Accepted author's copy (postprint) of a manuscript published in ***Current Biology***, 2020, 30: 304-311.
doi: 10.1016/j.cub.2019.11.032
Publisher copy available at https://doi.org/10.1016/j.cub.2019.11.032

An ancient and eroded social supergene is widespread across *Formica* ants

Alan Brelsford [1,3*], Jessica Purcell [2,3*], Amaury Avril [3], Patrick Tran Van[3], Junxia Zhang[2, 4], Timothée Brütsch [3], Liselotte Sundström [5], Heikki Helanterä [6], Michel Chapuisat [3]

[*] These authors contributed equally

[1] Department of Evolution, Ecology and Organismal Biology
University of California Riverside
Riverside, CA 92521
USA

[2] Department of Entomology
University of California Riverside
Riverside, CA 92521
USA

[3] Department of Ecology and Evolution
University of Lausanne
1015 Lausanne
Switzerland

[4] College of Life Sciences
Hebei University
Baoding, Hebei, 071002
China

[5] Organismal and Evolutionary Biology research Programme
Faculty of Biological and Environmental Sciences
University of Helsinki
00014 Helsinki
Finland

[6] Ecology and Genetics Research Unit
University of Oulu
90014 Oulu
Finland

Correspondence to: alan.brelsford@ucr.edu, jessica.purcell@ucr.edu, or
michel.chapuisat@unil.ch

Running title: Supergene evolution in ants

Summary: Supergenes, clusters of tightly linked genes, play a key role in the evolution of complex adaptive variation [1,2]. While supergenes have been identified in many species, we lack an understanding of their origin, evolution and persistence [3]. Here, we uncover 20-40 MY of evolutionary history of a supergene associated with polymorphic social organization in *Formica* ants [4]. We show that five *Formica* species exhibit homologous divergent haplotypes spanning 11 Mbp on chromosome 3. Despite the supergene's size, only 142 single nucleotide polymorphisms (SNPs) consistently distinguish alternative supergene haplotypes across all five species. These conserved trans-species SNPs are localized in a small number of disjunct clusters distributed across the supergene. This unexpected pattern of divergence indicates that the *Formica* supergene does not follow standard models of sex chromosome evolution, in which distinct evolutionary strata reflect an expanding region of suppressed recombination (e.g. [5]). We propose an alternative "eroded strata model," in which clusters of conserved trans-species SNPs represent functionally important areas maintained by selection in the face of rare recombination between ancestral haplotypes. The comparison of whole genome sequences across 10 additional *Formica* species reveal that the most conserved region of the supergene contains a transcription factor essential for motor neuron development in *Drosophila* [6]. The discovery that a very small portion of this large and ancient supergene harbors conserved trans-species SNPs linked to colony social organization suggests that the ancestral haplotypes have been eroded by recombination, with selection preserving differentiation at one or a few genes generating alternative social organization.

Results and Discussion:

Each year, new systems with tightly linked clusters of genes are discovered, pointing to the importance of supergenes in the evolution of certain classes of complex traits, including mimetic coloration in butterflies, self-incompatibility in plants, mating strategies in birds, mating types in fungus, and social organization in ants [1, 2, 7-10]. While the prevalence and impact of supergenes is increasingly clear, there are still large

gaps in our understanding of how they evolve, whether they tend to be transient or stable, and how much of the non-recombining region actually shapes the trait of interest.

Using a comparative approach, we investigate the evolutionary history of an autosomal supergene associated with colony social organization in the Alpine silver ant *Formica selysi* [4]. First, we examine whether this supergene system is stable or ephemeral by investigating whether it is present and has a similar function in five socially polymorphic *Formica* species, representing an estimated 20-40 MY of independent evolutionary history (Figure S1). This divergence time exceeds the age of inversion-based autosomal supergenes described so far [3]. Second, we use phylogenetic comparisons across the five species to infer how the supergene evolved. Specifically, we assess whether recombination was suppressed at different times across the length of the supergene and identify conserved trans-species single-nucleotide polymorphisms (SNPs) associated with social organization.

In *F. selysi,* alternative haplotypes of the supergene are associated with alternative colony social organization, namely whether the colony is headed by one queen (= monogyne) or by multiple queens (=polygyne) [4]. Monogyne colonies exclusively harbor individuals carrying one haplotype, Sm, whereas polygyne colonies always harbor individuals bearing at least one copy of the alternative haplotype, Sp [4, 11]. Queen number is also associated with a suite of individual and colony-level traits, including body size, colony size and reproductive strategy [12].

Many other *Formica* species are socially polymorphic [13-18]. So far, no genetic polymorphism associated with colony social organization has been documented outside of *F. selysi*. This absence may reflect phenotypic plasticity in colony queen number. Alternatively, a genomic basis to social organization may have remained undetected in previous studies based on few genetic markers [15, 16, 18].

We tested whether social organization was controlled by a conserved ancestral supergene across socially polymorphic *Formica* species. We collected ddRADseq population genomic data on five focal polymorphic species (Table S1): *F. truncorum (*subgenus *Formica* sensu stricto, 20 individuals, 24,431 sites, mean depth 17.9), *F. exsecta* (Coptoformica, 41 individuals, 24,577 sites, mean depth 17.2), *F. selysi* (Serviformica, 83 individuals, 21,554 sites, mean depth 14.2), *F. cinerea* (Serviformica,

161 individuals, 44,427 sites, mean depth 27.7), and *F. lemani* (Serviformica, 65 individuals, 64,260 sites, mean depth 15.3). In each species, we find elevated differentiation between individuals of monogyne and polygyne origin at chromosome 3 compared with other chromosomes (Figure 1), suggesting that an ancestral supergene is present and associated with colony queen number in the five species. Through principal component analysis (PCA) of variation on this chromosome, we show that individuals of monogyne origin are usually homozygous for one supergene haplotype, while individuals of polygyne origin are usually either heterozygous or are homozygous for an alternative haplotype (Figure S2). This association is perfect in *F. selysi*, *F. exsecta*, and *F. truncorum*, while mismatches between social structure and supergene genotype are observed in 5 of 39 *F. lemani* and 35 of 96 *F. cinerea* individuals with known social structure. Nonetheless, association between the presence of an Sp haplotype and polygyne social structure was significant even in the latter two species (Fisher's exact test $p = 0.00002$ in *F. cinerea*, $p = 0.000002$ in *F. lemani*). To further investigate the relationship between each haplotype across species, we selected homozygous workers or haploid males for subsequent whole genome sequencing. For *F. lemani,* we included three individuals, representing homozygotes for three alternative haplotypes (Figure S2).

Sex chromosomes are the most widely known and best understood class of supergenes [2, 19], and may provide a model for the evolution of autosomal supergenes. In the old and highly conserved sex chromosomes of birds and mammals, the regions of suppressed recombination have expanded over time, as new adjacent regions were inverted or otherwise rearranged [20, 21]. Blocks of the Z/W and X/Y chromosomes wherein recombination ceased at the same time during their evolutionary history are known as evolutionary 'strata.' For instance, comparisons of genome sequences from 17 bird species distributed across the phylogeny revealed that avian sex chromosomes have one small region where suppressed recombination predates the divergence of ratites from other birds. A large region of the Z and W chromosomes continues to recombine in ratites, while additional non-recombining strata accumulated over time in other avian lineages [5]. Whether this "expanding strata" model applies to autosomal supergenes remains an open question.

The supergene shared by multiple *Formica* species provides a great opportunity to reconstruct how alternative haplotypes evolved. We identify regions of the supergene that are consistently differentiated between social forms across the *Formica* species. By mapping conserved trans-species SNPs associated with social organization and reconstructing the phylogenetic topology across the *Formica* social supergene, we investigate whether recombination was suppressed at different times across the length of the supergene, forming evolutionary strata.

If the *Formica* supergene evolves according to the expanding strata model, one region of the supergene is expected to exhibit an 'old strata' topology, wherein the haplotypes of all five species cluster by social form. Other regions might exhibit intermediate strata topologies, wherein alternative supergene haplotypes cluster among closely related species but not distantly related species. The recombining ends of the supergene are expected to follow a 'young strata' topology, wherein individuals cluster by species regardless of social form. Moreover, the expanding strata model predicts that the old and intermediate strata would span entire inversions, such that each inversion would be acquired sequentially during the evolutionary history of the supergene. In contrast, models of genome evolution within single inversions predict that only inversion breakpoints and loci under selection will remain differentiated in very old inversion polymorphisms [22, 23].

We sequenced the genomes of representatives of each social form from the five focal *Formica* species, aligned them to a new chromosome-level genome assembly for *F. selysi*, and plotted the number of trans-species fixed differences per 1 kbp window between the monogyne- and polygyne-associated haplotypes (Figure 2). Moreover, we identified transitions in phylogenetic topology across the supergene with a hidden Markov model implemented in Saguaro [24]. Contrary to the predictions of the expanding strata model, we found multiple very small regions containing 142 conserved trans-species SNPs that clustered by social form (Figure 2A). These regions matched sections of the supergene with 'old strata' topologies (Figure 2B). The cumulative length of these small disjunct conserved regions with 'old strata' topologies was 136 kbp, which amounts to only 1.2% of the non-recombining supergene or 0.96% of the entire

chromosome. No such conserved trans-species SNPs were found on any other chromosome, across a total of 11.4 million SNPs genome-wide.

