p = 0.789) (Table S3).

255 Coalescence modelling revealed that the demographic histories of the supergene and the rest of
256 the genome only diverged within the last 5,000 years. Then, cross-coalescence quickly increased
257 across the genome, consistent with a secondary onset of gene flow between H- and P-populations,
258 while remaining nearly zero in the supergene (Figure 3E). These findings suggest that Sh and Sp
259 haplotypes initially diverged during the isolation period of the H- and P-population between
260 approximately 200,000 and 5,000 years ago. During this period, one or multiple mutations
261 suppressing recombination between Sh and Sp likely occurred (e.g. inversions), rendering the two
262 haplotypes isolated, despite the admixture between the two populations. Consistent with a very
263 recent origin of the supergene, phylogenetic analysis of Sh and Sp alleles with *P. subnitidus* as
264 outgroup showed that Sh and Sp diverged less than 0.46 (95% highest probability density (HPD)
265 0.26-0.66) MYA (Figure 3F).

266 In addition to the supergene region, our genome scans revealed several regions under divergent
267 selection between H- and P-populations. We first identified highly differentiated (outlier) regions by
268 estimating F_{ST} in 100 kb non-overlapping windows across the genome (Figures S4 and S11) and
269 subsequently compared estimates of Tajima's *D* and nucleotide diversity in the two populations to
270 identify signatures of selective sweeps (Figure S4). After clustering adjacent windows, we retained
271 51 clusters with negative Tajima's *D* and reduced nucleotide diversity consistent with signatures of
272 selective sweeps, of which 49 showed signatures of positive selection in the P-population and two
273 in the H-population (Dataset S3). Cross-population extended haplotype homozygosity (*xpEHH*)
274 tests [42] (Figure S4) identified 340 outlier SNPs in 32 genomic locations under selection (Dataset
275 S4). Overlapping both analyses resulted in 17 remaining candidate regions under selection in the

276 P-population (Table S4). Exonic SNPs in all 17 candidate regions included 70 synonymous
277 changes and 73 non-synonymous variants affecting the protein sequence (Table 1).

278

279 **Table 1.** Effects of single nucleotide polymorphisms (SNPs) found in candidate regions that show
280 strong signatures of selection in the pleometrotic population. We annotated 6,077 putative effects
281 on protein coding genes of 3,347 SNPs.
282

Effect	Count	Proportion [%]
3′UTR	13	0.21
5′UTR	1	0.02
Downstream	1391	22.89
Intergenic	3022	49.73
Intron	181	2.98
Nonsynonymous variant	73	1.20
Synonymous variant	70	1.15
Upstream	1319	21.70
Other	7	0.12

283

284

285 Twelve of these 17 regions (on scaffolds 14, 22 and 23) were in the supergene, while the remaining
286 five regions with clear signatures of selective sweeps in the P-population were clustered on
287 scaffolds 1, 9, 15, and 16 (Figure S4). These regions could initially not be placed into a linkage
288 group due to the lack of ddRAD markers. To exclude regions potentially in linkage with the
289 supergene, we constructed a relaxed linkage map with 11,986 (instead of 2,980) leniently filtered
290 ddRAD markers (MAF ≥ 0.4 and 30% missingness), In this map, one of the candidate regions
291 (scaffold 9) was contained in the supergene. The four remaining regions on scaffolds 1, 15 and 16
292 (Figure 4A), were placed into three different linkage groups, indicating that these regions are
293 independent of the supergene.

294 Among the 48 genes in these four remaining regions, 34 are likely functional as they are not
295 encoding TE proteins, are expressed, or homologous to other known insect proteins (Dataset S5).
296 Six of these 34 genes are particularly promising candidates to sustain the social niche
297 polymorphism, as they carry non-synonymous variants with a nearly perfect association of
298 phenotype and genotype, in that >90% of haplometrotic individuals carry the alternative and >90%
299 of pleometrotic individuals carry the reference allele (Figure 4B). Further, regions harboring the six
300 genes exhibit *EHHS* scores revealing a pattern of increased extended haplotype homozygosity in
301 the pleometrotic population (Figure S12), suggesting selective sweeps.

302 Visual inspection of variant positions showed that the most common pattern for all six genes is that
303 pleometrotic queens are fixed for the reference allele (except for the hybrid P99P queen), while
304 haplometrotic queens are either homozygous for alternate alleles or heterozygous (Figures S13A-
305 S13F). This is consistent with a selective sweep fixing the reference allele in the P-population.

