

RESEARCH ARTICLE

Open Access



Evolutionary genomics of socially polymorphic populations of *Pogonomyrmex californicus*

Mohammed Errbii¹, Ulrich R. Ernst^{1,3,4}, Aparna Lajmi², Eyal Privman², Jürgen Gadau^{1*†} and Lukas Schrader^{1*†} 

Abstract

Background 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 also 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.

Results 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*).

Conclusions Altogether, our 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.

Keywords Haplometrosis, Pleometrosis, Colony founding, Supergene, Social niche polymorphism, Polygenic trait, *Pogonomyrmex*

[†]Jürgen Gadau and Lukas Schrader contributed equally to this work.

*Correspondence:

Jürgen Gadau

gadauj@uni-muenster.de

Lukas Schrader

lukas.schrader@uni-muenster.de

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

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., [17, 20–23]). 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 [15, 16]. 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.

Results

Pogonomyrmex californicus genome assembly

To improve the available fragmented genome assembly, we combined minION long-read sequencing and 10× sequencing [34] of pleometrotic individuals to produce a highly contiguous reference genome assembly for *P. californicus*. The assembly spans 252.3 Mb across 199 scaffolds (N50 = 10.4 Mb, largest scaffold 20.75 Mb). We recovered 98.5% complete BUSCOs (S: 97.8%, D: 0.7%, F: 1.1%, M: 0.4%, n: 4415), suggesting a nearly complete genome assembly. Genome annotation identified 22.79% repetitive sequences and 15,899 protein-coding genes, resembling repeat and gene contents reported for other published ant genomes [35] (Additional file 1).

The haplometrotic and pleometrotic populations are genetically distinct

For our population genomic analyses, we performed whole-genome sequencing of 35 founding queens (19 from the P-population and 16 from the H-population). Principal component analyses (PCA) using 314,756 single-nucleotide polymorphisms (SNPs) clearly separated queens of the two populations (Fig. 1a, Additional file 2: Fig. S1) [34, 36–90]. Further, the PCA suggested higher genetic diversity in the H-population compared to the P-population. One founding queen (P99P) collected from the P-population was positioned between both populations in the PCA (Fig. 1a), likely representing an F1 hybrid between a haplo- and a pleometrotic individual.

A detailed analysis of coancestry showed that queens share more haplotypes in within-population pairs than in between-population pairs (Fig. 1b). Consistent with

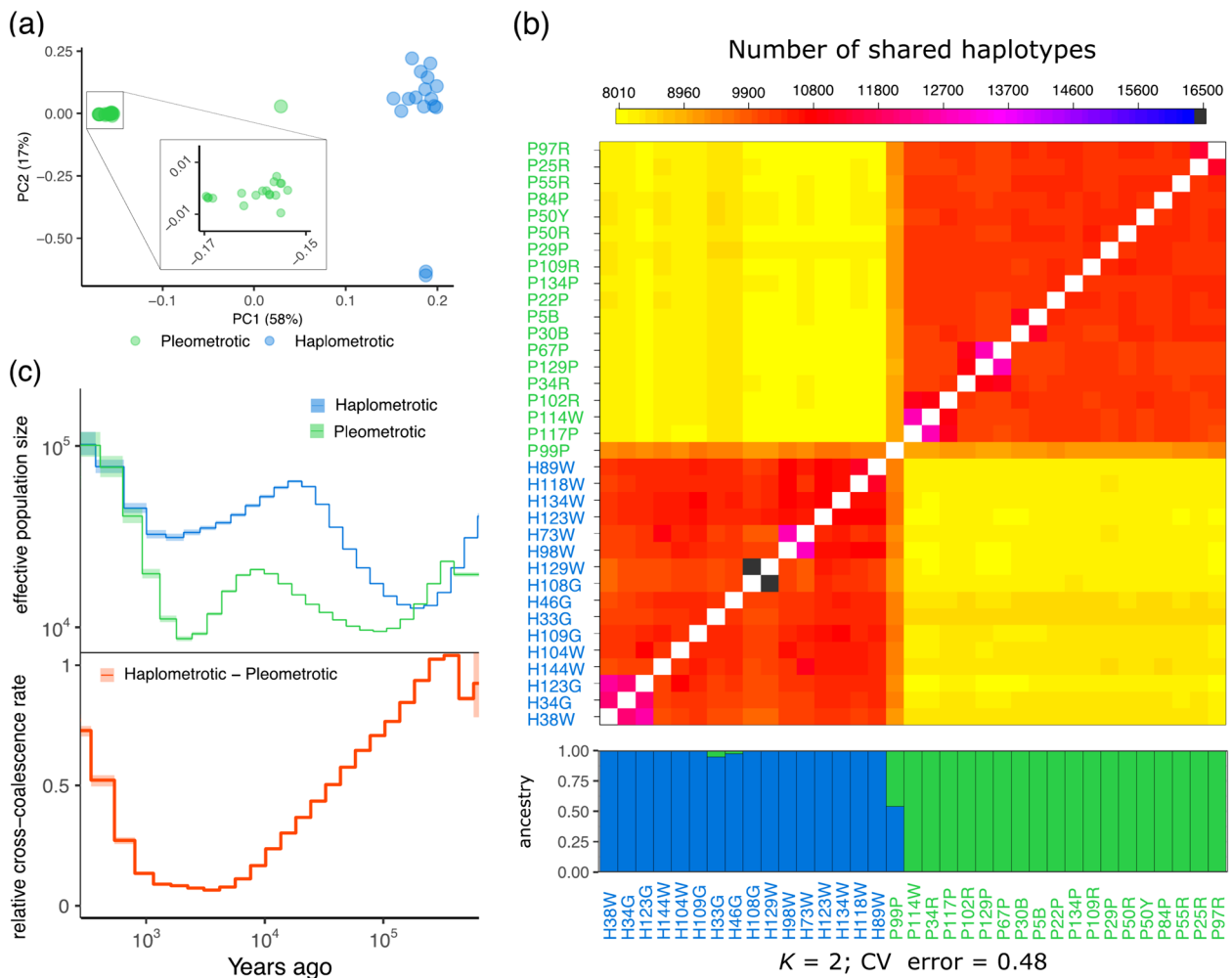


Fig. 1 Population structure and demographic history in *P. californicus* populations. **a** Principal component analysis based on genome-wide SNP data of queens of the pleometrotic and haplometrotic populations (a scatterplot matrix showing four principal components can be found in Additional file 2: Fig. 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 Additional file 2: Fig. 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

ongoing admixture between the two populations, we found significant between-population haplotype-sharing in some individuals (e.g., P99P, H33G, and H46G) corroborated by ADMIXTURE analyses (Fig. 1b, Additional file 2: Fig. S2).

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

We used MSMC2 [48] to reconstruct the demographic history of the P- and H-populations (Fig. 1c). Both populations showed a similar trajectory with a zigzag-like pattern in effective population size (N_e) between

approximately 500,000 and 400 years ago, indicative of two separate bottlenecks: an ancient one about 400,000 until about 150,000 years ago and a recent one between 15,000 to roughly 2000 years ago. Following the first bottleneck, N_e consistently remained smaller in the pleometrotic population until between 2000 and 400 years ago, when N_e started to increase again in both populations, reaching similarly high estimates in recent times.

We also inferred how these populations separated over time by estimating relative cross-coalescence rate (rCCR) [49] and found that the two populations started diverging during the first bottleneck at ~300,000 years

ago (Fig. 1c). Around 5000 years ago, both populations were almost completely isolated ($rCCR \approx 0$). However, until 400 years ago, $rCCR$ increased rapidly, suggesting a growing genetic exchange, coinciding with population expansions in both populations. This predicted recent gene flow is corroborated by evidence of recent genetic admixture (Fig. 1a,b).

Genetic diversity is significantly lower in the pleometrotic population

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

The social polymorphism in *P. californicus* shows characteristics of a polygenic trait involving a supergene

Genome scans uncovered two scaffolds (22 and 23) that showed a very distinct pattern from the rest of the genome. Both scaffolds had high levels of genetic differentiation between the two social morphs, reduced

nucleotide diversity, negative Tajima's D , as well as very high cross-population extended haplotype homozygosity ($xpEHH$) scores in the pleometrotic population (Additional file 2: Fig. S4). Further analyses revealed that SNPs were in high linkage disequilibrium (LD), with little to no LD decay in these scaffolds (Additional file 2: Fig. S5).

