Balancing selection at the 'social supergene' in the Alpine silver ants

Aakanksha Singh
Erasmus Mundus Masters in Evolutionary Biology (MEME)
S4 Research Project
August 2021
Supervisors- Hélène Boulain and Michel Chapuisat

Table of Contents

Abstract

Many populations are polymorphic in nature with multiple alternative morphs which differ in a variety of traits. The 'supergene' architecture which consists of multiple co-adapted linked genes maintains the specific combination of traits in each morph in the face of recombination. For maintaining stable polymorphisms, it is crucial that the frequency of genetic variants underlying the different morphs is maintained by balancing selection. Thus, the supergenes provide ample opportunities to study balancing selection. The supergene determining social organisation in the Alpine silver ant, *Formica selysi*, presents two haplotypes, M and P which determine whether the colony is headed by a single queen (monogynous colony) or multiple queens (polygynous colony). In this report, we aim to detect signatures of balancing selection that might be operating on the supergene using whole-genome resequencing of 48 individuals of *F. selysi* from different geographical origins. We first studied the clustering of the samples on the supergene and the rest of the genome, then parameters such as nucleotide diversity, divergence and departure from neutral evolution were used to test for signatures of balancing selection. We found that the samples clustered according to their social organisation on the supergene and according to their geographical origins on the rest of the genome. Nucleotide diversity, F_{ST} and Tajima's D had significantly higher values on the supergene as compared to the rest of the genome. Hence, we can conclude that there is evidence of balancing selection operating on the supergene that controls social organisation in *Formica selysi*, which differs sufficiently from the selection operating on the rest of the genome.

Keywords: *Formica selysi*, polymorphism, selection, ants, social organisation

Introduction

Diversity results from the varying interactions between organisms and their environments over long periods of time. Phenotypic polymorphism refers to the presence of two or more relatively distinct phenotypic categories within the same population. These categories are referred to as morphs or alternative phenotypes. Studying polymorphisms has been crucial to our understanding of evolution as the presence of different phenotypic morphs is sometimes considered a precursor to speciation in which morphs evolve into different species (Gray and McKinnon, 2007). Dobzhansky (1951) stated that the 'absolute equality of adaptive values of two biological forms is highly unlikely' and so one morphological form eventually replaces the other and polymorphism is lost. Hence, understanding the mechanisms that maintain polymorphisms in nature is challenging as both directional natural selection and genetic drift should eliminate alternate alleles and thus erode genetic diversity (Lewontin, 1974; Charlesworth and Hughes, 2000; Nielsen, 2005). Nevertheless, many examples of persistent polymorphisms occur in nature such as the blood groups in primates, coat colour in cats, coloured morphs of side-blotched lizards, different types of social organisation in ants, and many others (Hedrick, 1986; Mallet and Joron, 1999; Richman, 2000; Carius et al., 2001; Hedrick et al., 2002; Delph and Kelly, 2014; Gill et al., 2009). Phenotypic polymorphisms provide excellent opportunities to study mechanisms regulating genetic variation because they represent distinct and easily measurable variation maintained by selection (Jamie and Meier, 2020). The early evolutionary biologists hypothesised that owing to the discrete nature of the phenotypic differences observed in the natural populations of a species, the polymorphic variations between the alternate morphs had genetic origin. Moreover, the problem was further complicated by complex phenotypic traits such as, morphology, sexual-compatibility, social organisation etc. that differed between the alternate morphs. Maintenance of a specific combination of traits in each alternate morph in the face of recombination, led to the hypothesis that the genetic basis of this discrete variation is a functional unit containing multiple co-adapted linked loci, i.e., 'multiple linked genes' that are inherited as a 'single gene'. A 'supergene' architecture was thus proposed to explain the complex adaptive phenotypes that segregate within a species.

According to Thompson and Jiggins (2014) for a genetic element to qualify as a supergene, it should (a) exhibit clear evidence of a complex phenotype of multiple co-adapted elements, (b) have a pattern of inheritance essentially identical to alternative alleles at a single locus, and (c) be maintained in a stable polymorphism in a population. The sex chromosomes, which are morphologically and genetically distinct and have evolved independently in many groups of animals and plants (Bull 1983; Charlesworth 1996) are the best-known examples of supergenes. More recent studies have identified supergenes controlling wing patterns in *Heliconius numata.* In *H. numata*, up to seven morphs coexist in local populations and each morph accurately mimics of one of several available model species in another butterfly genus *Melinaea* (Brown and Benson, 1974b) to protect itself from predators. Shell colour polymorphism in the snail *Cepaea nemoralis* (Gonzalez et al., 2019), social organisation in ants and plumage and behavioural variation in the white-throated sparrow (Davis et al., 2011) are other examples of complex traits controlled by supergenes. Fisher (1930) predicted that there should be strong selection for mechanisms suppressing recombination between loci affecting different traits when mixed trait combinations are detrimental. Since then, it has been argued that supergenes evolve because natural selection reduces recombination between interacting loci, thus producing blocks of co-adapted genes and eliminating blocks of genes which are not co-adapted (Sheppard, 1953; Bodmer and Parsons, 1962; Ford, 1964). Physical linkage and lack of recombination in the supergene reduces the probability that particular trait combinations are rearranged to give offspring with novel combinations. An increased linkage in the region with co-adapted alleles is advantageous as it maintains traits in favourable combinations.

For the polymorphism to persist in a population, it is crucial that the frequency of genetic variants underlying the different morphs be stabilised by balancing selection (Gray and McKinnon, 2007). Balancing selection describes the suite of adaptive forces that maintain genetic variation for longer than expected by random chance. Mechanisms such as negative frequency-dependent selection (Ayala and Campbell, 1974), heterozygote advantage (Hedrick, 2012), selection pressures fluctuating through space and time (Van Valen, 1973; Levene, 1953) or sexual antagonism (Gouyon et al., 1991) may be involved in balancing selection. The core concept of each balancing selection model is that the selective value of an allele, whether it is beneficial or detrimental, depends on the environmental context (Dobzhansky, 1982; Clarke et al., 1988). That is, alleles are

advantageous and deleterious in different circumstances. However, in many cases where balanced polymorphism is observed, the underlying mechanisms remain a mystery. Supergenes provide thus ample opportunity to study a complex balanced polymorphism that is maintained in a population without the generation of maladaptive intermediates.

It has been recently revealed that social organization is controlled by a supergene in at least two independent ant lineages, *Solenopsis* and *Formica* (Purcell et al., 2014; Wang et al., 2014) and may possibly be present in a third ant lineage, *Leptothorax* (Braim, 2015). In these species, the genotype at the supergene determines whether colonies are headed by one or by multiple queens (Wang el al., 2013; Purcell et al., 2014). Our study species, *Formica selysi* (Bondroit, 1918), commonly known as the Alpine silver ant, is abundant in the southern half of France and the Western Alps (Seifert, 2002). In this species, the almost chromosome wide supergene determining social organization (Purcell et al., 2014) presents two haplotypes, M and P. The supergene is almost 10,000 kb long and is constituted by three inversions on chromosome 3 (Brelsford et al., 2020). In single-queen (monogynous) colonies, all females are homozygous for the M haplotype and mate with haploid M males. In contrast, all individuals produced by multiple-queen (polygynous) colonies bear at least one copy of the P haplotype (PP or MP) and can mate with either a M or a P male (Fig. 1). Avril et al. (2020) explained that maternal killing effect was responsible for the absence of MM workers in a polygynous colony despite observing matings between the heterozygous mothers (MP) with a M male from a monogynous colony. Fontcuberta et al. (2021) further demonstrated that females of monogynous origin (MM) did mate with males of polygynous origin (P) in natural conditions. This cross accounted for 20% of the matings by monogyne females in swarms in natural conditions, whereas it has never been detected in mature field colonies and in laboratory conditions. The two types of colonies differ in many ecological traits such as colonization skills, body size, fat storage, dispersal, and investment in reproductive progeny, mating behaviour, colony size and habitat in a geographical area (Keller and Passera, 1989; Chapuisat et al., 2004; Purcell et al, 2015; de Gasperin et al., 2020; Zahnd et al., 2021; Fontcuberta et al., 2021). This social supergene is widespread across the *Formica* genus and evolved 20-40 million years ago (Brelsford et al., 2020).