Small regions with trans-species fixed SNPs could be due to balancing selection or physical constraints (e.g. inversion breakpoints) that prevent recombination from homogenizing these genomic regions [22]. To distinguish between these hypotheses and further test the expanding strata model, we identified genomic rearrangements between the alternative supergene haplotypes. We constructed high density linkage maps using ddRAD genotypes from the female offspring of two Sp/Sp *F. selysi* queens (112 offspring total, 1792 and 3688 markers, mean sequence depth 36.6). We also constructed a linkage map from the male offspring of one Sm/Sm *F. exsecta* queen (67 offspring, 4603 markers, mean sequence depth 17.3) to determine whether the structure of the Sm haplotype is conserved across species. We mapped the positions of the old, intermediate, and young strata onto the *F. selysi* Sm genome. We then aligned the genome to the linkage maps. The Sm haplotype of *F. exsecta* was collinear with that of *F. selysi (*Figure 3). The conserved gene order on the Sm haplotype suggests that this haplotype is ancestral. In contrast, the alignment of the Sp haplotype of *F. selysi* to the Sm genome revealed at least four inversions along the length of the supergene (Figure 3). Regions of the supergene exhibiting the 'old strata' topologies were not localized on a single inversion, but instead were distributed across the supergene (Figures 2, 3), suggesting that haplotypes spanning the entire non-recombining region began to diverge prior to the divergence of all the *Formica* species we examined. At least some trans-species fixed SNPs were not close to inversion breakpoints based on a qualitative assessment of the linkage maps, suggesting that balancing selection, and not exclusively structural constraint, plays a role in maintaining these SNPs. Occasional recombination is suggested by intermediate strata topologies that were patchily distributed across the length of the supergene (Figure 3). As expected, the recombining regions at the ends of the supergene followed the 'young strata' pattern (Figures 2, 3).

Overall, the pattern of differentiation within the *Formica* supergene differs strikingly from the predictions of the expanding strata model (Figures 2-4). We propose that the *Formica* supergene results from a long history of rare recombination [25] and/or gene conversion [26, 27]) between alternative haplotypes in different lineages (Figure 4).

We previously observed evidence of rare recombination between Sm and Sp haplotypes in *F. selysi* [4], and similar observations have been recorded in the fire ant supergene system [28] and in a newly described inversion polymorphism in the great tit [29]. Under this alternative "eroded strata model", an initial event, such as an inversion, greatly reduced recombination across the length of the supergene in the common ancestor of the focal species (Figure 4). Next, occasional recombination homogenized the monogyne- and polygyne- associated haplotypes in portions of the chromosome, while selection on functionally important genes and regulatory regions, or structural constraints at inversion breakpoints, maintained small regions with the old strata topology. Over time, rare recombination events in regions not under selection eroded the ancestral strata, breaking up associations between alleles within each alternative haplotype and leaving only small disjunct areas with conserved trans-species polymorphisms (Figure 4). This model is consistent with analytical results obtained in models of genome evolution on single inversions [22]; our results provide empirical support for this model and scale it to a large supergene harboring multiple inversions.

Trans-species SNPs associated with a trait of interest can point to genomic regions responsible for the trait [30-32]. To characterize the most conserved trans-species SNPs and identify candidate genes determining alternative social organization, we sequenced the genomes of 10 additional European *Formica* species (Figure 2). Six of these additional species spanning three subgenera matched the Sm haplotype for 126 out of the 142 conserved SNPs associated with social organization in the initial comparative analysis of 5 focal species. One *Formica* sensu stricto matched the Sp haplotype across 113 of 135 conserved SNPs. Finally, three species had excess heterozygosity across the whole supergene and were heterozygous at a subset of the conserved SNPs (Figure 2). Overall, only 20 SNPs were conserved across all 15 *Formica* species. All but one of these conserved SNPs were located in the last exon and 3' untranslated region of the gene *Knockout* (Figure 2). The gene *Knockout* is a storkhead-box transcription factor essential for motor neuron development in *Drosphila* [6]. Additional SNPs conserved across all species except *F. picea* occurred in an intron of serine-threonine kinase *STK32B*, an exon of mitochondrial ribosomal protein *MRPL34*, and regions just downstream of the genes *RPUSD4* and *G9A*.

Whether other supergenes follow the eroded strata model is not yet clear, but several common characteristics suggest that some might do so. The *Formica* social supergene and the ruff autosomal supergene appear to differ from ancient sex chromosomes by the occurrence of rare events of recombination between alternative haplotypes [4, 33, 34]. The independent supergenes underlying coloration and mating strategies in ruffs and white-throated sparrows likely originated from inversions [33-35]. Both of these avian supergenes are much younger than the *Formica* supergene and are only found in a single species, which limits the possibility to test whether they follow an eroded strata model. Alternative haplotypes at the supergene underlying mimetic coloration in *Heliconius numata* apparently evolved sequentially, with one alternative haplotype containing a single inversion and a second alternative haplotype harboring the initial inversion and an adjacent second inversion [36]. The diversity of color patterns can be traced to a relatively small number of genetic 'modules' that underlie different color patches on butterfly wings [37]. As in the *Formica* supergene, the 'modules' often span very small portions of the genome and exhibit a different evolutionary history from one another and from whole genome patterns. Jay et al. [10] demonstrate that these alternative topologies result in some cases from introgression of modules between species. The contribution of introgression to evolutionary patterns in the *Formica* supergene remains to be investigated.

Recent studies discovered that independent, convergent supergenes underlie polymorphisms in social organization in at least three ant lineages (*Solenopsis invicta,* [38]; *Formica selysi,* [4]; *Leptothorax acervorum,* [39]). We do not yet know the extent of similarities in the evolutionary history of these convergent 'social' supergenes [4, 38]. An analysis of divergence between *S. invicta* SB and Sb haplotypes revealed no evidence of evolutionary strata [40], despite the presence of at least two inversions [41]. However, an early analysis of the odorant binding protein gene Gp9, which was subsequently found to be contained within the *Solenopsis* supergene, identified conserved polymorphisms across several *Solenopsis* species [42], and this was confirmed in a recent comparative genomic analysis of *S. invicta, S. richteri*, and *S. quinquecuspis* [43]. The combination of a lack of strata, multiple inversions, and trans-species polymorphism suggests that an

expanded multi-species analysis in *Solenopsis* would provide an interesting point of comparison with the *Formica* supergene.

Our study of the *Formica* supergene suggests several directions for future research. So far, we have investigated the DNA sequence differences between alternative supergene haplotypes in multiple species; comparison of gene expression patterns between individuals with each genotype across different species, both in general and within the candidate genes identified herein, could provide insights into the functional differences of each haplotype. Moreover, we have not analyzed copy-number variation in the *Formica* supergene haplotypes, but identifying haplotype-specific duplication or deletion of genes, or insertion of transposable elements, could point to variants that affect the different functions of the Sm and Sp haplotypes (e.g., [33, 43]). Both of these future directions would be enhanced by the development of high quality genome assemblies for additional *Formica* species, which would allow more precise identification of inversion breakpoints on the Sp haplotype (e.g., [43]) and enable researchers to test the robustness of our results when aligning to different genomes. Given the variation in genetic control and haplotype diversity uncovered in *F. lemani* and *F. cinerea*, it would also be valuable to examine non-genetic influences on social structure in these species, and to more broadly investigate geographic variation in the strength of association between the supergene and social organization by sampling a larger number of species across their range.

Overall, this comparative analysis revealed that at least five species of the genus *Formica* separated by up to 20-40 MY of independent evolution harbor an ancient supergene that contributes to polymorphism in social organization. This ancestral supergene followed an unusual evolutionary trajectory. We suggest that rare recombination between alternative haplotypes in different lineages reduced trans-species divergence, resulting in patterns of genetic differentiation that differ markedly from the expanding strata expected under standard models of sex chromosome evolution. The genomic signature of this novel "eroded strata model" is the presence of very small clusters of conserved trans-species SNPs that consistently differ between alternative haplotypes across multiple species. Across the *Formica* genus, these conserved trans-species SNPs highlight regions of the supergene that likely have an important function

both in its inception and in the ongoing control of colony social organization. The great diversity in origin, structure, size, and evolution of autosomal and sex-linked supergenes is intriguing. Further comparisons will reveal which key biological differences send supergenes on divergent evolutionary trajectories.

Acknowledgments

Funding was provided by Swiss National Science Foundation grants 31003A-146641 and 31003A-173189 to M.C., and by University of Lausanne and University of California Riverside. Computational resources were provided by the Vital-IT computing cluster at the Swiss Institute for Bioinformatics and the High Performance Computing Cluster at University of California Riverside. We thank Purcell, Brelsford, and D. Schluter lab members for helpful comments and discussions, and K. Loope and K. Thompson for detailed feedback on an earlier version of the manuscript.

Author Contributions

Conceptualization, AB, JP, MC; Formal Analysis, AB, JP, PTV, JZ; Investigation, AB, JP, AA, JZ, TB; Resources, JP, HH, LS, AA, AB; Writing – original draft, JP, AB; Writing – review and editing, JP, AB, MC, HH, JZ, PTV; Visualization, AB, JP, JZ; Supervision, MC, JP; Funding acquisition, MC, JP.

Declaration of Interests

The authors declare no competing interests.

Figure Legends

Figure 1: An ancestral supergene is associated with colony social organization across five polymorphic *Formica* species. In each of the five species (A-E), elevated differentiation (F_{ST}) occurred between individuals of monogyne and polygyne origin across much of chromosome 3, in contrast to lower levels of differentiation in the rest of the genome based on population ddRAD data. The phylogenetic relationships between the species based on genome-wide SNP data, excluding chromosome 3, is indicated on the right (F). Note that the maximum differentiation between monogynes and polygynes is influenced by the ploidy and the population genetic structure of the sequenced individuals (Table S1). See also Figures S1, S2 and Table S1.

Figure 2: Evolution of alternative haplotypes of the social supergene across species of the genus *Formica*. Fixed differences between Sm and Sp haplotypes across five focal species, based on whole-genome sequence data, are concentrated in small regions across chromosome 3 (A; number of conserved trans-species SNPs associated with social organization in 1 kbp windows; n=142). In several small regions distributed across the center of the chromosome, the sequences cluster by social form (red, "old strata" topology), while in large regions at the chromosome ends the sequences cluster by species (blue, "young strata" topology) (B; Hidden Markov Model of tree topology implemented in Saguaro). The 142 SNPs with fixed differences between Sm and Sp haplotypes across the five focal species were sequenced in single representatives of 10 additional species (C), with alleles matching the Sm haplotype shown in green and alleles matching the Sp haplotype shown in orange. Only a single region of 1,021 bp (positions 11,910,116 – 11,911,137) harbors SNPs that are consistently fixed between Sm and Sp haplotypes across all 15 species. See also Figure S1 and Table S1.

