Supplemental Information

An Ancient and Eroded Social Supergene

Is Widespread across *Formica* Ants

Alan Brelsford, Jessica Purcell, Amaury Avril, Patrick Tran Van, Junxia Zhang, Timothée Brütsch, Liselotte Sundström, Heikki Helanterä, and Michel Chapuisat
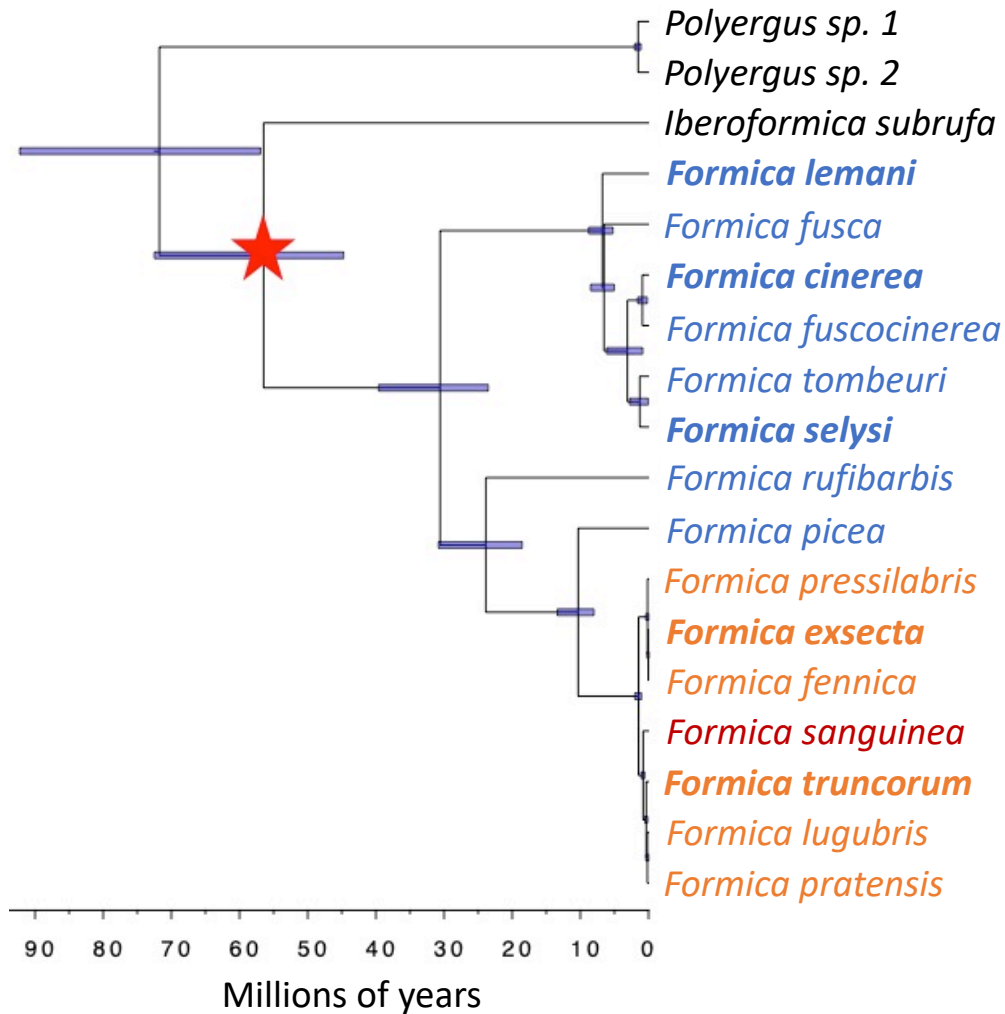
Figure S1. Time-calibrated Phylogeny of 15 Formica Species, Related to Figures 1, 2, 3. The *Formica* species examined here span an estimated 30 million years of evolutionary history. This phylogeny, implemented in BEAST (see STAR Methods), shows the 15 species investigated here as well as three outgroups. 