

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

2
3 Alan Brelsford ^{1,3*}, Jessica Purcell ^{2,3*}, Amaury Avril ³, Patrick Tran Van ³, Junxia Zhang²,
4 ⁴, Timothée Brütsch ³, Liselotte Sundström ⁵, Heikki Helanterä ⁶, Michel Chapuisat ³

5
6 * These authors contributed equally

7
8 ¹ Department of Evolution, Ecology and Organismal Biology
9 University of California Riverside
10 Riverside, CA 92521
11 USA

12
13 ² Department of Entomology
14 University of California Riverside
15 Riverside, CA 92521
16 USA

17
18 ³ Department of Ecology and Evolution
19 University of Lausanne
20 1015 Lausanne
21 Switzerland

22
23 ⁴ College of Life Sciences
24 Hebei University
25 Baoding, Hebei, 071002
26 China

27
28 ⁵ Organismal and Evolutionary Biology research Programme
29 Faculty of Biological and Environmental Sciences
30 University of Helsinki
31 00014 Helsinki
32 Finland

33
34 ⁶ Ecology and Genetics Research Unit
35 University of Oulu
36 90014 Oulu
37 Finland

38
39
40 Correspondence to: alan.brelsford@ucr.edu, jessica.purcell@ucr.edu, or
41 michel.chapuisat@unil.ch

42
43 Lead contact: Alan Brelsford

44
45 **Running title:** Supergene evolution in ants

46 **Summary:** Supergenes, clusters of tightly linked genes, play a key role in the evolution
47 of complex adaptive variation [1,2]. While supergenes have been identified in many
48 species, we lack an understanding of their origin, evolution and persistence [3]. Here, we
49 uncover 20-40 MY of evolutionary history of a supergene associated with polymorphic
50 social organization in *Formica* ants [4]. We show that five *Formica* species exhibit
51 homologous divergent haplotypes spanning 11 Mbp on chromosome 3. Despite the
52 supergene's size, only 142 single nucleotide polymorphisms (SNPs) consistently
53 distinguish alternative supergene haplotypes across all five species. These conserved
54 trans-species SNPs are localized in a small number of disjunct clusters distributed across
55 the supergene. This unexpected pattern of divergence indicates that the *Formica*
56 supergene does not follow standard models of sex chromosome evolution, in which
57 distinct evolutionary strata reflect an expanding region of suppressed recombination (e.g.
58 [5]). We propose an alternative "eroded strata model," in which clusters of conserved
59 trans-species SNPs represent functionally important areas maintained by selection in the
60 face of rare recombination between ancestral haplotypes. The comparison of whole
61 genome sequences across 10 additional *Formica* species reveal that the most conserved
62 region of the supergene contains a transcription factor essential for motor neuron
63 development in *Drosophila* [6]. The discovery that a very small portion of this large and
64 ancient supergene harbors conserved trans-species SNPs linked to colony social
65 organization suggests that the ancestral haplotypes have been eroded by recombination,
66 with selection preserving differentiation at one or a few genes generating alternative
67 social organization.

68

69 **Results and Discussion:**

70

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

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

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

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

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

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

107 161 individuals, 44,427 sites, mean depth 27.7), and *F. lemani* (Serviformica, 65
108 individuals, 64,260 sites, mean depth 15.3). In each species, we find elevated
109 differentiation between individuals of monogyne and polygyne origin at chromosome 3
110 compared with other chromosomes (Figure 1), suggesting that an ancestral supergene is
111 present and associated with colony queen number in the five species. Through principal
112 component analysis (PCA) of variation on this chromosome, we show that individuals of
113 monogyne origin are usually homozygous for one supergene haplotype, while individuals
114 of polygyne origin are usually either heterozygous or are homozygous for an alternative
115 haplotype (Figure S2). This association is perfect in *F. selysi*, *F. exsecta*, and *F.*
116 *truncorum*, while mismatches between social structure and supergene genotype are
117 observed in 5 of 39 *F. lemani* and 35 of 96 *F. cinerea* individuals with known social
118 structure. Nonetheless, association between the presence of an Sp haplotype and
119 polygyne social structure was significant even in the latter two species (Fisher's exact test
120 $p = 0.00002$ in *F. cinerea*, $p = 0.000002$ in *F. lemani*). To further investigate the
121 relationship between each haplotype across species, we selected homozygous workers or
122 haploid males for subsequent whole genome sequencing. For *F. lemani*, we included
123 three individuals, representing homozygotes for three alternative haplotypes (Figure S2).

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

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

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

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

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

169 Small regions with trans-species fixed SNPs could be due to balancing selection
170 or physical constraints (e.g. inversion breakpoints) that prevent recombination from
171 homogenizing these genomic regions [22]. To distinguish between these hypotheses and
172 further test the expanding strata model, we identified genomic rearrangements between
173 the alternative supergene haplotypes. We constructed high density linkage maps using
174 ddRAD genotypes from the female offspring of two Sp/Sp *F. selysi* queens (112
175 offspring total, 1792 and 3688 markers, mean sequence depth 36.6). We also constructed
176 a linkage map from the male offspring of one Sm/Sm *F. exsecta* queen (67 offspring,
177 4603 markers, mean sequence depth 17.3) to determine whether the structure of the Sm
178 haplotype is conserved across species. We mapped the positions of the old, intermediate,
179 and young strata onto the *F. selysi* Sm genome. We then aligned the genome to the
180 linkage maps. The Sm haplotype of *F. exsecta* was collinear with that of *F. selysi* (Figure
181 3). The conserved gene order on the Sm haplotype suggests that this haplotype is
182 ancestral. In contrast, the alignment of the Sp haplotype of *F. selysi* to the Sm genome
183 revealed at least four inversions along the length of the supergene (Figure 3). Regions of
184 the supergene exhibiting the ‘old strata’ topologies were not localized on a single
185 inversion, but instead were distributed across the supergene (Figures 2, 3), suggesting
186 that haplotypes spanning the entire non-recombining region began to diverge prior to the
187 divergence of all the *Formica* species we examined. At least some trans-species fixed
188 SNPs were not close to inversion breakpoints based on a qualitative assessment of the
189 linkage maps, suggesting that balancing selection, and not exclusively structural
190 constraint, plays a role in maintaining these SNPs. Occasional recombination is suggested
191 by intermediate strata topologies that were patchily distributed across the length of the
192 supergene (Figure 3). As expected, the recombining regions at the ends of the supergene
193 followed the ‘young strata’ pattern (Figures 2, 3).

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

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

213 Trans-species SNPs associated with a trait of interest can point to genomic
214 regions responsible for the trait [30-32]. To characterize the most conserved trans-species
215 SNPs and identify candidate genes determining alternative social organization, we
216 sequenced the genomes of 10 additional European *Formica* species (Figure 2). Six of
217 these additional species spanning three subgenera matched the Sm haplotype for 126 out
218 of the 142 conserved SNPs associated with social organization in the initial comparative
219 analysis of 5 focal species. One *Formica* sensu stricto matched the Sp haplotype across
220 113 of 135 conserved SNPs. Finally, three species had excess heterozygosity across the
221 whole supergene and were heterozygous at a subset of the conserved SNPs (Figure 2).
222 Overall, only 20 SNPs were conserved across all 15 *Formica* species. All but one of these
223 conserved SNPs were located in the last exon and 3’ untranslated region of the gene
224 *Knockout* (Figure 2). The gene *Knockout* is a storkhead-box transcription factor essential
225 for motor neuron development in *Drosophila* [6]. Additional SNPs conserved across all
226 species except *F. picea* occurred in an intron of serine-threonine kinase *STK32B*, an exon
227 of mitochondrial ribosomal protein *MRPL34*, and regions just downstream of the genes
228 *RPUSD4* and *G9A*.

