

## GENOTYPING-BY-SEQUENCING FOR GENETIC IMPROVEMENT IN HONEYBEES

G. E. L. Petersen<sup>1,2</sup>, P. F. Fennessy<sup>2</sup>, T. C. Van Stijn<sup>3</sup>, S. M. Clarke<sup>3</sup> and P. K. Dearden<sup>1</sup>

<sup>1</sup> Department of Biochemistry, University of Otago, Dunedin, New Zealand

<sup>2</sup> AbacusBio Limited, PO Box 5585, Dunedin, New Zealand

<sup>3</sup> AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand

### SUMMARY

The polyandrous mating habits of honeybee queens, the small size of the animals themselves and the eusocial organisation of honeybee colonies present unique challenges for the establishment of genetic improvement programs.

Here we present a method to genotype honeybee queens using pools of their male offspring as proxies in a Genotyping-by-Sequencing (GBS) protocol.

GBS makes use of restriction site-associated short reads of DNA sequencing, effectively sampling the genome. Aligning these short reads to the reference genome provides a reliable and repeatable, but cost-effective genotyping protocol for honeybee queens. We found contamination of drone pools with unrelated drones to be an issue which can be overcome by using drone larvae directly from the brood comb.

### INTRODUCTION

The Western Honeybee, *Apis mellifera*, has shifted into the focus of applied animal science due to the increase in importance of pollination services and rising prices for monofloral honeys. Simultaneously, honeybee breeding efforts have become more sophisticated, leading to the development of specialised elite queen breeding operations. While elite queen breeders have been able to adopt a number of breeding techniques from other livestock species, such as performance testing and artificial insemination (AI), their work is subject to a number of biological limitations.

Honeybees are organised in colonies of up to 60,000 or so individuals, with only one reproducing female per colony, the queen, and a handful of reproducing males, the drones. The bulk of a honeybee colony is made up of functionally sterile female workers. As descendants of the same queen mother, workers are either half- or full sisters that are constantly being replenished. Honeybee queens produce up to 2,000 eggs per day, both fertilized and unfertilized. Fertilized eggs develop into females, while unfertilised eggs produce haploid males. The sperm stores are accumulated during one virginal mating flight, where the queen mates with 6 to 25 drones.

Polyandry and the resulting diversity within the work force have been shown to be crucial factors for colony fitness (Mattila *et al.* 2007). However, they severely limit the accuracy of hive/queen pedigree, making it impossible to record pedigree beyond maternal line and a list of potential drone sources (queens presiding over colonies in the same area or used to gather drones for AI). AI can be a valuable tool to control mating, but if from a single drone source, it will diminish the performance of the resulting colony. Thus, queen genotyping is the only way to accurately determine a queen's ancestry without compromising colony performance. Unfortunately, the continuous policing of the physical state of the queen by her workers complicates sampling for genotyping, since the sampling of hemolymph or body parts (e.g. tarsus, antennae) will result in death, either directly or by enticing workers to overthrow the impaired queen.

In light of these obstacles, a new approach to the genotyping of breeder queens is necessary. With the haploid male drones arising from unfertilised eggs, they potentially provide a sample of the queen's genome without interfering with her physically. This paper describes results of a pilot study to evaluate the efficacy of assessing the DNA of a queen through genotyping her haploid drone offspring.

## MATERIALS AND METHODS

**Honeybee samples.** 7 samples of multiple drones and corresponding queens were obtained from a commercial beekeeping operation in Marlborough, New Zealand, between January 2016 (drones) and April-May 2016 (queens). In January, 10-14 adult drones were removed from each colony, immediately put on ice in the field and subsequently frozen at  $-20^{\circ}$  C. Simultaneously, basic information about the colonies was collected.

At arrival in the laboratory, drones were transferred into a  $-80^{\circ}$  freezer until required for DNA extraction. The same collection and storage procedures were followed for the queens. In addition, 1 old breeder queen (sample 8) was sampled with no corresponding drones, resulting in 8 samples.

**DNA extraction and Drone Pooling.** Genomic DNA (gDNA) was extracted using the “ZR-96 Tissue & Insect DNA Kit” (Zymo Research, Irvine, CA, USA). Due to known issues with PCR amplification of DNA extracted from honeybee heads (Boncristiani *et al.* 2011), thorax was chosen as the standard substrate for gDNA extraction from both drones and queens.

The standard protocol for the ZR-96 Tissue & Insect DNA Kit was followed, with a 2010 GenoGrinder® (SPEX® SamplePrep, Metuchen, NJ, USA) serving as tissue homogenizer, set to 1,200 rpm for 8 minutes.