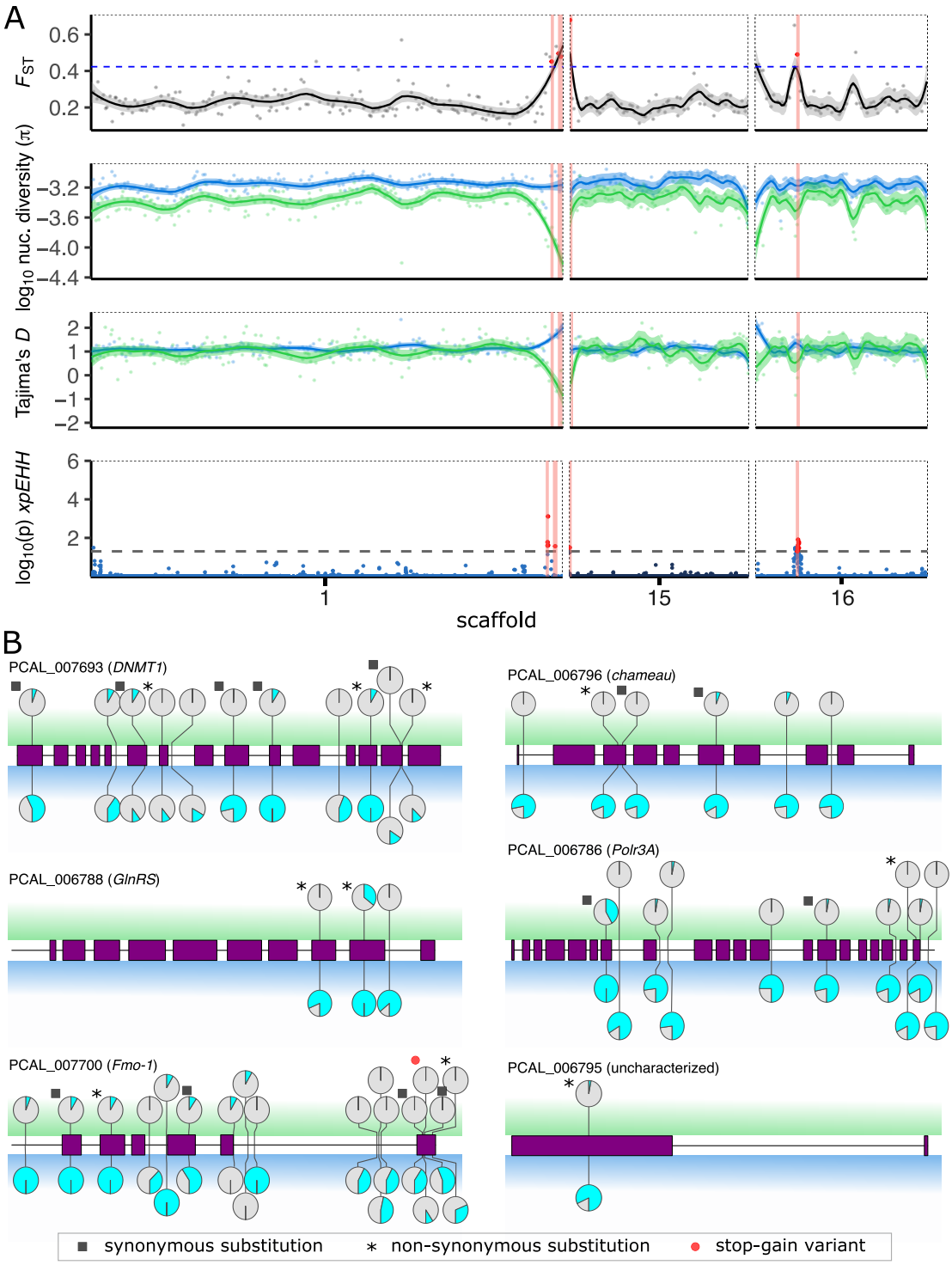

Figure 4. Signatures of selection in non-supergene regions. (**A**) Sliding window analyses of scaffolds 1, 15 and 16 showing (from top to bottom) genetic differentiation between populations (F_{ST}), nucleotide diversity (π), Tajima's D and the cross-population extended haplotype homozygosity (*xpEHH*). Horizontal blue dashed line indicates F_{ST} threshold ($F_{ST} > 0.423$; red dots), while horizontal black dashed line represents threshold of significance for outlier SNPs ($p < 0.05$; red dots) in the *xpEHH* analysis. Four candidate regions in three scaffolds outside the supergene showing signatures of selection in the pleometrotic populations are highlighted in red in all panels.

314 These regions are highly differentiated, have reduced diversity and Tajima's *D* in the pleometrotic
315 population and show a significant *xpEHH* peak. Lines and shaded areas are means and 95%
316 confidence intervals, respectively. Data for other scaffolds (1-25) are shown in Figure S4. (**B**)
317 Lollipop plots showing all variants affecting each of the six candidate genes (gene models in
318 purple). The pie charts display frequencies of the reference (grey) and alternate (cyan) alleles in
319 the pleometrotic (top; green background) and haplometrotic population (bottom; blue background).

320 Apart from one uncharacterized protein (PCAL_006795) without functional annotation, the
321 candidates encode evolutionarily well conserved proteins with functionally characterized orthologs
322 in *D. melanogaster*. PCAL_006788 is orthologous to *GlnRS* (FBgn0027090), which encodes a
323 tRNA synthetase involved in neurogenesis [43]. PCAL_006786 encodes an RNA polymerase III,
324 orthologous to the transcriptional regulator Polr3A (FBgn0030687). PCAL_007700 encodes a
325 flavin-containing monooxygenase homologous to *Drosophila* Fmo-1 (FBgn0034943) and Fmo-2
326 (FBgn0033079), which are suspected to be involved in xenobiotic detoxification and ageing [44].
327 The two remaining genes PCAL_006796 and PCAL_007693 particularly caught our attention, as
328 they are orthologs of genes involved in epigenetic regulation of transcriptional activity.
329 PCAL_006796 encodes chameau (ortholog to FBgn0028387), a MYST histone acetyl transferase
330 involved in transcriptional regulation via histone modification [45]. PCAL_007693 encodes Dnmt1,
331 a gene that is absent in Diptera. Dnmt1's primary function is to maintain methylation patterns, and
332 in social insects, inter-individual differences in DNA methylation have been linked to the epigenetic
333 regulation of phenotypic plasticity, including caste determination in bees [46].
334

335 **Discussion**

336 The social niche polymorphism [5] in colony founding of *P. californicus* was first reported over 20
337 years ago [28]. This polymorphism arises from differences in the social interaction among founding
338 queens, which will eventually lead to fundamentally different colony structures, where either
339 unrelated queens cooperate (i.e. primary polygyny) or a single queen monopolizes reproduction
340 (i.e. primary monogyny) [11]. The principal change underlying the emergence of pleometrosis in *P.*
341 *californicus* is hence the evolution of a lifelong tolerance for unrelated co-founding queens. Given
342 that the majority of North American *Pogonomyrmex sensu stricto* are primarily monogynous,
343 pleometrosis is likely a derived trait, adaptive under resource limitation and increased territorial
344 conflicts [29]. Since its original description, we have learned a lot about behavioral, physiological,
345 ecological, and socio-genetic correlates of haplo- and pleometrosis [11,26,27,29,32,33,47].
346 However, the genomic architecture and evolutionary history was unknown. The tight association of
347 social morph and geographic location in *P. californicus* poses a unique challenge when exploring
348 the genetic basis of this social polymorphism. As haplometrotic and pleometrotic individuals often
349 occur in allopatry, any signature of genetic differentiation between both social morphs could be a
350 product of genetic drift separating the populations. We therefore emphasize: While our findings are
351 consistent with a young supergene underlying the social polymorphism in this species, our results
352 are a characterization of genetic differentiation between two population that are almost fixed for
353 distinct behavioral phenotypes. Removing this geographic confounder will require studying
354 populations carrying both haplo- and pleometrotic individuals. However, despite our best efforts
355 during several field seasons in the last years, we have not yet been able to find suitable populations.
356 Thus, our study for now provides a first look at the genetic basis of social polymorphism in *P.*
357 *californicus* under these limitations.

358 We identify an approximately 8 Mb non-recombining region, reminiscent of social supergenes
359 described for social polymorphisms in *Solenopsis* fire ants [15] and *Formica* wood ants [16].
360 However, we also find genomic, population genetic, and ecological characteristics unique to the
361 *Pogonomyrmex* system, highlighting the independent evolutionary histories of the social niche
362 polymorphism in *Pogonomyrmex*, *Formica* and *Solenopsis*. The 8 Mb non-recombining region we

363 identify is similar in size to the supergenes in *Formica* and *Solenopsis* (8 Mb vs. 11 and 12.7 Mb).
364 According to estimates of F_{ST}, genetic divergence between individuals from the two social forms is
365 significantly higher within this region compared to other parts of the genome, which parallels the
366 patterns reported for the other supergenes [48]. Tajima's *D* and haplotype homozygosity statistics
367 suggest a selective sweep of the Sp haplotype, although it could also be the result of a population
368 bottleneck, similar to the *Solenopsis* Sb haplotype [49]. Very low LD decay and absence of
369 recombinant F1 offspring of a heterozygous queen (Sh/Sp) suggest little to no recombination
370 between Sh and Sp haplotypes at this locus. Further, TEs are significantly enriched at the
371 supergene, as expected in the absence of recombination following a Muller's ratchet dynamic [50].

