The *Pogonomyrmex californicus* social niche polymorphism is a polygenic trait involving a young supergene

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, August-von-Hartmann-Str. 13, 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. We discover a recently evolved (< 200 k years), 8 Mb non-recombining region segregating with the observed social niche polymorphism, showing characteristics of a supergene comparable to those underlying social polymorphisms in other ant species. However, we also find remarkable differences to the other so far 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*). Our results suggest that social morph in this species is a polygenic trait including an incipient supergene that evolved less than 200 000 years ago.

Introduction

Colonies of social Hymenoptera are led by a single (monogyny) or multiple reproductive females (polygyny) (Boomsma et al., 2014; Hölldobler and Wilson, 1977; Wilson, 1971). Such variations of social structure are suggested to have an adaptive value in certain conditions (Hölldobler and Wilson, 1977; Keller, 1993), and are conceptualized as alternative "social niches", i.e. "the set of social environments in which the focal individual has non-zero inclusive fitness" (Saltz et al., 2016). 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 (e.g. multiple *Formica* and *Myrmica* species (Seppä et al., 1995), *Solenopsis invicta* (Fletcher, 1983; Ross et al., 2007), *Messor pergandei* (Helms et al., 2013), *Pogonomyrmex californicus* (Overson et al., 2016), *Leptothorax acevorum* (Gill et al., 2009) and *L. longispinosus* (Herbers, 1986)).

Species that show 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 (Wang et al., 2013; Yan et al., 2020) and *Formica* wood ants (Brelsford et al., 2020; Purcell et al., 2021, 2014) 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 (Schwander et al., 2014), 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). In *Solenopsis* fire ants, monogyne queens are homozygous for the supergene SB haplotype, whereas polygyne queens are heterozygous SB/Sb (Keller and Ross, 1998; Ross and Keller, 1998). Homozygous Sb/Sb queens on the other hand, almost never reach reproductive maturity (Keller and Ross, 1998). Conversely, in *Formica* species, monogyne queens are always Sm/Sm homozygous, whereas polygyne queens can be either Sm/Sp or Sp/Sp (Avril et al., 2019; Purcell et al., 2014).

The California harvester ant *Pogonomyrmex californicus* (Buckley, 1867) occurs throughout the Southwestern USA and Northwestern Mexico (Johnson, 2002) and varies in terms of monogyny and polygyny, with primary monogyny as the predominant social organization (Johnson, 2004). 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”) (Haney and Fewell, 2018; Johnson, 2004; Overson et al., 2016; Rissing et al., 2000), and not re-adoption of queens or colony budding (Glancey and Lofgren, 1988; Rosengren et al., 1993).

Haney and Fewell (Haney and Fewell, 2018) found that higher colony densities and lower resource availabilities, i.e. harsher conditions, correlate with the frequency of pleometrosis and polygyny in *P. californicus*, suggesting that different social niches (haplo- and pleometrosis) are adaptations to different ecological niches. Hence, *P. californicus* queens construct their social niche (Saltz et al., 2016) during colony founding through social interactions (Clark and Fewell, 2014; Overson et al., 2014), analogous to ecological niche construction where organisms adaptively shape their environmental conditions (Odling-Smee et al., 1996; Saltz, 2011; Saltz and Nuzhdin, 2014).

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 almost fixed for either haplometrosis (H-population) or pleometrosis (P-population) (Overson et al., 2014; Rissing et al., 2000). We identify a non-recombining region of ca. 8 Mb that segregates with the observed polymorphism, showing all characteristics of a supergene comparable to the ones underlying previously described social niche polymorphisms in ants (Wang et al. 2013, Purcell et al. 2014). However, unlike in *Solenopsis* and *Formica* 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 relatively 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 10X sequencing (Bohn et al., 2021) 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 (Boomsma et al., 2017) (Dataset S1).

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) that were tested for their founding behavior (haplo- or pleometrotic; Supplemental file, Table S1) in standardized sandwich assays (Clark and Fewell, 2014). Principal Component Analyses (PCA) using 314,756 SNPs clearly separated queens of the two populations (Figure 1A and Supplemental file, Figure S1). 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 exactly between both populations in the PCA (Figure 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 (Figure 1B). Consistent with an ongoing admixture between the two populations, we found significant between-population haplotype-sharing in some individuals (e.g. P99P and H38W) corroborated by ADMIXTURE analyses (Figure 1B and Supplemental file, Figure S2).

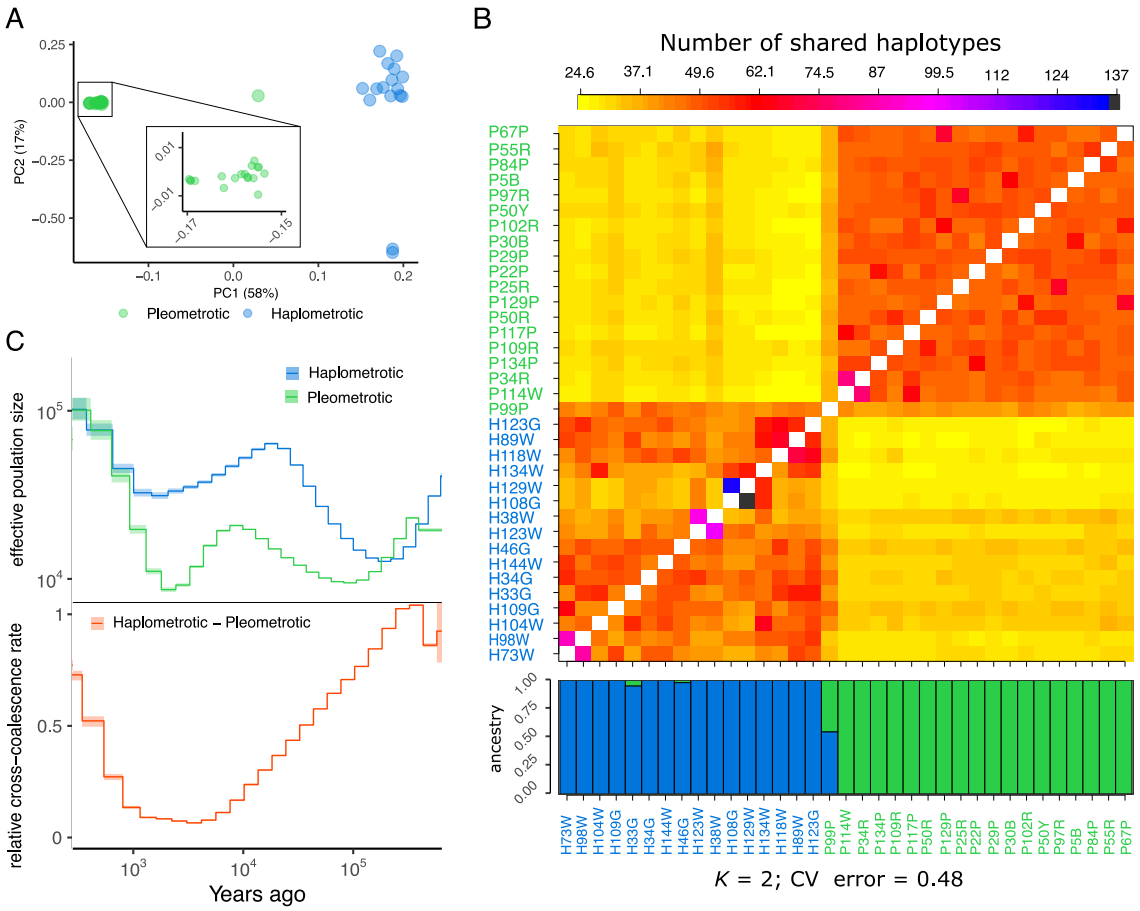

Figure 1. Population structure and demographic history in *P. californicus* populations. (A) Principal Component Analysis based on genome-wide SNP data of queens of the pleometrotic and haplometrotic populations (a scatterplot matrix showing four principal components can be found in Supplemental file, 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 Supplemental file, 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.

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

We used MSMC2 (Schiffels and Wang, 2020) to reconstruct the demographic history of the P- and H-populations (Figure 1C). Both populations showed a similar trajectory with a zigzag-like pattern in effective population size (N_e) between approximately 500,000 to 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 2,000 years ago. Following the first bottleneck, N_e consistently remained smaller in the pleometrotic population until between 2,000 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 (rCCR) (Schiffels and Durbin, 2014) and found that the two populations started diverging during the first bottleneck at ~300,000 years ago (Figure1C). Around 5,000 years ago, both populations were almost completely isolated (rCCR ≈ 0) but between 2,000 and 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 (Figures 1A and 1B).