For each of the 7 drone/queen samples, DNA from 5, 6, 7 and 8 drones was pooled, as well as from all drones in the sample. DNA concentrations for the pools via Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific, Waltham, MA, USA) ranged from 11.1µg/ml to 62.9µg/ml.

**Evaluation of different body parts.** For sample 8, DNA was extracted from thorax, legs, wings and head, to evaluate the option of using a part of the body that would require a less time-consuming insect dissection protocol.

**Genotyping-by-Sequencing.** A 96-well plate, containing DNA from the following was prepared for subsequent genotyping by GBS: 7 samples with 5 drone pools (5x, 6x, 7x, 8x, all drones) and 1 (sample 1), 2 (sample 4) or 3 (samples 2-3, 5-7) repeats of corresponding queens, as well as a set of DNA from queen sample 8 (3x legs, 1x head, 1x thorax, 1x wings). The plate also contained DNA from 20 individual drones from samples 6 and 7.

Following enzyme selection and adapter optimisation (data not shown), GBS was carried out using a double digest with ApeK1 and Msp1, following an optimised version of the original GBS protocol (Elshire *et al.* 2011, Dodds *et al.* 2015). 100 ng of DNA per “sample” was utilised to prepare an 80-sample indexed GBS library that was subsequently further purified using the Pippin Prep (SAGE Science, Beverly, MA, USA) with size selection to generate a 150-500bp DNA sequencing library. Single-end sequencing (1x100) was performed on an Illumina HiSeq2500 with v4 chemistry, yielding approximately 25Gb of raw sequence data per lane. Raw fastq files were quality checked using FastQC v0.10.1 (Andrews 2010).

**Data analysis, SNP calling and alignment.** Approximately 260 Million raw reads were processed with UNEAK, Tassel version 3.0.170, (Lu *et al.* 2013) to detect variants and report reference and alternative allele counts at variant sites. The resulting ~27,800 called SNPs were further processed to construct a relationship matrix using R software (KGD; Dodds *et al.* 2015). KGD unbiased estimates of relatedness were calculated via method 1 of VanRaden (2008) adjusted to account for sequence read depth at each individual SNP location including SNPs with zero/missing reads.

In order to create a more consistent methodology that can be evaluated across GBS runs with potentially different restriction enzyme cut patterns, the sequencing data were additionally analysed using TASSEL 5 GBSv2 (Glaubitz *et al.* 2014) and the Burrows-Wheeler Aligner (Li *et al.* 2010). In this process, the short reads were aligned to the *A. mellifera* reference genome (Weinstock *et al.* 2006) before SNP calling. After alignment, ~46,400 SNPs could be fed into the KGD R software.

**RESULTS AND DISCUSSION**

**GBS of honeybee drone pools and corresponding queens.** Samples 1 and 2 showed consistent internal relationships (> 0.98), while for sample 3, the relationship dropped markedly with the addition of drone #6 (to ~0.86) and recovered with the addition of more drones (to ~1.1). In all three cases, the observed relationships between drone pools and queen were similar to the relationships between queen repeats.

Two samples (4 and 5) showed strong relationships both within the drone pools and within the queen repeats (~1.1), but not between the drone pools and the queen from the respective hives (~0.6).

For samples 6 and 7, all drones contributing to the pools were genotyped individually to determine if they had been sampled correctly. These individual assessments showed that for sample 6, only 6 out of the 10 drones were in fact sons of the queen from the corresponding hive; for sample 7, this was true only for 4 out of the 10 drones. As a result, relationships between drone pools and the queen diminished with the addition of more drones.

**Drone pools as queen proxies.** Evaluation of the relationships within and between samples consisting of pools of varying numbers of adult drones and corresponding queen mothers showed that there is considerable variation in the accuracy with which the drone pools reflect the genome of the hive queen.

Under ideal conditions, when the sampled drones are descendants of the targeted queen, drone pools appear to be a valid way to genotype their queen mother (see Table 1).

**Table 1. Relationships between drone pools and their mother queen (results shown for sample 2)**

	5x	6x	7x	8x	total	queen	queen	queen
5x	1.33	1.00	0.98	0.99	1.00	0.98	1.01	0.98
6x		1.31	0.99	0.99	0.99	0.98	0.99	0.99
7x			1.31	0.99	0.99	0.97	0.98	0.97
8x				1.25	0.99	0.99	0.99	0.99
total					1.30	0.98	1.02	1.00
queen						1.28	1.00	0.99
queen							1.32	1.00
queen								1.27

Most of the variation in the results could be traced back to the accidental sampling of unrelated drones as outlined below.

