PROCEEDINGS
OF THE ROYAL SOCIETY B
BIOLOGICAL SCIENCES

Supporting Online Material for

Multiple large inversions and breakpoint rewiring of gene expression in the evolution of the fire ant social supergene

Yu-Ching Huang, Việt Đài Đặng, Ni-Chen Chang, and John Wang*

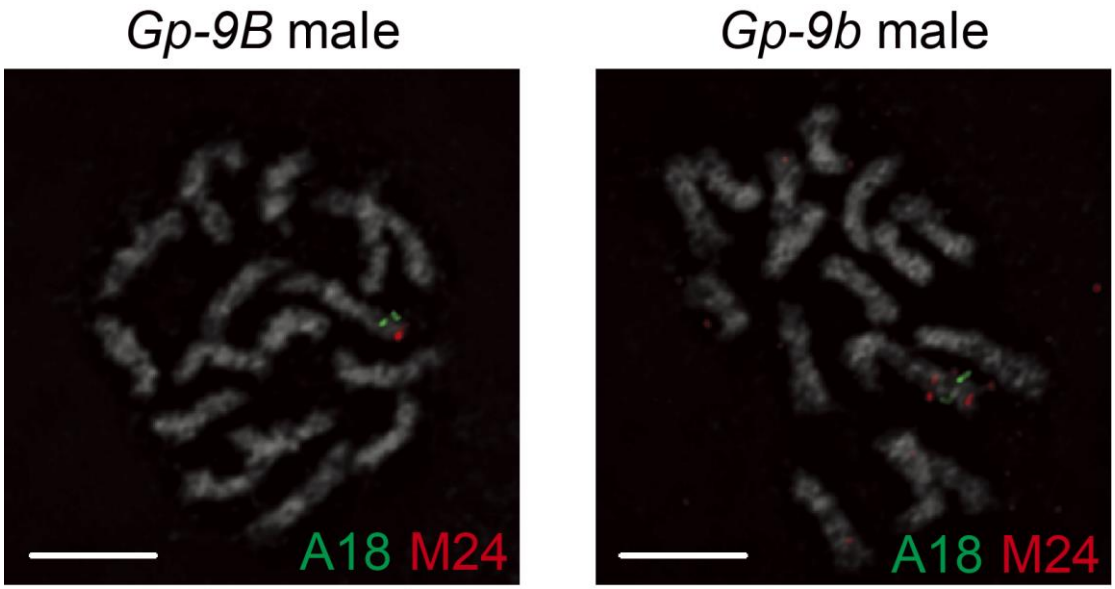
*To whom correspondence should be addressed. E-mail: johnwang@gate.sinica.edu.tw

DOI: 10.1098/rspb.2018.0221

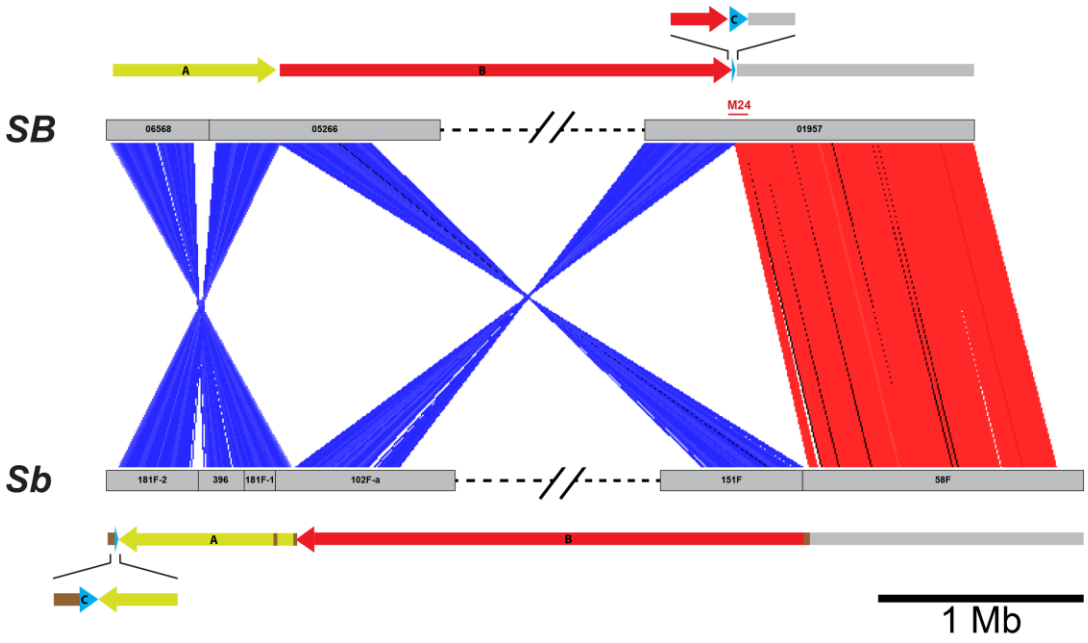
This PDF file includes:

