**Supplementary material for ‘Adaptive, caste-specific changes to recombination rates in a thelytokous honeybee population’**

Table of contents

[Supplementary methods 2](#_Toc66800290)

[Figure S1. Homozygosity within 10 kb windows in 14 thelytokous daughters of one Capensis worker. 4](#_Toc66800291)

[Figure S2. Heterozygosity within 10 kb windows in 14 thelytokous daughters of one Capensis worker. 5](#_Toc66800292)

[Table S1. Microsatellite primers used in this study 6](#_Toc66800293)

[Table S2 Microsatellite genotypes of the virgin queen and her thelytokous progeny and the four workers and their thelytokous progeny, including loss of heterozygosity calculations. 7](#_Toc66800294)

[Table S3. Loss of heterozygosity in a virgin queen subjected to CO2 narcosis. 7](#_Toc66800295)

[Table S4. Sequencing statistics. 8](#_Toc66800296)

[Table S5. Minimal loss of heterozygosity in the offspring of Capensis workers (SNPs). 9](#_Toc66800297)

[Table S6. Minimal loss of heterozygosity in the offspring of Capensis workers (windows). 10](#_Toc66800298)

# Supplementary methods

The quality of the raw data with was checked with FastQC 0.11.7 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). We used Trimmomatic 0.38 [1] to trim low quality reads and adapter sequences using the following parameters: ILLUMINACLIP:/path/to/NexteraPE-PE.fa:2:30:10:8:TRUE HEADCROP:16 LEADING:30 TRAILING:30 SLIDINGWINDOW:4:30 MINLEN:36. Trimmed reads were mapped to the honey bee genome assembly Amel\_HAv3.1 [2] using BWA 0.7.12 [3] with default parameters. We excluded reads that did not map uniquely, and then sorted and indexed alignments with SAMtools 1.9 [4]. We marked PCR duplicates with Picard 1.119 (http://broadinstitute.github.io/picard). We realigned sequences around insertions and deletions with GATK 3.8-1-0 [5] and called single nucleotide polymorphisms (SNPs) using GATK.

We excluded low quality (QUAL < 30) SNPs, as well as those found within unplaced and non-nuclear regions as follows: those with a Fisher strand bias (FS) > 20, a mapping quality (MQ) < 40, a quality by depth (QD) < 2, those that fell within 10 kb of insertions or deletions, and those that fell within 1 kb of SNPs called heterozygous in the father’s sample when set as diploid [6] using GATK. We excluded SNPs with a genotype quality (GQ) < 20, a read depth (DP) < 20, those with more than two alleles, those with a read depth falling outside 1.5 times the inter-quartile range [7], and those found within centromeres using VCFtools 0.1.14 [8].

We then determined the positions of centromeres as in [9]. We excluded heterozygous SNPs with an allele ratio < 0.15 using BCFtools 1.9 [10] and BEDtools 2.26.0 [11].

We used SnpEff 4.3 [12] to extract all heterozygous SNPs and all homozygous SNPs found in the mother worker and the 14 offspring larvae. We used R 3.3.3 [13] to determine the number of heterozygous and homozygous SNPs overlapping between the mother worker and each offspring larvae. We estimated the loss of heterozygosity in each offspring larvae as the number of homozygous SNPs being heterozygous in the mother worker, divided by the total number of SNPs being heterozygous in the mother and overlapping with that particular offspring. We then used BEDtools to create 10 kb windows along the entire genome [9]. We then retrieved, for each individual, the number of heterozygous and homozygous SNPs in each window with BEDtools, only considering windows that had at least 10 SNPs. We calculated the level of heterozygosity per window as the number of heterozygous SNPs divided by the total number of SNPs in each window. Windows with a heterozygosity level of zero were deemed homozygous. All other windows were deemed heterozygous. We used R to determine the number of heterozygous windows overlapping between the mother worker and the 14 offspring larvae. We calculated the proportion of the genome always found to be heterozygous as the number of overlapping heterozygous windows divided by the total number of overlapping windows. We used BEDtools to retrieve the number of genes intersecting with all overlapping heterozygous windows.

We tested the effect of distance from the centromere on the probability that windows would lose heterozygosity using a generalised linear mixed-effect model (GLMM) with a binomial distribution and a logic link function using the R package lme4 [14]. Log(distance of the centre of a window from the centromere) was modelled as a fixed effect and offspring larva was modelled as a random effect.

1. Bolger A.M., Lohse M., Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15), 2114-2120. (doi:10.1093/bioinformatics/btu170).

2. Wallberg A., Bunikis I., Pettersson O.V., Mosbech M.-B., Childers A.K., Evans J.D., Mikheyev A.S., Robertson H.M., Robinson G.E., Webster M.T. 2019 A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *BMC Genomics* **20**(1), 275. (doi:10.1186/s12864-019-5642-0).

3. Li H., Durbin R. 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**(14), 1754-1760. (doi:10.1093/bioinformatics/btp324).

4. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., Subgroup G.P.D.P. 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16), 2078-2079. (doi:10.1093/bioinformatics/btp352).

5. McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., et al. 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Gen Res* **20**(9), 1297-1303. (doi:10.1101/gr.107524.110).

6. Wallberg A., Han F., Wellhagen G., Dahle B., Kawata M., Haddad N., Simoes Z.L.P., Allsopp M.H., Kandemir I., De la Rua P., et al. 2014 A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature Genetics* **46**, 1081-1088. (doi:10.1038/ng.3077).

7. Harpur B.A., Dey A., Albert J.R., Patel S., Hines H.M., Hasselmann M., Packer L., Zayed A. 2017 Queens and workers contribute differently to adaptive evolution in bumble bees and honey bees. *Genome Biol Evol* **9**(9), 2395-2402. (doi:10.1093/gbe/evx182).

8. Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., et al. 2011 The variant call format and VCFtools. *Bioinformatics (Oxford, England)* **27**(15), 2156-2158. (doi:10.1093/bioinformatics/btr330).

9. Smith N.M.A., Wade C., Allsopp M.H., Harpur B.A., Zayed A., Rose S.A., Engelstadter J., Chapman N.C., Yagound B., Oldroyd B.P. 2019 Strikingly high levels of heterozygosity despite 20 years of inbreeding in a clonal honey bee *J Evol Biol* **33**(2), 144-152. (doi:10.1111/jeb.13397).

10. Li H. 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**(21), 2987-2993. (doi:10.1093/bioinformatics/btr509).

11. Quinlan A.R., Hall I.M. 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6), 841-842. (doi:10.1093/bioinformatics/btq033).

12. Cingolani P., Platts A., Wang L.L., Coon M., Nguyen T., Wang L., Land S.J., Lu X., Ruden D.M. 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**(2), 80-92. (doi:10.4161/fly.19695).

13. R Core Team. 2000 R language definition. (Viena, Austria, R Foundation for Statistical Computing).

14. Bates D., Machler M., Bolker B.M., Walker S.C. 2015 Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1-48. (doi:10.18637/JSS.V067.I01).

Figure S1. Homozygosity within 10 kb windows in 14 thelytokous daughters of one Capensis worker.

Blue lines represent homozygous windows. Orange lines represent homozygous windows that are heterozygous in the mother worker (i.e., regions of loss of heterozygosity). Each chromosome in the outer circle is represented with a different colour. Scale represents chromosome size (Mb). The first concentric circle is the mother worker. Each of the 14 inner concentric circles (below the dashed line) is a different progeny of the mother. Grey areas correspond to centromeric regions.

Chart, radar chart

Description automatically generated

Figure S2. Heterozygosity within 10 kb windows in 14 thelytokous daughters of one Capensis worker.

Blue lines represent heterozygous windows. Orange lines represent heterozygous windows that are present in all individuals. Each chromosome in the outer circle is represented with a different colour. Scale represents chromosome size (Mb). The first concentric circle is the mother worker. Each of the 14 inner concentric circles (below the dashed line) is a different progeny of the mother. Grey areas correspond to centromeric regions.

Chart, sunburst chart

Description automatically generated

|  |  |  |
| --- | --- | --- |
| Locus | Forward | Reverse |
| A8 | CGAAGGTAAGGTAAATGGAAC | GGCGGTTAAAGTTCTGG |
| A24 | CACAAGTTCCAACAATGC | CACATTGAGGATGAGCG |
| A79 | CGAAGGTTGCGGAGTCCTC | GTCGTCGGACCGATGCG |
| B124 | GCAACAGGTCGGGTTAGAG | CAGGATAGGGTAGGTAAGCAG |
| A113 | CTCGAATCGTGGCGTCC | CCTGTATTTTGCAACCTCGC |
| A88 | CGAATTAACCGATTTGTCG | GATCGCAATTATTGAAGGAG |
| HB-The-03 | TAACTGGTCGTCGGTGTT | ACGTAGAGAATCCCATTGT |
| A28 | GAAGAGCGTTGGTTGCAGG | GCCGTTCATGGTTACCACG |
| Ap43 | GGCGTGCACAGCTTATTCC | CGAAGGTGGTTTCAGGCC |

# Table S1. Microsatellite primers used in this study

# Table S2 Microsatellite genotypes of the virgin queen and her thelytokous progeny and the four workers and their thelytokous progeny, including loss of heterozygosity calculations.

This excel file is provided as a separate supplementary data file

# Table S3. Loss of heterozygosity in a virgin queen subjected to CO2 narcosis.

This excel file is provided as a separate supplementary data file

# Table S4. Sequencing statistics.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sample | Raw reads | Data yield (Gb) | Depth of coverage (X) | Heterozygous SNPs | Homozygous SNPs |
| Drone | 60,147,207 | 18.16 | 34.0 | 0 | 1,299,423 |
| Mother worker | 53,114,914 | 16.04 | 29.1 | 993,778 | 585,404 |
| Offspring larvae #1 | 57,074,195 | 17.24 | 34.9 | 1,114,134 | 663,352 |
| Offspring larvae #2 | 62,352,341 | 18.83 | 35.3 | 1,118,499 | 666,806 |
| Offspring larvae #3 | 59,290,292 | 17.91 | 33.4 | 1,094,070 | 650,492 |
| Offspring larvae #4 | 66,836,635 | 20.18 | 38.2 | 1,142,363 | 682,596 |
| Offspring larvae #5 | 65,949,155 | 19.92 | 38.5 | 1,126,244 | 695,309 |
| Offspring larvae #6 | 58,642,762 | 17.71 | 35.2 | 1,104,720 | 677,966 |
| Offspring larvae #7 | 61,741,651 | 18.65 | 37.2 | 1,133,008 | 676,630 |
| Offspring larvae #8 | 61,095,810 | 18.45 | 38.5 | 1,129,857 | 672,523 |
| Offspring larvae #9 | 66,762,702 | 20.16 | 42.3 | 1,156,860 | 691,518 |
| Offspring larvae #10 | 71,449,405 | 21.58 | 41.4 | 1,108,669 | 721,496 |
| Offspring larvae #11 | 63,333,639 | 19.13 | 39.6 | 1,139,238 | 690,310 |
| Offspring larvae #12 | 57,848,834 | 17.47 | 32.6 | 1,063,258 | 647,987 |
| Offspring larvae #13 | 71,139,349 | 21.48 | 33.4 | 1,077,616 | 640,096 |
| Offspring larvae #14 | 59,799,030 | 18.06 | 38.5 | 1,086,977 | 702,084 |