ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

SHORT COMMUNICATION

doi: 10.1111/age.12599

Genome-wide scans between two honeybee populations reveal putative signatures of human-mediated selection

M. Parejo*†, D. Wragg‡§, D. Henriques¶, A. Vignal‡ and M. Neuditschko*

*Agroscope, Swiss Bee Research Centre, 3003 Bern, Switzerland. †Institute of Bee Health, Vetsuisse Faculty, University of Bern, 3003 Bern, Switzerland. ‡Institut National de la Recherche Agronomique, 31326 Castanet-Tolosan, France. §The Roslin Institute, University of Edinburgh, EH25 9RG Edinburgh, UK. ¶Mountain Research Centre (CIMO), Polytechnic Institute of Bragança, 5301-855 Bragança, Portugal.

Summary

Human-mediated selection has left signatures in the genomes of many domesticated animals, including the European dark honeybee, *Apis mellifera mellifera*, which has been selected by apiculturists for centuries. Using whole-genome sequence information, we investigated selection signatures in spatially separated honeybee subpopulations (Switzerland, $n = 39$ and France, $n = 17$). Three different test statistics were calculated in windows of 2 kb (fixation index, cross-population extended haplotype homozygosity and cross-population composite likelihood ratio) and combined into a recently developed composite selection score. Applying a stringent false discovery rate of 0.01, we identified six significant selective sweeps distributed across five chromosomes covering eight genes. These genes are associated with multiple molecular and biological functions, including regulation of transcription, receptor binding and signal transduction. Of particular interest is a selection signature on chromosome 1, which corresponds to the *WNT4* gene, the family of which is conserved across the animal kingdom with a variety of functions. In *Drosophila melanogaster*, *WNT4* alleles have been associated with differential wing, cross vein and abdominal phenotypes. Defining phenotypic characteristics of different *Apis mellifera* ssp., which are typically used as selection criteria, include colour and wing venation pattern. This signal is therefore likely to be a good candidate for human mediated-selection arising from different applied breeding practices in the two managed populations.

Keywords composite selection score, *Apis mellifera*, selection signatures, whole-genome

The Western honeybee, *Apis mellifera*, is the most economically valuable pollinator for agriculture (Gallai *et al.* 2009). Its domestication began at least 3000 years ago in the Near East (Crane 1999). Today managed honeybees are selected mostly for specific characteristics suitable for apiculture such as docility, productivity and swarming behaviour (Crane 1999). Another important criterion is the breeding of specific honeybee subspecies. Currently, more than 27 subspecies are recognized, differing in morphology and behaviour (Meixner *et al.* 2013). The European dark honeybee, *Apis mellifera mellifera*, has been selected by apiculturists for a few centuries based on various characteristics, including colour, hair length and wing morphology (Ruttner 1988). In particular, the cubital index, which measures the ratio between two vein segments that are split by a cross vein, is used for pure race breeding (Ruttner 1988). The pattern of the fore wing veins is heritable and specific for each breed of honeybees and therefore widely applied for breeding purposes (Ruttner 1988).

In a previous study, we identified genetic substructures in two geographically isolated *A. m. mellifera* populations from Switzerland and France (Parejo *et al.* 2016). The samples from Switzerland originated predominantly from conservatories, where conservation breeding efforts for *A. m. mellifera* began in the 1970s. The introduction of non-native honeybees, such as the Carniolan bee, *A. m. carnica*, preferred by apiculturists due to their docile nature and higher productivity, threatens the genetic composition of the native type through introgression (Parejo *et al.* 2016). To distinguish native from introduced honeybees, breeders have typically referred to wing morphology, in particular the cubital index, although recently DNA-testing based on

