doi:10.1038/nature11832

Materials and Methods

Ant collections and general analysis tools

S. invicta monogyne and polygyne colonies were collected from the field and reared in the laboratory under standard conditions[1]. All individuals were genotyped at *Gp-9* using an established RFLP protocol[2]. Analyses were performed at the Swiss Institute of Bioinformatics Vital-IT center for High Performance Computing and the EPSRC-funded MidPlus HPC centre. In addition to the pre-existing tools described below, custom Ruby scripts[3] and the bioruby library[4] were used to transform input and output files, and R[5] scripts utilizing the ggplot2[6] and genomicRanges[7] packages were used for data analysis and visualization.

RAD sequencing and analysis

We used one monogyne and four polygyne colonies collected near Athens, Georgia (USA) and two monogyne colonies collected near Gainesville, Florida (USA) for genetic mapping. We followed the RADseq protocol described in Baird *et al*.[8]. In brief, genomic DNA was extracted from single adult males using the Qiagen DNeasy kit or using the Agencourt DNadvance kit and then 0.1 – 1.0 ug of the DNA digested with EcoRI (New England Biolabs [NEB]). Each digested sample was purified with the Qiaquick purification kit (Qiagen) and ligated to one of 64 barcoded P1 adapters for >1 h (T4 DNA Ligase, NEB or Fermentas; Supplementary Table 15). Ligated samples were heat inactivated for 20 min at 65^{o}C, pooled (16-24 samples), and randomly sheared to an average size of 500 bp (Bioruptor or Covaris sonicator). Sheared DNA was purified and concentrated using the MinElute DNA purification kit (Qiagen) and then run on an agarose gel. DNA ranging from 300-700 bp was size selected by gel purification using the MinElute Gel Extraction Kit with special care to avoid unligated or concatemerized P1 adapters. DNA ends were polished using the Quick Blunting Kit (NEB) and purified with the QIAquick kit. Subsequently, 3-prime dA overhangs were created by incubating dATP with Klenow (3' exo-, NEB) for 30 min at 37^{o}C and then re-purified (QIAquick purification kit). P2 adapters were ligated to these samples (T4 DNA ligase, NEB or Fermentas, Supplementary Table 15) for >60 min and then purified (QIAquick purification kit). DNA samples having ligated P1 and P2 adapters were amplified for 18 cycles using Phusion Master Mix (NEB). Samples were then gel purified (range 300-700 bp, Qiagen) and diluted to ~10 nM for sequencing on the Illumina GA2x platforms (76 bp or 120 bp read lengths, Supplementary Table 1).

We used the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) and custom Perl scripts to process raw sequence reads. RAD data were split into individual-specific data subsets based on perfect barcode matches and then the barcodes were removed. Next, sequences were truncated to 50 bp and those having a FASTQ quality score of Q ≥20 over the 50 bp retained. To simplify alignments, identical sequences were collapsed, counted, and aligned to the *S. invicta* genome[9] (assembly F) using BWA[10]. After alignment, sequences having a perfect EcoRI site were considered as RAD tags. For each family the RAD tag data (BWA output) were loaded into an individual MySQL database to permit structured querying.

For each of the seven families, we assessed individual sample data quality. To guide this analysis, we first created a preliminary genotype table using all family members except those with extremely low raw read counts (total RAD tags <100,000). The table consisted of loci having exactly 2 alleles in the family and giving a genotype for more than 25% of the individuals. Per individual, a locus could be either

monoallelic (ant males are haploid) or apparently heterozygous, presumably because of slightly divergent repeats. To exclude spurious RAD tag alleles (sequencing errors) but retain real alleles found only in a few individuals we only considered those with family-wise allele read count (the number of sequence reads per allele can be more than 1 per individual) greater than threshold T. We chose T equal to one quarter of the individuals in the family to permit retention of alleles linked to *Gp-9B* in the non-recombining region for family P033 because it had only 10% *Gp-9B* (5/48) individuals, and empirically verified to work well (based on genetic map construction, below) for all seven families studied. Using this preliminary genotype table, we discarded presumably contaminated samples as follows. Because there is a high negative correlation between the counts of loci scored as heterozygous versus missing, contaminated samples were identified by examining a scatter plot of these variables and looking for samples with excessive heterozygous counts (outliers above the line). A sample was considered an outlier if its residual value (difference between observed heterozygous loci count and regression line) was more than two standard deviations from the mean of all residual values in the scatter plot. Finally, we removed males having genotypic data at less than 50% of the loci. Using only males that fulfilled all these criteria, we reprocessed the raw RAD tag data to identify the final alleles and loci to be used for building genetic maps. For alleles, we used the same total read count T-threshold as above. For loci, we retained a locus if it was biallelic in a family and monoallelic (*i.e.*, never heterozygous) in individuals.

We used MSTmap[11] to create genetic maps, with the following parameters: distance_function = kosambi, no_map_dist = 15.0, and no_map_size = 2. Two parameters had a large impact on map quality, the missing_threshold (for filtering loci with missing data) and the cut_off_p_value (for clustering markers into LGs). These two parameters were adjusted to create an optimal map where inappropriate fissions and fusions of LGs were minimized (*S. invicta* is known to have 16 chromosomes[12]). We found that the use of missing_threshold ≤25% and cut_off_p_value = 1e-8 gave 16 main LGs for four families (M013, M047, M173, and P034; Supplementary Table 1, Fig. 1, and Supplementary Figs. 1-3). For the remaining three families (P008, P016, and P033), these parameters did not generate a genetic map with 16 LGs probably due to lower sample sizes (all ≤46 individuals). Nevertheless for family P033, these same parameters were still optimal, resulting in 22 main LGs (Supplementary Fig. 6). For family P016, only 161 biallelic loci passed a missing_threshold of 25% which resulted in very few loci clustering into LGs. Relaxing missing_threshold to ≤40% in this family allowed the inclusion of 8 times more loci (n=1,318) and produced 23 main LGs (Supplementary Fig. 5). Finally, for family P008, the preceding cut-off_p_value failed to merge any loci into LGs, whereas, by lowering the threshold to 2.5e-6 with a missing_threshold ≤25% gave a map with 33 main LGs (Supplementary Fig. 4). Because the phase along each chromosome for the two RAD tag alleles of each locus was not known beforehand, we assigned the allele with greater counts as allele A and the other as allele B. When the parental allele phase was unknown, this resulted in adjacent loci that were randomly assigned opposite phases being split into separate LGs. To avoid this problem, we used a haplotype doubling method for phasing. By simply adding a duplicate dataset consisting of the mirror image genotypes (A and B are reversed) an initial map with twice the number of linkage groups was generated[13]. Subsequently, we reduced the data to half to restore the proper number linkage groups.

We chose family M013 as the reference family for all genetic map comparisons for two reasons. First, the mother queens had a *Gp-9BB* genotype hence permitting ascertainment of the gene order along the social chromosome given that recombination also occurrs between the two SB chromosomes (see results). Second, this monogyne family possessed the highest number of individuals that could be used for the analyses (Supplementary Table 1). We reordered the linkage groups (LG) from LG1 to LG15 in decreasing genetic map size, except for the *Gp-9* containing LGS (social chromosome) which we placed

last. The LGs from all other families were ordered accordingly. Genetic maps were displayed in Mapchart[14].

De novo sequencing and assembly of a *Gp-9b* male

To compare the SB and Sb chromosomes, we took advantage of the fact that the published *S. invicta* genome was based on a *Gp-9B* male[9] produced by a *Gp-9Bb* queen for which we also had collected *Gp-9b* sons. We sequenced and assembled the genome of one of these *Gp-9b* haploid sons. Following our previously described approach[9], we obtained 9,470,826 reads (3,397,368,189 bp) of Roche 454 shotgun sequence as well as 171,604,909 pairs of 100 bp Illumina HiSeq reads separated by an average distance of 321 bp (estimated using MAQ[15]) for a total of 34,320,981,800 bp from the focal *Gp-9b* male. In addition, we obtained 1,682,629 (resp. 751,453) reads with paired-end information separated by 8,000 bp (resp. 20,000 bp) for a total of 2,434,082 bp from Roche 454 paired-end. Illumina reads were first processed with FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) and fastq.cpp (https://github.com/brentp/bio-playground) to 1) trim the first and last 3 bp of each read (these bp had low qualities or biased nucleotide distributions); 2) remove duplicate read pairs (such duplication can bias the assembly); and 3) remove reads containing one or more unresolved "N" base (such reads may be of low quality). Subsequently, the cleaned data were assembled using SOAPdenovo 1.04 with "-K 31 -D 3 -R -L200" and map_len=64 and pair_num_cutoff=5 options. The resulting scaffolds were split at every N and chopped into 350 bp long sequences with 300 bp overlap using EMBOSS splitter[16]. These chopped sequences were provided as an external FASTA file alongside 454 shotgun SFF files to Roche 454 Newbler 2.3 (091027_1459) with the options "-cpu 4 -ace-infoall -m -nrm -large -mi 98 -ml 100 -rip". Finally, SFF files containing Roche 454 paired-end data were added and runProject was performed with the options "-cpu 4 -ace -pairt -infoall -m -large -mi 98 -ml 200". The final resulting assembly consisted of 104,395 contigs grouped into 9,121 scaffolds representing 335,325,025 bp with an N50 scaffold size of 560,112 bp.

Comparisons between assemblies

Most comparisons between the *Gp-9B* and *Gp-9b* genome assemblies were performed before RAD data were available for polygynous families other than P034. Thus, we proceeded conservatively with the P034 data: Based on visual inspection of counts of genotyped RAD loci per male, we restricted analyses to the 38*B* + 38*b* males with highest RAD coverage. For subsequent analyses, we retained only scaffolds for which five biallelic RAD loci were genotyped in at least 80% of the males. Seventeen of the 189 retained scaffolds from the *Gp-9B* assembly and 67 of the 229 retained scaffolds from the *Gp-9b* assembly were in complete linkage disequilibrium with *Gp-9.* Importantly, the linkage patterns of these scaffolds were conserved in subsequent RAD analyses of the additional polygynous families.

We identified orthologies between scaffolds by aligning the retained scaffolds from the *Gp-9b* assembly to those from the *Gp-9B* assembly using blastn[17] ($E < 10^{-10}$) and the Fast Statistical Alignment program[18]. These alignments were used for preliminary visualisation with Circoletto[19] and the Artemis Comparison Tool[20]. Subsequently, lastz[21] and custom scripts were used to generate data tables for more complete visualisation in Circos[22]. These visualisation approaches helped to identify differences between Sb and SB as well as differences between the non-recombining regions and the rest of the genome.

To identify haplotype-specific insertions of repetitive sequences, pairwise nucleotide alignments between all pairs of orthologous scaffolds were constructed using FSA[18] with the options "--kmer 51 --anchored". To minimize the impact of potential problems with assembly or alignment, we retained only putative insertions matching the following four criteria: size > 100 bp because very short putative insertions may be due to misassembly of repetitive sequences (*e.g.*, microsatellites); size < 1000 bp

because very long putative insertions may be artefacts stemming from unscaffolded contigs; presence of resolved nucleotide sequence because we could not ascertain that putative insertions composed exclusively of unresolved stretches of "N" were real; and distance of at least 1000 bp to putative deletions because misalignment could lead to apparently neighbouring insertions and deletions. Using other criteria yielded results similar to those reported in the main text.

To calculate intron sizes, we mapped the official *S. invicta* gene set models[9] to the *Gp-9b* assembly and to the *Gp-9B* assembly with Scipio[23] because this software provides extensive intron and exon statistics for each gene. Then, we determined intron orthologies by matching gene identities and locations of splice sites within coding sequences.

To identify genes present in *Gp-9B* but at least partially missing in *Gp-9b* we first mapped one lane of paired Illumina reads from the *Gp-9b* male (lane "100830_s_7") to the *Gp-9B* genome using BWA[10]. RSamtools was used to determine the coverage depth of each predicted exon of the *S. invicta* 2.2.3 geneset (25.47±0.22 median±SE). Non-null coverage of a truly missing exon could be due to incorrectly mapped reads, thus we retained exons within the lowest 2% quantiles (*i.e.*, lower than 5.1x mean coverage) and larger than 120 bp as candidates missing from *Gp-9b*. Subsequently, we obtained 6x genome coverage paired Illumina reads from one *Gp-9B* son and one *Gp-9b* son from each of seven unrelated *Gp-9Bb* queens, mapped the reads to the *Gp-9B* genome and estimated coverage for each exon for each sequenced individual as above. This identified ten exons representing five genes that were putatively missing from Sb; these are reported in the manuscript. Finally, we randomly selected one exon from each of these genes for PCR confirmation (not shown). For all five tested exons we obtained strong PCR products only from *Gp-9B* males but not from *Gp-9b* males.

PCR confirmations

We performed PCR confirmations on four *Gp-9B* and four *Gp-9b* males from each of five colonies collected near Athens, Georgia (USA) as follows. Previously generated pairwise nucleotide-level FSA alignments of SB and Sb scaffolds were examined with Jalview[24]. We then copied from Jalview to Primer3Plus[25] sequences flanking putative insertions or rearrangements to design primer pairs expected to yield haplotype-specific PCR products. PCR mixes were performed in 20 µL volumes with 1 µL of DNA template, 1x PCR Buffer (QIAGEN), 0 to 5 mM of MgCl (QIAGEN), 0x to 1x of QIAGEN Q solution, 0.2 mM of dNTPs, 0.2 uM of each primer, and 0.05 U of QIAGEN Taq polymerase. The PCR protocol consisted of an initial denaturation for 5 min at 95°C followed by 35 cycles of 30 min denaturation at 95 °C, 30 min annealing at 55°C and between 45 sec and 2 min 15 sec extension at 72°C (extension time depended on expected PCR product size). PCR products were run on 1.5% agarose gels stained with 1% ethidium bromide and photographed with a Syngene EF2 imaging system. PCR product sizes were visually scored.

Gene expression microarrays

For cDNA microarray gene expression experiments comparing young queens of alternate *Gp-9* genotype, we collected colonies near Athens, Georgia (USA) and obtained 1-day old virgin queens as follows. First, because mother queens partly inhibit the production of sexuals[26,27] we made queenless sub-colonies consisting of ~10,000 haphazardly chosen workers with young brood to maximize the yield of virgin queens. Then, each day we checked these sub-colonies and transferred newly eclosed virgin queens (0 - 24 h) to sub-colonies with 300 – 500 workers without brood. Next, we let the transferred virgin queens mature for another day and then snap-froze them in liquid nitrogen, after which we stored them at -80°C until DNA and RNA extraction. To avoid pseudo-replication and to minimize day effects, we only used a single *Gp-9BB* and a single *Gp-9Bb* virgin queen (always eclosed on the same day in the same sub-

colony) from each of 16 unrelated polygyne colonies for a total of 32 queen individuals that were used for expression profiling.

For comparing males of alternate *Gp-9* genotypes, we collected colonies near Athens, Georgia (USA) and established laboratory colonies to obtain males of three time points: brown-eyed-all-white-bodied pupae, adults one day after eclosion, and adults eleven days after eclosion. To ensure that all males were derived from a single mother queen, we established laboratory colonies with workers, brood, and a single *Gp-9Bb* queen and waited for a minimum of six months before collecting males. Because the brown-eyed-all-white-bodied period during the pupal stage is short and synchronized to within 24 hrs (personal observation), we collected those individuals directly based on visual morphology. For the adult males, we proceeded like for the queens above where we transferred newly eclosed males to sub-colonies and waited one or 11 additional days before snap-freezing them in liquid nitrogen and storing at -80°C. For each time point, we only used a single *Gp-9B* and a single *Gp-9b* male (always sampled on the same day in the same colony or sub-colony for pupae or adults, respectively) from each of five independent *Gp-9Bb* mother queen colonies. In total, we used 30 male individuals (5 colonies x 3 time points x 2 genotypes) for expression profiling.