*Sampling of adult drones in a recently re-queened colony.* Samples 2 and 3 showed a pattern of strong relationships between drone pools, but weaker relationships between drone pools and queen. Queens in these colonies had been replaced prior to sample collection, and sampled drones

were descendants of the old queen, not the one presiding over the colony at collection. This highlights the importance of knowing the history of the hive before sampling.

*Displacement of adult drones.* Samples 5 and 7 showed limited relationships between drone pools and corresponding queens due to the fact that only 6 of 10 and 4 of 10 drones respectively were sons of their putative mothers. Both of these colonies were situated in a very tightly-packed yard in which adult drones returning to the hive could potentially drift over to another hive and end up in foreign colonies.

These two problems with the accidental sampling of mismatched drones can be overcome by switching from adult drones to drone brood as the source of DNA.

**GBS of different honeybee body parts.** There were 3 repeats of DNA extracted from the legs of queen 8; one of these failed to give a GBS result for unknown reasons. However other than this, relationships between different genotypes generated based on DNA extracted from different body parts were consistently over 0.87 (see Table 2).

**Table 2. Internal relationships for queen 8**

	legs	legs	head	wings	thorax
legs	1.17	0.90	0.88	0.90	0.91
legs		1.21	0.87	0.90	0.90
head			1.30	0.90	0.89
wings				1.24	0.92
thorax					1.20

These results are consistent with repeats of the same DNA extract. Moreover, the good GBS result based on DNA extracted from the queen's head suggests that with the use of the ZR-96 Insect & Tissue DNA Kit, DNA of a sufficient purity can be generated to avoid the PCR-inhibiting effects of honeybee compound eyes.

The use of wings as a non-lethal way to genotype bees has been suggested previously (Chaline *et al.* 2004). If sampling of an unmated queen is desired (e.g. to plan an AI mating), wings could present a valuable alternative to the proxy-based protocol presented here. Wing-origin DNA from queen 8 showed similar relationships to the other body parts both within and between samples (data not shown), but only 3.4 µg DNA per ml. Due to the low DNA concentration, wing clippings (~1/3 of the wing, removed to mark a queen) are unlikely to be a reliable source of DNA for GBS, but our findings show that it is possible to use whole sets of wings in cases where wings are not necessary for a queen's success (e.g. virgin AI queens).

## CONCLUSION

GBS of pooled drones is a reliable non-invasive genotyping method for honeybee queens, provided that a certain time has passed after a new queen is introduced into the hive and drones are sampled before emerging from the brood comb. Furthermore, by aligning the short reads generated via GBS to the *Apis mellifera* reference genome, this method can be used across populations. These findings will enable breeders of elite honeybees to take their breeding programs to the next level; for example, by controlling inbreeding without the use of restrictive breeding techniques such as single-drone AI.

## ACKNOWLEDGEMENTS

G. Petersen is currently a Callaghan Innovation PhD fellow in the Department of Biochemistry at the University of Otago, in cooperation with AbacusBio Limited and Taylor Pass Honey Company, Blenheim, New Zealand, with additional funding being provided by Taylor Pass Honey Co. This work was carried out in collaboration with the 'Genomics for Production & Security in a Biological Economy' C10X1306 MBIE AgResearch program.

## REFERENCES

- Andrews S. (2010). Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Boncrisiani, H. F., Li, J., Evans, J. D., Pettis, J. S. and Chen, Y. (2011) *Apidologie* **42**(4): 457–60.
- Chaline, N., Ratnieks, F. L. W., Raine, N. E. and Badcock, N. S. (2004). *Apidologie* **35**: 311–18.
- Dodds, K. G., McEwan, J. C., Brauning, R., Anderson, R. M., van Stijn, T. C., Kristjansson, T. and Clarke, S. (2015) *BMC Genomics* **16**(1047): 1–15.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. and Mitchell, S. E. (2011) *PLoS ONE* **6**(5): e19379.
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R., Sun, Q. and Buckler, E. S. (2014) *PLoS ONE* **9**(2).
- Li, H. and Durbin, R. (2010) *Bioinformatics*, Epub. [PMID: 20080505].
- Lu, F., Lipka, A. E., Glaubitz, J. C., Elshire, R., Cherney, J. H., Casler, M. D., Buckler, E. S. and Costich, D. E. (2013) *PLoS Genetics* **9**(1).
- Mattila, H. R. and Seeley, T. D. (2007) *Science* **317**: 362–64.
- Weinstock, G. M. *et al.* (2006) *Nature* **443**(7114):931-949.