Supplementary figures S1 to S3

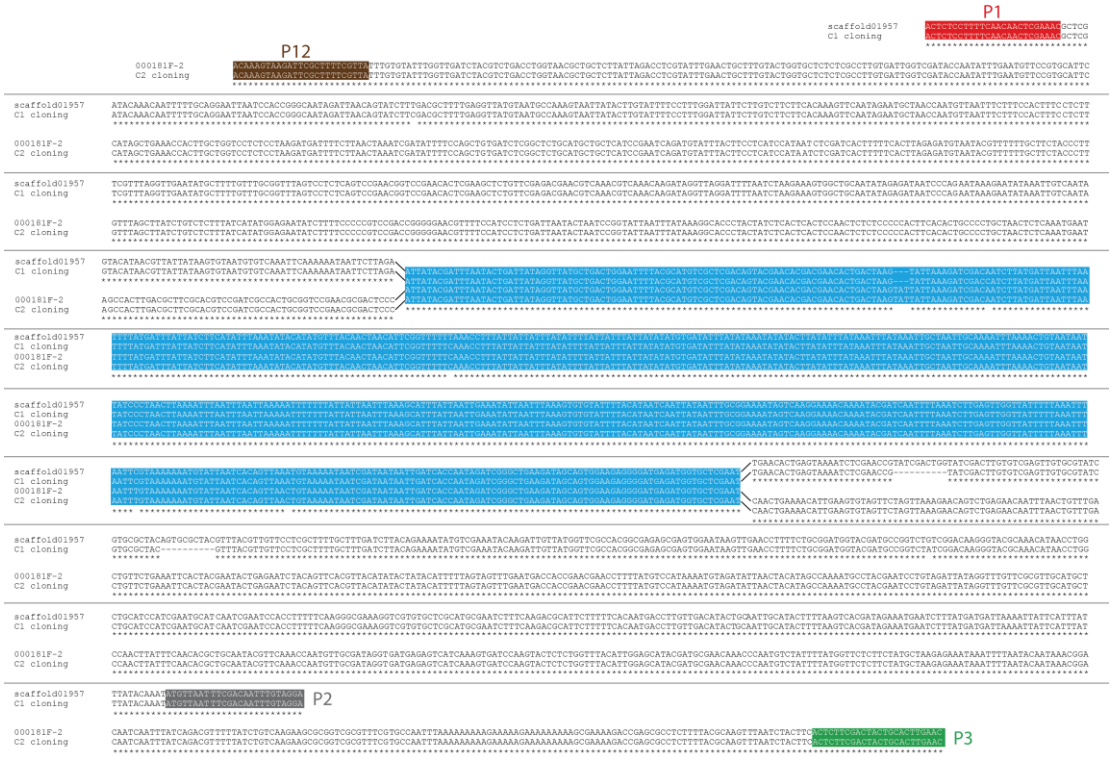
Supplementary tables S1 to S6



Supplementary figure 1. The original images of BAC-FISH in figure 1b. A18 (green signals) is an inner marker of the supergene for both *SB* and *Sb* social chromosomes. M24 probe (red) hybridized to the right side of the A18 signals on the *SB* social chromosome, whereas it hybridized to both sides of the A18 signal on the *Sb* chromosome. Scale bars, 5 μm.