372 Several other characteristics of the *P. californicus* social polymorphism differ from the well-studied
373 examples in *Solenopsis* and *Formica* as well. In both, *S. invicta* and *F. selysi*, there is little evidence
374 for genetic differentiation of polygyne and monogyne populations, with both forms often occurring
375 in sympatry [51,52]. In *P. californicus* however, populations are fixed for either the haplo- or the
376 pleometrotic strategy [11], with little gene flow between them. Accordingly, genetic differentiation
377 between social forms in *P. californicus* is higher than in *S. invicta* and *F. selysi* [51,52], but it is
378 similar to that observed in other *Formica* species, such as *F. exsecta* or *F. truncorum* [17]. The
379 onset of gene flow observed between the haplometrotic and pleometrotic population starting ~5,000
380 years ago could secondarily reduce genetic differentiation between both forms everywhere but in
381 the non-recombining supergene, eventually resulting in the pattern also observed in *Solenopsis*
382 *invicta* and *Formica selysi*, where genetic differentiation between both social morphs is restricted
383 to the supergene [16,48].

384 We did not find significant deviations from HWE at the non-recombining region in either population,
385 suggesting the absence of lethal haplotype combinations. This differs from *Formica* and *Solenopsis*
386 where intrinsic (such as recessive lethal effects in *Solenopsis* and maternal killing effects in
387 *Formica*) and extrinsic (such as green beard effects in *Solenopsis*) lethal combinations exist and
388 are presumed crucial for the maintenance of the social niche polymorphism [53,54]. Further unlike
389 in *Solenopsis* and *Formica*, we find additional genomic regions that segregate with the social

390 polymorphism in *P. californicus*. Theoretical models for supergene evolution predict gradual
391 establishment of supergenes by integration of formerly unlinked modifier loci [41,55]. According to
392 these models, natural selection will favor genomic rearrangements that eventually join the
393 supergene with previously unlinked but co-evolved loci carried on the same chromosome [56,57].
394 Hence, an alternative fate for the unlinked loci we describe here could be that they in fact persist
395 as (e.g. population-specific) 'modifier loci', similar to the genetic architecture for refined Batesian
396 mimicry described in *Papilio dardanus* [58–60] and *Heliconius numata* [61]. The presence of
397 unlinked modifier loci influencing the expression and ameliorating negative fitness effects of major
398 quantitative trait loci is not unusual [62], yet none have been described so far for the known social
399 supergenes.

400 The association between the genotype at the non-recombining region and the social form is
401 intriguingly imperfect in *P. californicus*. Pleometrotic queens were predominantly homozygous
402 (85%) for a single haplotype (Sp), while most haplometrotic queens (93%) carried at least one copy
403 of the Sh haplotype. This highlights that the genotype at this locus alone is insufficient to determine
404 social form – a fundamental difference to the supergenes described in other ants [17,63]! The
405 relatively young age of the suspected *Pogonomyrmex* supergene (*Pogonomyrmex*: ~0.2 MYA,
406 *Solenopsis*: 0.39–1.1 MYA [9,15,49] and *Formica*: 20–40 M years [17]) may contribute to this
407 disparity. Alternatively, the regulation of social morphs in this species may involve a polygenic
408 architecture, where the expression of haplometrosis or pleometrosis is influenced by social context,
409 as well as environmental and ecological conditions [29]. This aligns with previous studies showing
410 that environmental conditions and interactions with co-founding queens can modulate the
411 expression of aggression and tolerance in different populations [32,33].

412 The four differentiated genomic regions outside the non-recombing region contain 34 genes that
413 have undergone a selective sweep and potentially contribute to haplometrosis or pleometrosis. The
414 nature of their interaction with the presumed supergene, whether additive or epistatic, remains
415 unclear. Notably, the DNA methyltransferase *Dnmt1* and the histone acetyltransferase *chameau*
416 exhibit a near-perfect correlation between genotype and phenotype, with all P-individuals carrying

417 the reference allele and all H-individuals carrying the alternative allele. While these candidates
418 suggest that epigenetic regulation modulates the social niche polymorphisms in *P. californicus*, in-
419 depth functional studies will be necessary to explore how *chameau* or *Dnmt1* are indeed involved
420 in the expression of haplo- and pleometrosis. As pleometrotic queens show a higher behavioral
421 [32,33] and physiological [40] plasticity than haplometrotic queens during colony founding,
422 differential epigenetic regulation of plastic traits like division of labor could maintain the two distinct
423 social morphs.

424 Conclusion

425 Pleometrosis and primary polygyny in *P. californicus* likely evolved as adaptation to challenging
426 environments [29]. The genetic architecture associated with this social niche polymorphism shows
427 similarities, but also significant differences compared to those described in e.g. *Formica* and
428 *Solenopsis*. *Pogonomyrmex* also harbors a suspected non-recombining supergene, which is
429 similar in size and has two major segregating haplotypes, as in the other systems. Additionally, four
430 genomic regions showing signatures of a selective sweep in the pleometrotic population, potentially
431 influence the expression of haplo- or pleometrosis. Further investigations are needed to fully
432 understand the relevance and interactions of these loci. Unless both social morphs can be studied
433 in sympatry, it will however remain challenging to conclusively resolve whether the suspected
434 supergene and/or other genetic differences are indeed drivers of social polymorphism in this
435 species or whether they have been shaped by demographic effects or local adaptations to some
436 unknown ecological factor. Regardless, our study provides a first assessment of the genetic
437 differentiation between haplo- and pleometrotic populations of *P. californicus*. It also showcases
438 that the challenges with resolving the genetic basis for social niche polymorphisms, which may
439 often be much more complex and diverse than previously appreciated.

440 **Materials and Methods**

441 Further details about Materials and Methods are provided in the Supporting information file.

442 *Ant Collection*

443 We collected gynes of *P. californicus* after their nuptial flights in 2018 and 2019 at Pine Valley
444 (32.822 N, -116.528 W; pleometrotic population, P-population) and Lake Henshaw resort (33.232
445 N, -116.760 W; haplometrotic population, H-population), California, USA, which are 50 km apart.

446

447 *Genome sequencing*

448 DNA libraries (NEBNext® Ultra™ II FS DNA Libraries) of 16 queens from the H-population and 19
449 from the P-population were paired-end sequenced (2x 75 bp, average coverage 12x) at the
450 Genomic Core Facility of the University of Münster (Germany), using a NextSeq 500 (Table S1).