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

248 Recent studies discovered that independent, convergent supergenes underlie
249 polymorphisms in social organization in at least three ant lineages (*Solenopsis invicta*,
250 [38]; *Formica selysi*, [4]; *Leptothorax acervorum*, [39]). We do not yet know the extent
251 of similarities in the evolutionary history of these convergent ‘social’ supergenes [4, 38].
252 An analysis of divergence between *S. invicta* SB and Sb haplotypes revealed no evidence
253 of evolutionary strata [40], despite the presence of at least two inversions [41]. However,
254 an early analysis of the odorant binding protein gene Gp9, which was subsequently found
255 to be contained within the *Solenopsis* supergene, identified conserved polymorphisms
256 across several *Solenopsis* species [42], and this was confirmed in a recent comparative
257 genomic analysis of *S. invicta*, *S. richteri*, and *S. quinquecupis* [43]. The combination of
258 a lack of strata, multiple inversions, and trans-species polymorphism suggests that an

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

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

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

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

294

295 **Acknowledgments**

296 Funding was provided by Swiss National Science Foundation grants 31003A-146641 and
297 31003A-173189 to M.C., and by University of Lausanne and University of California
298 Riverside. Computational resources were provided by the Vital-IT computing cluster at
299 the Swiss Institute for Bioinformatics and the High Performance Computing Cluster at
300 University of California Riverside. We thank Purcell, Brelsford, and D. Schluter lab
301 members for helpful comments and discussions, and K. Loope and K. Thompson for
302 detailed feedback on an earlier version of the manuscript.

303

304 **Author Contributions**

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

309

310 **Declaration of Interests**

311 The authors declare no competing interests.

312

313 **Figure Legends**

314

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

324

325 **Figure 2: Evolution of alternative haplotypes of the social supergene across species**
326 **of the genus *Formica*.** Fixed differences between Sm and Sp haplotypes across five focal
327 species, based on whole-genome sequence data, are concentrated in small regions across
328 chromosome 3 (A; number of conserved trans-species SNPs associated with social
329 organization in 1 kbp windows; n=142). In several small regions distributed across the
330 center of the chromosome, the sequences cluster by social form (red, “old strata”
331 topology), while in large regions at the chromosome ends the sequences cluster by
332 species (blue, “young strata” topology) (B; Hidden Markov Model of tree topology
333 implemented in Saguaro). The 142 SNPs with fixed differences between Sm and Sp
334 haplotypes across the five focal species were sequenced in single representatives of 10
335 additional species (C), with alleles matching the Sm haplotype shown in green and alleles
336 matching the Sp haplotype shown in orange. Only a single region of 1,021 bp (positions
337 11,910,116 – 11,911,137) harbors SNPs that are consistently fixed between Sm and Sp
338 haplotypes across all 15 species. See also Figure S1 and Table S1.
339

340 **Figure 3: Structural rearrangements between alternative haplotypes of the *Formica***
341 **supergene.** The chromosome-level *Formica selysi* genome assembly for the Sm
342 supergene haplotype (middle; PacBio long read sequencing combined with linkage map)
343 is collinear with the Sm haplotype of *F. exsecta* (top; linkage map from a *F. exsecta*
344 Sm/Sm family). In contrast, the Sp haplotype of *F. selysi* reveals several inversions and
345 rearrangements compared to the Sm haplotype (bottom; merged linkage map from two *F.*
346 *selysi* Sp/Sp families). Lines between bars connect the RADtags in the linkage maps to
347 their position in the *F. selysi* genome assembly. Colored bars along the Sm haplotype of
348 *F. selysi* indicate the strata topologies inferred by Saguaro from whole-genome sequence
349 data across five *Formica* species. Blue bars represent regions of the supergene where
350 sequences cluster by species (young strata, B). Red bars show sections of the supergene
351 where sequences cluster by social form across all five species (old strata, C). Purple bars
352 show sections where sequences cluster by social form in the three *Serviformica* species,
353 and, separately, cluster by social form in *F. truncorum* and *F. exsecta* (D). Green bars
354 represent sections where sequences cluster by social form in the three *Serviformica*
355 species, but cluster by species for *F. truncorum* and *F. exsecta* (E). See also Figure S1.
356

357 **Figure 4: Comparison of the eroded strata model and the expanding strata model.** In
358 the eroded strata model (left panel), an initial inversion in one chromosome (basal blue
359 rectangle) greatly reduces recombination between two alternative haplotypes (red bars).
360 As new species form, this ancestral polymorphism is maintained, but occasional
361 recombination or gene conversion events (colored lines) homogenize sections of the
362 region in some lineages. The time series of plots at left represent the trans-species
363 divergence pattern expected under the eroded strata model, with disjunct regions
364 containing conserved trans-species polymorphisms. In contrast, in the expanding strata
365 model (right panel), new non-recombining regions appear sequentially in diverging
366 lineages, resulting in a pattern wherein young strata exhibit lower trans-species
367 differentiation than old strata (time series of plots at right). The topologies in the central
368 panel show the relationships between haplotypes and species for the young and old strata
369 scenarios, with colors matching the strata colors shown in each tree.
370

371 **STAR Methods:**

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

373 Key Resources Table

374 Lead Contact and Materials Availability

375 Experimental Model and Subject Details

376 Method Details

377 Supergene presence in multiple species

378 Linkage maps

379 Genome assembly

380 Whole-genome resequencing

381 Phylogeny and dating

382 Quantification and Statistical Analysis

383 Data and software availability

384

385 **Key Resource Table**

386 Attached as separate document

387

388 **Lead Contact and Materials Availability**

389 Further information and requests for resources and reagents should be directed to and will

390 be fulfilled by the Lead Contact, Alan Brelsford (alan.brelsford@ucr.edu). There are

391 restrictions to the availability of tissue and DNA samples due to the lack of an external

392 centralized repository for their distribution and our need to maintain the stock. We are

393 glad to share oligonucleotides with reasonable compensation by requestor for processing

394 and shipping.

395

396 **Experimental Model and Subject Details**

397 With the exception of the linkage map of the *F. selysi* Sp haplotype, all ants used in this
398 study were collected in the wild (sample sizes and localities for each species in Table S1).

399 For the *F. selysi* Sp haplotype linkage map, we obtained captive-reared offspring of two

400 mature queens from polygynous field colonies in Finges, Switzerland. The supergene

401 genotype of these two queens had been previously determined to be Sp/Sp [11]. Queens

402 were kept in isolated plastic nest boxes (15 x 13 x 6 cm) containing a tube with water and

403 ad libitum access to ant food consisting of agar, egg, and sugar, with at least 20 nestmate

404 workers, and left to produce eggs. These queens produced 35 and 77 newly emerged

405 worker offspring, respectively, and we collected these for linkage mapping. All ants used

406 in this study were stored in 100% ethanol prior to DNA extraction.