To determine if these scaffolds belong to a non-recombining region similar to the supergene associated with social polymorphisms in *Solenopsis invicta* and *Formica selysi* [15, 16], we used double-digest restriction-site associated DNA (ddRAD) sequencing of 108 haploid males from a single monogynous queen belonging to the H-population. While the initial genotype of the selected queen for ddRAD was unknown, we anticipated her to be heterozygous at these non-recombining scaffolds, based on our examination of the genotypes of haplometrotic queens (Additional file 2: Fig. S6). Linkage analysis using 2980 ddRAD markers with no missing data put all markers in 29 linkage groups (data not shown). The non-recombining region was found in the linkage group (LG) 14 (339 ddRAD markers, 255 cM over 26 Mb; Fig. 3a) and included scaffolds 22 and 23 as well as three other small scaffolds. This ~8-Mb region contains 69 mapped markers in complete linkage (Additional file 2: Fig. S7). Hence, this is a non-recombining region co-segregating with the P- and H-populations, spanning across five complete physical scaffolds (22, 23, 38, 42, 48), and parts of scaffolds 6 and 14 (Additional file 2: Table S1). Visual inspection of 182 markers from both males and queens showed that distinct haplotypes extended along the entire region, indicating absence of recombination (Fig. 3a) across at least ~8 Mb, consistent with patterns described for the supergenes in *Solenopsis* and *Formica* ants of non-recombining regions of 12.7 and 11 Mb, respectively [15, 17].

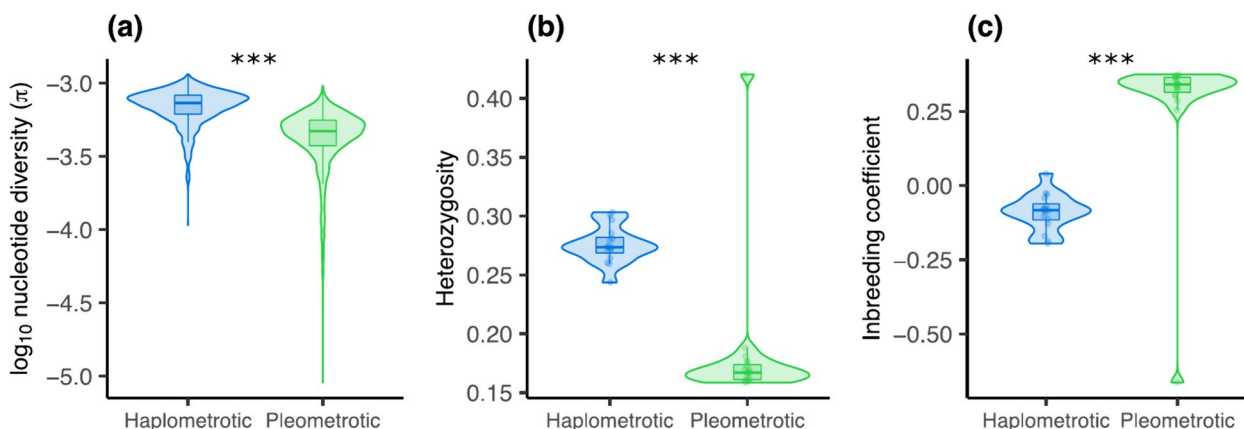


Fig. 2 Higher Inbreeding and lower heterozygosity in the pleometrotic population. **a** Nucleotide diversity (π), **b** heterozygosity, and **c** inbreeding coefficient in the haplometrotic (blue) and pleometrotic (green) populations. Heterozygosity and inbreeding coefficient were calculated on a per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

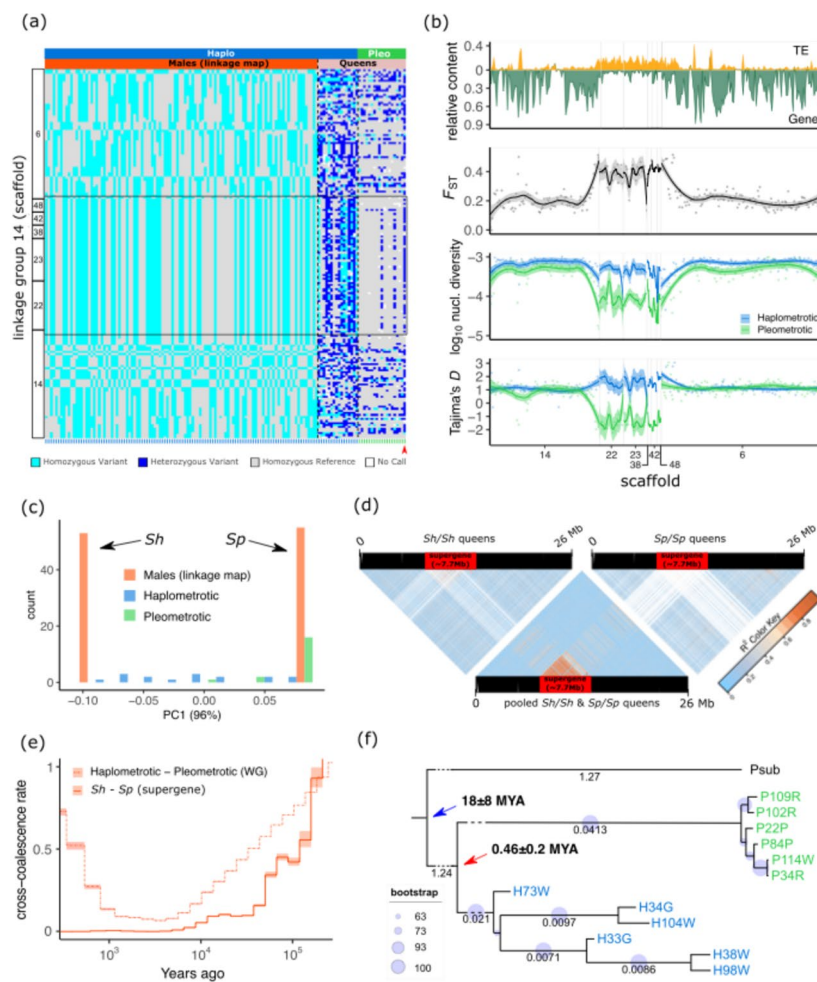


Fig. 3 Genomic architecture and recombination rates at the linkage group (LG) 14 harboring a putative supergene. **a** Genotypes for 108 sons from a single monogynous queen (haploretrotic population) based on ddRAD markers and for the 35 queen samples from the pleometrotic and haploretrotic populations based on sites corresponding to ddRAD 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 haploretrotic) represent individual samples grouped by population (pleometrotic and haploretrotic) and dataset (males and queens). The black rectangle highlights the non-recombining putative supergene region, spanning scaffolds 22, 23, 38, 42, and 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 (F_{ST}) between populations, nucleotide diversity (π), and Tajima's D estimates in the pleometrotic (green) and haploretrotic (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 haploretrotic queens, blue bars). **d** Linkage disequilibrium (LD) estimates (r^2) across LG 14 for *Sh* homozygous queens ($n=6$), pooled queens homozygous for 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 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 Additional file 7). The blue arrow indicates the upper bond estimate for the speciation of *P. subnitidus* and *P. californicus* 18 (95% highest probability density (HPD) ± 8) million years ago (MYA). The red arrow indicates the age estimate for the formation of the *Sh*-*Sp* haplotype groups 0.46 (95% HPD ± 0.2) MYA

Dense population genomic data using 5363 SNPs genotyped across all 35 queens further confirmed this finding (Additional file 2: Fig. S8).

Similar to the pattern described for the supergenes in *Solenopsis* and *Formica*, transposable element (TE) content in the *P. californicus* non-recombining region was significantly increased (Wilcoxon test: $W=14448$, $p<2.2e-16$), whereas exon content was significantly reduced compared to the rest of the same linkage group (Fig. 3b; Additional file 2: Fig. S9a; Wilcoxon test: $W=1602.5$, $p<2.2e-16$). This is consistent with the prediction that TEs accumulate in regions of low recombination [91, 92]. Further, genetic differentiation as measured by genetic differentiation (F_{ST}) between the H- and P-population was significantly increased in the supergene compared to the rest of the linkage group (Fig. 3b; Additional file 2: Fig. S9b; Wilcoxon test: $W=12261$, $p<2.2e-16$). Nucleotide diversity and Tajima's D in this region were significantly reduced in the pleometrotic population (Fig. 3b; Additional file 2: Fig. S9c,d; Kruskal–Wallis rank sum test $\chi^2 = 254.77$, $df=3$, $p<2.2e-16$; Kruskal–Wallis rank sum test $\chi^2 = 177.81$, $df=3$, $p<2.2e-16$; results of pairwise Wilcoxon rank sum post hoc tests are displayed in the figure). Despite sharing comparable genome architectures, we found no homology between the putative supergene in *P. californicus* and that in *S. invicta* (Additional file 2: Fig. S10), consistent with the convergent genetic architecture underlying social organization in ants [16].

Gene Ontology analysis of the 682 genes annotated within the supergene revealed significant enrichments in functions associated with zinc and DNA binding, as well as metalloendopeptidase activity (Additional file 2: Table S2), which are involved in transcriptional regulation. This suggests potential regulatory roles that could influence the expression of social phenotypes [93], consistent with the role transcriptional regulation has played in the evolution of sociality [94]. Furthermore, among the 156 expressed transcripts from a previously published RNA-seq dataset [77] mapped to the supergene (Additional file 3), 13 (8.3%) were reported to be significantly associated with aggression and the social context of founding queens (no significant enrichment, Additional file 3).