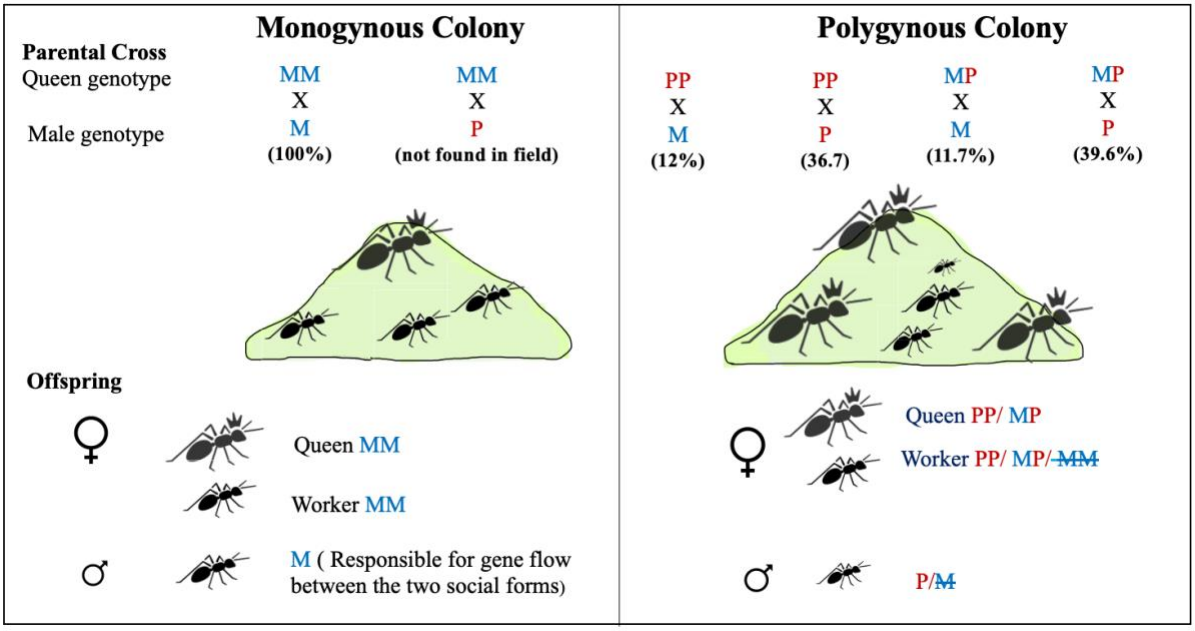

Fig 1. Social organisation in *Formica selysi* (adapted from Avril et al., 2020). The monogynous colony consists of ♀MM and ♂ M. The cross between ♀MM and ♂ P has not been found in the field. The ♂ M from monogynous colonies are responsible for gene flow between the two social forms. The polygynous colony consists of MP and PP females and the P males. All the ♀ MM and ♂ M produced in polygynous colonies are absent due to maternal effect killing.

To understand the selection regime that might be operating on the supergene giving rise to the unique social organisation in *F. selysi*, we address the following biological questions- Can we detect signatures of balancing selection operating on the supergene in *F. selysi*? How does the genetic diversity differ between the two social forms on the supergene and the rest of the genome? We hypothesize that we can detect clear signatures of balancing selection operating on the supergene that differs significantly from the rest of the genome. To test predictions based on our understanding of supergene evolution in each of the monogynous and polygynous social forms we analysed the complete genomes of forty-eight workers of *F. selysi* sampled from four geographical locations (Finges and Derborence in Switzerland and, Les Busset and Sallanches in France). I used the recently assembled high-quality reference genome of the monogynous form (MM) to call variants in all the sampled individuals. I first clustered the samples then used the divergence and diversity indices such as fixation index (F_{ST}), nucleotide diversity (π) to characterize the genetic

diversity and Tajima's D to look for departure from neutral evolution which act as signatures of balancing selection.

Materials and methods

---

a. Ant sampling and genotyping

Formica selysi, commonly known as 'the Alpine silver ant' is abundant in the floodplains throughout the Alps and the Pyrenees (Lude et al., 1999; Seifert, 2003). Sasha Zahnd and Massimo Bourquin collected worker ants from four locations in France and Switzerland, from June to August 2018. Forty-eight workers were collected from both the monogynous (MM=21) and polygynous colonies (MP=12, PP=15) from four geographical locations, Finges (N=16) Derborence (N=16) in Switzerland as well as from Les Bussets (N=8) and Sallanches (N=8) in France. Two samples of *Formica cinerea* (one monogyne and one polygyne) from Switzerland were used as outgroups.

b. Sequencing

Prior to my master's project, the DNA was extracted using the Qiagen DNeasy kit (Qiagen, Hilden, Germany) for insects, following the manufacturer protocol. The DNA library was prepared by the Genomic Technologies Facility at the University of Lausanne, using the NEBNext Ultra II DNA kit. The sequencing was performed on an Illumina HiSeq2500 system, with 100 bp paired end reads. The sequencing resulted in approximately 30 million reads per individual, corresponding to a mean coverage of 20X before pre-processing.

c. Data pre-processing

The pipeline to detect the signatures of balancing selection is summarised in Fig 2. First, Trimmomatic version 0.39 (Bolger et al., 2014) was used to trim the paired-end raw reads from individual sequences. After trimming, all the reads with a sequence length shorter than 80 were discarded. Reads were mapped and aligned on the homozygous monogyne reference genome (MM), using BWA-mem version 0.7.17 (Li, 2013), with default parameters. The samples that were

initially processed in two lanes while sequencing, were merged with the Picard Tools MergeSamFiles (Broad Institute, 2018). The indels were realigned with RealignerTargetCreator and IndelRealigner from the GenomeAnalysisToolkit (GATK) version 3.5.0 (Poplin et al., 2017). Duplicates were removed using Picard Tools MarkDuplicates (Broad Institute, 2018). At each step, the statistics of the output files were calculated using the sambamba version 0.7.1 (Tarasov et al., 2015). The bam files were indexed using Samtools version 1.20 (Li et al., 2009) after each step to make sure that no unusual number of reads were lost at any step. (Supplementary Table S1).

d. Variant Calling and filtering

The variants were called and filtered using GATK (version 4.1.3.0) tool, HaplotypeCaller in the gvcf mode, the individual gvcf files were merged using the GATK GenomicsDBImport to create a database before the joint genotyping using the GATK GenotypeGVCF tool. The joint VCF file was then filtered using GATK SelectVariants to retain the biallelic Single Nucleotide Polymorphisms (SNPs). The VCF was further filtered to remove the false positive hits and select the high-quality SNPs, using the hard filtering parameters for SNPs, as suggested by GATK best practices. GATK VariantFiltration was used to select SNPs that had: i) Quality (QUAL) value of greater than 30, ii) QualByDepth(QD) value greater than 2 to normalize the variant quality in order to avoid inflation caused by deep coverage, iii) FisherStrand (FS) value less than 60 to remove strand bias, iv) StrandOddsRatio (SOR) less than 4, to remove strand bias from exons, v) RMSMappingQuality (MQ) value of greater than 30, vi) MQRankSum greater than -8.5 to remove mapping quality bias between the reference and alternate alleles, vii) ReadPosRankSumTest (ReadPosRankSum) value greater than -0.5 to remove position bias between the reference and alternate alleles and viii) Depth (DP) value of greater than 5 .

The resulting filtered Variant Calling Format (VCF) file for the 48 individuals of *F. selysi* and 2 individuals of *F. cinerea*, mapped on the MM genome contained 3,672,131 SNPs. This filtered VCF file was used for all further analyses.

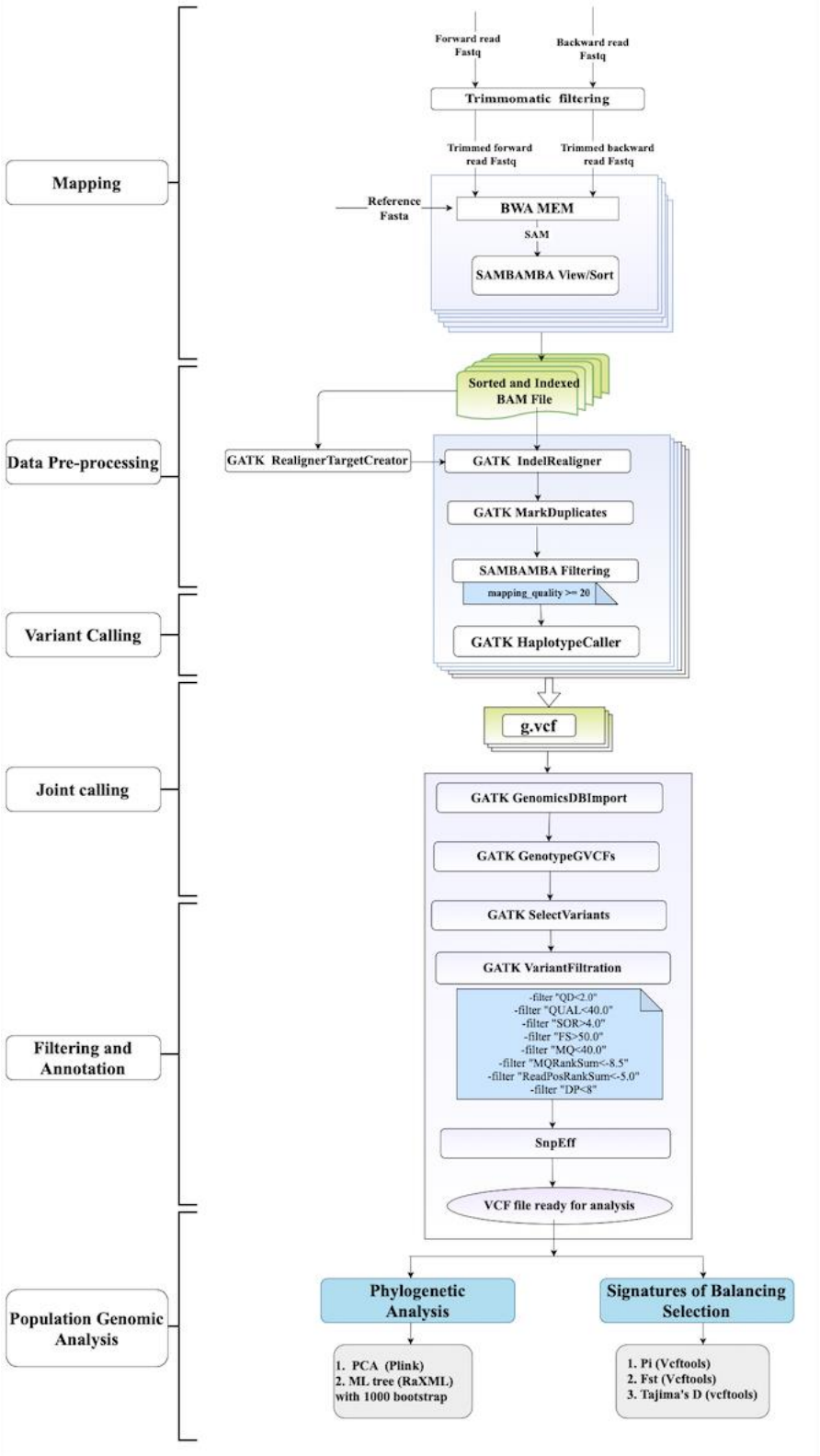

Fig 2. Schematic representation of the pipeline followed.