407

408 **Method details**

409 *Supergene presence in multiple species*

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

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

436 *Linkage maps*

437 We collected ddRAD sequence data on offspring of two *F. selysi* homozygous
438 Sp/Sp queens (77 and 35 newly emerged workers, respectively) and 67 males collected
439 from a monogyne *F. exsecta* colony. Library preparation, sequencing, and SNP calling
440 were carried out as described above in the *Supergene presence in multiple species*
441 section. We filtered raw variant calls separately for each mapping family using VCFtools
442 version 0.1.13 [47], retaining genotypes of SNP and indel variants with quality score >20,
443 and variants with <20% missing data per family and per-family minor allele frequency
444 >15%. We then inferred linkage maps for each family using MSTmap [49], using the
445 Kosambi mapping function and p-value cutoffs of 5e-5 for the smaller *F. selysi* family
446 and 5e-6 for the *F. exsecta* family and larger *F. selysi* family; full parameter sets are
447 reported in Table S2. Linkage maps for two Sp/Sp families were merged using
448 MergeMap [50], weighting each map by the number of individuals used to construct it.
449 *Genome assembly*

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

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

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

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

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

480 *Whole-genome sequences*

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

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

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

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

507 *Phylogeny and dating*

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

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

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

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

530 above IQ-TREE analysis. Using the known fossil records of *Formica* in Baltic ambers
531 [65], we placed one calibration point at the MRCA of *Iberoformica* and *Formica*
532 (lognormal distribution with offset = 42 Ma, median = 60 Ma, 95% quantile = 90 Ma; see
533 [66]). The analysis was run for 60 000 000 generations (trees sampled at every 2000
534 generations). Tracer v1.7.1 [59] was used to check when the MCMCs had reached a
535 stationary distribution by visual inspection of plotted posterior estimates. Trees sampled
536 during the first 12 000 000 generations (20%) were removed as burn-in and the remaining
537 trees (24 001 in total) were summarized in TreeAnnotator v2.5.2 [57] using the
538 ‘Maximum clade credibility tree’ and ‘Mean heights’ options, and then displayed with
539 age in millions of years using FigTree v1.4.3. The 95% highest probability density (95%
540 HPD) values were summarized.

541 **Quantification and Statistical Analysis**

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

547 **Data and code availability**

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

556

557

558 **References Cited**

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

- 561 2. Charlesworth, D. (2016). The status of supergenes in the 21st century:
562 Recombination suppression in Batesian mimicry and sex chromosomes and other
563 complex adaptations. *Evol. Appl.* *9*, 74-90.
- 564 3. Wellenreuther, M., and Bernatchez, L. (2018). Eco-evolutionary genomics of
565 chromosomal inversions. *Tr. Ecol. Evol.* *33*, 427-440.
- 566 4. Purcell J., Brelford A., Wurm Y., Perrin N., and Chapuisat M. (2014).
567 Convergent genetic architecture underlies social organization in ants. *Curr. Biol.*
568 *24*, 2728–2732.
- 569 5. Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E.D., Gilbert,
570 M.T.P., and Zhang, G. (2014). Complex evolutionary trajectories of sex
571 chromosomes across bird taxa. *Science* *346*, 1246338.
- 572 6. Hartmann, C., Landgraf, M., Bate, M., and Jäckle, H. (1997). *Krüppel* target
573 gene *knockout* participates in the proper innervation of a specific set
574 of *Drosophila* larval muscles. *EMBO J.* *16*, 5299-5309.
- 575 7. Yang, Q., Zhang, D., Li, Q., Cheng, Z., and Xue, Y. (2007). Heterochromatic and
576 genetic features are consistent with recombination suppression of the self-
577 incompatibility locus in *Antirrhinum*. *Plant J.* *51*, 140-151.
- 578 8. Branco, S., Carpentier, F., Rodríguez de la Vega, R.C., Badouin, H., Snirc, A., Le
579 Prieur, S., Coelho, M.A., de Vienne, D.M., Hartmann, F.E., Begerow, D., et al.
580 (2018). Multiple convergent supergene evolution events in mating-type
581 chromosomes. *Nat. Commun.* *9*, 2000.
- 582 9. Iijima, T., Kajitani, R., Komata, S., Lin, C-P., Sota, T., Itoh, T., and Fujiwara, H.
583 (2018). Parallel evolution of Batesian mimicry supergene in
584 two *Papilio* butterflies, *P. polytes* and *P. memnon*. *Sci. Adv.* *18*, eaao5416.
- 585 10. Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M.A., Nowell, R.W., Mallett,
586 J., Dasmahapatra, K.K., and Joron, M. (2018). Supergene evolution triggered by
587 introgression of a chromosomal inversion. *Curr. Biol.* *28*, 1839-1845.
- 588 11. Avril, A., Purcell, J., Brelford, A., and Chapuisat, M. (2019). Asymmetric
589 assortative mating and queen polyandry are linked to a supergene controlling ant
590 social organization. *Mol. Ecol.* *28*, 1428-1438.
- 591 12. Rosset, H., and Chapuisat, M. (2007). Alternative life-histories in a socially
592 polymorphic ant. *Evol. Ecol.* *21*, 577–588.
- 593 13. Sundström, L. (1993). Genetic population structure and sociogenetic organisation
594 in *Formica truncorum* (Hymenoptera; Formicidae). *Behav. Ecol. Sociobiol.* *33*,
595 345-354.
- 596 14. Goropashnaya, A.V., Seppä, P., and Pamilo, P. (2001). Social and genetic
597 characteristics of geographically isolated populations in the ant *Formica cinerea*.
598 *Mol. Ecol.* *10*, 2807–2818.
- 599 15. DeHeer, C.J., and Herbers, J.M. (2004). Population genetics of the socially
600 polymorphic ant *Formica podzolica*. *Insect. Soc.* *51*, 309-316.
- 601 16. Seppä, P., Gyllenstrand, N., Corander, J., and Pamilo, P. (2004). Coexistence of
602 the social types: genetic population structure in the ant *Formica exsecta*.
603 *Evolution* *58*, 2462-2471.
- 604 17. Gyllenstrand, N., Seppä, P., and Pamilo, P. (2005), Restricted gene flow between
605 two social forms in the ant *Formica truncorum*. *J. Evol. Biol.* *18*, 978–984.

- 606 18. Bargum, K., Helanterä, H., and Sundström, L. (2007). Genetic population
607 structure, queen supersedure and social polymorphism in a social Hymenoptera. *J.*
608 *Evol. Biol.* *20*, 1351–1360.
- 609 19. Thompson, M. J., and Jiggins, C. D. (2014). Supergenes and their role in
610 evolution. *Heredity* *113*, 1-8.
- 611 20. Lahn, B.T., and Page, D.C. (1999). Four evolutionary strata on the human X
612 chromosome. *Science* *286*, 964-967.
- 613 21. Lawson-Handley, L-J., Ceplitis, H., and Ellegren, H. (2004). Evolutionary strata
614 on the chicken Z chromosome: implications for sex chromosome evolution.
615 *Genetics* *167*, 367-376.
- 616 22. Guerrero, R.F. Rousset, F., and Kirkpatrick, M. (2012). Coalescent patterns for
617 chromosomal inversions in divergent populations. *Philos. T. R. Soc. B* *367*,
618 20110246.
- 619 23. Kapun, M., and Flatt, T. (2019). The adaptive significance of chromosomal
620 inversion polymorphisms in *Drosophila melanogaster*. *Mol. Ecol.* *28*, 1263-1282.
- 621 24. Zamani, N., Russell, P., Lantz, H., Hoepfner, M.P., Meadows, J.R.S., Vijay, N.,
622 Mauceli, E., di Palma, F., Lindblad-Toh, K., Jern, P., et al. (2013). Unsupervised
623 genome-wide recognition of local relationship patterns. *BMC Genomics* *14*, 347.
- 624 25. Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLOS*
625 *Biol.* *8*, e1000501.
- 626 26. Korunes, K.L., and Noor, M.A.F. (2017). Gene conversion and linkage: effects on
627 genome evolution and speciation. *Mol. Ecol.* *26*, 351-364.
- 628 27. Korunes, K.L., and Noor, M.A.F. (2019). Pervasive gene conversion in
629 chromosomal inversion heterozygotes. *Mol. Ecol.* *28*, 1302-1315.
- 630 28. Ross, K.G., and Shoemaker, D. (2018). Unexpected patterns of segregation
631 distortion at a selfish supergene in the fire ant *Solenopsis invicta*. *BMC Genet.* *19*,
632 101.
- 633 29. da Silva, V.H., Laine, V.N., Bosse, M., Spurgin, L.G., Derks, M.F.L., van Oers,
634 K., Dibbitts, B., Slate, J., Crooijmans, R.P.M.A., Visser, M.E., et al. (2019). The
635 genomic complexity of a large inversion in great tits. *Genome Biol. Evol.* *11*,
636 1870-1881.
- 637 30. Brelsford, A., Dufresnes, C., and Perrin, N. (2016). Trans-species variation in
638 *Dmrt1* is associated with sex determination in four European tree-frog species.
639 *Evolution* *70*, 840-847.
- 640 31. Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M.,
641 Suetake, H., Suzuki, S., Hosoya, S., et al. (2012). A trans-species missense SNP
642 in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu*
643 *rubripes* (Fugu). *PLOS Genet.* *8*, e1002798.
- 644 32. Zhang, W., Westerman, E., Nitzany, E., Palmer, S., and Kronforst, M.R. (2017).
645 Tracing the origin and evolution of supergene mimicry in butterflies. *Nat.*
646 *Commun.* *8*, 1269.
- 647 33. Küpper, C., Stocks, M., Risse, J.E., dos Remedios, N., Farrell, L.L., McRae, S.B.,
648 Morgan, T.C., Karlionova, N., Pinchuk, P., Verkuil, Y.I., et al. (2016). A
649 supergene determines highly divergent male reproductive morphs in the ruff. *Nat.*
650 *Genet.* *48*, 79-83.