Figure 3: Structural rearrangements between alternative haplotypes of the *Formica* supergene. The chromosome-level *Formica selysi* genome assembly for the Sm supergene haplotype (middle; PacBio long read sequencing combined with linkage map) is collinear with the Sm haplotype of *F. exsecta* (top; linkage map from a *F. exsecta* Sm/Sm family). In contrast, the Sp haplotype of *F. selysi* reveals several inversions and rearrangements compared to the Sm haplotype (bottom; merged linkage map from two *F. selysi* Sp/Sp families). Lines between bars connect the RADtags in the linkage maps to their position in the *F. selysi* genome assembly. Colored bars along the Sm haplotype of *F. selysi* indicate the strata topologies inferred by Saguaro from whole-genome sequence data across five *Formica* species. Blue bars represent regions of the supergene where sequences cluster by species (young strata, B). Red bars show sections of the supergene where sequences cluster by social form across all five species (old strata, C). Purple bars show sections where sequences cluster by social form in the three Serviformica species, and, separately, cluster by social form in *F. truncorum* and *F. exsecta* (D). Green bars represent sections where sequences cluster by social form in the three Serviformica species, but cluster by species for *F. truncorum* and *F. exsecta* (E). See also Figure S1.

Figure 4: Comparison of the eroded strata model and the expanding strata model. In the eroded strata model (left panel), an initial inversion in one chromosome (basal blue rectangle) greatly reduces recombination between two alternative haplotypes (red bars). As new species form, this ancestral polymorphism is maintained, but occasional recombination or gene conversion events (colored lines) homogenize sections of the region in some lineages. The time series of plots at left represent the trans-species divergence pattern expected under the eroded strata model, with disjunct regions containing conserved trans-species polymorphisms. In contrast, in the expanding strata model (right panel), new non-recombining regions appear sequentially in diverging lineages, resulting in a pattern wherein young strata exhibit lower trans-species differentiation than old strata (time series of plots at right). The topologies in the central panel show the relationships between haplotypes and species for the young and old strata scenarios, with colors matching the strata colors shown in each tree.

STAR Methods:

Detailed methods are provided in the online version of this paper and include the following:

Key Resources Table
Lead Contact and Materials Availability
Experimental Model and Subject Details
Method Details
    Supergene presence in multiple species
    Linkage maps
    Genome assembly
    Whole-genome resequencing
    Phylogeny and dating
Quantification and Statistical Analysis
Data and software availability

Key Resource Table

Attached as separate document

Lead Contact and Materials Availability

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Alan Brelsford (alan.brelsford@ucr.edu). There are restrictions to the availability of tissue and DNA samples due to the lack of an external centralized repository for their distribution and our need to maintain the stock. We are glad to share oligonucleotides with reasonable compensation by requestor for processing and shipping.

Experimental Model and Subject Details

With the exception of the linkage map of the *F. selysi* Sp haplotype, all ants used in this study were collected in the wild (sample sizes and localities for each species in Table S1). For the *F. selysi* Sp haplotype linkage map, we obtained captive-reared offspring of two mature queens from polygynous field colonies in Finges, Switzerland. The supergene genotype of these two queens had been previously determined to be Sp/Sp [11]. Queens were kept in isolated plastic nest boxes (15 x 13 x 6 cm) containing a tube with water and ad libitum access to ant food consisting of agar, egg, and sugar, with at least 20 nestmate workers, and left to produce eggs. These queens produced 35 and 77 newly emerged worker offspring, respectively, and we collected these for linkage mapping. All ants used in this study were stored in 100% ethanol prior to DNA extraction.

Method details

Supergene presence in multiple species

We collected workers and males from colonies of four *Formica* species (*F. cinerea, F. exsecta, F. lemani, F. truncorum;* sample sizes and countries of origin in Supplementary Materials Table S1). DNA was isolated from the head and thorax of each ant using a DNeasy Blood and Tissue kit (Qiagen). We collected ddRAD sequence data on these individuals using the protocol of [59], using restriction enzymes EcoRI and MseI. Briefly, we digested genomic DNA with EcoRI and MseI, ligated barcoded adapters to the resulting fragments, removed short fragments with AMPure magnetic beads, amplified fragments using PCR primers incorporating an index sequence, pooled the resulting amplicons, selected fragments of 300-500 bp by agarose gel electrophoresis, and performed a final AMPure bead cleanup on the pooled, size-selected library. Libraries were sequenced at the Lausanne Genomic Technologies Facility on an Illumina HiSeq 2500 with 100bp single-end reads. For a subset of individuals, colony social structure had been previously determined through parentage analysis of microsatellite genotypes [13, 16, 18, 60, 61] or by direct observation of multiple queens during sample collection. For subsequent steps, we reanalyzed previously published data for male *F. selysi* [4] as well as new data from the four additional species.

Reads were demultiplexed using the process_radtags module of Stacks 1.19 [44]. We mapped reads to the *F. selysi* genome using Bowtie 2.3.4.1 [45], called variants separately for each species with Samtools 0.1.19 [46], and filtered the resulting variants with VCFtools 0.1.13 [47], excluding indels and retaining SNP markers with missing data <20%, and minor allele frequency >5%. For each species, we extracted variants on linkage group 3, which contains the social supergene in *F. selysi*, and performed a principal component analysis using PLINK 1.90 [48]. Additionally, we estimated heterozygosity (F_{IS}) per individual and Weir and Cockerham's [62] F_{ST} between workers from monogynous and polygynous colonies across the entire genome in sliding 400 kbp windows with 300 kbp overlap between adjacent windows, using VCFtools 0.1.13 [47].

Linkage maps

We collected ddRAD sequence data on offspring of two *F. selysi* homozygous Sp/Sp queens (77 and 35 newly emerged workers, respectively) and 67 males collected from a monogyne *F. exsecta* colony. Library preparation, sequencing, and SNP calling were carried out as described above in the *Supergene presence in multiple species* section. We filtered raw variant calls separately for each mapping family using VCFtools version 0.1.13 [47], retaining genotypes of SNP and indel variants with quality score >20, and variants with <20% missing data per family and per-family minor allele frequency >15%. We then inferred linkage maps for each family using MSTmap [49], using the Kosambi mapping function and p-value cutoffs of 5e-5 for the smaller *F. selysi* family and 5e-6 for the *F. exsecta* family and larger *F. selysi* family; full parameter sets are reported in Table S2. Linkage maps for two Sp/Sp families were merged using MergeMap [50], weighting each map by the number of individuals used to construct it.

Genome assembly

We collected 20 males from a single monogyne colony. High molecular weight DNA from head and thorax of the males was extracted following [63]. Briefly, cells were lysed with an SDS-based lysis buffer, proteins precipitated by addition of potassium acetate, DNA bound to SeraMag beads and washed with ethanol before elution. PacBio sequencing libraries were prepared with a SMRTbell Template Prep Kit sequenced on 26 SMRT cells of PacBio RSII (Pacific Biosciences) using P6-C4 chemistry at the Lausanne Genomic Technologies Facility.

Raw PacBio reads were error corrected, trimmed and de novo assembled with CANU v1.7 [51] using default parameters. The genome assembly was decontaminated with BlobTools v1.0 [52] under the taxrule 'bestsumorder'. The hit file was obtained by blastn v2.7.1+ alignment to the NBCI nt database, searching for hits with an e-value below 1e-25 (Parameters: -max_target_seqs 10 -max_hsps 1 -evalue 1e-25). Coverage information was taken from the contig headers supplied by CANU. Only contigs with no hits or at least one arthropod hit were retained in the decontaminated assembly.

Subreads were mapped against the decontaminated genome assemblies using pbalign v0.3.0 and Samtools v1.4 [46] in order to perform a polishing step.
The polishing step was done using the GenomicConsensus v2.2.2 package with the Quiver method. Finally, additional filtering steps were applied: redundant polished

contigs were removed using Redundans v0.13c [53] and low-coverage (<15X) contigs were removed. Output statistics are provided in Table S3.

Assembled contigs were joined into chromosome-level scaffolds using a consensus linkage map, constructed using MergeMap [50] on three *F. selysi* families (one SmSm, [4]; two SpSp, this study) and one *F. exsecta* family (SmSm, this study), weighting each map by the number of individuals used to construct it. The two SpSp families were excluded for Scaffold 3. We extracted 1 kbp of sequence surrounding each mapped marker from the highly fragmented Illumina genome assembly [4], and aligned these sequences to the PacBio contigs using Blastn. All contigs containing at least two markers with different positions on the linkage map were placed and oriented on the linkage map; scaffolds were constructed manually based on contig order and orientation on the linkage map.

Whole-genome sequences

Based on the PCA results, we selected haploid or homozygous exemplars of the Sm and Sp haplotypes in each species for whole-genome sequencing (Tables S1, S4). We sequenced one individual for each of the two Sp haplotypes found in *F. lemani*. Additionally, we sequenced the genomes of one individual from each of ten additional species and three outgroup species (*Iberoformica subrufa, Polyergus vinosus, Polyergus mexicanus*) to an average depth of 9.6x. Library preparation and sequencing were performed at the Lausanne Genomic Technologies Facility and the UC Berkeley Vincent Coates Genome Sequencing Laboratory (see Table S4 for sample ID, sequencing platform, and read depth).

We mapped reads to the *F. selysi* genome using Bowtie2 2.3.4.1 [45], called variants with Samtools 0.1.19 [46], and filtered variants with VCFtools 0.1.13 [47], excluding indels and retaining SNP variants with sequence depth >2 in all 11 Formica individuals. We used VCFtools to identify SNPs with fixed differences between the Sm and Sp haplotypes in the five focal species by calculating Weir and Cockerham's [62] F_{ST} between the six Sp and five Sm individuals, selecting the SNPs with F_{ST} equal to 1. To identify the overlapping or nearby genes for these SNPs, we extracted 10 kbp surrounding each SNP from the *F. selysi* reference genome using the getfasta command in Bedtools 2.27 [54], and queried these sequences against the *Camponotus floridanus*

reference genome and the NCBI nr database using blastn v2.7.1+. Finally, we extracted the genotypes of these fixed SNPs in the ten additional *Formica* species, to determine which regions of the supergene continue to exhibit an "old strata" pattern even with increased species sampling.

