Radiation and hybridization underpin the spread of the fire ant social supergene

Quentin Helleu[a,1], Camille Roux[b], Kenneth G. Ross[c] 🄳, and Laurent Keller[a,1] 🄳

Supergenes are clusters of tightly linked genes that jointly produce complex phenotypes. Although widespread in nature, how such genomic elements are formed and how they spread are in most cases unclear. In the fire ant *Solenopsis invicta* and closely related species, a "social supergene controls whether a colony maintains one or multiple queens. Here, we show that the three inversions constituting the *Social b* (*Sb*) supergene emerged sequentially during the separation of the ancestral lineages of *S. invicta* and *Solenopsis richteri*. The two first inversions arose in the ancestral population of both species, while the third one arose in the *S. richteri* lineage. Once completely assembled in the *S. richteri* lineage, the supergene first introgressed into *S. invicta*, and from there into the other species of the socially polymorphic group of South American fire ant species. Surprisingly, the introgression of this large and important genomic element occurred despite recent hybridization being uncommon between several of the species. These results highlight how supergenes can readily move across species boundaries, possibly because of fitness benefits they provide and/or expression of selfish properties favoring their transmission.

fire ants | social polymorphism | supergene | introgression | hybridization

Understanding how complex traits requiring multiple novel mutations arise and are maintained in populations is a long-standing question in evolutionary biology (1–3). Supergenes are clusters of multiple, tightly linked genes that collectively produce complex phenotypes (2). They have evolved independently in many taxa and are responsible for intraspecific polymorphism in several important morphological and behavioral traits. The most prominent examples of supergenes are heteromorphic sex chromosomes, which drive the alternate development of males or females. Other examples of supergenes underpinning alternate phenotypes include Batesian mimicry in numerous butterfly species (4–7), self-incompatibility and floral heteromorphy in plants (S locus) (8–11), male meiotic drive (sperm killers) (12), bird plumage color polymorphism (13), and alternative social organization in ant colonies (14–17). In most known cases, the structural integrity of the supergene results from chromosomal rearrangements that suppress local recombination and thereby prevent dissociation of the genetic elements responsible for the integrated expression of complex character suites (5, 10, 13, 16).

The first supergene producing alternative social organization was identified in the fire ant *Solenopsis invicta* (16). In this species, colonies contain either a single egg-laying queen (monogyne social form) or multiple queens (polygyne social form), a fundamental distinction associated with a suite of important individual- and colony-level phenotypic differences (15, 18–21). Studies of invasive populations in the United States have revealed that this supergene comprises two haplotypes, the *Social b* (*Sb*) and the *Social B* (*SB*), which regulate the polymorphism (16, 22). In *S. invicta*, monogyne colonies invariably have a single homozygous *SB/SB* queen and only *SB/SB* workers, while polygyne colonies always have multiple heterozygous (*SB/Sb*) queens as well as workers of all three genotypes (15). Moreover, in invasive US populations, the *Sb* haplotype is responsible for a selfish "green beard" effect whereby *SB/Sb* (polygyne) workers recognize the cuticular chemical profiles of *SB/SB* queens and selectively eliminate them as they mature sexually and initiate reproduction (23, 24). Recent studies revealed that social organization also is regulated by the presence/absence of workers with the *Sb* haplotype in six closely related species of the *Solenopsis saevissima* species group (i.e., *Solenopsis richteri, Solenopsis megergates, Solenopsis macdonaghi, Solenopsis quinquecuspis*, the undescribed *Solenopsis AdRX*, and the undescribed *Solenopsis near interrupta*) (22, 25, 26). These species collectively are referred to as the socially polymorphic South American fire ants (22).

The extant *Sb* haplotype invariably consists of three inversions that together span a region of ~11.4 Mb on chromosome 16 containing >476 described genes. As a result, recombination is greatly reduced between the *Sb* and the *SB* haplotypes (16, 22, 27). The largest inversion, *In(16)1*, spans 9.48 Mb of chromosome 16. A second inversion, *In(16)2*, spans 0.84 Mb between *In(16)1* and a third inversion, *In(16)3*. *In(16)2* likely emerged

Author affiliations: [a]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland; [b]Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France; and [c]Department of Entomology, University of Georgia, Athens, GA 30605

[1]To whom correspondence may be addressed. Email: quentin.helleu@unil.ch or laurent.keller@unil.ch.

after *In(16)1*, given that a small fragment of *In(16)1*—telomeric in *SB* but central in *Sb*—is inverted again in the *In(16)2 Sb* haplotype (22). *In(16)3* (1.07 Mb) is located ~25 kb distant from *In(16)2* in a pericentric region of chromosome 16 (22, 28). Previous analyses of the divergence between the *Sb* and *SB* haplotypes within species suggested that the three inversions arose over an evolutionarily short time span (22).

Reconstructing the evolutionary history of supergene evolution is essential to understanding how such remarkable genomic elements originate and, possibly, the sequence of incorporation of key phenotypic elements into complex traits. The socially polymorphic species are hypothesized to have diverged from the outgroup species *S. saevissima* and *Solenopsis metallica*, ~0.75 to 1.75 million y ago and then radiated within the past ~500,000 y (22, 29). While a single origin of the *Sb* haplotype is robustly supported (30, 31), it remains unclear whether this unique element originated in the ancestral population before the radiation, or whether it has invaded the known socially polymorphic species through introgression events (22, 29, 32). Multiple studies have documented recent hybridization involving three of the socially polymorphic *Solenopsis* species in both the native and introduced ranges (33–37), highlighting a demographic context conducive to such genomic invasion in at least some cases.

Here, we present a comprehensive analysis of the history of the fire ant supergene intended to shed light on when and in what order each inversion occurred, and whether introgression of the *Sb* haplotype between species best explains observed patterns in the comparative genomic data. Using complementary approaches, we conclude that *In(16)3* is the oldest inversion and that it likely emerged in the common ancestor of *S. richteri* and the *S. invicta/AdRX* lineage. *In(16)1* emerged next, during the divergence between the *S. richteri* and *S. invicta/AdRX* lineages. Finally, the youngest [*In(16)2*] inversion emerged in the *S. richteri* lineage. The supergene comprising all three inversions thus emerged in

S. richteri, only to spread to the five most closely related species and confer the social polymorphism to them.

Results

Emergence of the Supergene Inversions in the *S. invicta/AdRX–S. richteri* Subgroup. To understand the evolutionary history of the *Sb* haplotype, we first reconstructed the species phylogeny using whole-genome data from 107 *SB* and 58 *Sb* haploid males from six species that are socially polymorphic (*S. invicta*, the undescribed *S. AdRX*, *S. richteri*, *S. megergates*, *S. macdonaghi*, and the undescribed *S. nr. interrupta*) as well as 20 haploid males from two outgroup species (*S. saevissima* and *S. metallica*). We applied two phylogenetic approaches, a maximum-likelihood phylogenetic analysis of a concatenated alignment of the 185 individuals' chromosomes 1 to 15 (supergene-bearing chromosome 16 not included) and a gene-tree coalescent-based method using randomly selected nonoverlapping genomic fragments (Fig. 1*A* and *SI Appendix, Text* and Fig. S1). The two approaches yielded concordant topologies in which the six species fall into five nested monophyletic groups. However, the phylogenomic analyses also revealed widespread phylogenetic discordance due to incomplete lineage sorting (ILS) and/or introgression (see below and Fig. 1*A* and *SI Appendix*, Figs. S1 and S2).

