Supporting Information

Supporting Information Corrected October 15, 2014

Harpur et al. 10.1073/pnas.1315506111

SI Materials and Methods

We developed the following bioinformatic pipeline: (*i*) FASTQ files were initially aligned to the *Apis mellifera* genome assembly AMELv4.5 using the default parameters of BWA (1) and alignments were then imported into SAMtools (2) in BAM format. (*ii*) We remapped each bee's sequence using Stampy (3) at a substitution rate of 0.02 to better align divergent sequences. (*iii*) We subsequently realigned sequences with GATK's RealignerTargetCreator followed by IndelRealigner to reduce any potential erroneous alignments close to indels (4). We detected SNPs and created variant calling files (i.e., VCF) using mpileup (–Q 20 option), bcftools (mutation rate of 0.05), and varfilter (–d 3 –Q 15 –D64) (2). We filtered out highly repetitive regions and recently duplicated genes from our analyses by first performing a blastn match of 50-bp segments of the *A. mellifera* genome back to the reference genome; we excluded any 50-bp segment matching two or more locations with fewer than three mismatches and blastn E-value of 2E-20. An average of 3.2% of SNPs were masked with this protocol. We also excluded 6.47 Mb of sequence from unmapped scaffolds (scaffolds 17.2000 and above in AMELv4.5) because of low sequencing coverage in these small (mean 1,957 bp) and gene-poor scaffolds. We aligned the *Apis cerana* sequences to the reference *A. mellifera* genome using the same methods as above, except we set the Stampy divergence threshold to $d = 0.05$.

1. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
2. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
3. Lunter G, Goodson M (2011) Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* 21(6):936–939.
4. McKenna A, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.

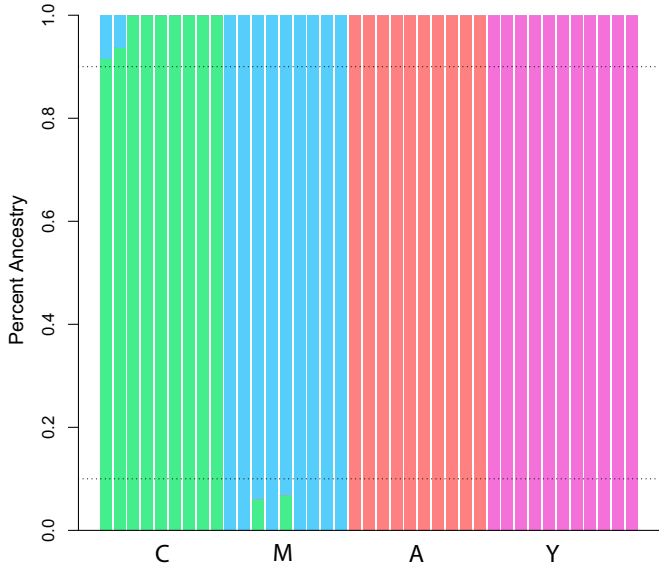

Fig. S1. Population structure and ancestry for bees used in this study. We tested $K = 1$–6 populations (100 times per K) assuming no prior knowledge of the population origin using 25,000 randomly selected SNPs from the bee genome (*Materials and Methods*). The dataset best fit a model with four distinct populations (A, Y, M, and C; statistics from a single run: $K = 4$, CV error = 0.18078, LL = –891627). Each column represents the relative ancestry of each sampled bee to the four populations delineated by the Bayesian analysis. The sampled bees were very pure, and on average each bee had a 99.39% ancestry to its inferred population.

