Received: 31 December 2021 | Revised: 15 March 2022 | Accepted: 15 May 2022

DOI: 10.1111/jeb.14038

JOURNAL OF Evolutionary Biology eseb

RESEARCH ARTICLE

A socially polymorphic *Formica* ant species exhibits a novel distribution of social supergene genotypes

Daniel Pierce[1] | Penglin Sun[2] | Jessica Purcell[2] | Alan Brelsford[1]

[1]Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, California, USA

[2]Department of Entomology, University of California, Riverside, California, USA

Correspondence
Jessica Purcell, Department of Entomology, University of California, Riverside 900 University Ave., Riverside, California 92521, USA.
Email: jpurcell@ucr.edu

Alan Brelsford, Department of Evolution, Ecology, and Organismal Biology, University of California, 900 University Ave, Riverside, California 92521, USA.
Email: alan.brelsford@ucr.edu

Funding information
Division of Environmental Biology, Grant/Award Number: 1754834 and 1942252; National Institute of Food and Agriculture, Grant/Award Number: CA-R-ENT-5126-H

Abstract

Most supergenes discovered so far are young, occurring in one species or a few closely related species. An ancient supergene in the ant genus *Formica* presents an unusual opportunity to compare supergene-associated phenotypes and the factors that influence the persistence of polymorphism in different species. We investigate the genetic architecture of social organization in *Formica francoeuri*, an ant species native to low- and mid-elevation semiarid regions of southern California, and describe an efficient technique for estimating mode of social organization using population genomic data. Using this technique, we show that *F. francoeuri* exhibits polymorphism in colony social organization and that the phenotypic polymorphism is strongly associated with genotypes within the *Formica* social supergene region. The distribution of supergene haplotypes in *F. francoeuri* differs from that of related species *Formica selysi* in that colonies with multiple queens contain almost exclusively workers that are heterozygous for alternative supergene haplotypes. Moreover, heterozygous workers exhibit allele-specific expression of the polygyne-associated haplotype at the candidate gene *Knockout*, which is thought to influence social organization. We also report geographic population structure and variation in worker size across a large fraction of the species range. Our results suggest that, although the *Formica* supergene is conserved within the genus, the mechanisms that maintain the supergene and its associated polymorphisms differ among species.

KEYWORDS
coadapted gene complex, Formicinae, polygyny syndrome, population genetics, queen number

1 | INTRODUCTION

Recombination allows deleterious alleles to be efficiently purged by selection but counteracts fixation of advantageous combinations. By suppressing recombination between haplotypes, inversions can maintain tightly linked groups of genes, or 'supergenes', that contribute to suites of complex morphological and behavioural phenotypes (Charlesworth, 2016; Schwander, Libbrecht, & Keller 2014). Investigating the origin, evolution and maintenance of these supergenes may provide insight into the processes that promote adaptation and generate biological diversity.

Inversion-based supergenes are known to underlie polymorphisms in traits with potentially large contributions to fitness, such as reproductive strategy (Kupper et al., 2016; Tuttle et al., 2016),

anti-predator defences (Joron et al., 2011) and organization of insect societies (Purcell, Brelsford, Wurm, Perrin, & Chapuisat, 2014; Wang et al., 2013). There is some evidence that balancing selection on the traits determined by alternative haplotypes can facilitate the persistence of supergenes, as in *Heliconius* butterflies (Joron et al., 2011), ruffs (Kupper et al., 2016) and white-throated sparrows (Tuttle et al., 2016). Some supergene haplotypes act as selfish genetic elements, propagating themselves at the expense of alternative haplotypes. In fire ants (*Solenopsis invicta*), for example, one supergene haplotype is favoured via a green-beard effect (Keller & Ross, 1998) and meiotic drive also leads to transmission ratio distortion of supergene haplotypes (Ross & Shoemaker, 2018). In the *Formica selysi* supergene, a different transmission ratio distortion mechanism, a lethal maternal effect, causes only the offspring with the selfish supergene haplotype to survive when their mother has the selfish haplotype (Avril, Purcell, Béniguel, & Chapuisat, 2020). The fact that supergene polymorphisms persist in these systems suggests that the spread of selfish genetic elements is countered by some form of negative frequency dependent selection (Larracuente & Presgraves, 2012; Huang & Wang, 2014; Wang et al., 2013; Lyon, 2003; Keller & Ross, 1998).

In many supergene systems, the inverted haplotype rarely or never occurs in the homozygous state and frequently shows evidence of degeneration (Jay et al., 2021; Stolle et al., 2019; Tuttle et al., 2016), leading to the observation that supergene evolution resembles the evolution of heteromorphic sex chromosomes (Tuttle et al., 2016; Wang et al., 2013). It is less clear whether degeneration in autosomal supergenes wherein homozygotes with both haplotypes are present would occur and, if present, what would trigger the initial degeneration of one haplotype and/or lead to the lower fitness of homozygous individuals bearing that haplotype (Schwander, Libbrecht & Keller, 2014).

In ants, supergenes associated with social organization have independently evolved at least twice: in *Solenopsis* and *Formica* (Purcell et al. 2014; Wang et al. 2013). In both cases, one supergene haplotype is associated with single-queen colonies (monogyne) and the alternative haplotype is associated with colonies harbouring multiple reproductive queens (polygyne). The supergene architecture in *Formica* appears to have persisted through much of the radiation of the genus (Purcell, Lagunas-Robles, Rabeling, Borowiec, & Brelsford, 2021; Brelsford et al., 2020). In addition to controlling social organization, the supergenes of both *S. invicta* and *F. selysi* are associated with a suite of other characteristics, known as the 'polygyny syndrome' (sensu Keller, 1993). These traits include worker and queen body size, queen fecundity, and colony size (Ross & Keller, 1995; Rosset & Chapuisat, 2007; Schwander, Rosset, & Chapuisat, 2005), and could contribute to the maintenance of polymorphism via frequency dependent or spatially variable selection. Although there are similarities between the convergent *Formica* and *Solenopsis* supergenes, some key differences suggest that the polymorphisms may be maintained through distinct mechanisms. The supergene of *S. invicta* shows some hallmarks of degeneration, including a stark deficit of homozygotes for the polygyne-associated haplotype and an increase in repetitive sequences (Stolle et al., 2019). In contrast, the distribution of supergene genotypes in *F. selysi* indicates that workers and queens homozygous for the polygyne-associated haplotype are present at high frequency (Avril, Purcell, Brelsford, & Chapuisat, 2019; Purcell et al. 2014). Understanding the distribution of supergene genotypes in colonies and their association with individual- and colony-level phenotypes sheds light on the role of recombination (both within and between haplotypes) and degeneration on the fate of these autosomal supergenes. Nevertheless, little is known about how alternative haplotypes function in other congeneric *Formica* and *Solenopsis* species.

In the present study, we investigate social organization and its genetic underpinnings in the species *Formica francoeuri*. We address four complementary questions by sequencing workers from dozens of colonies sampled across a large fraction of the species range. First, we ask whether this species exhibits social polymorphism by inferring the number of queens per colony using genome-wide single nucleotide polymorphism (SNP) data. Second, we investigate whether variation in colony social structure is linked to variation on the *Formica* social supergene using two complementary approaches. We carry out a genome-wide association study to assess markers associated with variation in colony social structure, and we examine allele-specific expression patterns at one highly conserved candidate gene with alleles associated with single- and multiple-queen colonies in other *Formica* species (Brelsford et al. 2020). Third, we examine variation in worker size and ask whether size varies consistently with queen number. Finally, we assess the population structure in this little-studied xeric species.

2 | METHODS

2.1 | Sampling and study species

Formica francoeuri inhabits arid and semiarid regions of southern California from sea level up to about 1500 m. This species nests mainly in sandy soils, often in dry riverbeds. We sampled 8 workers from each of 76 colonies of *F. francoeuri* at 11 localities between Ventura and San Diego, CA (Table S1). To ensure the nest identity of the collected workers, we selected only workers seen exiting the nest or within 20 cm of the nest entrance and sampled nests at least three metres apart. We stored the sampled workers in 100% ethanol.

2.2 | Laboratory methods

DNA extraction

We extracted DNA from 8 workers from each colony using the Qiagen DNeasy Blood & Tissue Kit protocol for insect tissue with several custom modifications (Purcell et al., 2014). Briefly, we detached the head and thorax of each worker, which we then placed

in a 1.7 mL centrifuge tube for pulverization in liquid nitrogen and subsequent extraction. We eluted purified DNA with 30 μL of AE buffer rather than 200 μL per the protocol to obtain higher DNA concentrations.