- 651 34. Lamichhaney, S. Fan, G., Widemo, F., Gunnarsson, U., Schwochow Thalmann,
652 D., Hoepfner, M.P., Kerje, S., Gustafson, U., Shi, C., Zhang, H., et al. (2016).
653 Structural genomic changes underlie alternative reproductive strategies in the ruff
654 (*Philomachus pugnax*). *Nat. Genet.* 48, 84-88.
- 655 35. Tuttle, E.M., Bergland, A.O., Korody, M.L., Brewer, M.S., Newhouse, D.J.,
656 Minx, P., Stager, M., Betuel, A., Cheviron, Z.A., Warren, W.C., et al. (2016).
657 Divergence and functional degradation of a sex chromosome-like supergene.
658 *Curr. Biol.* 26, 344-350.
- 659 36. Joron, M., Frezal, L., Jones, R.T., Chamberlain, N.L., Lee, S.F., Haag, C.R.,
660 Whibley, A., Becuwe, M., Baxter, S.W., Ferguson, L., et al. (2011).
661 Chromosomal rearrangements maintain a polymorphic supergene controlling
662 butterfly mimicry. *Nature* 477, 203-206.
- 663 37. Van Belleghem, S.M., Rastas, P., Papanicolaou, A., Martin, S.H., Arias, C.F.,
664 Supple, M.A., Hanly, J.J., Mallet, J., Lewis, J.J., Hines, H.M., et al. (2017).
665 Complex modular architecture around a simple toolkit of wing pattern genes. *Nat.*
666 *Ecol. Evol.* 1, 0052.
- 667 38. Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.C.,
668 Shoemaker, D., and Keller, L. (2013). A Y-like social chromosome causes
669 alternative colony organization in fire ants. *Nature* 493, 664-668.
- 670 39. Braim, B.S. (2015). Exploring the regulatory role of behaviour and genome
671 architecture in the socially polymorphic ant, *Leptothorax acervorum*. Unpublished
672 doctoral thesis, University of Leicester. <http://hdl.handle.net/2381/36076>
- 673 40. Pracana, R., Priyam, A., Levantis, I., Nichols, R.A., and Wurm, Y. (2017). The
674 fire ant social chromosome supergene variant Sb shows low diversity but high
675 divergence from SB. *Mol. Ecol.* 26, 2864-2879.
- 676 41. Huang, Y-C., Dang, V.D., Chang, N-C., and Wang, J. (2018). Multiple large
677 inversions and breakpoint rewiring of gene expression in the evolution of the fire
678 ant social supergene. *P. R. Soc. B* 285: 20180221.
- 679 42. Krieger, M.J.B., and Ross, K. (2002). Identification of a major gene regulating
680 complex social behavior. *Science* 295, 328-332.
- 681 43. Stolle, E., Pracana, R., Howard, P., Paris, C.I., Brown, S.J., Castillo-Carrillo, C.,
682 Rossiter, S.J., and Wurm, Y. (2019). Degenerative expansion of a young
683 supergene. *Mol. Biol. Evol.* 36, 553-561.
- 684 44. Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. (2013).
685 Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124-3140.
- 686 45. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with
687 Bowtie 2. *Nat. Methods* 9, 357.
- 688 46. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
689 Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and
690 SAMtools. *Bioinformatics* 25, 2078-2079.
- 691 47. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M.A.,
692 Handsaker, R.E., Lunter, G., Marth, G. T., Sherry, S. T., et al. (2011). The variant
693 call format and VCFtools. *Bioinformatics* 27, 2156-2158.
- 694 48. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D.,
695 Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J., et al. (2007) PLINK: a tool set

- 696 for whole-genome association and population-based linkage analyses. *Am. J.*
697 *Hum. Genet.* *81*, 559-575.
- 698 49. Wu, Y., Bhat, P.R., Close, T.J., and Lonardi, S. (2008a). Efficient and accurate
699 construction of genetic linkage maps from the minimum spanning tree of a graph.
700 *PLoS Genet.* *4*, e1000212.
- 701 50. Wu, Y., Close, T.J., and Lonardi, S. (2008b). On the accurate construction of
702 consensus genetic maps. *Comput. Sys. Bioinform.* *7*, 285-296.
- 703 51. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy,
704 A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer
705 weighting and repeat separation. *Genome Res.* *27*, 722-736.
- 706 52. Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: Interrogation of genome
707 assemblies. *F1000Research* *6*, 1287.
- 708 53. Prysycz, L.P., and Gabaldón, T. (2016). Redundans: An assembly pipeline for
709 highly heterozygous genomes. *Nucleic Acids Res.* *44*, 1-10.
- 710 54. Quinlan, A.R. (2014). BEDTools: the Swiss-army tool for genome feature
711 analysis. *Curr. Protoc. Bioinformatics.* *47*, 11-12.
- 712 55. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-
713 TREE: a fast and effective stochastic algorithm for estimating maximum-
714 likelihood phylogenies. *Mol. Biol. Evol.* *32*, 268-274.
- 715 56. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and
716 post-analysis of large phylogenies. *Bioinformatics* *30*, 1312-1313.
- 717 57. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., Suchard,
718 M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: A Software Platform
719 for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* *10*, e1003537.
- 720 58. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018).
721 Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.*
722 *67*, 901-904.
- 723 59. Brelsford, A., Dufresnes, C., and Perrin, N. (2016). High-density sex-specific
724 linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome
725 without information on offspring sex. *Heredity* *116*, 177-181.
- 726 60. Chapuisat, M., Bocherens, S., and Rosset, H. (2004). Variable queen number in
727 ant colonies: no impact on queen turnover, inbreeding, and population genetic
728 differentiation in the ant *Formica selysi*. *Evolution* *58*, 1064-1072.
- 729 61. Purcell, J., and Chapuisat, M. (2013). Bidirectional shifts in colony queen number
730 in a socially polymorphic ant population. *Evolution* *67*, 1169-1180.
- 731 62. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis
732 of population structure. *Evolution* *38*, 1358-1370.
- 733 63. Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C.,
734 Langlade, N., and Muñoz, S. (2016). Extraction of high-molecular-weight
735 genomic DNA for long-read sequencing of single molecules. *BioTechniques* *61*,
736 203-205.
- 737 64. Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., and Vinh, L.S.
738 (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol.*
739 *Evol.* *35*, 518-522.
- 740 65. Dlussky, G.M. (1997). Genera of ants (Hymenoptera: Formicidae) from Baltic
741 amber. *Paleontol. J.* *31*, 616-627.

