

Current Biology, Volume 24

Supplemental Information

Convergent Genetic Architecture

Underlies Social Organization in Ants

Jessica Purcell, Alan Brelsford, Yannick Wurm, Nicolas Perrin, and Michel Chapuisat

Supplementary Data

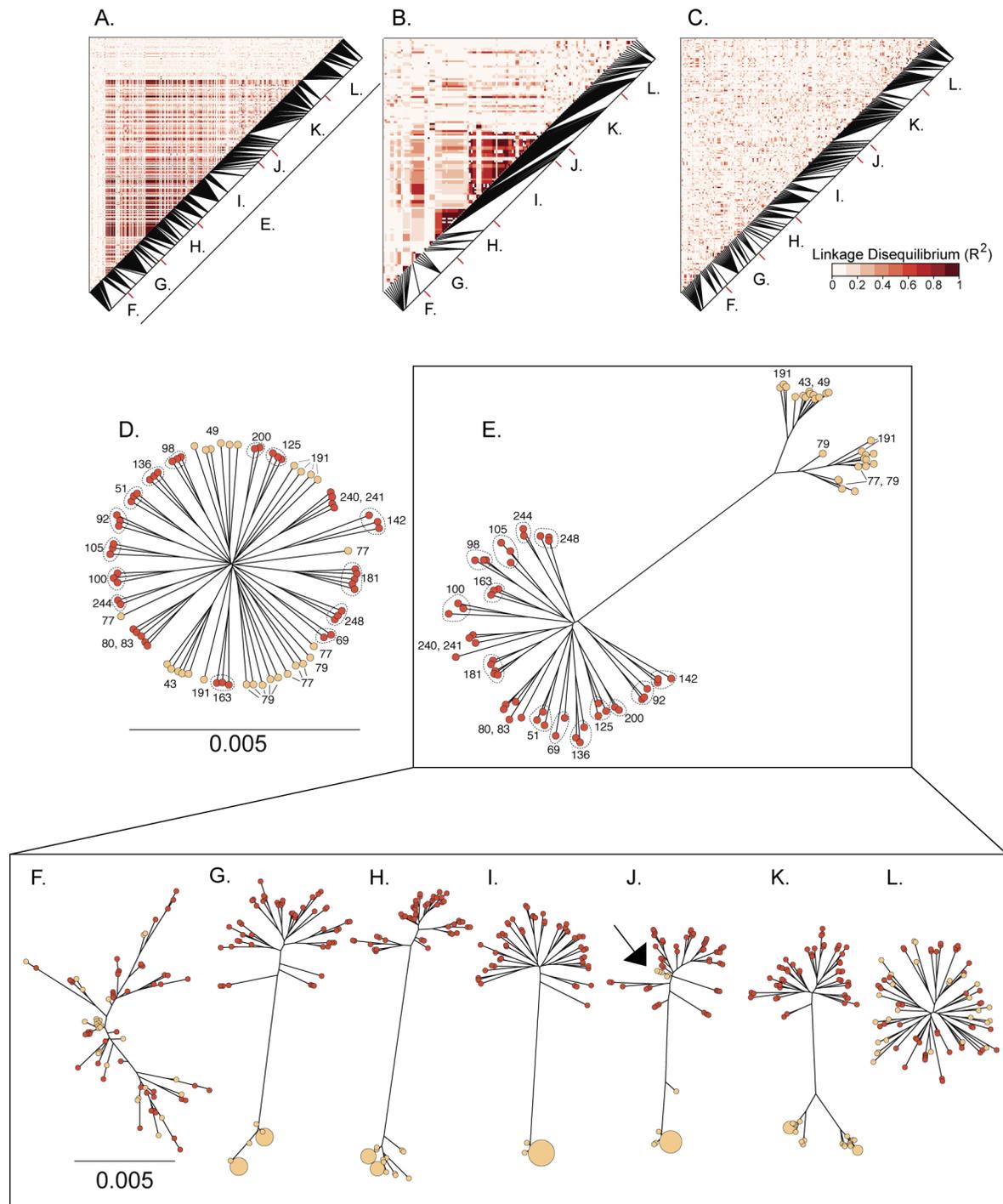


Fig. S1. Heat maps showing the linkage disequilibrium along LG3 for 3 individuals from each of 5 monogynous and 5 polygynous colonies (A), 3 individuals from each of 5 polygynous colonies (B) and 3 individuals from each of 5 monogynous colonies (C). Unrooted neighbor-joining trees based on SNPs located throughout the genome excluding the social chromosome (D) and on the social chromosome (E) in 79 males. For most of the genome,