RADseq library preparation

We performed restriction site-associated DNA sequencing (RADseq) following the method of Brelsford et al. (2016), which incorporated elements of other established protocols (Parchman, Gompert, Mudge, Schilkey, Benkman, & Buerkle, 2012; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). To digest the DNA, we used SbfI and MseI restriction enzymes. We then ligated MseI adaptors and uniquely barcoded SbfI adaptors for each sample. We purified DNA using Sera-Mag magnetic beads (Rohland & Reich, 2012) in a 0.8:1 ratio (beads: sample solution). We divided each sample into four replicates in separate wells, amplified the fragments by PCR with indexed Illumina primers and then pooled the replicate PCR products for each sample for a final PCR cycle with added primer. We ran the amplified DNA fragments in a 1.5% agarose gel and excluded samples that failed to amplify from subsequent pooling. We pooled each indexed library of up to 96 samples and removed small DNA fragments and adapter dimer with Sera-Mag magnetic beads (0.8:1 ratio). Our libraries were sequenced on an Illumina HiSeq 4000 at the University of California Berkeley sequencing core facility. The *F. francoeuri* samples were sequenced as a part of two libraries comprised of up to 2725 samples for this and other projects. We obtained 150 bp paired-end reads for 290 individuals and 150 bp single-end reads for 310 individuals of the focal species. We removed individuals with a high proportion of missing data (≥30%). We retained samples from colonies with at least six sequenced workers after quality filtering, leaving a total of 582 workers for our genetic analyses.

2.3 | Genotype and expression at *knockout*, a candidate gene for social structure

Primer development

Previous studies comparing the conserved regions of the *Formica* social supergene across species suggested that alternative alleles within one gene, *Knockout*, consistently differentiate monogyne and polygyne colonies (Brelsford et al., 2020, Purcell et al., 2021). To follow up on this discovery and to add another facet to our search for a possible supergene in *F. francoeuri*, we developed allele-specific primers for a variable portion of this gene based on alignment of *Knockout* sequences from monogyne- and polygyne-associated haplotypes of *F. selysi*, *Formica truncorum* and *Formica exsecta* (haplotypes are *Sm* and *Sp*, respectively; Table S2). We tested the specificity of these primers in a sample of 12 *F. francoeuri* workers that were initially assessed using our RADseq markers. This sample included five *Sm/Sm* homozygotes, six heterozygotes and one *Sp/Sp* homozygote.

RNA isolation and reverse transcription quantitative PCR

Workers collected from colonies in the socially polymorphic population from Lytle Creek, CA were used to assess expression of *Knockout* alleles. In 14 *Sm/Sm* workers and 22 *Sm/Sp* workers, we isolated total RNA using TRI Reagent (Molecular Research Center). We used DNase I (ThermoFisher Scientific) on 200 ng of total RNA to remove any residual DNA, and reverse-transcribed using oligo-dT and RevertAid Reverse Transcriptase following the manufacturer's protocol (ThermoFisher Scientific). To measure allele-specific expression of *Knockout*, we performed quantitative PCR (qPCR) with two or three technical replicates on a CFX96 Real-Time System (Bio-Rad) using 5 ng of cDNA for each technical replicate and Luna Universal qPCR Master Mix (NEB) following the manufacturer's protocol in 20 μL reactions. The reactions were run for 40 cycles, with an annealing temperature of 60°C, followed by the Melt Curve stage. Primer sequences are shown in Table S2. We used GAPDH as a control (Morandin, Havukainen, Kulmuni, Dhaygude, Trontti, & Helanterä, 2014). To calculate the ratio of the two target alleles' expression in each sample, we used the 'delta–delta Ct' method. We performed a control without reverse transcriptase for each sample by qPCR to confirm that there was no DNA contamination in the RNA sample.

2.4 | Data analysis

2.4.1 | Data preparation

We used *Stacks* v2.1 to demultiplex our data (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013), *PEAR* v0.9.10 (Zhang, Kobert, Flouri, & Stamatakis, 2014) to merge paired-end reads and remove adaptor sequences, and *Bowtie2* v2.3.4.1 (Langmead & Salzberg, 2012) to align reads to the *Formica selysi* genome (Brelsford et al., 2020). We obtained a read mapping rate of 91.6% from our alignment to the *F. selysi* reference genome. We called SNPs using *SAMtools mpileup* v1.9 (Li, 2011), and filtered the genotypes for a minimum genotype quality score of 20, a minimum read depth of 1, a minor allele count of 3, and excluded indels and sites with over 20% missing data using *VCFtools* v0.1.13 (Danecek et al. 2011). We produced a catalogue of 955 SNPs with average read depth of 86× per individual obtained from this filtering process to use for further analysis.

To obtain data suitable for a Genome Wide Association Study (GWAS) using *Gemma* v0.94 (Zhou & Stephens, 2012), we imputed missing genotypes using the full dataset of SNPs with *Beagle* v4.1 (Browning & Browning, 2016). Although the full data contains genotype calls that do not meet our quality standards, it provides the

maximum available information for imputation for the best estimation of missing genotypes at loci that pass the quality filters. *Gemma* requires that no missing genotypes are present in the data. Once imputation was completed, we filtered the imputed data to retain only loci that passed our pre-imputation quality filters.

2.4.2 | Determination of colony social organization

To infer the social organization of a colony from our sequence data, we first calculated metrics for identity-by-descent (IBD) and opposing homozygosity for each colony, then examined the empirical distributions of these metrics together in order to determine the putative social organization for each colony. This method is based on the following assumptions: (1) in a single-queen colony, all workers are sisters so the average length of genomic segments that are identical-by-descent (IBD) will be larger compared to a colony where workers descend from multiple queens. This is because the genomes of workers from different mothers are more generations removed from their most recent common ancestor, and their genomes are products of generations of recombination that breaks apart IBD segments (Thompson, 2013). The degree of relationship between mothers in a polygyne colony will affect the length of IBD segments detected. (2) In a colony with one singly mated queen, there should be no instances of loci that are represented by homozygotes for two alternative alleles in full-sisters, except due to genotyping errors. On the contrary, homozygotes for two alternative alleles will be common in a colony comprised of multiple queens. In reality, measures of opposing homozygosity are likely to be imperfect for several reasons. Polyandry is known to occur in *Formica* (Pamilo, 1982). A single polyandrous queen will produce offspring that have intermediate levels of opposing homozygosity (i.e. higher than offspring of a single monandrous queen and lower than offspring of multiple queens). In parallel, IBD will be low in offspring of polyandrous single queens relative to a colony produced by a monogynous, monandrous queen. Sampling errors (where a roaming worker from another colony is sampled from the intended colony) and genotyping errors will also increase opposing homozygosity and decrease IBD.

For IBD, we used *Beagle* v4.1 (Browning & Browning, 2013) to infer the length of segments identical-by-descent between pairs of workers within a colony, using a LOD score of 2 (allowing a 1 in 100 chance of false positives). These values were then summed and divided by the number of pairs of workers for each colony, to obtain an intra-colony measure of IBD segment length. For opposing homozygosity, we counted the number of loci for which homozygotes were present for both the reference and alternative alleles within a colony, for bi-allelic SNPs. Colonies with high IBD and low opposing homozygosity are likely monogyne, while colonies that exhibit low IBD and high opposing homozygosity are likely polygyne. The empirical distribution of both IBD and opposing homozygosity is expected to be bimodal for a data set composed of monogyne and polygyne colonies. We plotted histograms for both metrics, identified low points between modes (opposing homozygosity = 30, mean pairwise IBD = 5 Mbp) and used the values corresponding to bins that contain these low points as the thresholds for distinguishing putative monogyne and polygyne colonies.

As an additional measure to confirm our assessment of social organization for each colony, we estimated pairwise relatedness between workers in each colony. In summary, we pruned the SNP dataset, using *Plink* v1.90b3.38, to exclude one SNP from a pair of loci in close proximity (within 200 bp) with an $r^2 > 0.1$, resulting in 709 loci of the original 955. Examining the distribution of r^2 values for pairs of loci revealed that very few exhibited a strong correlation (the $r^2 > 0.1$ threshold bounds the top 3% of r^2 values for pairs of loci) between alleles and that nearly all of the loci in these pairs are within 200 bp. We then estimated pairwise relatedness among workers (specifying the Huang estimator) using the program *polyrelatedness* v1.11b (Huang, Ritland, Dunn, Qi, Guo, & Li, 2016), and we compared these results to our assessment of social organization based on opposing homozygosity and IBD.

2.4.3 | Genome-wide association study (GWAS)

To identify genomic regions associated with colony social structure, we fit a linear mixed model to the data using *Gemma* v0.94 (Zhou & Stephens, 2012). *Gemma* uses a relatedness matrix generated from the sample genetic data to correct for non-independence of the samples due to population structure. In this case, population structure may be due to geographic structure and sampling that includes related nestmates within a colony. Only colonies that could be clearly identified as polygyne or monogyne were used in the GWAS. To correct for multiple testing within our SNP data set, we applied a false discovery rate adjustment (Benjamini & Hochberg, 1995) to the *p*-values obtained from *Gemma* using the R package *p.adjust*.