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 (Figures 2A and 2B). Both estimates were significantly lower in the P- (median π = 4.71e-4, median het = 0.167) compared to the H-population (median π = 7.33e-4, median het = 0.274) (Figures 2A and 2B; 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) (Figure 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 (Supplemental file, Figure S3). This pattern is consistent with the repeated colonization and extinction of ephemeral habitats by *P. californicus*.

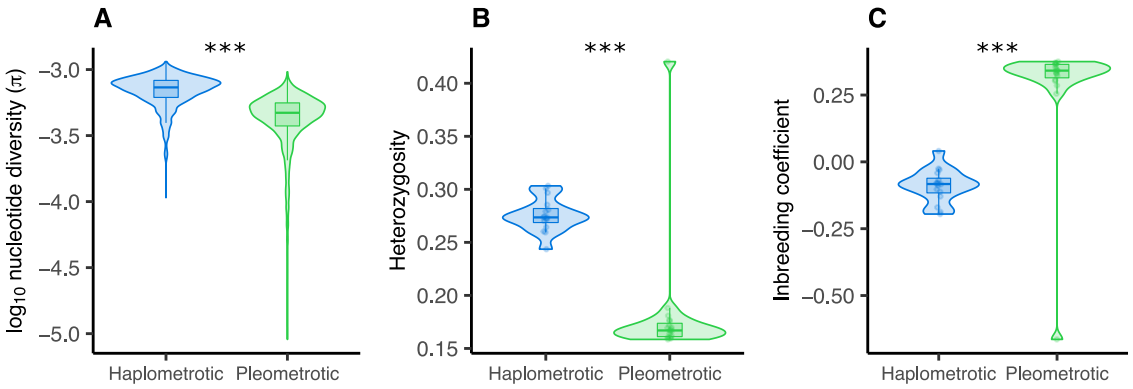

Figure 2. Higher Inbreeding and lower heterozygosity in the pleometrotic population. (A) Nucleotide diversity (π), (B) heterozygosity and (C) inbreeding coefficient in the haplometrotic (blue) and pleometrotic (green) populations. Heterozygosity and inbreeding coefficient were calculated on a per individual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

The social polymorphism in P. californicus is 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 extended haplotype scores (*xpEHH)* in the pleometrotic population (Supplemental file, Figure S4). Further analyses revealed that SNPs were in high linkage disequilibrium (LD), with little to no LD decay in these scaffolds (Supplemental file, Figure S5).

Linkage analysis on 108 haploid males from a monogynous colony belonging to the H-population using 2,980 ddRAD markers (double-digest Restriction-site Associated DNA sequencing) with no missing data put all markers in 29 Linkage groups (data not shown). The supergene was found in linkage group 14 (339 RAD markers, 255 cM over 26 Mb; Figure 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 (Supplemental file, Figure S6). Hence, this is a non-recombining region co-segregating with our social polymorphism, spanning across five complete physical scaffolds (22, 23, 38, 42, 48), and parts of scaffolds 6 and 14 (Supplemental file, Table S2). Visual inspection of 182 markers from both males and queens showed that distinct haplotypes extended along the entire region, indicating absence of recombination (Figure 3A) across at least ~8 Mb, consistent with patterns described for the supergenes in *Solenopsis* and *Formica* ants of non-recombining regions of ~13 and ~10 Mb, respectively (Brelsford et al., 2020; Purcell et al., 2014; Wang et al., 2013). Dense population genomic data using 5,363 SNPs genotyped across all 35 queens further confirmed this finding (Supplemental file, Figure S7).

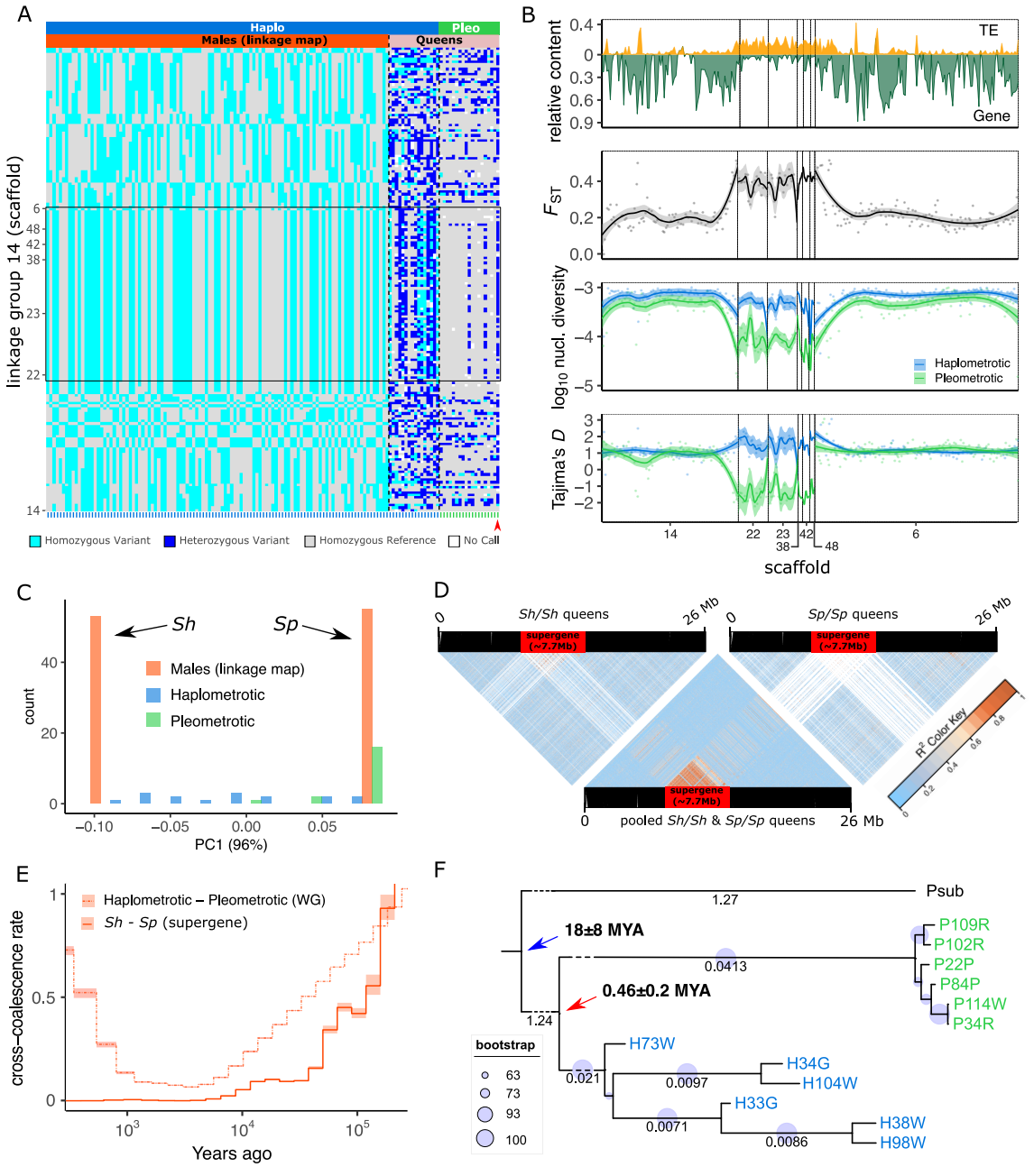

Figure 3. Genomic architecture and recombination rates at linkage group (LG) 14 harboring a putative supergene. (A) Genotypes for 108 sons from a single monogynous queen (haplometrotic population) based on RADseq markers and for the 35 queen samples from the pleometrotic and haplometrotic populations 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 upper bond estimate for the formation of the Sh–Sp haplotype groups 0.46 (95% HPD ±0.2) MYA.