Address for correspondence

M. Parejo, Agroscope, Swiss Bee Research Centre, 3003 Bern, Switzerland.
E-mail: melanie.parejo@alumni.ethz.ch

Accepted for publication 05 August 2017

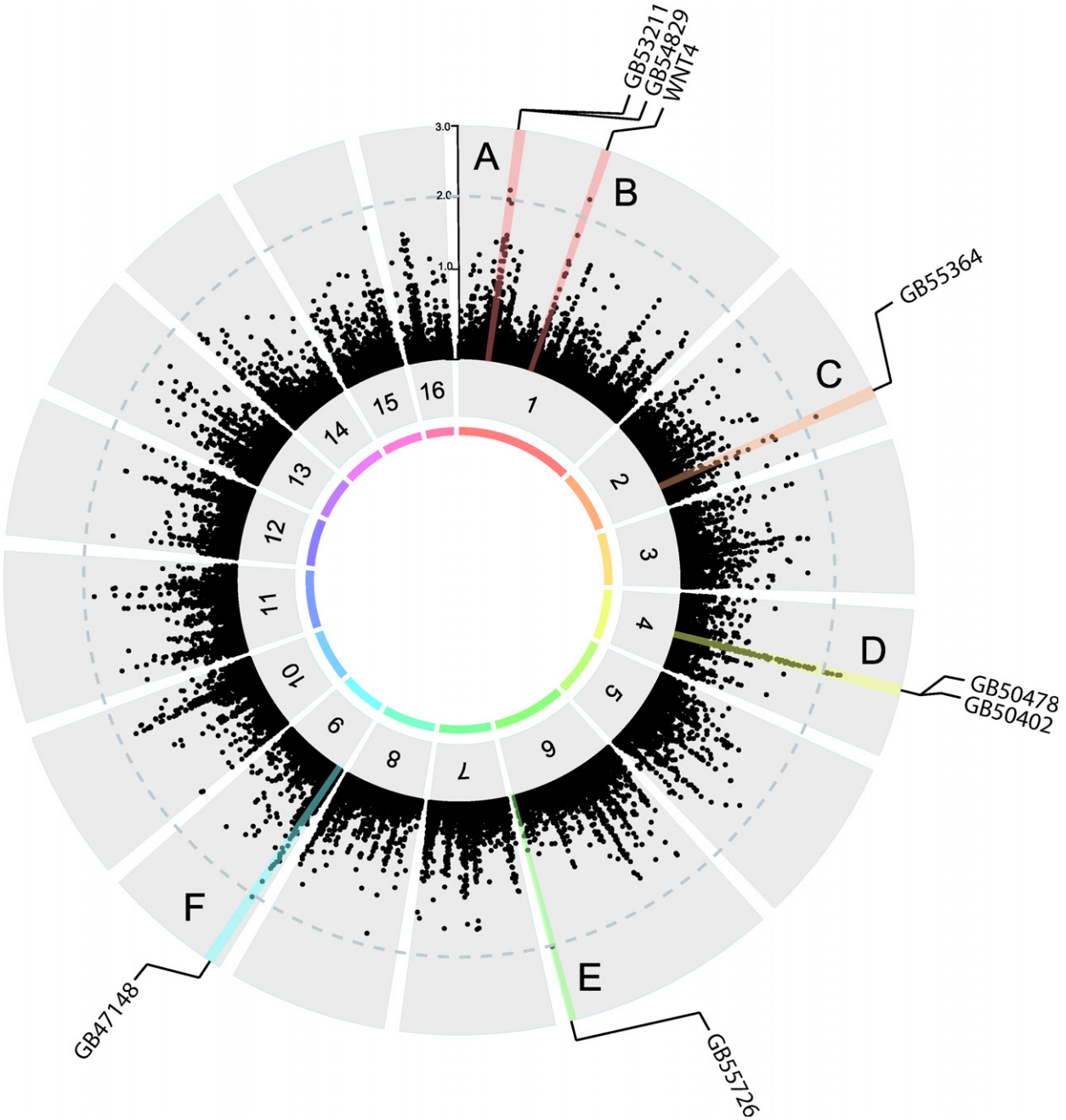

Figure 1 Analysis of genome-wide selection signatures identified six putative sweep regions (A–F). Composite selection scores combined F_{ST}, XP-EHH and XP-CLR test statistics in windows of 2 kb across the 16 honeybee chromosomes. The grey dashed line indicates the genome-wide significance threshold (false discovery rate = 0.01).

microsatellite markers is increasingly being used. The second population originated from Savoy, France, from a conservation breeding centre established in 1997. Here, selection was much more recent and based mainly on wing morphological parameters. Genetic diversity of this population is slightly higher than in the Swiss population (Parejo *et al.* 2016), potentially relating to the longer selective pressure against introgression from introduced bees in Switzerland. Thus, putative signals of selection could relate to differences in current and historical breeding regimes and the efficiency of breeding efforts to purge introduced alleles.