To reconstruct the evolutionary history of the *Sb* haplotype, we examined the phylogenetic relationships of the *Sb* and *SB* haplotypes by conducting maximum-likelihood analyses of complete alignments of each inversion in the 185 sequenced haploid male genomes. Each of the three inversions yielded a different, well-supported topology, yet for all three inversions the *Sb* haplotypes always formed a well-supported monophyletic group (Fig. 1*B*). Examination of these topologies along with the species tree revealed that *In(16)3* is likely to be the oldest inversion; specifically, the *Sb* haplotype clade with this
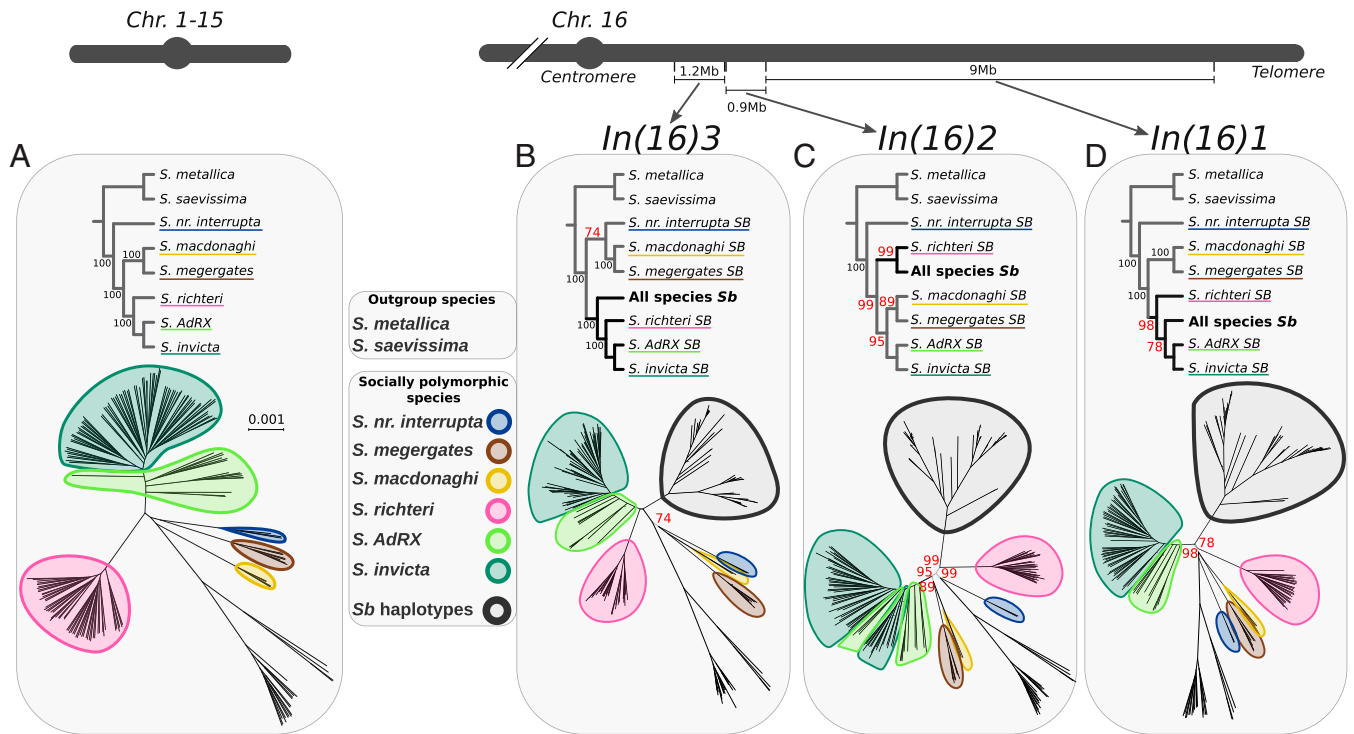
Fig. 1. Phylogenetic hypotheses for six socially polymorphic fire ant species and their three *Sb* supergene inversions. Rooted cladograms (*Top*) and unrooted phylogenies (*Bottom*) of chromosomes 1 to 15 (*A*) and of inversion *In(16)3* (*B*), inversion *In(16)2* (*C*), and inversion *In(16)1* (*D*) based on complete genomes of 185 (haploid) males. For each inversion phylogeny, *Sb* haplotypes are circled with a thick black line. Sequences from the outgroup species are not circled. Bootstrap support values <100 on inner branches between species lineages are shown in red type.

inversion forms a sister group to the *SB* haplotypes of *S. invicta/AdRX* and *S. richteri*, the branching pattern of which mirrors the branching pattern in the species tree (Fig. 1*B*). This result is consistent with inversion *In(16)3* having emerged before the initial divergence of the three species lineages commenced.

In(16)1 was likely the next inversion to appear. This inversion haplotype forms a poorly supported sister group to the *SB* haplotypes of *S. AdRX* and *S. invicta*, with low statistical support for this inner branch (bootstrap value 78; Fig. 1*D*) as well as generally shallow node depths in this part of the tree. To investigate whether the low support for the focal branch was an artifact of concatenating segments of *In(16)1* having different evolutionary histories, we used a sliding window–based phylogenetic analysis. This analysis indicates that *In(16)1* is not a mosaic of genomic segments differing in their evolutionary histories (*SI Appendix*, Fig. S3). Instead, the poorly resolved node(s) in this area of the *In(16)1* tree may be the product(s) of other historical processes, including ILS (expected with short branch lengths) and/or recombination between the *SB* and *Sb* haplotypes.

Finally, *In(16)2* appears to be the youngest inversion. Haplotypes of *In(16)2* form a sister group to the *S. richteri SB* haplotypes, suggesting that it emerged in the lineage leading to *S. richteri* (Fig. 1*C*). It is worth noting that the internal branch leading to the *S. richteri SB* haplotypes and all *Sb* haplotypes, although very short, is strongly supported by the bootstrapping results (bootstrap value 99; Fig. 1*C*). Importantly, the *S. richteri SB* haplotype lineage forms a sister group to the *S. megergates*–*S. macdonaghi*–*S. invicta/AdRX SB* haplotypes lineage rather than the *S. invicta/AdRX SB* haplotype lineage (predicted from the species tree). The fact that the *In(16)2* tree disagrees substantially with the species tree suggests that *SB* haplotypes of *In(16)2* experienced a different evolutionary history than the other inversions as well as chromosomes 1-15, particularly with regard to its passage in *S. richteri*. Such a pattern could be the product of ancestral balancing selection occurring in the *S. richteri* lineage during consolidation of the supergene via emergence of the *In(16)2* inversion, which effectively locked all three inversions together.

To assess whether these results derived from the phylogenetic analyses might be biased because of the large variation in sample sizes among species and haplotypes in our dataset (range of 1 to 60 individuals per haplotype/species combination; *SI Appendix*, Table S1), we conducted rarefaction analyses by generating phylogenetic trees based on one randomly selected *SB* and *Sb* individual from each species as well as two randomly selected individuals from the two outgroup species *S. saevissima* and *S. metallica* (1,000 resampling iterations). The topology of these trees was highly congruent with the topology obtained from the complete dataset (*SI Appendix*, Fig. S4), indicating that our phylogenies are not biased by the uneven sample sizes.

The topology of the *SB* haplotypes was largely congruent with the species topology (Fig. 1*A*), indicating that the *SB* variant of chromosome 16 shares a similar evolutionary history with the remaining 15 chromosomes comprising the rest of the genome (*SI Appendix*, Fig. S5). In contrast, and similar to what was found previously for the entire supergene (22), the topologies of each of the three *Sb* inversions display different lineage relationships from that of the species phylogeny (Fig. 1 and *SI Appendix*, Fig. S6). Surprisingly, phylogenies from each of the three *Sb* inversion haplotypes were almost identical to one another (*SI Appendix*, Fig. S6). They featured two main lineages, one of which included all the *S. richteri Sb* haplotypes while the other comprised the *Sb* haplotypes from all the other species. The fact that the *Sb* phylogenies are very different from the species phylogeny does not support the conclusion that this

supergene originated in the common ancestor of the socially polymorphic species and traversed multiple speciation events to confer the social polymorphism across the group (18, 22). This raises the important issue of what evolutionary process can explain the current distribution of the *Sb* supergene in fire ants as well as the observed patterns of variation in the genomic data.

To further test the reliability of the phylogenetic conclusions, we used the *D* statistics (ABBA-BABA) method to independently quantify the relationships between the *SB* and *Sb* haplotypes from each species at each inversion. Specifically, we determined whether the haplotype of each *Sb* inversion in each species shares derived alleles more commonly with its conspecific *SB* haplotype or with a heterospecific *SB* haplotype (Fig. 2*A*). These analyses revealed that the haplotypes of *Sb* inversions *In(16)3* and *In(16)1* in *S. macdonaghi*, *S. megergates*, and *S. nr. interrupta* share significant excesses of derived alleles with the *S. invicta/AdRX* and *S. richteri SB* haplotypes relative to conspecific *SB* haplotypes (Fig. 2*B*; positive *D* values, *P* value < 0.001; *SI Appendix*, Table S2). Such excesses of derived alleles mean that *S. macdonaghi*, *S. megergates*, and *S. nr. interrupta Sb* haplotypes are evolutionarily closer to *S. invicta/AdRX* and *S. richteri SB* haplotypes than to conspecific *SB* haplotypes. This finding supports the view that these two inversions emerged in the ancestral *S. invicta/AdRX*–*S. richteri* lineage and that there has been only limited intraspecific recombination between *SB* and *Sb* haplotypes (Fig. 2*B*). Moreover, in line with the view that *In(16)2* emerged in the *S. richteri* lineage, our analyses further show that *Sb* inversion *In(16)2* haplotypes of all species share a greater number of derived alleles with *S. richteri SB* haplotypes than with conspecific *SB* haplotypes.