variation is structured by colony but not by social form, whereas markers in the social chromosome show a strong separation between monogyne (red) and polygyne (yellow) males as well as low variability in polygyne. Numbers represent colonies of origin, and dotted lines encircle colony members when they form a single clade in both trees. Unrooted neighbor-joining trees constructed from 422 SNPs in seven regions of the social chromosome (a subset of the tree shown in E) show evidence of recombination between Sm and Sp haplotypes on each side of the region of suppressed recombination (F, L) as well as evidence of one recombination event in the center of the region (J, marked with black arrow). This finding suggests that there are at least two Sp haplotypes segregating in the population. The dendrograms for the remaining regions of the linkage group show evidence of divergence and suppressed recombination between Sm and Sp (G-I, K), and essentially no Sp polymorphism in a large, central portion of the non-recombining region (I). The size of colored circles in trees F-L is proportional to the number of males sharing an identical haplotype for the corresponding chromosomal region. The scale bars show per-SNP nucleotide divergence. These data relate to fig. 2.

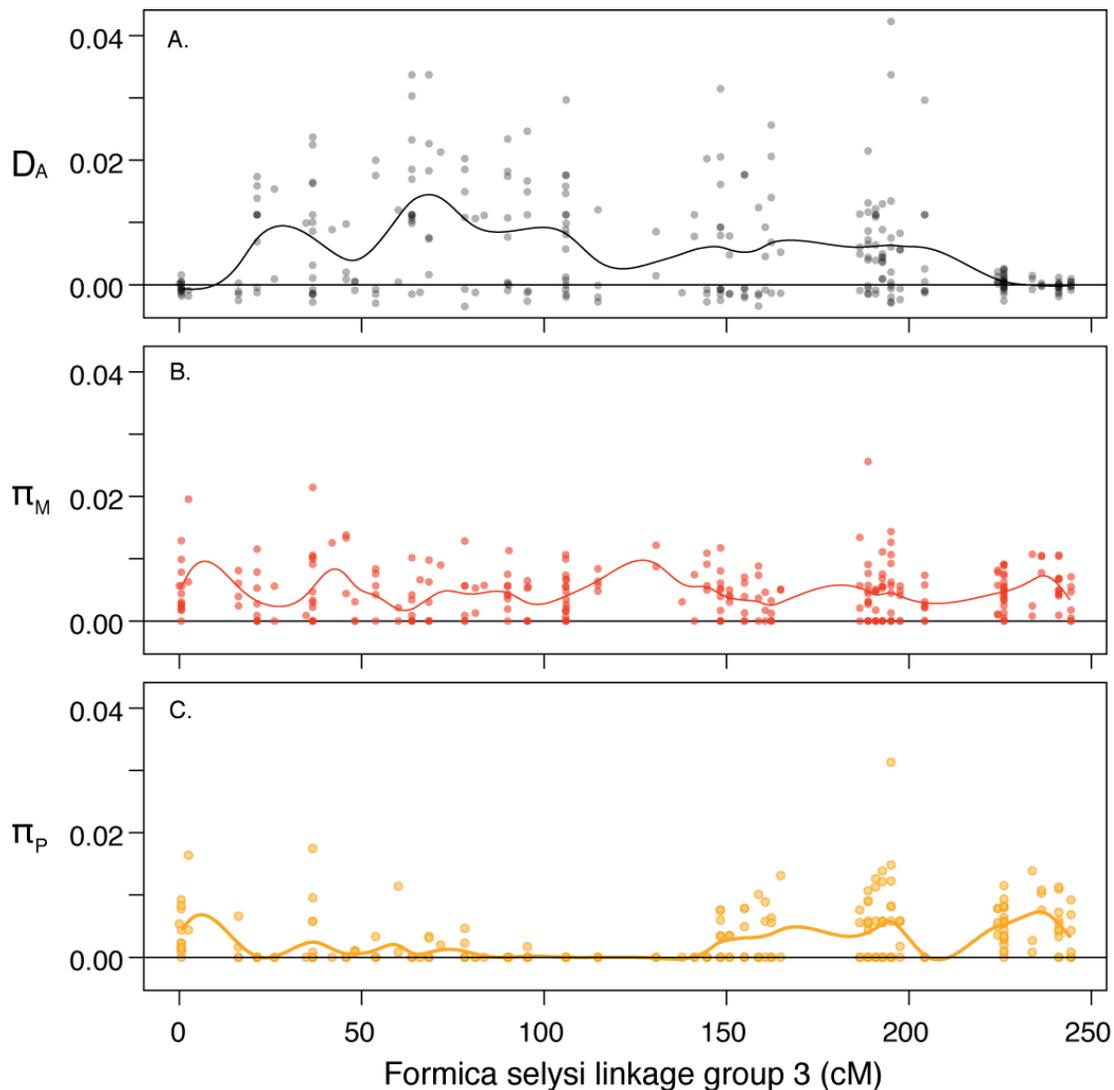


Fig. S2. Net sequence divergence (D_A) between males from monogynous and polygynous colonies is close to zero at the ends of the social chromosome, but relatively high across the central region (A). Polymorphism in monogynes (π_M) is consistent across the length of the chromosome (B), while polymorphism in polygyne (π_P) is much lower between cM 20 and 145 (C). These data support fig. 2.