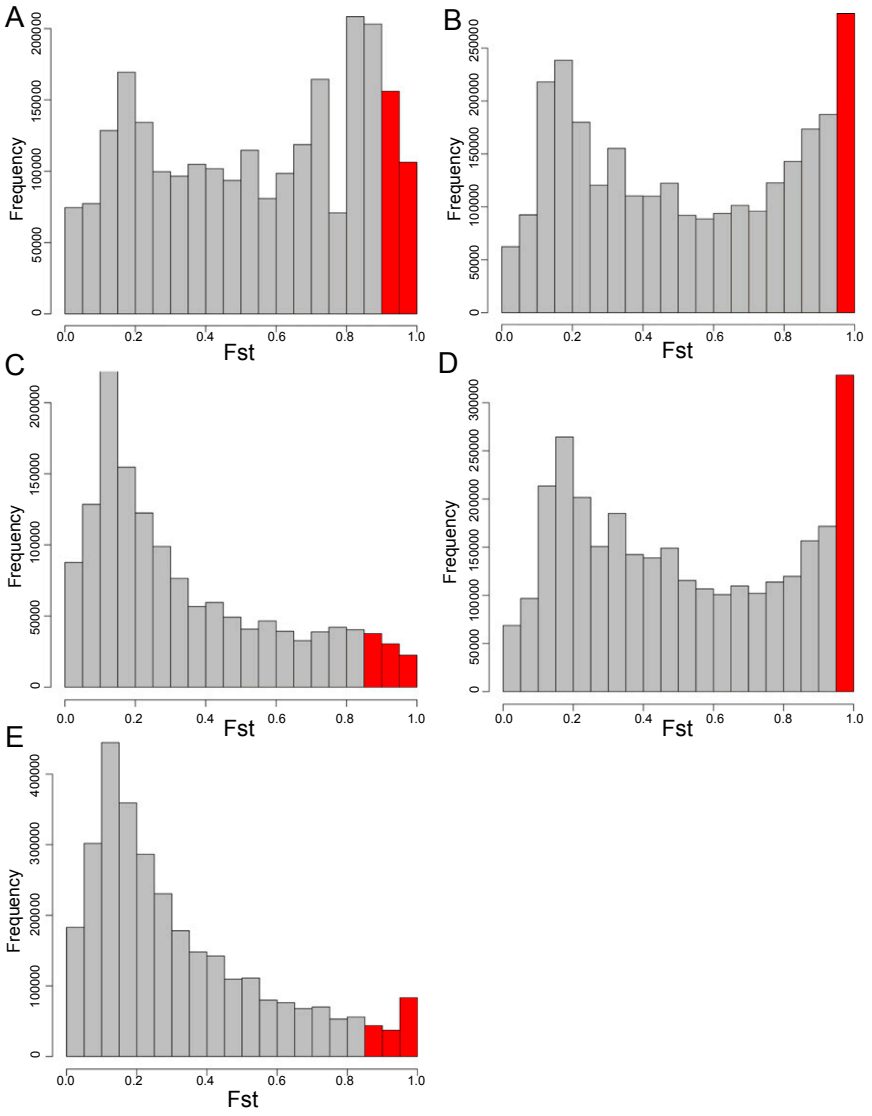

Fig. S2. Histograms of pairwise genetic differentiation (F_{ST}) between the following population pairs: (*A*) C vs. M, (*B*) C vs. Y, (*C*) C vs. A, (*D*) M vs. Y, and (*E*) A vs. Y. Areas in red indicate SNPs with high F_{ST} values (>95% of data).