e. Sample clustering

Principal Component Analysis (PCA) allows us to calculate principal components (PCs) that explain the differences between individuals in genetic data. The eigen decomposition of the pairwise differences between the SNPs give a PCA-like overview of genetic distances between the workers of *F. selysi* of 3 different genotypes, MM, MP, and PP from four geographical origins. A PCA was performed on the supergene as well on the rest of the genome devoid of the supergene for all the individuals using the PLINK version 1.90 (Purcell et al., 2007). We also studied the phylogenetic relationships between the samples of *F. selysi* and *F. cinerea* through a SNP-based phylogenetic analysis. The VCF file was converted to the phylip file format using vcf2phylip version 2.0 (Ortiz, 2019) prior to a phylogenetic analysis with RAxMLversion 8.2.12 (Stamatakis, 2014). A Maximum-likelihood tree was generated using the GTRGAMMA model with 1000 rapid bootstraps.

f. Signatures of balancing selection

Population genomic analyses often use summary statistics to describe patterns of genetic variation and provide insight into evolutionary processes. Among the most fundamental of these summary statistics is the Nucleotide Diversity (π) which is employed to study the diversity in a subpopulation. I calculated π in each of the three genotypes, MM, MP, and PP subpopulations. The Fixation Index (F_{ST}) was employed to study the divergence between the monogynous and the polygynous populations. Wright's F_{ST} was originally defined as the correlation between two randomly sampled gametes from the same subpopulation when the correlation of two randomly sampled gametes from the total population is set to zero. π and F_{ST} were calculated with VCFtools version 0.1.15 (Danecek et al., 2011) in non-overlapping sliding windows of 100kb.

Tajima's D was used to detect the deviation from neutral evolution on the supergene (Tajima, 1989). This statistic measures the difference between pairwise differences and segregation sites between two populations. Under neutrality, the means of both the above measured estimators should be approximately equal. Therefore, the expected value of Tajima's D for populations conforming to a standard neutral model is zero. Significant deviations from zero indicate a skew in the allele frequency distribution relative to neutral expectations. Positive values of Tajima's D

arise from an excess of intermediate frequency alleles and can result from population bottlenecks, structure and/or balancing selection. Negative values of Tajima's D indicate an excess of low frequency alleles and can result from population expansions or positive selection. VCFtools was used to calculate the values of Tajima's D in on non-overlapping sliding windows of 100kb. A summary representation of the pipeline followed for the analysis is presented in Fig. 2.

To test the significance of the results obtained *P-values* were calculated with the custom iterative test that compared the values found in the supergene to the one obtained from the same number of sliding windows resampled in the rest of the genome. 108 sliding-windows values were sampled randomly in the rest of the genome and their median compared to the one found in the supergene. The *P-value* was calculated as the proportion of trials having a median equal or higher than the supergene's one after 10,000 iterations. R version 4.0.4. (R Core Team, 2019) was used for all statistical analyses and plotting the graphs.

Results

---

a. **Sample clustering**

To understand how the samples of *F. selysi* cluster, we plotted PCAs based on the supergene (Fig 3A) and the rest of the genome without supergene (Fig 3B). On the supergene, PC1 explains most of the variation of differences across the individual workers (27.6 %), where the individual workers cluster according to their supergene genotype (along the x axis). On the rest of the genome, PC1 explains 12.8% differentiation between the individual workers, and we see well defined clusters associated with the geographical origin of the workers. The PCA with the two homozygous MM and PP samples of the outgroup, *F. cinerea* is presented in Supplementary Fig S1.

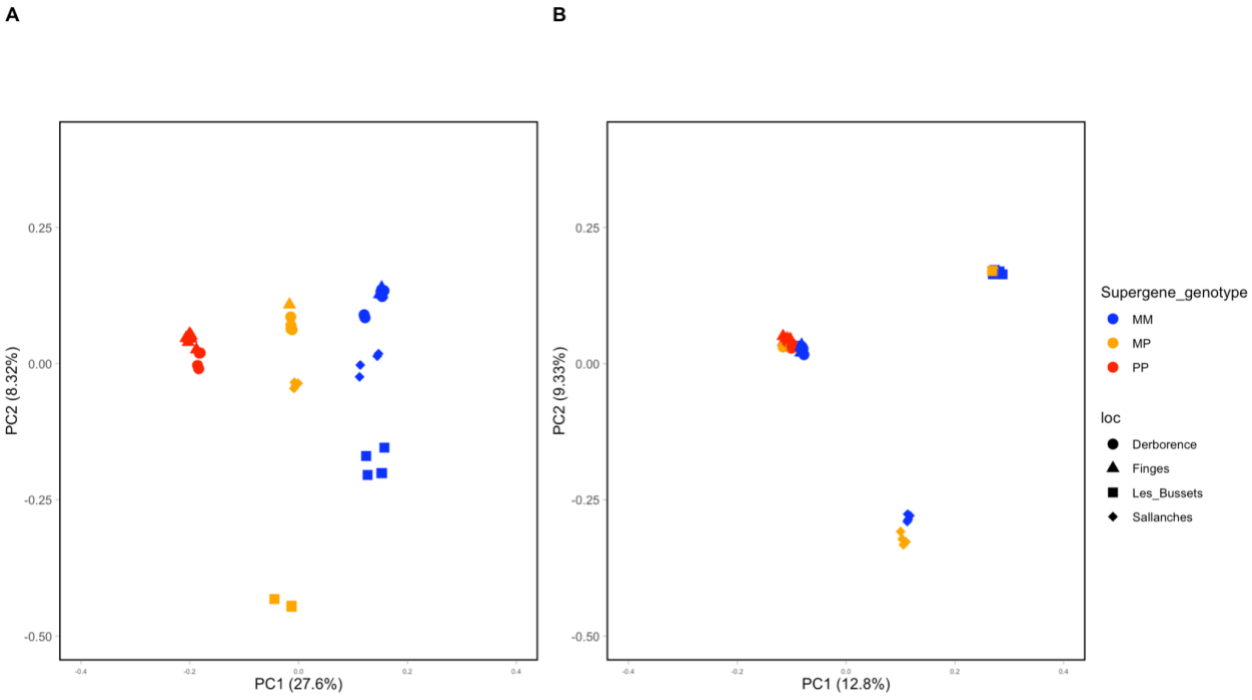

Figure 3. PCA of the forty-eight worker individuals of *F. selysi* collected from four geographical regions namely, Derborence, Finges, Les Bussets and Sallanches. (A) PCA using the 202,517 SNPs present on the supergene, and (B) PCA using the 3,672,131 SNPs present on the rest of the genome i,e. the genome excluding the supergene.

A maximum-likelihood analysis was performed using 2 individuals of *F. cinerea* along with 48 individuals of *F. selysi*. The analysis was carried out on the supergene and the rest of the genome without the supergene, using a bootstrap value of 1000. As shown in the tree (Fig. 4A), the workers cluster in three major groups: in accordance with their supergene haplotype with MM of *F. cinerea* clustered alongside MM of *F. selysi* and vice versa for the PP individuals. However, for the tree of the rest of the genome (Fig. 4B), the two main branches diverge according to the two species, *F. selysi* and *F. cinerea*, and *F. selysi* individuals clustered according to their geographical location.