For RNA and DNA extraction, we first homogenized each sample in 600 µl of RLT buffer (Qiagen RNeasy kit) in 2 ml tubes with ~8 ceramic beads for 1 min at maximum speed in a FastPrep-24 tissue homogenizer (MP Biomedicals). We used a 6 µl aliquot of the lysate for DNA extraction and purification with the Agencourt DNAdvance kit (Beckman Coulter), followed by PCR-RFLP for *Gp-9*[2]. For RNA purification from queens and male pupae, we precipitated the RNA from the RLT buffer with 50% ethanol before continuing with the manufacturer's protocol for the RNeasy kit. For adult males, because of an unknown inhibitor presumably in the cuticle that inhibited binding of RNA to the Qiagen columns, we precipitated total RNA from the RLT buffer with isopropanol overnight before dissolving the RNA pellet in 1 ml of TRIzol® reagent (Life Technologies), and then followed the manufacturer's instructions for RNA purification. For all RNA samples, we removed residual DNA from total RNA isolated with the Turbo DNA-free kit (Life Technologies). We then amplified the mRNA ~10^1-10^2 fold using the MessageAmp II aRNA kit (Ambion).

For microarray experiments, we labeled each amplified RNA sample with Cy3 dye and combined it with Cy5-labeled common reference RNA prior to hybridization onto *S. invicta* cDNA microarrays[28,29]. After hybridization, we scanned the microarrays using an Agilent scanner and extracted foreground and background intensities for Cy3 and Cy5 at each cDNA spot using Axon GenePix® Pro. After excluding cDNA spots that failed to yield a single PCR product and spots that had no blastx similarity to proteins in the NCBI non-redundant database we retained 15,132 spots representing 5,956 genes. Subsequently, we used limma[30] for normexp background correction, print-tip loess normalization within arrays, and aquantile normalization between arrays. Finally, we separately compared expression profiles between individuals of the same sex and stage of alternative genotypes (*i.e., Gp-9BB* versus *Gp-9Bb* for queens and *Gp-9B* versus *Gp-9b* for males) using empirical Bayes moderated paired *t*-statistics (as recommended for limma[30]) with a 1% (queens) or 5% (males) false discovery rate[31]. Between 1-day virgin queens of alternate genotypes, we identified 38 differentially expressed genes. For male pupae there were five differentially expressed genes, however, no genes were differentially expressed for either 1- or 11-day old adult males.

To determine the physical locations of the genes differentially expressed between males and queens of alternate genotype as well as for the 39 genes previously shown to be differentially expressed between workers of alternate genotypes[29], we used blastn[17] to compare the 67 non-redundant genes to the fire

ant genome and found 47 sequences mapping unambiguously ($E < 10^{-50}$ followed by visual inspection) to scaffolds on the *Gp-9B* genome. Of these, 39 genes were on a scaffold integrated onto the genetic map. We also assigned SiJWF04BEA, which encodes a piggyBac transposon, to the non-recombining region of the social chromosome because this transposon is fully linked to the *Gp-9b* allele[29]. To illustrate the relative expression pattern and the chromosomal locations of these differentially expressed genes, we used Circos[22].

Molecular evolution analyses

Molecular evolution analyses required alignments of orthologous genes from the *Gp-9B* and *Gp-9b* assemblies. Because the *S. invicta* official gene set 2.2.3 is based on the *Gp-9B* genome[9] we generated gene models for the *Gp-9b* genome assembly with MAKER2[32] based on assembled transcriptome sequence and the *S. invicta* official gene set. Orthology between gene models from the two assemblies was established using reciprocal best blastn searches[17,33]. A codon-level alignment was performed for each pair of orthologs using PRANK[34]. To reduce noise and eliminate genes with identical alleles in both assemblies, we retained only genes longer than 500 bp having at least one synonymous difference. The dN/dS ratios were calculated using codeml in PAML[35].

Bacterial artificial chromosome fluorescent in situ hybridization

We used two monogyne and four polygyne colonies collected near Taoyuan, Taiwan for BAC FISH analyses. For haploid nuclei, we dissected testes imaginal discs from 4th-instar male larvae. For diploid nuclei, we dissected brain (Fig. 2b) and ovary imaginal (Supplementary Fig. 8e) discs from diploid 4th-instar female (queen) larvae. In total, we examined five *Gp-9B* males from two monogyne colonies, six *Gp-9b* males from two polygyne colonies, and three *Gp-9Bb* females from two polygyne colonies. The genotype of each individual was determined using DNA extracted from the remainder of the body. We performed metaphase spreading according to Yoshido *et. al.*[36] with the following modifications. First, we pretreated the dissected tissues with hypotonic solution (1% sodium citrate) and then fixed the cells in freshly prepared fixation solution (methanol:acetic acid, 3:1). After chromosome spreading, we passed slides through a graded ethanol series (70, 80, and 100%) and subsequently aged them in 100% ethanol at -20^{o}C for at least 1 day.

To identify BACs specific to the social chromosomes, we performed end sequencing of random clones from plate 73 of the SW_Ba BAC library (Clemson University Genomics Institute, Georgia, USA) and then compared the sequence to the genetic linkage map using blastn[17]. We used the following five BACs for this analysis: A18, A22, E03, E17, and G23. BAC DNA was isolated by culturing BAC clones in LB medium containing 12.5 μg/ml chloramphenicol at 37^{o}C for 16 hr and then extracting the DNA with the Midi Kit (QIAGEN). We labelled BACs by nick translation using the Tag DNA Multicolor Kit (Invitrogen) with Alexa Fluor 488, 594, 647 (Invitrogen), or Dylight (Thermo) fluorochromes.

For hybridization, we first removed aged slides from -20^{o}C and air dried them at room temperature. Next, we treated slides with 0.05% pepsin for 10 min at 37^{o}C, washed them with PBS for 5 min at room temperature, and dehydrated them with an ethanol series (70, 80, and 100%). Then, we denatured chromosomes at 72^{o}C for 2 min in 70% formamide/2x SSC and stopped the reaction in an ice-cold ethanol series (70, 80, and 100%). The DNA probe/hybridization buffer contained 50 ng labeled probe, 50% formamide, 10% dextran sulfate, 2x SSC, 5 μg sonicated salmon sperm DNA, 0.2% SDS, and 1x Denhart's solution. After overnight hybridization in a humid chamber at 37^{o}C, we washed slides three times in 0.1x SSC/1% Triton X-100 at 62^{o}C. Next, we dipped the slides quickly in ddH_2O and air dried

them. Finally, we counterstained and mounted DNA in VECTASHIELD Mounting Medium containing 0.5 μg/ml DAPI.

We acquired images using a DeltaVision imaging system (Applied Precision Inc.) with DAPI, FITC, A594, and Cy5 filters, and then processed the images through softWoRx deconvolution. We pseudo-colored the original gray-scale images for each probe as indicated in the respective figures and superimposed multiple image overlays using Photoshop CS5.

Allele-specific gene expression

We sequenced DNA from one *Gp-9B* son and one *Gp-9b* son from each of seven unrelated *Gp-9Bb* queens collected near Athens, Georgia (USA) using two 101 bp paired-end multiplexed Illumina HiSeq 2000 lanes and retained only reads passing Illumina quality filters. To control for the reduced mapping-efficiency of sequence reads containing non-reference alleles[37], we constructed an SB- and an Sb-specific reference genome from the published fire ant genome assembly[9]. For this, we aligned the grouped sequences of the seven *Gp-9B* males to the published genome assembly using Bowtie2 version 2-2.0.0-beta7[38] and called genotypes with samtools mpileup and VarScan version v2.3.2[39] while eliminating reads that were duplicated, that matched more than one genomic location, that contained three or more unresolved "N" basepairs, or that differed from the published genome by three or more mismatching basepairs. Finally, VarScan mpileup2cns and custom scripts were used to build an SB-specific consensus genome sequence from SNPs with 5x coverage or more, while retaining the regions without coverage. The same procedure was repeated using the sequences of the seven *Gp-9b* males to construct an Sb-specific consensus genome sequence.

We used DNA and RNA sequence data from *Gp-9B* and *Gp-9b* males to identify SNPs in linkage disequilibrium with *Gp-9*. DNA samples included the above-mentioned seven *Gp-9B* and seven *Gp-9b* males as well as the additional single *Gp-9B* [9] and *Gp-9b* males used for initial genome assembly of each. The RNA samples included one pool of five *Gp-9B* males and one pool of five *Gp-9b* males from three unrelated colonies collected near Athens, Georgia (USA). All samples were aligned using Bowtie2 [38] to SB- and Sb-specific references. For each of the two sets of alignments, we then eliminated reads as described in the previous paragraph and normalised read counts by randomly retaining 10,313,520 aligned reads per DNA sample and 7,467,230 per RNA sample. Afterwards, we called SNPs on each reference using samtools mpileup and VarScan mpileup2snp, retaining only biallelic SNPs that were haploid in each genomic male sample, but overall had 1) at least 12x coverage of the major allele and 8x coverage of the minor allele, 2) a basepair quality of 20, and 3) a minor allele fraction of at least 35%. Overall, we retained only the 742,195 SNPs found in both SB and Sb reference alignments (*i.e.,* 89% of total SNPs).

We called a genotype in each sample for each of these 742,195 SNPs, and constructed a contingency table summarizing the number of *Gp-9B* samples and the number of *Gp-9b* samples with each of the two possible alleles. Allele-distributions for 31,273 SNPs deviated from random ($P < 0.05$ Fisher's Exact Test, FDR-corrected[31]). Among these significant SNPs, 23,924 (76.5%) occurred in genomic regions with known linkage groups, and 93.5% of these regions were in linkage disequilibrium with *Gp-9* (based on the RAD results). We retained the latter set to study allele-specific expression (ASE) in *Gp-9Bb* queens.

We performed RNA-seq on pools of four *Gp-9Bb* queens from each of six unrelated colonies to identify genes with ASE in *Gp-9Bb* queens. The resulting sequence data were aligned with Bowtie2[38] to the SB-specific consensus genome sequence, and subsequently to the Sb-specific consensus sequence. Alignments were filtered as was done for SNP identification, and read counts were normalised by

retaining 6,373,590 randomly selected alignments per sample. Furthermore, we only examined the 22,358 SNPs previously identified as being in linkage disequilibrium with *Gp-9.* To determine the expression level of the SB (resp. Sb) allele for each SNP in each sample, we extracted reference allele counts from the alignments to the SB consensus genome sequence (resp. Sb consensus genome sequence) using samtools mpileup and VarScan readcounts. We subsequently retained the 7,640 SNPs (34% out of 22,358 identified SNPs) that were biallelic as predicted and were present in transcribed regions. Finally, we further conservatively filtered SNPs by only retaining the 1,286 SNPs that had at least 20x coverage in all samples and did not contain insertions or deletions at SNP positions.

We assessed ASE using the Expectation Maximization algorithm for allele-specific events from the iASeq R Bioconductor package[40] with 100,000 iterations and relative error tolerance of 1e-3. Allele-biased expression was identified for 63 SNPs (posterior probability < 5%). For genes having multiple SNPs, we retained only those genes where all SNPs had consistent allele-biased expression patterns. This resulted in 21 genes (33 SNPs) with higher or monoallelic expression of the Sb allele and 10 genes (15 SNPs) with higher or monoallelic expression of the SB allele.

Fire ant repetitive elements

The sequence reads from each of 14 males (7 *Gp-9B* and 7 *Gp-9b* sons used for ASE, above) were mapped individually to a library of 1,461 fire ant repetitive elements (Christopher D. Smith, personal communication). The single best blastn hit per read with $E<10^{-10}$ was retained. Two-sided paired *t*-tests on numbers of reads matching each repetitive element were performed using a threshold of $P < 0.05$ after FDR correction.

Estimating when recombination suppression occurred

We were able to obtain a gross estimate of divergence time between SB and Sb using a molecular clock approach[41] calibrated using the divergence time of the two most closely related ant species that have sequenced genomes, namely, the leafcutter ants *Atta cephalotes* and *Acromyrmex echinatior,* which diverged ~10 million years ago[42]. Our approach assumes synonymous mutations are neutral and follow a Jukes-Cantor model, and more crudely assumes that leafcutter ants and the fire ant social chromosomes are under identical evolutionary regimes. We determined the distribution of synonymous substitution rates between the Sb-alleles and SB-alleles for all genes in the non-recombining region of the social chromosome and between all one-to-one orthologs of leafcutter ant genomes. Subsequently, we performed maximum likelihood comparisons of the two distributions using mle in R. These comparisons suggest that recombination in the non-recombining region of the social chromosome likely ceased 388,018 years ago (95% confidence interval: 350,000 to 424,000 years). This estimate is less than the 1.1 million year old *Drosophila miranda* neo-X and neo-Y sex chromosomes[43].

References

1 Jouvenaz, D. P., Allen, G. E., Banks, W. A. & Wojcik, D. P. Survey for pathogens of fire ants, *Solenopsis* spp., in the Southeastern United States. *Fla. Entomol.* **60**, 275-279 (1977).

2 Krieger, M. J. B. & Ross, K. G. Identification of a major gene regulating complex social behavior. *Science* **295**, 328-332 (2002).

3 Aerts, J. & Law, A. An introduction to scripting in Ruby for biologists. *BMC Bioinformatics* **10** (2009).

4 Goto, N. *et al.* BioRuby: bioinformatics software for the Ruby programming language. *Bioinformatics* **26**, 2617-2619 (2010).

5 Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

6 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York Inc, 2009).

7 Aboyoun, P., Pages, H. & Lawrence, M. GenomicRanges: Representation and manipulation of genomic intervals. (2012). <bioconductor.jp/packages/2.9/bioc/html/GenomicRanges.html>.

8 Baird, N. A. *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers. *Plos One* **3**, e3376 (2008).

9 Wurm, Y. *et al.* The genome of the fire ant *Solenopsis invicta*. *Proc. Natl. Acad. Sci. USA* **108**, 5679-5684 (2011).

10 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

11 Wu, Y., Bhat, P. R., Close, T. J. & Lonardi, S. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *Plos Genet.* **4**, e1000212 (2008).

12 Glancey, B. M., Romain, M. K. S. & Crozier, R. H. Chromosome numbers of red and black imported fire ants, *Solenopsis invicta* and *Solenopsis richteri*. *Ann. Entomol. Soc. Am.* **69**, 469-470 (1976).

13 Gadau, J. *et al.* A linkage analysis of sex determination in *Bombus terrestris* (L.) (Hymenoptera : Apidae). *Heredity* **87**, 234-242 (2001).

14 Voorrips, R. E. MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **93**, 77-78 (2002).

15 Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851-1858 (2008).

16 Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276-277 (2000).

17 Camacho, C. *et al.* BLAST plus : architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

18 Bradley, R. K. *et al.* Fast Statistical Alignment. *Plos Comput. Biol.* **5**, e1000392 (2009).

19 Darzentas, N. Circoletto: visualizing sequence similarity with Circos. *Bioinformatics* **26**, 2620-2621 (2010).

20 Carver, T., Harris, S. R., Berriman, M., Parkhill, J. & McQuillan, J. A. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464-469 (2012).

21 Harris, R. S. *Improved pairwise alignment of genomic DNA* PhD thesis, The Pennsylvania State University, (2007).

22 Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639-1645 (2009).

23 Keller, O., Odronitz, F., Stanke, M., Kollmar, M. & Waack, S. Scipio: using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *BMC Bioinformatics* **9** (2008).

24 Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).

25 Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**, W71-W74 (2007).

26 Tschinkel, W. R. *The fire ants*. (The Belknap Press of Harvard University Press, 2006).

27 Vargo, E. L. & Fletcher, D. J. C. Evidence of Pheromonal Queen Control over the Production of Male and Female Sexuals in the Fire Ant, Solenopsis-Invicta. *J Comp Physiol A* **159**, 741-749 (1986).

28 Wang, J. *et al.* An annotated cDNA library and microarray for large-scale gene-expression studies in the ant *Solenopsis invicta*. *Genome Biol* **8**, R9 (2007).

29 Wang, J., Ross, K. G. & Keller, L. Genome-wide expression patterns and the genetic architecture of a fundamental social trait. *Plos Genet.* **4**, e1000127 (2008).

30 Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, 3 (2004).

31 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc. B Met.*, 289-300 (1995).

32 Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12** (2011).

33 Wall, D. P., Fraser, H. B. & Hirsh, A. E. Detecting putative orthologs. *Bioinformatics* **19**, 1710-1711 (2003).

34 Loytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl Acad. Sci. USA* **102**, 10557-10562 (2005).

35 Yang, Z. H. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586-1591 (2007).

36 Yoshido, A., Marec, F. & Sahara, K. Resolution of sex chromosome constitution by genomic in situ hybridization and fluorescence in situ hybridization with (TTAGG)(n) telomeric probe in some species of Lepidoptera. *Chromosoma* **114**, 193-202 (2005).