We investigated selection signatures between the two subpopulations using 2 924 632 SNPs identified from whole-genome sequence information of 56 *A. m. mellifera* drones (Switzerland, $n = 39$ and France, $n = 17$) (Parejo *et al.* 2016; see Table S1 for further information on sample

origin and data availability). To confidently identify selection signals, we calculated three different test statistics in non-overlapping windows of 2 kb: (i) fixation index, F_{ST} (Weir & Cockerham 1984), (ii) cross-population extended haplotype homozygosity, XP-EHH (Sabeti *et al.* 2007), and (iii) cross-population composite likelihood ratio, XP-CLR (Chen *et al.* 2010) (Fig. S1). Subsequently, we estimated the composite selection score (CSS) following Randhawa *et al.* (2014), which combines different test statistics based on a joint fractional rank. Finally, we applied a false discovery rate (FDR = 0.01) to CSS. Details on the calculations of these test statistics can be found in Appendix S1.

In total, we identified six putative sweep regions (A–F) distributed across five chromosomes including eight genes (Fig. 1, Table 1). Unfortunately, the honeybee genome is still not very well annotated, such that in sweep regions A, E and F only uncharacterized loci are located.

Yet, of particular note is the most significant window on chromosome 1 (sweep region B), which covers the *WNT4* gene. Acting as intercellular signals, wnt proteins confer polarity and asymmetry to cells that are proliferating and thereby give shape to tissues (Loh *et al.* 2016). In *Drosophila melanogaster*, *WNT4* alleles have been associated with differential wing, wing hair, wing margin bristle, cross vein and abdominal phenotypes (Swarup & Verheyen 2012; Gramates *et al.* 2016). To have a better idea of the differences in this region between the two populations, we further examined the haplotype block in the most significant window and compared that to haplotypes found in other honeybee subspecies. Forty-one SNPs describing four major haplotype blocks were found within this window (Fig. S2b). Even though sample size in the French population was lower ($n = 17$), we observed higher haplotype diversity (Fig. S2a), whereas in the Swiss population one haplotype (1a) was dominant (28 out of 39 drones). Moreover, we identified one additional haplotype (5Clin) in the French population that is predominantly present in C-lineage bees (Parejo *et al.* 2016; D. Wragg & A. Vignal, unpublished data). These findings suggest that conservation breeding efforts have not entirely purged all foreign alleles from the French population, whereas the longer and intensive selective pressure in the Swiss population has led to reduced haplotype diversity. Given that the honeybee has been selected on wing morphological characteristics, the signal found in this gene could thus be human-mediated and a result of differently applied breeding practices within the two subpopulations.

Among the eight genes located in the significant sweep regions, two [*LOC725294* (GB55364) and *LOC724717* (GB50478)] encode tyrosine-protein phosphatases and are found on two different chromosomes. These enzymes are key regulatory components in signal transduction pathways by regulating enzyme activity and controlling cell growth and differentiation (Tonks 2006). Therefore, given their relevance, the selection signal found in the two tyrosine-

Table 1 Candidate genes located within and around (±2 kb) the six sweep regions.

Sweep region	BeeBase/Gene ID	Gene name	Molecular function	Biological function	Cellular component	Drosophila ortholog
A	GB53211/*LOC100576626*	*Uncharacterized LOC100576626*				
	GB54829/*LOC408614*	*Uncharacterized LOC408614*		Neuropeptide signaling pathway	Secretory granule	FBgn0041707, *7B2*
B	GB44787/*WNT4* also known as *LOC552374*	*Wnt family member 4*, also known as *protein Wnt-1*	Receptor binding	Multicellular organism development, Wnt signaling pathway	Extracellular region, proteinaceous extracellular matrix	FBgn0010453, *WNT4*
C	GB55364/*LOC725294*	*Tyrosine-protein phosphatase 99A-like*	Protein tyrosine phosphatase activity, protein binding	Protein dephosphorylation	Integral component of membrane	FBgn0004369, *PTP99A*
D	GB50478/*LOC724717*	*Uncharacterized LOC724717* (Tyrosine-protein phosphatase gamma)			Integral component of membrane	FBgn0038749, *Xport-A*
	GB50402/*LOC412801*	*Mediator of RNA polymerase II transcription subunit 12-like*	Guanyl-nucleotide exchange factor activity	Small gtpase mediated signal transduction	Intracellular	FBgn0037188
E	GB55726/*LOC107964680*	*Uncharacterized LOC107964680*				
F	GB47148/*LOC552326*	*Uncharacterized LOC552326*				FBgn0032913

protein phosphatase genes have the potential to manifest in differential phenotypes. However, further research is needed to identify the trait(s) associated with these genes in the honeybee.