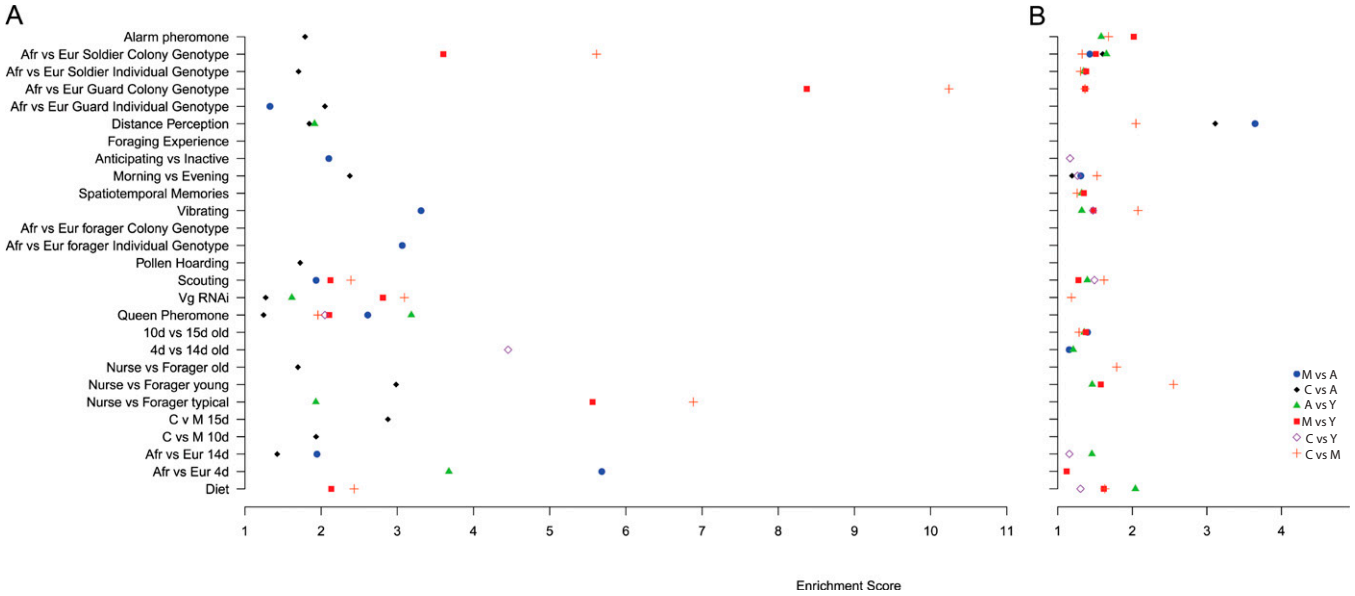

Fig. S3. Enrichment of F_{ST} outlier SNPs in (*A*) 500-bp regions upstream of genes and (*B*) nonsynonymous SNPs in exons for 27 experiments performed in the BeeSpace project. The 27 experiments were labeled according to the behavioral/genotypic contrast (1). Afr, Africanized bees; Eur, North/South American bees, likely of European descent; C, East European bees; and M, West/Northern European Bees. Two microarray experiments compared brain gene expression in 10- and 15-d-old worker bees from East Europe (C group) vs. West/Northern Europe (M group bees). Differentially expressed genes in these experiments were not enriched for coding or regulatory loci with signs of adaptive divergence between the M and C populations studied herein. This lack of enrichment can arise if adaptive changes in a relatively small number of genes are driving the observed shifts in brain gene expression between M and C group bees. Alternatively, it is likely that the M and C bees used for these microarray experiments were admixed (2) and do not adequately reflect the nonadmixed bees used in our study.

1. Chandrasekaran S, et al. (2011) Behavior-specific changes in transcriptional modules lead to distinct and predictable neurogenomic states. *Proc Natl Acad Sci USA* 108(44):18020–18025.
2. Harpur BA, Minaei S, Kent CF, Zayed A (2012) Management increases genetic diversity of honey bees via admixture. *Mol Ecol* 21(18):4414–4421.

Table S1. Sample information for the sequenced honey bees

Bee ID	Country	Species/ population group	% Purity	Coverage*
11	Chiang Mai, Thailand	*A. cerana*	NA	28.040
181	North Rhine-Westphalia, Germany	C	0.916	39.730
182	North Rhine-Westphalia, Germany	C	0.937	38.270
195	Novalija, Pag Island, Croatia	C	0.999	39.719
196	Komorovac, Pag Island, Croatia	C	0.999	38.136
199	Baric Draga, Croatia	C	0.999	44.745
200	Sugarje, Croatia	C	0.999	35.372
235	Slovenia	C	0.999	39.543
236	Slovenia	C	0.999	34.769
273	Slovenia	C	0.999	40.925
207	Zalewo, Poland	M	0.999	42.524
217	Rudawka, Poland	M	0.999	43.180
218	Rudawka, Poland	M	0.940	36.760
226	Sucha Rzeczka, Poland	M	0.999	36.919
227	Sucha Rzeczka, Poland	M	0.932	37.390
233	Cordoba, Spain	M	0.999	34.477
234	Murcia, Spain	M	0.999	29.040
248	Murcia, Spain	M	0.999	38.925
256	Cordoba, Spain	M	0.999	37.318
340	Riyadh, Saudi Arabia	Y	0.999	38.087
341	Riyadh, Saudi Arabia	Y	0.999	38.168
342	Riyadh, Saudi Arabia	Y	0.999	36.040
343	Riyadh, Saudi Arabia	Y	0.999	37.113
344	Riyadh, Saudi Arabia	Y	0.999	37.219
345	Riyadh, Saudi Arabia	Y	0.999	35.384
346	Riyadh, Saudi Arabia	Y	0.999	37.258
347	Sana'a, Yemen	Y	0.999	35.758
348	Sana'a, Yemen	Y	0.999	33.570
349	Sana'a, Yemen	Y	0.999	28.280
279	East Cape, South Africa	A	0.999	40.207
280	East Cape, South Africa	A	0.999	40.852
284	NW. Province, South Africa	A	0.999	37.158
285	N. Cape, South Africa	A	0.999	39.176
286	NW. Province, South Africa	A	0.999	40.791
290	West Cape, South Africa	A	0.999	40.552
296	Freestate, South Africa	A	0.999	38.507
299	N. Cape/NW. Province, South Africa	A	0.999	39.523
300	NW. Province, South Africa	A	0.999	40.822
301	East Cape, South Africa	A	0.999	41.290
302	East Cape, South Africa	A	0.999	41.076

*Coverage based on alignments with *A. mellifera* reference genome Amel4.5.

Table S2. Genome wide average pairwise genetic differentiation (F_{ST}) between the honey bee's four major population groups

Population	C	M	Y	A
C	—			
M	0.540	—		
Y	0.525	0.513	—	
A	0.343	0.335	0.324	—

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (XLSX)
Dataset S4 (XLSX)
Dataset S5 (XLSX)