Table S1. We observed no difference in the head size of Sm/Sp and Sp/Sp workers or mature queens, measured as the maximum width across the eyes following Schwander et al [S1]. In contrast, Sm/Sm workers were larger than individuals with a Sp haplotype, and a previous study showed the same pattern for monogynous compared to polygynous queens [S2]. Oligogynous colonies have a small number of closely related reproductive queens, confirmed through parentage analysis and analysis of worker relatedness with microsatellite loci. These colonies had previously been identified based on small effective queen number and monogyne-like traits of colony members [S1]. Our measurements of worker body size are consistent with these previous data: we found no difference between workers from monogynous and oligogynous colonies. Oligogynous colonies also resemble monogynous ones in their allocation to sexual production, and they most likely arise through adoption of a daughter queen while the mother queen is still present in the colony (Purcell & Chapuisat, unpublished data). These data relate to table 1.

Colony structure & individual genotype	Mean head size (mm)	95% Confidence Intervals	Statistical comparison	Statistical comparison
Workers				
Monogynous Sm/Sm	1.268	± 0.037	F _{1,13} = 0.095, p = 0.76	F _{3,292} = 15, p < 0.0001
Oligogynous Sm/Sm	1.270	± 0.032		
Polygynous Sm/Sp	1.208	± 0.017	F _{1,199} = 2.5, p = 0.12	
Polygynous Sp/Sp	1.196	± 0.028		
Mature Queens				
Polygynous Sm/Sp	1.925	± 0.015	F _{1,19} = 0.84, p = 0.37	
Polygynous Sp/Sp	1.916	± 0.020		

Table S2. Results of mixed models identifying genetic markers associated with social structure (in separate excel file). At the less stringent significance cut-off shown in Fig. 1 (per-SNP $\alpha = 0.01$, genome-wide FDR 0.167), 907 markers were significantly associated with social organization. These data support fig. 1.