PCA on 69 markers along this region showed that PC1, which explains 96% of the variation, separates the male dataset of our mapping population into two distinct clusters, which we labeled as the Sh and Sp haplotype groups, indicating that these males are indeed the progeny of a heterozygous Sh/Sp mother queen (Fig. 3c; Additional file 2: Fig. S11). Superimposing SNP data from H- and P-queens to the PCA revealed that all but two pleometrotic queens (besides the hybrid queen) were

homozygous for the Sp allele, while most haplometrotic queens were either heterozygous (Sh/Sp) or homozygous for the Sh allele (Fig. 3c; Additional file 2: Fig. S11). Further support for the suppression of recombination between Sh and Sp alleles came from LD analyses across queens. The region showed stronger LD in a mixed pool of Sh/Sh and Sp/Sp queens, as well as heterozygous Sh/Sp, in comparison to separate pools of Sh/Sh or Sp/Sp queens (Fig. 3d; Additional file 2: Fig. S12), indicating recombination in homozygous but not in heterozygous queens. The observed LD in homozygous Sh and Sp queens may be attributed to genomic rearrangements similar to *Formica* [16]. Suppression of recombination in heterozygotes is a key feature of supergenes [9, 15, 19, 95]. Taken together, these data indicate the presence of an ~8-Mb non-recombining supergene with two major haplotypes (Sh and Sp) segregating with the two populations of *P. californicus*. Unlike other social supergene systems (e.g., in *Solenopsis*) [7], there were no deviations from Hardy–Weinberg equilibrium (HWE) regarding the frequency of Sh/Sh, Sh/Sp, and Sp/Sp genotypes at the supergene locus in either population (pleometrotic: χ^2 ($df=1$, $N=19$) = 0.1396, $p=0.713$; haplometrotic: χ^2 ($df=1$, $N=16$) = 0.0711, $p=0.789$) (Additional file 2: Table S3).

Coalescence modeling revealed that the demographic histories of the supergene and the rest of the genome only diverged within the last 5000 years. Then, cross-coalescence quickly increased across the genome, consistent with a secondary onset of gene flow between H- and P-populations, while remaining nearly 0 in the supergene (Fig. 3e). These findings suggest that Sh and Sp haplotypes initially diverged during the isolation period of the H- and P-population between approximately 200,000 and 5,000 years ago. During this period, one or multiple mutations suppressing recombination between Sh and Sp likely occurred (e.g., inversions), rendering the two haplotypes isolated, despite the admixture between the two populations. Consistent with a very recent origin of the supergene, phylogenetic analysis of Sh and Sp alleles with *P. subnitidus* as outgroup showed that Sh and Sp diverged less than 0.46 (95% highest probability density (HPD) 0.26–0.66) million years ago (MYA) (Fig. 3f).

In addition to the supergene region, our genome scans revealed several regions under divergent selection between H- and P-populations. We first identified highly differentiated (outlier) regions by estimating F_{ST} in 100-kb non-overlapping windows across the genome (Additional file 2: Fig. S4 and S13) and subsequently compared estimates of Tajima's D and nucleotide diversity in the two populations to identify signatures of selective sweeps (Additional file 2: Fig. S4). After clustering adjacent windows, we retained 51 clusters with negative Tajima's D

and reduced nucleotide diversity consistent with signatures of selective sweeps, of which 49 showed signatures of positive selection in the P-population and two in the H-population (Additional file 4). Cross-population *xpEHH* tests [56] (Additional file 2: Fig. S4) identified 340 outlier SNPs in 32 genomic locations under selection (Additional file 5). Overlapping both analyses resulted in 17 remaining candidate regions under selection in the P-population (Additional file 2: Table S4). Exonic SNPs in all 17 candidate regions included 70 synonymous changes and 73 non-synonymous variants affecting the protein sequence (Table 1).

Twelve of these 17 regions (on scaffolds 14, 22, and 23) were in the supergene, while the remaining 5 regions with clear signatures of selective sweeps in the P-population were clustered on scaffolds 1, 9, 15, and 16 (Additional file 2: Fig. S4). These regions could initially not be placed into a linkage group due to the lack of ddRAD markers. To exclude regions potentially in linkage with the supergene, we constructed a relaxed linkage map with 11,986 (instead of 2980) leniently filtered ddRAD markers (minor allele frequency (MAF) ≥ 0.4 and 30% missingness). In this map, one of the candidate regions (scaffold 9) was contained in the supergene. The four remaining regions on scaffolds 1, 15, and 16 (Fig. 4a) were placed

into three different linkage groups, indicating that these regions are independent of the supergene.

Among the 48 genes in these 4 remaining regions, 34 are likely functional as they are not encoding TE proteins, are expressed, or homologous to other known insect proteins (Additional file 6). Six of these 34 genes are particularly promising candidates to sustain the social niche polymorphism, as they carry non-synonymous variants with a nearly perfect association of phenotype and genotype, in that $>90\%$ of haplotrotic individuals carry the alternative and $>90\%$ of pleometrotic individuals carry the reference allele (Fig. 4b). Further, regions harboring the six genes exhibit site-specific extended haplotype homozygosity (*EHHS*) scores revealing a pattern of increased extended haplotype homozygosity in the pleometrotic population (Additional file 2: Fig. S14), suggesting selective sweeps.

Visual inspection of variant positions showed that the most common pattern for all six genes is that pleometrotic queens are fixed for the reference allele (except for the hybrid P99P queen), while haplotrotic queens are either homozygous for alternate alleles or heterozygous (Additional file 2: Fig. S15a-f). This is consistent with a selective sweep fixing the reference allele in the P-population.

Apart from one uncharacterized protein (PCAL_006795) without functional annotation, the candidates encode evolutionarily well-conserved proteins with functionally characterized orthologs in *D. melanogaster*. PCAL_006788 is orthologous to *GlnRS* (FBgn0027090), which encodes a tRNA synthetase involved in neurogenesis [96]. PCAL_006786 encodes an RNA polymerase III, orthologous to the transcriptional regulator Polr3A (FBgn0030687). PCAL_007700 encodes a flavin-containing monooxygenase homologous to *Drosophila* Fmo-1 (FBgn0034943) and Fmo-2 (FBgn0033079), which are suspected to be involved in xenobiotic detoxification and aging [97]. The two remaining genes PCAL_006796 and PCAL_007693 particularly caught our attention, as they are orthologs of genes involved in epigenetic regulation of transcriptional activity. PCAL_006796 encodes chameau (ortholog to FBgn0028387), a MYST histone acetyl transferase

Table 1 Effects of single-nucleotide polymorphisms (SNPs) found in candidate regions that show strong signatures of selection in the pleometrotic population. We annotated 6077 putative effects on protein coding genes of 3347 SNPs

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
Non-synonymous variant	73	1.20
Synonymous variant	70	1.15
Upstream	1319	21.70
Other	7	0.12

(See figure on next page.)

Fig. 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 (F_{ST}) between populations, 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 population are highlighted in red in all panels. These regions are highly differentiated, have reduced diversity and Tajima's D in the pleometrotic population, and show a significant *xpEHH* peak. Lines and shaded areas are means and 95% confidence intervals, respectively. Data for other scaffolds (1–25) are shown in Additional file 2: Fig. S4. **b** Lollipop plots showing all variants affecting each of the six candidate genes (gene models in purple). The pie charts display frequencies of the reference (grey) and alternate (cyan) alleles in the pleometrotic (top; green background) and haplotrotic population (bottom; blue background)

