

## Scientific Report on

Exchange Grant

*'SNP chip data analysis for traditional Hungarian sheep and cattle breeds'*

within the framework of the ESF activity entitled  
'Advances in Farm Animal Genomic Resources'

by

Levente Czeglédi

University of Debrecen, Institute of Animal Science  
Hungary

25<sup>th</sup> June 2012

### 1. Purpose of the visit

The aim of this exchange grant is to learn bioinformatic tools for the molecular characterization of Hungarian autochthonous domestic breeds with new molecular genetic methods and to assist in the genetic management of these breeds that has so far been lacking. Our research group and institute will benefit from this exchange grant because of learning the state-of-art SNP based genetic analyses. Acquisition of this skill-set will dramatically improve population genetic studies at the University of Debrecen, Institute of Animal Science, which has focused only on using microsatellite markers up to now.

Molecular methods provide significant tools in population and conservation genetics for characterisation and genetic evaluation of species, breeds, varieties of wild and domestic animals. In recent years, microsatellite based analysis were dominant, but nowadays, methods based on genotyping for point mutations in the genome are more frequently used. Illumina 50k SNP chips data set provides huge amount of data of which handling, analysing was planned to carried out at the Organisms and Environment Research Division, Cardiff School of Biosciences, Cardiff, Wales. The high-density 50k-marker panel comprises approximately 50,000 single nucleotide polymorphisms (SNP) that have been identified in the bovine and ovine genome.

The results can enable in-depth analysis of the patterns of genome-wide diversity and differentiation across each breed, which can be used for genome-wide association studies and for future marker-assisted selection. Our aim was to use a variety of bioinformatic tools to analyse the high volume of data in the context of population genetics. This kind of analysis is carried out routinely within the host lab, thus Cardiff provided an ideal location for this training.

## 2. Description of the work carried out during the visit

Before the visit I have collected blood samples from Hungarian Grey cattle and Racka sheep as well. Animals of each species were divided into 2 groups: a) moderate performance individuals, production level of these animals is as the average of the breed; b) high performance individuals, these animals had the highest weight gain in the investigated population. Data of individuals – all of them are adult females before first calving – are date of birth, sire, dam, daily gain (from birth until weaning), additional information for sheep: weight at the beginning of grazing season, at end of grazing season, at age of 1 year.

DNA was extracted from blood and its quality was checked. Samples are on Illumina 50K SNP chip for lab analyses. All the costs of SNP chip work is financed by Prof. Mike Bruford's grant.

The primary aim of the visit was to study the application of data handling and bioinformatic tools for huge DNA data analyses. Prof. Mike Bruford provided me the 50K Sheep SNP chip data of Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, et al. (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biol* 10(2): e1001258. doi:10.1371/journal.pbio.1001258 published in February 2012 by the members of International Sheep Genomic Consortium. This dataset consisted of more than 49000 point mutation results per individual, 2819 animals belong to 83 breeds.

At first, I have started to be familiar with UNIX system. Commands of UNIX were used to manage, organize SNP chip result files. Unix operating systems are widely used and it has a special language.

The most important and appropriate software tools were PLINK and Haploview for chip dataset analysis.

PLINK is a whole genome association analysis toolset, freely available and developed by Center for Human Genetic Research, Massachusetts General Hospital, and the Broad Institute of Harvard & MIT (Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.). The software was used in UNIX environment.

The options which were used for working on .ped and .map files of chip data are grouped as follows: basic processes on data files; selection SNPs, individuals, breeds; creating summary statistics for whole dataset and subgroups; using thresholds; stratification and clustering; association analysis procedures; linkage disequilibrium pruning.

gPLINK is a free software as well, developed by the same groups. It is a simplified, user friendly form of PLINK, it contains the most commonly used PLINK commands. One can use gPLINK by clicking, however for PLINK correct command typing is required.

Haploview is a free software to visualize the analysis of PLINK results. It was developed and available from the Broad Institute (Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005 Jan 15, PubMed ID: 15297300). Haploview is able to use any output file from PLINK, it can run linkage disequilibrium and haplotype analysis, permutation testing for association, SNP and haplotype association tests, haplotype population frequency estimation.

### **3. Description of the main results obtained**

Main result of exchange visit is the knowledge of state-of-art bioinformatic tools. All the softwares mentioned in „description of work...” section provides the possibility to work on huge amount of SNP data. The experience I got at Mike Bruford’ lab is enables me to handle and work on my own genotype data of domestic animals and the gain of these skills makes it possible to apply for research grants in the field.

### **4. Future collaboration with host institution**

Mike Bruford’s lab and the Animal Genetics lab at University of Debrecen, Institute of Animal Science plan to continue the work that has been started with this grant. At present, the SNP chip data of Hungarian samples (cattle and sheep) will be analysed as a pilot study. The future collaboration will focus on domestic animal species, that could include autochthonous breeds for diversity studies or/and breeds with economic importance/intensive production for SNPs correlated to phenotype data, animal performance.

### **5. Projected publications / articles resulting or to result from the grant**

The grant gave the possibility to study bioinformatic tools for genome wide mutation analyses at the host institution. It provides a strong background to apply for research grants in the future to finance the cost of consumables, materials required for large scale studies. We do believe that this type of common work will result in publications.