Proc. Assoc. Advmt. Anim. Breed. Genet. 26: 315-318

USING GENOTYPING BY SEQUENCING TO MONITOR THE GENETIC DIVERSITY OF AUSTRALIAN HONEY BEES

P. Alexandri[1], S.P. Miller[1], E.A. Frost[2], N.C. Chapman[2], S.F. Walkom[1], R.G. Banks[1] and K.L. Bunter[1]

[1] Animal Genetics Breeding Unit*, University of New England, Armidale, NSW, 2351 Australia
[2] NSW Department of Primary Industries and Regional Development, Paterson, NSW, 2421 Australia

SUMMARY

European honey bees in Australia are largely genetically isolated due to strict border controls. They contribute significantly to the economy through honey production and pollination, yet national efforts for their genetic improvement have historically been limited. This study estimated the genetic diversity of Australian bees using low-pass genome sequencing on 711 samples from 26 locations, provided by breeders participating in the national PlanBee project. Most genotypes were obtained from pooled drone samples. Results showed average observed heterozygosity (Ho: 0.20) was lower than the expected one. Bias tended towards low Ho, with some pooled drone samples conversely exhibiting unexpectedly high values. These high values were potentially attributed to multiple queen ancestries of the drone pool. Low FST values (0-0.07) between sampling locations indicated minimal population structure, likely due to gene flow through the exchange of genetic material within the PlanBee project. Findings suggest a need for broader sampling and better documentation of hive history and queen lineage.

INTRODUCTION

European honey bees were initially introduced to Australia from Western Europe with later introductions from Eastern Europe and the Mediterranean. These introductions resulted in an admixed population (Chapman *et al.* 2016), which has subsequently been isolated due to geographical isolation, strict border control protocols, and limited importation of bee genetic material. European honey bees contribute to the Australian economy both directly through honey production and indirectly by providing pollination services for various crops (Chapman *et al*. 2022). Despite their importance, limited attempts have been made to genetically improve bee populations. Furthermore, previous studies of Australian bee diversity used a limited number of areas and markers (i.e. 95 single nucleotide polymorphisms -SNPs, Chapman *et al.* 2016; Chapman *et al*. 2019). Accurately benchmarking and monitoring the genetic diversity of Australian bees is important because bees are kept in highly variable environments influencing their productivity and viability. The recent arrival of the Varroa mite (*Varroa destructor*) in Australia poses an additional challenge as loss of genetic diversity may limit the potential to select for varroa mite resistance (Büchler *et al.* 2010). This work aimed to assess genetic diversity in the Australian bee population using low-pass genome sequencing data, with a focus on identifying population structure and gene flow between different bee-keeping locations.

MATERIALS AND METHODS

Queen bee tissues and pooled drone larvae and pupae tissues were collected from 26 different locations or apiaries across Australia. The number of drones for pooled samples ranged from 5 to

* A joint venture of NSW Department of Primary Industries and Regional Development and the University of New England

10. A total of 1235 samples were submitted to Gencove (https://gencove.com/) for DNA extraction and low-pass sequencing at two different depths (4x and 10x) using a commercially developed method. DNA was successfully extracted and sequence data were obtained for 717 samples.

Table 1. Number of drones (N_D), queens (N_Q) and total number of samples (N) from each location retained after quality control

State	Locations	N_D	N_Q	N
WA	1-10	41		41
NSW	11-16	211	152	374
SA	17-19	18		18
QLD	20	30		30
TAS	21-24	219		219
VIC	25-26	40		40
Total	26	559	152	711

During the initial analysis of sequencing data, paired-end reads were aligned to the *Apis mellifera* genome Amel_HAv3.1 (GCF_003254395.2, Wallberg *et al*. 2019) using Sentieon Driver's BWA-MEM implementation (Freed *et al*. *unpublished data*). At this stage, 6 samples were excluded because of low (<60%) alignment. Haplotype calling was performed using Sentieon's `Haplotyper` with genome variant call format (gVCF, with information for both variant and non-variant positions) output per sample. Joint variant calling for processed haplotypes was performed using Sentieon's `GVCFtyper` with a maximum number of 12 alternate alleles reported. Resulting variant call format (vcf) files containing 2,827,009 variable sites were combined and further quality control was applied to remove variants with a missing rate of more than 10%, Minor Allele Frequency (MAF) lower than 5% as well as variants deviating from Hardy – Weinberg Equilibrium ($P < 0.0005$). After quality control 711 (Table 1) samples genotyped in 48,874 SNPs could be used for further analysis. To estimate levels of genetic diversity in the sampled locations, heterozygosity was estimated with PLINK 2.0 (Chang *et al*. 2015). The existence of more than one distinct population was assessed using pairwise genetic differentiation (F_{ST}) using the method implemented in Arlequin 3.5.2 (Excoffier *et al*. 2010).