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.
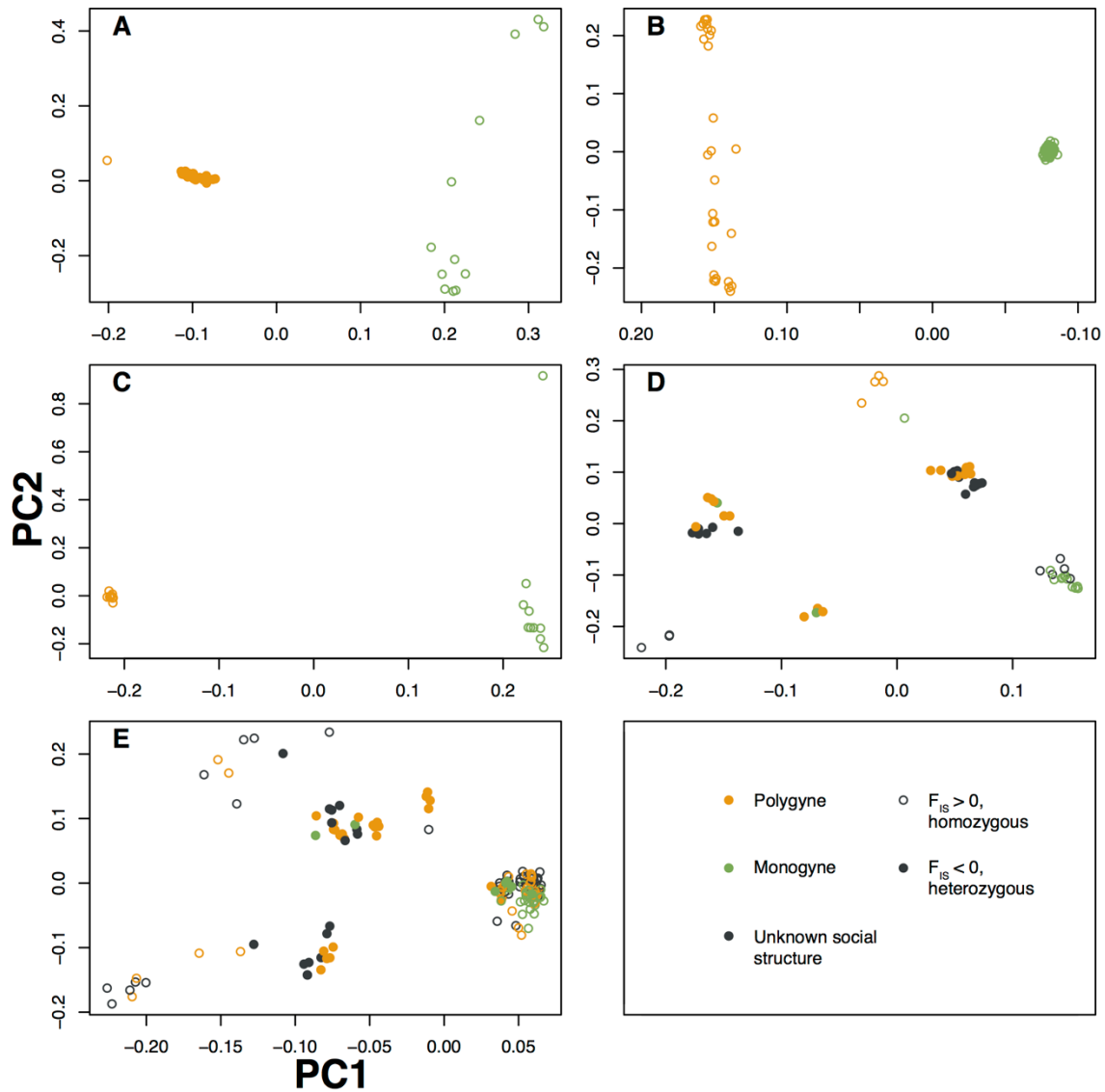
Figure S2. Principal Component Analyses of SNPs from Population ddRAD Data on Chromosome 3, Related to Figure 1. PCAs show two to six clusters in each species, with cluster membership strongly associated with colony social organization. Each panel represents an independent PCA in one species: *F. exsecta* (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