37 Degner, J. F. *et al.* Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics* **25**, 3207-3212 (2009).

38 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359 (2012).

39 Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568-576 (2012).

40 Wei, Y. Y., Li, X., Wang, Q. F. & Ji, H. K. iASeq: integrating multiple ChIP-seq datasets for detecting allele-specific binding. (2012). <www.bioconductor.org/packages/2.10/bioc/html/iASeq.html>.

41 Kumar, S. Molecular clocks: four decades of evolution. *Nat. Rev. Genet.* **6**, 654-662 (2005).

42 Nygaard, S. *et al.* The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Res.* **21**, 1339-1348 (2011).

43 Bachtrog, D. & Charlesworth, B. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**, 323-326 (2002).

44 Ometto, L., Shoemaker, D., Ross, K. G. & Keller, L. Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. *Mol. Biol. Evol.* **28**, 1381-1392 (2011).

Overview of Supplementary Figures

Supplementary Fig. 1. Genetic linkage map for family M013.

Supplementary Fig. 2. Genetic linkage map for family M047.

Supplementary Fig. 3. Genetic linkage map for family M173.

Supplementary Fig. 4. Genetic linkage map for family P034.

Supplementary Fig. 5. Genetic linkage map for family P008.

Supplementary Fig. 6. Genetic linkage map for family P016.

Supplementary Fig. 7. Genetic linkage map for family P033.

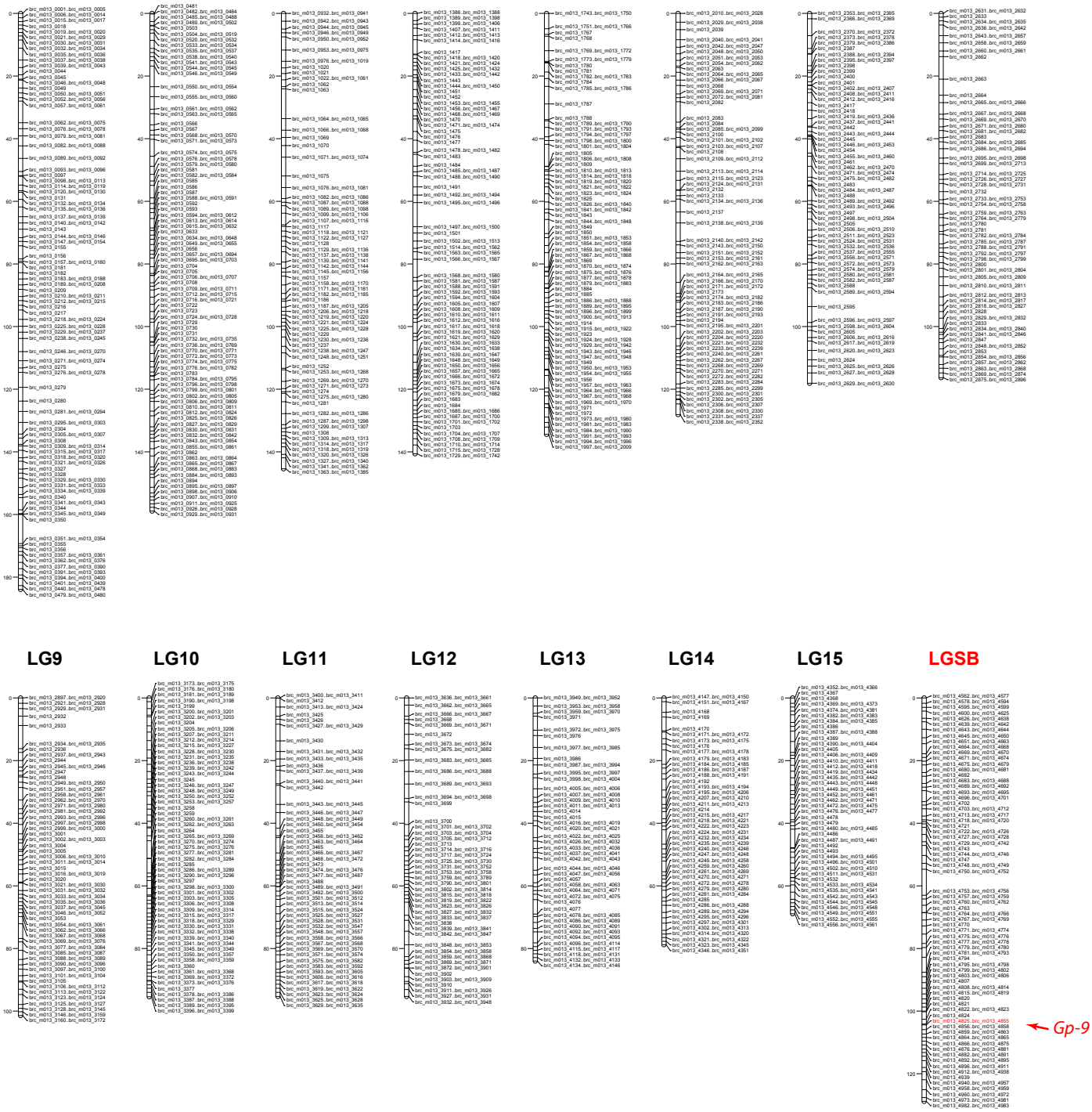
Supplementary Fig. 8. BAC FISH showing complete chromosome complement.

Supplementary Fig. 9. BAC FISH using additional samples and probes.

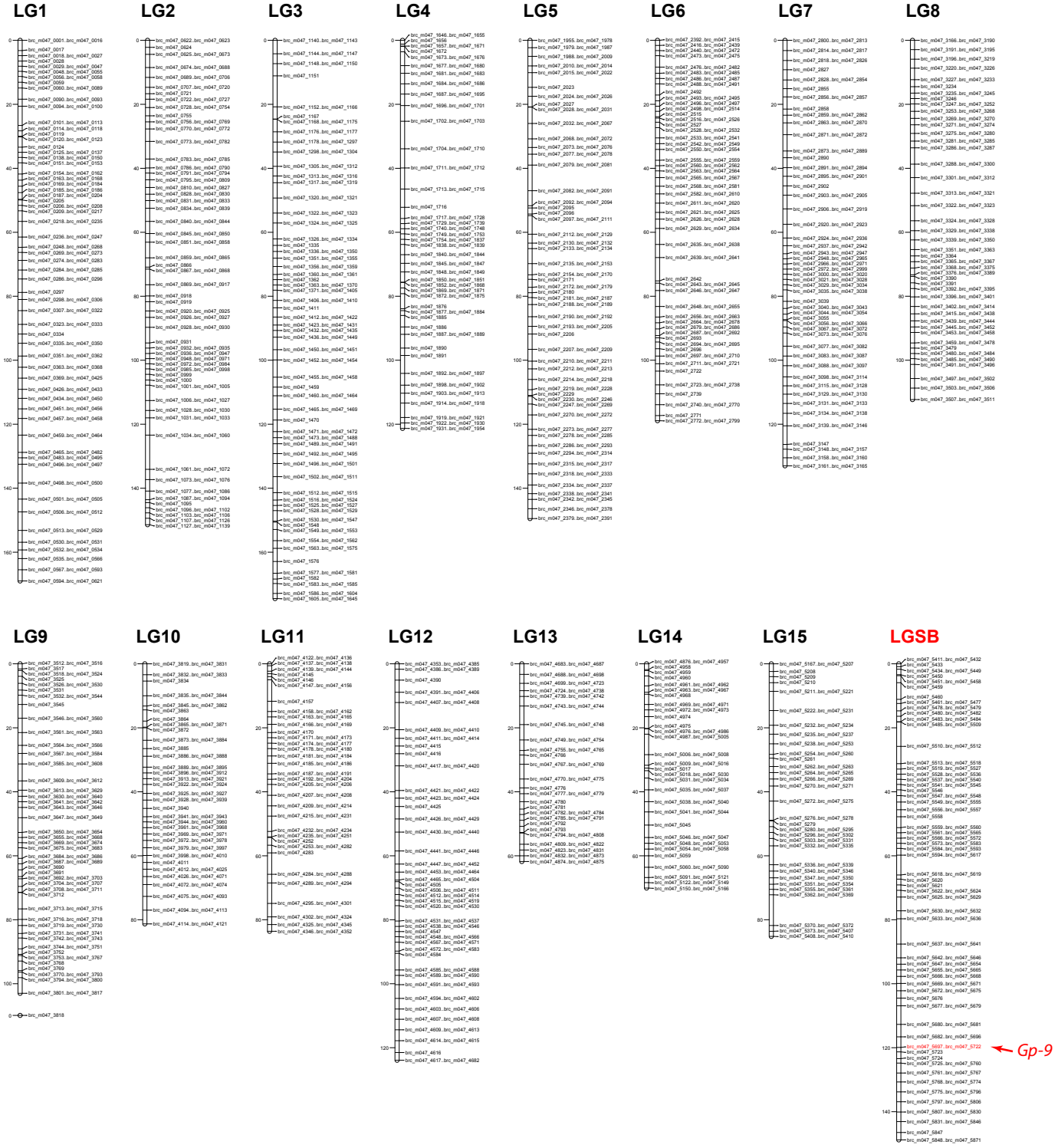
Supplementary Fig. 10. Structural rearrangement between the non-recombining SB (top) and Sb (bottom) regions.

Supplementary Fig. 11. Distribution of the relative fraction of gene expression from *B*-specific alleles for 288 genes in *Gp-9Bb* queens

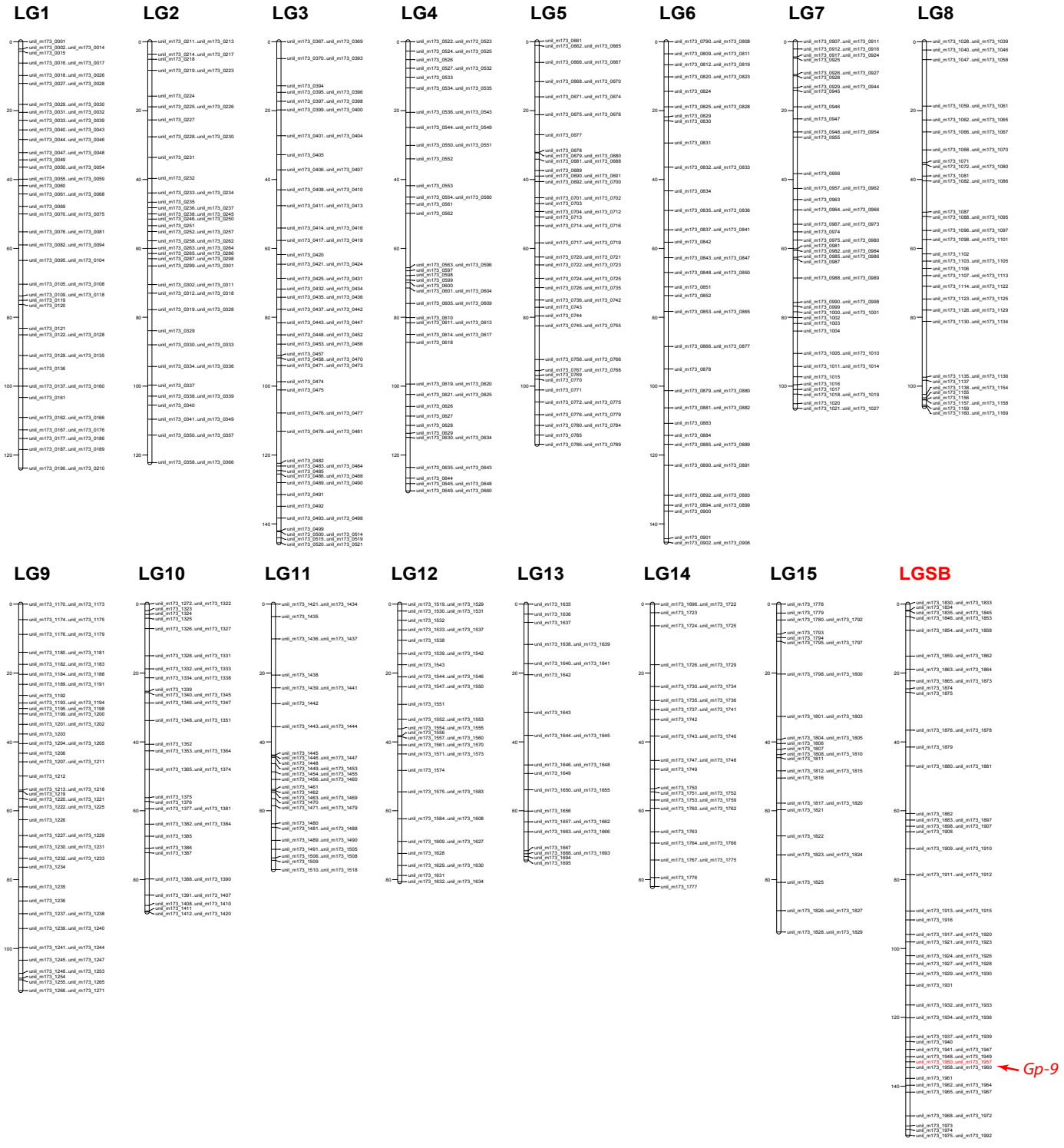
Supplementary Fig. 12. Distributions of frequencies of synonymous substitutions (dS) in a comparison of leafcutter ants and genes on the non-recombining portion of the social chromosome.



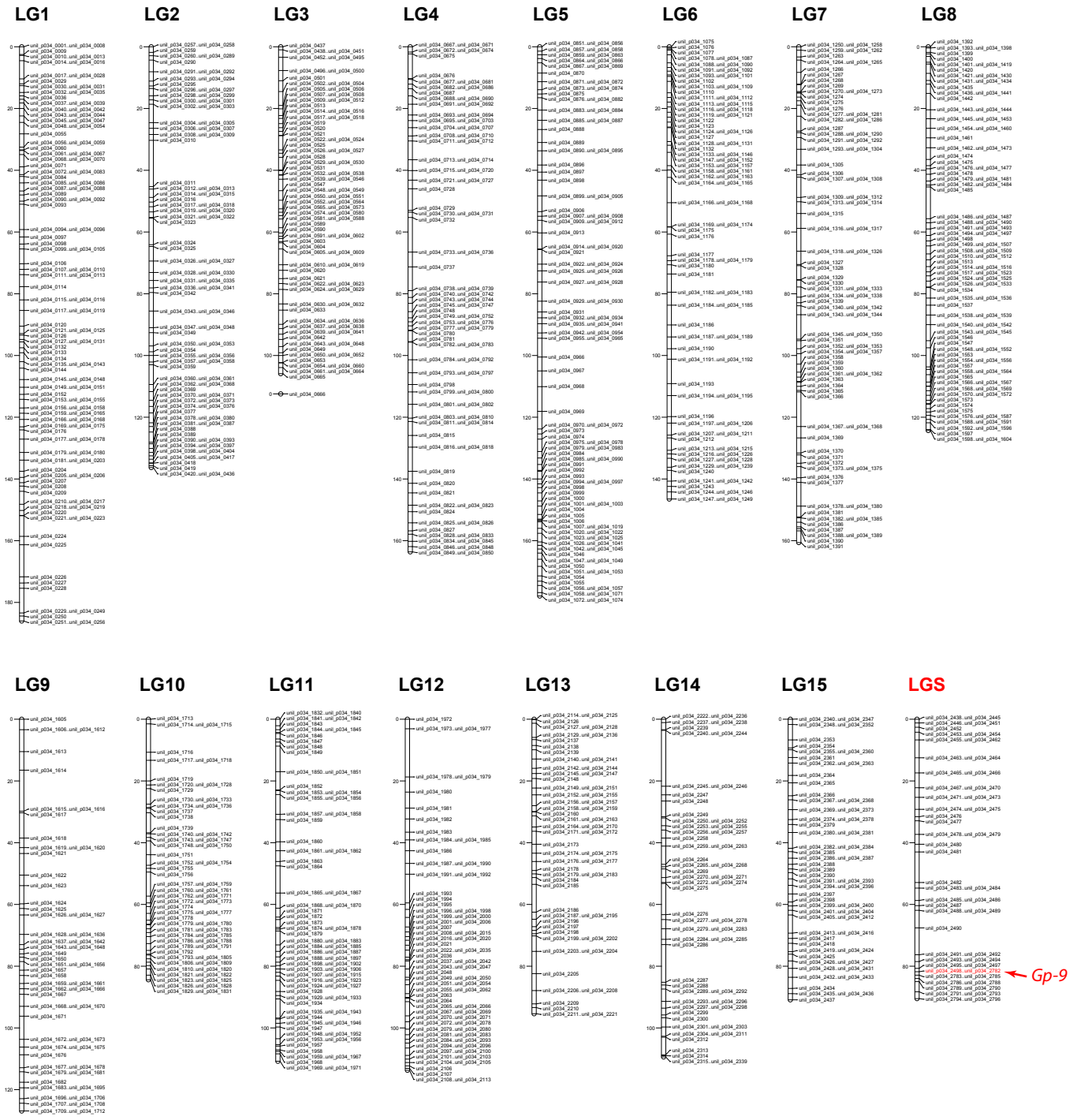
Supplementary Figure 1. Reference genetic linkage map for *Solenopsis invicta*. Genetic linkage map derived from RADseq analysis of male offspring from a monogyne *Gp-9BB* queen (M013). Linkage groups (LGs) are assigned 1 to 15 according to decreasing genetic length except for the social chromosome, LGS, which is placed last. Genetic positions in centimorgans (cM) are given on the left of each LG and RAD marker names on the right. Each genetic marker has a prefix (i.e., brc_m013_) and a number based on its serial position on the map. Multiple markers that have the same map position are indicated using "first .. last" marker notation. The position of *Gp-9* is indicated in red.