Table S3. PCR primers for microsatellite and SNP markers on the *F. seelysi* social chromosome. PCR recipes for microsatellites followed Chapuisat et al. [S3], and SNP PCR recipes followed Brelford & Irwin [S26]. Microsatellite thermal cycling programs included 35 cycles, and SNP programs 30 cycles. All reactions were run at 10 μ l volume using an annealing temperature of 58°C and extension time of 30s. For SNP markers, the product of the first PCR reaction was diluted 10x and used as a template for the second PCR with two internal primers. Three μ l of PCR product was then digested with 0.7 units of the corresponding restriction enzyme in the appropriate buffer (total volume 9 μ l) and visualized on agarose gel. This information supports table 1.

Marker ID	Position on Sm/Sm linkage map (cM)	Forward Primer	Reverse Primer			
<i>microsatellite</i>				<u>Multiplex</u>	<u>Adjusted Primer Volume</u>	
Fs3-01	0.64	Atto- CTCCCGAAGAT CCCTTTGT	AGAAATCGCG CATCTGAAAT	b	0.25	
Fs3-02	36.6	Hex- GTCCTGCGCTT CTCTCATTC	CAACGGCATCA AACATAACG	a	0.27	
Fs3-03	63.8	Fam- CGTGTGCGCAC GATAATACT	ACGTAGACTTC CGCTTCTCG	a	0.23	
Fs3-04	95.4	Hex- CCCCCTCCACC ATTTTCTAT	TCCATTTATCG CCCTTCTTG	a	0.23	
Fs3-05	144.6	Hex- GCTTACTGCGC CATTATCA	CCGCCATTCCA TCTATCAAT	b	0.3	
Fs3-06	150.8	Fam- CCACGTCACGG TATGATTTG	GCATACTGTCG AGACGACGA	b	0.3	
Fs3-07	164.8	Fam- ACATCGTCGTG TCACCGTCT	CCGGGAGAAC AAGAGAGATG	a	0.23	
Fs3-08	188.8	Atto- ATTGCCGATGC GAAATAGTC	TCAACCGGAA GTGTCTTTCC	a	0.32	
Fs3-09	195.1	CGTCTGTTCTA GAGAGATTCAG CA	Atto- GCCAAAGCCTT GACTCAAAA	b	0.25	
<i>SNP</i>		<u>PCR 1: Forward Primer</u>	<u>PCR 1: Reverse Primer</u>	<u>PCR 2: Forward Primer</u>	<u>PCR 2: Reverse Primer</u>	<u>Enzyme</u>
fsGBS-40656	241	GCCATCGCTCT CCTTAGTTG	AATGGTTCGAA ATTCGGATG	CGTGGCTGGTT ACATACGTG	TTCGGATGCA CATTATTTCG	HinPII
fsGBS-13220	224.3	GACTTCGACGG CAAAATGAT	TCTCCCATATC TTCGCGATT	CAATGGTCCGG CTTGAAATA	AGTCCGCGA TGATCTTGTT	DpnII
fsGBS-12943	68.4	ATCCCCTATC CCGCTATAC	CTGCCAATTTT TCGGAAGAG	ACGCTATTCCC CCTCGATAG	GGTGAGGAGA CTTCGTACCG	MspI
FsGBS-8125	~48	TGGGTTTCAAG TTGGAGGAG	ATTGTCAGATG CGATGAACG	AGAAGTACCG CGACAGAGGA	TGCGATGAAC GCAAGTATCT	MspI
FsGBS-8924	~106	AAAGATCGACA GATGGCTTGA	CTTCCATCGCG ATCAATTTT	CGACAGATGG CTTGAGAAAG	AATTGATGGT CCCCACTGAA	MspI

Supplementary experimental procedures

I. Study population and sampling

Formica selysi is abundant in flood plains in the Alps and the Pyrenees of Europe. Our study site is located along the river Rhône between Sierre and Susten, Switzerland (7°36'30"E, 46°18'30"N, altitude 565 m). The population is socially polymorphic, and colonies with one queen (= monogynous), two closely related queens (= oligogynous) and multiple queens (= polygynous) have been monitored for over 10 years [S3, S4]. Previous studies have identified a suite of characteristics that differ between monogynous and polygynous colonies, including body size of workers and queens, colony size, queen longevity, and investment in reproductive offspring ([S1, S2, S5, S6], Table S1). We independently determined the social organization (monogynous, oligogynous or polygynous) of each of the colonies included in the present study by performing a parentage analysis using eight workers genotyped at eight microsatellite loci [S4].