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 (Figure 3B and Supplemental file, Figure S8A; Wilcoxon test: W = 1602.5, p < 2.2e-16). This is consistent with the prediction that TEs accumulate in regions of low recombination (Dolgin and Charlesworth, 2008; Kent et al., 2017). Further, genetic differentiation as measured by F_{ST} between the H- and P-population was significantly increased in the supergene compared to the rest of the linkage group (Figure 3B and Supplemental file, Figure S8B; Wilcoxon test: W = 12261, p < 2.2e-16). Additionally, nucleotide diversity and Tajima's *D* in this region were significantly reduced in the pleometrotic population (Figure 3B and Supplemental file, Figures S8C and S7D; Kruskal–Wallis rank sum test $_{\text{nucleotide diversity}}$, χ^2 = 254.77, df = 3, p < 2.2e-16; Kruskal–Wallis rank sum test $_{\text{Tajima's } D}$, χ^2 = 177.81, df = 3, p < 2.2e-16; results of pairwise Wilcoxon rank sum *post hoc* tests are displayed in the figure). Finally, 156 expressed transcripts from a previously published RNA-seq dataset (Helmkampf et al., 2016) mapped to the supergene. 13 of those transcripts (8.3 %) were reported to be significantly associated with aggression depending on the social context of the queens (no significant enrichment, Dataset S2).

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 the progeny of a heterozygous Sh/Sp mother queen (Figure 3C and Supplemental file, Figure S9). 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 (Figure 3C and Supplemental file, Figure S9). Further support for the suppression of recombination between Sh and Sp alleles came from LD analyses across queens. Along the region, there was strong LD in a mixed pool of Sh/Sh and Sp/Sp queens as well as heterozygous Sh/Sp, but not in separate pools of Sh/Sh queens or Sp/Sp queens (Figure 3D and Supplemental file, Figure S10), indicating recombination in homozygous, but not in heterozygous queens. Such suppression of recombination in heterozygotes is a common

characteristic of supergenes (Charlesworth, 2016; Schwander et al., 2014; Wang et al., 2013; Yan et al., 2020). Taken together, these data indicate the presence of an ~8 Mb non-recombining supergene with two major haplotypes (Sh and Sp) in the two populations of *P. californicus*, segregating with the social niche polymorphism in colony founding. Unlike other social supergene systems (e.g. in *Solenopsis*) (Ross et al., 2007), 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. 0.0711, p = 0.789) (Supplemental file, Table S3).

Coalescence modelling revealed that the demographic histories of the supergene and the rest of the genome only diverged within the last 2,000 years. Since then, cross-coalescence quickly increased across the genome, consistent with a secondary onset of gene flow between H- and P-populations, while remaining nearly zero in the supergene (Figure 3E). These findings suggest that Sh and Sp haplotypes initially diverged due to allopatric separation 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 (which occurs in sympatry with *P. californicus*) showed that Sh and Sp diverged less than 0.46 (95% highest probability density (HPD) 0.26-0.66) MYA (Figure 3F).

In addition to the supergene region, our genome scans revealed several regions under divergent selection between H- and P-queens. We first identified highly differentiated (outlier) regions by estimating F_{ST} in 100 kb non-overlapping windows across the genome (Supplemental file, Figures S4 and S11) and subsequently compared estimates of Tajima's *D* and nucleotide diversity in the two populations to identify signatures of selective sweeps (Supplemental file, Figure 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 (Dataset S3). Cross-population extended haplotype homozygosity (*xpEHH*) tests (Sabeti et al., 2007) (Supplemental file, Figure S4) identified 340 outlier SNPs in 32 genomic locations under selection (Dataset S4). Overlapping both analyses resulted in 17 remaining candidate regions under selection in the P-population (Supplemental file, Table S4). Exonic SNPs in all 17 candidate regions included 70 synonymous changes and 73 non-synonymous variants affecting the protein sequence (Table 1).

288 **Table 1.** Effects of single nucleotide polymorphisms (SNPs) found in candidate regions that show
289 strong signatures of selection in the pleometrotic population. We annotated 6,077 putative effects
290 on protein coding genes of 3,347 SNPs.
291

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

292

293

Intriguingly, the supergene contained 12 of these 17 regions (on scaffolds 14, 22 and 23). The remaining five regions showing clear signatures of selective sweeps in the P-population were clustered on scaffolds 1, 9, 15, and 16 (Supplemental file, Figure 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 2,980) leniently filtered ddRAD markers (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 (Figure 4A), were placed into three different linkage groups, indicating that these regions are independent of the supergene.

Among the 48 genes in these four remaining regions, 34 are likely functional as they are not encoding TE proteins, are expressed, or homologous to other known insect proteins (Dataset S5). 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 haplometrotic individuals carry the alternative and >90% of pleometrotic individuals carry the reference allele (Figure 4B). Further, regions harboring the six genes exhibit *EHHS* scores revealing a pattern of increased extended haplotype homozygosity in the pleometrotic population (Supplemental file, Figure S12), 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 haplometrotic queens are either heterozygous or homozygous for alternate alleles (Supplemental file, Figures S13A-S13F), which is indeed consistent with a selective sweep fixing the reference allele in the P-population.

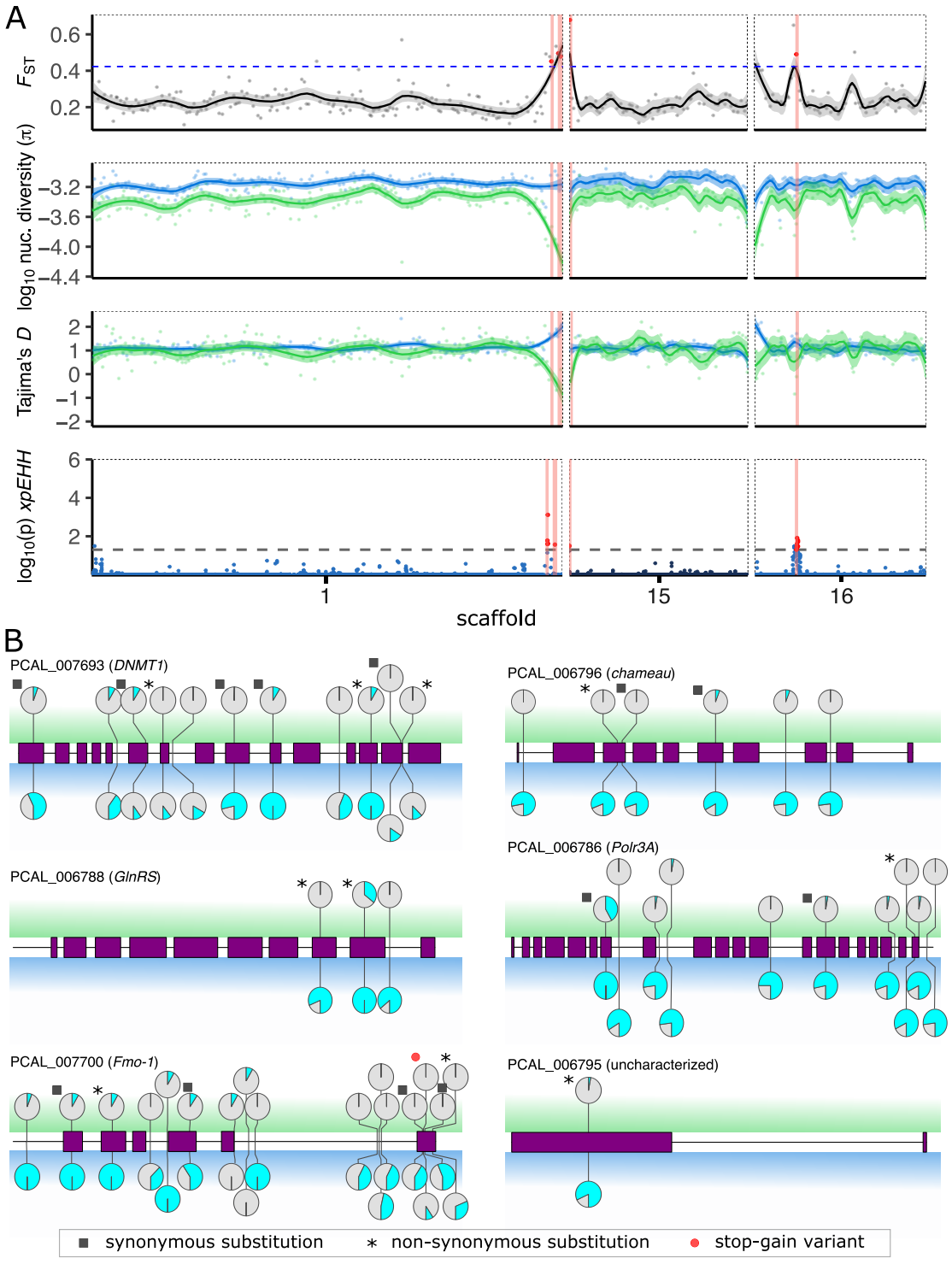

Figure 4. Signatures of selection in non-supergene regions. (**A**) Sliding window analyses of scaffolds 1, 15 and 16 showing (from top to bottom) genetic differentiation between populations (F_{ST}), nucleotide diversity (π), Tajima's *D* and the cross-population extended haplotype homozygosity (*xpEHH*). 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.