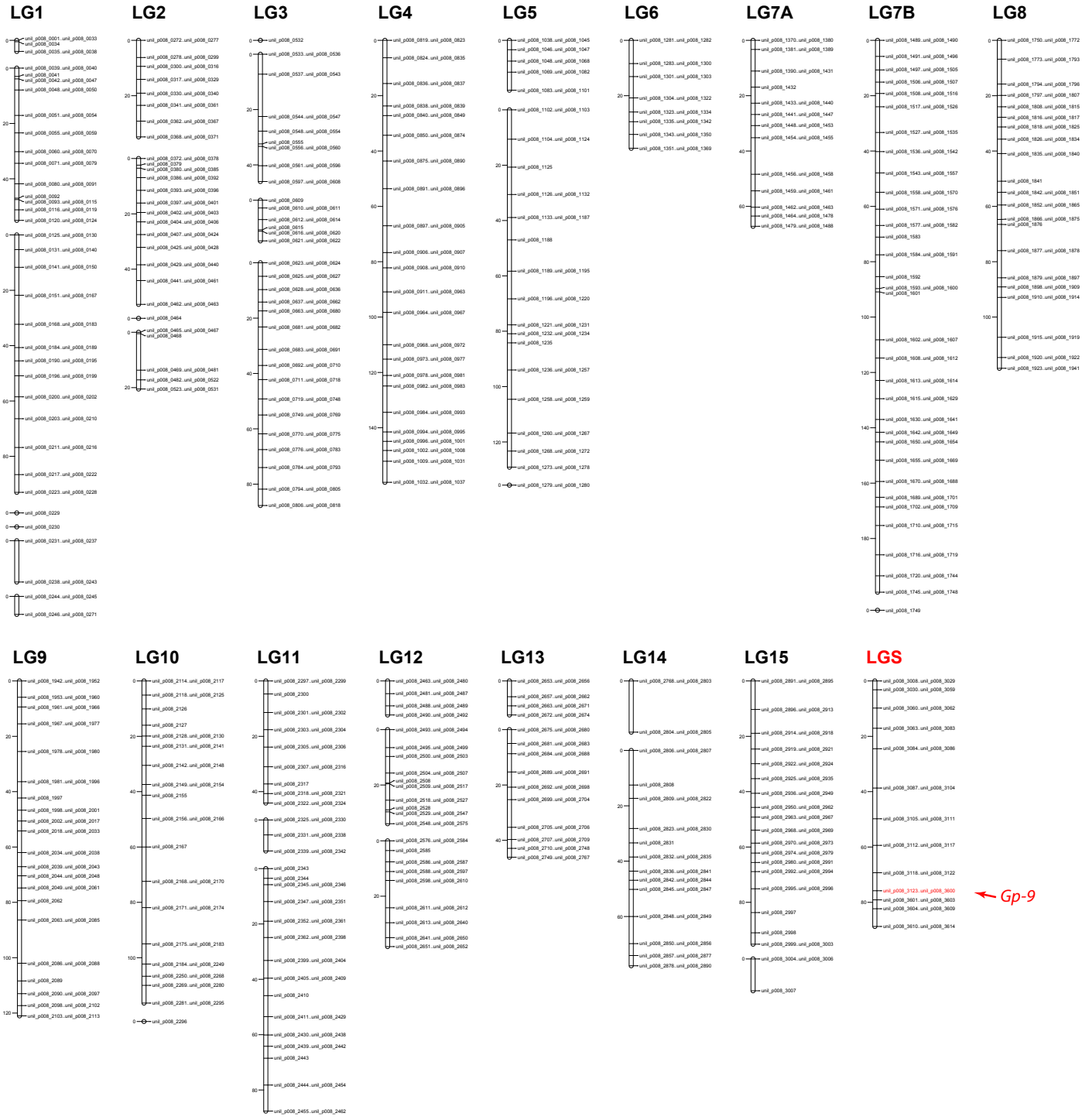
Supplementary Figure 2. Genetic linkage map for family M047.
LGs are arranged according to the reference genetic map (family M013).



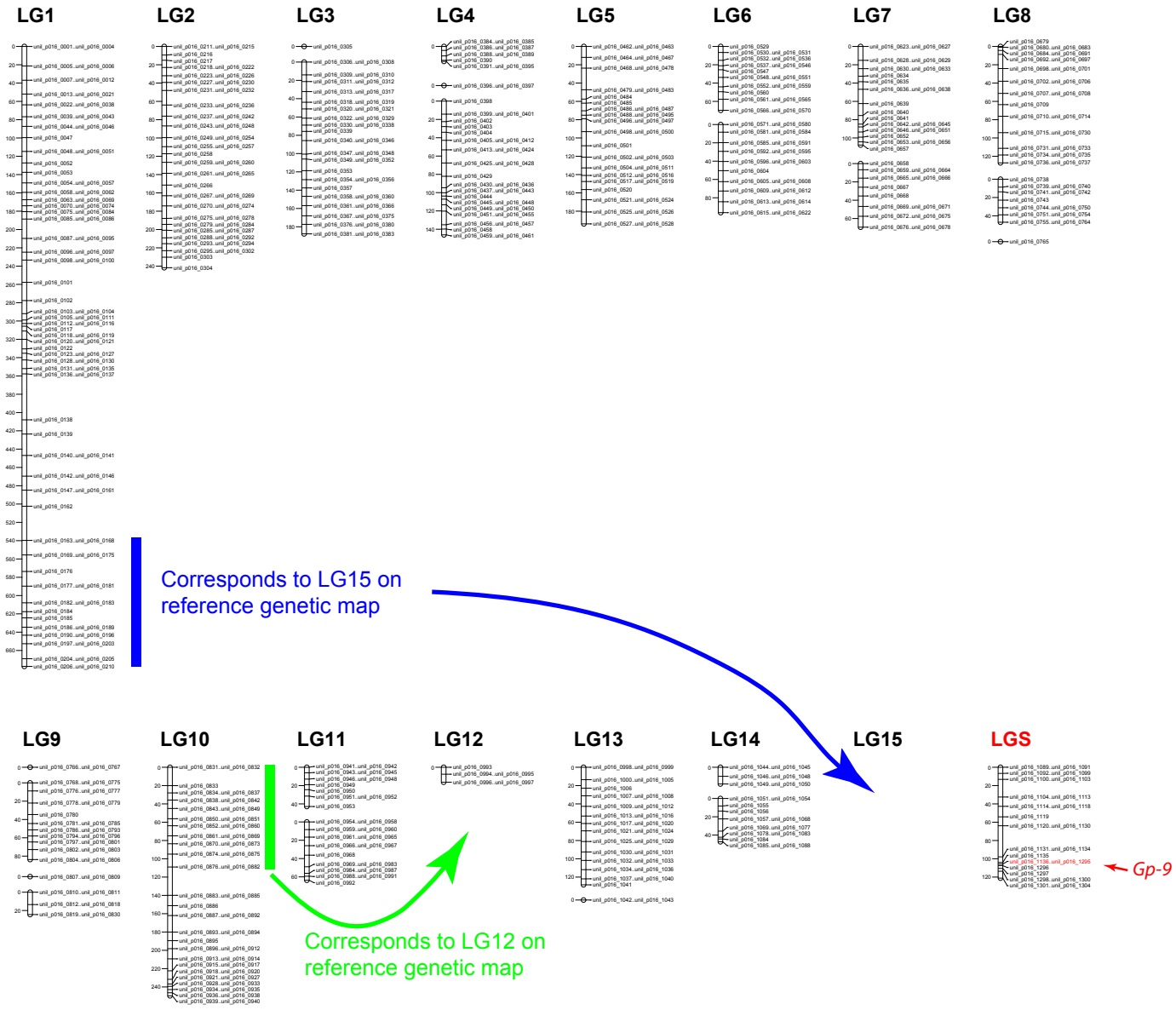
Supplementary Figure 3. Genetic linkage map for family M173.
LGs are arranged according to the reference genetic map (family M013).



Supplementary Figure 4. Genetic linkage map for family P034.
Of the 2,796 biallelic RAD markers in this family, 285 (10%) are non-recombining (red). LGs are arranged according to the reference genetic map (family M013).



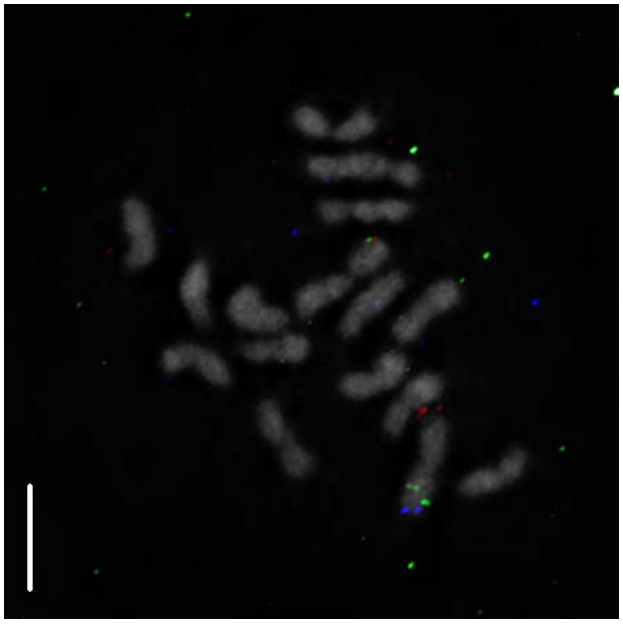
Supplementary Figure 5. Genetic linkage map for family P008.
Of the 3,614 biallelic RAD markers in this family, 478 (13%) are non-recombining (red). LGs are arranged according to the reference genetic map (family M013). Some reference LGs are split in this family likely due to low sample size (n=31). Sub-LGs are arranged vertically, space permitting.



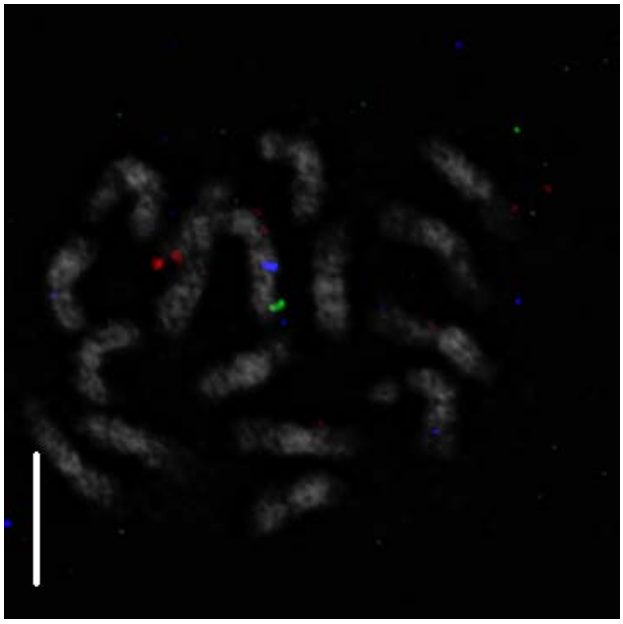
Supplementary Figure 6. Genetic linkage map for family P016.
Of the 1,380 biallelic RAD markers in this family, 160 (12%) are non-recombining (red). LGs are arranged according to the reference genetic map (family M013). Some reference LGs are split or merged (blue and green) in this family likely due to low sample size (n=46). Sub-LGs are arranged vertically.

Supplementary Figure 7. Genetic linkage map for family P033
Of the 7,360 biallelic RAD markers in this family, 812 (11%) are non-recombining (red). LGs are arranged according to the reference genetic map (family M013). Some reference LGs are split in this family likely due to low sample size (n=46). Sub-LGs are arranged vertically.