Supplementary figure 2. Comparison of the scaffold order between *SB* and *Sb* supergenes. The *SB* scaffolds (06568, 05266, and 01957) were assigned with genetic linkage evidence, whereas the *Sb* contigs (000181F-2, 000396F, 000181F-1, 000102F-a, 000151F, and 000058F) were oriented based on the identification of sequence fusions or inferred with the most likely neighboring contigs. Sequence comparisons were filtered and show only sequence fragments with score >1,400 and identity >90%; therefore the comparison for fragment C between *SB* and *Sb* is hidden due to its short length and low score (~600 bp, score = 1,051, identity = 99%). Sequence identities are color coded from light to bright (90% to 100%) for forward (red) and reverse (blue) alignments. Scaffold lengths are in scale, except the junctions (dash lines) between scaffold05266 and scaffold01957 in *SB*, and contigs 000102F-a and 000151F in *Sb*. Locations of BAC probes M24, and fragments A, B, and C depicted in figure 1a are shown.



Supplementary figure 3. Sequence alignments of fragment C between the *SB* and *Sb* genomes. The *SB*- and *Sb*-specific fragment C with ~500 bp upstream and downstream flanking sequences from BigB_G_scaffold01957 and littleb_000181F-2 were aligned independently with their corresponding cloned sequences C1 and C2. These two *SB*- and *Sb*-specific alignments were manually merged and re-aligned with respect to the locations of the fragment C sequences (blue background). Primers used for PCR amplifications of the cloning regions C1 (P1 and P2) and C2 (P12 and P3) are color coded as in figure 1a.

Supplementary table 1. Bacterial artificial chromosome clones spanning scaffolds of the social chromosome

Plate	Well	Scaffold*	Location	Reference
73	A18	Si_gnG.scaffold00480:286949-399624	Supergene	This study and Wang et al., 2013
73	A22	Si_gnG.scaffold02940:1084875-1181235	Chromosome left end	Wang et al., 2013
73	E03	Si_gnG.scaffold01957:332331-441480	Supergene	Wang et al., 2013
73	E17	Si_gnG.scaffold07090:311160-399319	Supergene	Wang et al., 2013
73	G23	Si_gnG.scaffold00899:696857-795039	Supergene	Wang et al., 2013
145	M24	Si_gnG.scaffold01957:401094-493803	Breakpoint	This study

*End sequencing of the BAC clones with T7P and SP6 primers. Sequence ends were trimmed for high quality using DNA Baser (v4.36.0).

Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang YC, Shoemaker D, and Keller L. 2013. AY-like social chromosome causes alternative colony organization in fire ants. Nature 493:664-668

Supplementary table 2. Primers information used in this study

Name	Primer sequences (5'-3')	*SB* scaffold	*Sb* contig
P1	ACT CTC CTT TTC AAC AAC TCG AAA C	Si_gnG.scaffold01957	000102F-a
P2	TCC TAC AAA TTG TCG AAA TTA ACA T	Si_gnG.scaffold01957	000058F
P3	GTT CAA GTG CAG TAG TCG AAG AGT	Si_gnG.scaffold05266	000181F-2
P4	TGA AGA ACT GTT ACT TTT GTG CAT C	Si_gnG.scaffold05266	000058F
P5	TTT TTC ACG TAT TCT GCG TTT TC	Si_gnG.scaffold06568	000102F-a
P6	TAG AGG AAA AGT TTG CTT ATT CTG C	Si_gnG.scaffold06568	000181F-1
P7	AGT GGA GAG GAA GGT ATC CGT AG		000058F
P8	TTG ATA TTC TGT AAC AAT TTG TGG A		000058F
P9	GAG AGG GCA AAG AAT TAG AGA GAG A		000102F-a
P10	GAT CAA TTT CTT ATG GTA ACG GCT A		000102F-a
P11	CCT CTT ACT CTC CTG GAT CTC ATG T		000181F-1
P12	ACA AAG TAA GAT TCG CTT TTC GTT A		

Junction	Primer pair	Size (bp)	*SB*-specific	*Sb*-specific
C1	P1 + P2	1,502	x	
C2	P12 + P3	1,751		x
J1	P3 + P4	888	x	
J2	P5 + P6	1,078	x	
J3	P7 + P2	787		x
J4	P4 + P8	840		x
J5	P9 + P1	771		x
J6	P10 + P5	1,057		x
J7	P6 + P11	1,066		x

Supplementary table 3. Summary of the expression of the three target genes in the worker antennae RNA-seq datasets.