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 Supplemental file, Figure 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 haplometrotic population (bottom; blue background).

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 (Chihara et al., 2007). 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 ageing (Rossner et al., 2017). 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 involved in transcriptional regulation via histone modification (Carrozza et al., 2003). PCAL_007693 encodes *DNMT1* (which was lost in Diptera), the principal methyl transferase in invertebrates for epigenetic regulation via DNA methylation (Lyko, 2018). While these candidates suggest that epigenetic transcriptional regulation modulates the social niche polymorphisms 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.

Discussion

The social niche polymorphism (Saltz et al., 2016) in colony founding of *P. californicus* was first reported over 20 years ago (Rissing et al., 2000). 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) (Overson et al., 2016). 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 (Haney and Fewell, 2018). Since its original description, we have learned a lot about behavioral, physiological, ecological, and socio- genetic correlates of haplo- and pleometrosis (Clark and Fewell, 2014; Haney and Fewell, 2018; Johnson, 2004, 2002; Overson et al., 2016, 2014; Shaffer et al., 2016). However, the genomic architecture and evolutionary history was unknown.

Our study reveals that the social polymorphism in *P. californicus* is associated with an approximately 8 Mb non-recombining region, reminiscent of social supergenes described for social polymorphisms in *Solenopsis* fire ants (Wang et al., 2013) and *Formica* wood ants (Purcell et al., 2014). 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*. These differences could possibly relate to the *Pogonomyrmex* supergene being much younger (*Pogonomyrmex*: ~0.2 MYA, *Solenopsis*: 0.39–1.1 MYA (Cohen and Privman, 2020; Wang et al., 2013; Yan et al., 2020) and *Formica*: 20–40 M years (Brelsford et al., 2020)).

Our analyses provide several lines of evidence that the 8 Mb non-recombining region is indeed a supergene. This region is similar in size to the supergenes in *Formica* and *Solenopsis* (8 Mb vs. 11 and 12.7 Mb). 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 (Pracana et al., 2017). Tajima's *D* and haplotype homozygosity statistics suggest a selective sweep of the Sp haplotype, similar to the *Solenopsis* Sb haplotype (Cohen and Privman, 2020). Very low LD decay and absence of recombinant F1 offspring 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 (Bachtrog, 2008).

Intriguingly, the association between social form and supergene genotype is not perfect. While pleometrotic queens were predominantly homozygous (85%) for a single supergene haplotype (Sp), most of the haplometrotic queens (93%) carried at least one copy of the Sh haplotype. This suggests that supergene genotype alone is insufficient to determine the social form in this species - a fundamental difference to the supergenes described in other ants (Brelsford et al., 2020; Ross and Keller, 1998)! 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 (Chapuisat et al., 2004; Shoemaker et al., 2006). In *P. californicus* however, certain populations are nearly fixed for either the haplo- or the pleometrotic strategy (Overson et al., 2016). The current study further corroborates Overson's et al. (2016) finding of significantly reduced gene flow between the two focal populations both at the nuclear and mitochondrial level. Genetic differentiation between haplo- and pleometrotic *P. californicus* is hence significantly higher than between social forms in *S. invicta* and *F. selysi*, both in sympatry and in allopatry (Chapuisat et al., 2004; Shoemaker et al., 2006). Similar levels of differentiation have however been described between monogyne and polygyne individuals in other *Formica* species, such as *F. exsecta* or *F. truncorum* (Brelsford et al., 2020). As we do not find significant deviations from HWE at the supergene in either population, there appears to be no intrinsically lethal haplotype combinations, unlike in *Formica* and *Solenopsis* where lethal combinations exist (either due to intrinsic factors or behavioral exclusion of one genotype) and are presumed crucial for the maintenance of the social niche polymorphism in these species (Avril et al., 2020; Ross, 1997).

In *Formica* and *Solenopsis*, the derived polygyne form is associated with the dominant supergene haplotype, while in *P. californicus* it is associated with the recessive haplotype (Sp). That is, polygyne *P. californicus* queens are homozygous for the derived Sp haplotype, while monogyne queens are homozygous for the ancestral haplotype in the other species. This difference has implications for the evolutionary dynamics of both neutral and adaptive processes of the supergene, and the evolution of genes contributing to the novel traits on the derived haplotype. Notably, the *P. californicus* Sp haplotype cannot accumulate as many deleterious mutations as the *Solenopsis* Sb haplotype, because it has to be functional in a homozygous state in the pleometrotic queens.

Further, unlike in *Solenopsis* and *Formica*, we find additional genomic regions other than the supergene that segregate with the social polymorphism in *P. californicus*. Theoretical models for supergene evolution predict gradual elaboration of supergenes by integration of formerly unlinked modifier loci, particularly in heterogeneous environments (Yeaman, 2013). According to these models, natural selection will favor genomic rearrangements that eventually join the supergene with previously unlinked but co-evolved loci in linkage (Thompson and Jiggins, 2014). This is consistent with our analyses showing that recombination at the supergene ceased only between approximately 200,000 to 5,000 years ago, almost overlapping with the 95% HPD estimates of the Sh and Sp divergence time (i.e. 0.26-0.66 MYA). However, Charlesworth and Charlesworth (1975) already argued that a scenario of coalescence of unlinked modifier loci with a supergene is credible only if all loci are already on the same chromosome. 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* (Clark et al., 2008; Clarke and Sheppard, 1963; Thompson and Timmermans, 2014) and *Heliconius numata* (Jones et al., 2012). The presence of unlinked modifier loci influencing the expression and ameliorating negative fitness effects of major quantitative trait loci is not unusual (Flint and Mackay, 2009), yet none have been described so far for the known social supergenes. Future studies that genetically and phenotypically characterize additional haplometrotic, pleometrotic, and mixed populations of *P. californicus* will be vital to resolve whether the same unlinked loci we describe

429 are involved in social polymorphism within other populations as well or have indeed evolved as
430 modifier loci unique to our two focal populations.

431 The four differentiated (physically unlinked) genomic regions outside the supergene that potentially
432 contribute to the expression of haplometrosis or pleometrosis, either by epistatic or additive
433 interactions with the supergene, contain 34 genes showing evidence of a selective sweep.
434 Intriguingly, correlation of genotype and phenotype were nearly perfect for the DNA
435 methyltransferase *DNMT1* and the histone acetyltransferase *chameau* (i.e. all P-individuals carry
436 the reference allele, all H-individuals carry the alternative allele), both of which could be relevant
437 for epigenetically regulating the expression of the alternative phenotypes. As pleometrotic queens
438 show a higher behavioral (e.g. division of labor) (Clark and Fewell, 2014; Overson et al., 2014) and
439 physiological (gene expression) (Helmkampf et al., 2016) plasticity than haplometrotic queens
440 during colony founding, differences in epigenetic regulation of plastic traits like division of labor
441 could be highly relevant for maintaining the two divergent social morphs.

442 With regard to the behavioral syndrome underlying the social polymorphism, previous population
443 screens and behavioral experiments showed that not all individuals from haplometrotic populations
444 act aggressively and, *vice versa*, not all queens from the pleometrotic population are tolerant
445 against co-foundresses (Clark and Fewell, 2014; Overson et al., 2014), indicating that social morph
446 and the underlying genetic loci are not fixed in either population. In our study, the four queens that
447 showed an apparent mismatch between the social supergene genotype and population, behaved
448 according to their population of origin and not their supergene genotype (Supplemental file, Table
449 S1).

450 These observations suggest that environmental conditions, in particular the social context and
451 interactions with co-founding queens (indirect genetic effects), might shift the expression of certain
452 behavioral traits involved in this complex social niche polymorphism (Clark and Fewell, 2014;
453 Overson et al., 2014; Saltz et al., 2016). We note that the margin of error of our behavioral assays
454 to discriminate haplometrotic and pleometrotic individuals is certainly larger than for e.g.