The phylogenetic analyses support a scenario in which the three inversions of the *Sb* supergene emerged in the following order: *In(16)3→In(16)1→In(16)2*. The *Sb* haplotypes of the *In(16)3* inversion form a branch basal to the *S. invicta/AdRX*–*S. richteri SB* haplotype lineage, suggesting its emergence before the separation of these three lineages. For the *In(16)1* inversion, we obtained two alternative topologies. The *Sb* haplotypes group with *S. invicta/AdRX SB* haplotypes in the first one and with *S. richteri SB* haplotypes in the second, suggesting its emergence during the period of divergence of these species. Finally, the *Sb* haplotypes of the *In(16)2* inversion form a sister clade to all *S. richteri SB* haplotypes, suggesting its relatively late emergence in *S. richteri*, after separation of this lineage from *S. invicta/AdRX*. This scenario of the order of emergence is also supported by the *D* statistics and Twisst analyses conducted for each inversion (Fig. 2 and *SI Appendix*, Figs. S7 and S8). Together, the analyses imply that the *Sb* supergene was completely assembled in its current form with three inversions in the *S. richteri* lineage and that it later introgressed into the other species as a fully functional, modular supergene. Two lines of evidence suggest that the complete supergene moved first from *S. richteri* to *S. invicta* and then to the other species. First, the separate phylogenetic analyses of each inversion show that *S. macdonaghi* and *S. nr. interrupta Sb* haplotypes are more similar to the *S. invicta/AdRX Sb* haplotypes than to the *S. richteri Sb* haplotypes (*SI Appendix*, Fig. S6), despite the fact that the former two species are equally related evolutionarily to all three of the latter species according to the species tree. Second, the *D* statistics reveal that, for each inversion, the *S. nr. interrupta* and *S. macdonaghi Sb* haplotypes share a greater number of derived alleles with the *S. invicta/AdRX Sb* haplotypes than with the *S. richteri Sb* haplotypes (*SI Appendix*, Fig. S9). Data from *S. megergates* are not sufficient to reach any conclusion about the origin of its *Sb* haplotype that is consistent between the phylogenetic and *D* statistics analyses.
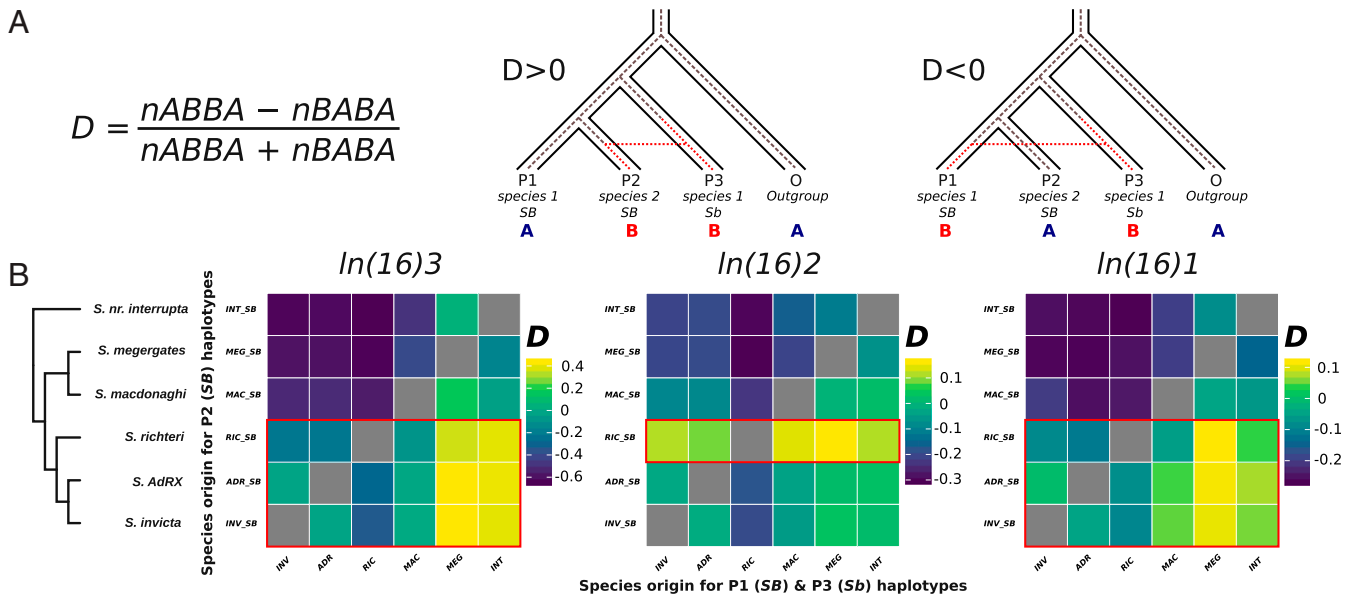
Fig. 2. *D* statistics analyses of the fire ant *SB* and *Sb* supergene haplotypes. (*A*) Schematic representation of the method. Only polymorphic sites with the ABBA and BABA combination of alleles in the focal species were analyzed. (*B*) Summary depiction of the *D* statistics for the *In(16)3* inversion (*Left*), *In(16)2* inversion (*Middle*), and *In(16)1* inversion (*Right*) for all combinations of *SB* and *Sb* haplotypes across species (species tree, *Far Left*). Increasingly negative (blue) *D* values indicate that *Sb* haplotypes share increasingly more of the derived alleles with *SB* haplotypes from the same species than with *SB* haplotypes of other species. Increasingly positive (yellow) values indicate increasingly more alleles are shared with the other species than with conspecifics. Red boxes highlight comparisons for which the highest *D* values are predicted for each inversion based on the phylogenetic results.

Genomewide Traces of Hybridization among the Socially Polymorphic Fire Ant Species. We assessed potential opportunities for hybridization and introgression between the study species by inferring their native ranges using both the sampling locations of the individuals used in this study (from ref. 26) and additional sample data from the Global Biodiversity Information Facility (GBIF) and AntWeb databases (22, 38) (*SI Appendix*, Tables S1, S3, and S4). Remarkably, the ranges of *S. macdonaghi*, *S. megergates*, and *S. AdRX* are completely encompassed by the range of *S. invicta* and partially overlap that of *S. richteri* (Fig. 3*A*) (see also refs. 26 and 39). Moreover, the distribution of *S. nr. interrupta* partially overlaps that of *S. macdonaghi*, *S. invicta/AdRX*, and *S. richteri*. Given that the socially polymorphic species exhibit largely overlapping native ranges, we next searched for traces of hybridization between species in both the mitochondrial and nuclear genomes. A phylogeny of the complete mitochondrial genomes revealed that most species are highly para- or polyphyletic for this maternally inherited, cytoplasmic organelle DNA (see also ref. 36). For instance, multiple mitochondrial sequences from *S. megergates* and *S. richteri* are embedded in a monophyletic clade dominated mostly by *S. invicta* sequences. This and multiple other examples of mitochondrial paraphyly with respect to the species designations support the conclusion that hybridization occurred at least occasionally between *S. invicta/AdRX* and the three species *S. macdonaghi*, *S. megergates*, and *S. richteri*, as well as between *S. richteri* and the pair of species *S. megergates* and *S. nr. interrupta* (*SI Appendix*, Fig. S10).

We next quantified phylogenetic heterogeneity using chromosomes 1 to 15 (supergene-bearing chromosome 16 not included), which can be caused by disturbing forces such as introgressive hybridization following lineage separation as well as ILS during speciation. We used a topology weighting method with 1,000-SNP (single-nucleotide polymorphism) sliding windows (40) (*SI Appendix*, Fig. S2). While the most frequently recovered tree has the identical topology as the species phylogeny (Fig. 1), it accounts for only 8% of the windows (Fig. 3*B*). Surprisingly, the

second most frequent alternative topology, in which *S. invicta/AdRX* is sister group to *S. macdonaghi/S. megergates*, is nearly as frequent as the species topology but occurs far less frequently with a weighting of 1 [i.e., maximal weighting for a single topology (41)] (Fig. 3*B* and *SI Appendix*, Figs. S1 and S2). Among other common topologies, *S. invicta/AdRX* is frequently sister group to *S. macdonaghi* or *S. richteri* to *S. nr. interrupta*. Overall, the topology weighting analyses revealed widespread phylogenetic discordance, suggesting potential hybridization between multiple lineages across the socially polymorphic fire ant clade. To establish that such phylogenetic discordance along chromosomes 1 to 15 results from introgression events rather than ILS, we measured the proportion of shared derived alleles between the different species, again using *D* statistics (42, 43). Significantly positive *D* values were obtained between *S. invicta/AdRX* and *S. macdonaghi* ($D = 0.18$, $P < 0.001$; Fig. 3*C* and *SI Appendix*, Table S5) and, to a lesser extent, between *S. invicta/AdRX* and *S. megergates* ($D = 0.03$, $P < 0.001$; Fig. 3*C* and *SI Appendix*, Table S5), strongly supporting the conclusion of historical gene flow between these species. Moreover, the *D* statistics results suggest possible gene flow between *S. nr. interrupta* and both *S. richteri* and *S. invicta/AdRX* (Fig. 3*C* and *SI Appendix*, Table S5).