To investigate the genetic basis of queen number, we sampled freshly emerged male offspring from monogynous and polygynous colonies. Male ants are haploid, which facilitates SNP calling and linkage analysis. For the genome-wide association study of colony social organization, we analyzed two to five males from each of 18 monogynous colonies and six males from each of five polygynous colonies, respectively (few polygynous colonies produce males each year [S5]). For the linkage map, we sampled 59 males from one monogynous colony.

We collected individuals from 91 field colonies of known social organization to study the distribution of the social chromosomes across colonies [S4]. We collected freshly emerged males and gynes from 41 colonies, workers from 71 colonies, and one mature, mated queen from each of 23 polygynous colonies. We measured the head size (as the maximum width across the eyes [S1]) of all oligogynous and polygynous workers and a subset of the monogynous workers (Table S1). We also collected 40 mature polygynous queens and measured their head sizes (Table S1). These queens are being maintained in the lab for further investigation, but we ascertained their genotype at socially diagnostic markers (see section VI) by typing at least 16 eggs from each individual. Individuals that produced 100% Sm/Sp or Sp/Sp offspring were deemed to be Sp/Sp individuals (mated to an Sm or Sp male, respectively), while those that produced a mix of genotypes were deemed to be heterozygous. While a Sm/Sm queen mated with a Sp male would also be expected to produce 100% Sm/Sp offspring, evidence from direct genotyping of queens and workers from polygynous colonies indicates that Sm/Sm queens are not present in these colonies (Table 1).

II. Genotyping-by-sequencing -- association study & mapping libraries

Laboratory methods

We constructed two genotyping-by-sequencing libraries following Parchman et al. [S7] with slight modifications. Briefly, we digested DNA samples with restriction enzymes EcoRI and MseI, ligated double-stranded adapters to the fragments (including individually barcoded EcoRI adapters), PCR amplified each individual sample, and size-selected the region between 300-500bp of the pooled library by agarose gel extraction. After ethanol precipitation, we quantified libraries by fluorimetry and sequenced them on an Illumina HiSeq 2000 at the Genomic Technologies Facility at the University of Lausanne using single-end 101bp reads. The library to test the association with social organization, which included males from 23 field colonies, was sequenced on a single flow cell lane. The mapping library, which included 59 haploid male offspring from one monogyne colony, was sequenced on two flow cell lanes, along with 101 samples from a different project. For the mapping library, we used modified MseI adapters allowing addition of Illumina index sequences [S8].

Data processing

We extracted SNP genotypes from raw Illumina sequence data using Stacks version 0.998 [S9]. Reads were demultiplexed and quality-filtered with `process_radtags`, with barcode distance 3 and default values for all other parameters. Genotypes were extracted using the `denovo_map` pipeline. We used minimum stack depth (`-m`) 3 and default values for all other parameters.

For the association study library, we exported the Stacks output in “genomic” format and used custom shell scripts to convert all heterozygous genotype calls to missing data, remove all loci with >10% missing data, remove loci with minor allele frequency <2%, and format the resulting matrix for import to R. During filtering, we noticed one individual with excessive heterozygosity, which we excluded from further analysis as a rare diploid male. After filtering, we retained 18199 SNPs, distributed within 13668 “GBS tags,” in 78 individuals.

For the linkage mapping library, before running the `denovo_map` pipeline we generated a “pseudo-queen” sample by concatenating subsamples of 50,000 reads per male. Since males result from unfertilized eggs, sequence data from 59 male offspring accurately reflects their mother’s genotype. We exported Stacks results in “haplotypes” format and used custom shell scripts to retain only loci that were heterozygous in the pseudo-queen, remove loci with male heterozygosity >5%, convert the remaining heterozygous genotypes to missing, remove loci with >20% missing data, and remove loci with a minor allele frequency of <10%. After filtering, we retained 2409 SNPs in 59 offspring.