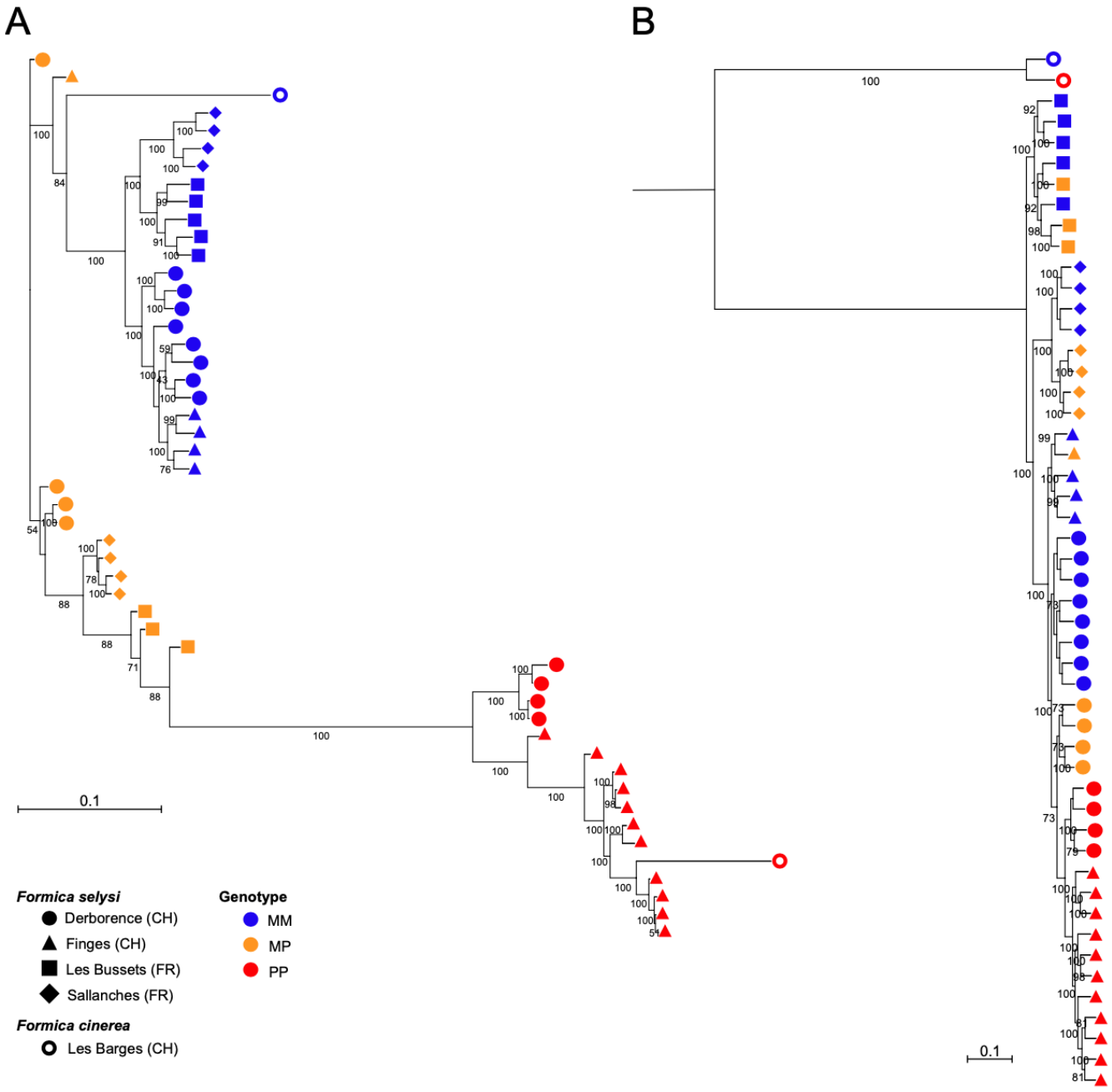

Figure 4. Maximum-likelihood tree generated with 1000 bootstrap demonstrating the relationship between the forty 48 workers of *F. selysi* and 2 workers of *F. cinerea* used as outgroup, based on (A) 202,517 SNPs present on the supergene and (B) 3,672,131 SNPs present on the genome excluding the supergene.

b. **Signatures of balancing selection on the supergene**

Nucleotide diversity ranged from 0 to 0.008 (Fig. 5A, B & C) on the 27 chromosomes in the three supergene genotypes. The median nucleotide diversity per non overlapping sliding window on the supergene, within each of the three genotypes, MM, MP, and PP was calculated to be 0.00188, 0.00422 and 0.00162 respectively (Table 1). However, the median nucleotide diversity per non overlapping sliding window on the rest of the genome, devoid of the supergene, was calculated to be 0.00168, 0.00173 and 0.00146 within each of the MM, MP, and PP genotypes respectively. The respective *p*-values suggest that in the three supergene genotypes, the supergene had significantly higher nucleotide diversity than the rest of the genome. The heterozygous MP had the greatest median nucleotide diversity on the supergene as well as on the rest of the genome. Though the homozygous PP individuals had the least median nucleotide diversity on the supergene among the three cases (Table 1), some non-overlapping windows had values as high as 0.0060 (Fig. 5C).

The divergence between the two social forms which was measured by F_{ST} was significantly increased in the supergene region compared to the rest of the genome, with a median F_{ST} as high as 0.40 on the supergene (Fig. 5C, Table 1). The supergene was characterized by F_{ST} estimates higher than 0.20 for consecutive 100 kbp non-overlapping windows, which contrasts sharply with the rest of the genome F_{ST}. We also tested divergence in our samples based on their geographic origin, (Supplementary fig. S2) where median F_{ST} values were calculated to be 0.17 on the supergene and 0.13 for the rest of the genome.

To investigate deviations from neutrality, we estimated Tajima's D in 100 kb non-overlapping windows in the two studied social morphs (Fig. 5D). There was a significant difference between the positive Tajima's D observed on the supergene (median = 0.723) and the negative Tajima's D observed on the rest of the genome (median = -0.860). According to estimates of nucleotide diversity, F_{ST} and Tajima's D genetic differentiation between individuals from the two social forms is significantly higher on the supergene compared to most of the remaining genome.

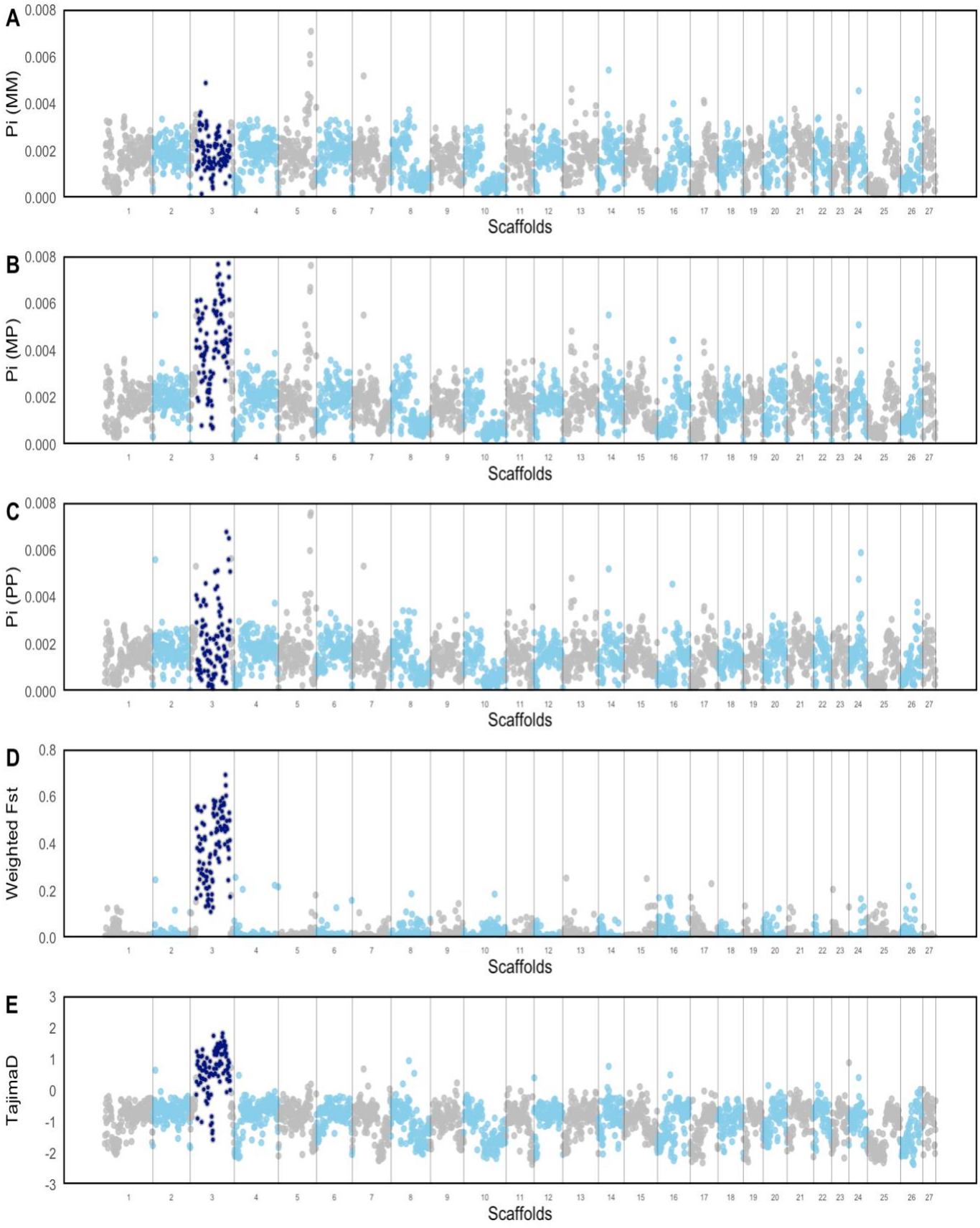
A
Pi (MM)
0.008
0.006
0.004
0.002
0.000
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Scaffolds
B
Pi (MP)
0.008
0.006
0.004
0.002
0.000
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Scaffolds
C
Pi (PP)
0.008
0.006
0.004
0.002
0.000
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Scaffolds
D
Weighted Fst
0.8
0.6
0.4
0.2
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Scaffolds
E
TajimaD
3
2
1
0
-1
-2
-3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Scaffolds

Figure 5. Manhattan plot of **(A)** Nucleotide diversity (π) within the monogynous (MM) subpopulation **(B)** Nucleotide diversity (π) within the heterozygous polygynous (MP) **(C)** Nucleotide diversity (π) within homozygous polygynous (PP) subpopulation **(D)** The divergence between the monogynous (MM) and polygynous (MP and PP) social forms is shown by the weighted F_{ST} values **(E)** Tajima's D to detect selection on the genome. All the values of π, F_{ST} and Tajima's D are calculated on the non-overlapping 100 kb sliding windows. The windows that lie within the supergene on chromosome 3 are highlighted in navy blue.

