GCRI Trust Travel Report (Jan 2017)

Tucson Plant Breeding Institute 2017 and Plant and Animal Genome Conference XXV

Joe Q He, PhD candidate, NIAB EMR, University of Reading

I am a PhD candidate associated with the University of Reading, based at NIAB EMR, investigating the implementation of genomic selection (GS) in a commercial strawberry breeding program. Genomic selection is an advanced breeding technique that requires a densely genotyped and phenotyped training population. It utilises a statistical model to regress the genotype against the phenotype allowing the prediction of performance of agronomically important traits in a breeding population. This technique is particularly suitable for polygenic traits where each loci has a small additive effect, which is the case for many traits in strawberry. GS is likely to deliver improved genetic gain per unit time and a reduction in the duration of the breeding cycle.

The Tucson Plant Breeding Institute 2017 took place in Tucson, Arizona, USA on the 9th – 13th January 2017. It offered state of the art instruction and training in statistical, molecular and computational plant breeding for practising breeders. The course was split into two modules: "Introduction to Plant Quantitative Genetics" and "Advanced Statistical Plant Breeding". The course was led by Bruce Walsh, Mike Gore and Lucia Gutierrez.

The first module provided an introduction to basic Mendelian genetics, population genetics and statistics. It continued with a discussion on the use of pedigree data and plot design in breeding and concluded with a description of the effects of inbreeding depression and heterosis. I was particularly interested in the lecture on field design by Gutierrez as it impressed upon me the importance of trials across multiple environments when attempting to deduce causative alleles for particular traits. Good field experimental design also allows the simplification of statistical models and improves power to detect real genetic effects by reducing confounding environmental noise. As I do not have a plant breeding background, this module also provided me with a broad overview of traditional plant breeding methods.

The second module focussed on advanced techniques that have been developed over the past two decades, which leverages the vast improvement in genome data collection rate to identify causative alleles and make predictions regarding plant performance. The first portion of this module recapped the basics of matrix algebra and mathematical methods to account for multiple traits of importance when selecting. Next, we discussed a range of models, based on the general linear model, which allows the prediction of plant genetic values, covariance of traits in plants and plant performance. We concluded by discussing the importance and difficulties with incorporating genotype by environment and epistasis interactions. I found the lecture on GS particularly interesting as this technique does not require significance testing (as marker assisted selection requires), removing the bias in prediction for large effect loci. It was also interesting to see that the most important factors in improving GS, in a case study of wheat, was to increase marker density, training

population size, and reduce training population age rather than to use more complex models that account for more biological phenomena.

The Plant and Animal Genome (PAG XXV) conference took place in San Diego, California, USA on the 14th – 18th of January 2017. This annual conference brought together over 3000 delegates with expertise in a range of fields to discuss the latest findings in genomic research, technologies and challenges. Due to the wide scope of the conference, multiple sessions were organised in parallel with each focussed on a particular topic. There was also a range of practical workshops, computational tutorials, organisational exhibits and a poster session with delegates free to move between sessions.

I was interested in sessions relating to the mathematical modelling of breeding populations to predict future performance, in particular GS. Whilst there were no examples of GS in strawberries, Fabio Cericola presented on "Optimising Training Population Size and Genotyping Strategy for Genomic Prediction Using Association Study Results and Pedigree Information. A Case of Study in Advanced Wheat Breeding Lines.". Utilising an advanced selection population of around 1000 wheat lines and a 15K SNP wheat genotyping array, he was able to create a genotyped and phenotyped training population to predict performance. Using genomic prediction based on a linear model, he was able to predict with an accuracy of around 0.6, a significant improvement on the 0.4 to be expected from phenotypic prediction alone. Reducing population size to 700 and marker density to just 1000 did not significantly affect the accuracy of prediction. Selecting specific markers on the basis of genome wide association studies could still yield high accuracies with as few as 100 markers allowing for more cost effective deployment of genomic prediction strategies. This adds further evidence that GS is an effective breeding tool and should be investigated for deployment in strawberries.

I was also able to attend sessions relating to the general importance of plant breeding in the context of human welfare. Gary Atlin presented on behalf of the Bill and Melinda Gates Foundation on "Designing Breeding and Seed Systems that Deliver Climate Change Adaptation.". Atlin highlighted that in many developing countries, the mean age of cultivars in subsistence farming is between 15 and 20 years old. These cultivars are unlikely to be well adapted to the current, probably different climate, leaving these farmers vulnerable to crop losses. Whist the technology and protocols to implement breeding programs to overcome this problem exist, lack of Governmental coordination, local expertise and financial support means that implementation remains slow.

Over 1200 posters on diverse topics were presented at the conference ranging from mathematical models on the effect of mis-genotyped bases in genomic studies to understanding chromosomal behaviour in the beefalo. I was particularly interested in discussing ideas and challenges with peers working on deploying genomic selection and understanding strawberry biology. Takanari Tanabata presented a poster on "Development of 3D Shape Measuring Method for Strawberry Fruits" where he utilised three RGB cameras directed at a strawberry atop a revolving platform to capture data. He demonstrated that this approach allowed him to generate in silico models of the berry and measure its dimensions and shape. Further work would allow other agronomically important traits, such colour and achene density to be determined from the same model. Quantitative,

standardised and accurate measurements of these traits is likely to improve the reliability of GS.

A range of commercial and charity organisations exhibited their technologies and services at PAG XXV. Data2Bio is a company, started in 2010 by academics, that specialises genotyping and managing the generated data. Interestingly, they offer support for breeders by performing GS on their behalf, suggesting that industry is increasingly realising the value of genomic prediction to aid breeding. However, Data2Bio have no proprietary GS models and is unlikely to outperform models already available; and their quote of approximately $65 per sample is likely to significantly exceed the cost of a breeder's in-house implementation.

The conference was concluded with a conference banquet, where I was able to discuss ideas and network with other delegates, including breeders who had also attended the Tucson Plant Breeding Institute 2017. I intend to maintain contact with them throughout my PhD and beyond.

I would like to take this opportunity to thank the AHDB and the GCRI Trust in generously supporting my travel to the USA.