Association tests

For each of the 18199 SNPs genotyped in multiple colonies, we tested for association with colony social organization using a mixed effects model (R package nlme). The model fit SNP genotype to social organization, with colony ID as a random factor (Table S2). We computed the uncorrected α corresponding to a genome-wide false discovery rate of 0.01 following Benjamini & Hochberg [S10]. We calculated F_{ST} following Hudson et al. [S11] using a custom R script.

Linkage map

A linkage map allows us to infer the relative position of genetic markers in the genome based on the amount of recombination observed between them. Markers are placed in linkage groups, which have similar patterns of segregation. For high quality maps, linkage groups will generally correspond to physical chromosomes. We constructed a linkage map based on 59 male offspring of a Sm/Sm queen using MSTmap [S12], following the procedure of Gadau [S13] to account for the unknown phase of maternal genotypes. We used the following parameter values: population_type DH, distance_function kosambi, cut_off_p_value 1e-6, no_map_dist 15, no_map_size 2. Linkage groups containing gaps of >50 cM (n=2) were split into two linkage groups at the gaps. Of 2409 markers, all but one were placed on 27 linkage groups; the one remaining marker was removed from the map.

Measurement of Linkage Disequilibrium

We estimated linkage disequilibrium (R^2), a measure of the correlation between alleles at different loci, for all pairs of SNPs that could be placed on linkage group 3 using the “cor” function in R. Separate LD estimates were generated for all 23 colonies of both social types, 18 monogyne colonies, and 5 polygyne colonies, and were displayed using the R package LDHeatmap [S14]. The higher LD observed within polygyne colonies compared to monogyne colonies is not an artifact of the smaller number of polygyne colonies sampled, as we obtained similar results with a random subsample of 3 individuals from 5 colonies of each social form (Fig. S1). We note that these estimates of LD are contingent on the individuals sampled, which include some full siblings. Nevertheless, these values allow us to visualize correlated blocks of markers across LG3.

Neighbor joining tree preparation

To visually display patterns of relatedness among individuals at different regions of the genome, we constructed distance trees based on SNP data. We exported SNP data for each individual in Fasta format from the following genomic regions: linkage group 3, all other linkage groups, and 7 partitions of linkage group 3 (Fig. S1). We then estimated unrooted neighbor-joining trees using the R package ape [S15] with uncorrected pairwise distances.

Estimates of divergence and polymorphism within LG3

We calculated polymorphism within monogyne and within polygyne haplotypes of the social chromosome following Nei and Li [S16] using a custom R script (Fig. S2). Net divergence was calculated using a modified version of Nei and Li's method [S16], subtracting monogyne polymorphism instead of the average of monogyne and polygyne polymorphism from the raw sequence divergence. Subtracting the average polymorphism may overestimate the divergence between Sm and Sp because of the loss of variability in the non-recombining region of Sp.

III. Genome sequencing

Laboratory methods

To facilitate inter-species synteny assessment, we collected 101bp paired-end whole-genome sequence data from a single male from a monogynous colony. The library was prepared by standard procedures using a 500bp insert size, multiplexed with one other sample, and sequenced on a single flow cell lane.