	Supergene	Rest of the genome	P value
π in MM	0.00188	0.00168	0.011
π in MP	0.00422	0.00173	0.000
π in PP	0.00162	0.00146	0.014
Fst	0.40644	0.00384	0.000
Tajima's D	0.72271	-0.86078	0.000

Table 1. Median values of Diversity (π), F_{ST} and *Tajima's D* for all 100 kbp non-overlapping sliding-windows located within the supergene and in the rest of the genome. There were 108 sliding-windows in the supergene and 2564 sliding-windows in the rest of the genome.

Discussion

In a species where there is no physical inhibition and all the individuals can mate randomly, polymorphism is hypothesized to be a result of balancing selection. Balancing selection leaves several types of signatures on the genome that form the basis of the statistical tests used to detect the process (Charlesworth 2006; Charlesworth and Charlesworth 2010). However, with these signatures we can only confirm if balancing selection may be operating on the part of the genome responsible for maintaining a population in a state of stable polymorphism, but we cannot predict which mechanisms of balancing selection may be involved. In the Alpine silver ant, a supergene system with two variants (M and P) controls whether colonies have one or multiple queens. Our

objective was to detect signatures of balancing selection operating on the supergene which could explain the presence of two social forms in the species.

Phylogenetic relationship between monogynous and polygynous individuals

The PCA analysis on supergene SNPs helped us verify that all the individuals were correctly genotyped in accordance with their social form and that all the samples were aptly marked during collection as they clustered according to their geographical location when analysing SNPs in the rest of the genome. The ML tree further helped us to establish the relationships between all the samples. We observed that the samples from Switzerland i.e. Finges and Derborence are very closely related which could be because of their close geographical origin. We used *F. cinerea* as an outgroup for building the phylogenetic tree and it is interesting to note that when based on the supergene, the MM and PP individuals of the outgroup are closer to the respective genotype of *F. selysi.* This could be a result of trans-specific polymorphism, i.e. the polymorphisms that are shared between the species. Indeed, Brelsford et al. (2020) found that the social organization was controlled by ancestral supergene (20-40 million years old) across socially polymorphic *Formica* species and found a small set of conserved trans-specific SNPs (~ 142) that consistently distinguish alternative supergene haplotypes across five *Formica* species, despite differences in the size of the supergene across the species. Ancestral shared polymorphisms or trans-species polymorphisms are considered as signatures of balancing selection because they arise when a balanced polymorphism is maintained for a very long time. As balancing selection minimizes the effects of drift, it slows the process of coalescence and thus lineage sorting. The matings between males from a monogynous colony with the queens of the polygynous colonies account for 23.7% (Fig. 1) of the matings in a polygynous nest. The resulting gene flow from the monogynous colonies to the polygynous colonies may also be responsible for reducing speciation and maintaining balanced polymorphism in the species. As the rate of drift is reduced by long term balancing selection at the locus of selection (supergene in our case), it pushes the most recent common ancestor to a more distant past (Takahata and Nei, 1990).

Supergene and nucleotide diversity

The diversity on the supergene in the samples belonging to the homozygous monogynous (MM) and the polygynous (PP) genotypes was more than that observed on the rest of the genome. The

heterozygous polygynous (MP) samples had the highest values of π in the three supergene genotypes. However, this higher diversity of the PP samples may not reflect reality. P-specific duplication events in the supergene could explain the higher diversity observed in PP samples when mapped on the MM genome, which would carry only one copy of certain genes. Reads corresponding to several duplicated genes would map on the single copy present in the MM genome increasing the apparent diversity if mutations accumulated after duplication in the P haplotype. We expected lower values of diversity in PP samples as the P allele is expected to have a lower effective population size due to its presence in only the polygynous colonies (the gene flow is reported to occur from monogynous colonies to polygynous colonies through M males but not *vice versa*).

Genetic differentiation between the two social forms

We used F_{ST} to test the genetic differentiation between the social forms. The elevated differentiation (F_{ST}) occurring between individuals of monogyne and polygyne origin across the supergene, in contrast to lower levels of differentiation in the rest of the genome. Balancing selection may also influence the geographic structuring of genetic variation; however, in our case we see that divergence of the populations based on their geographic origins (Supplementary Fig. S2) is less than the value of the divergence based on the social form, which suggests that the balancing selection is operating on the supergene which determines the social form and the populations do not show divergence according to geographic origin, as one would expect under neutral conditions.

Departure from neutral evolution

Balancing selection can produce an even frequency spectrum of segregating sites relative to expectations under neutral equilibrium, which is evident in cases of persistent polymorphisms. This pattern is caused by the overrepresentation of long internal branches in gene genealogies and can be detected with statistics such as Tajima's D. The test can thus be used for the identification of relatively recent instances of balancing selection. In our case, the Tajima's D has comparatively greater values on the supergene as compared to the rest of the genome. Positive value of Tajima's D on the supergene is a signature of balancing selection, however the negative value for the rest

of the genome implies a recent selective sweep and suggests that a different selection process is operating on the rest of the genome.

In some cases, balancing selection may not necessarily produce stable polymorphisms. For example, during co-evolution with pathogens, negative frequency-dependent selection results in highly variable allele frequency distributions that are often indistinguishable from neutral equilibrium expectations (Ejsmond et al., 2010). It is commonly assumed that the null model of the Tajima's D test, against which the focal locus is tested, is represented by neutrally evolving sequences; thus, skewed allele frequency spectra or increased nucleotide diversity can be attributed solely to balancing selection. However, if genes forming a supposedly neutral background are constrained, that is, under purifying selection, then the outlier status may be attributed either to balancing selection or to the relaxation of purifying selection. (Charlesworth et al., 1997; Engle and Fay, 2013; Harpur and Zayed, 2013). The significant increase in the nucleotide diversity on the supergene in the three genotypes rejects the hypothesis that positive Tajima's D values could have been a result of relaxed purifying selection and reaffirms that the spike in the Tajima's D values on the supergene may solely be a result of balancing selection.

Criticism of the work

Unevenly distributed genotypes (MM, MP, and PP) across the four locations could have caused sampling bias. Moreover, although the mapping rates for all the samples on the MM reference genome were similar (Supplementary Table S1), features of the P supergene haplotype such as specific gene duplications could have generated an increase in apparent diversity for the PP samples. We used Tajima's D to test for departure from neutral evolution on the supergene which may be extremely sensitive to confounding effects of nonequilibrium demography and population structure. Recent theoretical analyses indicate that although balancing selection could be far more widespread than previously thought and it is next to impossible to assess its importance via an analysis of polymorphism and divergence data. In such cases, a combination of experimental evolution and high throughput fitness assays may be reasonable, which may not be feasible in our case as the lifespan of a colony could be more than 10 years (Rosset and Chapuisat, 2006).

Conclusion and future prospects

There is strong evidence of balancing selection operating on the supergene controlling social organisation in *F. selysi*, that differs significantly from the selection operating on the rest of the genome. For any further work and a more in-depth population structure analysis, it would be useful to have samples that are uniformly distributed between the three supergene genotypes and locations. Calculating other statistics such as Direction of selection and pN/pS ratio would provide further insight to identify the regions of the supergene responsible for the differentiation and the maintenance of polymorphism in the population. Indels and copy number variants could also be tested for similar signatures of balancing selection as only biallelic SNPs were analysed in this project. Finally, identifying the genes which are targeted by balancing selection in the supergene could reveal key functions linked to the monogynous and polygynous phenotypes.

Acknowledgement

I would like to thank Michel Chapuisat and Hélène Boulain, my research supervisors, for their continuous guidance, encouragement, and useful critiques of this research work. I am grateful to Leo Beukeboom for his assistance in keeping my progress on schedule. I would also like to extend my gratitude to the ‘Social Evolution Group’ for the insightful discussions and for making my stay at UNIL a very pleasant experience.