451

452 *Genome assembly and annotation*

453 We sequenced and assembled the genome of *P. californicus* ("Pcal2") using data from MinION
454 long reads (MinION, Oxford Nanopore Technologies, Oxford, UK), Illumina short reads, and
455 previously published 10X Chromium linked reads [34]. For long read library preparation (Oxford
456 Nanopore kit SQK-LSK109), we used 2 µg of DNA extracted from three white worker pupae from
457 a pleometrotic colony ("Pcal18-02"). The library was sequenced for 48 h on a MinION, generating
458 1,650,688 reads that were trimmed and filtered with PoreChop (v.0.2.4) and FiltLong (v.0.2.0). We
459 assembled the genome from ~30x filtered Nanopore data using the following strategy. We created
460 one assembly using CANU (v.2.0) [64] and a second assembly using wtdbg2 (v.2.5). Both
461 assemblies were processed and optimized as follows: After contig assembly, we used SSPACE
462 (v.3.0) [65] for scaffolding, LR_closer for gap filling, and ntedit for polishing with long-read data. We
463 then used short-read, short-insert Illumina data from three samples (P50Y, P67P, P55R) for
464 scaffolding with SSPACE before three rounds of polishing with Pilon [66]. The polished assembly
465 was super-scaffolded with publicly available 10X Chromium data for *P. californicus* (accession
466 SRX8051306) [34] using scaff10x and subsequent correction with break10x. Finally, we combined

467 both assemblies with ntjoin, using the CANU-based assembly as reference and removed
468 redundancies with funannotate clean (v.1.7.1).
469 We combined de-novo libraries from RepeatModeler2 (v.2.0.1) and EDTA (v.1.8.3) with public
470 resources from RepBase (RepBase25.04.invrep.fa) and Dfam (Dfam3.0.arthropod.fa) for repeat
471 annotation with RepeatMasker (v.4.0.7). Protein-coding and tRNA genes were annotated with
472 funannotate v.1.7.1 in the soft-masked genome, including homology-based gene predictions from
473 GeMoMa (v.1.7) [67] and supported by RNAseq data of 18 heads of *P. californicus* (Ernst et al.,
474 unpublished) and previously published long-read RNAseq data of a pool of workers [34].
475
476 *Read mapping and variant calling*
477 Short-read data was cleaned with Trimmomatic (v.0.38) [68]. Paired reads were mapped to the
478 genome assembly using BWA-MEM (v.0.7.17-r1188) [69]. After marking duplicates using Picard's
479 MarkDuplicates (v.2.20.0), we performed joint-variant calling across all samples with the GATK [70]
480 and filtered variant calls with VCFtools (v.0.1.16) [71]. After a visual inspection of mapping quality
481 of reads across the genome (Figure S14), we excluded fragmented scaffolds with low mapping
482 quality and considered 613,411 SNPs identified in the 25 largest scaffolds (representing 88.2% of
483 the assembly) for further analyses.
484
485 *Statistical phasing*
486 We used SHAPEIT (v.2.r904) [72] to statistically phase the 25 largest scaffolds of the *P. californicus*
487 genome based on 210,013 SNPs with no missing data (i.e. genotyped in all 35 queens), as no
488 reference panel is available for *P. californicus*. Scaffolds 22 and 23 displayed an atypical
489 architecture potentially representing a supergene (discussed further below) and were thus
490 excluded (along with three other small scaffolds) for fineSTRUCTURE and MSMC2 analyses.
491
492 *Population structure analyses*
493 LD-pruning with PLINK yielded 314,756 SNPs that we used for Principal Component Analyses
494 (PCA), considering four eigenvectors. LD-pruned SNPs were further used for ADMIXTURE

495 (v.1.3.0) analyses [73] with 2 to 4 clusters. To better characterize the relationships among queens
496 in our sample, we used the haplotype-based approach implemented in fineSTRUCTURE [74]. In
497 our model, we assumed a uniform recombination rate, based on the estimate of 14 cM/Mb found
498 in *Pogonomyrmex rugosus* [75].
499
500 *Demographic population history analysis*
501 We inferred demographic population history and the relative cross-coalescence rate (rCCR) using
502 Multiple Sequential Markovian Coalescence modeling (MSMC2) [37] on four queens (i.e. eight
503 phased haplotypes) from the haplometrotic (H123G, H118W, H73W and H123W) and pleometrotic
504 populations (P22P, P134P, P117P and P109R). Mappability masks were generated using
505 SNPABLE [http://lh3lh3.users.sourceforge.net/snpable.shtml]. We ran 1,000 bootstrap replicates,
506 each time randomly sampling 60 Mb for each of the analyzed genomes. To obtain the time in years,
507 we assumed a generation time of 4 years and a similar mutation rate as in honey bees [76], i.e. 3.6
508 * 10^{-9} mutations * nucleotide^{-1} * generation^{-1}.
509
510 *Population genomic estimates*
511 We used VCFtools (v.0.1.16) [71] to estimate nucleotide diversity (π) (*--window-pi)* and Tajima's *D*
512 *(--TajimaD*) within each population using 100 kb non-overlapping windows across the genome.
513 VCFtools' *--het* function was employed to calculate heterozygosity and inbreeding coefficient on a
514 per individual basis. Linkage disequilibrium (LD) decay was estimated by calculating pairwise r^2
515 values between all markers within 200 kb windows using VCFtools' *--hap-r2* function in each of the
516 two populations. We used SNP markers with a minor allele frequency (MAF) of 0.2 to minimize the
517 effects of rare variants and averaged r^2 values between all pairs of markers in 100 bp distance bins
518 for visualization.
519
520 *Genome scan for selection*
521 We screened genomes of both populations for evidence of selection using an F_{ST} outlier approach
522 [77–80] and by identifying regions of extended haplotype homozygosity (selective sweeps) [42,81],

523 comparing the haplometrotic and pleometrotic populations. We excluded a total of six queens from
524 pairs with relatedness > 0.4 and a potential F1 hybrid (P99P) from the analyses. We used mean
525 F_{ST}, nucleotide diversity π, and Tajima's D estimated in 100 kb non-overlapping windows across
526 the genome and from the empirical distribution of F_{ST} values, we considered the top 5% of windows
527 as differentiated.
528 F_{ST} estimates can be affected due to variation in recombination rate across the genome [82].
529 Hence, we further applied the cross-population extended haplotype homozygosity (*xpEHH*) and
530 the site-specific extended haplotype homozygosity (*EHHS*) tests [42,81], to detect signatures of
531 selective sweeps. The *xpEHH* analyses were performed on the phased data from unrelated
532 queens using the R package rehh (v.3.1.2) [83]. We used the Benjamani & Hochberg method for
533 FDR correction and set a p-value of 0.05 as a threshold to identify outlier SNPs. Outliers occurring
534 within 100 kb of each other were grouped, resulting in 32 clusters distributed across multiple
535 scaffolds (Dataset S4). Regions within -/+ 100 kb around the most significant SNP of each cluster
536 were regarded as candidates for positive selective sweeps. Regions identified by F_{ST} and by *xpEHH*
537 overlapped at 17 genomic loci (Table S4).
538
539 *Characterization of candidate genes and SNPs*
540 For functional characterization of genes, we performed BLAST (blastx) [84] searches against the
541 nr database (January 2020), used OrthoFinder (v.2.3.1) [85] to identify *Drosophila melanogaster*
542 orthologs, and used RepeatProteinMask to identify genes encoding TE proteins. We used SnpEff
543 (v.5.0) [86] to annotate putative effects on protein-coding genes of all 3,347 SNPs contained in the
544 candidate regions (Table 1). For the manual review of genes contained in genomic regions showing
545 evidence for a selective sweep, we removed TE-encoded genes as well as genes without RNAseq
546 support and no known homology to any proteins in the nr database, resulting in 34 putatively
547 functional genes. Of these, 18 genes showed non-synonymous variants in our dataset (Dataset
548 S5). In six of these 18 genes at least one non-synonymous variant showed a pattern where over
549 90% of the individuals from the H-population carried the alternative allele (i.e. either 0/1 or 1/1) and