In order to place the hybridization events on the species tree, we used *f* statistics to fit admixture graphs to the tree-based shared-alleles frequency data (42, 43). The best-fitting admixture graphs all suggest gene flow between the *S. macdonaghi* and *S. invicta/AdRX* lineages but differ on whether introgression is inferred to have occurred between *S. megergates* and *S. invicta/AdRX* or between *S. megergates* and *S. macdonaghi* (after hybridization with *S. invicta*). The admixture graphs also disagree on whether gene flow occurred between the *S. nr. interrupta* and *S. invicta/S. richteri* lineages or between the *S. nr. interrupta* and *S. invicta* lineages (Fig. 3*D* and *SI Appendix*, Fig. S11). Given this uncertainty, we used an approximate Bayesian computation (ABC) approach to identify the most likely evolutionary history by comparing explicit alternative models of speciation, with and without gene flow between the species. Our ABC
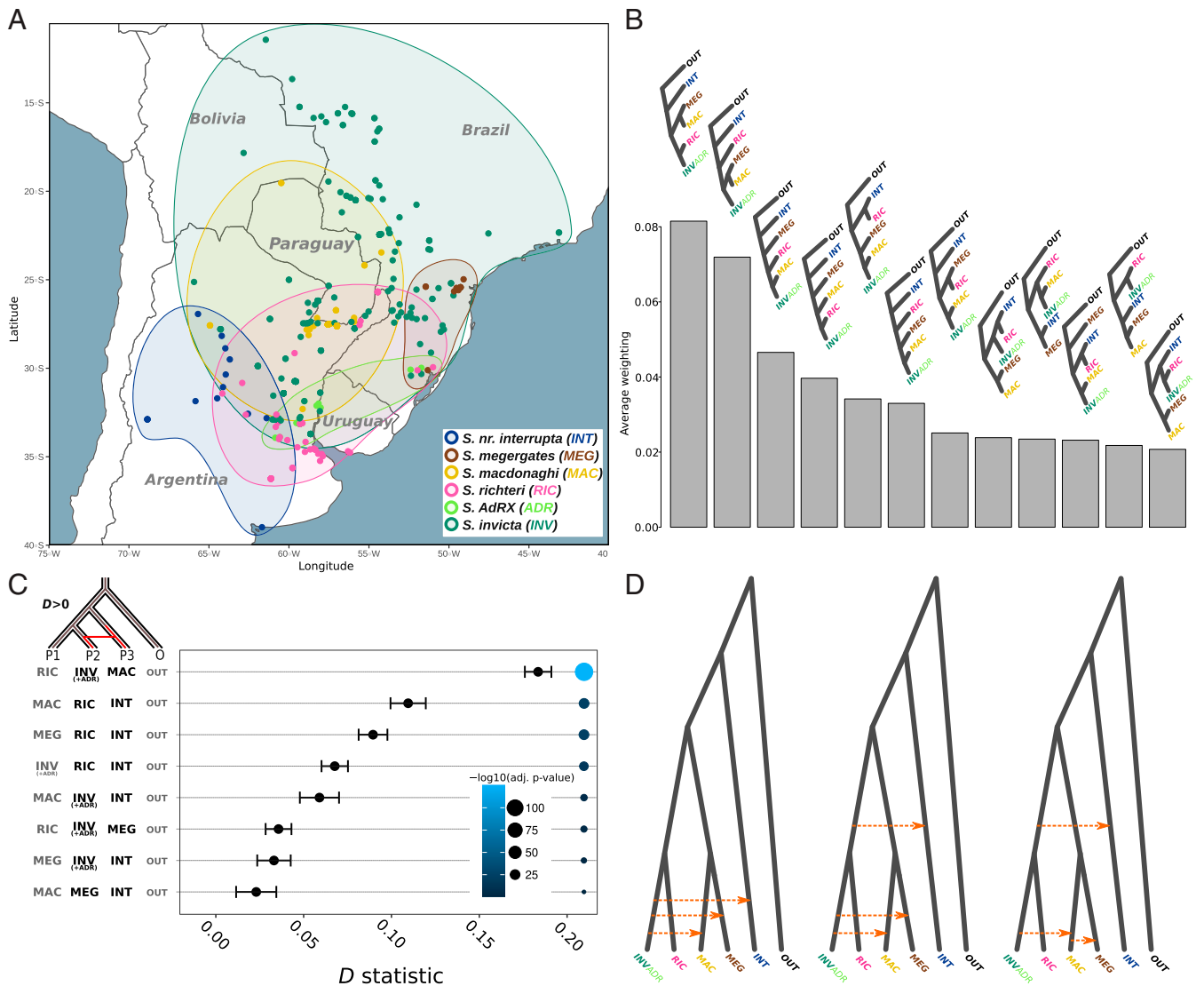
Fig. 3. Geographic distributions of the six socially polymorphic fire ant species and evidence of gene flow among them. (*A*) Estimated native species ranges. (*B*) Frequencies of the 12 most commonly recovered topologies inferred from topology weighting using windows of 1,000 SNPs in the nuclear genomes (chromosomes 1 to 15). (*C*) Heatmap summarizing the *D* statistics results for different combinations of socially polymorphic species using *S. metallica* and *S. saevissima* as outgroups (*S. AdRX* samples are included with *S. invicta*). Higher *D* values (warmer colors) indicate higher estimates of introgression (admixture). Error bars represent SD. (*D*) Admixture graphs representing the evolutionary history of the socially polymorphic species. The three graphs show all proposed admixture events (orange dashed lines) included in the six best-fitting admixture graphs (*SI Appendix,* Fig. S11).

analysis, using the *D* statistic and another 17 summary statistics (*Methods*), strongly supports the idea that gene flow occurred between *S. invicta* and several other species (*S. macdonaghi, S. megergates,* and *S. nr. interrupta*), in accord with one of the six best-fitting admixture graphs (*SI Appendix,* Fig. S11 and Table S6). Thus, four distinct but complementary analyses (mitochondrial DNA [mtDNA] phylogeny reconstruction, nuclear genome topology weighting, *D* statistics analyses, and ABC tests) jointly implicate some level of historical gene flow between *S. invicta/AdRX* and both *S. macdonaghi* and *S. megergates*. The analyses provide weaker support as well for historical introgression events involving several other pairs of lineages.

Discussion

We investigated the evolutionary history of the supergene that regulates colony social organization, a fundamental feature of social life in ants, in *S. invicta* and closely related species of South American fire ants. Based on the results from multiple complementary

approaches, we conclude that the *Sb* supergene underlying the social polymorphism was formed by sequential incorporation of three inversions. The two initial inversions, *In(16)3* and *In(16)1*, occurred in the ancestral population of *S. invicta/AdRX* and *S. richteri*, and the third, *In(16)2*, arose in the derived *S. richteri* lineage (Fig. 4). Once assembled in its current form, the supergene was first transferred from the *S. richteri* lineage to its sister *S. invicta/AdRX* lineage, presumably via hybridization, and from there to three other lineages (represented by modern-day *S. nr. interrupta, S. macdonaghi,* and *S. megergates*). This complex evolutionary history of the supergene, which spanned multiple lineages across the socially polymorphic fire ant clade and involved repeated bouts of hybridization, left telltale traces across the genomes of these ants such as excesses of shared derived alleles in remotely related species and numerous discordant genealogies at different genomic locations.