2.4.4 | Allele-biased expression of *knockout*

We carried out two analyses on qPCR data. In the first analysis, we compared the expression levels of the *Sm* allele of *Knockout* in workers with the *Sm/Sm* and *Sm/Sp* genotypes. We calculated fold change of the *Sm* allele in comparison with the GAPDH expression for each individual, and we formally compared log-transformed fold change between the two genotype groups using a linear model with fold change as the response variable, genotype as the explanatory variable and qPCR batch as a random effect using the lmer command in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R v3.6.1 (R Core Team, 2019). In the second analysis, we compared fold change expression of the *Sp* allele of *Knockout* to *Sm* expression in *Sm/Sp* heterozygous workers. We used a one-sample *t*-test to compare the log-transformed fold change of *Sp* to the expected expression level of 1.4 relative to the *Sm* allele (if the two alleles are expressed in equal proportions). This expected value was calculated from primer efficiency

tests, which showed slightly greater *Sp* primer efficiency (89.83% and 99.97% primer efficiencies for *Sm* and *Sp* primers, respectively).

2.4.5 | Morphometrics

To assess whether *Formica francoeuri* exhibits the consistent worker size variation expected based on the polygyny syndrome (Keller, 1993), we tested for a difference in mean worker size by social form via a mixed effects model implemented in the package lme4 of R v3.4.0 (Bates et al., 2015; R Core Team, 2019). We used head width as a proxy for worker size as this measure has a strong positive correlation with other body dimensions in *Formica* species (Schwander et al., 2005; Tawdros, West, & Purcell, 2020). We used a *Leica* S8APO stereomicroscope and *Leica* DMC2900 camera to photograph the anterior view of the head at 25× magnification. Using *Leica Application Suite* v4.6.2 software, we measured a distance line at the maximum width across the eyes in 495 workers from 57 colonies. Most colonies were assigned to social structure based on our initial analysis of IBD and opposing homozygosity. We included additional colonies with intermediate IBD or opposing homozygosity based on the presence or absence of the *Sp* supergene haplotype, which was tightly associated with polygyne social structure. We used a linear mixed model to test whether social structure predicted worker size. The model we constructed included worker size as the response variable, social structure as the explanatory variable, and genetic cluster (identified by our Admixture analysis; see Population Structure section) and colony nested within locality as random effects. We performed a likelihood ratio test to determine whether the two random effects, genetic cluster and colony nested within locality, have a significant effect on worker size. To perform the likelihood ratio tests, we compared the full model to alternative models with one of the random effect variables excluded.

2.4.6 | Population structure

To assess the genetic structure of sampled populations, we selected the worker with the least missing data from each colony to carry out a principal component analysis implemented in *Plink* v1.90b3.38 (Purcell et al., 2007) and plotted the first two principal components in R v3.4.0. Using the same dataset, we used *ADMIXTURE* v1.3.0 (Alexander, Novembre, & Lange, 2009) to infer genetic clusters in our dataset for *K* values from 2 to 10.

3 | RESULTS

3.1 | Colony social organization is strongly associated with SNPs on chromosome 3

By combining an assessment of opposing homozygosity and identity-by-descent among nestmates, we infer that 24 colonies are monogynous, 46 are polygynous, and we were unable to assign six colonies a putative phenotype (Figure 1a). Within-colony opposing homozygosity appears bimodal with modes at 0–6 and 60–66 SNPs. The distribution of values for IBD segment length also appears bimodal with modes at 0–1 Mbp and 10–11 Mbp, although the second mode appears much weaker (Figure 1a).

We discovered five SNPs significantly associated with inferred social form via GWAS ($\alpha = 0.05$). *p*-Values for four of these loci are well below the significance threshold at $\alpha = 0.01$ (Figure 2, Table 1). All four of these loci can be mapped to the social supergene region in *F. selysi*, which is strongly associated with social structure (Purcell et al. 2014).

Among the four SNPs with the strongest association with social structure, genotype distributions reveal that workers from monogyne colonies are exclusively *Sm/Sm*, while the polygyne colonies are almost exclusively *Sm/Sp*. We find that among all workers, only one was found with the *Sp/Sp* genotype—this worker was sampled from a polygyne colony (Figures 3, S1).

Our estimates of pairwise relatedness for each colony are shown in Figures 1 and S1. The average pairwise relatedness of colonies categorized as monogyne is 0.648 (SD = 0.197), and the average relatedness of colonies categorized as polygyne is 0.17 (SD = 0.134). The median relatedness between workers from different colonies within the same locality was 0.13 and between workers from different localities was −0.17, showing that geographic population structure inflates pairwise relatedness estimates within locality.

3.2 | *Knockout* gene expression is higher in polygyne-associated haplotype

A recent study revealed that *F. francoeuri* has two distinct haplotypes at the gene *Knockout*, which is located on the social supergene in *F. selysi* (Purcell et al., 2021; Brelsford et al. 2020). We refer to the monogyne-associated variant as *Knockout Sm* and the polygyne-associated variant as *Knockout Sp*. We observed a significant difference in the expression of the *Knockout Sm* allele in comparing *Sm/Sm* homozygotes and *Sm/Sp* heterozygotes (Figure 4a). Specifically, expression of *Knockout Sm* relative to the control gene was approximately twice as high in homozygotes as in heterozygotes, as we would expect from the copy number difference between the two groups (linear mixed effects model $F_{1,\ 33.66} = 10.9$, $p = 0.0023$). Analysis of fold change in *Knockout Sp* expression relative to *Knockout Sm* expression in each heterozygous worker demonstrated higher expression of the *Sp* allele in all individuals, even after accounting for modest differences in primer efficiency (Figure 4b,c; one sample *t*-test $t_{21} = 9.36$, $p < 0.0001$; excluding outlier $t_{20} = 17.40$, $p < 0.0001$).

3.3 | Colony social organization does not predict worker size

We used morphological data for 495 individuals from 57 colonies with at least 6 workers. Social structure did not have a significant effect on worker size (linear mixed model $t_{30.88209} = -0.338$,