We used a Hidden Markov Model implemented in Saguaro [24] to identify regions of linkage group 3 with phylogenetic tree topologies matching the "old strata" expectation, and regions with topologies matching the species tree, in the five focal species.

Phylogeny and dating

To obtain aligned sequences in fasta format suitable for phylogenetic analyses, we ran the vcf2fq command in the vcfutils.pl module of Samtools 0.1.19 [46] on each bam file resulting from the previously described Bowtie2 alignment of whole-genome sequence data to the *F. selysi* reference genome. We extracted the chromosome 1 consensus sequence from each individual and concatenated these into a single aligned fasta file.

The phylogeny of the 18 species (15 ingroup species of *Formic*a and three outgroup species of *Polyergus* and *Iberoformica*) was reconstructed using the chromosome 1 sequence alignment. Phylogenetic reconstruction was performed using maximum likelihood (ML) criterion with IQ-TREE version 1.6.3 [55] and the model GTR+G+I. Ultrafast bootstrap analysis with 1000 replicates was conducted to assess node support in IQ-TREE version 1.6.3 [64].

To generate a small dataset for BEAST analysis, we first split the scaffold one sequence alignment into 10 kbp non-overlapping windows. After removing the windows that only contain uncalled bases or one taxon, 1532 windows were retained for further analyses. The ML tree and 100 rapid bootstrap replicates were then inferred for each window in RAxML version 8.2.8 [56] using the model GTR+G. The BEAST analysis was conducted on a dataset that contains the top 50 windows with the highest average bootstrap support and all 18 taxa.

Divergence times were estimated by Bayesian Markov Chain Monte Carlo (MCMC) analysis using the relaxed (uncorrelated lognormal) molecular clock model and GTR+G+I model in BEAST v2.4.5 [57] with the topology fixed to the ML tree from the

above IQ-TREE analysis. Using the known fossil records of *Formica* in Baltic ambers [65], we placed one calibration point at the MRCA of *Iberoformica* and *Formica* (lognormal distribution with offset = 42 Ma, median = 60 Ma, 95% quantile = 90 Ma; see [66]). The analysis was run for 60 000 000 generations (trees sampled at every 2000 generations). Tracer v1.7.1 [59] was used to check when the MCMCs had reached a stationary distribution by visual inspection of plotted posterior estimates. Trees sampled during the first 12 000 000 generations (20%) were removed as burn-in and the remaining trees (24 001 in total) were summarized in TreeAnnotator v2.5.2 [57] using the 'Maximum clade credibility tree' and 'Mean heights' options, and then displayed with age in millions of years using FigTree v1.4.3. The 95% highest probability density (95% HPD) values were summarized.

Quantification and Statistical Analysis

For the two species with observed mismatches between supergene genotype and social structure (*F. lemani* and *F. cinerea*), we tested the significance of association between the presence of an Sp haplotype (Sp/Sp homozygotes and Sm/Sp heterozygotes were both coded as "present") and polygynous social origin using Fisher's exact test implemented in R 3.3.1.

Data and code availability

The *F. selysi* genome assembly has been deposited to NCBI Genome (Bioproject PRJNA557079). PacBio sequence data has been deposited to NCBI SRA (Bioproject PRJNA559791). All new ddRAD and whole-genome sequence data has been deposited to NCBI SRA (Bioproject PRJNA557080). Previously published *F. selysi* sequence data for used in this study is available on NCBI SRA under Bioprojects PRJNA260443 (whole-genome) and PRJNA260459 (ddRAD). Linkage maps and a table of oligonucleotides used in ddRAD library preparation have been deposited to the Dryad data repository (DOI 10.6086/D1KD40).

References Cited

1. Schwander, T., Libbrecht, R., and Keller, L. (2014). Supergenes and complex phenotypes. Curr. Biol. *24*, R288-R294.

2. Charlesworth, D. (2016). The status of supergenes in the 21st century: Recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. Evol. Appl. *9*, 74-90.
3. Wellenreuther, M., and Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. Tr. Ecol. Evol. *33*, 427-440.
4. Purcell J., Brelsford A., Wurm Y., Perrin N., and Chapuisat M. (2014). Convergent genetic architecture underlies social organization in ants. Curr. Biol. *24*, 2728–2732.
5. Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E.D., Gilbert, M.T.P., and Zhang, G. (2014). Complex evolutionary trajectories of sex chromosomes across bird taxa. Science *346*, 1246338.
6. Hartmann, C., Landgraf, M., Bate, M., and Jäckle, H. (1997). *Krüppel* target gene *knockout* participates in the proper innervation of a specific set of *Drosophila* larval muscles. EMBO J. *16*, 5299-5309.
7. Yang, Q., Zhang, D., Li, Q., Cheng, Z., and Xue, Y. (2007). Heterochromatic and genetic features are consistent with recombination suppression of the self-incompatibility locus in *Antirrhinum.* Plant J. *51*, 140-151.
8. Branco, S., Carpentier, F., Rodríguez de la Vega, R.C., Badouin, H., Snirc, A., Le Prieur, S., Coelho, M.A., de Vienne, D.M., Hartmann, F.E., Begerow, D., et al. (2018). Multiple convergent supergene evolution events in mating-type chromosomes. Nat. Commun. *9*, 2000.
9. Iijima, T., Kajitani, R., Komata, S., Lin, C-P., Sota, T., Itoh, T., and Fujiwara, H. (2018). Parallel evolution of Batesian mimicry supergene in two *Papilio* butterflies, *P. polytes* and *P. memnon*. Sci. Adv. *18*, eaao5416.
10. Jay, P., Whibley, A., Frézal, L., Rodriguez de Cara, M.A., Nowell, R.W., Mallett, J., Dasmahapatra, K.K., and Joron, M. (2018). Supergene evolution triggered by introgression of a chromosomal inversion. Curr. Biol. *28*, 1839-1845.
11. Avril, A., Purcell, J., Brelsford, A., and Chapuisat, M. (2019). Asymmetric assortative mating and queen polyandry are linked to a supergene controlling ant social organization. Mol. Ecol. *28*, 1428-1438.
12. Rosset, H., and Chapuisat, M. (2007). Alternative life-histories in a socially polymorphic ant. Evol. Ecol. *21*, 577–588.
13. Sundström, L. (1993). Genetic population structure and sociogenetic organisation in *Formica truncorum* (Hymenoptera; Formicidae). Behav. Ecol. Sociobiol. *33*, 345-354.
14. Goropashnaya, A.V., Seppä, P., and Pamilo, P. (2001). Social and genetic characteristics of geographically isolated populations in the ant *Formica cinerea*. Mol. Ecol. *10*, 2807–2818.
15. DeHeer, C.J., and Herbers, J.M. (2004). Population genetics of the socially polymorphic ant *Formica podzolica.* Insect. Soc. *51*, 309-316.
16. Seppä, P., Gyllenstrand, N., Corander, J., and Pamilo, P. (2004). Coexistence of the social types: genetic population structure in the ant *Formica exsecta.* Evolution *58*, 2462-2471.
17. Gyllenstrand, N., Seppä, P., and Pamilo, P. (2005), Restricted gene flow between two social forms in the ant *Formica truncorum*. J. Evol. Biol. *18*, 978–984.

18. Bargum, K., Helanterä, H., and Sundström, L. (2007). Genetic population structure, queen supersedure and social polymorphism in a social Hymenoptera. J. Evol. Biol. *20*, 1351–1360.
19. Thompson, M. J., and Jiggins, C. D. (2014). Supergenes and their role in evolution. Heredity *113*, 1-8.
20. Lahn, B.T., and Page, D.C. (1999). Four evolutionary strata on the human X chromosome. Science *286*, 964-967.
21. Lawson-Handley, L-J., Ceplitis, H., and Ellegren, H. (2004). Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. Genetics *167*, 367-376.
22. Guerrero, R.F. Rousset, F., and Kirkpatrick, M. (2012). Coalescent patterns for chromosomal inversions in divergent populations. Philos. T. R. Soc. B *367*, 20110246.
23. Kapun, M., and Flatt, T. (2019). The adaptive significance of chromosomal inversion polymorphisms in *Drosophila melanogaster.* Mol. Ecol. *28*, 1263-1282.
24. Zamani, N., Russell, P., Lantz, H., Hoeppner, M.P., Meadows, J.R.S., Vijay, N., Mauceli, E., di Palma, F., Lindblad-Toh, K., Jern, P., et al. (2013). Unsupervised genome-wide recognition of local relationship patterns. BMC Genomics *14*, 347.
25. Kirkpatrick, M. (2010). How and why chromosome inversions evolve. PLOS Biol. *8*, e1000501.
26. Korunes, K.L., and Noor, M.A.F. (2017). Gene conversion and linkage: effects on genome evolution and speciation. Mol. Ecol. *26*, 351-364.
27. Korunes, K.L., and Noor, M.A.F. (2019). Pervasive gene conversion in chromosomal inversion heterozygotes. Mol. Ecol. *28*, 1302-1315.
28. Ross, K.G., and Shoemaker, D. (2018). Unexpected patterns of segregation distortion at a selfish supergene in the fire ant *Solenopsis invicta.* BMC Genet. *19*, 101.
29. da Silva, V.H., Laine, V.N., Bosse, M., Spurgin, L.G., Derks, M.F.L., van Oers, K., Dibbits, B., Slate, J., Crooijmans, R.P.M.A., Visser, M.E., et al. (2019). The genomic complexity of a large inversion in great tits. Genome Biol. Evol. *11*, 1870-1881.
30. Brelsford, A., Dufresnes, C., and Perrin, N. (2016). Trans-species variation in Dmrt1 is associated with sex determination in four European tree-frog species. Evolution *70*, 840-847.
31. Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., et al. (2012). A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). PLOS Genet. *8*, e1002798.
32. Zhang, W., Westerman, E., Nitzany, E., Palmer, S., and Kronforst, M.R. (2017). Tracing the origin and evolution of supergene mimicry in butterflies. Nat. Commun. *8*, 1269.
33. Küpper, C., Stocks, M., Risse, J.E., dos Remedios, N., Farrell, L.L., McRae, S.B., Morgan, T.C., Karlionova, N., Pinchuk, P., Verkuil, Y.I., et al. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. Nat. Genet. *48*, 79-83.