Supplementary

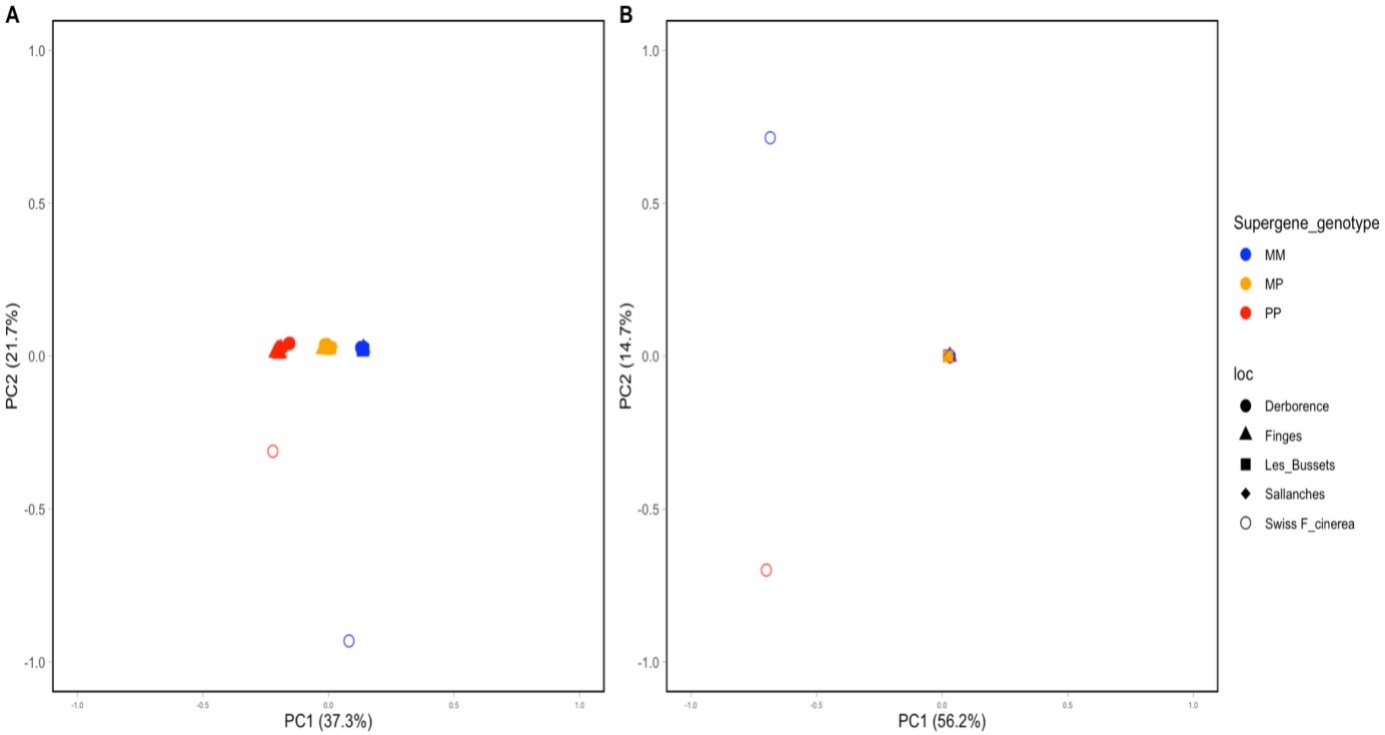

Fig S1. PCA of the 48 individuals of the *F. selysi* with two individuals of *F. cinerea* as the outgroup on (A) the supergene and (B) the rest of the genome. On the supergene, PC1 differentiates based on the social form while PC2 differentiates based on the species. However, on the rest of the genome, the maximum variation explained by PC1 (56.2 %) is based on the species.

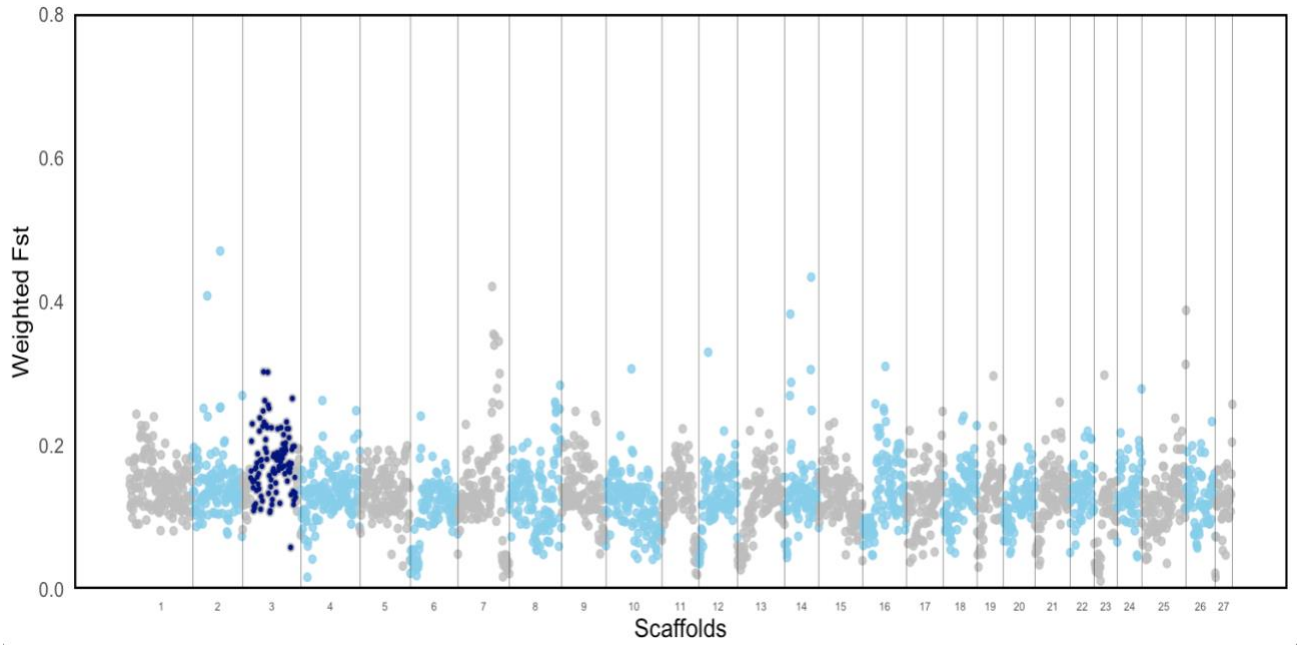

Fig S2. The divergence between populations from Finges, Derborence in Switzerland and Les Bussets, Sallanches in France, is shown by the weighted F_{ST} values. Supergene $_{median}$ = 0.17; Rest_of_the_genome $_{median}$ = 0.13.

Sample name	Species	Genotype	Social form	Origin	Raw Reads	Trimmomatic	Mapping	Pre-processing	Qualtity filter	Coverage
674W1_L2	F. selysi	MM	Monogynous	Finges	57837048	95.34	97.40	94.62	71.44	14.65
677W1_L2	F. selysi	MM	Monogynous	Finges	55009540	95.56	97.70	93.16	68.14	13.28
700W3_L2	F. selysi	MM	Monogynous	Finges	53464450	95.50	97.49	93.65	71.25	13.52
701W1_L7	F. selysi	MM	Monogynous	Finges	65580416	92.49	96.94	88.48	54.32	12.2
De115W1_L2	F. selysi	MM	Monogynous	Derborence	49525732	96.40	96.19	93.78	73.05	13.01
De125W1_L2	F. selysi	MM	Monogynous	Derborence	30879532	96.09	92.36	90.52	66.71	7.4
De172W1_L2	F. selysi	MM	Monogynous	Derborence	58911110	96.28	94.48	91.86	69.64	14.69
De397W1_L2	F. selysi	MM	Monogynous	Derborence	41175296	96.78	91.47	88.68	66.98	10.01
De434W1_L1	F. selysi	MM	Monogynous	Derborence	55543306	95.78	96.07	92.34	67.62	13.37
De44W1_L1	F. selysi	MM	Monogynous	Derborence	53478130	96.05	96.07	93.48	74.46	14.31
De67W1_L1	F. selysi	MM	Monogynous	Derborence	37814522	95.92	96.34	N/A	N/A	19.72
De67W1_L8	F. selysi	MM	Monogynous	Derborence	15216174	91.97	95.43	N/A	N/A	7.46
De67W1_merged	F. selysi	MM	Monogynous	Derborence	N/A	N/A	96.08	96.08	72.39	13.56
De84W1_L1	F. selysi	MM	Monogynous	Derborence	71874630	96.14	96.01	91.76	67.76	17.5
LB1W2_L6	F. selysi	MM	Monogynous	Les_Bussets	64326468	93.86	97.66	92.71	63.73	14.23
LB2W2_L6	F. selysi	MM	Monogynous	Les_Bussets	71478876	94.08	97.75	93.52	67.73	16.85
LB3W2_L6	F. selysi	MM	Monogynous	Les_Bussets	64824832	93.84	97.55	92.33	63.55	14.3
LB4W2_L6	F. selysi	MM	Monogynous	Les_Bussets	64640444	93.74	97.33	92.09	62.85	14.1
LB7W2_L6	F. selysi	MM	Monogynous	Les_Bussets	64693370	93.37	98.06	95.62	73.85	16.49
Sal1W1_L7	F. selysi	MM	Monogynous	Sallanches	68935972	93.44	97.48	92.87	65.72	15.66
Sal2W1_L7	F. selysi	MM	Monogynous	Sallanches	67655620	93.85	97.53	93.31	67.66	15.9
Sal3W1_L7	F. selysi	MM	Monogynous	Sallanches	68520664	93.48	97.51	92.71	64.73	15.34
Sal4W1_L7	F. selysi	MM	Monogynous	Sallanches	64799560	93.05	97.49	93.05	67.35	15.04
174W1_L2	F. selysi	MP	Polygynous	Finges	48014582	95.78	97.53	94.13	67.13	11.45
De107W1_L1	F. selysi	MP	Polygynous	Derborence	51481986	96.82	96.35	93.50	71.85	13.43
De193W1_L2	F. selysi	MP	Polygynous	Derborence	63072008	96.49	96.52	90.04	63.99	14.52
De254W1_L2	F. selysi	MP	Polygynous	Derborence	101490632	96.72	96.15	88.40	61.76	22.63
De287W1_L1	F. selysi	MP	Polygynous	Derborence	50097310	96.98	96.75	91.23	61.72	11.14
LB5W2_L6	F. selysi	MP	Polygynous	Les_Bussets	62342282	92.24	96.42	91.78	63.79	13.58
LB6W2_L6	F. selysi	MP	Polygynous	Les_Bussets	62956778	94.09	96.41	91.35	62.53	13.71
LB8W2_L6	F. selysi	MP	Polygynous	Les_Bussets	80044828	94.07	97.05	91.89	67.17	18.71
Sal10W6_L1	F. selysi	MP	Polygynous	Sallanches	48844726	97.08	95.35	90.01	62.45	11.02
Sal5W3_L1	F. selysi	MP	Polygynous	Sallanches	64225148	96.67	96.90	91.07	62.71	14.48
Sal6W5_L1	F. selysi	MP	Polygynous	Sallanches	53490916	96.99	73.85	71.17	46.60	8.84
Sal9W4_L1	F. selysi	MP	Polygynous	Sallanches	67036034	97.05	96.82	90.38	63.94	15.52
508W1_L1	F. selysi	PP	Polygynous	Finges	57028150	96.34	96.27	91.52	64.01	13.15
703W4_L1	F. selysi	PP	Polygynous	Finges	52121756	95.79	77.58	76.98	67.73	12.62
706W4_L1	F. selysi	PP	Polygynous	Finges	52596572	95.91	96.79	93.13	68.24	12.83
710W2_L7	F. selysi	PP	Polygynous	Finges	60852002	93.46	97.14	92.58	64.56	13.61
713W3_L1	F. selysi	PP	Polygynous	Finges	37857742	96.32	96.37	N/A	N/A	18.89
713W3_L8	F. selysi	PP	Polygynous	Finges	14811142	93.01	95.51	N/A	N/A	7.01
713W3_merged	F. selysi	PP	Polygynous	Finges	N/A	N/A	96.13	96.13	68.79	12.92
715W2_L7	F. selysi	PP	Polygynous	Finges	67859186	93.88	97.02	91.57	62.31	14.71
716W1_L2	F. selysi	PP	Polygynous	Finges	52355190	95.56	93.27	90.30	66.68	12.57
722W1_L2	F. selysi	PP	Polygynous	Finges	54734532	95.49	96.96	94.54	71.38	13.88
733W1_L2	F. selysi	PP	Polygynous	Finges	51794176	95.86	97.02	93.54	67.86	12.54
748W1_L2	F. selysi	PP	Polygynous	Finges	65005654	95.75	96.88	91.00	64.37	14.89
750W2_L7	F. selysi	PP	Polygynous	Finges	66089568	93.59	97.28	92.22	63.77	14.61
De259W2_L1	F. selysi	PP	Polygynous	Derborence	41891668	96.77	96.12	92.40	63.70	9.62
De267W1_L2	F. selysi	PP	Polygynous	Derborence	57378094	96.69	96.13	91.25	63.33	13.12
De299W1_L2	F. selysi	PP	Polygynous	Derborence	53546946	96.35	95.99	90.78	64.22	12.4
De446W2_L1	F. selysi	PP	Polygynous	Derborence	59218802	96.74	94.62	92.15	70.34	15.03
Bar1-w1_L6	F. cinerea	MM	Monogynous	Les_Barges	45934274	96.01	93.03	90.55	65.88	11.19
Bar21-w2_L6	F. cinerea	PP	Polygynous	Les_Barges	53956634	95.32	91.97	89.07	63.68	12.73