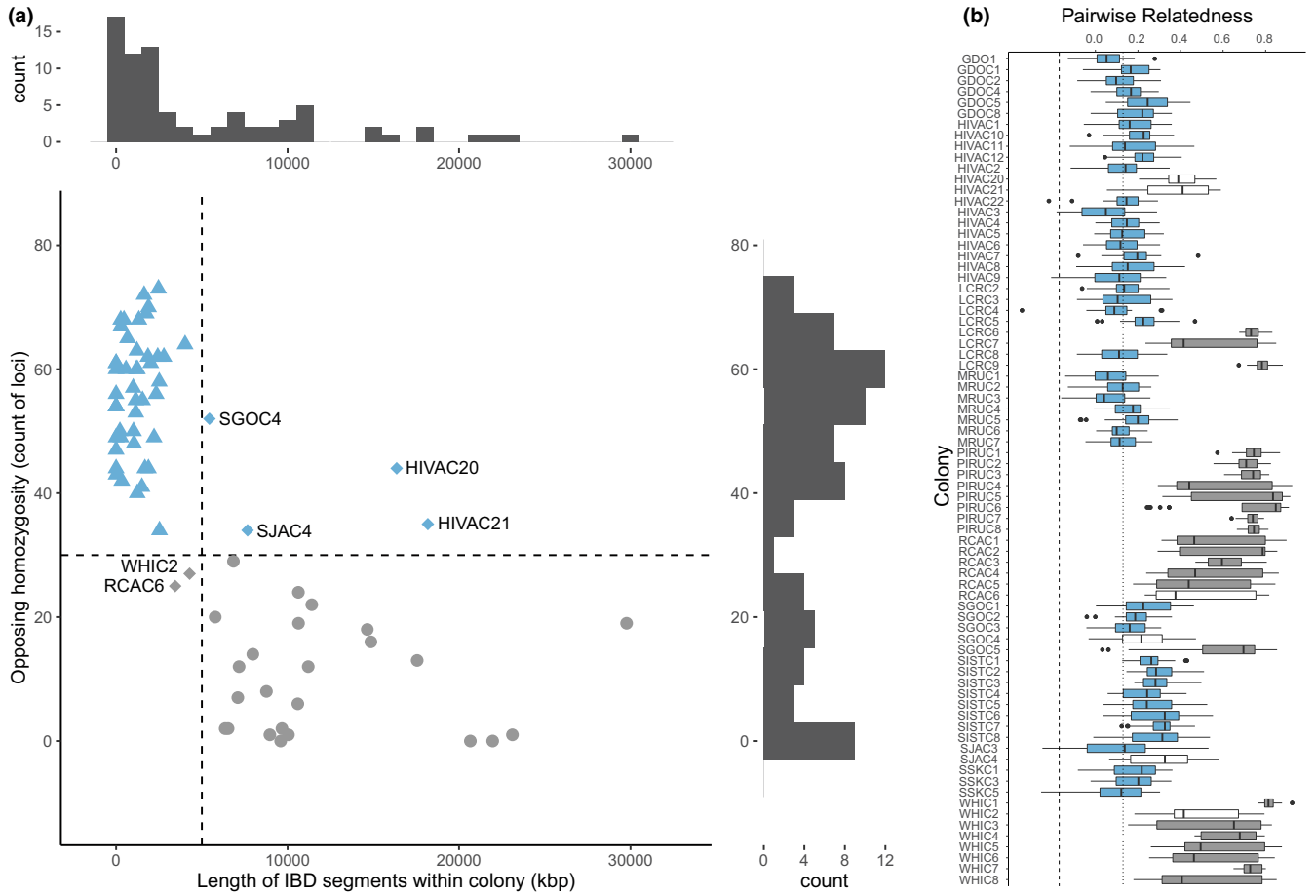

FIGURE 1 (a) Assessment of social organization by length of IBD segments and opposing homozygosity within a colony. Dashed lines represent cut-off values for assigning social form based on the distributions of values for opposing homozygosity (30) and IBD segment length (5 Mbp). Inferred monogynous and polygynous colonies are shown in the bottom right and top left quadrants, respectively, while the other two quadrants contain ambiguous colonies. Shapes represent inferred social organization based on IBD and opposing homozygosity (triangles represent polygynous colonies, circles represent monogynous colonies, and diamonds represent ambiguous colonies). Colour indicates social form determined by evaluating genotypes at the SNPs most significantly associated with inferred social organization; blue indicates polygynous colonies; and grey indicates monogynous colonies. (b) Pairwise relatedness values for each colony estimated by *polyrelatedness*. Colour represents mode of social organization based on opposing homozygosity and IBD in (a); blue indicates polygynous colonies, grey indicates monogynous colonies, and ambiguous colonies are shown in white. The dotted line represents the median pairwise relatedness of workers from different colonies within the same locality, and the dashed line represents the median pairwise relatedness of workers from different localities

p-value = 0.738; Figure 5c). However, model comparison indicated that worker size differed significantly based on the random effect colony nested within locality (likelihood ratio test $X^2_{4,6} = 73.4$, $p < 0.0001$; Figure 5a), but not based on genetic cluster (likelihood ratio test $X^2_{5,6} = 0$, $p = 1$; Figure 5b).

3.4 | Population structure is independent of social organization and supergene genotype

Principal components 1 and 2 explain 17.3 and 9.65 per cent of the variation, respectively, and separate the colonies into three genetic clusters (Figure 6b). Fifty-seven of the colonies we sampled grouped within a large cluster including the northernmost populations (lowest values on PC1). The second largest cluster is represented by 11 colonies from central populations (middle values on PC1), and the third cluster is represented by eight colonies from the southern coastal population (highest values on PC1).

Results from Admixture with $K = 3$ (lowest cross validation error, Figure S2) show population structure that corresponds with clustering in the PCA (Figure 6). Admixture results based on K values 2–5 are shown in Figure S3. Within genetic clusters and within some localities (i.e. LCR, HIVA and SJA), colonies with both forms of social organization are found.

4 | DISCUSSION

Recent evidence points to an ancient origin of some supergenes, with alternative haplotypes persisting through numerous speciation

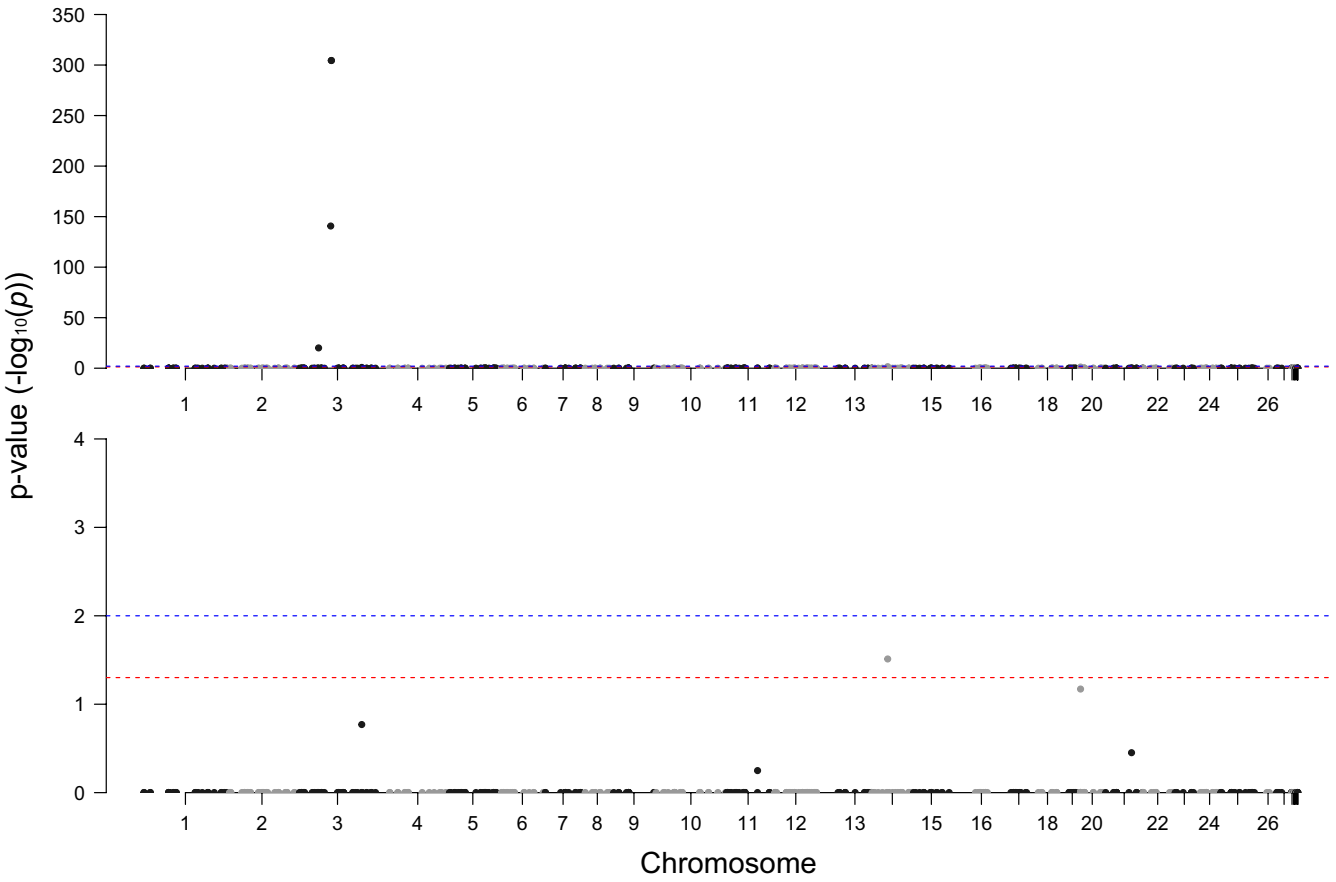

FIGURE 2 Manhattan plots show FDR adjusted *p*-values of GWAS for social organization. Red dashed lines represent the $\alpha = 0.05$ threshold for significance. Blue dashed lines represent the $\alpha = 0.01$ threshold for significance. (a) all SNPs (two of the four significant SNPs are tightly linked and have identical *p*-values). (b) all SNPs excluding the 4 most significantly associated SNPs on chromosome 3

TABLE 1 SNPs with the 10 lowest *p*-values reported by *GEMMA* and false discovery rate adjusted *p*-values

Chromosome	Position	*p*-Value	FDR adjusted *p*-value
3	6412530	1.26E-307	3.34E-305
3	6412531	1.26E-307	3.34E-305
3	6304361	1.13E-143	2.00E-141
3	4197683	6.23E-23	8.26E-21
14	2763363	2.90E-04	0.031
20	876259	7.63E-04	0.067
3	11689706	2.24E-03	0.17
21	4778164	5.34E-03	0.35
11	5551192	9.57E-03	0.56
1	527308	3.18E-01	0.997

events. However, we still know little about how the function of these haplotypes differs across species. In this study, we analysed genotype and gene expression data from colonies of *F. francoeuri* in order to determine whether this species exhibits polymorphism in colony social structure and, if so, whether the phenotype is associated with the *Formica* social supergene. This species is not closely related to other congeneric species with a known association between social polymorphism and supergene variation (Purcell et al., 2021), so our study greatly expands the phylogenetic context of research on supergene evolution in *Formica*.

Our results indicate that *F. francoeuri* is socially polymorphic and that social structure in this species is associated with SNPs on chromosome 3. However, in contrast to another well-studied species, *F. selysi*, we find some differences in how supergene genotypes are distributed within colonies and which phenotypes are influenced by the supergene. These findings hint at divergent processes that maintain a conserved supergene and the polymorphic phenotypes it influences in individual species.