550 over 90% of the individuals from the P-population carried the reference allele (i.e. either 0/1 or 0/0)
551 (Figure 4B).
552
553 *Haplotype frequencies*
554 We calculated haplotype frequencies for the putative supergene region and tested for potential
555 departure from Hardy-Weinberg equilibrium using a χ^2 test (Table S3). Haplotypes were assigned
556 manually, based on visual inspection of the suggested supergene region (Figure S7);
557 predominantly homozygous for the reference alleles (i.e. from a pleometrotic population) and
558 alternate allele were assigned "Sp/Sp" and "Sh/Sh", respectively, and predominantly heterozygous
559 regions were assigned "Sh/Sp".
560
561 *RAD sequencing and analysis*
562 DNA of males collected from a monogynous colony at Lake Henshaw (H-population) in 2019 was
563 used to construct reduced representation genomic libraries following Inbar et al. [87], using a
564 modified double-digest (EcoR1 and Mse1) Restriction-site Associated DNA sequencing
565 (ddRADseq) protocol [88,89]. Briefly, genomic DNA was digested at 37°C for 8hrs and ligated with
566 uniquely barcoded adaptors. The adaptor ligated products were PCR amplified, pooled, and size-
567 selected using Beckman Coulter, Agencourt AMPure XP beads. Libraries were sequenced in one
568 lane of a paired-end, 150 bp reads, on a HiSeq X Ten Illumina sequencer.
569 We used fastQC (v.0.11.7), Trimmomatic (v.0.39) [68], and *process_radtags* (STACKS v.2.2) [90]
570 for quality control, filtering, and demultiplexing the raw data. Reads were mapped with Bowtie2
571 v2.3.4.2 (Langmead et al 2012) and alignments processed with *ref_map.pl* and *populations*
572 (STACKS v.2.2), resulting in 55,340 raw variants. After filtering with VCFtools (v.0.1.16) [71] and
573 excluding heterozygous markers we arrived at a final set of 2,980 markers used for building a
574 genetic map.
575 Markers were mapped with MultiPoint [http://www.multiqtl.com] using the backcross population
576 setting. A maximum threshold recombination frequency of 0.25 resulted in 29 linkage groups (data

577 not shown). A skeleton map was generated using delegate markers and Kosambi's mapping
578 function was used to convert recombination frequencies to centimorgans [91].
579
580 *Characterization of the supergene*
581 To investigate the potential supergene identified in the current study, we used SNP markers from
582 queens (whole genome resequencing) and males (RADseq). We used BCFtools' (v.1.9) *isec* and
583 *merge* functions to extract 182 shared SNPs between the two datasets found on LG14. Genotypes
584 of the males and queens at LG14 were visualized using VIVA (v.0.4.0) [92]. To further characterize
585 the two alleles of the supergene, we performed a PCA on 69 SNP markers identified in the non-
586 recombining area of LG14 (i.e. scaffolds 22, 23, 38, 42, 48, and parts of 6 and 14) in datasets of
587 males and queens.
588 To investigate recombination suppression in the supergene region, we used SNP data from six
589 queens homozygous for the Sh allele (H104W, H33G, H34G, H38W, H73W and H98W) and six
590 others homozygous for the Sp allele (P22P, P102R, P114W, P109R, P34R, P84P), assigned based
591 on visual inspection of their genotypes (Figure S7). After VCF lift over (on scaffolds 14, 22, 23, 38,
592 42, 48 and 6 into LG14) using flo [48] and picard's *LiftoverVcf* (v.2.20.0), we used VCFtools
593 (v.0.1.16) [71] for filtering and LDheatmap (v.1.0-4) [93] to visualize pairwise linkage disequilibrium
594 (LD). We calculated gene and TE content in 100 kb sliding windows across LG14 using our *P.*
595 *californicus* gene and TE annotations, respectively. To explore whether any of the transcripts
596 previously shown to be associated with the aggressive behavior of queens [40] belong to the
597 supergene, we aligned the assembled transcripts to the Pcal2 reference genomes using GMAP
598 [94]. We only retained the top hits with at least 70% query coverage, yielding 7303 (out of 7890)
599 transcripts unambiguously aligned to the reference genome. We then used BEDtools (v.2.29.2)
600 [95] to extract transcripts that mapped to the supergene region (Dataset S2).
601
602 *Dating the supergene formation*
603 We estimated the divergence between Sh and Sp by computing the rCCR, using MSMC2 [37] on
604 four Sh/Sh queens (H104W, H38W, H73W and H98W) and four Sp/Sp queens (P114W, P50R,

605 P84P and P97R). We limited the analysis to the supergene region (scaffolds 22, 23 and part of
606 scaffold 6) and ran 500 bootstraps by randomly sampling 2 Mb each time.
607 Further, we performed a phylogenetic analysis of Sh and Sp alleles with *P. subnitidus* as an
608 outgroup species. Sequencing data of one *P. subnitidus* individual was generated and processed
609 as described above for *P. californicus*. Joint SNP calling was performed using GATK [70] and
610 a maximum-likelihood (ML) tree was constructed based on 33,424 SNPs at the supergene region
611 (scaffolds 22, 23, 38, 42, 48 and parts of scaffolds 6 and 14), using RAxML [96] under the
612 GTRGAMMA model with bootstrapping.
613 We then used a fossil-calibrated dated phylogeny [97] to date the split between the Sh and Sp
614 haplotype groups using a phylogenetic approach. The Ward et al. [97] phylogeny dated the split
615 between *P. vermiculatus* and *P. imberbiculus* at 18 (95% HPD ±8) MYA. This is an upper bound
616 on the speciation date of *P, californicus* and *P. subnitidus* according to the most recent
617 *Pogonomyrmex* phylogeny [98]. We calculated a ratio of 1:39 between the speciation time of *P.*
618 *subnitidus* and *P. californicus*, and the Sh–Sb haplotypes divergence time. Based on our dating
619 reference (i.e. 18±8 MYA), we obtained 0.46 (95% HPD ±0.20) MYA as an upper bound for the
620 divergence between the two haplotype groups, which is consistent with the MSMC2 analysis. We
621 note that the confidence in the age inference is strengthened by the agreement between two distinct
622 approaches: the inference by MSMC2 is based on a coalescent model, which is a distinct approach
623 from the phylogenetic dating approach. The former approach depends on estimates of mutation
624 rate and generation time, whereas the latter depends on fossil calibration points. The tree figure
625 was produced using interactive Tree Of Life (iTOL) [99].
626

627 **Data Availability**

628 The genome assembly and the raw sequencing data prior to trimming and mapping are available
629 at NCBI (BioProject: PRJNA682388; release pending acceptance). A detailed description of our
630 bioinformatic analysis pipelines is publicly available at https://github.com/merrbii/Pcal-
631 SocialPolymorphism.

