Genotyping by Sequencing Enables Selective Breeding of Goat Herd

New Zealand–based research group uses GBS to overcome cost and resource hurdles associated with genomic selection for orphan crops and minor livestock species.

Introduction

Microarray-based genotyping enables herd managers and farmers to breed for specific genetic traits in a given species. Its use has grown rapidly in recent years, with successful applications in major species that are prevalent worldwide, such as dairy and beef cattle, sheep, corn, wheat, and sorghum. However, there remains significant cost, know-how, and resource barriers for developing and implementing microarray tools to support genomics-based breeding and selection programs in minor livestock and orphan crops species. The solution for these applications might be genotyping by sequencing (GBS).

AgResearch, a New Zealand–based research institute, was recently tasked by Meredith Dairy, an Australian goat farmer/cheese producer, to build a genomics-based selection program for its herd. Goats are a minor livestock species. Herd sizes are generally small and so are the budgets associated with farming them. The owners of Meredith Dairy, Sandy and Julie Cameron, had no genetic information about their herd, which consists of approximately 9000 goats. Principal Scientist John McEwan, and his AgResearch team used GBS to implement a genomic selection program for the Meredith Dairy goat herd. In less than six months, AgResearch provided the Camerons with highly predictive genomic estimated breeding values (GEBVs) for their herd.

iCommunity spoke with John McEwan to learn how GBS can aid those seeking genomic information for small herds or orphan crops, why it might be more cost effective than traditional arrays, and what the future holds for this technology in the improvement of breeding programs worldwide.

Q: What is AgResearch and what types of studies are you conducting?

John McEwan (JM): AgResearch is a privately owned company run by the New Zealand government as a Crown Research Institute. While it's a research institute, it's also a company, which allows us to be more commercial than if we were a government department. In our group, we have a trading name called GenomNZ* and we perform commercial genotyping for a

wide variety of species. In our lab, half of the people conduct agrigenomics research. Most of that is funded, or partially funded, by government or industry.

Q: How has agrigenomics research at AgResearch changed over the last 30 years?

JM: Genomics research at AgResearch began in 1985 and it has changed immensely over the years. For the first five years, there were two research focuses—DNA and protein. The protein research was used commercially to estimate breed composition in deer. The DNA research initially used methods such as Southern Blots. In 1990, we shifted to microsatellites and quantitative trait loci (QTL) mapping as well as DNA mix and match parentage. In the late 1990s, we performed more SNP[†] genotyping. However, we had to develop each marker, because we weren't studying animals that had a reference genome.

In the early 2000s, I was involved in the sequencing and development of the cattle genome, and pushed for completion of the sheep genome. We also performed SNP discovery. About 10 years ago, SNP microarrays or chips were developed and we adopted them rapidly in our research.

Q: Which species arrays did you develop?

JM: I have been directly involved in the development of Illumina 50K deer, 50K and 600K sheep arrays, and about five low-density sheep arrays.

Q: What is the value of arrays in the genomic selection of livestock?

JM: Illumina arrays are based on amazing technology. The major advantage is that they provide high-quality repeatable results. As long as the DNA isn't degraded, they are also robust and the sample processing is fast.

John McEwan is a Principal Scientist at AgResearch in New Zealand.

*GenomeNZ is the trademarked name of New Zealand's foremost commercial DNA testing laboratory, and is accredited by International Accreditation New Zealand (IIANZ) to the ISO 17025:2005.

[†]Single nucleotide polymorphism (SNP)

Meredith Dairy

Meredith Dairy is a family-owned, vertically integrated goat cheese and yogurt business. The farm is located in Meredith, Victoria, about a two-hour drive west of Melbourne. The total herd of approximately 9000 goats is a mix of three different breeds and has 6000 milking goats. The factory is on site, so the milk comes directly from the goat into the factory, where it is manufactured into cheese and yogurt. Meredith Dairy goat cheese is sold throughout the world.

Farm manager Nathan Ederton and Sandy Cameron, DVM, PhD in the milking parlor at Meredith Dairy.

There are approximately 9000 goats in the Meredith Dairy herd.

Q: What are the limitations of arrays?
JM: One of the limitations of arrays is that it is difficult to design arrays for species that have high genetic diversity. Some people also have very diverse species that they farm together. For example, they'll have Scottish red deer and Canadian elk that are so far diverged that we get severe ascertainment bias in most arrays unless we're extraordinarily careful in how we design them.