RESULTS AND DISCUSSION

Observed queen heterozygosity values for each location were estimated either directly from the queen's genotype or indirectly using pooled drone genotypes (Figure 1). The average expected heterozygosity in the data was 0.23. Across all locations, observed heterozygosity was 0.20 and ranged from 0.04 to 0.50 (Figure 1). For the queen genotypes only, the average observed heterozygosity was 0.21 and it ranged between 0.15 and 0.47. For pooled drone genotypes, the highest 10% of heterozygosity values were observed in samples with more than 6 individuals, indicating a higher number of heterozygous SNPs in relatively larger pooled samples of drones. Including more individuals in the pooled sample may give a better representation of the queen's genome and provide the opportunity to capture more of the queen's alleles. However, higher heterozygosity in pooled drone genotypes can also be a result of the drones descending from more than one queen. Petersen *et al.* (2017) reported drones of different ancestry were present in pooled samples that came from hives where the queens had recently been replaced, highlighting the importance of recording the hive history and the queen's age or time in that hive before sampling.

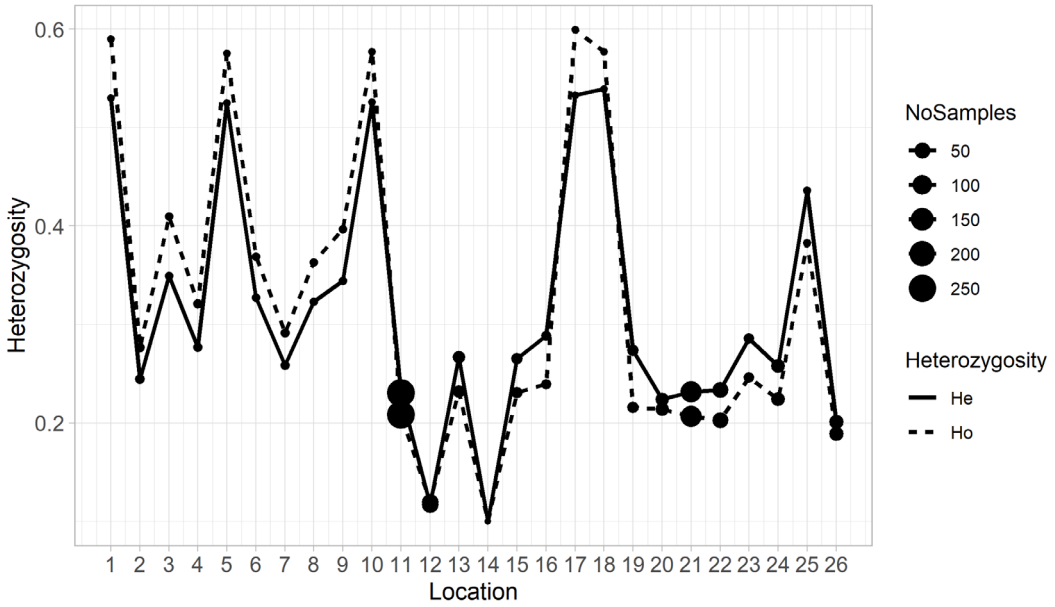

Figure 1. Observed (Ho) and expected (He) heterozygosity values for each location. Dot sizes indicate the number of samples from each location. Location codes correspond to Table 1

Previous studies have demonstrated the hybrid origin of Australian bees (Chapman *et al.* 2016) and the contribution of management and admixture to increased diversity in bee populations has been well described (Harpur *et al.* 2012). However, most samples in this study had lower heterozygosity than expected, potentially due to lower levels of genetic diversity within the sampled populations. A similar study in the New Zealand bee population revealed that although the overall levels of diversity were high enough to ensure long-term viability, within regions and companies the diversity was significantly reduced (Petersen *et al*. 2021). Moreover, low heterozygosity for Australian bees as a result of the sampling process cannot be excluded and more diverse sampling may be needed. Pairwise F_{ST} values estimated for all locations ranged from 0 to 0.07 (Figure 2), indicating no distinct subpopulations. Western Australian bee samples were expected to be more differentiated based on previous results using microsatellites (N. Chapman, unpublished data), reflecting a ban prohibiting importations to WA from other states. Genomic data used in this study captured greater variation on the genome and therefore would be more efficient at detecting population differentiation than using a small number of microsatellite markers. Further, several potential sub-populations (e.g. states) were represented by few samples, and some breeders were known to exchange genetic material between locations. Therefore, the low differentiation between Western Australia and the other sites likely reflects the sampling process and the existence of gene flow through queen transfer from Western Australia to Eastern states during the PlanBee project.