Table S1. Summary of the 50 samples used in the analysis. The numbers in the column-Trimmomatic, Mapping, Pre-processing and Quality filter are calculated as percentage of the raw reads after the respective step of the pipeline.

References

- **Avril, A.**, Purcell, J., Béniguel, S., and Chapuisat, M. (2020). Maternal effect killing by a supergene controlling ant social organization. *Proceedings of the National Academy of Sciences*, *117*(29), 17130–17134. https://doi.org/10.1073/pnas.2003282117

- **Ayala, F. J. and Campbell, C. A.** (1974). Frequency-Dependent Selection. *Annual Review of Ecology and Systematics*, *5*(1), 115–138. https://doi.org/10.1146/annurev.es.05.110174.000555

- **Barreiro, L. B.**, Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J. K., Bouchier, C., Tichit, M., Neyrolles, O., Gicquel, B., Kidd, J. R., Kidd, K. K., Alcaïs, A., Ragimbeau, J., Pellegrini, S., Abel, L., Casanova, J. L., and Quintana-Murci, L. (2009). Evolutionary Dynamics of Human Toll-Like Receptors and Their Different Contributions to Host Defense. *PLoS Genetics*, *5*(7), e1000562. https://doi.org/10.1371/journal.pgen.1000562

- **Bodmer, W. F. and Parsons, P. A.** (1962). Linkage and recombination in evolution. *Advances in Genetics*, *11*, 1–100.

- **Bolger, A. M.**, Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

- **Bondroit, J.** (1918). Les fourmis de France et de Belgique. *Annales de La Societe Entomologique de France*, *87*, 1–174.

- **Braim, B.S.** (2015). Exploring the regulatory role of behaviour and genome architecture in the socially polymorphic ant, *Leptothorax acervorum*. PhD thesis (University of Leicester).

- **Brelsford, A.**, Purcell, J., Avril, A., Tran Van, P., Zhang, J., Brütsch, T., Sundström, L., Helanterä, H., and Chapuisat, M. (2020). An Ancient and Eroded Social Supergene Is Widespread across Formica Ants. *Current Biology*, *30*(2), 304–311.e4. https://doi.org/10.1016/j.cub.2019.11.032

- **Broad institute.** (2018). https://github.com/broadinstitute/picard

- **Brown, K. S. and Benson, W. W.** (1974b). Adaptive Polymorphism Associated with Multiple Mullerian Mimicry in Heliconius numata (Lepid. Nymph.). *Biotropica*, *6*(4), 205. https://doi.org/10.2307/2989666

- **Bull, J. J.** (1983). *Evolution of sex determining mechanisms*. Benjamin/Cummings Pub. Co., Advanced Book Program.

- **Carius, H. J.**, Little, T. J., and Ebert, D. (2001). Genetic association in a host-parasite association: Potential for coevolution and frequency-dependent selection. *Evolution*, *55*(6), 1136–1145. https://doi.org/10.1111/j.0014-3820.2001.tb00633.x

- **Chapuisat, M.**, Bocherens, S., and Rosset, H. (2004). Variable queen number in ant colonies: No impact on queen turnover, inbreeding, and population genetic differentiation in ant Formica selysi. *Evolution*, *58*(5), 1064. https://doi.org/10.1554/03-351

- **Charlesworth, B.** (1996). The evolution of chromosomal sex determination and dosage compensation. *Current Biology*, *6*(2), 149–162. https://doi.org/10.1016/s0960-9822(02)00448-7

- **Charlesworth, B. and Charlesworth, D.** (2010). *Elements of Evolutionary Genetics* (1st ed.). W. H. Freeman.

- **Charlesworth, B. and Hughes, K. A.** (2000). *The maintenance of genetic variation in life-history traits*. Cambridge University Press.

- **Charlesworth, D.** (2006). Balancing Selection and Its Effects on Sequences in Nearby Genome Regions. *PLoS Genetics*, *2*(4), e64. https://doi.org/10.1371/journal.pgen.0020064

- **Charlesworth, J. and Eyre-Walker, A.** (2008). The McDonald-Kreitman Test and Slightly Deleterious Mutations. *Molecular Biology and Evolution*, *25*(6), 1007–1015. https://doi.org/10.1093/molbev/msn005

- **Cingolani, P.**, Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., and Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

- **Clarke, B. C.**, Shelton, P. R., and Mani, G. S. (1988). Frequency-dependent selection, metrical characters and molecular evolution. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *319*(1196), 631–640. https://doi.org/10.1098/rstb.1988.0070

- **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., and Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

- **Darlington, C. D. and Mather, K.** (1949). *Elements of Genetics.* George Allen & Unwin Ltd: London.

- **Davis, J. K.**, Mittel, L. B., Lowman, J. J., Thomas, P. J., Maney, D. L., Martin, C. L., and Thomas, J. W. (2011). Haplotype-Based Genomic Sequencing of a Chromosomal Polymorphism in the White-Throated Sparrow (*Zonotrichia albicollis*). *Journal of Heredity*, *102*(4), 380–390. https://doi.org/10.1093/jhered/esr043

- **de Gasperin, O.**, Blacher, P., Grasso, G., and Chapuisat, M. (2020). Winter is coming: harsh environments limit independent reproduction of cooperative-breeding queens in a socially polymorphic ant. *Biology Letters*, *16*(1), 20190730. https://doi.org/10.1098/rsbl.2019.0730

- **Delph, L. F. and Kelly, J. K.** (2013). On the importance of balancing selection in plants. *New Phytologist*, *201*(1), 45–56. https://doi.org/10.1111/nph.12441

- **Dobzhansky, T.** (1951). *Genetics and the origin of species*. New York, NY, USA: Columbia University Press.

- **Dobzhansky, T.** (1982). *Genetics and the Origin of Species (The Columbia Classics in Evolution)*. Columbia University Press.

- **Ejsmond, M.**, Babik, W., and Radwan, J. (2010). MHC allele frequency distributions under parasite-driven selection: A simulation model. *BMC Evolutionary Biology*, *10*(1), 332. https://doi.org/10.1186/1471-2148-10-332

- **Engle, E. K. and Fay, J. C.** (2013). ZRT1 Harbors an Excess of Nonsynonymous Polymorphism and Shows Evidence of Balancing Selection in Saccharomyces cerevisiae. *G3 Genes|Genomes|Genetics*, *3*(4), 665–673. https://doi.org/10.1534/g3.112.005082

- **Fijarczyk, A. and Babik, W.** (2015). Detecting balancing selection in genomes: limits and prospects. *Molecular Ecology*, *24*(14), 3529–3545. https://doi.org/10.1111/mec.13226

- **Fisher, R. A.** (1930). *The Genetical Theory of Natural Selection.* Oxford University Press.

- **Fontcuberta, A.**, de Gasperin, O., Avril, A., Dind, S., & Chapuisat, M. (2021). Disentangling the mechanisms linking dispersal and sociality in supergene-mediated ant social forms. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1949). https://doi.org/10.1098/rspb.2021.0118

- **Ford, E. B.** (1964). *Ecological Genetics*. Methuen: London.