Our results concerning the origin of the supergene contrast with those from three previous studies. The first study suggested that the *Sb* haplotype arose in the most recent common
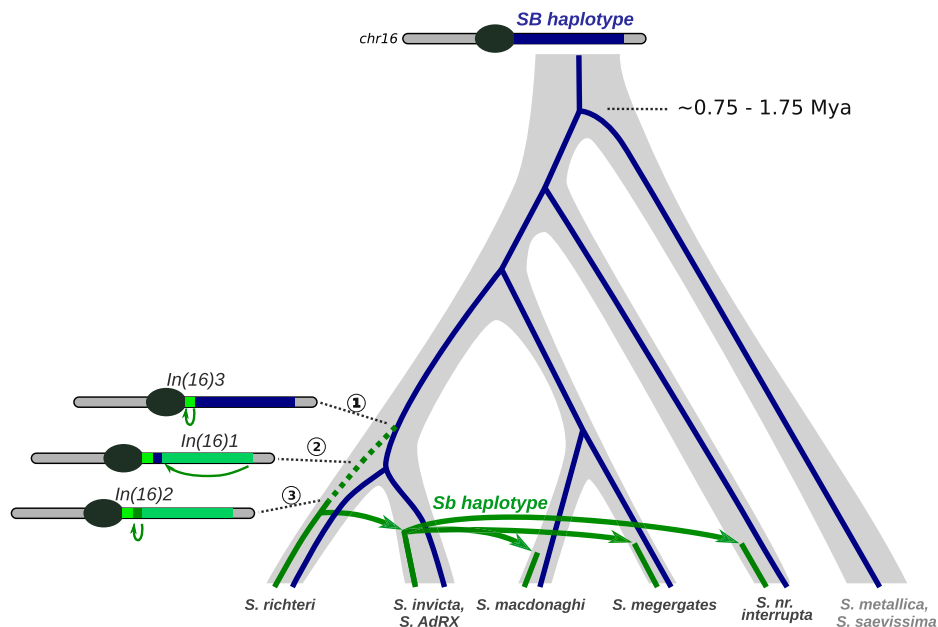
Fig. 4. Hypothesized evolutionary history of the *Sb* supergene haplotype. The green line represents the evolutionary history of *Sb*: dashed for the intermediate stages, and solid for the complete haplotype with the three inversions. Arrows illustrate introgression events between the socially polymorphic species inferred from *Sb* inversion phylogenies. The blue line represents the evolutionary history of the *SB* haplotype. The estimation of divergence time is from refs. 22 and 29.

ancestor of the socially polymorphic fire ant clade and spread across the radiating species via normal cladogenetic lineage splitting (22). The discrepancies in the conclusions of this study from ours partly stems from ref. 22 using a strong linkage disequilibrium (LD) filter that removed blocks of polymorphisms that represent highly informative phylogenetic signals for the supergene region. By relaxing the LD filtering parameter, ref. 22 obtained phylogenetic results for the supergene more in line with ours (ref. 22, extended data fig. 8). Two more recent studies suggested, instead, that the supergene evolved in the ancestor of one of the extant species and then spread to other species. Specifically, based on the analysis of *S. invicta* and *S. richteri*, ref. 32 concluded that the supergene probably arose in *S. richteri* and introgressed into *S. invicta*, while ref. 29 concluded, from an analysis including more species, that the supergene evolved in *S. invicta/S. macdonaghi* (note that the authors incorrectly labeled samples of *S. AdRX* as *S. macdonaghi* in this paper). A potential shortcoming of both studies is that they were based on sequences from a limited number of markers across the supergene and a highly fragmented genome assembly. Most importantly, these two studies, as well as the study in ref. 22, did not analyze the three inversions separately but instead as a single unique locus. This approach cannot produce a clear image of the supergene evolutionary history as it merges mixed signals from the different inversions. For example, the vast majority of the genes (87 out of 97) analyzed by ref. 29 are localized on *In(16)1* (*SI Appendix*, Table S7). Thus, this study effectively reconstructed the evolutionary history of the largest inversion [*In(16)1*], which our analyses indeed revealed to be a (moderately supported) sister group to the *SB* haplotypes of *S. AdRX* and *S. invicta*. Thus, it appears that while requiring more computational resources, the approach of leveraging the abundant genetic information contained in each of the three inversions is necessary to understand the distinct events that generated the supergene. Similarly, considering genetic information separately for each inversion has been crucial to understanding the complex evolutionary history of the *Heliconius* butterfly mimicry supergene (6).

The finding that all study individuals had either all three inversions or none at all, as has been found for all socially polymorphic fire ants appropriately sequenced (22, 25, 29, 44, 45), is in line with the view that many supergenes may originate by the sequential addition of beneficial "modifier" mutations when adding new inversions to the nonrecombining region such that the intermediate states of only one or two inversions are strongly selectively disadvantageous relative to the fully formed element. Nonetheless, multiple examples of other supergenes formed through the sequential accumulation of inversions exist in which intermediate states are fully adaptive alternate variants. The best-known example is the Batesian mimicry supergene identified in the butterfly *Heliconius numata*, in which various combinations of the three inversions yield different wing-color patterns (6, 46). On the other hand, gradually formed supergene systems in which intermediate states are absent in natural populations are common, such as the supergene controlling mate choice in the seaweed fly, *Coelopa frigida*, which comprises three overlapping inversions (47). The supergene in the white-throated sparrow, *Zonotrichia albicollis*, regulates alternative mating syndromes and is composed of at least two nested pericentric inversions (48), while the social supergene in the socially polymorphic ant *Formica selysi* is formed by at least four successive inversions (49). Large sequential structural change in a chromosome is likely an efficient process giving rise to such complex phenotypes, because it allows stepwise optimization of the phenotype by "freezing" in place large numbers of favorable allelic combinations that contribute cumulatively to the novel alternate complex phenotype. It does, however, raise a long-standing question—how do the intermediate stages in the evolution of complex structures, which presumably are not highly fit because they display some sort of intermediate phenotype, persist sufficiently long in enough individuals to allow the appearance of the requisite subsequent mutations for expression of the fully formed alternate phenotype? Studies such as this one that define discrete intermediate stages in the acquisition of the necessary components of such complex trait regulators can play pivotal roles in addressing this fundamental evolutionary

question once gene inventories and functional annotations for each inversion are available.

Recent studies showed that the clade of socially polymorphic South American fire ants diversified very rapidly over the last 0.1 to 2.1 million y (22, 32). Such rapid species radiation is frequently associated with substantial gene flow between incipient independent lineages because adjacent or overlapping geographical distributions generate the opportunities for interbreeding, and genomic divergence causing incompatibilities and hybrid breakdown is not yet sufficiently developed [e.g., wild and domestic cattle (50), *Anopheles* mosquitoes (51, 52), *Panthera* cats (53), cichlid fishes (54, 55), and *Heliconius* butterflies (6, 56–58)]. Our analyses indeed revealed evidence of gene flow between several of the studied species, as suggested also by earlier research (35, 36). However, the extent of hybridization apparently has been quite limited between some species, in particular between *S. richteri* and *S. invicta/AdRX* (see also refs. 35, 37, and 59) as well as *S. invicta/AdRX* and *S. nr. interrupta*. A possible reason for only limited traces of hybridization between these species is that introgression of the *Sb* haplotype occurred via an intermediate species, such as *S. quinquecuspis*, a species that has experienced extensive historical introgression of heterospecific nuclear DNA and mtDNA from both *S. invicta* and *S. richteri* in some parts of its range (35, 36). Unfortunately, this hypothesis could not be tested because we do not have access to sequenced individuals from relevant populations of this species.

Alternatively, a possible explanation for the transfer of the supergene between these species that left limited traces along the remainder of the nuclear genome is simply that it was subjected to strong positive selection after the transfer while the remainder of the genome had become increasingly resistant to introgression because of growing genomic incompatibilities (52, 60, 61). Such adaptive introgression in the face of well-developed species barriers has been suggested to be common for supergenes because they comprise combinations of coadapted genes that may readily provide benefits across multiple closely related species (2, 62). In the case of fire ants, the benefits are probably related to the ecological advantages that polygyny provides in massively disturbed, harsh habitats and/or highly competitive environments (63, 64). Such advantages include insurance against loss of colony queens and bypassing of the risky independent colony-founding phase of the life cycle characterizing the monogyne form. Moreover, it is also possible that the selfish effects of the *Sb* haplotype, which have been described in introduced populations of *S. invicta*, also manifest themselves in the related species and thus promote the spread of the *Sb* haplotype. The selective elimination of *SB/SB* queens by workers bearing the *Sb* haplotype (23, 24), as well as a bias toward sexualization of the female brood bearing it (65), results in increased reproductive success of *SB/Sb* queens and thus in the number of copies of the *Sb* haplotype transmitted each generation, similar to the effects of meiotic drive elements that invade a population by distorting the expected outcomes of meiosis (66). Such selective advantages to the *Sb* haplotype at the level of the colony or of the gene thus may explain the presumed rapid expansion of this haplotype despite hybridization being limited.

A supergene regulating social organization evolved independently in the ant *F. selysi*, a species in a lineage that diverged from the lineage giving rise to *Solenopsis* about 110 million y ago. The inversions in the two genera are fundamentally distinct, with few genes in common and different chromosomal locations (17). Phylogenetic and population genomic analyses revealed that homologous haplotypes of the *F. selysi* supergene (*Sp* haplotypes) also regulate social organization in four other *Formica* species (49). In contrast to the *Solenopsis* supergene, the *Formica* supergene apparently did not spread between species via introgression but appeared in a common ancestor of these five socially polymorphic species 20 to 40 million y ago. Importantly, the *Formica* supergene also exhibits selfish properties, with the *Sp* haplotype producing an embryo-killing maternal effect. As a result, *Sp/Sm* queens never produce *Sm/Sm* daughters (67). This dynamic is very similar to the selective advantage provided by the selfish *Sb* haplotype in *S. invicta* and may also help account for the presence of this supergene in several *Formica* species.