The major limitation of arrays is that the cost of developing arrays is high for minor species where there is a small number of individuals to be genotyped. We also need to genotype many individuals in order for arrays to be cost effective.

Q: When did you begin using GBS?
JM: Beginning in 2011, we realized that SNP chips were unlikely to be cost effective in minor species, particularly species with high genetic diversity. There were several GBS strategies and the one we selected was developed by the Buckler Maize Genetics and Diversity Lab at Cornell University.[1] We've modified that method, and made it cheaper and simpler to use.

Q: How does GBS compare to arrays on a cost and resource basis?
JM: On a cost basis, GBS using 30,000–100,000 SNPs is about the same price as using low-density arrays in a major species like cattle. If we analyze genomic selection data in the right way with GBS, there's also no need for imputation, which reduces resource and time costs. For example, people often use a mixture of high-density and low-density arrays when genotyping herds. Imputation is necessary to transfer the low-density arrays to high-density genotypes before performing genomic selection analysis. We don't have to do that for GBS because we have higher density genotypes in the first place. That's a significant savings in processing time and it means that we don't need large, high-density training sets.

> "We chose GBS over arrays because it has a low entry cost, is inexpensive for genotyping, and we don't have to do any imputation."

Q: What's the value of GBS in genomic selection?
JM: It has significant advantages, yet it's not for the faint hearted. We invested a significant amount of time in assay optimization and software development, and are now reaping the rewards of those efforts.

The first advantage of GBS is its low development cost. Second, it handles species with high divergence. Third, it can be used when there isn't a reference genome for the species we're analyzing. Fourth, if we use an appropriate discovery step for the variants, we can perform discovery and genotyping simultaneously, which reduces ascertainment bias issues. Finally, it's cheaper to run GBS after we have the workflow going, because we can use the same system to analyze many species simultaneously. That enables us to use the same lab protocol for multiple species. The result is that we can have samples from different species in the same library and in the same sequencing lanes. We use bioinformatics to deconvolute the samples and species.

Q: Which species have you used GBS to inform genomic selection?
JM: We initially used GBS in cattle and deer when we were developing our expertise. We carried on with method development in red deer, rye grass, and white clover. By 2014, we started to branch out into other species. Today we have developed GBS workflows for approximately 50 species.

Q: How are you using GBS for the Meredith Dairy goat herd?
JM: We're using GBS to genotype all the individuals in the herd. It's very difficult to record pedigrees of a dairy goat herd when there are thousands of animals. As a result, there was no pedigree data available for the individuals in the herd, only milk production

records. The bucks (male goats) didn't have any breeding values so the dairy's buck selection was *ad hoc*. By genotyping the whole herd, they now know exactly which does (female goats) have the best breeding values for all the traits being measured. They also know which bucks have been leaving daughters that are producing the most milk. To our knowledge, this is the first time GBS has ever been used for a project this size in goats.

Q: Why did you choose GBS for this genomic selection project?
JM: We chose GBS over arrays because it has a low entry cost, is inexpensive for genotyping, and we don't have to do any imputation. The 50K goat array is expensive to use for genomic selection of a small goat herd. To obtain a good price, we would need to purchase many chips.

In contrast, it cost ~$3500 US to develop GBS in goats. We used more than 50,000 SNPs and genotyped the herd for ~$18 US per animal. That price includes DNA extraction and breeding value determination.

> "We completed the project in ~3 months. GBS development took 1–2 months and the sequencing turnaround time was 4–6 weeks."

Q: How did you use GBS for assessing the GEBVs of the goat herd?
JM: We put the raw GBS reads through a bioinformatic pipeline and processed the results to identify variants. GBS results are raw call numbers, rather than genotypes. For example, two reads of variant A and one of variant B, or three reads of variant A and none of variant B.

Low-coverage sequencing also picks up more than just SNPs. GBS data from the SNP calling pipelines includes structural variants, gene duplications, fixed variations, or unusual inheritance patterns. We process the data through software that Ken G. Dodds developed. The KGD software converts the imprecise SNP calls and heterozygote undercalling into a genomic relationship matrix.[2] The genomic relationship matrix is equivalent to a pedigree that's been developed from SNP chips. We use the genomic relationship information and production measurements such as milk yield, composition, and live weight to calculate the breeding values.