455 morphological or other obvious dimorphisms. In particular, if environmental, ecological and social
456 conditions that we can only poorly emulate under laboratory conditions also contribute to the
457 expression of the social morph. Regardless, we conclude from these patterns that the social morph
458 in this species is controlled or at least regulated by a polygenic architecture including an incipient
459 supergene, but that –ultimately– the expression of haplometrosis or pleometrosis is also impacted
460 by social context and environmental and ecological conditions (Haney and Fewell, 2018).

461 Our coalescent analysis indicates that the haplo- and pleometrotic populations started to separate
462 approximately 300,000 years ago, reaching almost complete isolation around 5,000 years ago,
463 before gene flow between these two populations increased again considerably. In accordance,
464 some individuals showed clear patterns of hybridization, recent co-ancestry and admixture (e.g.
465 potential F1 hybrid queen P99P; Figures 1A and 1B), consistent with ongoing admixture between
466 the two populations, which are only 50 km apart. The onset of gene flow between the haplometrotic
467 and pleometrotic population starting ~2,000 years ago could secondarily reduce genetic
468 differentiation between both forms everywhere but in the non-recombining supergene, eventually
469 resulting in the pattern also observed in *Solenopsis invicta* and *Formica selysi*, where genetic
470 differentiation between both social morphs is nearly exclusively restricted to the supergene
471 (Pracana et al., 2017; Purcell et al., 2014).

472 Supergenes have long been suspected to facilitate the emergence and maintenance of complex
473 phenotypes (Schwander et al., 2014). The success of the polygynous form in *S. invicta* in its
474 introduced range is attributed to its ability to efficiently colonize disturbed areas rather than its
475 higher competitive abilities (King and Tschinkel, 2006). This might be similar in *P. californicus*
476 where pleometrosis can be very successful and nearly reaching fixation in newly colonized
477 disturbed habitats.

478 Theoretical models suggest that pleometrotic populations might be stable against invasions from
479 haplometrotic individuals, if pleometrosis occurs above a certain frequency, because single
480 haplometrotic queens would be out-competed by polygynous colonies (Shaffer et al., 2016).

Conversely, the founding success of pleometrotic queens is likely negatively correlated with the frequency of haplometrotic individuals in a population, as pleometrotic/haplometrotic pairings commonly result in the death of the former. This would explain the rareness of high frequency pleometrotic populations (Johnson, 2002), as such populations can only evolve when the frequency of pleometrosis constantly remains above a certain threshold. Such social dynamics may also explain the geographical separation between populations that are nearly fixed for either social form, which allows for genetic divergence also outside of the supergene.

Conclusion

Pleometrosis and primary polygyny in *P. californicus* appear to have evolved as an adaptation to difficult environments, i.e. higher colony densities and resource limitation. However, this social niche is generally rare in *P. californicus* and has reached fixation in only a few populations. The genetic architecture underlying this social niche polymorphism shows similarities, but also striking differences to previously studied social niche polymorphisms in *Formica* and *Solenopsis*. *Pogonomyrmex* also harbors a supergene, which is similar in size and has two major segregating haplotypes, as in the other systems. However, there is currently no evidence for intrinsic lethality or the selective killing of individuals carrying a certain genotype combination in *P. californicus*. Whereas the dominant haplotype is associated with polygyny in *Solenopsis* (Sb) and *Formica* (Sp), our results suggest that the haplometrotic Sh allele is dominant in *P. californicus*. This might additionally limit the frequency of pleometrosis, but it also allows for its persistence in an otherwise predominantly haplometrotic species as a recessive allele, invisible to selection. Apart from the supergene, there are at least four additional regions showing signatures of a selective sweep in the pleometrotic population/social phenotype. Candidate genes linked to epigenetic modifications in these regions show strict co-segregation with the social form implicating them in the expression of haplo- or pleometrosis. Further studies are necessary to understand the relevance and interactions of these loci for the expression of the different social phenotypes, as they might turn out to be more important than the supergene itself. We are aware of the limitations of our current study, focusing on just two geographically separated, non-panmictic populations, as we cannot entirely exclude that the supergene and other genetic differences are demographic effects (i.e. fixation by drift) or

509 local adaptations to some unknown ecological factor rather than to social phenotype. However, a
510 strong association of the supergene and the additional loci with the social phenotype currently is
511 the most likely explanation. Further populations of *P. californicus*, where both haplo- and
512 pleometrotic forms co-occur sympatrically, are known (Haney, 2017), offering the necessary setting
513 for further studies to disentangle the different evolutionary forces and conclusively describe the
514 genetic architecture of this social niche polymorphism. Already, our study showcases that the
515 genetic architecture underlying social niche polymorphisms may be much more complex and
516 diverse than previously appreciated.

Materials and Methods

Further details about Materials and Methods are provided in the Supplemental file.

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; predominantly pleometrotic population, P-population) and Lake Henshaw resort (33.232 N, -116.760 W; predominantly haplometrotic population, H-population), California, USA, which are 50 km apart. We tested for aggressive interactions between queen pairs in sandwich assays (Supplemental file) (Clark and Fewell, 2014).

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 (2x 75 bp, average coverage 12x) at the Genomic Core Facility of the University of Münster (Germany), using a NextSeq 500 (Supplemental file, Table S1).

Genome assembly and annotation

We sequenced and assembled the genome of *P. californicus* ("Pcal2") using data from MinION long reads (MinION, Oxford Nanopore Technologies, Oxford, UK), Illumina short reads, and previously published 10X Chromium linked reads (Bohn et al., 2021). 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 ~30x filtered Nanopore data using the following strategy. We created one assembly using CANU (v.2.0) (Koren et al., 2017) 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) (Boetzer et al., 2011) 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 (Walker et al., 2014). The polished assembly was super-scaffolded with publicly available 10X Chromium data for *P. californicus* (accession SRX8051306) (Bohn et al., 2021) using scaff10x and subsequent correction with break10x. 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) (Keilwagen et al., 2016) 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 (Bohn et al., 2021).

Read mapping and variant calling

Short-read data was cleaned with Trimmomatic (v.0.38) (Bolger et al., 2014). Paired reads were mapped to the genome assembly using BWA-MEM (v.0.7.17-r1188) (Li and Durbin, 2009). After marking duplicates using Picard's MarkDuplicates (v.2.20.0), we performed joint-variant calling across all samples with the GATK (Van der Auwera et al., 2013) and filtered variant calls with VCFtools (v.0.1.16) (Danecek et al., 2011). After a visual inspection of mapping quality of reads across the genome (Supplemental file, Figure S14), 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) (Delaneau et al., 2012) 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 Principal Component Analyses (PCA), considering four eigenvectors. LD-pruned SNPs were further used for ADMIXTURE (v.1.3.0) analyses (Alexander et al., 2009) with 2 to 4 clusters. To better characterize the relationships among queens in our sample, we used the haplotype-based approach implemented in fineSTRUCTURE (Lawson et al., 2012). In our model, we assumed a uniform recombination rate, based on the estimate of 14 cM/Mb found in *Pogonomyrmex rugosus* (Sirviö et al., 2011).

Demographic population history analysis

We inferred demographic population history and the relative cross-coalescence rate (rCCR) using Multiple Sequential Markovian Coalescence modeling (MSMC2) (Schiffels and Durbin, 2014) 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 1,000 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 (Yang et al., 2015), i.e. $3.6 * 10^{-9}$ mutations * nucleotide^{-1} * generation^{-1}.

Population genomic estimates

We used VCFtools (v.0.1.16) (Danecek et al., 2011) to estimate nucleotide diversity (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. Linkage disequilibrium (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 minor allele frequency

(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 (Akey et al., 2010; Fabian et al., 2012; Kelley et al., 2006; Kolaczkowski et al., 2011) and by identifying regions of extended haplotype homozygosity (selective sweeps) (Sabeti et al., 2007; Tang et al., 2007), comparing the haplometrotic and pleometrotic 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 (Booker et al., 2020). Hence, we further applied the cross-population extended haplotype homozygosity (*xpEHH*) and the site-specific extended haplotype homozygosity (*EHHS*) tests (Sabeti et al., 2007; Tang et al., 2007), 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) (Gautier et al., 2017). We used the Benjamani & 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 (Dataset S4). 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 (Supplemental file, Table S4).