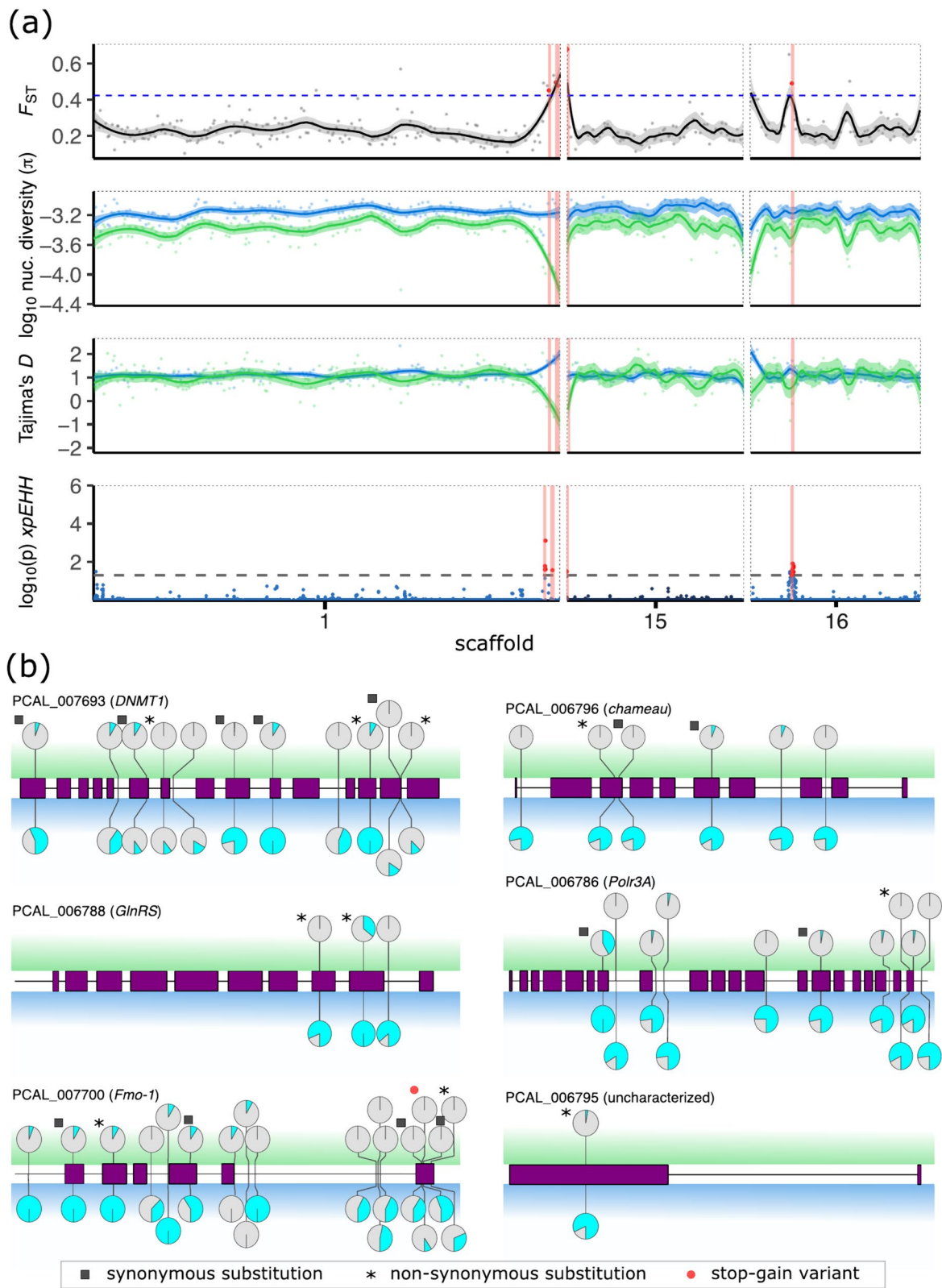


Fig. 4 (See legend on previous page.)

involved in transcriptional regulation via histone modification [98]. PCAL_007693 encodes Dnmt1, a gene that is absent in Diptera. Dnmt1's primary function is to maintain methylation patterns, and in social insects, inter-individual differences in DNA methylation have been linked to the epigenetic regulation of phenotypic plasticity, including caste determination in bees [99].

Discussion

The social niche polymorphism [5] in colony founding of *P. californicus* was first reported over 20 years ago [28]. This polymorphism arises from differences in the social interaction among founding queens, which will eventually lead to fundamentally different colony structures, where either unrelated queens cooperate (i.e., primary polygyny) or a single queen monopolizes reproduction (i.e., primary monogyny) [11]. The principal change underlying the emergence of pleometrosis in *P. californicus* is hence the evolution of a lifelong tolerance for unrelated co-founding queens. Given that the majority of North American *Pogonomyrmex sensu stricto* are primarily monogynous, pleometrosis is likely a derived trait, adaptive under resource limitation and increased territorial conflicts [29]. Since its original description, we have learned a lot about behavioral, physiological, ecological, and socio-genetic correlates of haplo- and pleometrosis [11, 26, 27, 29, 32, 33, 100]. However, the genomic architecture and evolutionary history was unknown.

The tight association of social morph and geographic location in *P. californicus* poses a unique challenge when exploring the genetic basis of this social polymorphism. As haplo- and pleometrotic individuals often occur in allopatry, any signature of genetic differentiation between both social morphs could be a product of genetic drift within these populations. Alternatively, rather than solely regulating social morph, the supergene and/or the modifier loci may be associated with adaptations to diverse ecological conditions between the allopatric habitats [29]. Ultimately resolving these uncertainties will be a challenge, as the two social morphs likely themselves represent adaptations to distinct ecological conditions [29]. Experimentally addressing this would require disentangling social morph from other putative local adaptations.

It is therefore important to emphasize that while our findings are consistent with a young supergene underlying the social polymorphism in this species, our results are a characterization of genetic differentiation between two populations that are almost fixed for distinct behavioral phenotypes. Addressing this geographic confounder will require studying populations carrying both haplo- and pleometrotic individuals at significant proportions. However, despite our best

efforts during several field seasons in the last years, we have not yet been able to find suitable populations. Thus, our study for now provides initial insight into the genetic basis of social polymorphism in *P. californicus* under these limitations.

We identify an approximately 8-Mb non-recombining region, reminiscent of social supergenes described for social polymorphisms in *Solenopsis* fire ants [15] and *Formica* wood ants [16]. However, we also find genomic, population genetic, and ecological characteristics unique to the *Pogonomyrmex* system, highlighting the independent evolutionary histories of the social niche polymorphism in *Pogonomyrmex*, *Formica*, and *Solenopsis*. The 8-Mb non-recombining region we identify is similar in size to the supergenes in *Formica* and *Solenopsis* (11 and 12.7 Mb) and shows no homology to social chromosomes described in other ant species consistent with the divergence in genetic architecture underlying social organization in ants [16]. According to estimates of F_{ST} , genetic divergence between individuals from the two social forms is significantly higher within this region compared to other parts of the genome, which parallels the patterns reported for the other supergenes [74]. Tajima's D and haplotype homozygosity statistics suggest a selective sweep of the Sp haplotype, although it could also be the result of a population bottleneck, similar to the *Solenopsis* Sb haplotype [101]. Very low LD decay and absence of recombinant F1 offspring of a heterozygous queen (Sh/Sp) suggest little to no recombination between Sh and Sp haplotypes at this locus. Further, TEs are significantly enriched at the supergene, as expected in the absence of recombination following a Muller's ratchet dynamic [102].

Several other characteristics of the *P. californicus* social polymorphism differ from the well-studied examples in *Solenopsis* and *Formica* as well. In both, *S. invicta* and *F. selysi*, there is little evidence for genetic differentiation of polygyne and monogyne populations, with both forms often occurring in sympatry [103, 104]. In *P. californicus*, however, populations are fixed for either the haplo- or the pleometrotic strategy [11], with little gene flow between them. Accordingly, genetic differentiation between social forms in *P. californicus* is higher than in *S. invicta* and *F. selysi* [103, 104], but it is similar to that observed in other *Formica* species, such as *F. exsecta* or *F. truncorum* [17]. The onset of gene flow observed between the haplo- and pleometrotic population starting ~5000 years ago could secondarily reduce genetic differentiation between both forms everywhere but in the non-recombining supergene, eventually resulting in the pattern also observed in *S. invicta* and *F. selysi*, where genetic differentiation between both social morphs is restricted to the supergene [16, 74].

We did not find significant deviations from HWE at the non-recombining region in either population, suggesting the absence of lethal haplotype combinations. This differs from *Formica* and *Solenopsis* where intrinsic (such as recessive lethal effects in *Solenopsis* and maternal killing effects in *Formica*) and extrinsic (such as green beard effects in *Solenopsis*) lethal combinations exist and are presumed crucial for the maintenance of the social niche polymorphism [105, 106]. Further, unlike in *Solenopsis* and *Formica*, we find additional genomic regions that segregate with the social polymorphism in *P. californicus*. Theoretical models for supergene evolution predict gradual establishment of supergenes by integration of formerly unlinked modifier loci [95, 107]. According to these models, natural selection will favor genomic rearrangements that eventually join the supergene with previously unlinked but co-evolved loci carried on the same chromosome [108, 109]. Hence, an alternative fate for the unlinked loci we describe here could be that they in fact persist as (e.g., population-specific) “modifier loci,” similar to the genetic architecture for refined Batesian mimicry described in *Papilio dardanus* [110–112] and *Heliconius numata* [113]. The presence of unlinked modifier loci influencing the expression and ameliorating negative fitness effects of major quantitative trait loci is not unusual [114], yet none have been described so far for the known social supergenes.