The analyses of worker morphology, population genetic structure and associations of both with social structure suggest that environmental selection gradients may shape the geographic distribution of social structure and worker size separately in this species. Moreover, we uncover substantial population genetic structure in this species with a small geographic range.

4.1 | GWAS and the social supergene

The SNPs that show the strongest association with variation in social structure in *F. francoeuri* are localized on chromosome 3, which

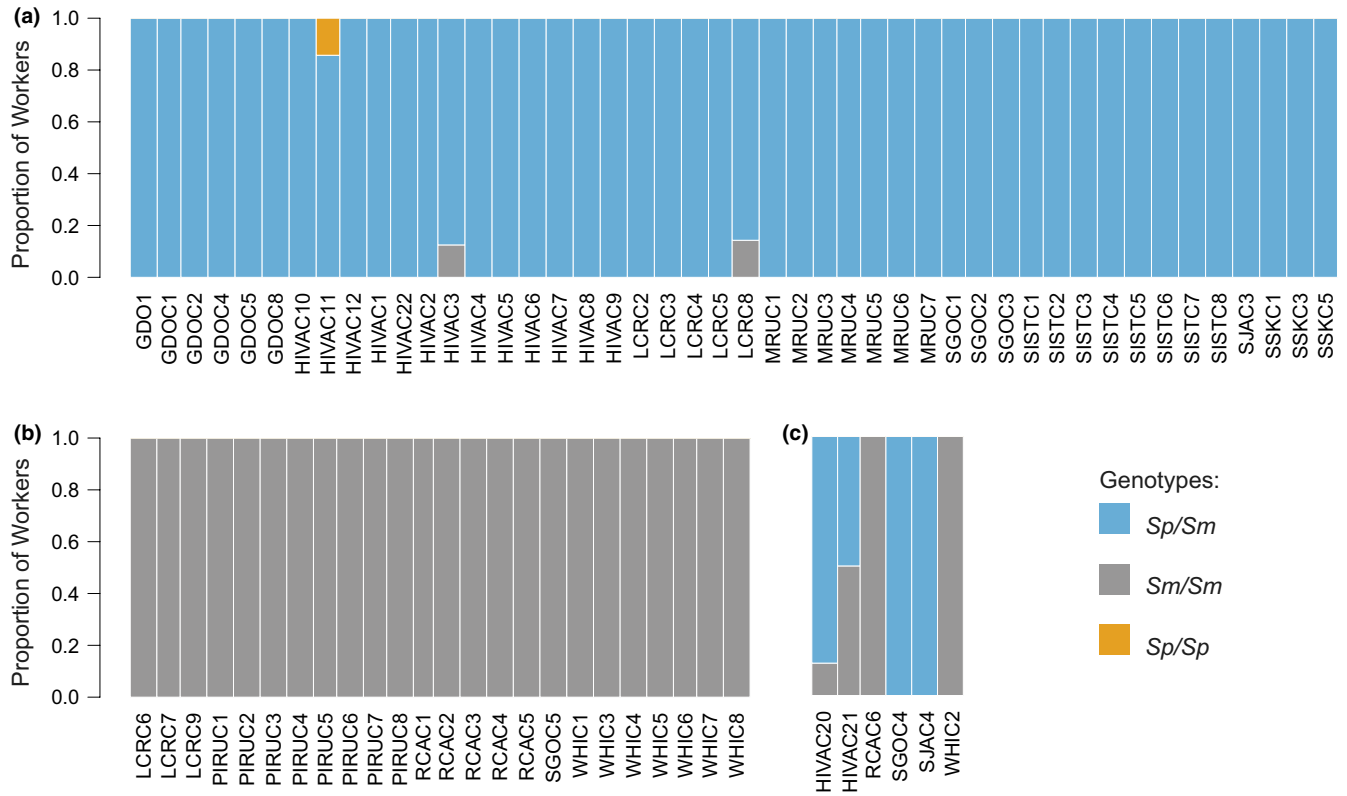

FIGURE 3 Genotype distributions at one of the most strongly associated loci (chromosome 03, position 6412530) for (a) polygyne colonies, (b) monogyne colonies and (c) ambiguous colonies. Colour indicates the number of *Sp* alleles; width of bars indicates the number of workers sequenced in a colony (6–8 workers in each colony)

harbours a supergene that influences social structure in other *Formica* species (Brelsford et al., 2020). We also find one SNP, on chromosome 14, that passes the significance thresholds. This may be a false positive or indicate a locus with a small effect on social structure. The strong association between alleles on chromosome 3 and inferred social structure suggests that *F. francoeuri* has a similar genetic basis for the social polymorphism as *F. selysi*. Lending further support, the *Sp* allele of the gene *Knockout* is expressed at a significantly higher level than the *Sm* allele in heterozygous *F. francoeri* workers. *Knockout* was identified as a candidate gene that may contribute to social organization based on conserved, haplotype-specific SNPs that are shared across many *Formica* species (Brelsford et al., 2020; Purcell et al., 2021). The low density of RADseq markers we obtained does not allow us to determine the boundaries of the supergene in *F. francoeuri*, but we can estimate that it extends from at least 4.2 to 11.9 Mbp on chromosome 3, defined by the SNPs strongly associated with social form in our GWAS and the position of *Knockout*. Although no reference genome is available for *F. francoeuri*, karyotype has remained stable throughout the radiation of the genus *Formica*: members of a derived clade of socially parasitic species tend to have 26 chromosomes while species in the paraphyletic '*Serviformica*' group tend to have 27 (Hauschteck-Jungen & Jungen, 1976). A comparison of linkage maps between *F. selysi* (27 chromosomes) and *F. exsecta* (26 chromosomes) revealed only one major chromosomal rearrangement genome-wide, and no rearrangement between the *Sm* haplotypes of the two species, providing additional evidence that synteny is highly conserved across the genus (Brelsford et al., 2020).

Although we find that the genetic architecture of social organization in *F. selysi* and *F. francoeuri* likely shares a common origin (see also Purcell et al., 2021), *F. francoeuri* exhibits several important differences with respect to the distribution of supergene genotypes. In particular, the near absence of *Sp/Sp* workers in *F. francoeuri* suggests that different mechanisms may contribute to the maintenance of the supergene polymorphism in this species (Table 2).

4.2 | Maintenance of the *Formica* supergene

If the *Sp* haplotype confers a green-beard effect or acts as a transmission ratio distorter in *F. francoeuri*, as has been found in other ant supergene systems (Avril, Purcell, Béniguel, and Chapuisat 2020; Ross & Shoemaker 2018; Keller & Ross 1998), this could explain the low occurrence of *Sm/Sm* workers in polygynous colonies. However, it does not explain the near absence of *Sp/Sp* workers. The rarity of *Sp* homozygotes may indicate low viability of workers with this genotype. Along these lines, the *Sb* haplotype in the socially polymorphic ant *S. invicta* is associated with reduced reproductive success in males (Fritz et al. 2006, Lawson et al. 2012) and reduced viability of homozygous gynes and workers (Wang et al. 2013). An absence of

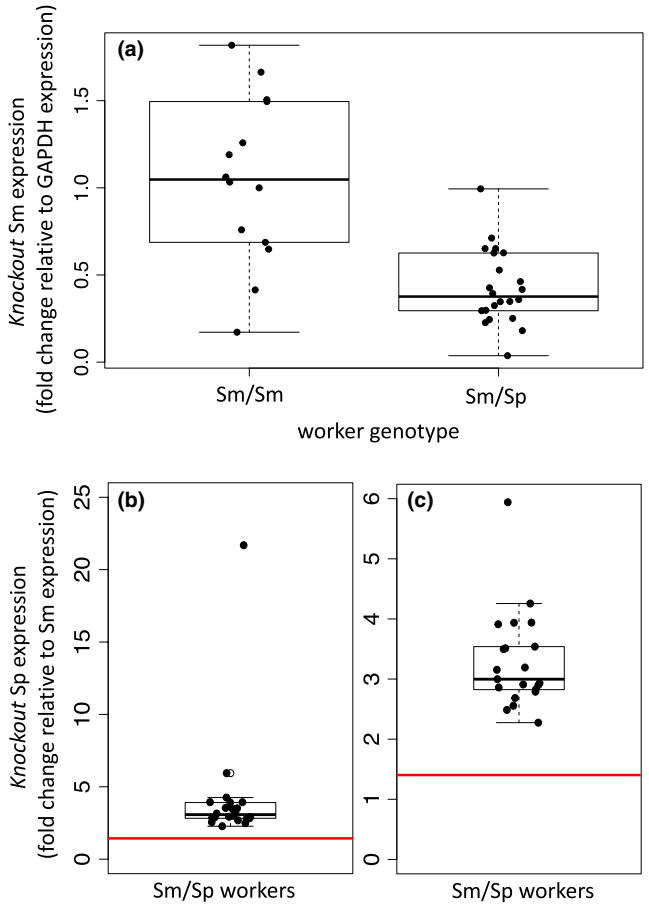

FIGURE 4 Using qPCR, we calculated the expression levels of the two alleles at the candidate gene *knockout* in *F. francoeuri* workers. Expression of the Sm allele was about twice as high in *Sm/Sm* workers relative to *Sm/Sp* workers (a, fold change calculated with reference to the housekeeping gene GAPDH). The *Sp* allele exhibits significantly elevated expression relative to the *Sm* allele in all heterozygous workers (b and c). In 22 *Sm/Sp* workers (b), the fold change of *Sp* is significantly greater than the null expectation of equal expression of both alleles, even when we account for differences in primer efficiency (the red line indicates the expected *Sp* fold change if allele expression rates were equal). Panel c shows the same data as b, excluding the outlier