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

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Genomic DNA; <i>Formica exsecta</i> (n=108), <i>F. cinerea</i> (n=161), <i>F. selysi</i> (n=195), <i>F. truncorum</i> (n=20), <i>F. lemani</i> (n=65), <i>F. tombeuri</i> (n=1), <i>F. fusca</i> (n=1), <i>F. fuscocinerea</i> (n=1), <i>F. lugubris</i> (n=1), <i>F. sanguinea</i> (n=1), <i>F. rufibarbis</i> (n=1), <i>F. picea</i> (n=1), <i>F. pressilabris</i> (n=1), <i>F. fennica</i> (n=1), <i>F. pratensis</i> (n=1), <i>Polyergus mexicanus</i> (n=1), <i>P. vinosus</i> (n=1).	This paper	SRA PRJNA557080
Genomic DNA; <i>Iberofornica subrufa</i> (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 μm, 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		
<i>Formica selysi</i> genome assembly	This study	PRJNA557079
Linkage mapping RADseq data, <i>F. selysi</i> (n=112) and <i>F. exsecta</i> (n=67).	This study	PRJNA557080
Population RADseq data, <i>F. selysi</i>	[4]	PRJNA260459
Whole-genome sequence data, <i>F. selysi</i> Sm haplotype	[4]	SRX695613
Whole-genome sequence data, <i>Formica exsecta</i> (n=2), <i>F. cinerea</i> (n=2), <i>F. selysi</i> (n=1), <i>F. truncorum</i> (n=2), <i>F. lemani</i> (n=3), <i>F. tombeuri</i> (n=1), <i>F. fusca</i> (n=1), <i>F. fuscocinerea</i> (n=1), <i>F. lugubris</i> (n=1), <i>F. sanguinea</i> (n=1), <i>F. rufibarbis</i> (n=1), <i>F. picea</i> (n=1), <i>F. pressilabris</i> (n=1), <i>F. fennica</i> (n=1), <i>F. pratensis</i> (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