Our conclusion that the *Sb* haplotype has introgressed across species adds to the increasing number of such instances documented. In *Heliconius* butterflies, introgression of an inversion triggered the formation of the *P* supergene (6). Similarly, the $2/2^m$ supergene controlling disassortative mating between male morphs in the white-throated sparrow may originate from a past hybridization event with other *Zonotrichia* species (13). The *S* locus supergene in *Arabidopsis*, which regulates production of two alternative floral morphs (pin and thrum), also shows traces of repeated introgressions between two closely related species (68).

In summary, this study strongly supports a scenario in which the fire ant *Sb* supergene was formed rapidly by sequential recruitment of three inversions, with the complete element subsequently introgressing into other closely related species despite limited evidence of hybridization between some of them. Following introgression, the same supergene has been maintained in each of the species of the socially polymorphic South American fire ant clade, showing that there are important selective forces associated with polygyny (balancing selection and/or selfish genetic effects) at work in these ants.

Methods

Sequence Data. Collection, DNA extraction, and sequencing methods of the 185 studied haploid fire ant males are described in ref. 22 (*SI Appendix*, Table S1; National Center for Biotechnology Information [NCBI] Sequence Read Archive [SRA] BioProjects PRJNA421367 and PRJNA821075).

Mapping, Calling, and Filtering. We mapped the Illumina paired-end whole-genome sequences of the 185 individuals on the *S. invicta SB* reference genome GCA_009650705 (22) using bwa-mem (v.0.7.17) (69). We then used SAMtools to manipulate, convert, and sort output files from bwa-mem (70). We added sample names and read-group tags to BAM alignments using bamaddrg. The SNP calling was done using FreeBayes (v.1.3.2) with -ploidy 1 (71) to identify sequence polymorphisms. We processed all individual BAMs simultaneously using -bam-list; we used a parallel approach, with each thread analyzing a distinct 150-kb region. Locus VCF files were then concatenated using BCFtools concat (v.1.10.2) and haplotypes were disassembled with vt (v.0.5772) (70, 72) to dissociate indels from multiple-nucleotide variants. Finally, using VCFtools (v.0.1.16), we removed indels and sites with a phred-scaled quality score <30 or missing individuals >0.5 (73).

Phylogeography. We obtained the sample locations of the male specimens whose sequences were analyzed in this study from ref. 26 and other specimens from the GBIF and AntWeb databases [AntWeb version 8.41, accessed 24 July 2020 (38)]. We collected the geographic coordinates of the samples and mapped the minimal inferred ranges of the study species using the geom_sf function from the R ggplot2 package (v.3.3.2; R Core Team, 2020).

Whole-Genome Phylogenetic Analyses. Because phylogenetic analyses conducted on SNP matrices (i.e., without invariant sites and singletons) are not optimal for estimating branch lengths and topologies (74), we conducted the phylogenetic analyses using complete sequences, excluding only indels. We

reconstituted the DNA multisequence alignments including monomorphic sites from the VCF file using alignment-from-vcf.py (https://github.com/qhelleu/Social-Chrom_Phylo). All maximum-likelihood (ML) analyses were conducted with IQ-TREE (versions 1.7.8 and 2.0.4), with branch support calculated from 1,000 iterations of the ultrafast bootstrap algorithm (75–78). We used multiple strategies to reconstruct the species tree: First, we conducted phylogenetic analyses on a concatenated alignment using IQ-TREE and the GTR+I+F+R3 and GTR+F+R3 substitution models, as well as the substitution model that had the smallest Bayes information criterion (BIC) score using ModelFinder (79). To take into account heterogeneity in rates of evolution across the genome, we partitioned the data into nonoverlapping windows of 100 kb (80). To account for stationary, reversible, and homogeneous (SRH) model violations, we used the symmetric test implemented in IQ-TREE to identify and remove partitions that diverge from the SRH assumptions (81). We reran the phylogenetic analysis using an edge-proportional partition model, using GTR+I+F+R3, GTR+F+R3, or the substitution model that had the smallest BIC score from ModelFinder. To confirm that the tree obtained was not biased by introgression between the species/populations analyzed, we randomly sampled 25% of the partitions throughout the genome, using the AMAS replicate tool (82). To assess the robustness of the identified species tree, we applied a resampling approach using nonoverlapping windows of 10, 20, 50, 80, and 200 kb. We analyzed each dataset with IQ-TREE and the parameters described above. To account for ILS along the genome, we also analyzed the multiple trees for each partitioned dataset with ASTRAL (v.5.7.1), using DiscoVista to plot the results (83–86). Tree visualization and manipulation were done in iTOL (87). We performed topology weighting along the genome using the Twisst method (40) on 50-, 100-, 200-, 500-, and 1,000-SNP sliding windows to explore how the inferred relationships of the focal species vary across the genome. Individual-locus trees were built using PhyML (v.3.3) with the neighbor-joining method (88).

Supergene Phylogenetic Analyses. We conducted supergene phylogenetic analyses on the complete sequences of the three inversions, again including invariant sites and excluding indels. We analyzed each inversion separately using IQ-TREE (versions 1.7.8 and 2.0.4), with the GTR+I+F+R3 and GTR+F+R3 substitution models, as well as the substitution model that had the smallest BIC score calculated using ModelFinder (76, 78, 79). We also used a partitioning approach to identify the most frequent alternative topologies to the ML tree. Individual-locus trees were built using windows of 1,000 SNPs using PhyML with the neighbor-joining method (88). Then, we estimated the frequency of alternative species tree topologies using the Twisst method (40) on 50-, 100-, 200-, 500-, and 1,000-SNP sliding windows. Tree topology exploration from the Twisst analyses, including identification of sister clades, was conducted using the R packages ape and phytools (89, 90). In parallel, to take into account the variable number of SB and Sb haplotypes in each species, we applied a rarefaction procedure involving randomized selection of individual sequences: One sequence of each haplotype (SB or Sb) was randomly selected for each socially polymorphic species, and two sequences were selected for each outgroup species. We repeated this randomization procedure 1,000 times for each inversion. For each group of random sequences, we analyzed the alignment using IQ-TREE with the GTR+I+F+R3 substitution model and 1,000 iterations of the ultrafast bootstrap algorithm (75, 77, 78).

Inferences of SB/Sb Haplotype Relationships Using D Statistics. We used D statistics (ABBA-BABA method) to quantify for each species and for each inversion the number of derived alleles shared between Sb and either conspecific or heterospecific SB haplotypes. In a hypothetical four-taxon phylogeny, an outgroup (O) is utilized to estimate numbers of shared derived alleles between one sequence from the focal lineage (P3) and a second sequence from either the same (P1) or a different (P2) lineage. We used AdmixTools 6.0 in the wrapper admixr (43, 91) to calculate the D statistics. We compared only four individuals at a time to measure the number of derived alleles shared between the SB and Sb haplotypes: two conspecific individuals from a socially polymorphic species carrying either the SB or Sb haplotype (P1 and P3), a third individual carrying an SB haplotype from a different socially polymorphic species (P2), and a single outgroup individual from one of the two species lacking the Sb supergene ($S. saevissima$ or $S. metallica$). To avoid biases due to the different numbers of haplotypes sequenced among species, we first randomly selected three different

species (two SB/Sb–carrying and one outgroup species), and then randomly selected individuals within each species for each run, using the python3 random module. We ran 5,000 iterations for each inversion. The significance of D statistics was assessed using the pairwise Dunn test with a Benjamini–Yekutieli correction (92).

Genomewide Tests for Introgression. We further used D statistics to help distinguish introgression from ILS as the primary cause of incongruence in inferred trees. We compared all triplet combinations of species using Admix-Tools v.6.0 in the wrapper admixr (43, 91), with $S. metallica$ and $S. saevissima$ individuals comprising the outgroup. With ILS only, the focal species (P3) should share an equal number of derived alleles ($D = 0$) with both compared species (P1 and P2), while an excess of shared derived alleles (D value significantly different from 0) is indicative of past gene flow between two species and is not expected under a scenario of ILS without gene flow. We kept only triplets for comparisons in which the calculated D statistic fell within the interval [0, 1] and the species triplet order was concordant with the species phylogeny, meaning that P1 and P2 were sister clades with respect to P3 for all comparisons. Among the remaining triplets, we retained only comparisons involving the same pair of species in P2 and P3 with the highest D value. The significance of D statistics was assessed using a jackknife procedure (42) on blocks of 100 SNPs. The function p.adjust in R v.3.6.3 (R Core Team, 2020) was used to apply a Benjamini–Yekutieli correction (92). Results were plotted using the Ruby script "plot_d.rb" from https://github.com/mmatschiner/tutorials/. To build the admixture graphs for the socially polymorphic fire ant species, we used qpgraph from AdmixTools v.6.0 through the wrappers qpBrute and admixture-graph (43, 93–95).