Sp/Sp workers has also been observed in *Formica glacialis* (Lagunas-Robles, Purcell & Brelsford, 2021), although their sample of polygynous colonies was small. This study of *F. francoeuri* shows that species may differ in their supergene genotype distributions despite conservation of the supergene and its association with polymorphic social organization. Our current data do not allow us to determine the cause of the paucity of *Sp/Sp* workers in *F. francoeuri* but we can look to both theory and empirical studies that indicate what kinds of evolutionary dynamics may be taking place.

When large genomic regions are tightly linked, they are subject to genetic hitchhiking, which reduces the effectiveness of purifying selection. Low recombination rate can also reduce standing genetic variation and lead to the accumulation of deleterious mutations and repetitive elements, gene loss, and other genetic modifications associated with degeneration (Gutiérrez-Valencia, Hughes, Berdan, & Slotte, 2021; Jay et al., 2021; Rice, 1987). Some of these signatures of degeneration are found in the polygyne-associated haplotype of the *S. invicta* supergene (Gutiérrez-Valencia et al., 2021; Martinez-Ruiz, et al., 2020; Stolle, et al. 2019).

Given this context, our data provide circumstantial evidence for the presence of recessive deleterious alleles on the social supergene of *F. francoeuri*. Whether degeneration is occurring in the social supergene region of some species of *Formica* is as yet unknown.

The mechanisms maintaining the *Formica* supergene and its associated phenotypic polymorphisms appear to vary among species. For example, *Sm/Sm* eggs laid by heterozygous *F. selysi* queens fail to hatch (Avril et al., 2020). The effect is that polygynous *F. selysi* colonies are comprised of *Sp/Sp* and/or *Sm/Sp* workers and gynes. Transmission distortion that favours the *Sp* haplotype and the advantages of the *Sm* haplotype (such as unidirectional gene flow from monogyne to polygyne colonies via *Sm* males) may contribute to balancing selection that maintains the polymorphism (Avril et al., 2019). The maintenance of the *S. invicta* social supergene may be determined by similar dynamics, in which the polygyne haplotype confers a green-beard mechanism while recessive lethality and reduced mating success of males (Lawson et al. 2012; Keller & Ross, 1998) prevents fixation of the selfish allele. A combination of these mechanisms may be at play in *F. francoeuri*, which lacks *Sm/Sm* and *Sp/Sp* workers in polygynous colonies (Table 2). Comparative analyses across a broader set of species with differing mechanisms maintaining supergene polymorphisms will allow researchers to broaden conclusions about what competing selective forces maintain complex phenotypes.

4.3 | *Knockout* & gene expression

We find significant allele-specific expression of the *Knockout Sp* allele in *F. francoeuri* heterozygotes. Identifying this expression pattern adds evidence that *Knockout* influences colony social organization in *Formica* (Brelsford et al. 2020; Purcell et al., 2021). In European *Formica* species, Brelsford et al. (2020) detected no differences in the protein-coding sequence of alternative *Knockout* alleles. They proposed that functional differences between individuals with alternative genotypes may depend on differential expression, which is consistent with the patterns observed in *F. francoeuri* heterozygotes (Figure 4). A more comprehensive analysis of *Knockout* expression at different developmental stages and in different castes may shed additional light on the role that *Knockout* may play in shaping alternative social structure.

4.4 | Social form inference

From our inference based on identity-by-descent and opposing homozygosity, we were able to determine the social structure of 70 of 76 colonies. This method of inferring social structure has some potential sources of error. In addition to the role of occasional

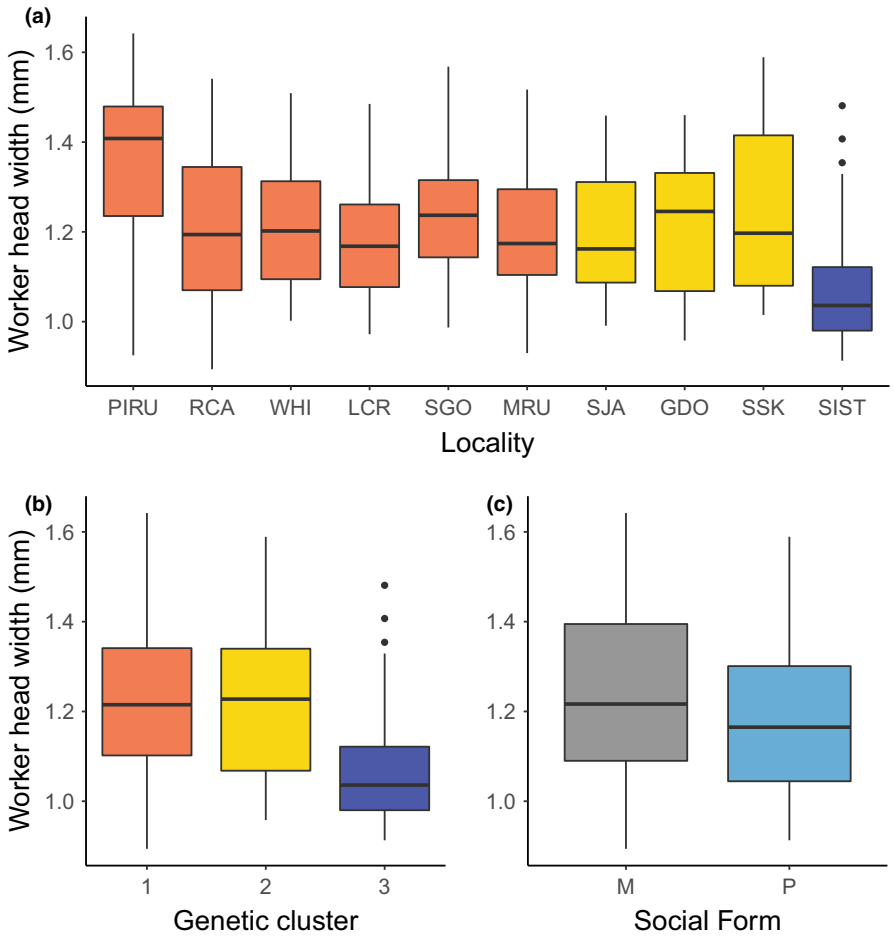

FIGURE 5 Worker head width by (a) locality, (b) genetic cluster (see Figure 6c) and (c) mode of social organization

genotyping errors that was described in the methods, intermediate values of opposing homozygosity leading to the six 'ambiguous' colonies may be produced by biologically relevant causes. For instance, queens from monogyne colonies may mate with more than one male, which would decrease IBD values and increase opposing homozygosity relative to singly mated monogyne colonies. Alternatively, two closely related queens may occasionally form a cooperative nest (Purcell et al., 2014) which would produce a similar effect. Finally, if workers drift between monogyne colonies, these may be misclassified as polygyne or ambiguous. Despite these uncertainties about the mating habits and the relationships of queens in each colony, we identified a very strong association between supergene genotype and inferred social structure.

Our estimation of pairwise relatedness within colonies supports our determination of social form. We find a clear difference in average pairwise relatedness among workers from colonies we categorized as monogyne and those we categorized as polygyne. Our estimates of relatedness among monogyne colonies are near the expected value of 0.75 for full sibling workers. Some of these monogyne colonies exhibit high variance in pairwise relatedness. This variance may be due to polyandry, or to sources of error described above. For polygyne colonies, we expect very low pairwise relatedness among workers. As a generous example, for a polygyne nest consisting of two sister-queens, each mated to only one male, the average relatedness of the first-cousin workers is approximately 0.19. This hypothetical colony would contain a mix of full siblings and cousins, resulting in a higher colony-level relatedness estimate. The average relatedness values we obtained for all polygyne colonies are lower than what we can expect in this example. However, the method of estimating relatedness we implemented is sensitive to population structure, resulting in an overestimation of relatedness in our polygyne colonies. We compared pairwise relatedness among workers from different localities to pairwise relatedness among workers from the same locality but different colonies. We found that relatedness among workers from different localities is near or below zero, while relatedness among workers from different colonies in the same locality is similar to the values we obtained for polygyne colonies (Figure 1b).