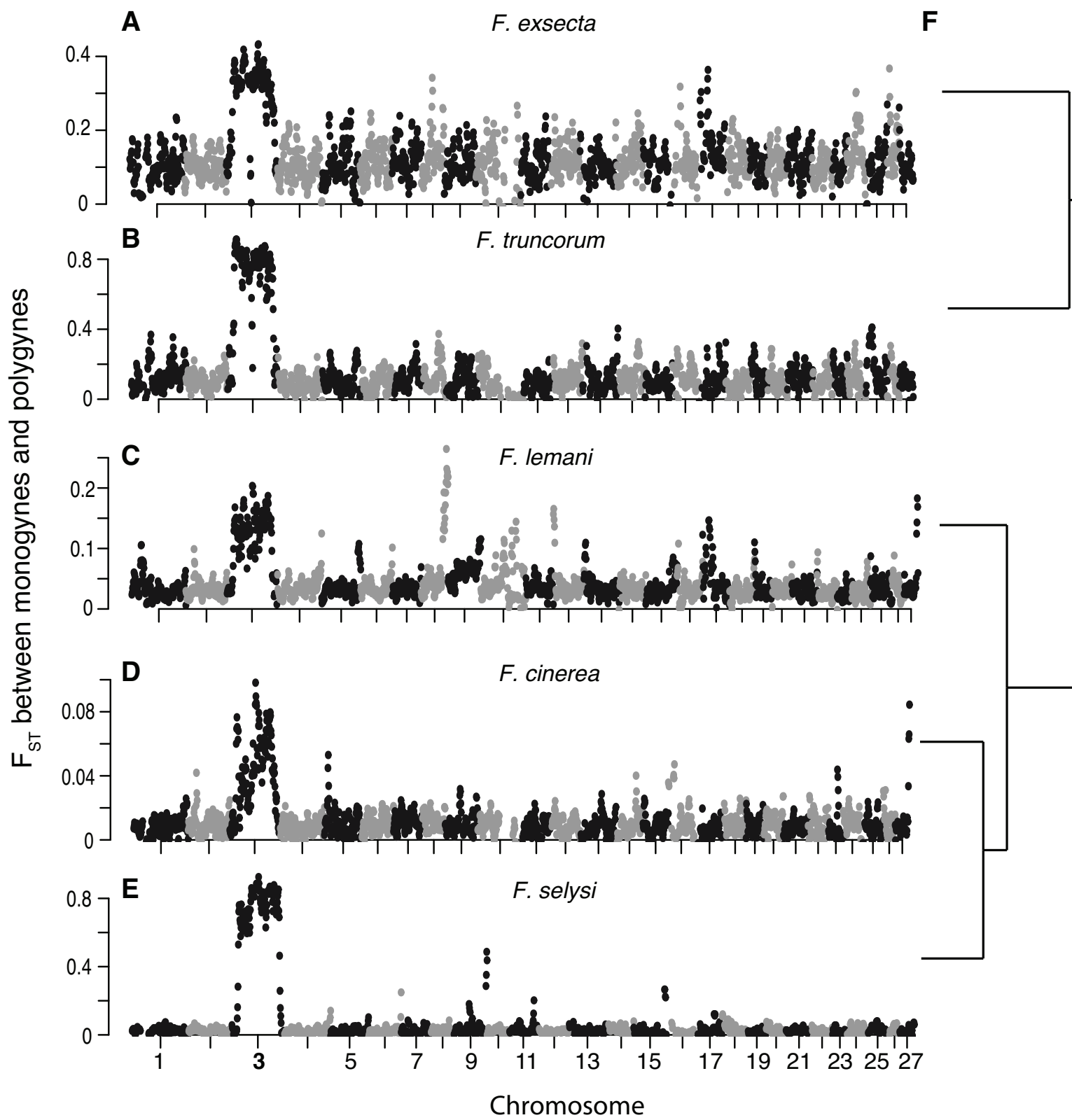


Figure 2

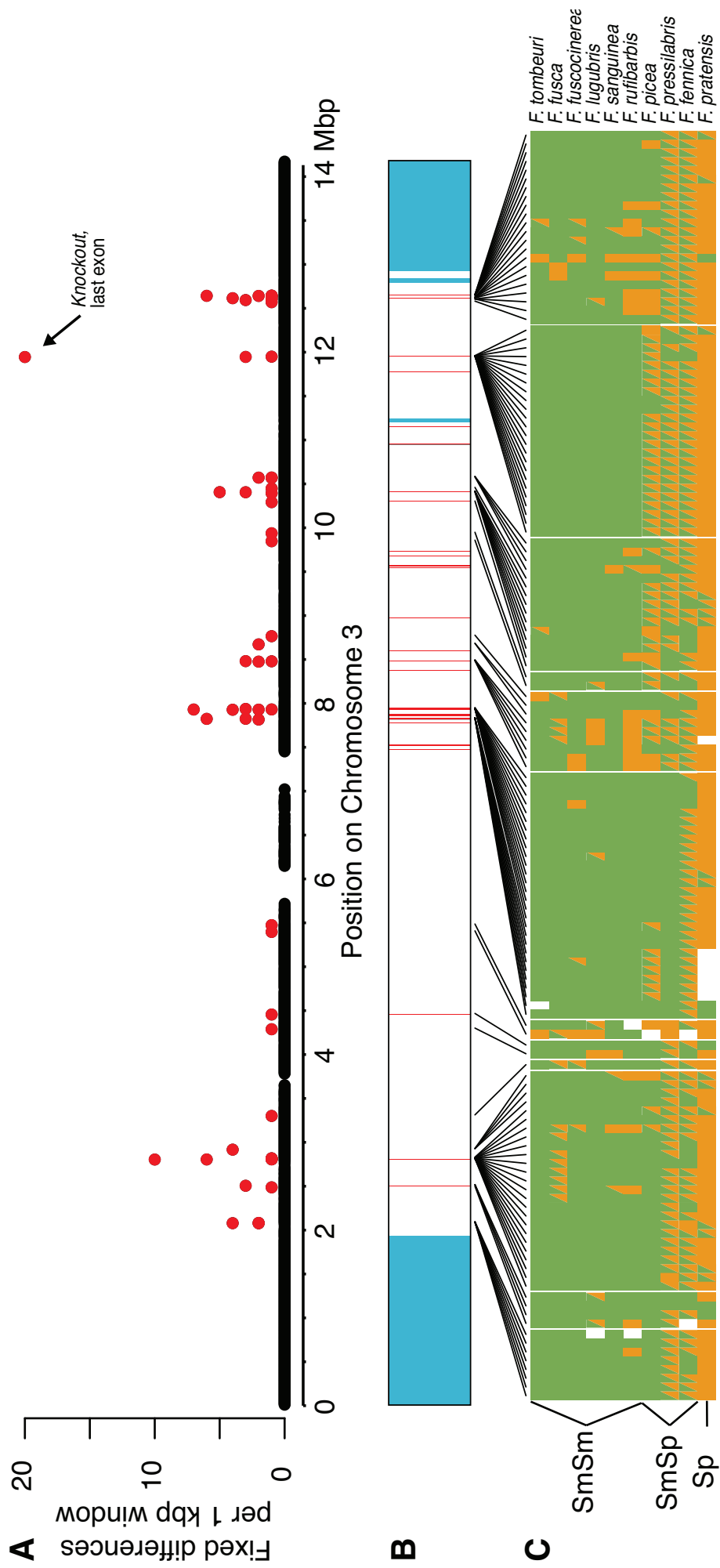


Figure 3

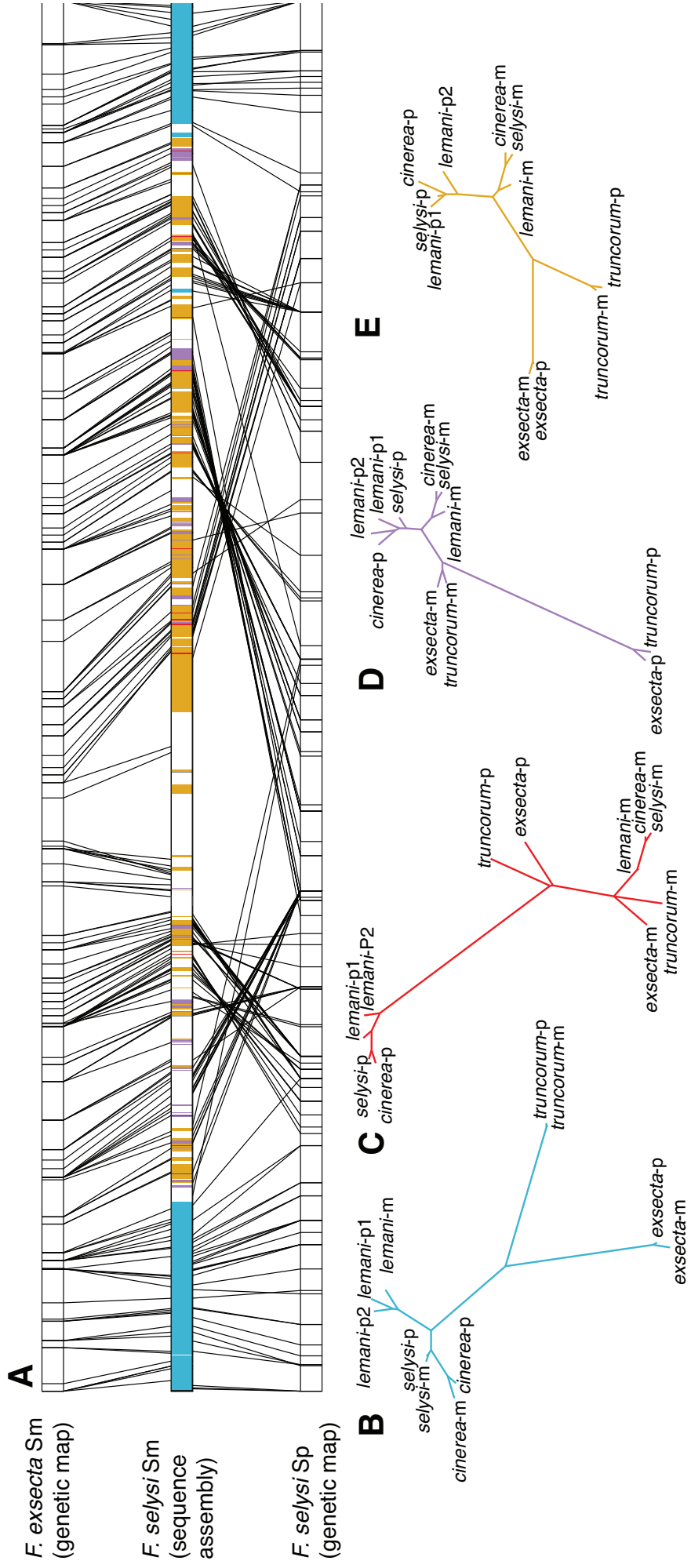
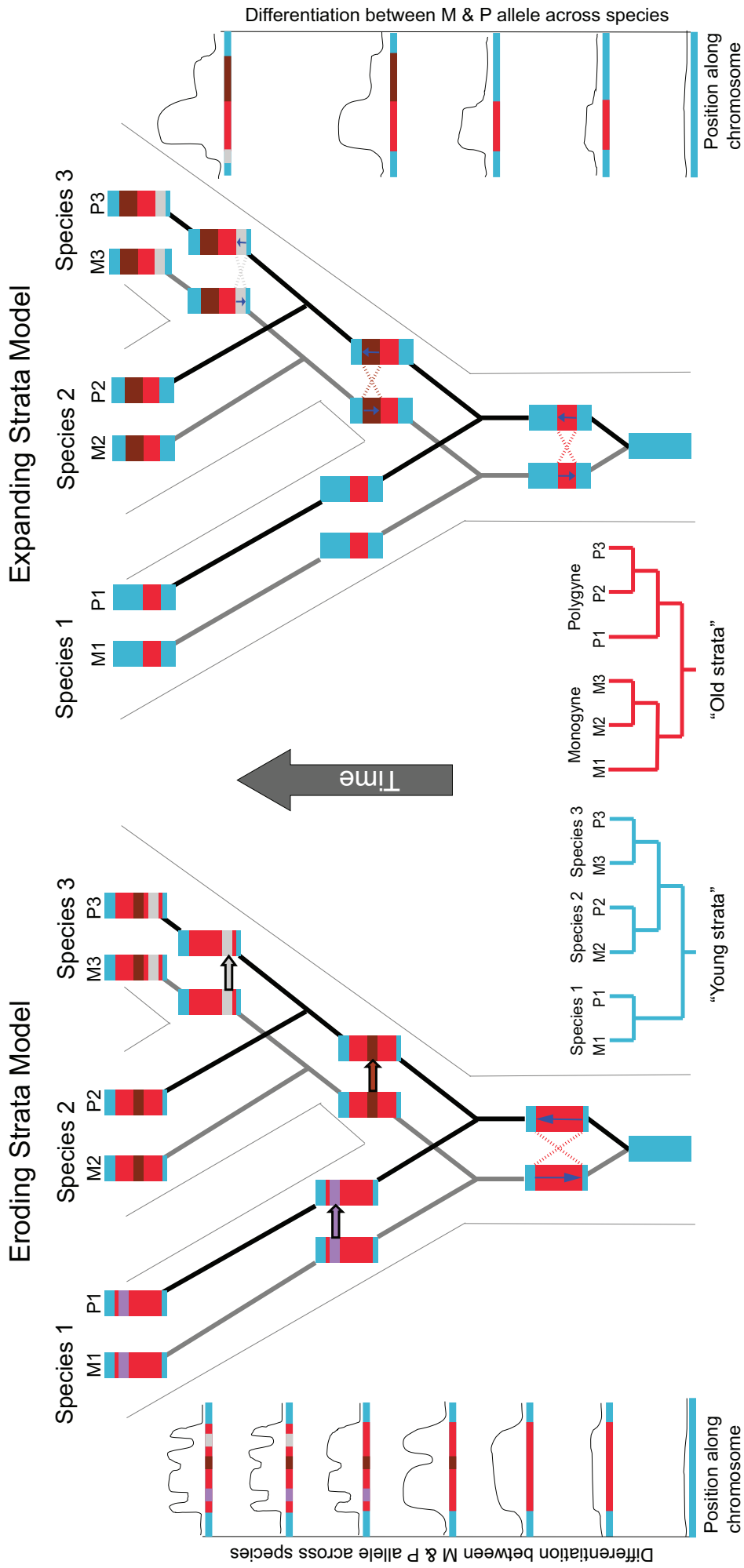