Gp-9B haploid male

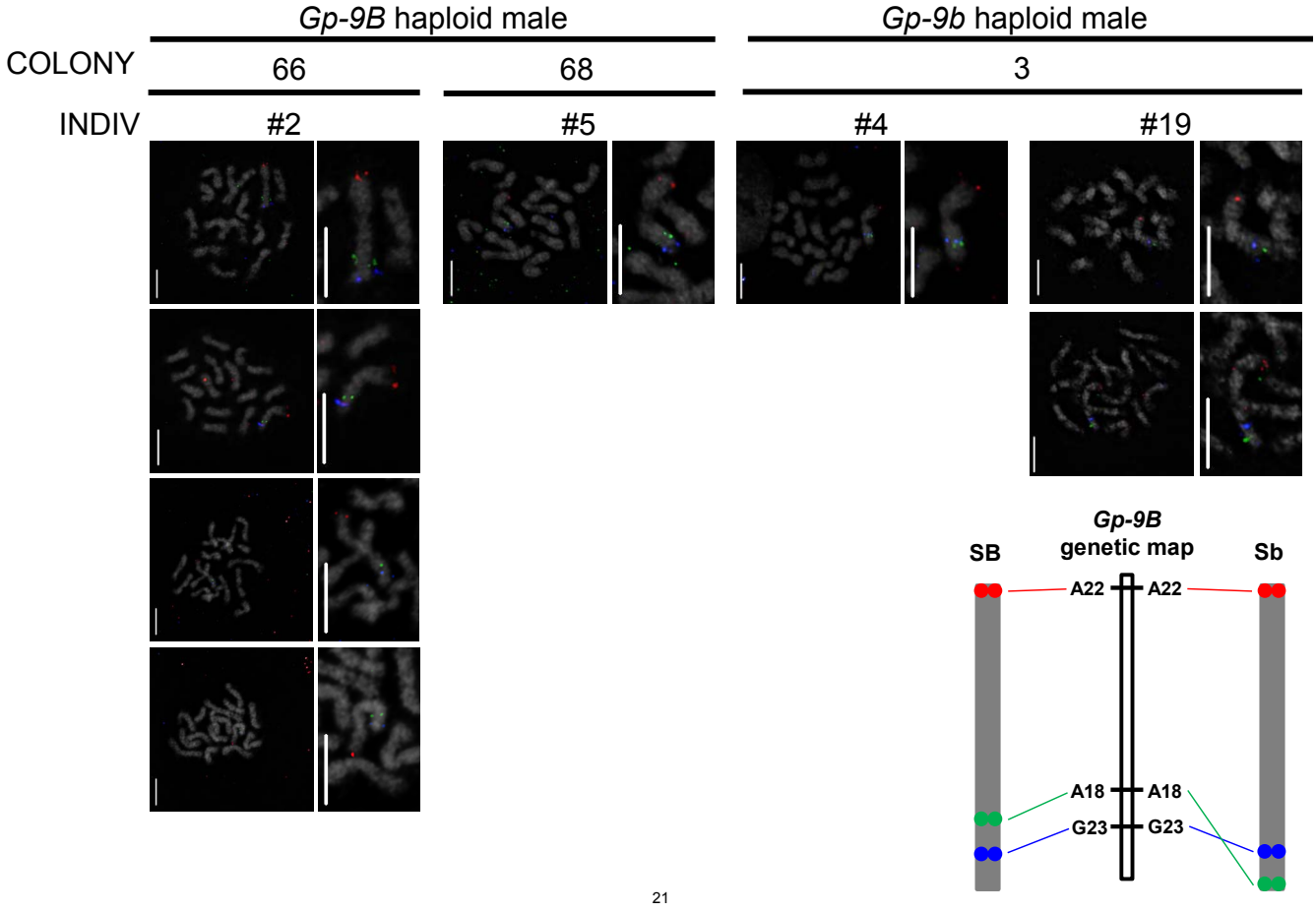


Gp-9b haploid male

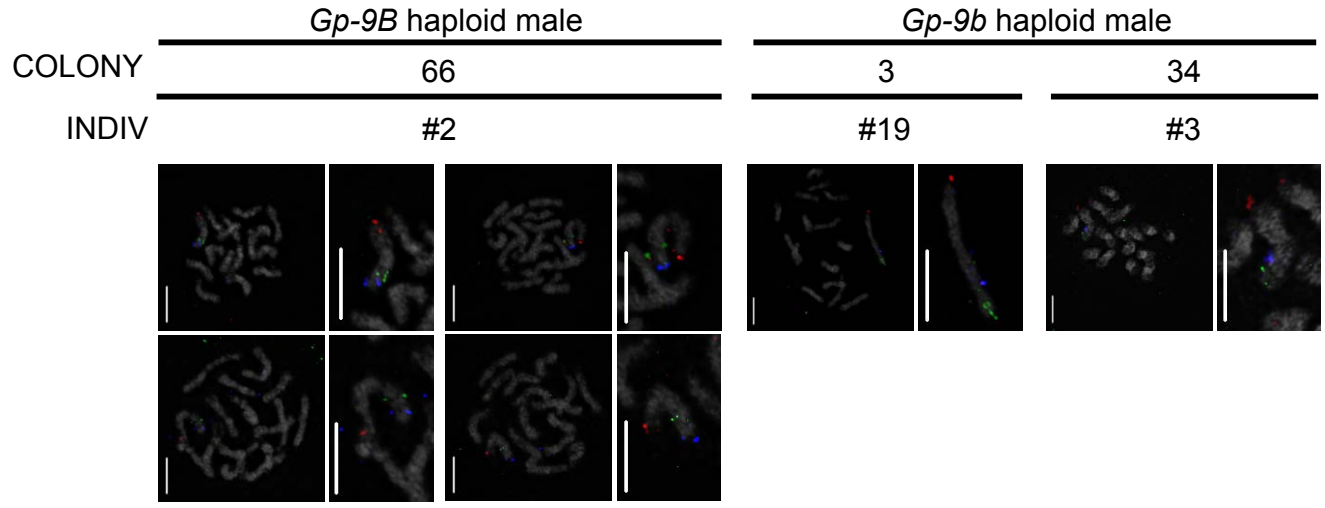


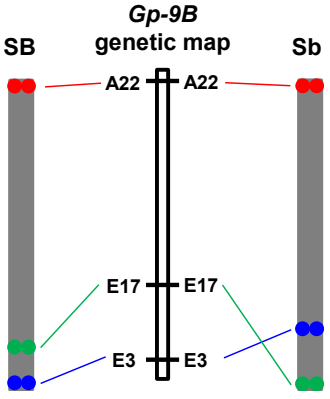
Supplementary Figure 8. BAC FISH showing complete chromosome complement. Magnified single chromosome images for Fig. 1b in main text are from these respective source cell images.