The association between the genotype at the non-recombining region and the social form is imperfect in *P. californicus*. Pleometrotic queens were predominantly homozygous (85%) for a single haplotype (Sp), while most haplometrotic queens (93%) carried at least one copy of the Sh haplotype. While the precise effects of this locus on social polymorphism require further validation, this highlights that the genotype at this locus alone is insufficient to determine social form—a fundamental difference to the supergenes described in other ants [17, 115]! The relatively young age of the suspected *Pogonomyrmex* supergene (*Pogonomyrmex*: ~0.2 MYA, *Solenopsis*: 0.39–1.1 MYA [9, 15, 101], and *Formica*: 20–40 M years [17]) may contribute to this disparity. Alternatively, the regulation of social morphs in this species may involve a polygenic architecture, where the expression of haplometrosis or pleometrosis is influenced by social context, as well as environmental and ecological conditions [29]. This aligns with previous studies showing that environmental conditions and interactions with co-founding queens can modulate the expression of aggression and tolerance in different populations [32, 33].

The four differentiated genomic regions outside the non-recombining region contain 34 genes that have undergone a selective sweep and potentially contribute to haplometrosis or pleometrosis. The nature of

their interaction with the presumed supergene, whether additive or epistatic, remains unclear. Notably, the DNA methyltransferase *Dnmt1* and the histone acetyltransferase *chameau* exhibit a near-perfect correlation between genotype and phenotype, with non-synonymous variants showing a distinct pattern; most haplometrotic individuals carry the alternative allele, while more than 90% of pleometrotic individuals possess the reference allele. As pleometrotic queens show a higher behavioral [32, 33] and physiological [77] plasticity than haplometrotic queens during colony founding, differential epigenetic regulation of plastic traits like division of labor could maintain the two distinct social morphs. While these candidates suggest that epigenetic regulation modulates the social niche polymorphism in *P. californicus*, in-depth functional studies will be necessary to explore how *chameau* or *Dnmt1* are indeed involved in the expression of haplo- and pleometrosis.

Conclusions

Pleometrosis and primary polygyny in *P. californicus* likely evolved as adaptation to challenging environments [29]. The genetic architecture associated with this social niche polymorphism shows not only similarities but also significant differences compared to those described in, e.g., *Formica* and *Solenopsis*. *Pogonomyrmex* also harbors a suspected non-recombining supergene, which is similar in size and has two major segregating haplotypes, as in the other systems. Additionally, four genomic regions showing signatures of a selective sweep in the pleometrotic population, potentially influence the expression of haplo- or pleometrosis. Further investigations are needed to fully understand the relevance and interactions of these loci. Unless both social morphs can be studied in sympatry, it will however remain challenging to conclusively resolve whether the suspected supergene and/or other genetic differences are indeed drivers of social polymorphism in this species or whether they have been shaped by demographic effects. Moreover, although climatic differences between the two study sites are small [29], we cannot dismiss the possibility that the observed genetic differences may represent local adaptations to some unknown ecological factor. Regardless, our study provides a first assessment of the genetic differentiation between haplo- and pleometrotic populations of *P. californicus*. It also showcases that the challenges with resolving the genetic basis for social niche polymorphisms, which may often be much more complex and diverse than previously appreciated.

Methods

Further details about the Methods are provided in the Additional file 2.

Ant collection

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

Genome sequencing

DNA libraries (NEBNext[®] Ultra[™] II FS DNA Libraries) of 16 queens from the H-population and 19 from the P-population were paired-end sequenced (2×75 bp, average coverage 12×) at the Genomic Core Facility of the University of Münster (Germany), using a NextSeq 500 (Additional file 2: Table S5).

Genome assembly and annotation

We sequenced and assembled the genome of *P. californicus* (“Pcal.v2”) using data from MinION long reads (MinION, Oxford Nanopore Technologies, Oxford, UK), Illumina short reads, and previously published 10×Chromium linked reads [34]. For long read library preparation (Oxford Nanopore kit SQK-LSK109), we used 2 µg of DNA extracted from three white worker pupae from a pleometrotic colony (“Pcal18-02”). The library was sequenced for 48 h on a MinION, generating 1,650,688 reads that were trimmed and filtered with PoreChop (v.0.2.4) and FiltLong (v.0.2.0). We assembled the genome from ~30×filtered Nanopore data using the following strategy. We created one assembly using CANU (v.2.0) [47] and a second assembly using wtdbg2 (v.2.5). Both assemblies were processed and optimized as follows: after contig assembly, we used SSPACE (v.3.0) [58] for scaffolding, LR_closer for gap filling, and ntedit for polishing with long-read data. We then used short-read, short-insert Illumina data from three samples (P50Y, P67P, P55R) for scaffolding with SSPACE before three rounds of polishing with Pilon [116]. The polished assembly was super-scaffolded with publicly available 10×Chromium data for *P. californicus* (accession SRX8051306 [34]) using scaff10× and subsequent correction with break10×. Finally, we combined both assemblies with ntjoin, using the CANU-based assembly as reference and removed redundancies with funannotate clean (v.1.7.1).

We combined de-novo libraries from RepeatModeler2 (v.2.0.1) and EDTA (v.1.8.3) with public resources from RepBase (RepBase25.04.invrep.fa) and Dfam (Dfam3.0.arthropod.fa) for repeat annotation with RepeatMasker (v.4.0.7). Protein-coding and tRNA genes were annotated with funannotate v.1.7.1 in the soft-masked genome, including homology-based gene predictions from GeMoMa (v.1.7) [117] and supported by RNAseq data of 18 heads of *P. californicus* (Ernst

et al. unpublished) and previously published long-read RNAseq data of a pool of workers [34].

Read mapping and variant calling

Short-read data was cleaned with Trimmomatic (v.0.38) [88]. Paired reads were mapped to the genome assembly using BWA-MEM (v.0.7.17-r1188) [89]. After marking duplicates using Picard’s MarkDuplicates (v.2.20.0), we performed joint-variant calling across all samples with the GATK [37] and filtered variant calls with VCFtools (v.0.1.16) [39]. After a visual inspection of mapping quality of reads across the genome (Additional file 2: Fig. S16), we excluded fragmented scaffolds with low mapping quality and considered 613,411 SNPs identified in the 25 largest scaffolds (representing 88.2% of the assembly) for further analyses.

Statistical phasing

We used SHAPEIT (v.2.r904) [40] to statistically phase the 25 largest scaffolds of the *P. californicus* genome based on 210,013 SNPs with no missing data (i.e., genotyped in all 35 queens), as no reference panel is available for *P. californicus*. Scaffolds 22 and 23 displayed an atypical architecture potentially representing a supergene (discussed further below) and were thus excluded (along with three other small scaffolds) for fineSTRUCTURE and MSMC2 analyses.

Population structure analyses

LD-pruning with PLINK yielded 314,756 SNPs that we used for PCA, considering four eigenvectors. LD-pruned SNPs were further used for ADMIXTURE (v.1.3.0) analyses [43] with 2 to 4 clusters. To better characterize the relationships among queens in our sample, we used the haplotype-based approach implemented in fineSTRUCTURE [44]. In our model, we assumed a uniform recombination rate, based on the estimate of 14 cM/Mb found in *Pogonomyrmex rugosus* [45].

Demographic population history analysis

We inferred demographic population history and the rCCR using Multiple Sequential Markovian Coalescence modeling (MSMC2) [49] on four queens (i.e., eight phased haplotypes) from the haplometrotic (H123G, H118W, H73W, and H123W) and pleometrotic populations (P22P, P134P, P117P, and P109R). Mappability masks were generated using SNPABLE (<http://lh3lh3.users.sourceforge.net/snpable.shtml>). We ran 1000 bootstrap replicates, each time randomly sampling 60 Mb for each of the analyzed genomes. To obtain the time in years, we assumed a generation time of 4 years and a similar mutation rate as in honey bees [50], i.e., 3.6×10^{-9} mutations × nucleotide⁻¹ × generation⁻¹.

Population genomic estimates

We used VCFtools (v.0.1.16) [39] to estimate π (*--window-pi*) and Tajima's D (*--TajimaD*) within each population using 100-kb non-overlapping windows across the genome. VCFtools' *--het* function was employed to calculate heterozygosity and inbreeding coefficient on a per individual basis. LD decay was estimated by calculating pairwise r^2 values between all markers within 200-kb windows using VCFtools' *--hap-r2* function in each of the two populations. We used SNP markers with a MAF of 0.2 to minimize the effects of rare variants and averaged r^2 values between all pairs of markers in 100 bp distance bins for visualization.

Genome scan for selection

We screened genomes of both populations for evidence of selection using an F_{ST} outlier approach [51–54] and by identifying regions of extended haplotype homozygosity (selective sweeps) [56, 57], comparing the haplotypic and pleiotropic populations. We excluded a total of six queens from pairs with relatedness >0.4 and a potential F1 hybrid (P99P) from the analyses. We used mean F_{ST} , nucleotide diversity π , and Tajima's D estimated in 100-kb non-overlapping windows across the genome and from the empirical distribution of F_{ST} values; we considered the top 5% of windows as differentiated.