Figure 4



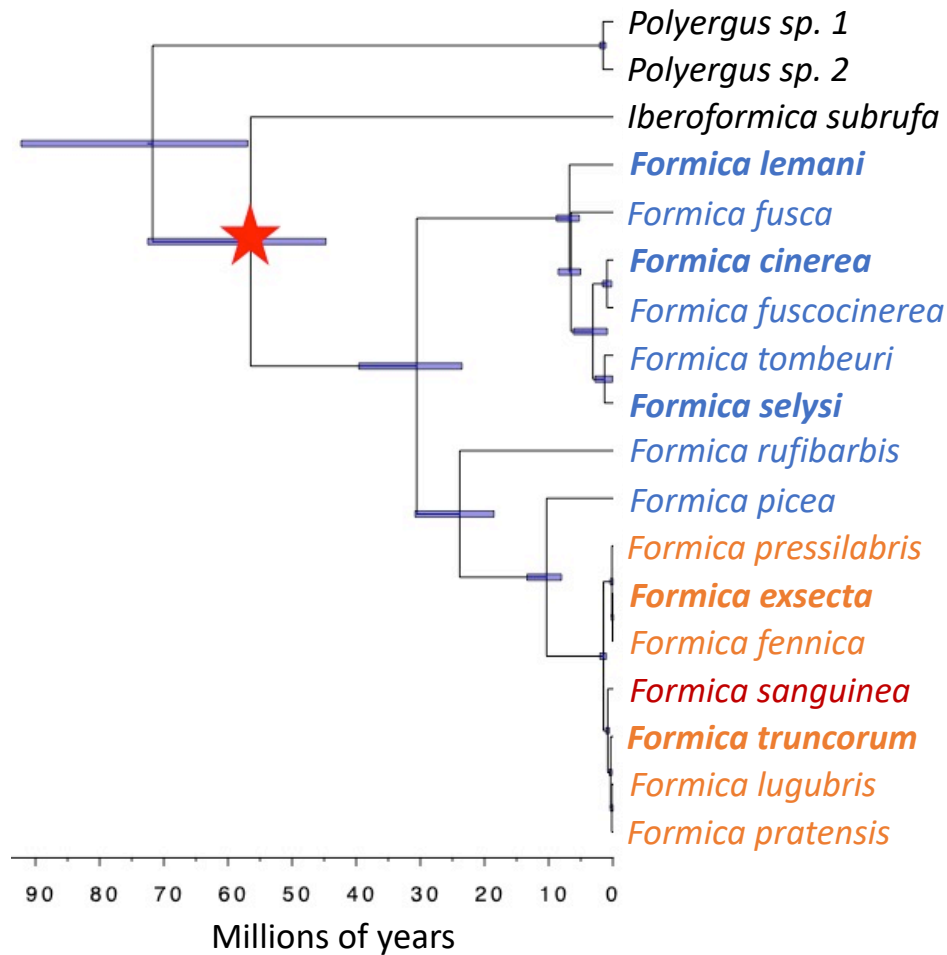


Figure S1. Time-calibrated Phylogeny of 15 *Formica* Species, Related to Figures 1, 2, 3. The *Formica* species examined here span an estimated 30 million years of evolutionary history. This phylogeny, implemented in BEAST (see STAR Methods), shows the 15 species investigated here as well as three outgroups. The calibration point is shown as the red star, and 95% highest posteriority 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.