- **Gill, R. J**., Arce, A., Keller, L., and Hammond, R. L. (2009). Polymorphic social organization in an ant. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1677), 4423–4431. https://doi.org/10.1098/rspb.2009.1408

- **Gonzalez, D. R.**, Aramendia, A. C., and Davison, A. (2019). Recombination within the Cepaea nemoralis supergene is confounded by incomplete penetrance and epistasis. *Heredity*, *123*(2), 153–161. https://doi.org/10.1038/s41437-019-0190-6

- **Gouyon, P. H.**, Vichot, F., and van Damme, J. (1991). Nuclear-cytoplasmic male sterility - single-point equilibria versus limit-cycles. *American Naturalist*, *137*(4), 498–514.

- **Gray, S. M. and McKinnon, J. S.** (2007). Linking color polymorphism maintenance and speciation. *Trends in Ecology & Evolution*, *22*(2), 71–79. https://doi.org/10.1016/j.tree.2006.10.005

- **Harpur, B. A. and Zayed, A.** (2013). Accelerated Evolution of Innate Immunity Proteins in Social Insects: Adaptive Evolution or Relaxed Constraint? *Molecular Biology and Evolution*, *30*(7), 1665–1674. https://doi.org/10.1093/molbev/mst061

- **Hedrick, P. W.** (1986). Genetic Polymorphism in Heterogeneous Environments: A Decade Later. *Annual Review of Ecology and Systematics*, *17*(1), 535–566. https://doi.org/10.1146/annurev.es.17.110186.002535

- **Hedrick, P. W.** (2007). Sex: Differences in mutation, recombination, selection, gene flow, and genetic drift. *Evolution*, *61*(12), 2750–2771. https://doi.org/10.1111/j.1558-5646.2007.00250.x

- **Hedrick, P. W.** (2012). What is the evidence for heterozygote advantage selection? *Trends in Ecology & Evolution*, *27*(12), 698–704. https://doi.org/10.1016/j.tree.2012.08.012

- **Hedrick, P. W.**, Lee, R. N., and Garrigan, D. (2002). Major histocompatibility complex variation in red wolves: evidence for common ancestry with coyotes and balancing selection. *Molecular Ecology*, *11*(10), 1905–1913. https://doi.org/10.1046/j.1365-294x.2002.01579.x

- **Jamie, G. A. and Meier, J. I.** (2020). The Persistence of Polymorphisms across Species Radiations. *Trends in Ecology & Evolution*, *35*(9), 795–808. https://doi.org/10.1016/j.tree.2020.04.007

- **Jay, P.,** Whibley, A., Frézal, L., Rodríguez De Cara, M. N., Nowell, R. W., Mallet, J., Dasmahapatra, K. K., and Joron, M. (2018). Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. *Current Biology*, *28*(11), 1839–1845.e3. https://doi.org/10.1016/j.cub.2018.04.072

- **Keller, L. and Passera, L.** (1989). Size and fat content of gynes in relation to the mode of colony founding in ants (Hymenoptera; Formicidae). *Oecologia*, *80*(2), 236–240. https://doi.org/10.1007/bf00380157

- **Kim, Y. and Stephan, W.** (2002). Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome. *Genetics*, *160*(2), 765–777. https://doi.org/10.1093/genetics/160.2.765

- **Klien, J.** (1987). The major histocompatibility complex and protein recognition by T lymphocytes. *Advances in Experimental Medicine and Biology*, *225*, 1–10.

- **Levene, H.** (1953). Genetic Equilibrium When More Than One Ecological Niche is Available. *The American Naturalist*, *87*(836), 331–333. https://doi.org/10.1086/281792

- **Lewontin, R. C.** (1974). *The genetic basis of evolutionary change*. New York, NY, USA: Columbia University Press.

- **Li, H.** (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997v2 [q-Bio.GN]*. Published.

- **Li, H.**, Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

- **Llaurens, V.**, Whibley, A., and Joron, M. (2017). Genetic architecture and balancing selection: the life and death of differentiated variants. *Molecular Ecology*, *26*(9), 2430–2448. https://doi.org/10.1111/mec.14051

- **Lude, A**., Reich, M., and Plachter, H. (1999). Life Strategies of Ants in Unpredictable Floodplain Habitats of Alpine Rivers (Hymenoptera: Formicidae). *Entomologia Generalis*, *24*(2), 75–91. https://doi.org/10.1127/entom.gen/24/1999/75

- **Mallet, J. and Joron, M.** (1999). Evolution of Diversity in Warning Color and Mimicry: Polymorphisms, Shifting Balance, and Speciation. *Annual Review of Ecology and Systematics*, *30*(1), 201–233. https://doi.org/10.1146/annurev.ecolsys.30.1.201

- **Nielsen, R.** (2005). Molecular Signatures of Natural Selection. *Annual Review of Genetics*, *39*(1), 197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420

- **Ortiz, E. M.** (2019). vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. *Github*. Published. https://doi.org/10.5281/zenodo.2540861

- **Poplin, R.**, Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., and Banks, E. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, *201178*. https://doi.org/10.1101/201178

- **Pracana, R.**, Priyam, A., Levantis, I., Nichols, R. A., and Wurm, Y. (2017). The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. *Molecular Ecology*, *26*(11), 2864–2879. https://doi.org/10.1111/mec.14054

- **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. https://doi.org/10.1016/j.cub.2014.09.071

- **Purcell, J.**, Pellissier, L., and Chapuisat, M. (2015). Social structure varies with elevation in an Alpine ant. *Molecular Ecology*, *24*(2), 498–507. https://doi.org/10.1111/mec.13042

- **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., and 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. https://doi.org/10.1086/519795

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

- **Rosset, H.**, & Chapuisat, M. (2006). Alternative life-histories in a socially polymorphic ant. *Evolutionary Ecology*, *21*(5), 577–588. https://doi.org/10.1007/s10682-006-9139-3

- **Richman, A.** (2000). Evolution of balanced genetic polymorphism. *Molecular Ecology*, *9*(12), 1953–1963. https://doi.org/10.1046/j.1365-294x.2000.01125.x

- **Sabeti, P. C.**, Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., and Lander, E. S. (2007). Genome-wide

detection and characterization of positive selection in human populations. *Nature*, *449*(7164), 913–918. https://doi.org/10.1038/nature06250

- **Schwander, T.**, Libbrecht, R., & Keller, L. (2014). Supergenes and Complex Phenotypes. *Current Biology*, *24*(7), R288–R294. https://doi.org/10.1016/j.cub.2014.01.056

- **Ségurel, L.**, Gao, Z., & Przeworski, M. (2013). Ancestry runs deeper than blood: The evolutionary history of ABO points to cryptic variation of functional importance. *BioEssays*, *35*(10), 862–867. https://doi.org/10.1002/bies.201300030

- **Seifert, B.** (2002). A taxonomic revision of the Formica cinerea group (Hymenoptera: Formicidae). *Abhandlungen Berichte Naturkundemus. Görlitz*, *74*(2), 245–272.

- **Seifert, B.** (2003). A taxonomic revision of the Formica cinerea group (Hymenoptera, Formicidae). *Abhandlungen Und Berichte Des Naturkundemuseseums*, *74*, 245–272.

- **Sheppard, P. M.** (1953). Polymorphism, linkage and the blood groups. *American Naturalist*, *87*, 283–294.

- **Stamatakis, A.** (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

- **Stoletzki, N. and Eyre-Walker, A.** (2010). Estimation of the Neutrality Index. *Molecular Biology and Evolution*, *28*(1), 63–70. https://doi.org/10.1093/molbev/msq249

- **Tajima, F.** (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

- **Takahata, N. and Nei, M.** (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, *124*(4), 967–978. https://doi.org/10.1093/genetics/124.4.967

- **Tarasov, A.**, Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics*, *31*(12), 2032–2034. https://doi.org/10.1093/bioinformatics/btv098

- **Thompson, M. J. and Jiggins, C. D.** (2014). Supergenes and their role in evolution. *Heredity*, *113*(1), 1–8. https://doi.org/10.1038/hdy.2014.20

- **Tuttle, E.**, Bergland, A., Korody, M., Brewer, M., Newhouse, D., Minx, P., Stager, M., Betuel, A., Cheviron, Z., Warren, W., Gonser, R., and Balakrishnan, C. (2016). Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Current Biology*, *26*(3), 344–350. https://doi.org/10.1016/j.cub.2015.11.069

- **Uyenoyama, M. K.** (2004). Evolution under tight linkage to mating type. *New Phytologist*, *165*(1), 63–70. https://doi.org/10.1111/j.1469-8137.2004.01246.x

- **van Valen, L.** (1973). A new Evolutionary Law. *Evolutionary Theory*, *1*, 1–30.

- **Wang, J.**, Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y. C., Shoemaker, D., and Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, *493*(7434), 664–668. https://doi.org/10.1038/nature11832

- **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., and Keller, L. (2020). Evolution of a supergene that regulates a trans-species social polymorphism. *Nature Ecology & Evolution*, *4*(2), 240–249. https://doi.org/10.1038/s41559-019-1081-1

- **Zahnd, S.**, Fontcuberta, A., Koken, M., Cardinaux, A., and Chapuisat, M. (2021). Fine-scale habitat heterogeneity favours the coexistence of supergene-controlled social forms in Formica selysi. *BMC Ecology and Evolution*, *21*(1). https://doi.org/10.1186/s12862-020-01742-0