Each library is a pool of 36 to 53 pairs of antennae of the same Gp-9 genotype and from the same colony.

[a]Number of reads produced by Illumina sequencing using the HiSeq platform

[b]Number of mapped reads by BWA-mem

[c]Total mapped reads for all genes as calculated by HTSeq-count.

[d]Number of reads mapped to the gene region

[e]The expression value of the target genes (Gene_read) normalized by the library size (Mapped_read)(count-per-million, CPM)

[(*)]Pairs of reads

[(**)]Both singly and properly paired mapped reads are calculated as 1 in HTSeq-count by default

							SINV23002-SINV23011		SINV22157		SINV22107	
Sample ID	Social form	Genotype	#pairs of antennae	Library[a(*)]	BWA_mapped[b(*)]	Mapped_read[c(**)]	Gene_read[d(**)]	CPM[e]	Gene_read[d(**)]	CPM[e]	Gene_read[d(**)]	CPM[e]
M_BB2	Monogyne	SB/SB	39.5	29,197,782	25,362,865	33,029,573	195	5.90	96	2.91	1,240	37.54
M_BB3			53	29,815,211	29,129,293	34,459,212	187	5.43	176	5.11	1,190	34.53
M_BB6			40	30,629,100	29,890,678	41,952,749	141	3.36	248	5.91	1,860	44.34
M_BB7			40.5	23,908,640	22,751,357	28,103,927	163	5.80	135	4.80	1,749	62.23
P_BB2	Polygyne	SB/SB	39.5	27,570,550	26,682,410	28,295,241	225	7.95	258	9.12	1,396	49.34
P_BB3			53	29,718,335	28,771,536	37,040,319	139	3.75	186	5.02	2,040	55.08
P_BB6			40	26,978,725	26,341,745	31,499,190	84	2.67	233	7.40	1,314	41.72
P_BB7			36	28,093,637	27,250,328	27,664,357	117	4.23	170	6.15	1,292	46.70
P_Bb2	Polygyne	SB/Sb	39.5	30,670,661	29,736,108	32,939,344	176	5.34	464	14.09	2,182	66.24
P_Bb3			53	24,020,182	23,038,506	30,015,219	156	5.20	390	12.99	1,718	57.24
P_Bb6			40	30,685,360	29,891,124	36,937,213	163	4.41	519	14.05	1,851	50.11
P_Bb7			36	28,858,541	28,082,134	14,064,349	80	5.69	192	13.65	768	54.61

Supplementary table 4. Statistical test of the expression of the three target genes in the worker antennae

[a]Log2-fold change

[b]*P* -value produced by applying quasi-likelihood F-test for differential expression of each gene by edgeR

[c]Adjusted p-value threshold for the 9 tests using Bonferroni method at 5% false discovery rate

[d]*P* -value adjusted for multiple testing using Benjamini-Hochberg method at 5% false discovery rate for whole antennal expressed gene set (FDR)

[*]P_Bb = Polygyne SB/Sb; P_BB = Polygyne SB/SB; M_BB = Monogyne SB/SB

Gene	Pairwise comparison[*]	logFC[a]	*P* -value[b]	9-test-FDR[c]	library-FDR[d]
SINV23002-SINV23011	P_Bb vs P_BB	0.15	0.70	0.0056	1.00
	P_Bb vs M_BB	0.0085	0.98	0.0056	1.00
	P_BB vs M_BB	-0.14	0.71	0.0056	1.00
SINV22157	P_Bb vs P_BB	0.99	0.0051	0.0056	1.00
	P_Bb vs M_BB	1.55	2.10E-005	0.0056	0.01
	P_BB vs M_BB	0.56	0.11	0.0056	1.00
SINV22107	P_Bb vs P_BB	0.24	0.39	0.0056	1.00
	P_Bb vs M_BB	0.35	0.21	0.0056	0.99
	P_BB vs M_BB	0.11	0.70	0.0056	1.00

Supplementary table 5. Allelic specific expression of the three genes in *SB/Sb* worker antennae

Data are merged RNA-seq data from four bioreplicates of polygyne worker antennae. Each bioreplicate was an amplified RNA pool of at least 40 pairs of antennae.