Sup Figure 9a

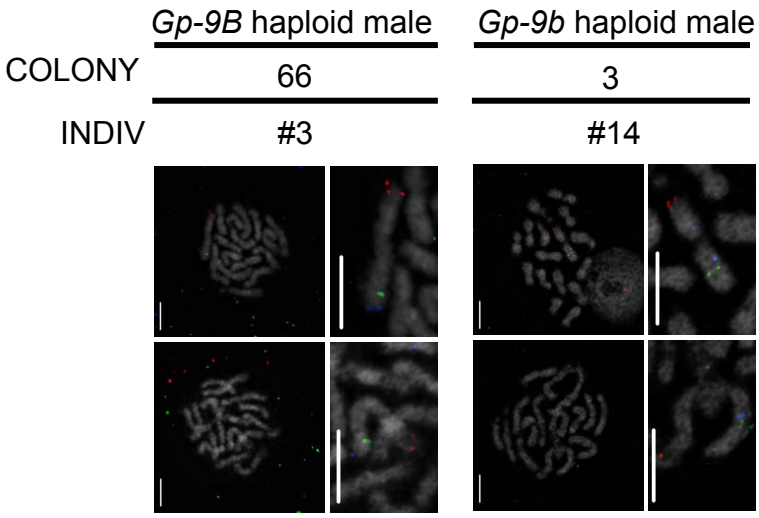


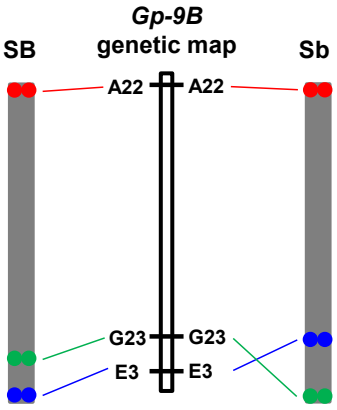
Sup Figure 9b



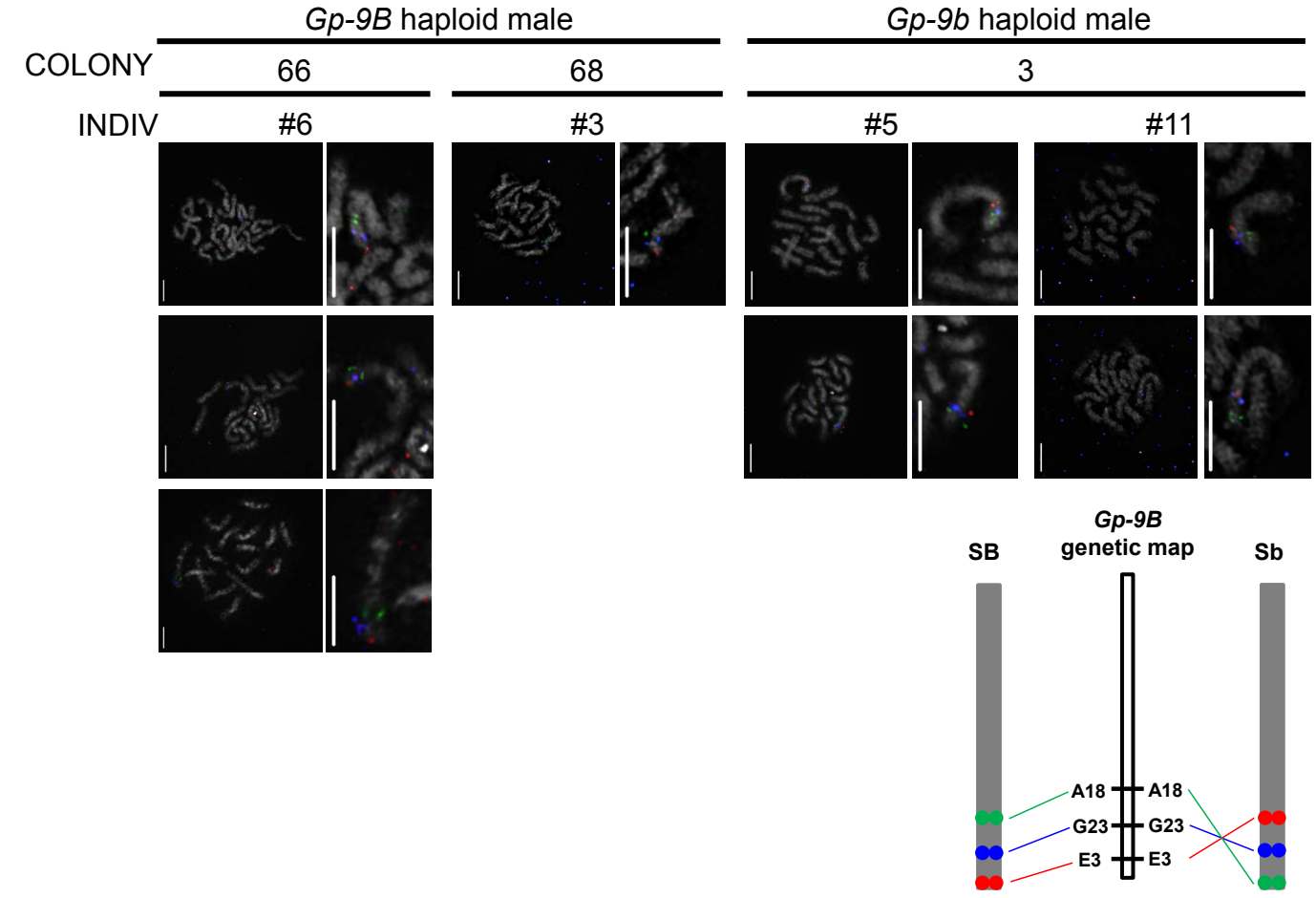


Sup Figure 9c

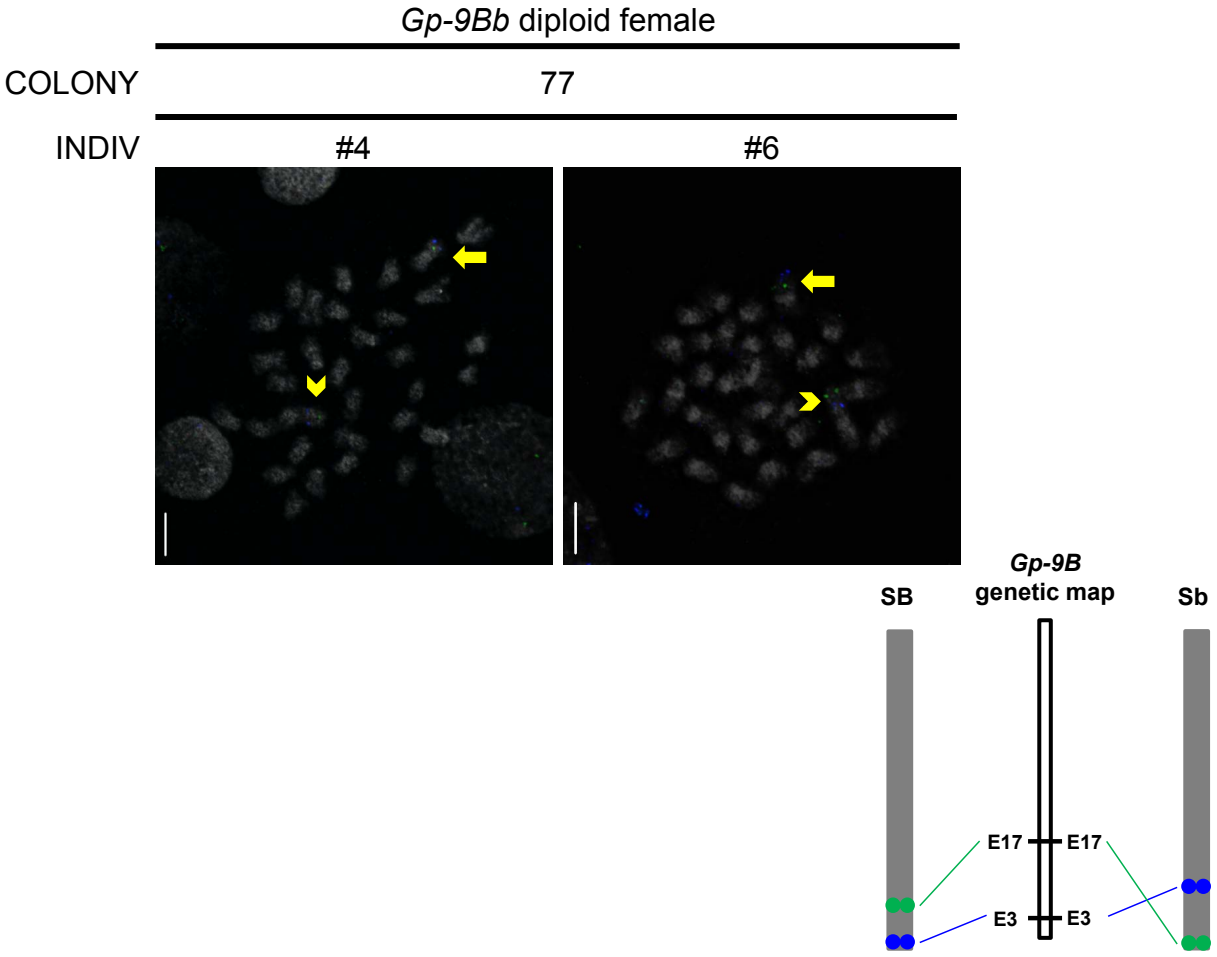




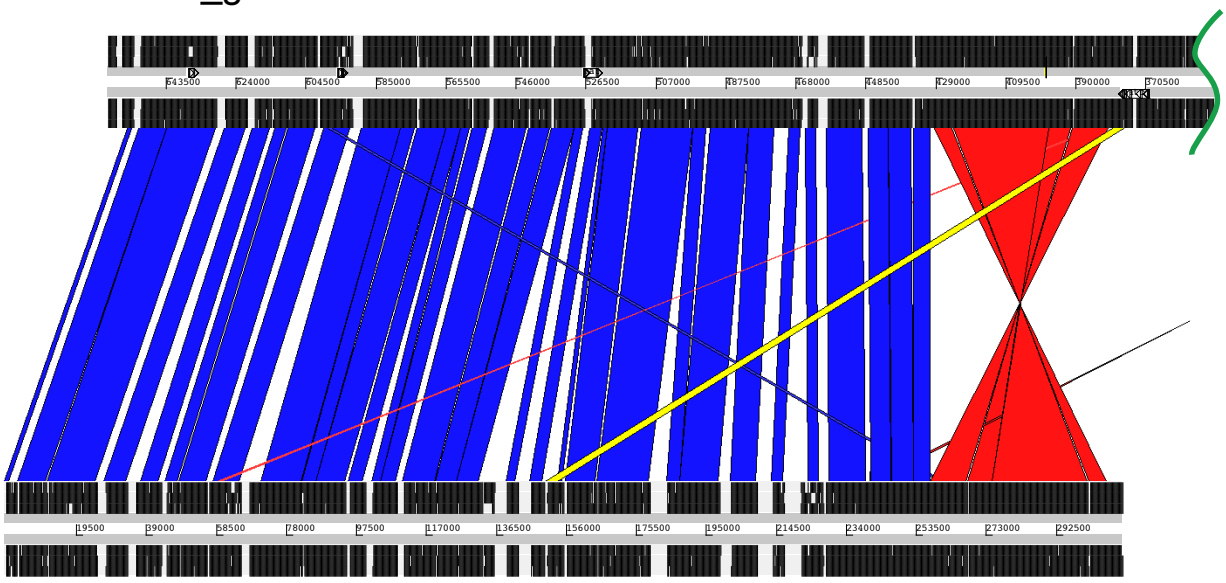
Sup Figure 9d



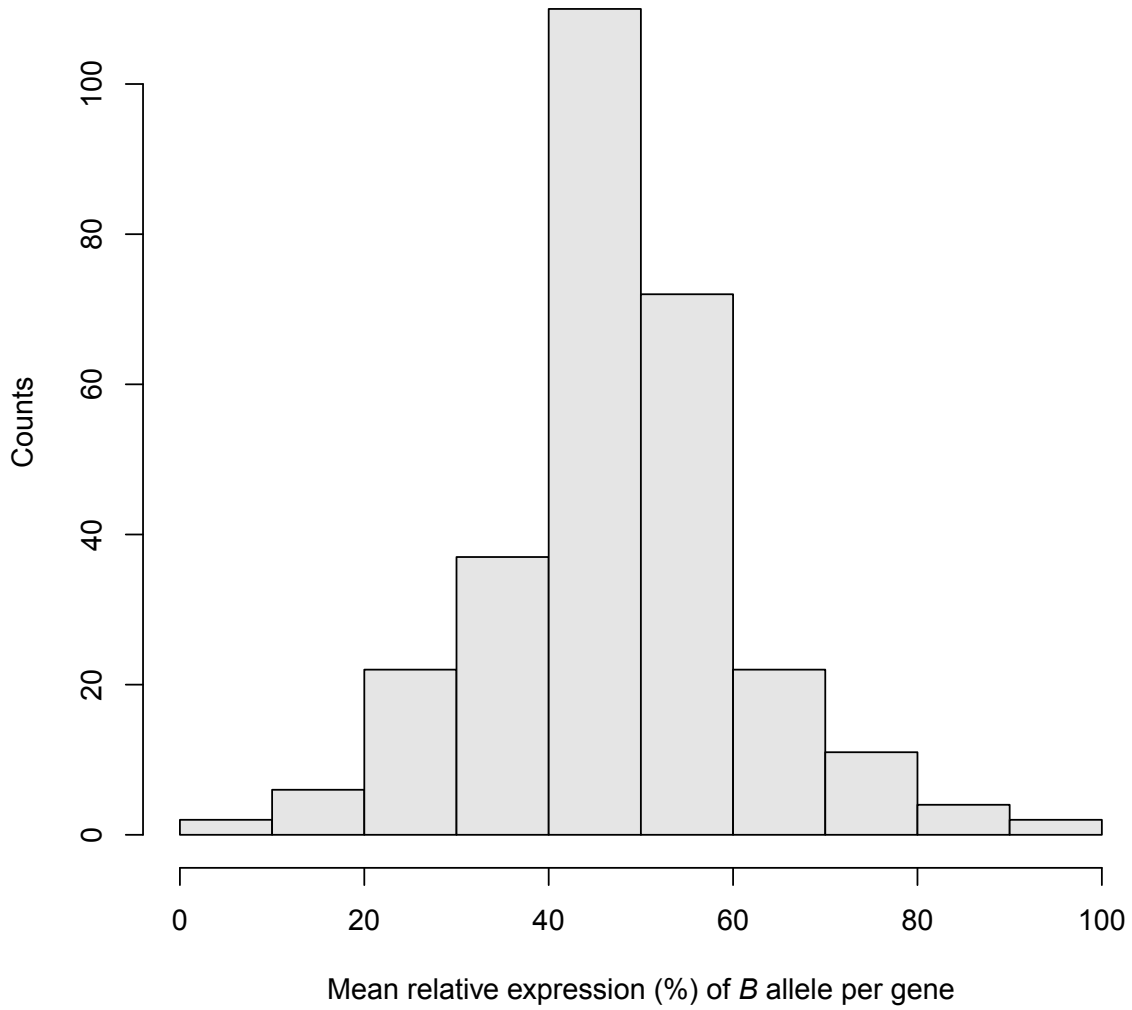
Sup Figure 9e



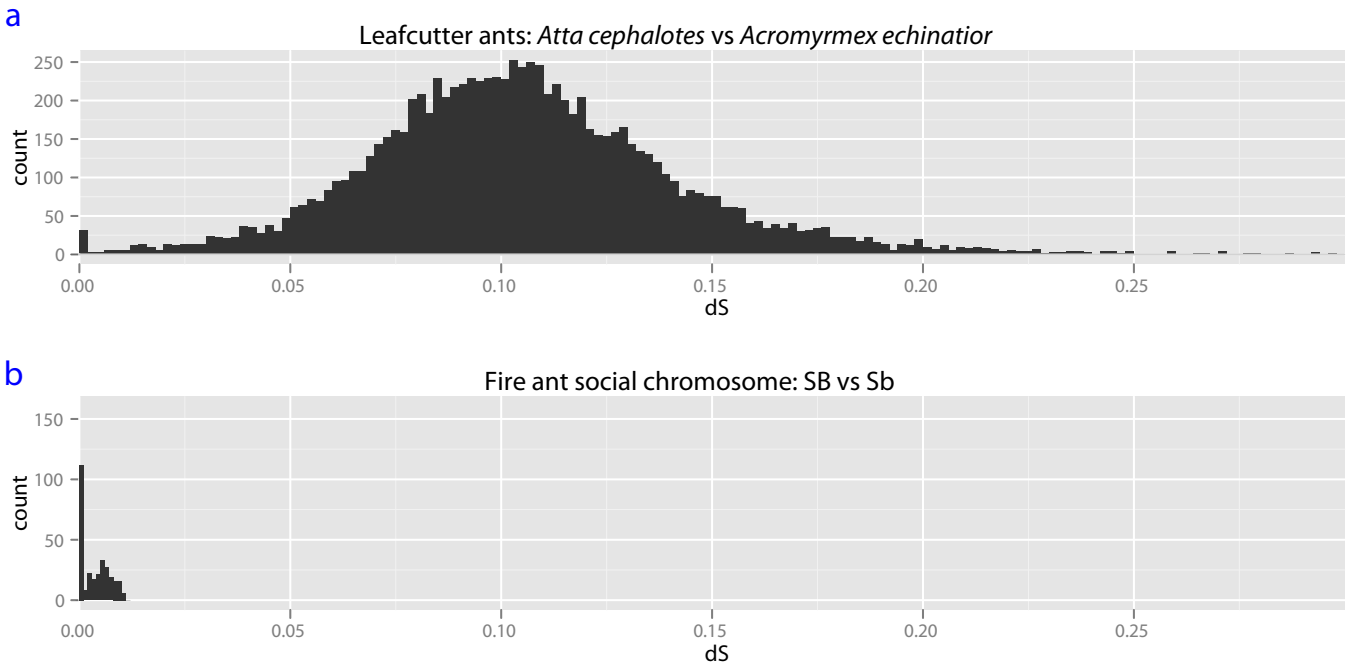
Supplementary Figure 9. BAC FISH using additional samples and probes.
a-d, Social chromosomes from *Gp-9B* haploid males (SB chromosome, left half) and *Gp-9b* haploid males (Sb chromosome, right half). For each cell sample, a pair of images are shown consisting of the complete chromosome set (left) and zoomed image of social chromosome (right). **e**, Full chromosome complement of single cells from *Gp-9Bb* diploid females (SB/Sb chromosomes) from a colony independent to that for the female in Figure 1c of the main text. Both orientations corresponding to SB (arrow) and Sb (chevron) can be observed. Note, only two probes used. For all subpanels, BAC probes, colonies, and individuals are as indicated as well as schematic interpretations of hybridization patterns and BAC positions on the SB genetic map. Chromosomes are counterstained with DAPI (white) and hybridized with fluorescently labeled BAC probes: A18, A22, E17, E03, and G23 as indicated.



Supplementary Figure 10. Structural rearrangement between the non-recombining SB (top) and Sb (bottom) regions. Red and blue ribbons indicate similarity; an inversion spanning at least 48 kb is highlighted in red.



Supplementary Figure 11. Distribution of the relative fraction of gene expression from *B*-specific alleles for 288 genes in *Gp-9Bb* queens



Supplementary Fig. 12. Distributions of frequencies of synonymous substitutions (dS) in a comparison of leafcutter ants (a) and genes on the non-recombining portion of the social chromosome (b)

Overview of Supplementary Tables

Supplementary Table 1. Summary of RADseq samples.

Supplementary Table 2. Short-range PCR assays in five families of four *Gp-9B* males and four *Gp-9b* males.

Supplementary Table 3. Genes significantly differentially expressed between *Gp-9BB* and *Bb* 1-day old queens and between *Gp-9B* and *b* male pupae.

Supplementary Table 4. Genes with dN/dS >1 between the non-recombining region of Sb and SB.

Supplementary Table 5. Differences in repeat content between *Gp-9b* and *Gp-9B* males.

Supplementary Table 6. Exons in SB missing from Sb.

Supplementary Table 7. Allele-specific expression differences for genes in the non-recombining region of the social chromosome in *Gp-9Bb* queens. [Separate file]

Supplementary Tables 8–14. RAD marker ID position information for the 7 families analyzed. [Separate files]

Supplementary Table 15. RAD primers used. [Separate file]

Supplementary Table 1. Summary of RADseq samples

					Males				
		Raw RAD tags			Number passing QC (starting)[†]				
	Raw Illumina reads	Total	Average per male			*Gp-9* genotype			Number of LGs[$]
Family[%]	x10[6]	x10[6]	x10[3]	Read length	Total	*B*	*b*	Number loci[&]	Main (+ minor)[#]
M013	153	122	1281	120	87 (95)	87 (95)	na	4983	16
M047	149	108	1135	120	62 (95)	62 (95)	na	5871	16 (+1)
M173	139	91	947	76	67 (96)	67 (96)	na	1992	16
P034	167	92	561	76	92 (110)	45 (54)	47 (56)	2796	16 (+1)
P008[*]	41	21	961	76	31 (38)	19 (25)	12 (13)	3614	33 (+7)
P016[*]	82	46	889	76	46 (48)	15	31 (33)	1318	23 (+10)
P033[*]	51	43	839	76	46 (48)	5	41 (43)	7360	22

[†] QC, quality control

[$] LG, linkage group

[%] M, monogyne; P, polygne social form from which queen originated

[&] Biallelic loci analyzed within each family present in ≥75% of males passing QC, except family P016 (≥60%)

[#] minor, orphan LGs composed of 1-3 RAD tag loci having the same genetic map position

[*] The number of main LGs >16 because the number of males was insufficient for better resolution

Supplementary Table 2. Short-range PCR assays in five families of four *Gp-9.B* males and four *Gp-9b* males.
Table cells show estimated PCR amplicon sizes or NA when no PCR product was obtained.
Alleles putatively specific to SB (italic) or Sb (bold) are indicated.

	Assay	Primer F	Primer R	Colony	Q2_P g346_ H	Q2_P g346_ H	Q2_P g346_ H	Q2_P g346_ H	Q2_P g346_ H	Q3_P g350_ H	Q3_P g350_ H	Q3_P g350_ H	Q3_P g350_ H	Q3_P g350_ H	Q7_3 29	Q7_3 29	Q7_3 29	Q7_3 29	Q4_9 7	Q4_9 7	Q4_9 7	Q4_9 7	Q7_3 45	Q7_3 45	Q7_3 45	Q7_3 45	Q2_Pg346 _H	Q2_Pg346 _H	Q2_Pg346 _H	Q2_Pg346 _H	Q3_Pg350 _H	Q3_Pg350 _H	Q3_Pg350 _H	Q3_Pg350 _H	Q7_329	Q7_329	Q7_329	Q7_329	Q4_97	Q4_97	Q4_97	Q4_97	Q7_345	Q7_345	Q7_345	Q7_345
				Individual	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male	Male 7	Male 8	Male 10	Male 11	Male 16	Male 18	Male 22	Male 23	Male 1	Male 2	Male 4	Male 5	Male 1	Male 2	Male 3	Male 5	Male 2	Male 3	Male 5	Male 7
				Gp-9 genotype	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b
Assays of insertions putatively specific to Sb	5	CGGATGCGTACGTGATCT	TCGCGAGTGTACGTCGAATA		*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	*243*	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**	**611**
	6	TCACCACACCATCACCCTTA	ATCGCAAGGTCTCGCTTCTA		*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	*631*	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**	**2400**
	8	ATGCGCATGCGTTCATATAG	CATTTGCTCTCGTGGGGTAT		*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	*212*	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**	**1287**
	9	TGCTGTTGCACCCAATACAT	GAAGGTCGAAAGAAGGAGCA		*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	*752*	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**	**1874**
	10	CGTATATGTCCCAGGCGTTC	TGTCCGGAAAAGAGATGTCC		*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	*316*	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**	**865**
	11	AGTGCTGCGTTTCGGTAACT	GGCGGAAAATAAATGTGTGG		*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	*853*	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**	**1423**
	12	GGATATGGGTGGAGGGAGTT	GAACACGGATTCTTAATGCAAG		*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	*467*	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**	**467 + 1400**
	13	TGAGCGCGACTTATATCGAC	TCACGACGCTACTTGTCCTG		*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	*288*	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**	**528**
	14	AGTTGGCAGTAATCGGTCGT	TGCTTACACAGAGCCGTCAT		*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	*430*	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**	**1586**
	15	GAAACGATCTCCGACGAAAA	CTTCGCGCTCTCGATCTACT		*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	*181*	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**	**400 + 500**
	16	ATGGGGAATCATCAGGATCT	TCCTAGAACGTGGAAAAAGCA		*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	*163*	**2158**	**2158**	**2158**	**2158**	**2158**	**2158**	**2158**	**NA**	**2158**	**2158**	**2158**	**2158**	**2158**	**NA**	**2158**	**NA**	**2158**	**2158**	**2158**	**2158**
	1	GAGCGAAAATTGAGCCTCTG	CGGTAAGACGTTTGCTCTCC		*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	*412*	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**	**1502**
	36	TAGTTCCCGACGTCGCTTC	CCATATATCGTTCCGGAGTGT		*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	*198*	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**	**650**
	37	TTGAAGCCTGTGGTAAAGACC	GTGATCGCCAAATTGTTACG		*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	*234*	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**
	38	TGTTCGTTATTGTTAAACTAGGA	TGAGAGAGAACGAAGTGAATCAG		*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	*229*	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**1800**	**1800**	**NA**	
	56	CCGGAAATAGTCGGGATACA	TTCCGTAAAACGCAAGGAAA		*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	*800*	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**
Assays spanning the edges of an inversion and transposition specific to Sb.	59	CCAAGGAAAGACGTGATGGT	GAAAAGGTGGCGGTTATGAA		*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**	**1200**
	44	GGTTTGCCGTAAGGAGATCA	GGCAGGAAAGCCCATTTAG		*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	*2500*	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**
	65	TTTTGCTTTGCCGAAAAGAT	GGAGCTTGCGGCTAACTATG		*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**	**1800**

Supplementary Table 3. Genes significantly differentially expressed between *Gp-9BB* and *Gp-9Bb* 1-day old queens and between *Gp-9B* and *Gp-9b* male pupae.

[#]log base 2; positive, gene expression in *Gp-9Bb* (or *Gp-9b*) > *Gp-9BB* (or *Gp-9B*); negative, *Gp-9BB* (or *Gp-9B*) > *Gp-9Bb* (or *Gp-9b*)

[†]FDR adjusted, 0.01 threshold for queen analysis, 0.05 for male analysis

[$]EST contigs from ref 28, some highly similar contigs collapsed

[&]un, unannotated gene, *i.e.*, not in the *S. invicta* official gene set 2.2.3; r, potential repeat sequence or gene with multiple copies in fire ant genome

[x]BlastX, except where noted; only hits better than 1e-5 are reported

[p]BlastP

[%]annotation from ref 29

1-day old queens							Similarity of *S. invicta* gene to UniProt	
microarray spot	fold change[#]	*P*-value[†]	EST assembly ID[$]	*S. invicta* OGS gene[&]	LG	E-value[x]	UniProt name	Description
SiJWF06ADC_pcr_1	0.67	7.93E-03	SI.CL.4.cl.412.SiJWF06ADC.scf	un	2	2E-15	E2BB27_HARSA	Four-domain proteases inhibitor
SiJWF03BAE_pcr_1	0.82	7.88E-03	SI.CL.1.cl.147.Contig1	un	2	2E-71	D7EKE8_TRICA	Uncharacterized protein
SiJWC10ADY_pcr_1	0.73	1.56E-03	SI.CL.22.cl.2279.Contig1	SI2.2.0_07124	3	2E-65	H9I0X0_ATTCE	Uncharacterized protein
SiJWC10ADY_pcr_newprimers	0.87	1.13E-03						
SiJWH06CAI_pcr_1	1.12	9.61E-06						
SiJWF12BBW_pcr_1	1.38	1.88E-06						
SiJWD11BDF_pcr_1	2.08	5.80E-08						
SiJWG09AAX_pcr_1	0.57	4.60E-03	SiJWG09AAX.scf	un	3			
SiJWA12ABC_pcr_2	0.81	1.69E-06	SI.CL.5.cl.576.SiJWA12ABC.scf	un	3			
SiJWG06ADO_pcr_1	1.62	5.81E-09	SI.CL.0.cl.070.Contig1	un	3	4E-33	E2BDZ6_HARSA	Uncharacterized protein
SiJWF06BAN_pcr_1	2.10	4.02E-11						
SiJWG06ADX_pcr_1	2.37	1.26E-14						
SiJWC07BAP_pcr_1	2.40	7.78E-15						
SiJWF02BAO_pcr_1	2.41	1.49E-16						
SiJWB05BDY_pcr_1	2.46	4.17E-15						
SiJWB10BAY_pcr_1	2.57	3.22E-18						
SiJWE02ACF_pcr_2	2.66	1.90E-16						
SiJWB10AAM_pcr_2	-0.78	9.18E-03	SI.CL.10.cl.1023.Contig1	un	6	3E-10	E2A6L5_CAMFO	Sequestosome-1
SiJWD04BCP_pcr_1	-0.78	4.22E-03						
SiJWD09BAZ_pcr_1	-0.82	4.25E-03						
SiJWE01ACW_pcr_2	-0.85	9.93E-03						
SiJWA02BDZ_pcr_1	-0.93	9.22E-04						
SiJWD06ABW_pcr_2	0.87	2.32E-03	SiJWD06ABW.scf	un	7	3E-20	E2C4N5_HARSA	Epidermal retinal dehydrogenase 2
SiJWA08BDL_pcr_1	-2.23	3.69E-05	SI.CL.7.cl.764.Contig1	un	7			
SiJWH01BCK_pcr_1	1.53	6.55E-09	SI.CL.23.cl.2354.SiJWH01BCK2.scf	SI2.2.0_02031	8	2E-20	E2AAL4_CAMFO	Uncharacterized protein
SiJWD02BCM_pcr_1	3.37	1.44E-17	SiJWD02BCM.scf	un	14			
SiJWF08AAP_pcr_2	1.16	1.60E-03	SI.CL.14.cl.1443.Contig1	un	S			
SiJWC05ACO_pcr_2	0.87	2.33E-04	SiJWC05ACO.scf	SI2.2.0_09037	S	1E-127	E2AR35_CAMFO	Growth arrest-specific protein 8
SiJWD08ABZ_pcr_2	0.47	2.16E-03	SI.CL.16.cl.1643.Contig1	SI2.2.0_10155	S	1E-147	F4X499_ACREC	Ubiquitin-like domain-containing CTD phosphatase 1
SiJWF04BEA_pcr_2	3.11	1.43E-11	SiJWF04BEA.scf	piggyBac[%]	S[%]	4E-17	Q75R41_BOMMO	piggyBac
SiJWF04BEA_pcr_1	3.30	2.11E-13						