The individual records are used across the entire herd to estimate the individual GEBVs based on how distant that individual is from all the others. Sometimes, we have a mixture of pedigree records on animals that have been genotyped. In that case, we use a single-step best linear unbiased predictor (ssBLUP) to join the pedigree matrix and the relationship matrix. Sometimes, the only information we have is from the animals that have been genotyped. In those cases, we use the genomic relationship matrix (GBLUP) alone.

Q: How long did it take for you to complete the Meredith Dairy project?
JM: We completed the project in ~3 months. GBS development took 1–2 months and the sequencing turnaround time was 4–6 weeks. There was also the time to determine project objectives and organize logistics, such as obtaining samples from about 9000 herd individuals.

> "GBS makes genomic selection possible and cost effective for any crop or species that someone wants to improve."

Q: How did Meredith Dairy feel about the data you provided about its herd?
JM: Sandy Cameron is very happy with the genomic information that we've provided using GBS. He's a veterinarian with a PhD in reproductive technology. Suddenly, he has a tool that he can use to inform embryo transfer and other reproductive decisions to support genomic selection.

Meredith Dairy is also vertically integrated. The dairy owns the goat herd and sells branded cheese. As a result, it obtains the full economic benefit of the genetic improvement.

> "The HiSeq™ 2500 System is performing well and we are very happy with the data that we've obtained. We rarely have to rerun samples, which means that the processing up front is good."

Q: What sequencing system do you use to perform GBS?
JM: We use single lanes of a HiSeq 2500 System, with 101 base pair reads. The HiSeq 2500 System is performing well and we are very happy with the data that we've obtained. We rarely have to rerun samples, which means that the processing up front is good. We're running 1–3 flow cells a week, with ~3000 samples a flow cell.

Q: What software do you use to analyze the data?
JM: The data are processed with standard Illumina software to produce the FASTQ files. We've developed the software from the QC pipeline through to the SNP calls, and it's on GitHub.[3] We've built most of the pipelines from existing software. The only new software we developed is the KGD genetics software to handle the raw GBS data. It is also available on GitHub.

Q: What are the next steps for this GBS project?
JM: There are six plant and animal species where we're using GBS for commercial genetic improvement. We're also using GBS to assist with conservation genetics of native New Zealand

species. For those projects, we're using GBS to determine if interisland populations interbreed and to track bycatch animals back to a particular island population.

Q: In the future, what do you think will be the tipping point for people to choose GBS over microarrays?

JM: As the software improves and the cost of sequencing declines, people will start sequencing individuals at low coverage. That's where cost reductions in sequencing will have a significant effect. My guess is we're at the US $18–20 price point at the moment. I believe the cost will slowly drop to about US $12 per sample over time. People will be busy over the next 5–10 years with the current opportunities.

Q: What benefit will GBS provide in implementing breeding selection programs for orphan crops and minor livestock species?

JM: The benefits of GBS are significant for orphan crops and minor livestock. GBS makes genomic selection possible and cost effective for any crop or species that someone wants to improve.

New Zealand is a small country that relies heavily on the export of high-quality food stuffs. We have many minor species that are the basis of products that we sell overseas, including dairy goat, venison, dairy sheep, Pacific salmon, green shell mussels, pinus radiata (Monterey pine), kiwi fruit, apples, and avocados. We need genomic improvement in all of them to be successful and remain competitive.

Q: In addition to genomic selection, where else does GBS provide an advantage in your research?

JM: GBS can be used in many ways. We can use it for linkage mapping to assist genome assemblies, and to perform GWAS or DNA parentage studies to estimate breed or strain composition. We can also use GBS for traceability or country of origin labeling. The significant challenge will be to develop the software for these applications to analyze GBS data efficiently.

Learn more about the Illumina systems and products mentioned in this article:

HiSeq 2500 System, www.illumina.com/systems/sequencing-platforms/hiseq-2500.html

References

1. Elshire RJ, Glaubitz JC, Sun Q, et al. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS One*. 2011; doi: 10.1371/journal.pone.0019379

2. Dodds KG, McEwan JC, Brauning R, et al. Construction of relatedness matrices using genotyping-by-sequencing data. *BMC Genomics*. 2015; doi: 10.1186/s12864-015-2252-2.

3. GitHub development platform. AgResearch/KGD. Kinship (genetic relatedness) using GBS (genotyping-by-sequencing) with Depth adjustment. www.github.com/AgResearch/KGD. Accessed October 10, 2017.

For Research Use Only. Not for use in diagnostic procedures.