[a]position to the coding DNA sequence (CDS)

[b]SNP/indel in sense strand

[c]Number of reads called by GATK software

[d]Two-tailed binomial test for difference from equality

[e]Two-tailed binomial test for difference from equality for the summed gene reads

[f]Explaination of invalid data

Gene name	Position			Allele (SNP or indel)[b]		Read count[c]				*P*-value[d]	Read sum across gene[c]				*P*-value[e]	Note[f]
	Si_gnG		CDS[a]	SB	Sb	SB	Sb	Total	Sb/Total		SB	Sb	Total	Sb/Total		
SINV22107	scaffold01957	439993	-67	C	T	-	-	-	NA	NA	401	605	1006	0.60	1.35E-10	reads with mutation, not called (total count: 89)
		440988	929	G	A	61	107	168	0.64	0.0004806						
		440991	932	C	T	61	104	165	0.63	0.0010159						
		441210	1151	G	A	85	122	207	0.59	0.0121579						
		441318	1259	T	C	93	127	220	0.58	0.0258664						
		441857	1798	T	C	101	145	246	0.59	0.0060015						
SINV22157	scaffold01957	439234	-617	C	G	14	149	163	0.91	1.14E-29	152	491	643	0.76	1.53E-42	
		439145	-529	CGTTCGGACT	T	26	52	78	0.67	0.0043349						
		439077	-460	C	T	23	45	68	0.66	0.0103377						
		438865	-248	T	G	23	55	78	0.71	0.0003778						
		438677	-60	T	C	20	108	128	0.84	8.59E-16						
		437405	1213	C	T	22	34	56	0.61	0.1408954						
		436803	1815	A	T	24	48	72	0.67	0.0063098						
SINV23002 SINV23011	scaffold05266	344039	5	G	A	-	-	-	NA	NA	248	351	599	0.59	2.95E-05	no polymorphism (total count: 17)
		344088	54	A	G	41	28	69	0.41	0.1480322						
		345505	1471	A	G	45	47	92	0.51	0.9170405						
		345926	1892	T	C	64	90	154	0.58	0.0435999						
		346219	2185	A	G	49	92	141	0.65	0.0003682						
		346238	2204	T	C	49	94	143	0.66	0.0002095						

Supplementary table 6. The expression of SINV22157, *Gp-9* and an elongation factor gene (NCBI assession EH413242.1) in public datasets.

All the RNAseq data are from whole bodies

Raw reads were mapped against the fire ant genome by TopHat. Expression level is shown in FPKM (fragments per kilobase of transcript per million mapped reads).

The expression of the target gene SINV22157 is shown in comparison with the expression of the well-known *Gp-9* and an elongation factor (NCBI assession EH413242.1)

Dataset	Social_form	Caste	SINV22157	EH413242.1	*Gp-9*	REF
SRR619836	Polygyne	Queen	16.4	57.7	2,736.5	Wang et al. (2013)
SRR619837			17.5	53.6	18,084.2	
SRR619838			19.6	53.0	7,479.5	
SRR619849			19.1	53.5	25,162.4	
SRR619956			16.0	60.8	6,245.8	
SRR619976			18.1	59.0	14,649.4	
DRS023309	Monogyne	Queen	19.5	62.8	8,470.5	Morandin et al. (2016)
DRS023310			29.8	30.6	2,014.9	
DRS023311			18.9	23.0	2,263.0	
DRS023312	Monogyne	Worker	10.0	15.5	13,802.6	Morandin et al. (2016)
DRS023313			4.1	11.7	53.8	
DRS023314			6.2	16.6	13,079.8	
DRS023315	Polygyne	Queen	25.7	15.4	1,531.7	Morandin et al. (2016)
DRS023316			10.3	17.1	6,508.6	
DRS023317			19.3	46.1	19,383.9	
DRS023318	Polygyne	Worker	11.1	22.4	23,875.4	Morandin et al. (2016)
DRS023319			7.5	35.9	15,195.0	
DRS023320			8.6	20.7	38,939.5	

Morandin C, Tin MMY, Abril S, Gómez C, Pontieri L, Schiøtt M, Sundström L, Tsuji K, Pedersen JS, Helanterä H and Mikheyev AS. 2013. Comparative transcriptomics reveals the conserved building blocks involved in parallel evolution of diverse phenotypic traits in ants. Genome Biology 17:43