34. Lamichhaney, S. Fan, G., Widemo, F., Gunnarsson, U., Schwochow Thalmann, D., Hoeppner, M.P., Kerje, S., Gustafson, U., Shi, C., Zhang, H., et al. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). Nat. Genet. *48*, 84-88.
35. 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., et al. (2016). Divergence and functional degradation of a sex chromosome-like supergene. Curr. Biol. *26*, 344-350.
36. 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., et al. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature *477*, 203-206.
37. Van Belleghem, S.M., Rastas, P., Papanicolaou, A., Martin, S.H., Arias, C.F., Supple, M.A., Hanly, J.J., Mallet, J., Lewis, J.J., Hines, H.M., et al. (2017). Complex modular architecture around a simple toolkit of wing pattern genes. Nat. Ecol. Evol. *1*, 0052.
38. 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*, 664–668.
39. Braim, B.S. (2015). Exploring the regulatory role of behaviour and genome architecture in the socially polymorphic ant, *Leptothorax acervorum*. Unpublished doctoral thesis, University of Leicester. http://hdl.handle.net/2381/36076
40. 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. Mol. Ecol. *26*, 2864–2879.
41. Huang, Y-C., Dang, V.D., Chang, N-C., and Wang, J. (2018). Multiple large inversions and breakpoint rewiring of gene expression in the evolution of the fire ant social supergene. P. R. Soc. B *285*: 20180221.
42. Krieger, M.J.B., and Ross, K. (2002). Identification of a major gene regulating complex social behavior. Science *295*, 328-332.
43. Stolle, E., Pracana, R., Howard, P., Paris, C.I., Brown, S.J., Castillo-Carrillo, C., Rossiter, S.J., and Wurm, Y. (2019). Degenerative expansion of a young supergene. Mol. Biol. Evol. *36*, 553-561.
44. Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. Mol. Ecol. *22*, 3124-3140.
45. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357.
46. 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*, 2078-2079.
47. 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., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156-2158.
48. 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., et al. (2007) PLINK: a tool set

for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559-575.
49. Wu, Y., Bhat, P.R., Close, T.J., and Lonardi, S. (2008a). Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet. *4*, e1000212.
50. Wu, Y., Close, T.J., and Lonardi, S. (2008b). On the accurate construction of consensus genetic maps. Comput. Sys. Bioinform. *7*, 285-296.
51. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. *27*, 722–736.
52. Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: Interrogation of genome assemblies. F1000Research *6*, 1287.
53. Pryszcz, L.P., and Gabaldón, T. (2016). Redundans: An assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. *44*, 1-10.
54. Quinlan, A.R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinformatics. *47*, 11-12.
55. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268-274.
56. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.
57. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. *10*, e1003537.
58. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. *67*, 901-904.
59. Brelsford, A., Dufresnes, C., and 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*, 177-181.
60. 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 the ant *Formica selysi*. Evolution *58*, 1064-1072.
61. Purcell, J., and Chapuisat, M. (2013). Bidirectional shifts in colony queen number in a socially polymorphic ant population. Evolution *67*, 1169-1180.
62. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. Evolution *38*, 1358-1370.
63. Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., Langlade, N., and Muños, S. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. BioTechniques *61*, 203-205.
64. Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518-522.
65. Dlussky, G.M. (1997). Genera of ants (Hymenoptera: Formicidae) from Baltic amber. Paleontol. J. *31*, 616-627.

66. Blaimer, B.B., Brady, S.G., Schultz, T.R., Lloyd, M.W., Fisher, B.L., and Ward, P.S. (2015). Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. BMC Evol. Biol. *15*, 271.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Genomic DNA; *Formica exsecta* (n=108), *F. cinerea* (n=161), *F. selysi* (n=195), *F. truncorum* (n=20), *F. lemani* (n=65), *F. tombeuri* (n=1), *F. fusca* (n=1), *F. fuscocinerea* (n=1), *F. lugubris* (n=1), *F. sanguinea* (n=1), *F. rufibarbis* (n=1), *F. picea* (n=1), *F. pressilabris* (n=1), *F. fennica* (n=1), *F. pratensis* (n=1), *Polyergus mexicanus* (n=1), P. vinosus (n=1).	This paper	SRA PRJNA557080
Genomic DNA; *Iberoformica subrufa* (n=1)	M. Borowiec, U. Idaho	SRA SAMN13065562
Chemicals, Peptides, and Recombinant Proteins		
EcoRI-HF	New England Biolabs	R3101
SbfI-HF	New England Biolabs	R3642
MseI	New England Biolabs	R0525
T4 DNA Ligase	New England Biolabs	M0202
Q5 Hot Start polymerase	New England Biolabs	M0493
ATP 100 mM	Thermo Fisher	R0441
dNTP mix	Thermo Fisher	R0192
Sodium Chloride Biotechnology Grade	VWR	97061
1M TRIS, pH 8.0 biotechnology grade	VWR	E199
0.5M EDTA, sterile solution biotechnology grade	VWR	BDH7830
Sodium Dodecyl Sulfate (SDS), 20% Solution Biotechnology Grade	VWR	97062
Polyvinylpyrrolidone average mol wt 40,000	Sigma	PVP40
Sodium metabisulfite ReagentPlus®, ≥99%	Sigma	S9000
Potassium acetate for molecular biology, ≥99.0%	Sigma	P1190
Polyethylene Glycol 8000 (PEG)	Fisher Scientific	BP233
Sera-Mag SpeedBead magnetic carboxylate modified particles, DSMG-CM, 1 um, 5% solids	Fisher Scientific	09-981
RNase A 100 mg/ml	Qiagen	19101
Ethanol absolute AnalaR NORMAPUR® ACS, Reag. Ph. Eur. analytical reagent	VWR	10107
Critical Commercial Assays		
SMRTbell Template Prep Kit 1.0	PacBio	100-259-100
DNeasy Blood and Tissue extraction kit	Qiagen	69506
Agencourt AMPure XP	Beckman Coulter	A63882
TruSeq Nano library preparation kit	Illumina	FC-121-4001

Deposited Data		
Formica selysi genome assembly	This study	PRJNA557079
Linkage mapping RADseq data, *F. selysi* (n=112) and *F. exsecta* (n=67).	This study	PRJNA557080
Population RADseq data, *F. selysi*	[4]	PRJNA260459
Whole-genome sequence data, *F. selysi* Sm haplotype	[4]	SRX695613
Whole-genome sequence data, *Formica exsecta* (n=2), *F. cinerea* (n=2), *F. selysi* (n=1), *F. truncorum* (n=2), *F. lemani* (n=3), *F. tombeuri* (n=1), *F. fusca* (n=1), *F. fuscocinerea* (n=1), *F. lugubris* (n=1), *F. sanguinea* (n=1), *F. rufibarbis* (n=1), *F. picea* (n=1), *F. pressilabris* (n=1), *F. fennica* (n=1), *F. pratensis* (n=1).	This study	PRJNA557080
Oligonucleotides		
ddRAD barcoded adapters and primers	[4, 59]	Dryad DOI 10.6086/D1KD40
Software and Algorithms		
Stacks 1.19	[44]	N/A
Bowtie 2.3.4.1	[45]	N/A
Samtools 0.1.19	[46]	N/A
VCFtools 0.1.13	[47]	N/A
Plink 1.90	[48]	N/A
MSTMap	[49]	N/A
MergeMap	[50]	N/A
Canu 1.7	[51]	N/A
BlobTools 1.0	[52]	N/A
Pbalign 0.3.0	https://github.com/PacificBiosciences/pbalign	N/A
Samtools 1.4	[45]	N/A
GenomicConsensus 2.2.2	https://github.com/PacificBiosciences/GenomicConsensus	N/A
Redundans0.13c	[53]	N/A
Bedtools 2.27	[54]	N/A
Saguaro 0.1	[22]	N/A
IQ-TREE 1.6.3	[55]	N/A
RAxML 8.2.8	[56]	N/A
BEAST 2.4.5	[57]	N/A
Tracer 1.7.1	[58]	N/A
TreeAnnotator 2.5.2	[57]	N/A
FigTree 1.4.3	http://tree.bio.ed.ac.uk/software/figtree/	N/A
R 3.3.1	http://www.R-project.org	N/A