632

633 **Acknowledgments**

634 We thank Pine Valley Academy (Pine Valley, California, USA) for support and the staff of Lake
635 Henshaw resort (Lake Henshaw, California, USA) for permission to collect ants on their premises.
636 Thanks to Jenny Märzhäuser, Tobias van Elst and Ti Eriksson for help in collecting the ants, and
637 to Jennifer Fewell for hosting us at Arizona State University, Tempe, Arizona, USA. We are grateful
638 for Hilde Schwitte's and Una Hadziomerovic's help in DNA extraction, barcoding, and library
639 construction. JG and UE were funded by the German Research Foundation (DFG) as part of the
640 SFB TRR 212 (NC^3) – TP C04 project numbers 316099922 and 396780988. ME was funded by
641 the German Research Foundation (DFG) – 403813881 with grant to LS (SCHR 1554/2-1) under
642 the priority program "Rapid evolutionary adaptation: Potential and constraints" (SPP 1819). ME
643 acknowledges support by the DFG Research Training Group 2220 "Evolutionary Processes in
644 Adaptation and Disease".

645

646 **Competing interests**

647 The authors have no conflict of interest to declare.

648

649 **Author Contributions**

650 JG, UE, ME and LS designed the research, UE organized sample collection and behavioral assays,
651 AL, LS, ME, UE, EP and JG analyzed data, ME, UE, LS and JG wrote the manuscript, JG and LS
652 coordinated the project, all authors revised the manuscript.

653

References

1. Wilson EO. The insect societies. Cambridge, Massachusetts, USA: Harvard University Press; 1971.
2. Hölldobler B, Wilson EO. The number of queens: An important trait in ant evolution. Naturwissenschaften. 1977;64: 8–15. doi:10.1007/BF00439886
3. Boomsma JJ, Huszár DB, Pedersen JS. The evolution of multiqueen breeding in eusocial lineages with permanent physically differentiated castes. Anim Behav. 2014;92: 241–252. doi:10.1016/j.anbehav.2014.03.005
4. Keller L. The Assessment of Reproductive Success of Queens in Ants and Other Social Insects. Oikos. 1993;67: 177. doi:10.2307/3545107
5. Saltz JB, Geiger AP, Anderson R, Johnson B, Marren R. What, if anything, is a social niche? Evol Ecol. 2016;30: 349–364. doi:10.1007/s10682-015-9792-5
6. Seppä P, Sundström L, Punttila P. Facultative polygyny and habitat succession in boreal ants. Biol J Linn Soc. 1995;56: 533–551. doi:10.1111/j.1095-8312.1995.tb01109.x
7. Ross KG, Krieger MJB, Keller L, Shoemaker DD. Genetic variation and structure in native populations of the fire ant *Solenopsis invicta*: Evolutionary and demographic implications. Biol J Linn Soc. 2007;92: 541–560. doi:10.1111/j.1095-8312.2007.00853.x
8. Fletcher DJC. Three newly-discovered polygynous populations of the fire ant, *Solenopsis invicta*, and their significance. J Georg Entomol Soc. 1983;18: 538–543.
9. Yan Z, Martin SH, Gotzek D, Arsenault S V., Duchen P, Helleu Q, et al. Evolution of a supergene that regulates a trans-species social polymorphism. Nat Ecol Evol. 2020;4: 240–249. doi:10.1038/s41559-019-1081-1
10. Helms KR, Newman NJ, Helms Cahan S. Regional variation in queen and worker aggression in incipient colonies of the desert ant *Messor pergandei*. Behav Ecol Sociobiol. 2013;67: 1563–1573. doi:10.1007/s00265-013-1568-8
11. Overson R, Fewell JH, Gadau J. Distribution and origin of intraspecific social variation in the California harvester ant *Pogonomyrmex californicus*. Insectes Soc. 2016;63: 531–541. doi:10.1007/s00040-016-0497-8
12. Brodetzki TR, Inbar S, Cohen P, Aron S, Privman E, Hefetz A. The Interplay between Incipient Species and Social Polymorphism in the Desert Ant Cataglyphis. Sci Rep. 2019;9: 1–14. doi:10.1038/s41598-019-45950-1
13. Gill RJ, Arce A, Keller L, Hammond RL. Polymorphic social organization in an ant. Proc R Soc B Biol Sci. 2009;276: 4423–4431. doi:10.1098/rspb.2009.1408
14. Herbers JM. Nest site limitation and facultative polygyny in the ant *Leptothorax longispinosus*. Behav Ecol Sociobiol. 1986;19: 115–122. doi:10.1007/BF00299946
15. Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang Y-C, Shoemaker D, et al. A Y-like social chromosome causes alternative colony organization in fire ants. Nature. 2013;493: 664–668. doi:10.1038/nature11832
16. Purcell J, Brelsford A, Wurm Y, Perrin N, Chapuisat M. Convergent genetic architecture underlies social organization in ants. Curr Biol. 2014;24: 2728–2732. doi:10.1016/j.cub.2014.09.071
17. Brelsford A, Purcell J, Avril A, Tran Van P, Zhang J, Brütsch T, et al. An Ancient and Eroded Social Supergene Is Widespread across *Formica* Ants. Curr Biol. 2020;30: 304-311.e4. doi:10.1016/j.cub.2019.11.032
18. Purcell J, Lagunas-Robles G, Rabeling C, Borowiec ML, Brelsford A. The maintenance of polymorphism in an ancient social supergene. Mol Ecol. 2021;00: 1– 13. doi:10.1111/mec.16196
19. Schwander T, Libbrecht R, Keller L. Supergenes and complex phenotypes. Current Biology. Cell Press; 2014. pp. R288–R294. doi:10.1016/j.cub.2014.01.056
20. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA, et al. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. Begun DJ, editor. PLoS Genet. 2010;6: e1000862. doi:10.1371/journal.pgen.1000862
21. Matschiner M, Barth JMI, Torresen OK, Star B, Baalsrud HT, Brieuc MSO, et al. Origin and fate of supergenes in Atlantic cod. bioRxiv. 2021; 1–34.