SiJWE03ABS_pcr_2	0.54	1.60E-03	SI.CL.0.cl.054.SiJWE03ABS.scf	SI2.2.0_11174	na	4E-13	H9I2N6_ATTCE	Uncharacterized protein
SiJWG03BAD_pcr_1	0.83	4.00E-03	SI.CL.0.cl.065.Contig4	r	na	1E-119	E2AQR4_CAMFO	Uncharacterized protein
SiJWF10ABX_pcr_2	0.77	4.82E-03	SI.CL.4.cl.420.Contig2	r	na			
SiJWA08AAI_pcr_2	0.82	2.34E-04	SI.CL.4.cl.455.Contig1	SI2.2.0_06410	na			
SiJWA02ABO_pcr_2	0.56	1.71E-03	SI.CL.5.cl.576.Contig1	SI2.2.0_13469	na	2E-84	F4X7X5_ACREC	Pre-mRNA branch site p14-like protein
SiJWG07AAQ_pcr_2	0.70	4.61E-03	SI.CL.8.cl.873.Contig1	r	na	1E-104	E0VZF7_PEDHC	Enzymatic polyprotein, putative
SiJWC01AAP_pcr_2	0.72	3.86E-03	SI.CL.9.cl.934.Contig1	r	na	2E-16	E2BF11_HARSA	Fatty acid synthase
SiJWF05BAS_pcr_1	2.17	2.60E-08	SI.CL.11.cl.1171.Contig1	un	na	3E-42	E2ARQ4_CAMFO	Uncharacterized protein
SiJWC01AAO_pcr_2	2.65	5.00E-13						
SiJWE10ABJ_pcr_2	0.49	4.70E-03	SI.CL.14.cl.1437.Contig1	r	na			
SiJWD10AAP_pcr_2	0.59	4.70E-04						
SiJWE02AAJ_pcr_2	0.69	1.59E-03						
SiJWE01BCG_pcr_1	1.04	3.07E-07	SI.CL.23.cl.2386.Contig1	r	na			
SiJWE01BCG_pcr_1	1.04	3.07E-07						
SiJWD11ADN_pcr_1	1.10	7.49E-09						
SiJWF05BBF_pcr_1	1.46	7.48E-06	SI.CL.26.cl.2690.Contig1	BEL-PAO transposon[%]	na	6E-35	Q4JS97_ANOGA	BEL12_AG transposon polyprotein
SiJWA06BAO_pcr_1	1.48	1.69E-06						
SiJWF08BAR_pcr_1	0.87	2.82E-03	SI.CL.27.cl.2710.Contig1	r	na			
SiJWE04BBQ_pcr_1	1.51	1.31E-05	SI.CL.28.cl.2816.Contig1	r	na	2E-27	E2B0Z4_CAMFO	Putative nuclease HARBI1
SiJWG02BBR_pcr_1	1.68	1.89E-06						
SiJWH06BAP_pcr_1	1.89	1.69E-06						
SiJWC11BBQ_pcr_1	0.97	1.56E-03	SI.CL.31.cl.3138.Contig1	r	na			
SiJWC04BBM_pcr_1	1.37	1.38E-04						
SiJWB03BCD_pcr_1	0.68	6.24E-05	SI.CL.32.cl.3240.Contig1	r	na	1E-16	H9IBR5_ATTCE	Uncharacterized protein
SiJWF03BCX_pcr_1	1.02	8.85E-06	SI.CL.34.cl.3410.Contig1	r	na	4E-18	E2AIF5_CAMFO	Transposable element Tc3 transposase
SiJWD11BDA_pcr_1	-0.49	8.63E-04	SI.CL.37.cl.3709.Contig1	r	na			
SiJWD03AAK_pcr_2	0.65	1.42E-04	SiJWD03AAK.scf	SI2.2.0_09871	na			
SiJWB03ABO_pcr_2	0.90	7.25E-04	SiJWB03ABO.scf	r	na			
SiJWC05ADA_pcr_1	0.53	4.22E-03	SiJWC05ADA.scf	r	na	5E-36	D2CFX1_TRICA	Uncharacterized protein
SiJWC07ADW_pcr_1	1.08	9.96E-08	SiJWC07ADW.scf	un	na	6E-16	F4X5W3_ACREC	Fatty acid synthase
SiJWH05BDX_pcr_1	2.29	1.36E-15	SiJWH05BDX.scf	un	na			
SiJWC03CAW_pcr_1	1.44	8.39E-10	SiJWC03CAW.scf	un	na	2E-18	F4WVA3_ACREC	Putative deoxyribonuclease TATDN1

Male pupae							**Similarity of *S. invicta* gene to UniProt**	
microarray spot	**fold change[#]**	***P* -value[†]**	**EST assembly[$]**	***S. invicta* OGS gene[&]**	**LG**	**E-value[x]**	**UniProt name**	**Description**
SiJWG02AAK_pcr_2	3.56	4.15E-02	SI.CL.4.cl.412.Contig1	un	2	6E-16	F4X0U9_ACREC	Trypsin inhibitor
SiJWB04AAX_pcr_1	-2.14	2.12E-03	SiJWB04AAX.scf	SI2.2.0_09412	S	0[p]	RFWD2_MOUSE	E3 ubiquitin-protein ligase RFWD2
SiJWB04AAX_pcr_2	-1.80	3.49E-03						
SiJWA09BAD_pcr_1	1.73	1.07E-03	SI.CL.18.cl.1885.Contig1	SI2.2.0_13217	S			
SiJWF10CAF_pcr_1	1.85	4.10E-03						
SiJWA04ABN_pcr_2	1.53	2.93E-02						
SiJWC01ADY_pcr_newprimers	2.19	1.51E-05	SiJWC01ADY.scf	SI2.2.0_15655	S			
SiJWC01ADY_pcr_1	1.50	4.53E-03						
SiJWF03BCX_pcr_1	3.58	2.45E-04	SI.CL.34.cl.3410.Contig1	r	na	1E-17	E2AIF5_CAMFO	Tc3 transposase

Supplementary Table 4. Genes with dN/dS >1 between the non-recombining region of Sb and SB

					Uniprot Swissprot similarity				Probability >0.99 that expression level in males is				
									higher than[#]:			lower than[#]:	
Gene id	dN/dS	dN	dS	scaffold	E-value	ID	long name		in Workers	in Queens		in Workers	in Queens
SI2.2.0_00181	1.0346	0.0061	0.0059	Si_gnF.scaffold00899	1.00E-06	ZF161_HUMAN	Zinc finger protein 161 homolog		unknown	unknown		unknown	unknown
SI2.2.0_00539	2.6127	0.0381	0.0146	Si_gnF.scaffold00899	NA	NA	NA		unknown	unknown		unknown	unknown
SI2.2.0_02112	1.0416	0.0023	0.0022	Si_gnF.scaffold05266	2.00E-16	NACH_DROME	Sodium channel protein Nach		-	P		A	-
SI2.2.0_02194	1.2148	0.0053	0.0044	Si_gnF.scaffold00255	4.00E-114	VASA1_DROME	ATP-dependent RNA helicase vasa, isoform A		unknown	unknown		unknown	unknown
SI2.2.0_03135	1.759	0.0106	0.006	Si_gnF.scaffold06914	6.00E-19	TTPAL_HUMAN	Alpha-tocopherol transfer protein-like		P	A,P,L		-	-
SI2.2.0_03570	1.9623	0.0254	0.013	Si_gnF.scaffold00690	2.00E-106	CP4C1_BLADI	Cytochrome P450 4C1		unknown	unknown		unknown	unknown
SI2.2.0_04915	2.2557	0.0105	0.0046	Si_gnF.scaffold06914	NA	NA	NA		P	-		-	-
SI2.2.0_05650	1.6115	0.0049	0.0031	Si_gnF.scaffold00899	NA	NA	NA		unknown	unknown		unknown	unknown
SI2.2.0_08197	1.1238	0.0034	0.0031	Si_gnF.scaffold00690	NA	NA	NA		-	-		-	P
SI2.2.0_09963	1.0814	0.0216	0.02	Si_gnF.scaffold00690	1.00E-12	NRF6_CAEEL	Nose resistant to fluoxetine protein 6		unknown	unknown		unknown	unknown
SI2.2.0_10460	1.0215	0.0052	0.0051	Si_gnF.scaffold00690	2.00E-06	CUO6_BLACR	Cuticle protein 6		A	A		P	P
SI2.2.0_11401	1.0903	0.002	0.0018	Si_gnF.scaffold00899	NA	NA	NA		A	P		-	-
SI2.2.0_15037	1.0344	0.0071	0.0068	Si_gnF.scaffold01573	NA	NA	NA		A	-		-	-
SI2.2.0_15786	1.1068	0.0036	0.0032	Si_gnF.scaffold05266	9.00E-28	XRCC5_MOUSE	X-ray repair cross-complementing protein 5		unknown	unknown		unknown	unknown
SI2.2.0_80708	2.1503	0.0052	0.0024	Si_gnF.scaffold03404	NA	NA	NA		-	-		A,P,L	L

[#]Expression comparisons are from Ometto *et al* 2011[44]. Gene expression comparisons were performed in adults (A), pupae (P), larvae (L); "-" indicates a probability of differential expression < 0.99; "unknown" indicates no available data.

Supplementary Table 5. Differences in repeat content between *Gp-9b* and *Gp-9B* males.

Repeat identifier	Median ratio of number of reads from *Gp-9b* to number of reads from *Gp-9B*	pvalues.raw	pvalues.fdr
sinv_100.46.aligned.centroid_PILER	1.08	0.001807	0.016197
sinv_11.144.aligned.centroid_PILER	1.39	0.000725	0.008545
sinv_112.34.aligned.centroid_PILER	1.14	0.000059	0.001726
sinv_114.31.aligned.centroid_PILER	1.08	0.003084	0.023968
sinv_122.18.aligned.centroid_PILER	1.18	0.000221	0.003833
sinv_124.16.aligned.centroid_PILER	1.11	0.000777	0.008937
sinv_125.15.aligned.centroid_PILER	1.04	0.008678	0.049761
sinv_126.14.aligned.centroid_PILER	1.06	0.005120	0.034752
sinv_13.142.aligned.centroid_PILER	0.87	0.000218	0.003833
sinv_131.9.aligned.centroid_PILER	1.15	0.001858	0.016454
sinv_136.3.aligned.centroid_PILER	1.87	0.000051	0.001540
sinv_147.163.aligned.centroid_PILER	0.86	0.007723	0.046057
sinv_17.138.aligned.centroid_PILER	1.15	0.000287	0.004457
sinv_24.131.aligned.centroid_PILER	4.23	0.000000	0.000084
sinv_27.128.aligned.centroid_PILER	1.44	0.000017	0.000889
sinv_33.122.aligned.centroid_PILER	1.13	0.000050	0.001540
sinv_34.121.aligned.centroid_PILER	1.27	0.000003	0.000500
sinv_4.151.aligned.centroid_PILER	1.07	0.003530	0.026511
sinv_57.94.aligned.centroid_PILER	1.13	0.000304	0.004527
sinv_63.86.aligned.centroid_PILER	1.20	0.001329	0.012861
sinv_67.82.aligned.centroid_PILER	0.87	0.001829	0.016294
sinv_68.81.aligned.centroid_PILER	1.43	0.002974	0.023362
sinv_81.68.aligned.centroid_PILER	1.19	0.000003	0.000500
sinv_87.61.aligned.centroid_PILER	1.82	0.000190	0.003424
sinv_88.59.aligned.centroid_PILER	1.21	0.000042	0.001491
sinv_99.47.aligned.centroid_PILER	1.20	0.000227	0.003855
sinv_rnd-1_family-0_RepeatScout	1.06	0.000467	0.006257
sinv_rnd-1_family-100_RepeatScout	1.10	0.005659	0.037583
sinv_rnd-1_family-108_RepeatScout	1.08	0.000048	0.001540
sinv_rnd-1_family-120_RepeatScout	1.08	0.003650	0.027092
sinv_rnd-1_family-136_RepeatScout	1.07	0.004447	0.031364
sinv_rnd-1_family-14_RepeatScout	1.07	0.000044	0.001501
sinv_rnd-1_family-156_RepeatScout	1.12	0.000496	0.006415
sinv_rnd-1_family-164_RepeatScout	1.07	0.006571	0.041027
sinv_rnd-1_family-178_RepeatScout	1.15	0.000006	0.000533
sinv_rnd-1_family-199_RepeatScout	1.16	0.001171	0.011640
sinv_rnd-1_family-204_RepeatScout	1.14	0.008163	0.047897
sinv_rnd-1_family-212_RepeatScout	1.12	0.003538	0.026511
sinv_rnd-1_family-217_RepeatScout	1.09	0.001136	0.011526
sinv_rnd-1_family-230_RepeatScout	1.11	0.000003	0.000500
sinv_rnd-1_family-23_RepeatScout	1.06	0.004068	0.029279
sinv_rnd-1_family-241_RepeatScout	1.16	0.004793	0.033028
sinv_rnd-1_family-250_RepeatScout	1.04	0.000311	0.004561
sinv_rnd-1_family-260_RepeatScout	1.11	0.000266	0.004190
sinv_rnd-1_family-268_RepeatScout	1.10	0.000012	0.000768