Figure 1

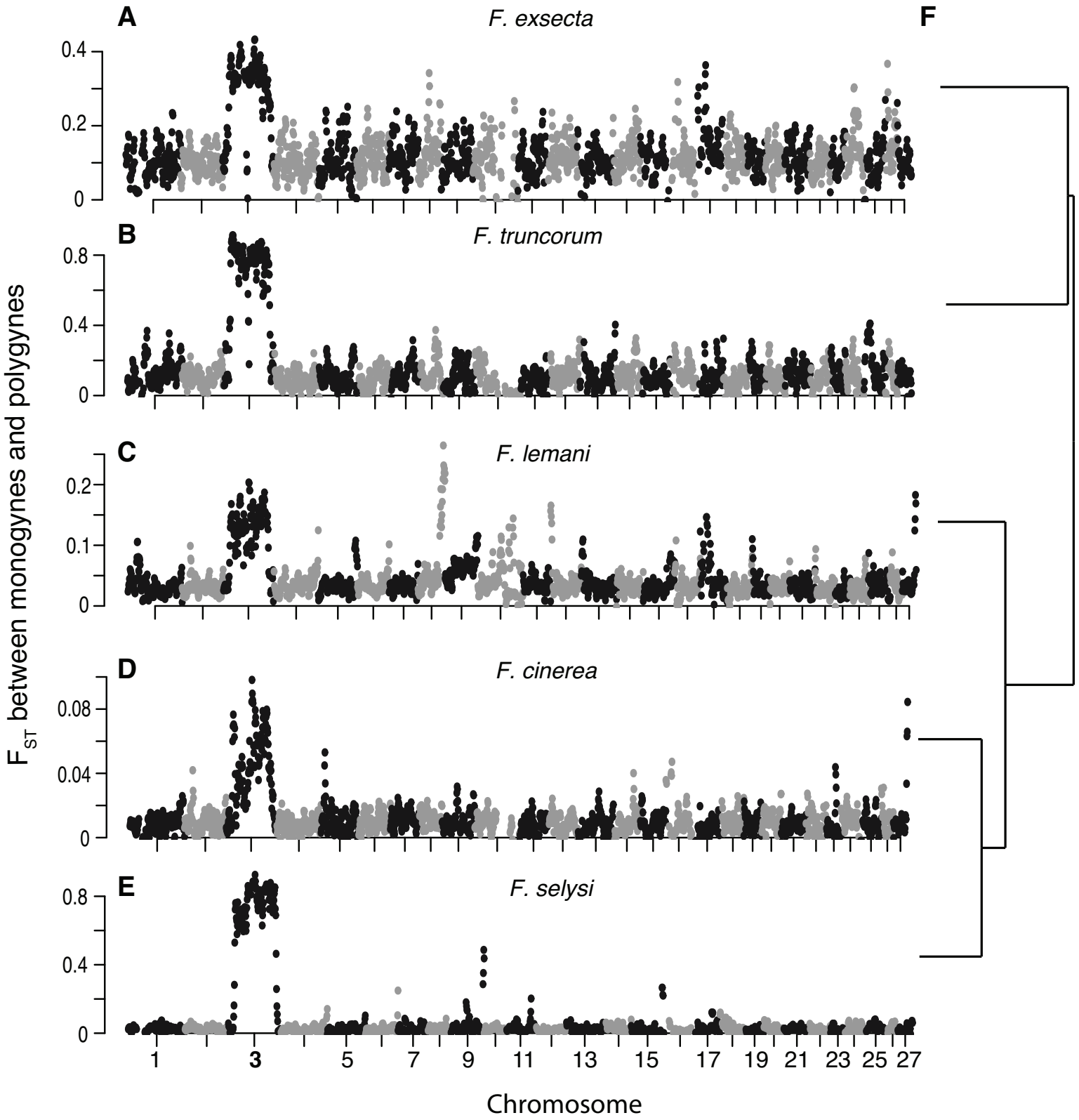

A
F. exsecta
F
0.4
0.2
0
B
F. truncorum
0.8
0.4
0
C
F. lemani
0.2
0.1
0
D
F. cinerea
0.08
0.04
0
E
F. selysi
0.8
0.4
0
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Chromosome
F_ST between monogynes and polygynes

Figure 2

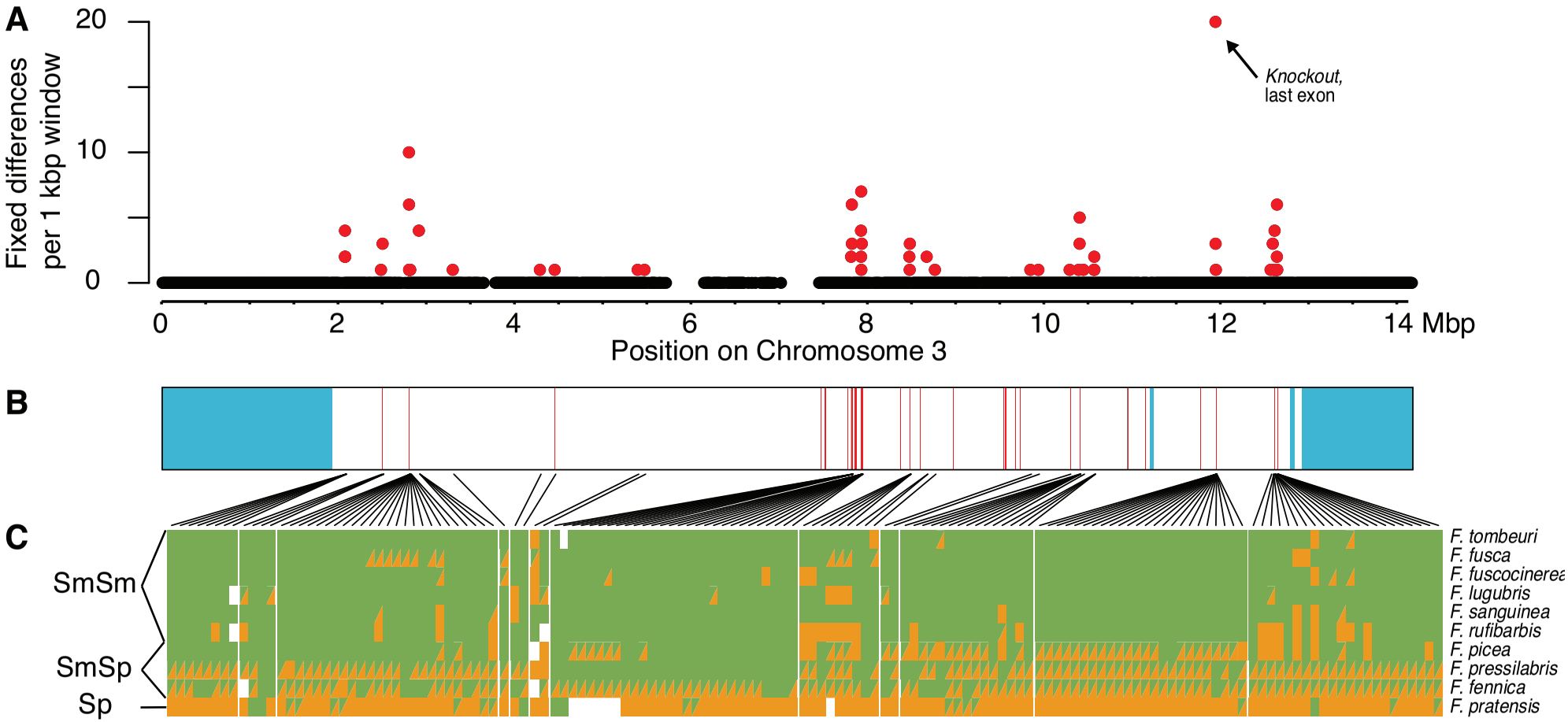

A
Fixed differences per 1 kbp window
20
10
0
Knockout, last exon
0
2
4
6
8
10
12
14 Mbp
Position on Chromosome 3
B
C
SmSm
SmSp
Sp
F. tombeuri
F. fusca
F. fuscocinerea
F. lugubris
F. sanguinea
F. rufibarbis
F. picea
F. pressilabris
F. fennica
F. pratensis

Figure 3

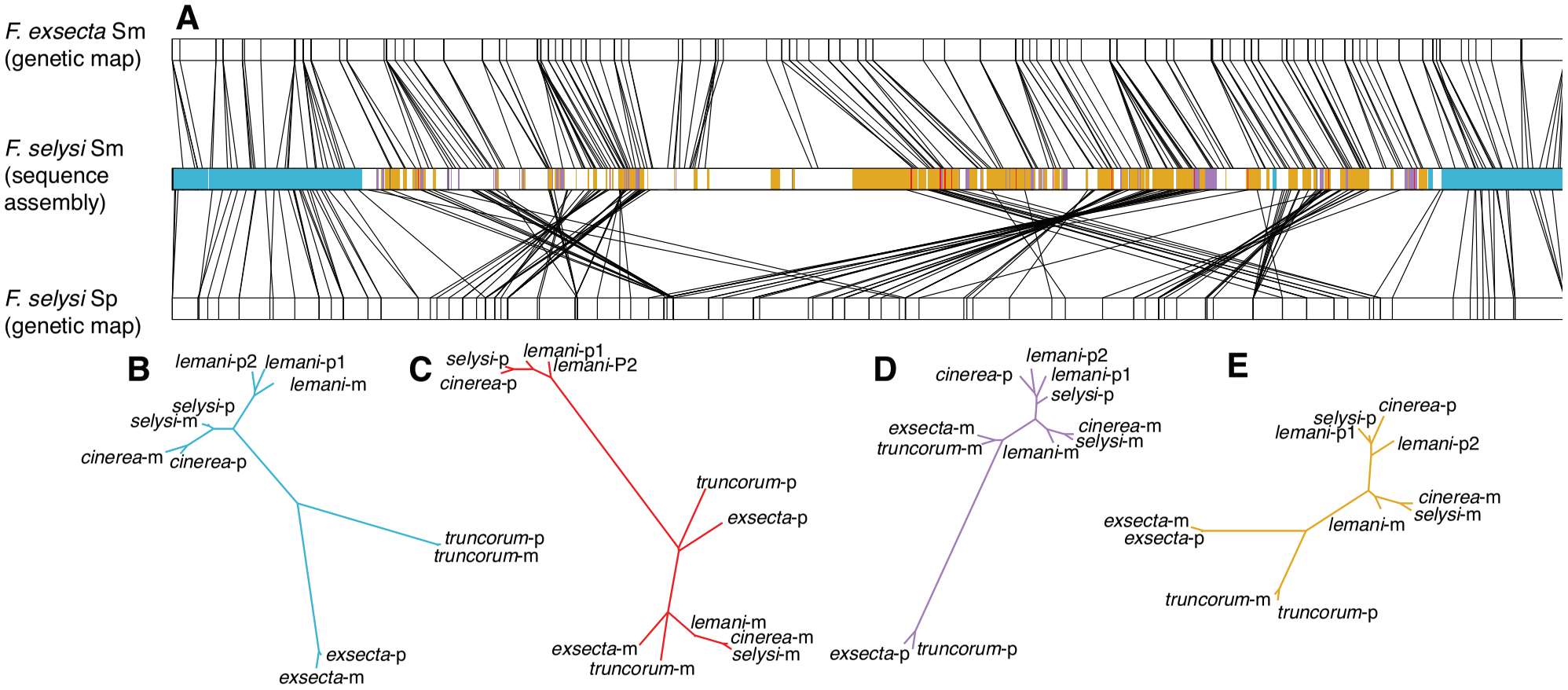
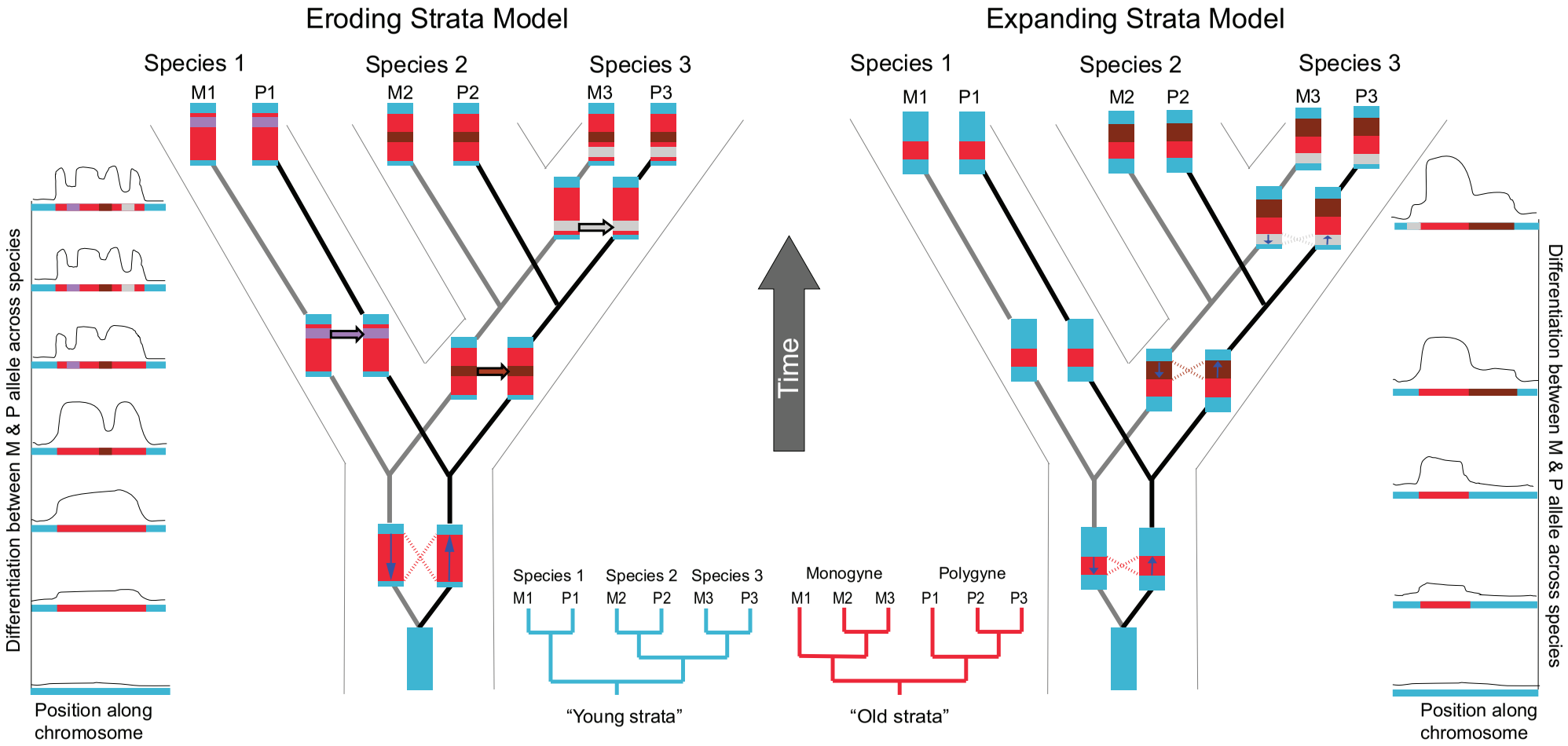