doi:10.1101/2021.02.28.433253
22. Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, et al. Genomic islands of divergence in hybridizing heliconius butterflies identified by large-scale targeted sequencing. Philos Trans R Soc B Biol Sci. 2012;367: 343–353. doi:10.1098/rstb.2011.0198
23. Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, Baxter S, et al. A conserved supergene locus controls colour pattern diversity in Heliconius butterflies. PLoS Biol. 2006;4: 1831–1840. doi:10.1371/journal.pbio.0040303
24. Favreau E, Lebas C, Stolle E, Priyam A, Pracana R, Aron S, et al. No supergene despite social polymorphism in the big-headed ant *Pheidole pallidula*. bioRxiv. 2022; 2022.12.06.519286. doi:10.1101/2022.12.06.519286
25. Buckley SB. Descriptions of new species of North American Formicidae. Proc Entomol Soc Philadelphia. 1867;6: 152–172.
26. Johnson RA. Semi-claustral colony founding in the seed-harvester ant *Pogonomyrmex californicus*: A comparative analysis of colony founding strategies. Oecologia. 2002;132: 60–67. doi:10.1007/s00442-002-0960-2
27. Johnson RA. Colony founding by pleometrosis in the semiclaustral seed-harvester ant *Pogonomyrmex californicus* (Hymenoptera: Formicidae). Anim Behav. 2004;68: 1189–1200. doi:10.1016/j.anbehav.2003.11.021
28. Rissing SW, Johnson RA, Martin JW. Colony Founding Behavior of Some Desert Ants: Geographic Variation in Metrosis. Psyche A J Entomol. 2000;103: 95–101. doi:10.1155/2000/20135
29. Haney BR, Fewell JH. Ecological drivers and reproductive consequences of non-kin cooperation by ant queens. Oecologia. 2018;187: 643–655. doi:10.1007/s00442-018-4148-9
30. Rosengren R, Sundström L, Fortelius W. Monogyny and polygyny in *Formica* ants: the results of alternative dispersal tactics. In: Keller L, editor. Queen number and sociality in insects. Oxford University Press; 1993. pp. 308–333.
31. Glancey BM, Lofgren CS. Adoption of Newly-Mated Queens: A Mechanism for Proliferation and Perpetuation of Polygynous Red Imported Fire Ants, *Solenopsis invicta* Buren. Florida Entomol. 1988;71: 581. doi:10.2307/3495016
32. Overson R, Gadau J, Clark RM, Pratt SC, Fewell JH. Behavioral transitions with the evolution of cooperative nest founding by harvester ant queens. Behav Ecol Sociobiol. 2014;68: 21–30. doi:10.1007/s00265-013-1618-2
33. Clark RM, Fewell JH. Social dynamics drive selection in cooperative associations of ant queens. Behav Ecol. 2014;25: 117–123. doi:10.1093/beheco/art093
34. Bohn J, Halabian R, Schrader L, Shabardina V, Steffen R, Suzuki Y, et al. Genome assembly and annotation of the California harvester ant *Pogonomyrmex californicus*. G3. 2021;11. doi:10.1093/g3journal/jkaa019
35. Boomsma JJ, Brady SG, Dunn RR, Gadau J, Heinze J, Keller L, et al. The Global Ant Genomics Alliance (GAGA). Myrmecological News. 2017;25: 61–66.
36. Schiffels S, Wang K. MSMC and MSMC2: The multiple sequentially Markovian coalescent. Methods in Molecular Biology. Humana Press Inc.; 2020. pp. 147–166. doi:10.1007/978-1-0716-0199-0_7
37. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. Nat Genet. 2014;46: 919–925. doi:10.1038/ng.3015
38. Dolgin ES, Charlesworth B. The effects of recombination rate on the distribution and abundance of transposable elements. Genetics. 2008;178: 2169–77. doi:10.1534/genetics.107.082743
39. Kent TV, Uzunović J, Wright SI. Coevolution between transposable elements and recombination. Philos Trans R Soc Lond B Biol Sci. 2017;372. doi:10.1098/rstb.2016.0458
40. Helmkampf M, Mikheyev AS, Kang Y, Fewell J, Gadau J. Gene expression and variation in social aggression by queens of the harvester ant *Pogonomyrmex californicus*. Mol Ecol. 2016;25: 3716–3730. doi:10.1111/mec.13700
41. Charlesworth D. The status of supergenes in the 21st century: Recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations.

Evol Appl. 2016;9: 74–90. doi:10.1111/eva.12291
42. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007;449: 913–918. doi:10.1038/nature06250
43. Chihara T, Luginbuhl D, Luo L. Cytoplasmic and mitochondrial protein translation in axonal and dendritic terminal arborization. Nat Neurosci. 2007;10: 828–837. doi:10.1038/nn1910
44. Rossner R, Kaeberlein M, Leiser SF. Flavin-containing monooxygenases in aging and disease: Emerging roles for ancient enzymes. Journal of Biological Chemistry. 2017. pp. 11138–11146. doi:10.1074/jbc.R117.779678
45. Carrozza MJ, Utley RT, Workman JL, Côté J. The diverse functions of histone acetyltransferase complexes. Trends Genet. 2003;19: 321–329. doi:10.1016/S0168-9525(03)00115-X
46. Lyko F. The DNA methyltransferase family: A versatile toolkit for epigenetic regulation. Nature Reviews Genetics. 2018. pp. 81–92. doi:10.1038/nrg.2017.80
47. Shaffer Z, Sasaki T, Haney B, Janssen M, Pratt SC, Fewell JH. The foundress's dilemma: Group selection for cooperation among queens of the harvester ant, *Pogonomyrmex californicus*. Sci Rep. 2016;6: 1–9. doi:10.1038/srep29828
48. Pracana R, Priyam A, Levantis I, Nichols RA, Wurm Y. The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. Mol Ecol. 2017;26: 2864–2879. doi:10.1111/mec.14054
49. Cohen P, Privman E. The social supergene dates back to the speciation time of two *Solenopsis* fire ant species. Sci Rep. 2020;10: 1–9. doi:10.1038/s41598-020-67999-z
50. Bachtrog D. The temporal dynamics of processes underlying Y chromosome degeneration. Genetics. 2008;179: 1513–1525. doi:10.1534/genetics.107.084012
51. Chapuisat M, Bocherens S, Rosset H. Variable queen number in ant colonies: No impact on queen turnover, inbreeding, and population genetic differentiation in the ant *Formica selysi*. Evolution. 2004;58: 1064–1072. doi:10.1111/j.0014-3820.2004.tb00440.x
52. Shoemaker DDW, Ahrens ME, Ross KG. Molecular phylogeny of fire ants of the *Solenopsis saevissima* species-group based on mtDNA sequences. Mol Phylogenet Evol. 2006;38. doi:10.1016/j.ympev.2005.07.014
53. Avril A, Purcell J, Béniguel S, Chapuisat M. Maternal effect killing by a supergene controlling ant social organization. Proc Natl Acad Sci U S A. 2020;117: 17130–17134. doi:10.1073/pnas.2003282117
54. Ross KG. Multilocus evolution in fire ants: Effects of selection, gene flow and recombination. Genetics. 1997;145. doi:10.1093/genetics/145.4.961
55. Gutiérrez-Valencia J, Hughes PW, Berdan EL, Slotte T. The Genomic Architecture and Evolutionary Fates of Supergenes. Genome Biol Evol. 2021;13: 1–19. doi:10.1093/gbe/evab057
56. Thompson MJ, Jiggins CD. Supergenes and their role in evolution. Heredity. 2014. pp. 1–8. doi:10.1038/hdy.2014.20
57. Charlesworth D, Charlesworth B. Theoretical genetics of batesian mimicry II. Evolution of supergenes. J Theor Biol. 1975;55: 305–324. doi:10.1016/S0022-5193(75)80082-8
58. Clark R, Brown SM, Collins SC, Jiggins CD, Heckel DG, Vogler AP. Colour pattern specification in the Mocker swallowtail Papilio dardanus: the transcription factor invected is a candidate for the mimicry locus H. Proc R Soc B Biol Sci. 2008;275: 1181–1188. doi:10.1098/RSPB.2007.1762
59. Thompson MJ, Timmermans MJTN. Characterising the Phenotypic Diversity of Papilio dardanus Wing Patterns Using an Extensive Museum Collection. PLoS One. 2014;9: e96815. doi:10.1371/JOURNAL.PONE.0096815
60. Clarke CA, Sheppard PM. Interactions between major genes and polygenes in the determination of the mimetic patterns of *Papilio dardanus*. Evolution. 1963;17: 404–413. doi:10.1111/j.1558-5646.1963.tb03297.x
61. Jones RT, Salazar PA, ffrench-Constant RH, Jiggins CD, Joron M. Evolution of a mimicry supergene from a multilocus architecture. Proc R Soc B Biol Sci. 2012;279: 316–325. doi:10.1098/RSPB.2011.0882

