

OECD Fellowship Summary Report

Peter Bradbury

Computational methods for haplotype identification, phasing, and imputation using short-read DNA sequence to enable low-cost genotyping

Theme number: theme 3, Transformational technologies and innovation, Advanced breeding tools/Genetic and genomic technologies

Host Institution: The University of Edinburgh

Host Collaborator: Dr. John Hickey

Dates: 22 September 2019 to 29 November 2019

I give my consent to post either the summary report or a short paragraph about it on the Co-operative Research Programme's website.

Summary Report

1. What were the objectives of the research project? Why is the research project important?

The objectives of the visit were to develop components for bioinformatic pipelines that would enable low cost genotyping for plant and animal breeders and geneticists and to work on methods for determining the maternal parent of seed from open pollination. My lab is developing a platform, called the Practical Haplotype Graph, that will be used for imputing genotypes from low coverage sequence, which will enable low cost genotyping. Target users for the application will be public sector plant breeders and geneticists. An important goal of the visit will be to incorporate methods and software developed by the host lab into that platform. Providing methods that enable breeders to make use of modern genomics to increase the rate of genetic improvements has great potential impact.

2. Were the objectives of the fellowship achieved?

The objectives of the fellowship were largely achieved. Software was written for imputing genotypes. It is anticipated that this software will become part of the Practical Haplotype Graph and be used by breeding and research programs for producing genotypes from low-cost DNA sequencing. In addition to writing software, I had valuable discussions with Hickey lab members about genotype phasing and parent identification.

3. What were the major achievements of the fellowship? (up to three)

The major achievements of the fellowship included identifying methods that would improve the efficiency and speed of genotype imputation by the Practical Haplotype Graph software, choosing one of those methods to implement, and writing software to do that.

4. Will there be any follow-up work?

The software that I wrote during my stay works well for small, simulated data sets, but is still too slow and uses too much memory for large, real data sets. I am continuing work on the software to resolve the remaining issues before it can be released. The work done while visiting the University of Edinburgh will probably not lead to a primary publication, though there are plans to present it as a poster at a scientific conference. There is a publication planned for the Practical Haplotype Graph to which this work will contribute. For that, the OECD will be acknowledged as a funding source. We also identified potential areas for future collaboration between the labs. In particular, both labs are interested in methods that make effective use of new long-read sequencing methods for phasing and imputation.

5. How might the results of your research project be important for helping develop regional, national or international agro-food, fisheries or forestry policies and, or practices, or be beneficial for society?

The work that I did during the visit funded by this fellowship, once completed, will result in improvements to the genotype imputation method currently implemented in the Practical Haplotype Graph. It will allow those methods to be effectively deployed at a scale useful to breeding programs, which might genotype tens of thousands individual samples or more each year. Improvements in the efficiency of plant and animal breeding programs worldwide will help breeders respond to an increasing demand for agricultural products and breed more resilient varieties to help cope with changing environments. Inexpensive genotyping and the software needed to generate and organize the data will be an important part of achieving those gains in efficiency.

6. How was this research relevant to the CRP?

The research funded by this travel grant falls directly under CRP theme 3, transformational technologies and innovation, advanced breeding tools/genetic and genomic technologies. Much progress has been made in recent years in methods for genomic assisted plant and animal breeding. Applying those

methods requires techniques for genotyping large numbers of individuals at a low cost per individual. Transformational technologies are those that enable basic research to be applied to practical problems on a large scale. A primary goal of this work was to provide software that enables the use of low-cost sequencing advances to be applied in breeding programs.

7. Satisfaction

The results of the visit funded by this fellowship exceeded my expectations. The members of my host lab were very welcoming and I had many useful discussions with them while I was there. One of the lab members in particular has worked extensively on related methods and gave me valuable guidance on adapting those approaches to my application. The visit will not impact my career opportunities because I am approaching retirement and not looking for new opportunities. I did not encounter any problems with the OECD fellowship grant.

8. Advertising the Co-operative Research Programme

I learned about the OECD co-operative research programme in an e-mail that was sent out by the international programs office of the USDA-ARS.