A
F. exsecta Sm (genetic map)
F. selysi Sm (sequence assembly)
F. selysi Sp (genetic map)
B
lemani-p2
lemani-p1
lemani-m
selysi-p
selysi-m
cinerea-m
cinerea-p
truncorum-p
truncorum-m
exsecta-p
exsecta-m
C
selysi-p
lemani-p1
lemani-P2
cinerea-p
truncorum-p
exsecta-p
lemani-m
cinerea-m
selysi-m
exsecta-m
truncorum-m
D
lemani-p2
cinerea-p
lemani-p1
selysi-p
exsecta-m
truncorum-m
cinerea-m
selysi-m
lemani-m
exsecta-p
truncorum-p
E
cinerea-p
selysi-p
lemani-p1
lemani-p2
cinerea-m
selysi-m
exsecta-m
exsecta-p
lemani-m
truncorum-m
truncorum-p

Figure 4


Eroding Strata Model
Species 1
Species 2
Species 3
M1
P1
M2
P2
M3
P3
Differentiation between M & P allele across species
Position along chromosome
Time
Species 1
Species 2
Species 3
M1
P1
M2
P2
M3
P3
"Young strata"
Monogyne
Polygyne
M1
M2
M3
P1
P2
P3
"Old strata"
Expanding Strata Model
Species 1
Species 2
Species 3
M1
P1
M2
P2
M3
P3
Differentiation between M & P allele across species
Position along chromosome

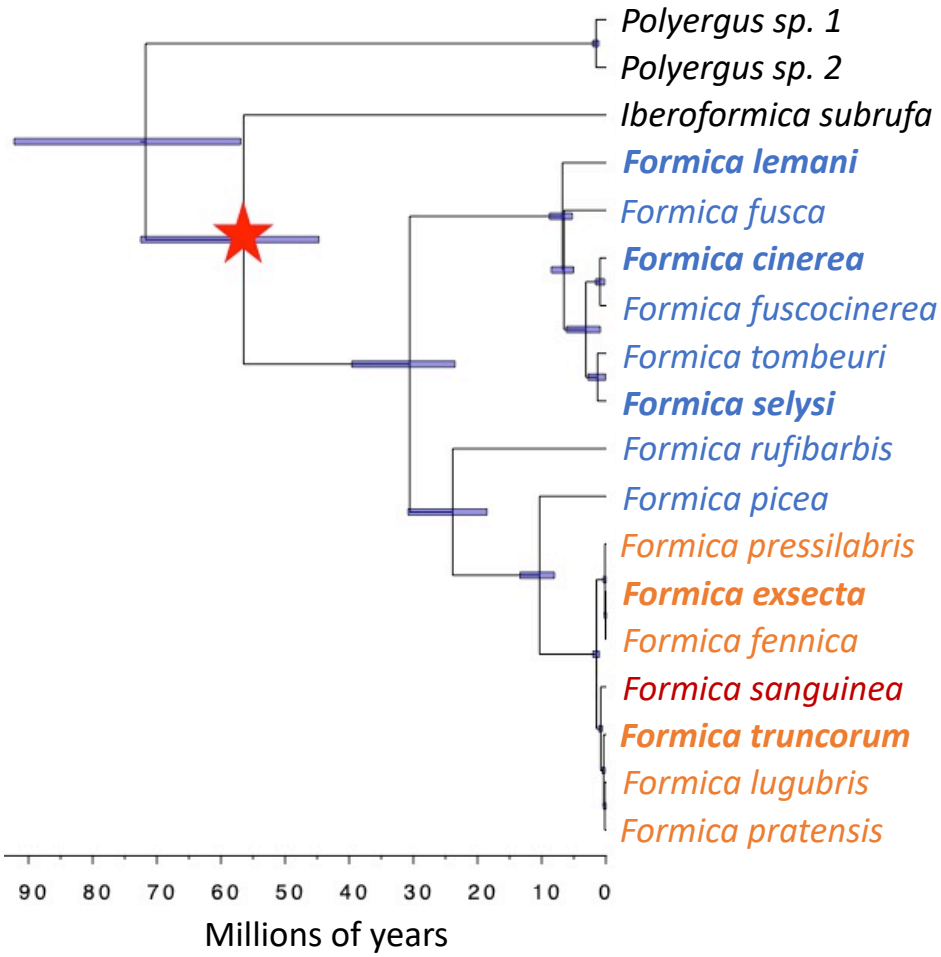


Figure S1. Time-calibrated Phylogeny of 15 Formica Species, Related to Figures 1, 2, 3. The *Formica* species examined here span an estimated 30 million years of evolutionary history. This phylogeny, implemented in BEAST (see STAR Methods), shows the 15 species investigated here as well as three outgroups. The calibration point is shown as the red star, and 95% highest posterity density (HPD) intervals are indicated with blue bars. Species examined in Figures 1, 2A and B, and S2 are shown in bold. The remaining 10 species, examined in Figure 2C, are also shown. Species that do not exhibit socially parasitic behaviors (sometimes called *Serviformica*) are shown in blue, temporary social parasites in orange (including both *Formica* sensu stricto and *Coptoformica*), and facultative slave-making species in red (*Raptiformica*). Outgroups are shown in black.

