HAL
open science

Sequencing haploid drones from royaljelly and honey bee populations for detection of differentiation and selective sweeps

David Wragg, Benjamin Basso, Yves Le Conte, Jean Pierre Bidanel, Alain Vignal

To cite this version:

David Wragg, Benjamin Basso, Yves Le Conte, Jean Pierre Bidanel, Alain Vignal. Sequencing haploid drones from royaljelly and honey bee populations for detection of differentiation and selective sweeps. Biology and Genomics of Social Insects, May 2015, New York, United States. hal-02792618

HAL Id: hal-02792618

https://hal.inrae.fr/hal-02792618v1

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ResearchGate

See discussions, stats, and author profiles for this publication at: http://www.researchgate.net/publication/276265614

Sequencing haploid drones from royal jelly and honey bee populations for detection of differentiation and selective sweeps

CONFERENCE PAPER · MAY 2015

DOWNLOADS
83

VIEWS
68

5 AUTHORS, INCLUDING:

Yves Le Conte
French National Institute for Agricultural Research
133 PUBLICATIONS **2,162** CITATIONS

SEE PROFILE

Jean-Pierre Bidanel
French National Institute for Agricultural Research
203 PUBLICATIONS **2,218** CITATIONS

SEE PROFILE

Available from: David Wragg
Retrieved on: 03 August 2015

Sequencing haploid drones from royal jelly and honey bee populations for detection of differentiation and selective sweeps

David Wragg[1], Benjamin Basso[2], Yves Le Conte[3], Jean-Pierre Bidanel[4], Alain Vignal[1]

[1]INRA, GenPhySE, Toulouse; [2]ITSAP, PrADE, Avignon; [3]INRA, PrADE, Avignon; [4]INRA, GABI, Jouy-en-Josas

Correspondance to david.wragg@toulouse.inra.fr or alain.vignal@toulouse.inra.fr

In France, until the middle of last century the endemic populations of honey bees were represented by a single subspecies, the black bee, *Apis mellifera mellifera*. More recently, apiculturists have demonstrated an interest in using other subspecies and hybrids between *A. m. ligustica*, *A. m. caucasica*, and *A. m. mellifera*, which have been found to be more efficient producers of honey and royal jelly.

Most population genetic studies to date in *A. mellifera* have been performed using microsatellite markers, and more recently with medium density SNP microarrays. However, recent advances in high-throughput sequencing and declining costs now make population studies at the genome level feasible and two recent publications[1,2] have proven its utility for worldwide surveys of bee populations. The SeqApiPop project is designed to study the structure of French bee populations by whole genome sequencing (Illumina™) of one drone per colony for 1000 colonies. The sequencing of haploid males rather than diploid worker individuals enables high confidence variant calling, reducing the need for greater depth of coverage and will also allow for easy determination of haplotypes.

METHODS

As a pilot study, one drone was sequenced per colony from a total of 60 colonies representing two distinct populations; one used in the **production of honey** (**HN**), and one for the **production of royal jelly** (**RJ**). Paired-reads were mapped to the reference (Amel4.5) using *BWA-MEM*[3]. Variants were called per individual using *GATK*[4], *Pileup*[5] and *Platypus*[6], and combined using *BAYSIC*[7]. VCF files were merged, and then filtered on call rate using Plink[8]. Missing genotypes were imputed with *BEAGLE*[9].