4.5 | Phenotype and population genetic structure

Differences in average worker size are not significantly associated with social form, but instead differ based on sampling locality. Even so, the average worker size among colonies from most localities is similar with the exceptions of PIRU and SIST, having the largest and smallest workers, respectively. The lack of a significant difference between the workers of polygyne and monogyne colonies contrasts with findings in other species of *Formica* and in *Solenopsis* which

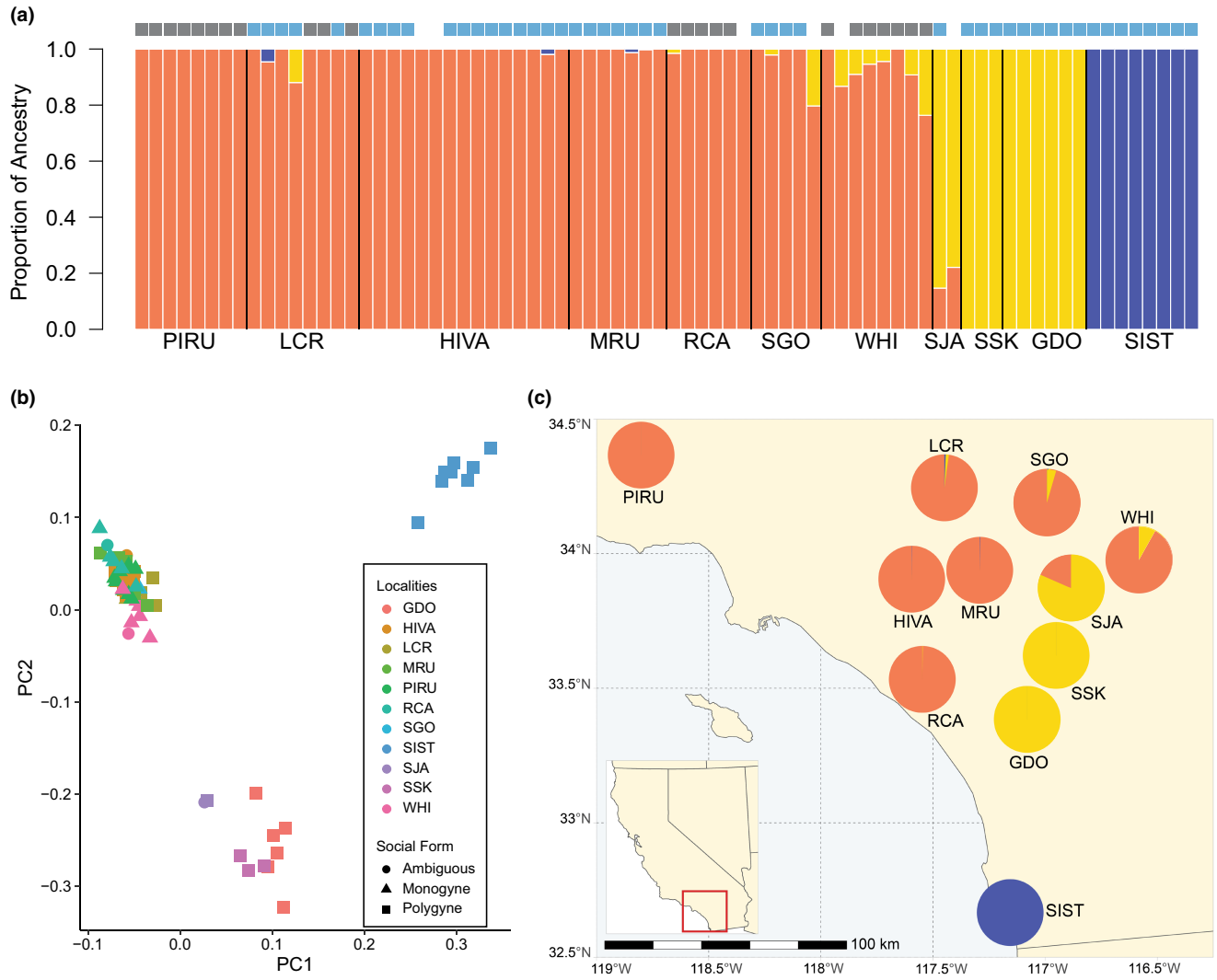

FIGURE 6 (a) Admixture results for $K = 3$. Each vertical bar represents one individual from a colony (one worker containing the least missing data). Colours indicate unique genetic clusters. The height of each bar represents the proportion of an individual's ancestry from each genetic cluster. Polygynous colonies are represented with a light blue square above, monogynous colonies with a grey square above and ambiguous colonies with no indicator of social form. (b) Principal components analysis. Each point represents one individual from a colony (one worker containing the least missing data). Colours distinguish localities, and shape indicates social organization. (c) Map of sampling localities and averaged proportion of ancestry from each of three genetic clusters identified by Admixture

TABLE 2 A summary of the genotype distributions in *Formica francoeuri* and two species used for comparison (Purcell et al. 2014; Avril et al. 2019; Wang et al. 2013, Ross 1997)

	Genotypes		Mechanisms affecting worker genotypes in polygyne colonies
	Monogyne	Polygyne	
Formica selysi	*Sm/Sm*	*Sm/Sp*, *Sp/Sp*	Maternal effect lethality of *Sm/Sm* eggs (when offspring of heterozygous queen)
Solenopsis invicta	*SB/SB*	*SB/SB*, *SB/Sb*	*Sb* allele highly deleterious in homozygotes
Formica francoeuri	*Sm/Sm*	*Sm/Sp*	Potentially deleterious *Sp* allele, Potential maternal effect lethality of *Sm/Sm* offspring in polygyne colonies

describe morphological differences associated with social form (Schwander et al., 2005; Ross & Keller, 1995; Keller, 1993). Analysis of worker size between social forms using additional samples from localities where both social forms are present (e.g. Lytle Creek, Hidden Valley, San Jacinto) will be more appropriate for disentangling the effect of geographically dependent variation in worker size

from variation associated with the polygyny syndrome. Additionally, other aspects of the polygyny syndrome, such as queen body mass, may be assessed with sampling of alate queens.

The convergence of polymorphic social organization and its underlying genetic architecture raises questions about the ecological and evolutionary importance of supergenes and the traits they control. It has been suggested that the suite of characteristics that differ between gynes of alternative social forms is an adaptation to different ecologies (Liautard and Keller 2001; Purcell et al. 2015). Alternatively, source-sink dynamics may determine the distribution of social forms across a heterogeneous landscape, according to the effects of morphological differences and social strategies (Purcell et al. 2015). The larger body mass and greater energy reserves of queens of monogyne origin may allow greater dispersal distances after mating relative to queens of polygyne origin. The ability to disperse greater distances, combined with higher fecundity and longer life span (Ross and Keller 1995), may enable more successful colonization of suboptimal habitat. While we were unable to assess the full range of life history traits that make up the polygyny syndrome in *F. francoueri*, our examination of worker phenotypes is inconsistent with the polygyny syndrome. In particular, we do not find a correlation between worker size and social organization. Additionally, elevation is not apparently correlated with social organization, as is the case in *F. selysi* (Purcell et al. 2015), given that polygyne colonies were found in high proportions at the highest (SGO: 770 m) and lowest (SIST: near sea level) elevation sites and that monogyne colonies were found at high frequency at the next highest and lowest elevation sites (WHI: 670 m; RCA: 105 m). Other as yet unknown ecological factors may favour each social form in different localities, however, since *F. francoeuri* populations frequently contained exclusively monogyne or exclusively polygyne colonies.

5 | CONCLUSION

Previous studies of ant societies have revealed that convergent genetic architecture can contribute to convergence of complex phenotypes across broad phylogenetic distances (Purcell et al., 2014; Wang et al., 2013). More recent studies that focus on closely related taxa show that supergenes and their associated phenotypes may be conserved for great lengths of time (Purcell et al., 2021; Brelsford et al., 2020; Yan et al., 2020). The present study highlights a peculiar example of how the phenotypic effect of a supergene may be conserved despite apparent differences in the mechanisms that maintain the polymorphism across species. The similarities and contrasts between *F. francoeuri* and *F. selysi* hint at the possibility that the *Formica* supergene, which is shared in divergent congeneric species, will provide unique insights into the causes and consequences of recombination suppression.

AUTHOR CONTRIBUTIONS

Study conception and design: D. Pierce and J. Purcell; data collection: all authors; analysis and interpretation of results: D. Pierce, J. Purcell, and A. Brelsford; draft manuscript preparation: D. Pierce, J. Purcell, and A. Brelsford; Review and approval of the final version of the manuscript: all authors.