Finally, sweep region D also entails a gene [*LOC412801* (GB50402)] that is involved in the regulation of transcription of RNA polymerase II-dependent genes. Acting as a co-activator in the mediator complex, it is vital to regulatory mechanisms with a broad and dynamic range of functions (Malik & Roeder 2010).

In conclusion, we identified six sweep regions across the genome including eight genes, of which four have unknown functions and four are annotated for important molecular and biological functions. Collectively, these findings suggest that differential selective pressures are acting on these genes in these two closely related populations. However, it needs to be mentioned that selection signature analyses can reveal only putative candidate genes whose functional relevance on phenotypic differences remains to be tested. The strongest selection signal was found in *WNT4*, a gene affecting wing vein patterns in *D. melanogaster*. In addition, *A. m. mellifera* has been intensely selected on wing veins for decades. This further evidence in the case of *WNT4* makes it a plausible candidate gene for wing venation patterns in *A. mellifera* and an exemplification of human-mediated selection in the Western honeybee.

Acknowledgements

We thank the beekeepers and beekeeping associations mellifera.ch and CETA for providing samples and the Swiss Federal Office for Agriculture and Sur-la-Croix Foundation for funding.

References

Chen H., Patterson N. & Reich D. (2010) Population differentiation as a test for selective sweeps. *Genome Research* **20**, 393–402.

Crane E. (1999) *The World History of Beekeeping and Honey Hunting*. Routledge, New York, NY.

Gallai N., Salles J.-M., Settele J. & Vaissière B.E. (2009) Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics* **68**, 810–21.

Gramates L.S., Marygold S.J., Santos G. *et al.* (2016) FlyBase at 25: looking to the future. *Nucleic Acids Research*, **45**, D663–71.

Loh K.M., van Amerongen R. & Nusse R. (2016) Generating cellular diversity and spatial form: Wnt signaling and the evolution of multicellular animals. *Developmental Cell* **38**, 643–55.

Malik S. & Roeder R.G. (2010) The metazoan mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Reviews Genetics* **11**, 761–72.

Meixner M.D., Pinto M.A., Bouga M., Kryger P., Ivanova E. & Fuchs S. (2013) Standard methods for characterising subspecies and ecotypes of *Apis mellifera*. *Journal of Apicultural Research* **52**, 1–28.

Parejo M., Wragg D., Gauthier L., Vignal A., Neumann P. & Neuditschko M. (2016) Using whole-genome sequence information to foster conservation efforts for the European Dark honeybee. *Frontiers in Ecology and Evolution* **4**, 140.

Randhawa I.A.S., Khatkar M.S., Thomson P.C. & Raadsma H.W. (2014) Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. *BMC Genetics* **15**, 34.

Ruttner F. (1988) *Breeding Techniques and Selection for Breeding of the Honeybee*. BIBBA, London.

Sabeti P.C., Varilly P., Fry B. *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–8.

Swarup S. & Verheyen E.M. (2012) Wnt/wingless signaling in Drosophila. *Cold Spring Harbor Perspectives in Biology* **4**, a007930.

Tonks N.K. (2006) Protein tyrosine phosphatases: from genes, to function, to disease. *Nature Reviews Molecular Cell Biology* **7**, 833–46.

Weir B.S. & Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–70.

Supporting information

Additional supporting information may be found online in the supporting information tab for this article:

Figure S1 Manhattan plots of the three employed statistics to infer selection signatures (F_{ST}, XP-EHH, and XP-CLR) in windows of 2 kb.

Figure S2 (a) Haplotype distribution of sweep region B in the Swiss and French populations, and (b) corresponding haplotype blocks per individual.

Table S1 Further information on sample origin and data availability.

Appendix S1 Materials and methods.