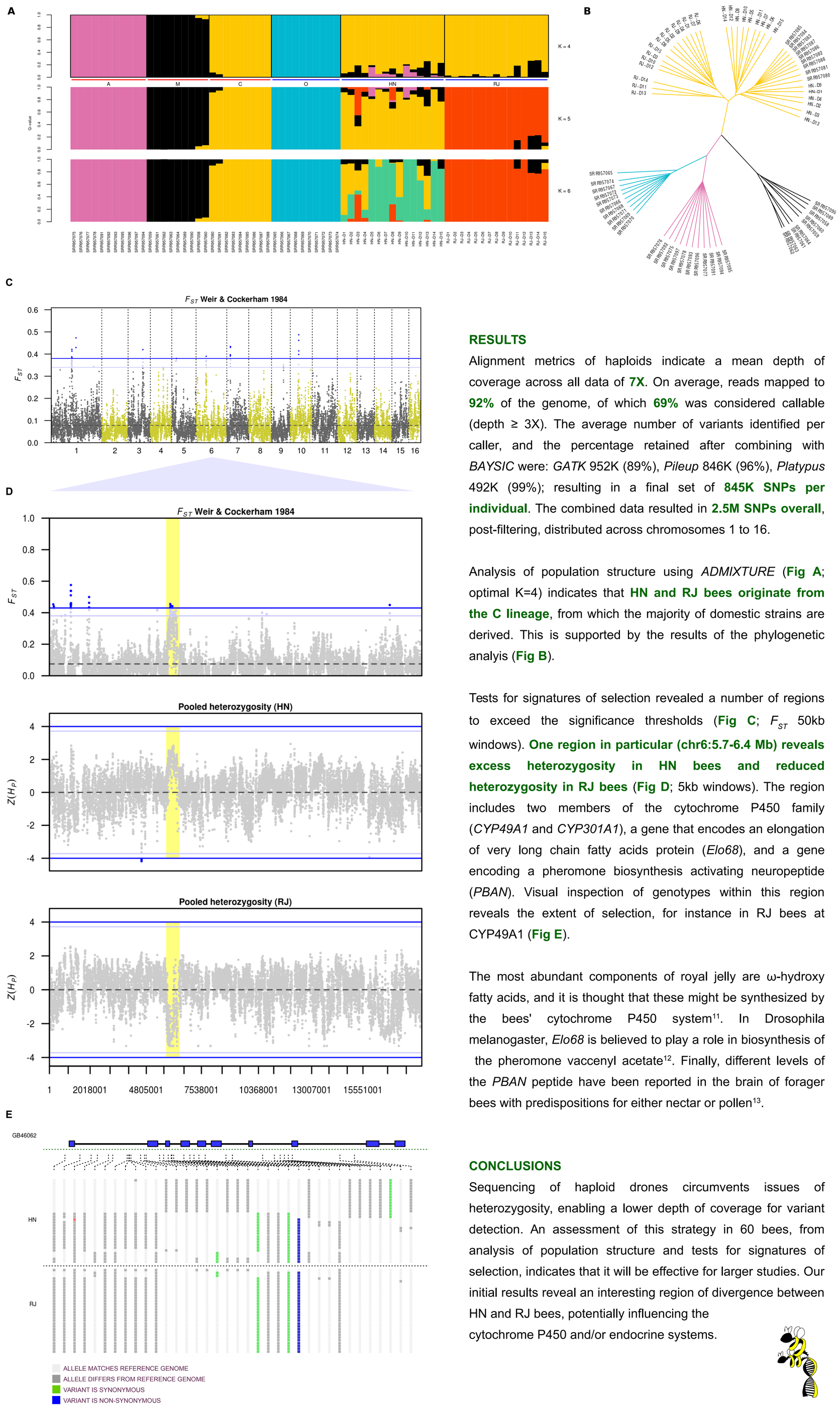
Diploids were constructed by combining haploids, population structure assessed using *ADMIXTURE*[10] with respect to a reference panel of diploid workers[1] from each of the main evolutionary lineages (A, M, C, O). The same data was used to construct a phylogentic tree, and the branches coloured according to the optimal *K* following *ADMIXTURE*.

Signatures of selection were detected by estimating F_{ST} and pooled heterozygosity (H_P) using a sliding window approach. Windows overlapped by 80% and were analysed at varying sizes (1-5, 10, 20, 50 kb). HP values were standardised to zero mean and unit variance. Significance thresholds were marked at the 99th and 99.9th percentiles for F_{ST}, and $|Z(H_P)| \geq 3.719$ and 4 for the H_P analysis. All plots were generated using R.