Characterization of candidate genes and SNPs

For functional characterization of genes, we performed BLAST (blastx) (Altschul et al., 1990) searches against the nr database (December 2019), used OrthoFinder (v.2.3.1) (Emms and Kelly, 2019) to identify *Drosophila melanogaster* orthologs, and used RepeatProteinMask to identify

genes encoding TE proteins. We used SnpEff (v.5.0) (Cingolani et al., 2012) to annotate putative effects on protein-coding genes of all 3,347 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 RNAseq 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 (Dataset S5). In six of these 18 genes at least one 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) (Figure 4B).

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 (Supplemental file, Table S3). Haplotypes were assigned manually, based on visual inspection of the suggested supergene region (Supplemental file, Figure S7); predominantly homozygous for the reference alleles (i.e. from a pleometrotic 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. (Inbar et al., 2020), using a modified double-digest (EcoR1 and Mse1) Restriction-site Associated DNA sequencing (ddRADseq) protocol (Parchman et al., 2012; Peterson et al., 2012). Briefly, genomic DNA was digested at 37°C for 8hrs 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) (Bolger et al., 2014), and *process_radtags* (STACKS v.2.2) (Rochette et al., 2019) for quality control, filtering, and demultiplexing the raw data. Reads were mapped with Bowtie2 v2.3.4.2 (Langmead et al 2012) 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) (Danecek et al., 2011) and excluding heterozygous markers we arrived at a final set of 2,980 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 (Kosambi, 1943).

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 (RADseq). 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) (Tollefson et al., 2019). 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, 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 (Supplemental file, Figure S7). After VCF lift over (on scaffolds 14, 22, 23, 38, 42, 48 and 6 into LG14) using flo (Pracana et al., 2017) and picard's *LiftoverVcf* (v.2.20.0), we used VCFtools (v.0.1.16) (Danecek et al., 2011) for filtering and LDheatmap (v.1.0-4) (Shin et al., 2006) 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

683 to be associated with the aggressive behavior of queens (Helmkampf et al., 2016) belong to the
684 supergene, we aligned the assembled transcripts to the Pcal2 reference genomes using GMAP
685 (Wu and Watanabe, 2005). We only retained the top hits with at least 70% query coverage, yielding
686 7303 (out of 7890) transcripts unambiguously aligned to the reference genome. We then used
687 BEDtools (v.2.29.2) (Quinlan and Hall, 2010) to extract transcripts that mapped to the supergene
688 region (Dataset S2).
689
690 *Dating the supergene formation*
691 We estimated the divergence between Sh and Sp by computing the rCCR, using MSMC2 (Schiffels
692 and Durbin, 2014) on four Sh/Sh queens (H104W, H38W, H73W and H98W) and four Sp/Sp
693 queens (P114W, P50R, P84P and P97R). We limited the analysis to the supergene region
694 (scaffolds 22, 23 and part of scaffold 6) and ran 500 bootstraps by randomly sampling 2 Mb each
695 time.
696 Further, we performed a phylogenetic analysis of Sh and Sp alleles with *P. subnitidus* as an
697 outgroup species. Sequencing data of one *P. subnitidus* individual was generated and processed
698 as described above for *P. californicus*. Joint SNP calling was performed using GATK (Van der
699 Auwera et al., 2013) and a maximum-likelihood (ML) tree was constructed based on 33,424 SNPs
700 at the supergene region (scaffolds 22, 23, 38, 42, 48 and parts of scaffolds 6 and 14), using RAxML
701 (Stamatakis, 2014) under the GTRGAMMA model with bootstrapping.
702 We then used a fossil-calibrated dated phylogeny (Ward et al., 2015) to date the split between the
703 Sh and Sp haplotype groups using a phylogenetic approach. The Ward et al. (Ward et al., 2015)
704 phylogeny dated the split between *P. vermiculatus* and *P. imberbiculus* at 18 (95% HPD ±8) MYA.
705 This is an upper bound on the speciation date of *P, californicus* and *P. subnitidus* according to the
706 most recent *Pogonomyrmex* phylogeny (Johnson and Moreau, 2016). We calculated a ratio of 1:39
707 between the speciation time of *P. subnitidus* and *P. californicus*, and the Sh–Sb haplotypes
708 divergence time. Based on our dating reference (i.e. 18±8 MYA), we obtained 0.46 (95% HPD
709 ±0.20) MYA as an upper bound for the divergence between the two haplotype groups, which is
710 consistent with the MSMC2 analysis. We note that the confidence in the age inference is

711 strengthened by the agreement between two distinct approaches: the inference by MSMC2 is
712 based on a coalescent model, which is a distinct approach from the phylogenetic dating approach.
713 The former approach depends on estimates of mutation rate and generation time, whereas the
714 latter depends on fossil calibration points. The tree figure was produced using interactive Tree Of
715 Life (iTOL) (Letunic and Bork, 2021).
716

Data Availability

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

Acknowledgments

We thank Pine Valley Academy (Pine Valley, California, USA) for support and the staff of Lake Henshaw resort (Lake Henshaw, California, USA) for permission to collect ants on their premises. Thanks to Jenny Märzhäuser, Tobias van Elst and Ti Eriksson for help in collecting the ants, and to Jennifer Fewell for hosting us at Arizona State University, Tempe, Arizona, USA. We are grateful for Hilde Schwitte's and Una Hadziomerovic's help in DNA extraction, barcoding, and library construction. JG and UE were funded by the German Research Foundation (DFG) as part of the SFB TRR 212 (NC³) – TP C04 project numbers 316099922 and 396780988. ME was funded by the German Research Foundation (DFG) – 403813881 with grant to LS (SCHR 1554/2-1) under the priority program "Rapid evolutionary adaptation: Potential and constraints" (SPP 1819). ME acknowledges support by the DFG Research Training Group 2220 "Evolutionary Processes in Adaptation and Disease".

Competing interests

The authors have no conflict of interest to declare.

Author Contributions

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

References

Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, Nicholas TJ, Neff MW. 2010. Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci U S A* **107**:1160–1165. doi:10.1073/pnas.0909918107

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**:1655–1664. doi:10.1101/gr.094052.109

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

Avril A, Purcell J, Béniguel S, Chapuisat M. 2020. Maternal effect killing by a supergene controlling ant social organization. *Proc Natl Acad Sci U S A* **117**:17130–17134. doi:10.1073/pnas.2003282117

Avril A, Purcell J, Brelsford A, Chapuisat M. 2019. Asymmetric assortative mating and queen polyandry are linked to a supergene controlling ant social organization. *Mol Ecol* **28**:1428–1438. doi:10.1111/MEC.14793

Bachtrog D. 2008. The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics* **179**:1513–1525. doi:10.1534/genetics.107.084012

Boetzer M, Henkel C V., Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**:578–579. doi:10.1093/bioinformatics/btq683

Bohn J, Halabian R, Schrader L, Shabardina V, Steffen R, Suzuki Y, Ernst UR, Gadau J, Makałowski W. 2021. Genome assembly and annotation of the California harvester ant Pogonomyrmex californicus. *G3 Genes, Genomes, Genet* **11**. doi:10.1093/G3JOURNAL/JKAA019

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170

Booker TR, Yeaman S, Whitlock MC. 2020. Variation in recombination rate affects detection of outliers in genome scans under neutrality. *Mol Ecol* **29**:4274–4279. doi:10.1111/MEC.15501

Boomsma JJ, Brady SG, Dunn RR, Gadau J, Heinze J, Keller L, Moreau CS, Sanders NJ, Schrader L, Schultz TR, Sundström L, Ward PS, Wcislo WT, Zhang G, The GAGA Consortium. 2017. The Global Ant Genomics Alliance (GAGA). *Myrmecological News* **25**:61–66.

Boomsma JJ, Huszár DB, Pedersen JS. 2014. The evolution of multiqueen breeding in eusocial lineages with permanent physically differentiated castes. *Anim Behav* **92**:241–252. doi:10.1016/j.anbehav.2014.03.005

Brelsford A, Purcell J, Avril A, Tran Van P, Zhang J, Brütsch T, Sundström L, Helanterä H, Chapuisat M. 2020. An Ancient and Eroded Social Supergene Is Widespread across Formica Ants. *Curr Biol* **30**:304-311.e4. doi:10.1016/j.cub.2019.11.032

Buckley SB. 1867. Descriptions of new species of North American Formicidae. *Proc Entomol Soc Philadelphia* **6**:152–172.