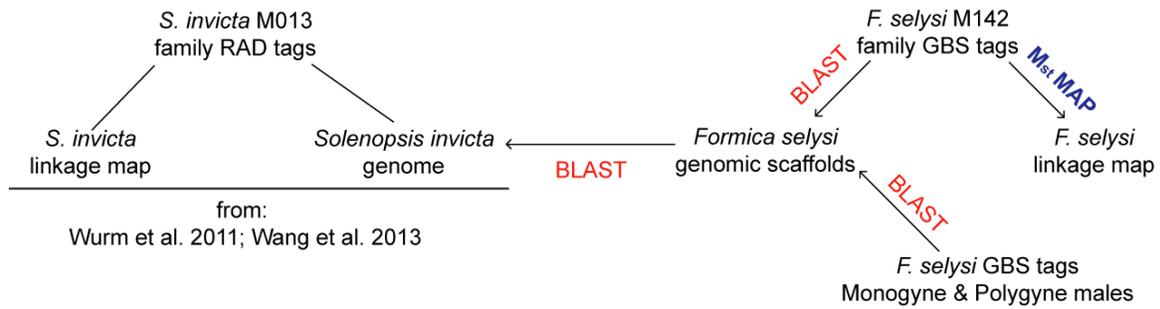
Data processing and assembly

We obtained 12.4 Gbp of raw whole-genome sequence data for one male. To clean the data prior to assembly, we removed all reads with the Illumina “failed-chastity” flag, filtered PCR duplicates using filterPCRDupl.pl [S17], collapsed overlapping paired-end reads to single reads using AdapterRemoval [S18], removed remaining adapter sequences using cutadapt [S19], trimmed each read to the longest segment containing bases with quality score 15 or greater using DynamicTrim.pl [S20], and removed all reads shorter than 40bp using LengthSort.pl [S20]. After filtering, we retained 9.8 Gbp of sequence data.

We performed de novo assembly with ABySS [S21], using all odd-numbered kmer values between 25 and 93, and used SSPACE [S22] to join contigs into scaffolds. In general, higher kmer values produced shorter assemblies with longer n50. We retained the assembly resulting from k=49 as the best compromise between completeness and contiguity. After removing contigs and scaffolds shorter than 200bp, this assembly was 224 Mbp in length, with a scaffold n50 of 16.1 kbp.

IV. Cross species comparison

We used BLASTN to align SNP-containing GBS sequences from both libraries to the *F. selysi* genome survey scaffolds, and to align the genome survey scaffolds to the *S. invicta* genome sequence, summarized as follows:



To avoid falsely inferring orthology of paralogous loci, we retained only blast hits with an e-value five orders of magnitude better than that of the second best hit. 13030 of 13873 GBS tags from the association study library (94%) aligned to the *F. selysi* genome survey, as did 2295 of 2409 loci from the mapping library (95%). The 1845 *F. selysi* scaffolds placed on the linkage map contained a total of 6672 SNPs from the association study library.

We placed *S. invicta* scaffolds [S23] on the *S. invicta* linkage map [S24] using data from Supplementary Table 8 of Wang et al. [S24]. Of the 1845 mapped *F. selysi* scaffolds, 1688 aligned to the *S. invicta* genome (91%). We could thus identify the approximate position on the *S. invicta* linkage map of 6328 SNPs from the *F. selysi* association study library. When an *F. selysi* scaffold aligned to an *S. invicta* scaffold spanning multiple map positions, we identified the *S. invicta* RAD marker that was located closest to the nucleotide position of the Blast hit, and assigned the *F. selysi* scaffold to the map position of that RAD marker.

In complement, we investigated the position on the fire ant genome of the remaining unmapped *F. selysi* markers that were significantly associated with social organization. The 458 SNPs that could be placed on the *F. selysi* genome sequence occurred on 216 unique scaffolds. None of these scaffolds aligned to the *S. invicta* social chromosome. Instead, the *F. selysi* scaffolds aligned to *S. invicta* scaffolds on *S. invicta* LG2, LG3, LG5, LG10, LG11, and LG12. These are six of the seven *S. invicta* linkage groups associated with *F. selysi* markers that map to the *F. selysi* social chromosome (Fig. 1).

V. Investigating suppressed recombination between Sm and Sp chromosomes

We designed nine microsatellite and five PCR-RFLP markers on the social chromosome to directly assess recombination between Sm and Sp haplotypes in heterozygous queens (Fig. 2A; primers, PCR conditions, and RFLP enzymes in table S3). These markers were genotyped in 80 worker offspring of four Sm/Sp queens. These queens were isolated in the laboratory in flunolined boxes (15x13x6 cm) with 10-20 nestmate workers, and provided with *ad libitum* access to standard ant food and water. Under these conditions, we could ensure that the brood was the offspring of a single queen; we collected late instar larvae and pupae for genotyping. No recombination was observed among the 11 markers located within the differentiated region of the

social chromosome. We calculated pairwise map distances between the three recombining markers and this non-recombining region using the Kosambi mapping function [S25]. We attempted to investigate recombination in the offspring of Sp/Sp queens, as well. However, the five queens that produced enough offspring to investigate so far were highly homozygous, and construction of a reliable linkage map was not possible.