REFERENCES

1. Harpur BA, *et al.* 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *PNAS* **111**:2614–2619
2. Wallberg A, *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
3. Li, H 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio*
4. McKenna A, *et al.* 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20:**1297–1303
5. Li H. *et al.* 2009. The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25:**2078–2079
6. Rimmer A, *et al.* 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**:912-918.
7. Cantarel BL, *et al.* 2014. BAYSIC: a Bayesian method for combining sets of genome variants with improved specificity and sensitivity. *BMC Bioinformatics* **15**:104
8. Purcell S & Chang C. PLINK v1.90ble. at <https://www.cog-genomics.org/plink2>
9. Browning SR & Browning BL 2007. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics* **81**:1084–1097
10. Alexander DH, *et al.* 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* doi:10.1101/gr.094052.109
11. Plettner E, *et al.* 1998. Biosynthesis of Mandibular Acids in Honey Bees (Apis mellifera): *De novo* Synthesis, Route of Fatty Acid Hydroxylation and Caste Selective β-Oxidation. *Insect Biochemistry and Molecular Biology* **28**:31–42
12. Chertemps T, *et al.* 2005. A new elongase selectively expressed in Drosophila male reproductive system. *Biochem Biophys Res Commun.* **333**:1066–1072
13. Farooqui T, 2008. Iron-induced oxidative stress modulates olfactory learning and memory in honeybees. *Behav Neurosci.* **122**:433–447

USDA NIFA, Foundational Programs, Plant-Associated Insects and Nematodes Program provided support for attending this conference.



RESULTS

Alignment metrics of haploids indicate a mean depth of coverage across all data of **7X**. On average, reads mapped to **92%** of the genome, of which **69%** was considered callable (depth ≥ 3X). The average number of variants identified per caller, and the percentage retained after combining with *BAYSIC* were: *GATK* 952K (89%), *Pileup* 846K (96%), *Platypus* 492K (99%); resulting in a final set of **845K SNPs per individual**. The combined data resulted in **2.5M SNPs overall**, post-filtering, distributed across chromosomes 1 to 16.

Analysis of population structure using *ADMIXTURE* (**Fig A**; optimal K=4) indicates that **HN and RJ bees originate from the C lineage**, from which the majority of domestic strains are derived. This is supported by the results of the phylogenetic analyis (**Fig B**).

Tests for signatures of selection revealed a number of regions to exceed the significance thresholds (**Fig C**; F_{ST} 50kb windows). **One region in particular (chr6:5.7-6.4 Mb) reveals excess heterozygosity in HN bees and reduced heterozygosity in RJ bees** (**Fig D**; 5kb windows). The region includes two members of the cytochrome P450 family (*CYP49A1* and *CYP301A1*), a gene that encodes an elongation of very long chain fatty acids protein (*Elo68*), and a gene encoding a pheromone biosynthesis activating neuropeptide (*PBAN*). Visual inspection of genotypes within this region reveals the extent of selection, for instance in RJ bees at CYP49A1 (**Fig E**).

The most abundant components of royal jelly are ω-hydroxy fatty acids, and it is thought that these might be synthesized by the bees' cytochrome P450 system[11]. In Drosophila melanogaster, *Elo68* is believed to play a role in biosynthesis of the pheromone vaccenyl acetate[12]. Finally, different levels of the *PBAN* peptide have been reported in the brain of forager bees with predispositions for either nectar or pollen[13].

CONCLUSIONS

Sequencing of haploid drones circumvents issues of heterozygosity, enabling a lower depth of coverage for variant detection. An assessment of this strategy in 60 bees, from analysis of population structure and tests for signatures of selection, indicates that it will be effective for larger studies. Our initial results reveal an interesting region of divergence between HN and RJ bees, potentially influencing the cytochrome P450 and/or endocrine systems.

UMR1388 GenPhySE
31326 Castanet Tolosan, France
https://www-lgc.toulouse.inra.fr/intranet