F_{ST} estimates can be affected due to variation in recombination rate across the genome [55]. Hence, we further applied the cross-population *xpEHH* and the site-specific *EHHs* tests [56, 57], to detect signatures of selective sweeps. The *xpEHH* analyses were performed on the phased data from unrelated queens using the R package *rehh* (v.3.1.2) [59]. We used the Benjamini and Hochberg method for FDR correction and set a p -value of 0.05 as a threshold to identify outlier SNPs. Outliers occurring within 100 kb of each other were grouped, resulting in 32 clusters distributed across multiple scaffolds (Additional file 5). Regions within ± 100 kb around the most significant SNP of each cluster were regarded as candidates for positive selective sweeps. Regions identified by F_{ST} and by *xpEHH* overlapped at 17 genomic loci (Additional file 2: Table S4).

Characterization of candidate genes and SNPs

For functional characterization of genes, we performed BLAST (blastx) [63] searches against the nr database (January 2020), used OrthoFinder (v.2.3.1) [64] to identify *Drosophila melanogaster* orthologs, and used Repeat-ProteinMask to identify genes encoding TE proteins. We used SnpEff (v.5.0) [62] to annotate putative effects on protein-coding genes of all 3347 SNPs contained in the candidate regions (Table 1). For the manual review of genes contained in genomic regions showing evidence

for a selective sweep, we removed TE-encoded genes as well as genes without RNA-seq support and no known homology to any proteins in the nr database, resulting in 34 putatively functional genes. Of these, 18 genes showed non-synonymous variants in our dataset (Additional file 6). In 6 of these 18 genes, at least 1 non-synonymous variant showed a pattern where over 90% of the individuals from the H-population carried the alternative allele (i.e., either 0/1 or 1/1) and over 90% of the individuals from the P-population carried the reference allele (i.e., either 0/1 or 0/0).

Haplotype frequencies

We calculated haplotype frequencies for the putative supergene region and tested for potential departure from Hardy–Weinberg equilibrium using a χ^2 test (Additional file 2: Table S3). Haplotypes were assigned manually, based on visual inspection of the suggested supergene region (Additional file 2: Fig. S8); predominantly homozygous for the reference alleles (i.e., from a pleiotropic population) and alternate allele were assigned “Sp/Sp” and “Sh/Sh”, respectively, and predominantly heterozygous regions were assigned “Sh/Sp”.

RAD sequencing and analysis

DNA of males collected from a monogynous colony at Lake Henshaw (H-population) in 2019 was used to construct reduced representation genomic libraries following Inbar et al. [65], using a modified double-digest (EcoRI and MseI) restriction-site associated DNA (ddRAD) sequencing protocol [66, 67]. Through visual inspection of the genotypes of haplotypic queens (Additional file 2: Fig. S6), we anticipated that the males were progeny of a queen heterozygous at the non-recombining region. Genomic DNA was digested at 37 °C for 8 h and ligated with uniquely barcoded adaptors. The adaptor ligated products were PCR amplified, pooled, and size-selected using Beckman Coulter, Agencourt AMPure XP beads. Libraries were sequenced in one lane of a paired-end 150 bp reads on a HiSeq X Ten Illumina sequencer.

We used fastQC (v.0.11.7), Trimmomatic (v.0.39) [88], and *process_radtags* (STACKS v.2.2) [68] for quality control, filtering, and demultiplexing the raw data. Reads were mapped with BWA-MEM (v.0.7.17-r1188) [89] with default parameters and alignments processed with *ref_map.pl* and *populations* (STACKS v.2.2), resulting in 55,340 raw variants. After filtering with VCFtools (v.0.1.16) [39] and excluding heterozygous markers, we arrived at a final set of 2980 markers used for building a genetic map.

Markers were mapped with MultiPoint (<http://www.multiqtl.com>) using the backcross population setting. A maximum threshold recombination frequency of 0.25

resulted in 29 linkage groups (data not shown). A skeleton map was generated using delegate markers and Kosambi's mapping function was used to convert recombination frequencies to centimorgans [72].

Characterization of the supergene

To investigate the potential supergene identified in the current study, we used SNP markers from queens (whole-genome resequencing) and males (ddRAD sequencing). We used BCFtools' (v.1.9) *isec* and *merge* functions to extract 182 shared SNPs between the two datasets found on LG14. Genotypes of the males and queens at LG14 were visualized using VIVA (v.0.4.0) [73]. To further characterize the two alleles of the supergene, we performed a PCA on 69 SNP markers identified in the non-recombining area of LG14 (i.e., scaffolds 22, 23, 38, 42, and 48, and parts of 6 and 14) in datasets of males and queens.

To investigate recombination suppression in the supergene region, we used SNP data from six queens homozygous for the Sh allele (H104W, H33G, H34G, H38W, H73W, and H98W) and six others homozygous for the Sp allele (P22P, P102R, P114W, P109R, P34R, P84P), assigned based on visual inspection of their genotypes (Additional file 2: Fig. S8). After VCF lift over (on scaffolds 14, 22, 23, 38, 42, 48, and 6 into LG14) using *flo* [74] and *picard's LiftoverVcf* (v.2.20.0), we used VCFtools (v.0.1.16) [39] for filtering and LDheatmap (v.1.0–4) [75] to visualize pairwise linkage disequilibrium (LD). We calculated gene and TE content in 100-kb sliding windows across LG14 using our *P. californicus* gene and TE annotations, respectively. To explore whether any of the transcripts previously shown to be associated with the aggressive behavior of queens [77] belong to the supergene, we aligned the assembled transcripts to the Pcal.v2 reference genomes using GMAP [78]. We only retained the top hits with at least 70% query coverage, yielding 7303 (out of 7890) transcripts unambiguously aligned to the reference genome. We then used BEDtools (v.2.29.2) [61] to extract transcripts that mapped to the supergene region (Additional file 3).

To evaluate the homology between the putative supergene in *P. californicus* and that in *S. invicta*, we conducted whole-genome alignment against the *S. invicta* assembly (GCA_016802725.1) using minimap2 (2.26-r1175) [79] with the "asm20" preset parameter. We then visualized the output generated by minimap2 in R (v.4.1.3) [46] and filtered out alignments shorter than 500 bp using the R script *asynt.R* accessible at <https://github.com/simonhmartin/asynt>.

To test whether genes found in the putative supergene are enriched for specific functions, we ran a GO enrichment analysis. We defined these regions. We used

BEDTools' intersect (v.2.30.0) [61] to extract 682 genes located in the non-recombining region and ran topGO [81] in R.

Dating the supergene formation

We estimated the divergence between Sh and Sp by computing the rCCR, using MSMC2 [49] on four Sh/Sh queens (H104W, H38W, H73W, and H98W) and four Sp/Sp queens (P114W, P50R, P84P, and P97R). We limited the analysis to the supergene region (scaffolds 22 and 23 and part of scaffold 6) and ran 500 bootstraps by randomly sampling 2 Mb each time.

Further, we performed a phylogenetic analysis of Sh and Sp alleles with *P. subnitidus* as an outgroup species. Sequencing data of one *P. subnitidus* individual was generated and processed as described above for *P. californicus*. Joint SNP calling was performed using GATK [37] and a maximum-likelihood tree was constructed based on 33,424 SNPs at the supergene region (scaffolds 22, 23, 38, 42, and 48 and parts of scaffolds 6 and 14), using RAxML [83] under the GTRGAMMA model with bootstrapping.

We then used a fossil-calibrated dated phylogeny [84] to date the split between the Sh and Sp haplotype groups using a phylogenetic approach. The Ward et al. [84] phylogeny dated the split between *P. vermiculatus* and *P. imberbiculus* at 18 (95% HPD \pm 8) MYA. This is an upper bound on the speciation date of *P. californicus* and *P. subnitidus* according to the most recent *Pogonomyrmex* phylogeny [85]. We calculated a ratio of 1:39 between the speciation time of *P. subnitidus* and *P. californicus* and the Sh–Sb haplotypes divergence time. Based on our dating reference (i.e., 18 \pm 8 MYA), we obtained 0.46 (95% HPD \pm 0.20) MYA as an upper bound for the divergence between the two haplotype groups, which is consistent with the MSMC2 analysis. We note that the confidence in the age inference is strengthened by the agreement between two distinct approaches: the inference by MSMC2 is based on a coalescent model, which is a distinct approach from the phylogenetic dating approach. The former approach depends on estimates of mutation rate and generation time, whereas the latter depends on fossil calibration points. The tree figure was produced using interactive tree of life (iTOL) [86].