ACKNOWLEDGEMENTS

We are grateful to Polly Campbell, Elisa Henderson, German Lagunas-Robles, Mari West and Madison Sankovitz for comments on a draft of this manuscript. We are also thankful for Aracely Ospina's assistance with DNA extraction and library preparation. This work was supported by an NSF-CAREER DEB-1942252 and U. S. Department of Agriculture NIFA Hatch #CA-R-ENT-5126-H to JP and NSF DEB-1754834 to AB and JP. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

The sequencing read data that support the findings of this study are available in the NCBI Sequence Read Archive under Bioproject PRJNA816105. Morphometric data are available on Dryad (doi: 10.6086/D1HM4F).

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jeb.14038.

ORCID

Daniel Pierce https://orcid.org/0000-0003-0941-2573
Penglin Sun https://orcid.org/0000-0003-4087-8267
Jessica Purcell https://orcid.org/0000-0001-8568-4162
Alan Brelsford https://orcid.org/0000-0002-5074-7765

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Avril, A., Purcell, J., Brelsford, A., & Chapuisat, M. (2019). Asymmetric assortative mating and queen polyandry are linked to a supergene controlling ant social organization. *Molecular Ecology*, *28*(6), 1428–1438.

Avril, A., Purcell, J., Béniguel, S., & Chapuisat, M. (2020). Maternal effect killing by a supergene controlling ant social organization. *Proceedings of the National Academy of Sciences*, *117*(29), 17130–17134.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological*, *57*(1), 289–300.

Brelsford, A., Dufresnes, C., & Perrin, N. (2016). High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity*, *116*(2), 177–181.

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. *Current Biology*, *30*(2), 304–311.e4.

Browning, B. L., & Browning, S. R. (2016). Genotype imputation with millions of reference samples. *American Journal of Human Genetics*, *98*(1), 116–126.

Browning, B. L., & Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*, *194*(2), 459–471.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140.

Charlesworth, D. (2016). The status of supergenes in the 21st century: Recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evolutionary Applications*, *9*(1), 74–90.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Fritz, G. N., Vander Meer, R. K., & Preston, C. A. (2006). Selective male mortality in the red imported fire ant, *Solenopsis invicta*. *Genetics*, *173*(1), 207–213.

Gutiérrez-Valencia, J., Hughes, P. W., Berdan, E. L., & Slotte, T. (2021). The genomic architecture and evolutionary fates of supergenes. *Genome Biology and Evolution*, *13*(5), evab057.

Hauschteck-Jungen, E., & Jungen, H. (1976). Ant chromosomes I.—The genusFormica. *Insectes Sociaux*, *23*(4), 513–524.

Huang, K., Ritland, K., Dunn, D. W., Qi, X., Guo, S., & Li, B. (2016). Estimating relatedness in the presence of null alleles. *Genetics*, *202*(1), 247–260.

Huang, Y. C., & Wang, J. (2014). Did the fire ant supergene evolve selfishly or socially? *BioEssays*, *36*(2), 200–208.

Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., & Joron, M. (2021). Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature Genetics*, *53*(3), 288–293.

Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., Whibley, A., Becuwe, M., Baxter, S. W., Ferguson, L., Wilkinson, P. A., Salazar, C., Davidson, C., Clark, R., Quail, M. A., Beasley, H., Glithero, R., Lloyd, C., Sims, S., … ffrench-Constant, R. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, *477*(7363), 203–206.

Keller, L. (1993). The assessment of reproductive success of queens in ants and other social insects. *Oikos*, *67*, 177–180.

Keller, L., & Ross, K. G. (1998). Selfish genes: A green beard in the red fire ant. *Nature*, *394*(6693), 573–575.

Küpper, C., Stocks, M., Risse, J. E., Dos Remedios, N., Farrell, L. L., McRae, S. B., & Kitaysky, A. S. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, *48*(1), 79–83.

Lagunas-Robles, G., Purcell, J., & Brelsford, A. (2021). Linked supergenes underlie split sex ratio and social organization in an ant. *Proceedings of the National Academy of Sciences*, *118*(46), e2101427118.

Langmead, B., & Salzberg, S. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods*, *9*, 357–359.

Larracuente, A. M., & Presgraves, D. C. (2012). The selfish segregation distorter gene complex of *Drosophila melanogaster*. *Genetics*, *192*(1), 33–53.

Lawson, L. P., Vander Meer, R. K., & Shoemaker, D. (2012). Male reproductive fitness and queen polyandry are linked to variation in the supergene Gp-9 in the fire ant *Solenopsis invicta*. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1741), 3217–3222.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993.

Liautard, C., & Keller, L. (2001). Restricted effective queen dispersal at a microgeographic scale in polygynous populations of the ant *Formica exsecta*. *Evolution*, *55*(12), 2484–2492.

Lyon, M. F. (2003). Transmission ratio distortion in mice. *Annual Review of Genetics*, *37*(1), 393–408.

Martinez-Ruiz, C., Pracana, R., Stolle, E., Paris, C. I., Nichols, R. A., & Wurm, Y. (2020). Genomic architecture and evolutionary antagonism drive allelic expression bias in the social supergene of red fire ants. *eLife*, *9*, e55862.

Morandin, C., Havukainen, H., Kulmuni, J., Dhaygude, K., Trontti, K., & Helanterä, H. (2014). Not only for egg yolk-functional and evolutionary insights from expression, selection, and structural analyses of *Formica* ant vitellogenins. *Molecular Biology and Evolution*, *31*, 2181–2193 pmid:24895411.

Pamilo, P. (1982). Genetic population structure in polygynous *Formica* ants. *Heredity*, *48*(1), 95–106.

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, *21*(12), 2991–3005.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*(5), e37135.

Purcell, J., Brelsford, A., Wurm, Y., Perrin, N., & Chapuisat, M. (2014). Convergent genetic architecture underlies social organization in ants. *Current Biology*, *24*(22), 2728–2732.

Purcell, J., Pellissier, L., & Chapuisat, M. (2015). Social structure varies with elevation in an alpine ant. *Molecular Ecology*, *24*(2), 498–507.

Purcell, J., Lagunas-Robles, G., Rabeling, C., Borowiec, M. L., & Brelsford, A. (2021). The maintenance of polymorphism in an ancient social supergene. *Molecular Ecology*, *30*(23), 6246–6258.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing URL https://www.R-project.org/

Rice, W. R. (1987). Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics*, *116*(1), 161–167.

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*(5), 939–946.

Ross, K. G. (1997). Multilocus evolution in fire ants: Effects of selection, gene flow and recombination. *Genetics*, *145*(4), 961–974.

Ross, K. G., & Keller, L. (1995). Ecology and evolution of social organization: Insights from fire ants and other highly eusocial insects. *Annual Review of Ecology and Systematics*, *26*(1), 631–656.

Ross, K. G., & Shoemaker, D. (2018). Unexpected patterns of segregation distortion at a selfish supergene in the fire ant *Solenopsis invicta*. *BMC Genetics*, *19*(1), 1–22.

Rosset, H., & Chapuisat, M. (2007). Alternative life-histories in a socially polymorphic ant. *Evolutionary Ecology*, *21*(5), 577–588.

Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology*, *24*(7), R288–R294.

Schwander, T., Rosset, H., & Chapuisat, M. (2005). Division of labour and worker size polymorphism in ant colonies: The impact of social and genetic factors. *Behavioral Ecology and Sociobiology*, *59*(2), 215–221.

Stolle, E., Pracana, R., Howard, P., Paris, C. I., Brown, S. J., Castillo-Carrillo, C., Rossiter, S. J., & Wurm, Y. (2019). Degenerative

expansion of a young supergene. *Molecular Biology and Evolution*, *36*(3), 553–561.

Tawdros, S., West, M., & Purcell, J. (2020). Scaling relationships in *Formica* ants with continuous worker size variation. *Insectes Sociaux*, *67*(4), 463–472.

Thompson, E. A. (2013). Identity by descent: Variation in meiosis, across genomes, and in populations. *Genetics*, *194*(2), 301–326.

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., Stager, M., Betuel, A., Cheviron, Z. A., Warren, W. C., Gonser, R. A., & Balakrishnan, C. N. (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*, *26*(3), 344–350.

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*(7434), 664–668.

Yan, Z., Martin, S. H., Gotzek, D., Arsenault, S. V., Duchen, P., Helleu, Q., Riba-Grognuz, O., Hunt, B. G., Salamin, N., Shoemaker, D., Ross, K. G., & Keller, L. (2020). Evolution of a supergene that regulates a trans-species social polymorphism. *Nature Ecology & Evolution*, *4*(2), 240–249.

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina paired-end reAd mergeR. *Bioinformatics*, *30*(5), 614–620.

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*(7), 821–824.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pierce, D., Sun, P., Purcell, J., & Brelsford, A. (2022). A socially polymorphic *Formica* ant species exhibits a novel distribution of social supergene genotypes. *Journal of Evolutionary Biology*, *35*, 1031–1044. https://doi.org/10.1111/jeb.14038