Mitochondrial Genome De Novo Assembly and Phylogeny Reconstruction. We used the seed extend–based assembler NOVOPlasty (v.4.3) tool (96) to assemble the mitochondrial genome of each individual, excluding potential sequences transposed into the nuclear DNA. The seed is iteratively extended bidirectionally until the assembly circularizes in the expected size range (97). Then, we aligned the circularized sequences using MARS and MAFFT (v.7.475) (98, 99). Finally, we used IQ-TREE (v.2.0.4) to infer the ML phylogenetic tree of all aligned sequences with 1,000 ultrafast bootstrap iterations and the best-fitting substitution model from ModelFinder (78, 79). Tree visualization and manipulation were done in iTOL (87). We plotted the nuclear and mitochondrial phylogenetic trees face to face with links using the cophyloplot function in the ape R package [v.5.5 (89)].

Approximate Bayesian Computation Analyses with Four Populations. We used an ABC framework with a supervised machine-learning model classification procedure (random forest) to identify possible secondary contacts and interbreeding between the nascent socially polymorphic species or ancestral populations of their lineage. In all simulations, we used the coalescent simulator msnsam (October 2007 version), a modified version of the ms program, allowing variation in sample sizes among loci under an infinite-sites mutation model (100). We ran simulations for a large set of genomic models–combinations of strict isolation and secondary contacts between the four species considered in each simulation–and set of demographic models–heterogeneous or homogeneous effective population size and/or migration rate–to examine the effects of selection against migrants. We produced a simulated dataset of 108 average values from 18 summary statistics and their SDs computed over loci for each four-species combination of tested genomic and demographic models [https://github.com/popgenomics/ABC_4pop (101)]. To limit computation time and avoid overfitting, we used a subsample of 1,000 randomly sampled 50-kb bins from chromosomes 1 to 15 for each of the 10 replicates. We then used the simulations to train a random forest, using the $abcrf$ R package (102). The trained forest was next used to predict the model that best explains the observed dataset. We built the random forest in $abcrf$ from 1,000 decision trees using nonnull summary statistics, and then we ran the prediction also with 1,000 decision trees, to estimate the posterior probability of each model. Each prediction was scored using the number of decision trees supporting each model and the associated posterior probability for the best one. We also implemented a post hoc test to quantify the strength (robustness) of the prediction, which judges the ability of a trained random forest to discriminate between the results from two sets of simulations. To do this, we used

simulations from the best-identified model corresponding to an observed dataset, combined with the simulation output from one alternative model. First, we used 15% of the simulated data from both models to train a random forest. Next, we used the random forest to predict the models with which the 85% remaining simulated data showed the highest concordance. We then measured the proportion of simulations accurately identified. If more than 10% of the simulations from the best-identified model of the two were wrongly predicted, we considered the results from the ABC analysis as being ambiguous (101).