Carrozza MJ, Utley RT, Workman JL, Côté J. 2003. The diverse functions of histone acetyltransferase complexes. *Trends Genet* **19**:321–329. doi:10.1016/S0168-9525(03)00115-X

Chapuisat M, Bocherens S, Rosset H. 2004. Variable queen number in ant colonies: No impact on queen turnover, inbreeding, and population genetic differentiation in the ant Formica selysi. *Evolution (N Y)* **58**:1064–1072. doi:10.1111/j.0014-3820.2004.tb00440.x

Charlesworth D. 2016. The status of supergenes in the 21st century: Recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evol Appl* **9**:74–90. doi:10.1111/eva.12291

Charlesworth D, Charlesworth B. 1975. Theoretical genetics of batesian mimicry II. Evolution of supergenes. *J Theor Biol* **55**:305–324. doi:10.1016/S0022-5193(75)80082-8

Chihara T, Luginbuhl D, Luo L. 2007. Cytoplasmic and mitochondrial protein translation in axonal and dendritic terminal arborization. *Nat Neurosci* **10**:828–837. doi:10.1038/nn1910

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. 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)* **6**:80–92. doi:10.4161/fly.19695

Clark R, Brown SM, Collins SC, Jiggins CD, Heckel DG, Vogler AP. 2008. 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* **275**:1181–1188. doi:10.1098/RSPB.2007.1762

Clark RM, Fewell JH. 2014. Social dynamics drive selection in cooperative associations of ant queens. *Behav Ecol* **25**:117–123. doi:10.1093/beheco/art093

Clarke CA, Sheppard PM. 1963. Interactions between major genes and polygenes in the determination of the mimetic patterns of Papilio dardanus. *Evolution (N Y)* **17**:404–413. doi:10.1111/j.1558-5646.1963.tb03297.x

Cohen P, Privman E. 2020. The social supergene dates back to the speciation time of two Solenopsis fire ant species. *Sci Rep* **10**:1–9. doi:10.1038/s41598-020-67999-z

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158. doi:10.1093/bioinformatics/btr330

Delaneau O, Marchini J, Zagury JF. 2012. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**:179–181. doi:10.1038/nmeth.1785

Dolgin ES, Charlesworth B. 2008. The Effects of Recombination Rate on the Distribution and Abundance of Transposable Elements. *Genetics* **178**:2169–2177. doi:10.1534/GENETICS.107.082743

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

Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlötterer C, Flatt T. 2012. Genome-wide patterns of latitudinal differentiation among populations of Drosophila melanogaster from North America. *Mol Ecol* **21**:4748–4769. doi:10.1111/j.1365-294X.2012.05731.x

Fletcher DJC. 1983. Three newly-discovered polygynous populations of the fire ant, Solenopsis invicta, and their significance. *J Georg Entomol Soc* **18**:538–543.

Flint J, Mackay TFC. 2009. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res* **19**:723–733. doi:10.1101/GR.086660.108

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

Gill RJ, Arce A, Keller L, Hammond RL. 2009. Polymorphic social organization in an ant. *Proc R Soc B Biol Sci* **276**:4423–4431. doi:10.1098/rspb.2009.1408

Glancey BM, Lofgren CS. 1988. Adoption of Newly-Mated Queens: A Mechanism for Proliferation and Perpetuation of Polygynous Red Imported Fire Ants, Solenopsis invicta Buren. *Florida Entomol* **71**:581. doi:10.2307/3495016

Haney BR. 2017. Ecological Drivers and Reproductive Consequences of Queen Cooperation in the California Harvester Ant Pogonomyrmex Californicus. Arizona State University, Tempe, USA.

Haney BR, Fewell JH. 2018. Ecological drivers and reproductive consequences of non-kin cooperation by ant queens. *Oecologia* **187**:643–655. doi:10.1007/s00442-018-4148-9

Helmkampf M, Mikheyev AS, Kang Y, Fewell J, Gadau J. 2016. Gene expression and variation in social aggression by queens of the harvester ant Pogonomyrmex californicus. *Mol Ecol* **25**:3716–3730. doi:10.1111/mec.13700

Helms KR, Newman NJ, Helms Cahan S. 2013. Regional variation in queen and worker aggression in incipient colonies of the desert ant Messor pergandei. *Behav Ecol Sociobiol* **67**:1563–1573. doi:10.1007/s00265-013-1568-8

Herbers JM. 1986. Nest site limitation and facultative polygyny in the ant Leptothorax longispinosus. *Behav Ecol Sociobiol* **19**:115–122. doi:10.1007/BF00299946

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA, Stiffler N, Bassham S, Hohenlohe PA. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genet* **6**:e1000862. doi:10.1371/journal.pgen.1000862

Hölldobler B, Wilson EO. 1977. The number of queens: An important trait in ant evolution. *Naturwissenschaften* **64**:8–15. doi:10.1007/BF00439886

Inbar S, Cohen P, Yahav T, Privman E. 2020. Comparative study of population genomic approaches for mapping colony-level traits. *PLOS Comput Biol* **16**:e1007653.

doi:10.1371/journal.pcbi.1007653

Johnson RA. 2004. Colony founding by pleometrosis in the semiclaustral seed-harvester ant Pogonomyrmex californicus (Hymenoptera: Formicidae). *Anim Behav* **68**:1189–1200. doi:10.1016/j.anbehav.2003.11.021

Johnson RA. 2002. Semi-claustral colony founding in the seed-harvester ant Pogonomyrmex californicus: A comparative analysis of colony founding strategies. *Oecologia* **132**:60–67. doi:10.1007/s00442-002-0960-2

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

Jones RT, Salazar PA, ffrench-Constant RH, Jiggins CD, Joron M. 2012. Evolution of a mimicry supergene from a multilocus architecture. *Proc R Soc B Biol Sci* **279**:316–325. doi:10.1098/RSPB.2011.0882

Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, Baxter S, Abanto M, Bermingham E, Humphray SJ, Rogers J, Beasley H, Barlow K, Ffrench-Constant RH, Mallet J, McMillan WO, Jiggins CD. 2006. A conserved supergene locus controls colour pattern diversity in Heliconius butterflies. *PLoS Biol* **4**:1831–1840. doi:10.1371/journal.pbio.0040303

Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. 2016. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res* **44**:89. doi:10.1093/nar/gkw092

Keller L. 1993. The Assessment of Reproductive Success of Queens in Ants and Other Social Insects. *Oikos* **67**:177. doi:10.2307/3545107

Keller L, Ross KG. 1998. Selfish genes: A green beard in the red fire ant. *Nature* **394**:573–575. doi:10.1038/29064

Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM. 2006. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res* **16**:980–989. doi:10.1101/gr.5157306

Kent T V, Uzunović J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Philos Trans R Soc B Biol Sci*. doi:10.1098/rstb.2016.0458

King JR, Tschinkel WR. 2006. Experimental evidence that the introduced fire ant, Solenopsis invicta, does not competitively suppress co-occurring ants in a disturbed habitat. *J Anim Ecol* **75**. doi:10.1111/j.1365-2656.2006.01161.x

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ. 2011. Genomic differentiation between temperate and tropical Australian populations of Drosophila melanogaster. *Genetics* **187**:245–260. doi:10.1534/genetics.110.123059

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Res* **27**:722–736. doi:10.1101/gr.215087.116

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

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of Population Structure using Dense Haplotype Data. *PLoS Genet* **8**:1002453. doi:10.1371/journal.pgen.1002453

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

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. doi:10.1093/bioinformatics/btp324

Lyko F. 2018. The DNA methyltransferase family: A versatile toolkit for epigenetic regulation. *Nat Rev Genet*. doi:10.1038/nrg.2017.80

Matschiner M, Barth JMI, Torresen OK, Star B, Baalsrud HT, Brieuc MSO, Pampoulie C, Bradbury I, Jakobsen KS, Jentoft S. 2021. Origin and fate of supergenes in Atlantic cod. *bioRxiv* 1–34. doi:10.1101/2021.02.28.433253

Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, Quail MA, Joron M, Ffrench-Constant RH, Blaxter ML, Mallet J, Jiggins CD. 2012. Genomic islands of divergence in hybridizing heliconius butterflies identified by large-scale targeted sequencing. *Philos Trans R Soc B Biol Sci* **367**:343–353. doi:10.1098/rstb.2011.0198

Odling-Smee PJ, Laland KN, Feldman MW. 1996. Niche construction. *Am Nat* **147**:641–648.