Abbreviations

PCA	Principal component analyses
SNPs	Single-nucleotide polymorphisms
N_e	Effective population size
rCCR	Relative cross-coalescence rate
π	Nucleotide diversity
$\chi pEHH$	Cross-population extended haplotype homozygosity
LD	Linkage disequilibrium
ddRAD	Double-digest restriction-site associated DNA
LG	Linkage group
F_{ST}	Genetic differentiation

HPD	Highest probability density
MYA	Million years ago
TE	Transposable element
HWE	Hardy–Weinberg equilibrium
MAF	Minor allele frequency
EHHS	Site-specific extended haplotype homozygosity
iTOL	Interactive tree of life

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-01907-z>.

Additional file 1. Detailed report on *Pogonomyrmex californicus* genome sequencing, assembly and annotation.

Additional file 2: Extended Methods. Further details about the Methods used in this study. **Fig. S1.** PCA based on 314,756 SNPs of pleometrotic and haplometrotic queens. **Fig. S2.** Admixture patterns among *P. californicus* populations using ADMIXTURE at $K = 3$ and $K = 4$. **Fig. S3.** Tajima's D estimates for the pleometrotic and haplometrotic populations. **Fig. S4.** Genetic differentiation, nucleotide diversity, Tajima's D and $xpEHH$ scores in the 25 largest scaffolds of the *P. californicus* genome. **Fig. S5.** Linkage disequilibrium decay in each of the 25 largest scaffolds in the two *P. californicus* populations. **Fig. S6.** Heatmap showing the genotypes (at scaffolds 22 and 23) of queens originating from pleometrotic and haplometrotic populations. **Fig. S7.** Linkage group (LG) 14 (253.38 cM). **Fig. S8.** Genomic architecture at LG 14 harboring the putative supergene in *P. californicus*. **Fig. S9.** Genetic architecture of the putative supergene in *P. californicus*. **Fig. S10.** Alignment of the putative *P. californicus* supergene to the genome of *Solenopsis invicta*. **Fig. S11.** PCA based on 69 markers spanning the supergene region. **Fig. S12.** Linkage disequilibrium estimates (r^2) across LG 14 containing the putative supergene for six heterozygous queens (Sh/Sp). **Fig. S13.** Distribution of pairwise F_{ST} between haplometrotic and pleometrotic queens. **Fig. S14.** Signatures of site-specific extended haplotype homozygosity at two outlier SNPs in the pleometrotic and haplometrotic populations. **Fig. S15.** Genotypes of haplo- and pleometrotic queens for candidate genes under selection. **Fig. S16.** Mapping quality across the 199 scaffolds of the Pcal.v2 assembly. **Fig. S17.** Quality control of 100 kb non-overlapping genomic windows used to compute population genomic metrics. **Table S1.** Genomic coordinates and sizes of regions belonging to the putative supergene on LG14. **Table S2.** Gene Ontology enrichment analyses for 682 genes annotated within the putative supergene. **Table S3.** Hardy–Weinberg calculations at the supergene locus for the pleometrotic and haplometrotic populations. **Table S4.** Genomic coordinates of regions showing patterns of selective sweeps in the pleometrotic population. **Table S5.** Quality control metrics for the sequencing data used in this study.

Additional file 3. List of all expressed transcripts found in the supergene region on LG14 that as well as those shown previously to be differentially expressed between queens of the two populations in the context of origin (pleometrotic/haplometrotic), behavior (aggressive/nonaggressive) or social context (please refer to [77] for details). Transcript comp23388_c0 showed differential expression in four distinct comparisons.

Additional file 4. Top 5% highly differentiated regions between the haplometrotic and pleometrotic populations. Genomic position, number of SNPs, F_{ST} , π , Tajima's D , mean coverage, and mean mapping quality (MQ) are displayed for each candidate region.

Additional file 5. Outlier SNPs showing clear peaks of $xpEHH$ ($p < 0.05$) using the Benjamani & Hochberg method for FDR correction. Positive values in the $xpEHH$ column indicate selection in the pleometrotic population, while negative values indicate selection in the haplometrotic population. SNPs within 100 kb of each other were grouped in the same cluster (32 clusters indicated by the cluster column).

Additional file 6. Candidate genes showing signatures of selection in the pleometrotic population identified by our two selection scan approaches (F_{ST} and $xpEHH$). 34 likely functional genes were kept as they are not encoding TE proteins, are expressed, or homologous to other known

insect proteins (see Methods). The six candidate genes depicted in Fig. 4b are highlighted in bold formatting for emphasis.

Additional file 7. Phylogenetic tree used to date the supergene, produced by RAXML using the GTRGAMMA model, in Newick format.

Acknowledgements

We thank Pine Valley Academy (Pine Valley, CA, USA) for the support and the staff of Lake Henshaw resort (Lake Henshaw, CA, USA) for the permission to collect ants on their premises. Thanks to Jenny Märzhäuser, Tobias van Elst, and Ti Eriksson for the help in collecting the ants and to Jennifer Fewell for hosting us at Arizona State University, Tempe, AZ, USA. We are grateful for Hilde Schwitte's and Una Hadziomerovic's help in DNA extraction, barcoding, and library construction. We are very grateful to the anonymous reviewers for their constructive comments.

Authors' contributions

J.G., U.E., M.E., and L.S. designed the research, U.E. organized the sample collection, A.L., L.S., M.E., U.E., E.P., and J.G. analyzed data, M.E., U.E., L.S., and J.G. wrote the manuscript, J.G. and L.S. coordinated the project, and all authors revised the manuscript. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. J.G. and U.E. were funded by the German Research Foundation (DFG) as part of the SFB TRR 212 (NC³)—TP C04 project numbers 316099922 and 396780988. M.E. was funded by the German Research Foundation (DFG) —403813881 with grant to L.S. (SCHR 1554/2–1) under the priority program "Rapid evolutionary adaptation: Potential and constraints" (SPP 1819). M.E. acknowledges support by the DFG Research Training Group 2220 "Evolutionary Processes in Adaptation and Disease".

Availability of data and materials

The genome assembly and the raw sequencing data prior to trimming and mapping are available at NCBI (BioProject: PRJNA682388 [118]). A detailed description of our bioinformatic analysis pipelines is publicly available at <https://github.com/merrbii/Pcal-SocialPolymorphism> [119].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have no conflict of interest to declare.

Author details

¹Molecular Evolution and Sociobiology Group, Institute for Evolution and Biodiversity, University of Münster, Hüfferstr. 1, Münster DE-48149, Germany. ²Department of Evolutionary and Environmental Biology, Institute of Evolution, University of Haifa, Haifa, Israel. ³Present Address: Apicultural State Institute, University of Hohenheim, Erna-Hruschka-Weg 6, Stuttgart DE-70599, Germany. ⁴Center for Biodiversity and Integrative Taxonomy (KomBioTa), University of Hohenheim, Stuttgart DE-70599, Germany.

Received: 3 October 2023 Accepted: 30 April 2024

Published online: 13 May 2024

References

- Wilson EO. The insect societies. Cambridge: Harvard University Press; 1971.
- Hölldobler B, Wilson EO. The number of queens: an important trait in ant evolution. Naturwissenschaften. 1977;64:8–15.