62. Flint J, Mackay TFC. Genetic architecture of quantitative traits in mice, flies, and humans. Genome Res. 2009;19: 723–733. doi:10.1101/GR.086660.108
63. Ross KG, Keller L. Genetic control of social organization in an ant. Proc Natl Acad Sci. 1998;95: 14232–14237. doi:10.1073/pnas.95.24.14232
64. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. Genome Res. 2017;27: 722–736. doi:10.1101/gr.215087.116
65. Boetzer M, Henkel C V., Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27: 578–579. doi:10.1093/bioinformatics/btq683
66. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. Wang J, editor. PLoS One. 2014;9: e112963. doi:10.1371/journal.pone.0112963
67. Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. 2016;44: 89. doi:10.1093/nar/gkw092
68. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170
69. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324
70. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr Protoc Bioinforma. 2013;43: 11.10.1-11.10.33. doi:10.1002/0471250953.bi1110s43
71. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27: 2156–2158. doi:10.1093/bioinformatics/btr330
72. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods. 2012;9: 179–181. doi:10.1038/nmeth.1785
73. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19: 1655–1664. doi:10.1101/gr.094052.109
74. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of Population Structure using Dense Haplotype Data. PLoS Genet. 2012;8: 1002453. doi:10.1371/journal.pgen.1002453
75. Sirviö A, Pamilo P, Johnson RA, Page RE, Gadau J. Origin and evolution of the dependent lineages in the genetic caste determination system of *Pogonomyrmex* ants. Evolution. 2011;65: 869–84. doi:10.1111/j.1558-5646.2010.01170.x
76. Yang S, Wang L, Huang J, Zhang X, Yuan Y, Chen JQ, et al. Parent-progeny sequencing indicates higher mutation rates in heterozygotes. Nature. 2015;523: 463–467. doi:10.1038/nature14649
77. Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, et al. Tracking footprints of artificial selection in the dog genome. Proc Natl Acad Sci U S A. 2010;107: 1160–1165. doi:10.1073/pnas.0909918107
78. Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlötterer C, et al. Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. Mol Ecol. 2012;21: 4748–4769. doi:10.1111/j.1365-294X.2012.05731.x
79. Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM. Genomic signatures of positive selection in humans and the limits of outlier approaches. Genome Res. 2006;16: 980–989. doi:10.1101/gr.5157306
80. Kolaczkowski B, Kern AD, Holloway AK, Begun DJ. Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. Genetics. 2011;187: 245–260. doi:10.1534/genetics.110.123059
81. Tang K, Thornton KR, Stoneking M. A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. Wolfe KH, editor. PLoS Biol. 2007;5: e171. doi:10.1371/journal.pbio.0050171
82. Booker TR, Yeaman S, Whitlock MC. Variation in recombination rate affects detection of

outliers in genome scans under neutrality. Mol Ecol. 2020;29: 4274–4279. doi:10.1111/MEC.15501

83. Gautier M, Klassmann A, Vitalis R. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Molecular Ecology Resources. 2017. pp. 78–90. doi:10.1111/1755-0998.12634

84. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-2

85. Emms DM, Kelly S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20: 238. doi:10.1186/s13059-019-1832-y

86. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin). 2012;6: 80–92. doi:10.4161/fly.19695

87. Inbar S, Cohen P, Yahav T, Privman E. Comparative study of population genomic approaches for mapping colony-level traits. Mikheyev A, editor. PLOS Comput Biol. 2020;16: e1007653. doi:10.1371/journal.pcbi.1007653

88. Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA. Genome-wide association genetics of an adaptive trait in lodgepole pine. Mol Ecol. 2012;21: 2991–3005. doi:10.1111/j.1365-294X.2012.05513.x

89. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 2012;7. doi:10.1371/journal.pone.0037135

90. Rochette NC, Rivera-Colón AG, Catchen JM. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol Ecol. 2019;28. doi:10.1111/mec.15253

91. Kosambi DD. The Estimation of Map Distances from Recombination Values. Ann Eugen. 1943;12: 172–175. doi:10.1111/j.1469-1809.1943.tb02321.x

92. Tollefson GA, Schuster J, Gelin F, Agudelo A, Ragavendran A, Restrepo I, et al. VIVA (VIsualization of VAriants): A VCF File Visualization Tool. Sci Rep. 2019;9: 1–7. doi:10.1038/s41598-019-49114-z

93. Shin J-H, Blay S, Mcneney B, Graham J. LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria between Single Nucleotide Polymorphisms. JSS J Stat Softw. 2006;16. doi:10.18637/jss.v000.i00

94. Wu TD, Watanabe CK. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. 2005;21: 1859–1875. doi:10.1093/bioinformatics/bti310

95. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

96. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30: 1312–1313. doi:10.1093/bioinformatics/btu033

97. Ward PS, Brady SG, Fisher BL, Schultz TR. The evolution of myrmicine ants: Phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). Syst Entomol. 2015;40: 61–81. doi:10.1111/syen.12090

98. Johnson RA, Moreau CS. A new ant genus from southern Argentina and southern Chile, *Patagonomyrmex* (Hymenoptera: Formicidae). Zootaxa. 2016;4139: 1–31. doi:10.11646/zootaxa.4139.1.1

99. Letunic I, Bork P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49: W293–W296. doi:10.1093/nar/gkab301