doi:10.1086/285870

Overson R, Fewell JH, Gadau J. 2016. Distribution and origin of intraspecific social variation in the California harvester ant Pogonomyrmex californicus. *Insectes Soc* **63**:531–541. doi:10.1007/s00040-016-0497-8

Overson R, Gadau J, Clark RM, Pratt SC, Fewell JH. 2014. Behavioral transitions with the evolution of cooperative nest founding by harvester ant queens. *Behav Ecol Sociobiol* **68**:21–30. doi:10.1007/s00265-013-1618-2

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

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

Pracana R, Priyam A, Levantis I, Nichols RA, Wurm Y. 2017. The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. *Mol Ecol* **26**:2864–2879. doi:10.1111/mec.14054

Purcell J, Brelsford A, Wurm Y, Perrin N, Chapuisat M. 2014. Convergent genetic architecture underlies social organization in ants. *Curr Biol* **24**:2728–2732. doi:10.1016/j.cub.2014.09.071

Purcell J, Lagunas-Robles G, Rabeling C, Borowiec ML, Brelsford A. 2021. The maintenance of polymorphism in an ancient social supergene. *Mol Ecol* **00**:1– 13. doi:10.1111/mec.16196

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

Rissing SW, Johnson RA, Martin JW. 2000. Colony Founding Behavior of Some Desert Ants: Geographic Variation in Metrosis. *Psyche A J Entomol* **103**:95–101. doi:10.1155/2000/20135

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

Rosengren R, Sundström L, Fortelius W. 1993. 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. pp. 308–333.

Ross KG. 1997. Multilocus evolution in fire ants: Effects of selection, gene flow and recombination. *Genetics* **145**. doi:10.1093/genetics/145.4.961

Ross KG, Keller L. 1998. Genetic control of social organization in an ant. *Proc Natl Acad Sci* **95**:14232–14237. doi:10.1073/pnas.95.24.14232

Ross KG, Krieger MJB, Keller L, Shoemaker DD. 2007. Genetic variation and structure in native populations of the fire ant Solenopsis invicta: Evolutionary and demographic implications. *Biol J Linn Soc* **92**:541–560. doi:10.1111/j.1095-8312.2007.00853.x

Rossner R, Kaeberlein M, Leiser SF. 2017. Flavin-containing monooxygenases in aging and disease: Emerging roles for ancient enzymes. *J Biol Chem*. doi:10.1074/jbc.R117.779678

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao Hongbin, Zhao Hui, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Yan, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Yayun, Sun W, Wang Haifeng, Wang Yi, Wang Ying, Xiong X, Xu L, Waye MMY, Tsui SKW, Xue H, Wong JTF, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, You QS, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P,

967 Bentley DR, Daly MJ, De Bakker PIW, Barrett J, Chretien YR, Maller J, McCarroll S,
968 Patterson N, Pe'Er I, Price A, Purcell S, Richter DJ, Saxena R, Sham PC, Stein LD,
969 Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ,
970 Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean
971 G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer
972 C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Johnson
973 TA, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Matsuda I, Fukushima Y, MacEr DR,
974 Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C,
975 Royal CDM, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF,
976 Knoppers BM, Foster MW, Clayton EW, Watkin J, Muzny D, Nazareth L, Sodergren E,
977 Weinstock GM, Yakub I, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP,
978 Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen
979 Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K,
980 Wang Hongguang, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE,
981 Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS,
982 Kennedy K, Jamieson R, Stewart J. 2007. Genome-wide detection and characterization of
983 positive selection in human populations. *Nature* **449**:913–918. doi:10.1038/nature06250
984 Saltz JB. 2011. Natural genetic variation in social environment choice: Context-dependent gene-
985 environment correlation in Drosophila melanogaster. *Evolution (N Y)* **65**. doi:10.1111/j.1558-
986 5646.2011.01295.x
987 Saltz JB, Geiger AP, Anderson R, Johnson B, Marren R. 2016. What, if anything, is a social
988 niche? *Evol Ecol* **30**:349–364. doi:10.1007/s10682-015-9792-5
989 Saltz JB, Nuzhdin S V. 2014. Genetic variation in niche construction: Implications for
990 development and evolutionary genetics. *Trends Ecol Evol*. doi:10.1016/j.tree.2013.09.011
991 Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple
992 genome sequences. *Nat Genet Vol* **46**. doi:10.1038/ng.3015
993 Schiffels S, Wang K. 2020. MSMC and MSMC2: The multiple sequentially Markovian
994 coalescentMethods in Molecular Biology. Humana Press Inc. pp. 147–166.
995 doi:10.1007/978-1-0716-0199-0_7
996 Schwander T, Libbrecht R, Keller L. 2014. Supergenes and complex phenotypes. *Curr Biol*.
997 doi:10.1016/j.cub.2014.01.056
998 Seppä P, Sundström L, Punttila P. 1995. Facultative polygyny and habitat succession in boreal
999 ants. *Biol J Linn Soc* **56**:533–551. doi:10.1111/j.1095-8312.1995.tb01109.x
1000 Shaffer Z, Sasaki T, Haney B, Janssen M, Pratt SC, Fewell JH. 2016. The foundress's dilemma:
1001 Group selection for cooperation among queens of the harvester ant, Pogonomyrmex
1002 californicus. *Sci Rep* **6**:1–9. doi:10.1038/srep29828
1003 Shin J-H, Blay S, Mcneney B, Graham J. 2006. LDheatmap: An R Function for Graphical Display
1004 of Pairwise Linkage Disequilibria between Single Nucleotide Polymorphisms. *JSS J Stat*
1005 *Softw* **16**. doi:10.18637/jss.v000.i00
1006 Shoemaker DDW, Ahrens ME, Ross KG. 2006. Molecular phylogeny of fire ants of the
1007 Solenopsis saevissima species-group based on mtDNA sequences. *Mol Phylogenet Evol*
1008 **38**. doi:10.1016/j.ympev.2005.07.014
1009 Sirviö A, Pamilo P, Johnson RA, Page RE, Gadau J. 2011. Origin and evolution of the dependent
1010 lineages in the genetic caste determination system of pogonomyrmex ants. *Evolution (N Y)*
1011 **65**:869–884. doi:10.1111/j.1558-5646.2010.01170.x
1012 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
1013 phylogenies. *Bioinformatics* **30**:1312–1313. doi:10.1093/bioinformatics/btu033
1014 Tang K, Thornton KR, Stoneking M. 2007. A New Approach for Using Genome Scans to Detect
1015 Recent Positive Selection in the Human Genome. *PLoS Biol* **5**:e171.
1016 doi:10.1371/journal.pbio.0050171
1017 Thompson MJ, Jiggins CD. 2014. Supergenes and their role in evolution. *Heredity (Edinb)*.
1018 doi:10.1038/hdy.2014.20
1019 Thompson MJ, Timmermans MJTN. 2014. Characterising the Phenotypic Diversity of Papilio
1020 dardanus Wing Patterns Using an Extensive Museum Collection. *PLoS One* **9**:e96815.
1021 doi:10.1371/JOURNAL.PONE.0096815
1022 Tollefson GA, Schuster J, Gelin F, Agudelo A, Ragavendran A, Restrepo I, Stey P, Padbury J,

Uzun A. 2019. VIVA (VIsualization of VAriants): A VCF File Visualization Tool. *Sci Rep* **9**:1–7. doi:10.1038/s41598-019-49114-z

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella K V., Altshuler D, Gabriel S, DePristo MA. 2013. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr Protoc Bioinforma* **43**:11.10.1-11.10.33. doi:10.1002/0471250953.bi1110s43

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One* **9**:e112963. doi:10.1371/journal.pone.0112963

Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang Y-C, Shoemaker D, Keller L. 2013. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**:664–668. doi:10.1038/nature11832

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

Wilson EO. 1971. The insects societies., Heredity.

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

Yan Z, Martin SH, Gotzek D, Arsenault S V., Duchen P, Helleu Q, Riba-Grognuz O, Hunt BG, Salamin N, Shoemaker DW, Ross KG, Keller L. 2020. Evolution of a supergene that regulates a trans-species social polymorphism. *Nat Ecol Evol* **4**:240–249. doi:10.1038/s41559-019-1081-1

Yang S, Wang L, Huang J, Zhang X, Yuan Y, Chen JQ, Hurst LD, Tian D. 2015. Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* **523**:463–467. doi:10.1038/nature14649

Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc Natl Acad Sci* **110**:E1743–E1751. doi:10.1073/PNAS.1219381110