3. Boomsma JJ, Huszár DB, Pedersen JS. The evolution of multiqueen breeding in eusocial lineages with permanent physically differentiated castes. *Anim Behav.* 2014;92:241–52.
4. Keller L. The assessment of reproductive success of queens in ants and other social insects. *Oikos.* 1993;67:177.
5. Saltz JB, Geiger AP, Anderson R, Johnson B, Marren R. What, if anything, is a social niche? *Evol Ecol.* 2016;30:349–64.
6. Seppä P, Sundström L, Puntila P. Facultative polygyny and habitat succession in boreal ants. *Biol J Linn Soc.* 1995;56:533–51.
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;92:541–60.
8. Fletcher DJC. Three newly-discovered polygynous populations of the fire ant, *Solenopsis invicta*, and their significance. *J Georg Entomol Soc.* 1983;18:538–43.
9. Yan Z, Martin SH, Gotzek D, Arsenault SV, Duchon P, Helleu Q, et al. Evolution of a supergene that regulates a trans-specific social polymorphism. *Nat Ecol Evol.* 2020;4:240–9.
10. Helms KR, Newman NJ, Helms CS. Regional variation in queen and worker aggression in incipient colonies of the desert ant *Messor pergandei*. *Behav Ecol Sociobiol.* 2013;67:1563–73.
11. Overson R, Fewell JH, Gadau J. Distribution and origin of intraspecific social variation in the California harvester ant *Pogonomyrmex californicus*. *Insectes Soc.* 2016;63:531–41.
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.
13. Gill RJ, Arce A, Keller L, Hammond RL. Polymorphic social organization in an ant. *Proceedings Biol Sci.* 2009;276:4423–31.
14. Herbers JM. Nest site limitation and facultative polygyny in the ant *Leptothorax longispinosus*. *Behav Ecol Sociobiol.* 1986;19:115–22.
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;493:664–8.
16. Purcell J, Brelsford A, Wurm Y, Perrin N, Chapuisat M. Convergent genetic architecture underlies social organization in ants. *Curr Biol.* 2014;24:2728–32.
17. Brelsford A, Purcell J, Avril A, Van Tran P, Zhang J, Brüttsch T, et al. An ancient and eroded social supergene is widespread across formica ants. *Curr Biol.* 2020;30:304–311.e4.
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.
19. Schwander T, Libbrecht R, Keller L. Supergenes and complex phenotypes. *Curr Biol.* 2014;24:R288–94.
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. *PLoS Genet.* 2010;6:e1000862.
21. Matschiner M, Barth JMI, Tørresen OK, Star B, Baalsrud HT, Briec MSO, et al. Supergene origin and maintenance in Atlantic cod. *Nat Ecol Evol.* 2022;6:469–81.
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;367:343–53.
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;4:1831–40.
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; <https://doi.org/10.1101/2022.12.06.519286>.
25. Buckley SB. Descriptions of new species of North American Formicidae. *Proc Entomol Soc Philadelphia.* 1867;6:152–72.
26. Johnson RA. Semi-claustral colony founding in the seed-harvester ant *Pogonomyrmex californicus*: a comparative analysis of colony founding strategies. *Oecologia.* 2002;132:60–7.
27. Johnson RA. Colony founding by pleometrosis in the semiclastral seed-harvester ant *Pogonomyrmex californicus* (Hymenoptera: Formicidae). *Anim Behav.* 2004;68:1189–200.
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.
29. Haney BR, Fewell JH. Ecological drivers and reproductive consequences of non-kin cooperation by ant queens. *Oecologia.* 2018;187:643–55.
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. p. 308–33.
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.
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.
33. Clark RM, Fewell JH. Social dynamics drive selection in cooperative associations of ant queens. *Behav Ecol.* 2014;25:117–23.
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:jkaa019.
35. Boomsma JJ, Brady SG, Dunn RR, Gadau J, Heinze J, Keller L, et al. The Global Ant Genomics Alliance (GAGA). *Myrmecol News.* 2017;25:61–6.
36. Gadau J. DNA isolation from ants. *Cold Spring Harb Protoc.* 2009;4:pdb.prot5245.
37. 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 SUPPL.43:11.10.1-11.10.33.
38. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43:491–501.
39. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8.
40. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2012;9:179–81.
41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
42. Chang C, Chow C, Tellier LC, Vattikuti S, Purcell S, Lee J. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
43. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
44. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. *PLoS Genet.* 2012;8:1002453.
45. 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 (NY).* 2011;65:869–84.
46. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2022. <https://www.R-project.org/>.
47. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27:722–36.
48. Schiffels S, Wang K. MSMC and MSMC2: The multiple sequentially Markovian coalescent. *Methods Mol Biol.* 2020;2090:147–66.
49. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet.* 2014;46:919–25.
50. 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;523:463–7.
51. 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;107:1160–5.
52. 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;21:4748–69.
53. 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;16:980–9.

54. Kolaczkowski B, Kern AD, Holloway AK, Begun DJ. Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics*. 2011;187:245–60.
55. Booker TR, Yeaman S, Whitlock MC. Variation in recombination rate affects detection of outliers in genome scans under neutrality. *Mol Ecol*. 2020;29:4274–9.
56. 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;449:913–8.
57. Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol*. 2007;5:e171.
58. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*. 2011;27:578–9.
59. Gautier M, Klassmann A, Vitalis R. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. In: *Molecular ecology resources*. 2017. p. 78–90.
60. Voight BF, Kudravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol*. 2006;4:0446–58.
61. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
62. 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.
63. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
64. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:238.
65. Inbar S, Cohen P, Yahav T, Privman E. Comparative study of population genomic approaches for mapping colony-level traits. *PLOS Comput Biol*. 2020;16:e1007653.
66. 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.
67. 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:e37135.
68. Rochette NC, Rivera-Colón AG, Catchen JM. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol Ecol*. 2019;28:4737–54.
69. Bailly-Bechet M, Haudry A, Lerat E. "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. *Mob DNA*. 2014;5:13.
70. Gadau J. Phase-unknown linkage mapping in ants. *Cold Spring Harb Protoc*. 2009;4:pdb.prot5251.
71. Mester D, Ronin Y, Minkov D, Nevo E, Korol A. Constructing large-scale genetic maps using an evolutionary strategy algorithm. *Genetics*. 2003;165:2269–82.
72. Kosambi DD. The estimation of map distances from recombination values. *Ann Eugen*. 1943;12:172–5.
73. 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.
74. 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;26:2864–79.
75. Shin J-H, Blay S, Mcnenny B, Graham J. LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *JSS J Stat Softw*. 2006;16:1–9.
76. Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics*. 2012;28:1919–20.
77. Helmkamp 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;25:3716–30.
78. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*. 2005;21:1859–75.
79. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–100.
80. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30:1236–40.
81. Alexa A, Rahnenführer J. Gene set enrichment analysis with topGO. 2020.
82. Gower G, Tuke S, Rohrlach ABA, Soubrier J, Llamas B, Bean N, et al. Population size history from short genomic scaffolds: how short is too short?. *bioRxiv*. 2018. <https://doi.org/10.1101/382036>.
83. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3.
84. 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.
85. Johnson RA, Moreau CS. A new ant genus from southern Argentina and southern Chile, *Patagonomyrmex* (Hymenoptera: Formicidae). *Zootaxa*. 2016;4139:1–31.
86. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49:W293–6.
87. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. EGGNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res*. 2016;44:D286–93.
88. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
89. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
90. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016;32:292–4.
91. Dolgin ES, Charlesworth B. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics*. 2008;178:2169–77.
92. Kent TV, Uzunović J, Wright SI. Coevolution between transposable elements and recombination. *Philos Trans R Soc Lond B Biol Sci*. 2017;372:20160458.
93. Toth AL, Robinson GE. Genomics of social insects. In: *Encyclopedia of social insects*. Cham: Springer; 2021. p. 425–34.
94. Simola DF, Wissler L, Donahue G, Waterhouse RM, Helmkamp M, Roux J, et al. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res*. 2013;23:1235–47.
95. 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.
96. Chihara T, Luginbuhl D, Luo L. Cytoplasmic and mitochondrial protein translation in axonal and dendritic terminal arborization. *Nat Neurosci*. 2007;10:828–37.
97. Rossner R, Kaeberlein M, Leiser SF. Flavin-containing monooxygenases in aging and disease: emerging roles for ancient enzymes. *J Biol Chem*. 2017;292:11138–46.
98. Carrozza MJ, Utley RT, Workman JL, Côté J. The diverse functions of histone acetyltransferase complexes. *Trends Genet*. 2003;19:321–9.
99. Lyko F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet*. 2018;19:81–92.
100. Shaffer Z, Sasaki T, Haney B, Janssen M, Pratt SC, Fewell JH. The founder's dilemma: group selection for cooperation among queens of the harvester ant, *Pogonomyrmex californicus*. *Sci Rep*. 2016;6:1–9.
101. 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.
102. Bachtrog D. The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics*. 2008;179:1513–25.
103. 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 (N Y)*. 2004;58:1064–72.
104. 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;38:200–15.

105. 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;117:17130–4.
106. Ross KG. Multilocus evolution in fire ants: effects of selection, gene flow and recombination. *Genetics*. 1997;145:961–74.
107. Gutiérrez-Valencia J, Hughes PW, Berdan EL, Slotte T. The genomic architecture and evolutionary fates of supergenes. *Genome Biol Evol*. 2021;13:1–19.
108. Thompson MJ, Jiggins CD. Supergenes and their role in evolution. *Heredity*. 2014;113:1–8.
109. Charlesworth D, Charlesworth B. Theoretical genetics of batesian mimicry II. Evolution of supergenes. *J Theor Biol*. 1975;55:305–24.
110. 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;275:1181–8.
111. Thompson MJ, Timmermans MJTN. Characterising the phenotypic diversity of *Papilio dardanus* wing patterns using an extensive museum collection. *PLoS One*. 2014;9:e96815.
112. Clarke CA, Sheppard PM. Interactions between major genes and polygenes in the determination of the mimetic patterns of *Papilio dardanus*. *Evolution (N Y)*. 1963;17:404–13.
113. 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;279:316–25.
114. Flint J, Mackay TFC. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res*. 2009;19:723–33.
115. Ross KG, Keller L. Genetic control of social organization in an ant. *Proc Natl Acad Sci*. 1998;95:14232–7.
116. 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. *PLoS One*. 2014;9:e112963.
117. 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.
118. Errbii M. Genome assembly and the raw sequencing data for *Pogonomymex californicus*. 2021. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA682388/>. Accessed 25 Apr 2024.
119. Errbii M. Bioinformatic pipelines for evolutionary genomics of socially polymorphic populations of *Pogonomymex californicus*. 2021. <https://github.com/merrbii/Pcal-SocialPolymorphism>. Accessed 25 Apr 2024.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.