1. D. Charlesworth, B. Charlesworth, Theoretical genetics of Batesian mimicry II. Evolution of supergenes. *J. Theor. Biol.* **55**, 305–324 (1975).
2. T. Schwander, R. Libbrecht, L. Keller, Supergenes and complex phenotypes. *Curr. Biol.* **24**, R288–R294 (2014).
3. M. J. Thompson, C. D. Jiggins, Supergenes and their role in evolution. *Heredity* **113**, 1–8 (2014).
4. K. Kunte *et al.*, doublesex is a mimicry supergene. *Nature* **507**, 229–232 (2014).
5. M. Joron *et al.*, Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–206 (2011).
6. P. Jay *et al.*, Supergene evolution triggered by the introgression of a chromosomal inversion. *Curr. Biol.* **28**, 1839–1845.e3 (2018).
7. W. Zhang, E. Westerman, E. Nitzany, S. Palmer, M. R. Kronforst, Tracing the origin and evolution of supergene mimicry in butterflies. *Nat. Commun.* **8**, 1269 (2017).
8. J. D. Labonne, F. Tamari, J. S. Shore, Characterization of X-ray-generated floral mutants carrying deletions at the S-locus of distylous *Turnera subulata*. *Heredity* **105**, 235–243 (2010).
9. E. Durand *et al.*, Dominance hierarchy arising from the evolution of a complex small RNA regulatory network. *Science* **346**, 1200–1205 (2014).
10. J. Li *et al.*, Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. *Nat. Plants* **2**, 16188 (2016).
11. C. N. Huu, B. Keller, E. Conti, C. Kappel, M. Lenhard, Supergene evolution via stepwise duplications and neofunctionalization of a floral-organ identity gene. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23148–23157 (2020).
12. A. M. Larracuente, D. C. Presgraves, The selfish segregation distorter gene complex of *Drosophila melanogaster*. *Genetics* **192**, 33–53 (2012).
13. E. M. Tuttle *et al.*, Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* **26**, 344–350 (2016).
14. K. G. Ross, Multilocus evolution in fire ants: Effects of selection, gene flow and recombination. *Genetics* **145**, 961–974 (1997).
15. K. G. Ross, L. Keller, Genetic control of social organization in an ant. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14232–14237 (1998).
16. J. Wang *et al.*, A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664–668 (2013).
17. J. Purcell, A. Brelsford, Y. Wurm, N. Perrin, M. Chapuisat, Convergent genetic architecture underlies social organization in ants. *Curr. Biol.* **24**, 2728–2732 (2014).
18. D. Gotzek, K. G. Ross, Genetic regulation of colony social organization in fire ants: An integrative overview. *Q. Rev. Biol.* **82**, 201–226 (2007).
19. C. J. DeHeer, M. A. Goodisman, K. G. Ross, Queen dispersal strategies in the multiple-queen form of the fire ant *Solenopsis invicta*. *Am. Nat.* **153**, 660–675 (1999).
20. C. J. DeHeer, A comparison of the colony-founding potential of queens from single- and multiple-queen colonies of the fire ant *Solenopsis invicta*. *Anim. Behav.* **64**, 655–661 (2002).
21. K. G. Ross, L. Keller, Ecology and evolution of social organization: Insights from fire ants and other highly eusocial insects. *Annu. Rev. Ecol. Syst.* **26**, 631–656 (1995).
22. Z. Yan *et al.*, Evolution of a supergene that regulates a trans-species social polymorphism. *Nat. Ecol. Evol.* **4**, 240–249 (2020).
23. L. Keller, K. G. Ross, Selfish genes: A green beard in the red fire ant. *Nature* **394**, 573–575 (1998).
24. W. Trible, K. G. Ross, Chemical communication of queen supergene status in an ant. *J. Evol. Biol.* **29**, 502–513 (2016).
25. E. Stolle *et al.*, Degenerative expansion of a young supergene. *Mol. Biol. Evol.* **36**, 553–561 (2019).
26. J. P. Pitts, G. P. Camacho, D. Gotzek, J. V. Mchugh, K. G. Ross, Revision of the fire ants of the *Solenopsis saevissima* species-group (Hymenoptera: Formicidae). *Proc. Entomol. Soc. Wash.* **120**, 308–411 (2018).
27. K. G. Ross, D. Shoemaker, Unexpected patterns of segregation distortion at a selfish supergene in the fire ant *Solenopsis invicta*. *BMC Genet.* **19**, 101 (2018).
28. Y. C. Huang, V. D. Dang, N. C. Chang, J. Wang, Multiple large inversions and breakpoint rewiring of gene expression in the evolution of the fire ant social supergene. *Proc. R. Soc. Lond. B Biol. Sci.* **285**, 20180221 (2018).
29. E. Stolle *et al.*, Recurring adaptive introgression of a supergene variant that determines social organization. *Nat. Commun.* **13**, 1180 (2022).
30. M. J. Krieger, K. G. Ross, Molecular evolutionary analyses of the odorant-binding protein gene Gp-9 in fire ants and other *Solenopsis* species. *Mol. Biol. Evol.* **22**, 2090–2103 (2005).
31. D. Gotzek, D. D. Shoemaker, K. G. Ross, Molecular variation at a candidate gene implicated in the regulation of fire ant social behavior. *PLoS One* **2**, e1088 (2007).
32. P. Cohen, E. Privman, The social supergene dates back to the speciation time of two *Solenopsis* fire ant species. *Sci. Rep.* **10**, 11538 (2020).
33. K. G. Ross, R. K. V. Meer, D. J. C. Fletcher, E. L. Vargo, Biochemical phenotypic and genetic studies of two introduced fire ants and their hybrid (Hymenoptera: Formicidae). *Evolution* **41**, 280–293 (1987).
34. K. G. Ross, J. L. Robertson, Developmental stability, heterozygosity, and fitness in two introduced fire ants (*Solenopsis invicta* and *S. richteri*) and their hybrid. *Heredity* **64**, 93–103 (1990).
35. K. G. Ross, D. D. Shoemaker, Species delimitation in native South American fire ants. *Mol. Ecol.* **14**, 3419–3438 (2005).
36. D. D. Shoemaker, M. E. Ahrens, K. G. Ross, Molecular phylogeny of fire ants of the *Solenopsis saevissima* species-group based on mtDNA sequences. *Mol. Phylogenet. Evol.* **38**, 200–215 (2006).
37. P. Cohen, E. Privman, Speciation and hybridization in invasive fire ants. *BMC Evol. Biol.* **19**, 111 (2019).
38. M. A. Lane, J. L. Edwards, "The Global Biodiversity Information Facility (GBIF)" in *Biodiversity Databases: Techniques, Politics, and Applications*, G. B. Curry, C. J. Humphries, Eds. (Systematics Association Special Volume Series, CRC Press, 2007), pp. 1–4.
39. W. R. Tschinkel, *The Fire Ants* (Belknap Press, 2006).
40. S. H. Martin, S. M. Van Belleghem, Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).
41. S. H. Martin, J. W. Davey, C. Salazar, C. D. Jiggins, Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol.* **17**, e2006288 (2019).
42. R. E. Green *et al.*, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
43. N. Patterson *et al.*, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
44. E. Privman *et al.*, Positive selection on sociobiological traits in invasive fire ants. *Mol. Ecol.* **27**, 3116–3130 (2018).
45. C. Martinez-Ruiz *et al.*, Genomic architecture and evolutionary antagonism drive allelic expression bias in the social supergene of red fire ants. *eLife* **9**, e55862 (2020).
46. Y. Le Poul *et al.*, Evolution of dominance mechanisms at a butterfly mimicry supergene. *Nat. Commun.* **5**, 5644 (2014).
47. S. Enge *et al.*, A supergene in seaweed flies modulates male traits and female perception. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.06.30.450538. Accessed 12 January 2022.
48. J. W. Thomas *et al.*, The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. *Genetics* **179**, 1455–1468 (2008).
49. A. Brelsford *et al.*, An ancient and eroded social supergene is widespread across *Formica* ants. *Curr. Biol.* **30**, 304–311.e4 (2020).
50. D. D. Wu *et al.*, Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat. Ecol. Evol.* **2**, 1139–1145 (2018).
51. M. C. Fontaine *et al.*, Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **347**, 1258524 (2015).
52. S. T. Small *et al.*, Radiation with reticulation marks the origin of a major malaria vector. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 31583–31590 (2020).
53. H. V. Figueiró *et al.*, Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Sci. Adv.* **3**, e1700299 (2017).
54. B. S. Meyer, M. Matschiner, W. Salzburger, Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. *Syst. Biol.* **66**, 531–550 (2017).
55. M. Malinsky *et al.*, Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* **2**, 1940–1955 (2018).
56. K. K. Dasmahapatra *et al.*; Heliconius Genome Consortium, Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012).
57. N. B. Edelman *et al.*, Genomic architecture and introgression shape a butterfly radiation. *Science* **366**, 594–599 (2019).
58. K. M. Kozak, M. Joron, W. O. McMillan, C. D. Jiggins, Rampant genome-wide admixture across the *Heliconius* radiation. *Genome Biol. Evol.* **13**, evab099 (2021).
59. K. G. Ross, J. C. Trager, Systematics and population genetics of fire ants (*Solenopsis saevissima* complex) from Argentina. *Evolution* **44**, 2113–2134 (1990).
60. A. Suarez-Gonzalez, C. Lexer, Q. C. B. Cronk, Adaptive introgression: A plant perspective. *Biol. Lett.* **14**, 20170688 (2018).
61. G. Dixon, J. Kitano, M. Kirkpatrick, The origin of a new sex chromosome by introgression between two stickleback fishes. *Mol. Biol. Evol.* **36**, 28–38 (2019).
62. M. Kirkpatrick, N. Barton, Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).
63. J. T. Vogt, M. L. Allen, B. Wallet, D. Boykin, W. A. Smith, Distribution patterns of imported fire ants (Hymenoptera: Formicidae) on a sheep and goat farm in Oklahoma. *Environ. Entomol.* **38**, 551–560 (2009).
64. M. Milks, J. Fuxa, A. Richter, E. Moser, Multivariate analyses of the factors affecting the distribution, abundance and social form of Louisiana fire ants, *Solenopsis invicta*. *Insectes Soc.* **54**, 283–292 (2007).
65. S. D. Buechel, Y. Wurm, L. Keller, Social chromosome variants differentially affect queen determination and the survival of workers in the fire ant *Solenopsis invicta*. *Mol. Ecol.* **23**, 5117–5127 (2014).
66. Y. C. Huang, J. Wang, Did the fire ant supergene evolve selfishly or socially? *BioEssays* **36**, 200–208 (2014).
67. A. Avril, J. Purcell, S. Béniguel, M. Chapuisat, Maternal effect killing by a supergene controlling ant social organization. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17130–17134 (2020).
68. V. Castric, J. Bechsgaard, M. H. Schierup, X. Vekemans, Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genet.* **4**, e1000168 (2008).
69. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
70. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
71. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing. arXiv [Preprint] (2012). https://arxiv.org/abs/1207.3907. Accessed 8 October 2021.
72. A. Tan, G. R. Abecasis, H. M. Kang, Unified representation of genetic variants. *Bioinformatics* **31**, 2202–2204 (2015).

73. P. Danecek *et al.*; 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
74. A. D. Leaché, B. L. Banbury, J. Felsenstein, A. N. M. de Oca, A. Stamatakis, Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. *Syst. Biol.* **64**, 1032–1047 (2015).
75. B. Q. Minh, M. A. T. Nguyen, A. von Haeseler, Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
76. L. T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
77. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, Ufboot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
78. B. Q. Minh *et al.*, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
79. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
80. J. Soubrier *et al.*, The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol. Biol. Evol.* **29**, 3345–3358 (2012).
81. S. Naser-Khdour, B. Q. Minh, W. Zhang, E. A. Stone, R. Lanfear, The prevalence and impact of model violations in phylogenetic analysis. *Genome Biol. Evol.* **11**, 3341–3352 (2019).
82. M. L. Borowiec, AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4**, e1660 (2016).
83. S. Mirarab *et al.*, ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–i548 (2014).
84. C. Zhang, M. Rabiee, E. Sayyari, S. Mirarab, ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19** (suppl. 6), 153 (2018).
85. M. Rabiee, E. Sayyari, S. Mirarab, Multi-allele species reconstruction using ASTRAL. *Mol. Phylogenet. Evol.* **130**, 286–296 (2019).
86. E. Sayyari, J. B. Whitfield, S. Mirarab, DiscoVista: Interpretable visualizations of gene tree discordance. *Mol. Phylogenet. Evol.* **122**, 110–115 (2018).
87. I. Letunic, P. Bork, Interactive tree of life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
88. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
89. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
90. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **2**, 217–223 (2012).
91. M. Petr, B. Vernot, J. Kelso, admixr–R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics* **35**, 3194–3195 (2019).
92. Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
93. K. Leppälä, S. V. Nielsen, T. Mailund, admixturegraph: An R package for admixture graph manipulation and fitting. *Bioinformatics* **33**, 1738–1740 (2017).
94. M. Ní Leathlobhair *et al.*, The evolutionary history of dogs in the Americas. *Science* **361**, 81–85 (2018).
95. L. Liu *et al.*, Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. *Nat. Commun.* **10**, 1992 (2019).
96. N. Dierckxsens, P. Mardulyn, G. Smits, NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **45**, e18 (2017).
97. D. Gotzek, J. Clarke, D. Shoemaker, Mitochondrial genome evolution in fire ants (Hymenoptera: Formicidae). *BMC Evol. Biol.* **10**, 300 (2010).
98. L. A. Ayad, S. P. Pissis, MARS: Improving multiple circular sequence alignment using refined sequences. *BMC Genomics* **18**, 86 (2017).
99. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
100. R. R. Hudson, Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002).
101. C. Fraïsse *et al.*, DILS: Demographic inferences with linked selection by using ABC. *Mol. Ecol. Resour.* **21**, 2629–2644 (2021).
102. P. Pudlo *et al.*, Reliable ABC model choice via random forests. *Bioinformatics* **32**, 859–866 (2016).