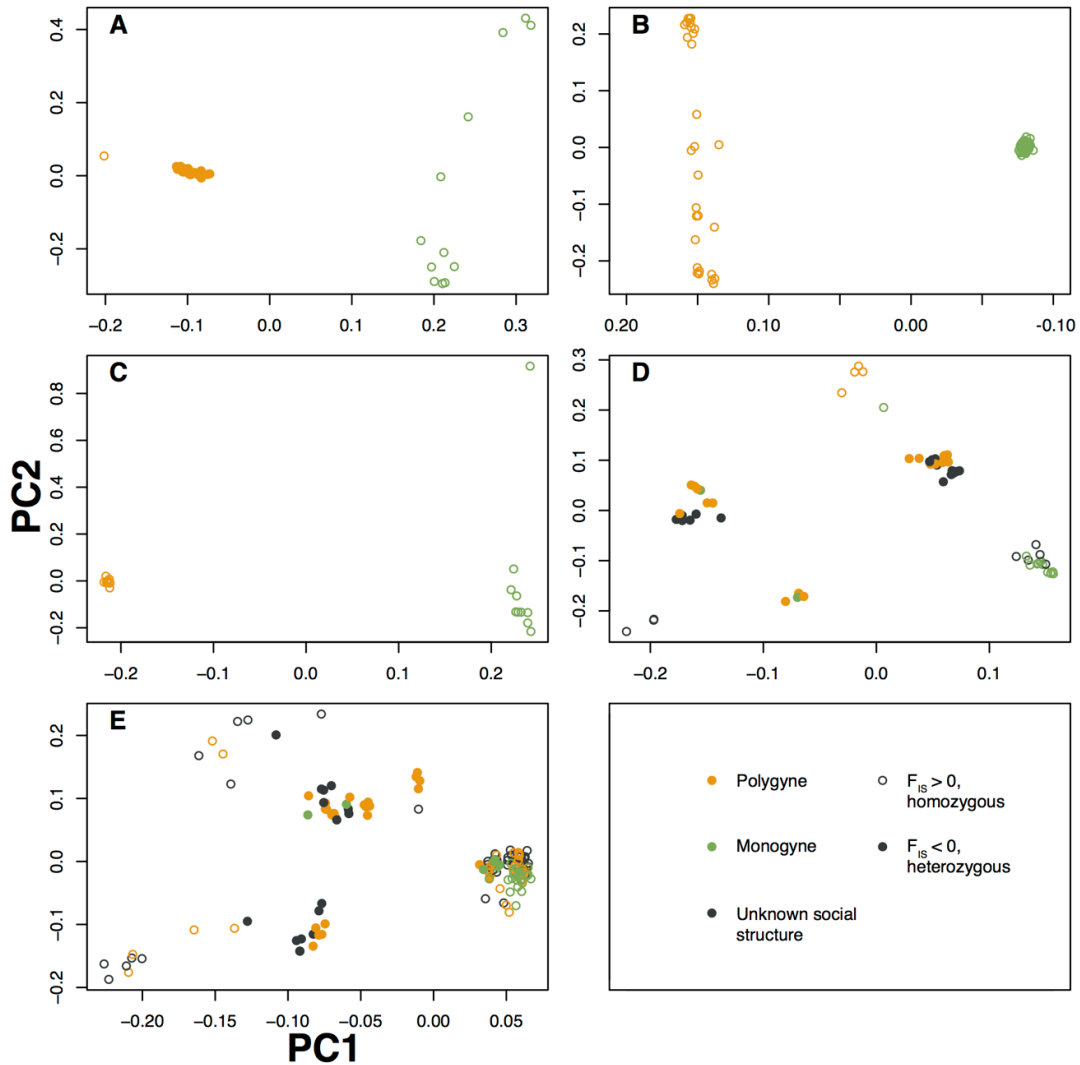


Figure S2. Principal Component Analyses of SNPs from Population ddRAD Data on Chromosome 3, Related to Figure 1. PCAs show two to six clusters in each species, with cluster membership strongly associated with colony social organization. Each panel represents an independent PCA in one species: *F. exsecta* (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
<i>Formica selysi</i>	Switzerland	83 males (monogyne and polygyne)	112 workers (polygyne)	2 males (Sm, Sp)
<i>F. cinerea</i>	Finland, Italy, Switzerland	161 workers (monogyne, polygyne, and unknown)	NA	1 male (from Switzerland: Sm) 1 worker (from Italy: Sp/Sp)
<i>F. lemani</i>	Finland, Spain, Switzerland	65 workers (monogyne, polygyne, and unknown)	NA	3 workers (from Switzerland: Sm/Sm, Sp ₁ Sp ₁ , Sp ₂ Sp ₂)
<i>F. exsecta</i>	Finland	12 males (monogyne) and 29 workers (polygyne)	67 males (monogyne)	1 male (Sm), 1 worker (Sp/Sp)
<i>F. truncorum</i>	Finland	5 males and 5 workers (monogyne) and 10 males (polygyne)	NA	2 males (Sm and Sp)
<i>F. tombeuri</i>	Spain	NA	NA	1 worker (inferred Sm/Sm)
<i>F. fusca</i>	Switzerland	NA	NA	1 worker (inferred Sm/Sm)
<i>F. fuscocinerea</i>	Switzerland	NA	NA	1 worker (inferred Sm/Sm)
<i>F. lugubris</i>	Switzerland	NA	NA	1 worker (inferred Sm/Sm)
<i>F. sanguinea</i>	Germany	NA	NA	1 worker (inferred Sm/Sm)
<i>F. rufibarbis</i>	Portugal	NA	NA	1 worker (inferred Sm/Sm)
<i>F. picea</i>	Finland	NA	NA	1 worker (polygyne, inferred Sm/Sp)
<i>F. pressilabris</i>	Finland	NA	NA	1 worker (polygyne, inferred Sm/Sp)
<i>F. fennica</i>	Finland	NA	NA	1 worker (polygyne, inferred Sm/Sp)
<i>F. pratensis</i>	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	<i>F. selysi</i> Sp/Sp queen, colony 191	<i>F. selysi</i> Sp/Sp queen, colony 192	<i>F. exsecta</i> 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
<i>F. selysi</i>	F92M2	Switzerland	M	Sm	HiSeq 2000, Lausanne	100bp PE	15.5
<i>F. selysi</i>	079M2	Switzerland	M	Sp	HiSeq 2500, Lausanne	100bp PE	19.6
<i>F. cinerea</i>	FcBra10	Switzerland	M	Sm	HiSeq 2500, Lausanne	100bp PE	12.5
<i>F. cinerea</i>	FcQuin3	Italy	F	Sp/Sp	HiSeq 2500, Lausanne	100bp PE	11.8
<i>F. exsecta</i>	FE-MM5	Finland	M	Sm	HiSeq 2500, Lausanne	100bp PE	10.9
<i>F. exsecta</i>	FE-PW10	Finland	F	Sp/Sp	HiSeq 2500, Lausanne	100bp PE	8.8
<i>F. lemani</i>	FL-BG13A	Switzerland	F	Sm/Sm	HiSeq 2500, Lausanne	100bp PE	9.2
<i>F. lemani</i>	FL-BG25W1	Switzerland	F	Sp/Sp	HiSeq 2500, Lausanne	100bp PE	10.5
<i>F. lemani</i>	FL-BG9W1	Switzerland	F	Sp/Sp	HiSeq 2500, Lausanne	100bp PE	9.9
<i>F. truncorum</i>	FT-MM1	Finland	M	Sm	HiSeq 2500, Lausanne	100bp PE	9.6
<i>F. truncorum</i>	FT-PM6	Finland	M	Sp	HiSeq 2500, Lausanne	100bp PE	10.4
<i>F. picea</i>	Fpic1-2	Finland	F	Sm/Sp	HiSeq 2500, Lausanne	100bp PE	13.0
<i>F. rufibarbis</i>	PortoA	Portugal	F	Sm/Sm	HiSeq 2500, Lausanne	100bp PE	10.2
<i>F. tombeuri</i>	ainc1w9	Spain	F	Sm/Sm	HiSeq 4000, Berkeley	150bp PE	9.3
<i>F. fusca</i>	bg22w1	Switzerland	F	Sm/Sm	HiSeq 4000, Berkeley	150bp PE	7.5
<i>F. fuscocinerea</i>	furka4w1	Switzerland	F	Sm/Sm	HiSeq 4000, Berkeley	150bp PE	9.6
<i>F. lugubris</i>	lucg1w1	Switzerland	F	Sm/Sm	HiSeq 4000, Berkeley	150bp PE	6.9
<i>F. sanguinea</i>	fsanw1	Germany	F	Sm/Sm	HiSeq 4000, Berkeley	150bp PE	6.8
<i>F. fennica</i>	ob6pol4w1	Finland	F	Sm/Sp	HiSeq 4000, Berkeley	150bp PE	6.7
<i>F. pressilabris</i>	br6pol4w1	Finland	F	Sm/Sp	HiSeq 4000, Berkeley	150bp PE	7.9
<i>F. pratensis</i>	fp43m1	Finland	M	Sp	HiSeq 4000, Berkeley	150bp PE	9.1
<i>Iberoformica subrufa</i>	D1135	Spain	F	n/a	HiSeq 4000, Berkeley	150bp PE	4.9
<i>Polyergus vinosus</i>	scrc2w15	USA	F	n/a	HiSeq 4000, Berkeley	150bp PE	3.8
<i>Polyergus mexicanus</i>	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