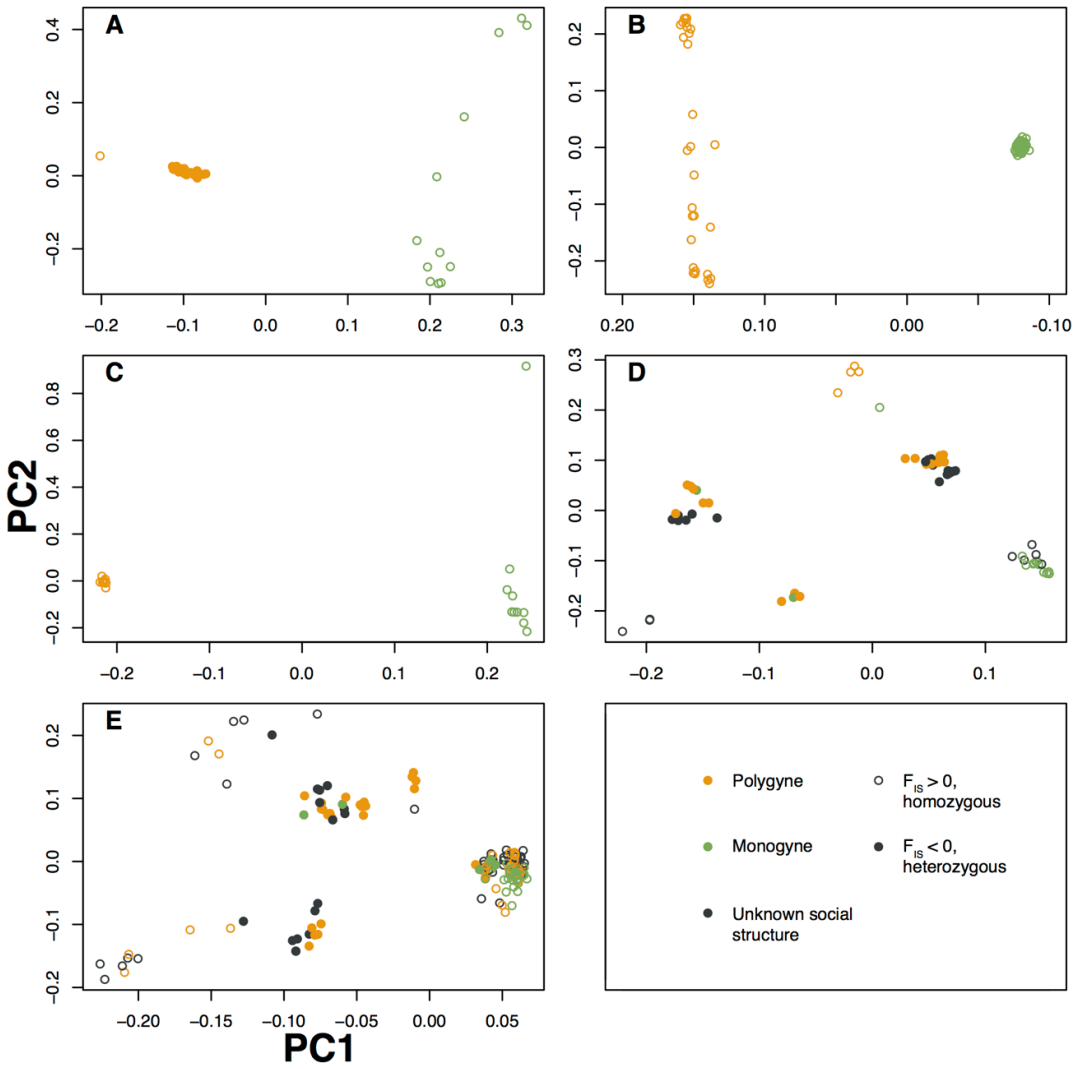


Figure S2. Principal Component Analyses of SNPs from Population ddRAD Data on Chromosome 3, Related to Figure 1. PCAs show two to six clusters in each species, with cluster membership strongly associated with colony social organization. Each panel represents an independent PCA in one species: *F. exsecta* (A), *F. selysi* (B), *F. truncorum* (C), *F. lemani* (D), and *F. cinerea* (E). Each dot corresponds to an individual worker or male; supergene heterozygotes tend to have excess heterozygosity resulting in a strongly negative F_{IS} value. In *F. lemani* and *F. cinerea*, we found evidence for a third supergene haplotype. Two of the three alternative supergene haplotypes in both species were much more common in individuals of polygyne origin compared to those of monogyne origin; we therefore infer that these systems contain two alternative Sp haplotypes and one Sm haplotype. In both systems, we also find mismatches between supergene genotype and colony social structure.

Species	Origin	RADseq (population)	RADseq (linkage map)	WGS
Formica selysi	Switzerland	83 males (monogyne and polygyne)	112 workers (polygyne)	2 males (Sm, Sp)
F. cinerea	Finland, Italy, Switzerland	161 workers (monogyne, polygyne, and unknown)	NA	1 male (from Switzerland: Sm) 1 worker (from Italy: Sp/Sp)
F. lemani	Finland, Spain, Switzerland	65 workers (monogyne, polygyne, and unknown)	NA	3 workers (from Switzerland: Sm/Sm, Sp_1Sp_1, Sp_2Sp_2)
F. exsecta	Finland	12 males (monogyne) and 29 workers (polygyne)	67 males (monogyne)	1 male (Sm), 1 worker (Sp/Sp)
F. truncorum	Finland	5 males and 5 workers (monogyne) and 10 males (polygyne)	NA	2 males (Sm and Sp)
F. tombeuri	Spain	NA	NA	1 worker (inferred Sm/Sm)
F. fusca	Switzerland	NA	NA	1 worker (inferred Sm/Sm)
F. fuscocinerea	Switzerland	NA	NA	1 worker (inferred Sm/Sm)
F. lugubris	Switzerland	NA	NA	1 worker (inferred Sm/Sm)
F. sanguinea	Germany	NA	NA	1 worker (inferred Sm/Sm)
F. rufibarbis	Portugal	NA	NA	1 worker (inferred Sm/Sm)
F. picea	Finland	NA	NA	1 worker (polygyne, inferred Sm/Sp)
F. pressilabris	Finland	NA	NA	1 worker (polygyne, inferred Sm/Sp)
F. fennica	Finland	NA	NA	1 worker (polygyne, inferred Sm/Sp)
F. pratensis	Finland	NA	NA	1 male (polygyne, Sp)

Table S1. List of Species Used in Analyses, Country of Origin, and Samples Used for ddRAD Sequencing (RADseq) and Whole Genome Sequencing (WGS), Related to Figures 1 and 2 and STAR Methods

MSTMap Parameter	*F. selysi* Sp/Sp queen, colony 191	*F. selysi* Sp/Sp queen, colony 192	*F. exsecta* sM brothers, colony FE63
Distance_function	Kosambi	Kosambi	Kosambi
Cut_off_p_value	0.00005	0.000005	0.000005
No_map_dist	30	30	30
No_map_size	1	1	1
Missing_threshold	0.1	0.1	0.1
Estimation_before_clustering	No	No	No
Detect_bad_data	Yes	Yes	Yes
Objective_function	ML	ML	ML
Number_of_loci	1792*	3688*	4603*
Number_of_individuals	35	77	63
Output			
Total number of linkage groups (incl. unplaced loci)	29*	48*	31*
Number of unplaced loci	3 in 2 LGs	15 in 9 LGs	5 in 5 LGs

* Number of loci and number of linkage groups shown here are the true input and output numbers; in order to account for the unknown allele phase in each queen, we duplicate each locus in the input file, recoding each allele as 'A' or 'B'. This results in duplicated linkage groups, which are then manually compared and removed.

Table S2. Parameters and Results for Linkage Map Construction, Related to STAR Methods

Output Statistic	**Value**
Pacbio sequence depth	100x
Assembly Length	290 Mbp
Contig N50	5.7 Mbp
Scaffold N50	7.9 Mbp
Number of Scaffolds Assigned to Chromosomes	27
Length of Scaffolds Assigned to Chromosomes	227 Mbp
Number of Scaffolds Not Assigned to Chromosomes	471
Length of Scaffolds Not Assigned to Chromosomes	63 Mbp

Table S3. Genome Assembly Results, Related to STAR Methods.

Species	Sample ID	Origin	Sex	Supergene genotype	Sequencer	read length	depth
F. selysi	F92M2	Switzerland	M	Sm	HiSeq 2000, Lausanne	100bp PE	15.5
F. selysi	079M2	Switzerland	M	Sp	HiSeq 2500, Lausanne	100bp PE	19.6
F. cinerea	FcBra10	Switzerland	M	Sm	HiSeq 2500, Lausanne	100bp PE	12.5
F. cinerea	FcQuin3	Italy	F	Sp/Sp	HiSeq 2500, Lausanne	100bp PE	11.8
F. exsecta	FE-MM5	Finland	M	Sm	HiSeq 2500, Lausanne	100bp PE	10.9
F. exsecta	FE-PW10	Finland	F	Sp/Sp	HiSeq 2500, Lausanne	100bp PE	8.8
F. lemani	FL-BG13A	Switzerland	F	Sm/Sm	HiSeq 2500, Lausanne	100bp PE	9.2
F. lemani	FL-BG25W1	Switzerland	F	Sp/Sp	HiSeq 2500, Lausanne	100bp PE	10.5
F. lemani	FL-BG9W1	Switzerland	F	Sp/Sp	HiSeq 2500, Lausanne	100bp PE	9.9
F. truncorum	FT-MM1	Finland	M	Sm	HiSeq 2500, Lausanne	100bp PE	9.6
F. truncorum	FT-PM6	Finland	M	Sp	HiSeq 2500, Lausanne	100bp PE	10.4
F. picea	Fpic1-2	Finland	F	Sm/Sp	HiSeq 2500, Lausanne	100bp PE	13.0
F. rufibarbis	PortoA	Portugal	F	Sm/Sm	HiSeq 2500, Lausanne	100bp PE	10.2
F. tombeuri	ainc1w9	Spain	F	Sm/Sm	HiSeq 4000, Berkeley	150bp PE	9.3
F. fusca	bg22w1	Switzerland	F	Sm/Sm	HiSeq 4000, Berkeley	150bp PE	7.5
F. fuscocinerea	furka4w1	Switzerland	F	Sm/Sm	HiSeq 4000, Berkeley	150bp PE	9.6
F. lugubris	lugc1w1	Switzerland	F	Sm/Sm	HiSeq 4000, Berkeley	150bp PE	6.9
F. sanguinea	fsanw1	Germany	F	Sm/Sm	HiSeq 4000, Berkeley	150bp PE	6.8
F. fennica	ob6pol4w1	Finland	F	Sm/Sp	HiSeq 4000, Berkeley	150bp PE	6.7
F. pressilabris	br6pol4w1	Finland	F	Sm/Sp	HiSeq 4000, Berkeley	150bp PE	7.9
F. pratensis	fp43m1	Finland	M	Sp	HiSeq 4000, Berkeley	150bp PE	9.1
Iberoformica subrufa	D1135	Spain	F	n/a	HiSeq 4000, Berkeley	150bp PE	4.9
Polyergus vinosus	scrc2w15	USA	F	n/a	HiSeq 4000, Berkeley	150bp PE	3.8
Polyergus mexicanus	slac1m1	USA	M	n/a	HiSeq 4000, Berkeley	150bp PE	6.1

Table S4. Details of Individual Samples Used for Whole-Genome Sequencing, Related to STAR Methods