sinv_rnd-1_family-271_RepeatScout	1.04	0.000938	0.010223
sinv_rnd-1_family-280_RepeatScout	1.15	0.007307	0.044116
sinv_rnd-1_family-290_RepeatScout	1.10	0.005521	0.036831
sinv_rnd-1_family-315_RepeatScout	1.13	0.000366	0.005182
sinv_rnd-1_family-324_RepeatScout	1.10	0.001257	0.012407
sinv_rnd-1_family-331_RepeatScout	1.04	0.005832	0.037700
sinv_rnd-1_family-335_RepeatScout	1.05	0.001078	0.011091
sinv_rnd-1_family-337_RepeatScout	1.11	0.005713	0.037589
sinv_rnd-1_family-345_RepeatScout	1.21	0.000552	0.007078
sinv_rnd-1_family-347_RepeatScout	1.11	0.001494	0.013994
sinv_rnd-1_family-35_RepeatScout	1.05	0.006840	0.042109
sinv_rnd-1_family-360_RepeatScout	1.11	0.000086	0.002092
sinv_rnd-1_family-369_RepeatScout	1.13	0.002490	0.019985
sinv_rnd-1_family-392_RepeatScout	1.07	0.001359	0.012896
sinv_rnd-1_family-417_RepeatScout	1.23	0.000137	0.002820
sinv_rnd-1_family-41_RepeatScout	1.04	0.001924	0.016736
sinv_rnd-1_family-422_RepeatScout	1.07	0.001277	0.012517
sinv_rnd-1_family-427_RepeatScout	1.11	0.000157	0.003010
sinv_rnd-1_family-429_RepeatScout	0.94	0.001030	0.010989
sinv_rnd-1_family-436_RepeatScout	1.16	0.000022	0.000975
sinv_rnd-1_family-443_RepeatScout	1.18	0.001966	0.016897
sinv_rnd-1_family-444_RepeatScout	1.09	0.000623	0.007651
sinv_rnd-1_family-44_RepeatScout	1.11	0.000218	0.003833
sinv_rnd-1_family-465_RepeatScout	1.13	0.001305	0.012711
sinv_rnd-1_family-481_RepeatScout	1.05	0.006617	0.041139
sinv_rnd-1_family-486_RepeatScout	1.16	0.002418	0.019517
sinv_rnd-1_family-495_RepeatScout	1.17	0.006413	0.040213
sinv_rnd-1_family-496_RepeatScout	1.16	0.002814	0.022225
sinv_rnd-1_family-500_RepeatScout	1.15	0.002261	0.018458
sinv_rnd-1_family-506_RepeatScout	1.15	0.004254	0.030317
sinv_rnd-1_family-512_RepeatScout	1.03	0.007777	0.046191
sinv_rnd-1_family-518_RepeatScout	1.25	0.000016	0.000851
sinv_rnd-1_family-519_RepeatScout	1.07	0.004465	0.031364
sinv_rnd-1_family-523_RepeatScout	1.15	0.000135	0.002811
sinv_rnd-1_family-531_RepeatScout	1.23	0.001338	0.012861
sinv_rnd-1_family-533_RepeatScout	1.11	0.000711	0.008516
sinv_rnd-1_family-535_RepeatScout	1.18	0.002156	0.017795
sinv_rnd-1_family-54_RepeatScout	1.06	0.001154	0.011626
sinv_rnd-1_family-575_RepeatScout	1.16	0.002009	0.017165
sinv_rnd-1_family-578_RepeatScout	1.17	0.000741	0.008662
sinv_rnd-1_family-580_RepeatScout	1.28	0.000112	0.002510
sinv_rnd-1_family-581_RepeatScout	1.17	0.000369	0.005182
sinv_rnd-1_family-591_RepeatScout	1.08	0.004850	0.033269
sinv_rnd-1_family-593_RepeatScout	1.15	0.000116	0.002552
sinv_rnd-1_family-604_RepeatScout	1.08	0.006860	0.042109
sinv_rnd-1_family-608_RepeatScout	2.15	0.000000	0.000123
sinv_rnd-1_family-609_RepeatScout	1.28	0.000097	0.002251
sinv_rnd-1_family-612_RepeatScout	1.20	0.000068	0.001812
sinv_rnd-1_family-613_RepeatScout	1.12	0.003286	0.025003
sinv_rnd-1_family-615_RepeatScout	1.19	0.000481	0.006377
sinv_rnd-1_family-62_RepeatScout	1.06	0.003653	0.027092
sinv_rnd-1_family-632_RepeatScout	1.36	0.000052	0.001540

sinv_rnd-1_family-635_RepeatScout	1.16	0.000237	0.003942
sinv_rnd-1_family-636_RepeatScout	1.14	0.001169	0.011640
sinv_rnd-1_family-637_RepeatScout	1.18	0.001059	0.011025
sinv_rnd-1_family-649_RepeatScout	1.15	0.000006	0.000533
sinv_rnd-1_family-64_RepeatScout	1.10	0.000860	0.009592
sinv_rnd-1_family-661_RepeatScout	1.17	0.001799	0.016197
sinv_rnd-1_family-665_RepeatScout	1.13	0.000678	0.008192
sinv_rnd-1_family-690_RepeatScout	1.21	0.002725	0.021755
sinv_rnd-1_family-692_RepeatScout	1.27	0.000034	0.001253
sinv_rnd-1_family-697_RepeatScout	1.21	0.000142	0.002847
sinv_rnd-1_family-700_RepeatScout	1.07	0.002067	0.017555
sinv_rnd-1_family-704_RepeatScout	1.29	0.000001	0.000278
sinv_rnd-1_family-70_RepeatScout	1.04	0.001708	0.015501
sinv_rnd-1_family-715_RepeatScout	1.28	0.000002	0.000487
sinv_rnd-1_family-723_RepeatScout	1.22	0.000030	0.001167
sinv_rnd-1_family-730_RepeatScout	1.14	0.003159	0.024163
sinv_rnd-1_family-747_RepeatScout	1.25	0.000010	0.000722
sinv_rnd-1_family-76_RepeatScout	1.14	0.001664	0.015234
sinv_rnd-1_family-86_RepeatScout	1.22	0.000000	0.000216
sinv_rnd-1_family-87_RepeatScout	1.12	0.000020	0.000944
sinv_rnd-1_family-88_RepeatScout	1.10	0.000667	0.008116
sinv_rnd-1_family-90_RepeatScout	1.07	0.005467	0.036636
sinv_rnd-1_family-91_RepeatScout	1.08	0.005077	0.034659
sinv_rnd-1_family-93_RepeatScout	1.07	0.005318	0.035807
sinv_rnd-1_family-94_RepeatScout	1.04	0.004376	0.031034
sinv_rnd-1_family-9_RepeatScout	1.06	0.003124	0.024022
sinv_rnd-4_family-1165_Recon	1.08	0.007852	0.046446
sinv_rnd-4_family-1621_Recon	1.18	0.000012	0.000768
sinv_rnd-4_family-19_Recon	1.07	0.000568	0.007094
sinv_rnd-4_family-865_Recon	1.10	0.005930	0.037830
sinv_rnd-5_family-1000_Recon	1.11	0.008437	0.049109
sinv_rnd-5_family-1114_Recon	1.29	0.001044	0.010993
sinv_rnd-5_family-1122_Recon	1.15	0.000930	0.010216
sinv_rnd-5_family-124_Recon	1.11	0.000399	0.005453
sinv_rnd-5_family-126_Recon	1.12	0.000083	0.002044
sinv_rnd-5_family-1380_Recon	1.12	0.004566	0.031764
sinv_rnd-5_family-138_Recon	1.11	0.002168	0.017795
sinv_rnd-5_family-1392_Recon	1.19	0.000064	0.001798
sinv_rnd-5_family-150_Recon	1.10	0.000033	0.001239
sinv_rnd-5_family-1528_Recon	1.26	0.000030	0.001167
sinv_rnd-5_family-171_Recon	1.25	0.000001	0.000281
sinv_rnd-5_family-173_Recon	1.15	0.000154	0.002997
sinv_rnd-5_family-176_Recon	1.15	0.002108	0.017598
sinv_rnd-5_family-1813_Recon	1.17	0.000005	0.000533
sinv_rnd-5_family-1919_Recon	1.24	0.000005	0.000533
sinv_rnd-5_family-1934_Recon	1.13	0.003805	0.027899
sinv_rnd-5_family-1999_Recon	1.14	0.000448	0.006067
sinv_rnd-5_family-2002_Recon	1.16	0.000485	0.006377
sinv_rnd-5_family-201_Recon	1.15	0.005690	0.037589
sinv_rnd-5_family-203_Recon	1.09	0.001046	0.010993
sinv_rnd-5_family-2187_Recon	1.16	0.000246	0.003994
sinv_rnd-5_family-2226_Recon	0.94	0.001064	0.011025

sinv_rnd-5_family-2228_Recon	1.37	0.006646	0.041141
sinv_rnd-5_family-2307_Recon	1.29	0.000016	0.000851
sinv_rnd-5_family-2421_Recon	1.43	0.000013	0.000821
sinv_rnd-5_family-2499_Recon	1.14	0.005747	0.037589
sinv_rnd-5_family-250_Recon	1.31	0.002098	0.017598
sinv_rnd-5_family-255_Recon	1.21	0.000787	0.008986
sinv_rnd-5_family-270_Recon	1.08	0.000166	0.003107
sinv_rnd-5_family-2714_Recon	1.26	0.000185	0.003376
sinv_rnd-5_family-2753_Recon	1.16	0.000251	0.004033
sinv_rnd-5_family-2778_Recon	1.11	0.005763	0.037589
sinv_rnd-5_family-280_Recon	1.23	0.000235	0.003942
sinv_rnd-5_family-2929_Recon	1.16	0.004022	0.029123
sinv_rnd-5_family-2943_Recon	1.07	0.003384	0.025620
sinv_rnd-5_family-2950_Recon	1.22	0.000562	0.007078
sinv_rnd-5_family-2964_Recon	0.95	0.008685	0.049761
sinv_rnd-5_family-3112_Recon	1.11	0.004790	0.033028
sinv_rnd-5_family-3242_Recon	1.18	0.000242	0.003965
sinv_rnd-5_family-3361_Recon	1.16	0.000772	0.008937
sinv_rnd-5_family-341_Recon	1.09	0.004501	0.031467
sinv_rnd-5_family-3466_Recon	1.16	0.001550	0.014428
sinv_rnd-5_family-424_Recon	1.17	0.003115	0.024022
sinv_rnd-5_family-43_Recon	1.18	0.000223	0.003833
sinv_rnd-5_family-4657_Recon	1.16	0.000117	0.002552
sinv_rnd-5_family-46_Recon	1.16	0.001494	0.013994
sinv_rnd-5_family-559_Recon	1.13	0.000050	0.001540
sinv_rnd-5_family-572_Recon	1.13	0.001123	0.011478
sinv_rnd-5_family-573_Recon	1.13	0.000139	0.002823
sinv_rnd-5_family-582_Recon	1.17	0.000066	0.001798
sinv_rnd-5_family-6007_Recon	1.18	0.000363	0.005182
sinv_rnd-5_family-608_Recon	1.05	0.006112	0.038656
sinv_rnd-5_family-6488_Recon	1.07	0.006020	0.038242
sinv_rnd-5_family-650_Recon	1.23	0.002282	0.018524
sinv_rnd-5_family-685_Recon	1.08	0.003819	0.027899
sinv_rnd-5_family-724_Recon	1.17	0.001355	0.012896
sinv_rnd-5_family-738_Recon	1.07	0.005138	0.034752
sinv_rnd-5_family-8274_Recon	1.17	0.000293	0.004461
sinv_rnd-5_family-8296_Recon	1.13	0.000290	0.004457
sinv_rnd-5_family-994_Recon	1.14	0.000398	0.005453
sinv_rnd-6_family-106_Recon	1.12	0.001631	0.015078
sinv_rnd-6_family-107_Recon	1.12	0.003009	0.023512
sinv_rnd-6_family-1120_Recon	1.21	0.000078	0.001955
sinv_rnd-6_family-11591_Recon	1.17	0.000050	0.001540
sinv_rnd-6_family-1255_Recon	1.06	0.002783	0.022095
sinv_rnd-6_family-1426_Recon	1.31	0.000170	0.003153
sinv_rnd-6_family-1449_Recon	1.20	0.000165	0.003107
sinv_rnd-6_family-1538_Recon	1.35	0.000894	0.009899
sinv_rnd-6_family-176_Recon	1.13	0.000021	0.000944
sinv_rnd-6_family-190_Recon	1.18	0.001920	0.016736
sinv_rnd-6_family-1925_Recon	1.10	0.002148	0.017795
sinv_rnd-6_family-2113_Recon	1.22	0.000028	0.001147
sinv_rnd-6_family-2129_Recon	1.14	0.004027	0.029123
sinv_rnd-6_family-216_Recon	1.04	0.008504	0.049272

sinv_rnd-6_family-225_Recon	1.33	0.000015	0.000851
sinv_rnd-6_family-2296_Recon	1.23	0.000025	0.001058
sinv_rnd-6_family-2310_Recon	1.25	0.000298	0.004495
sinv_rnd-6_family-231_Recon	1.67	0.000064	0.001798
sinv_rnd-6_family-258_Recon	1.07	0.000152	0.002997
sinv_rnd-6_family-2608_Recon	1.16	0.000023	0.001005
sinv_rnd-6_family-2619_Recon	1.31	0.005866	0.037757
sinv_rnd-6_family-2623_Recon	0.94	0.003788	0.027899
sinv_rnd-6_family-2683_Recon	1.23	0.000559	0.007078
sinv_rnd-6_family-282_Recon	1.11	0.000002	0.000487
sinv_rnd-6_family-2962_Recon	1.38	0.000075	0.001921
sinv_rnd-6_family-3056_Recon	1.19	0.000312	0.004561
sinv_rnd-6_family-318_Recon	1.17	0.000385	0.005356
sinv_rnd-6_family-3257_Recon	1.39	0.000105	0.002391
sinv_rnd-6_family-3260_Recon	1.30	0.000124	0.002641
sinv_rnd-6_family-3283_Recon	1.18	0.000007	0.000546
sinv_rnd-6_family-3356_Recon	1.28	0.006963	0.042562
sinv_rnd-6_family-3361_Recon	1.17	0.000004	0.000500
sinv_rnd-6_family-3480_Recon	1.19	0.000090	0.002127
sinv_rnd-6_family-3597_Recon	1.26	0.000090	0.002127
sinv_rnd-6_family-3610_Recon	1.22	0.000007	0.000546
sinv_rnd-6_family-3697_Recon	1.17	0.000346	0.005008
sinv_rnd-6_family-402_Recon	1.04	0.007304	0.044116
sinv_rnd-6_family-4092_Recon	1.32	0.000619	0.007651
sinv_rnd-6_family-4488_Recon	1.31	0.000006	0.000533
sinv_rnd-6_family-4571_Recon	1.22	0.001668	0.015234
sinv_rnd-6_family-4758_Recon	1.07	0.007441	0.044556
sinv_rnd-6_family-4787_Recon	1.05	0.005908	0.037830
sinv_rnd-6_family-4813_Recon	1.12	0.001881	0.016557
sinv_rnd-6_family-5261_Recon	1.11	0.001962	0.016897
sinv_rnd-6_family-5357_Recon	1.21	0.000073	0.001913
sinv_rnd-6_family-5619_Recon	1.26	0.000961	0.010405
sinv_rnd-6_family-5686_Recon	1.12	0.001016	0.010915
sinv_rnd-6_family-5820_Recon	1.18	0.000016	0.000851
sinv_rnd-6_family-6567_Recon	1.61	0.000018	0.000889
sinv_rnd-6_family-6878_Recon	1.06	0.006236	0.039274
sinv_rnd-6_family-6951_Recon	0.89	0.007130	0.043405
sinv_rnd-6_family-72_Recon	1.13	0.000044	0.001501
sinv_rnd-6_family-7332_Recon	1.15	0.002093	0.017598
sinv_rnd-6_family-780_Recon	1.16	0.000798	0.009037
sinv_rnd-6_family-914_Recon	1.15	0.007425	0.044556
sinv_rnd-6_family-9190_Recon	1.35	0.000004	0.000500
sinv_rnd-6_family-960_Recon	1.15	0.000721	0.008545
sinv_rnd-6_family-985_Recon	1.10	0.008532	0.049272
sinv_rnd-6_family-996_Recon	0.95	0.008014	0.047211

Supplementary Table 6. Exons in SB missing from Sb

Missing exon				Similarity to UniProt gene			
Gene ID	Genomic scaffold	Exon start	Exon end	Accession	blastp E-value	Name	Protein name
SI2.2.0_09248	Si_gnF.scaffold00255	396674	396992	P09052	4.00E-112	VASA1_DROME	ATP-dependent RNA helicase vasa, isoform A
SI2.2.0_00913	Si_gnF.scaffold06568	43720	43540	NA	NA	NA	NA
		43123	42805	NA	NA	NA	
SI2.2.0_07415	Si_gnF.scaffold06568	36319	36172	Q27594	4.00E-32	CP6A9_DROME	Cytochrome P450 6a9
		35617	35290				
SI2.2.0_03123	Si_gnF.scaffold06568	28183	28009	Q27902	2.00E-15	CP6B4_PAPGL	Cytochrome P450 6B4
		27949	27739				
SI2.2.0_11320	Si_gnF.scaffold07090	587420	587291	Q8T674	2.00E-68	NA	NA
		586436	586229				
		582905	582728				