CONCLUSION

The study found that the observed heterozygosity (0.20) across sampling locations was slightly lower than expected (0.23). Low FST values (0–0.07) between location indicate minimal population differentiation and suggest a more homogeneous population due to gene flow from queen movement. The lower-than-expected heterozygosity may reflect high homozygosity but also limited sample diversity, highlighting the need for broader sampling. Sufficient drones per pool are required for

better representation of the queen's genome, but better documentation of hive history and queen lineage is also needed to avoid bias from drones originating from multiple queens.

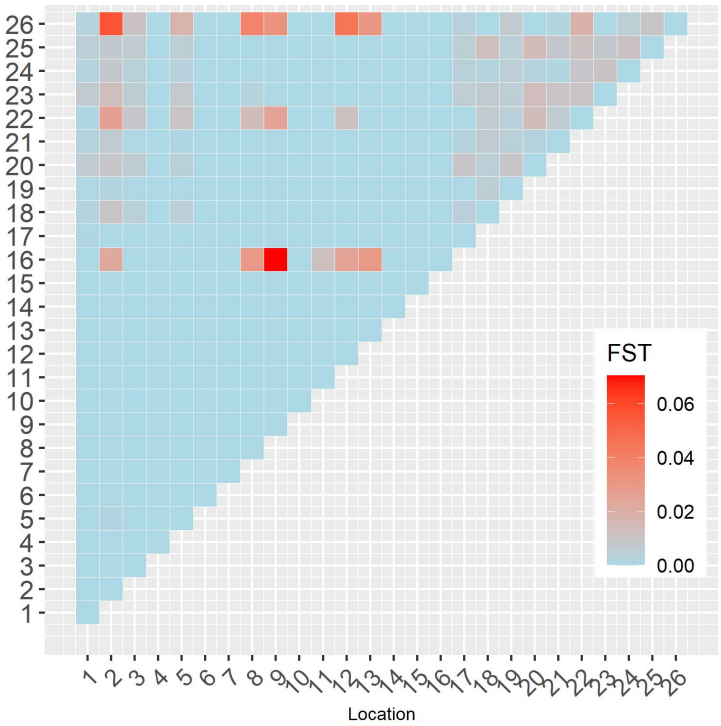

Figure 2. Pairwise genetic differences (FST) between locations estimated for all sequenced samples. Location codes correspond to Table 1

ACKNOWLEDGEMENTS

PlanBee project was supported by AgriFutures Australia through funding from the Australian Government Department of Agriculture, Water and the Environment as part of its Rural R&D for Profit program, participating research institutions and industry.

REFERENCES

Büchler R., Berg S. and Le Conte Y. (2010) *Apidologie* **41**: 393.
Chapman N.C., Harpur B.A., Lim J., *et al.* (2016) *Apidologie* **47**: 26.
Chapman N.C., Sheng S., Lim J., *et al*. (2019) *Apidologie* **50**: 28.
Chapman N.C., Frost E.A. and Banks R.G. (2022) *Proc. World Cong. Genet. Appl. Livest. Prod.* **12**: 2512.
Chang C., Chow C.C., Tellier L.C.A.M., *et al.* (2015) *GigaScience*, **4**: s13742–015–0047–8.
Excoffier L. and Lischer H.E.L. (2010) *Mol. Ecol. Resour.* **10**: 564.
Harpur B.A., Minaei S., Kent C.F. and Zayed A. (2012) *Mol. Ecol.* **21**: 4414.
Petersen G.E.L., Fennessy P.F., Van Stijn T.C.,*et al.* (2017). *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22**: 217.
Petersen G.E.L., Fennessy P.F. and Dearden P.K. (2021) *Anim. Prod. Sci.* **61**: 1856.
Wallberg A., Bunikis I., Vinnere Petterson O., *et al.* (2019) *BMC Genomics* **20**: 275.