VI. Distribution of genotypes within field colonies

We used three diagnostic PCR-RFLP markers to rapidly assess whether additional individuals carried the Sm, Sp, or both haplotypes. These markers were distributed across the non-recombining region of the social chromosome (Fig. 2A). One SNP occurred on a scaffold already assigned to a map position. For the other two SNPs, we aligned their associated scaffolds to the *S. invicta* genome and searched for the position of that scaffold on the *F. selysi* linkage map. SNP loci were amplified by nested PCR, digested using MspI and visualized by agarose gel electrophoresis. Primers and PCR conditions are shown in table S3. We applied this assay to the additional workers, queens and males collected from 91 field colonies of known social structure (see section I; Table 1).

Supplementary References

- S1. Schwander, T., Rosset, H., and Chapuisat, M. (2005). Division of labour and worker size polymorphism in ant colonies: the impact of social and genetic factors. *Behavioral Ecology and Sociobiology* 59, 215-221.
- S2. Meunier, J., and Chapuisat, M. (2009). The determinants of queen size in a socially polymorphic ant. *J. Evol. Biol.* 22, 1906-1913.
- S3. 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.
- S4. Purcell, J., and Chapuisat, M. (2013). Bidirectional shifts in colony queen number in a socially polymorphic ant population. *Evolution* 67, 1169-1180.
- S5. Rosset, H., and Chapuisat, M. (2006). Sex allocation conflict in ants: when the queen rules. *Curr. Biol.* 16, 328-331.
- S6. Rosset, H., and Chapuisat, M. (2007). Alternative life-histories in a socially polymorphic ant. *Evol. Ecol.* 21, 577-588.
- S7. Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W., and Buerkle, C.A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol.* 21, 2991-3005.
- S8. Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7, e37135.
- S9. Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J.H. (2011). Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3-Genes Genomes Genetics* 1, 171-182.

- S10. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 57, 289-300.
- S11. Hudson, R.R., Slatkin, M., and Maddison, W. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics* 132, 583-589.
- S12. Wu, Y., Bhat, P.R., Close, T.J., and Lonardi, S. (2008). Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLOS Genet* 4, e1000212.
- S13. Gadau, J. (2009). Phase-Unknown Linkage Mapping in Ants. Cold Spring Harbor Protocols 2009, pdb. prot5251.
- S14. Shin, J.-H., Blay, S., McNeney, B., and Graham, J. (2006). LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Softw.* 16, 1-9.
- S15. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289-290.
- S16. Nei, M., and Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl Acad. Sci. U.S.A.* 76, 5269-5273.
- S17. Smeds, L., and Künstner, A. (2011). ConDeTri-a content dependent read trimmer for Illumina data. *PLoS ONE* 6, e26314.
- S18. Lindgreen, S. (2012). AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* 5, 337.
- S19. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* 17, 10-12.
- S20. Cox, M.P., Peterson, D.A., and Biggs, P.J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11, 485.
- S21. Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19, 1117-1123.
- S22. Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27, 578-579.
- S23. Wurm, Y., Wang, J., Riba-Grognuz, O., Corona, M., Nygaard, S., Hunt, B.G., Ingram, K.K., Falquet, L., Nipitwattanaphon, M., Gotzek, D., et al. (2011). The genome of the fire ant *Solenopsis invicta*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5679-5684.
- S24. 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.
- S25. Kosambi, D. (1943). The estimation of map distances from recombination values. *Ann. Eug.* 12, 172-175.
- S26. Brelsford, A., and Irwin, D.E. (2009). Incipient speciation despite little assortative mating: the yellow-rumped warbler hybrid zone. *Evolution* 63, 3050-3060.