

## **Summer Institute in Statistical Genetics 2008 – Seattle, Washington.**

Andrew Clempson, PhD Student at the Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Hertfordshire AL9 7TA.

### **Funding**

Funding was provided from the British Society of Animal Science, which contributed towards attendance to the Summer Institute in Statistical Genetics 2008, held at the University of Washington, Seattle.

### **Introduction**

I am currently a PhD student at the Royal Veterinary College, London, studying genetic selection of cattle for fertility and longevity. My research has recently become focussed on SNP analysis in a number of genes that are thought to control metabolic and morphological phenotypes in dairy cows. A wide range of different traits have been measured, such as length, weight, height, girth, IGF, insulin and glucose measurements. These traits have are thought to have significant effects on growth, fertility and lactation in dairy cows.

Studies will be performed to determine if certain SNPs are linked with these phenotypic traits, thus advancing knowledge of animal genetics in a hope to improve growth, fertility and lactation in dairy cows. This will require the use of univariate and multivariate analysis in a range of statistical packages such as 'ASReml' and 'R'. Both of these programs are extremely powerful for the completion of statistical operations, especially in terms of genetic data, and therefore, their use is paramount in this analysis. In order to achieve a thorough understanding of these topics, it was essential that I attended this course in statistical genetics, which proved to be an excellent opportunity to introduce the basic concepts in this field.

### **Location and Course Structure**

The Summer Institute in Statistical Genetics was held at the University of Washington, Seattle, and consisted of a series of modules composed of lectures and computer lab exercises. I attended four modules over a two-week period, which covered –

- Computing for Statistical Genetics
- Regression and Analysis of Variance
- Quantitative Genetics
- QTL Mapping

These modules were delivered by many experts in biostatistics and genetics, including Dr Thomas Lumley – a core developer of the R statistical software, Dr Bill Barlow – an expert in Cancer genetics, Prof. Bruce Weir – a leading expert in quantitative genetics and Dr Rebecca Doerge – an acclaimed researcher in the mapping of quantitative trait loci.

### **Overview of Modules**

#### **Computing for Statistical Genetics 16<sup>th</sup> – 18<sup>th</sup> June**

This module introduced the basics of programming genetic data into the R statistical software. This included a significant aspect of data management, syntax writing and analysis of output data. Specific analyses were performed using genetic data throughout this module to ensure a thorough understanding of the basic concepts. Towards the end of the course, association mapping and haplotype inference was also discussed.

This module was particularly beneficial, as the R software was introduced from a basic level, ensuring that all participants had a thorough understanding of the core concepts.

### **Regression and Analysis of Variance 18<sup>th</sup> – 20<sup>th</sup> June.**

This module covered topics such as standard linear regression, analysis of variance and logistic regression, and consisted of a series of lectures and computer lab exercises using R. Specific data sets were used throughout this module in order to gain familiarity with the data and output results.

Question and answer sessions were also held throughout the course to ensure that participants had a thorough understanding of the basic concepts.

### **Quantitative Genetics 23<sup>rd</sup> – 25<sup>th</sup> June**

This module provided an excellent introduction to quantitative genetics and introduced the concepts of genetic and environmental impacts on phenotypic traits. This was followed by more specific discussion of breeding values, Fischer's variance decomposition, heritability, selection, and inbreeding. Statistical formulae relating to the calculation of these components were discussed in both the 'R' and 'ASReml' environment, which has proved to be very useful in the understanding my analyses.

This module was particularly beneficial, as a wide variety of scientists from many different fields attended the course. This resulted in many discussions that were specific to individual cases, however, demonstrated the dynamic applications of quantitative genetics in today's research.

### **QTL Mapping 25<sup>th</sup> – 27<sup>th</sup> June**

This module introduced the wide range of methods of mapping quantitative trait loci in experimental cross populations. A combination of lectures and computer lab exercises were delivered to effectively introduce the basic concepts behind QTL mapping. Discussion was also directed towards linkage maps and their construction, and the progression of QTL mapping from single-marker analysis and interval mapping to composite interval mapping. Furthermore, the future of eQTL mapping was introduced, demonstrating the fast pace at which this area of research is advancing.

Computer lab exercises were also held, which consisted of an introduction to Windows QTL Cartographer, followed by analysis of genetic data and interpretation of the output files.

### **Relevance to my Research**

Attendance to this course has significantly increased my understanding of statistical genetics, and therefore will be extremely helpful for completion of my analyses. I have gained valuable knowledge on the 'R' and 'ASReml' statistical packages, and how they can be used to analyse a wide range of genetic data. This knowledge will be extremely useful for analysing my data and understanding the results. This will enable conclusions to be drawn in relation to the effect of different SNPs on phenotypic traits related to cattle growth, fertility and lactation, thus advancing knowledge in this area of animal science.

Throughout this course, I also had many opportunities to discuss my work with a wide range of scientists from around the globe. This has added a sense of dynamicism to my thoughts on future work. I have also built a number of contacts with a wide range of experts in the field, many of which will be very useful in the forth-coming years.

### **Acknowledgments**

I would like to thank the British Society of Animal Science for providing funding towards this trip. Without their generosity, attendance to this course would not have been possible.

I would also like to thank my supervisor – Prof. Claire Wathes – for her enthusiasm and full support of this trip.