# Biomathematics and Statistics Scotland

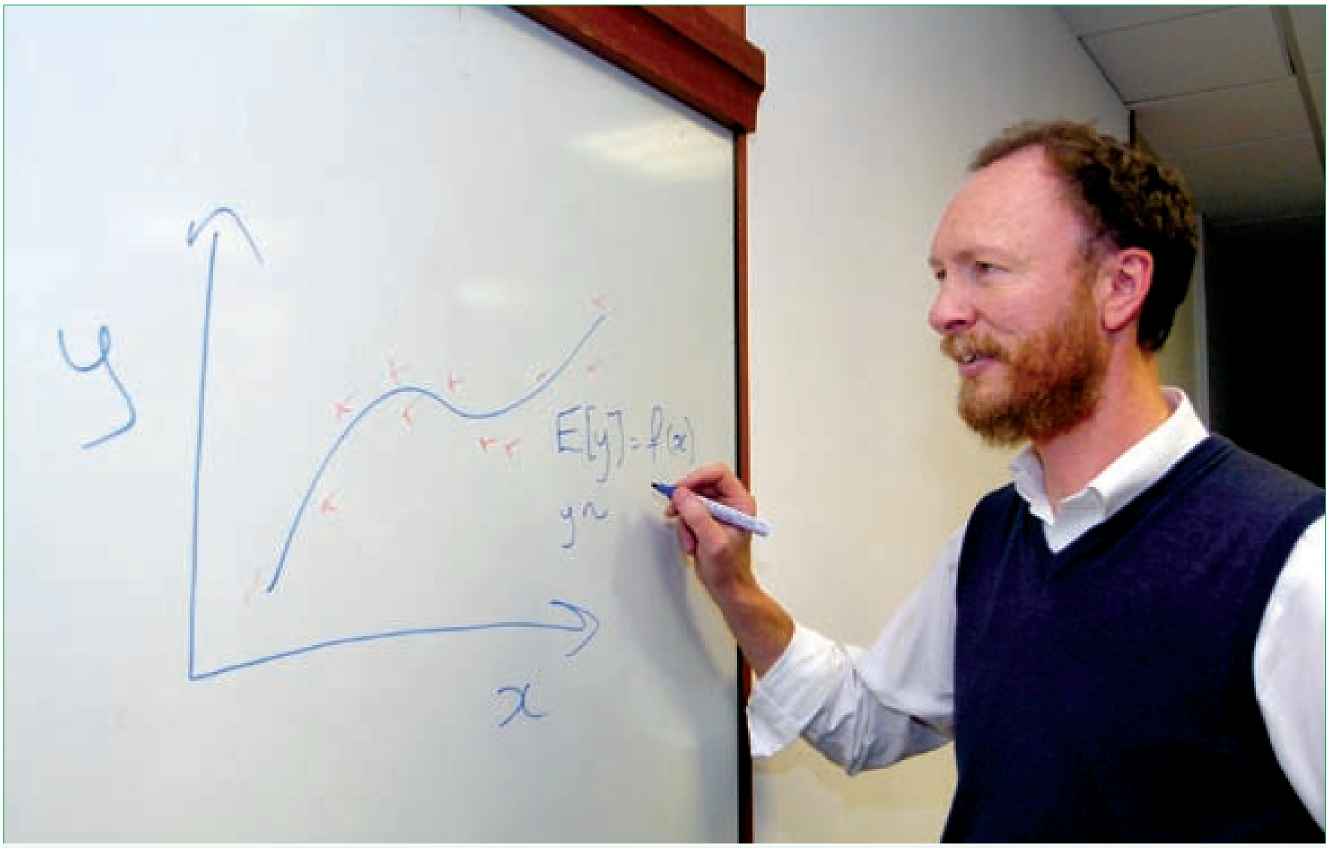
**David A. Elston**

*Biomathematics & Statistics Scotland (BioSS; www.bioss.ac.uk) delivers research, consultancy and training in statistics, mathematical modelling and bioinformatics. Although BioSS forms part of the SCRI Group, its work is equally relevant to a number of other scientific organisations. Indeed BioSS plays a unique role in the Scottish research community. It bridges the gap between research in the mathematically based and traditionally more qualitative sciences. BioSS staff work within four broad application areas: plant science; animal health and welfare; ecology & environmental science; and human health & nutrition. Our programme of applied strategic research addresses generic issues encountered in these application areas and is managed in three broad themes: statistical bioinformatics; systems & process modelling; and statistical methodology.*

## Improving Quantitative Trait Locus (QTL) analysis for key crop traits

**Chris A. Hackett**

During 2009 BioSS marked up 20 years of collaboration with SCRI scientists in the area of linkage analysis and QTL mapping. The central idea of estimating genetic distances between DNA markers on the basis of recombination between them, and relating variation at marker positions to variation in phenotypic traits to locate genetic regions affecting these traits, remains unchanged. However we have seen major changes in the types of population, the markers, the traits of interest and the statistical approaches needed. In this review we look back over this progression, and forward to new developments.



David Elston, Director of BioSS.

The first collaboration was an analysis of an $F_2$ barley population, relating height and ear emergence to two markers on chromosome 4. This required the development of a novel mixture model with two unlinked genes interacting epistatically to control the vernalisation requirement. Subsequent collaborations have involved many more markers, as different techniques have developed: Restriction Fragment Length Polymorphisms (RFLPs); Random Amplified Polymorphic DNA (RAPD); Amplified Fragment Length Polymorphisms (AFLPs); Simple Sequence Repeats (SSRs). Single Nucleotide Polymorphism (SNP) analysis and Diversity Arrays Technology (DArT) now enable linkage maps of barley to be constructed with more than 1000 markers.

Barley is an ideal crop for QTL analysis, as studies on barley typically begin with homozygous parents, where recombination can be estimated with few complications. However research at SCRI has expanded to include QTL studies for all the major crops here. In raspberries and blackcurrants, mapping populations are usually derived by crossing two (outbreeding) parents and studying their $F_1$ progeny. Potato has the further complication that the breeders' potato is autotetraploid. BioSS and SCRI obtained funding to develop the theory for mapping in autotetraploid species, and our methods have been incorporated in the software TetraploidMap for Windows. This has been used by other groups to develop linkage maps and locate QTLs in plants such as alfalfa, leek, coffee and rose and will be developed further to meet the needs of a new SCRI project on blueberry.

The phenotypic traits under investigation have also changed in number and type. The first collaboration involved two highly heritable characters. Subsequent studies have looked at yield, and yield components, which are less heritable. More recent traits have involved more complex analysis in advance of the QTL mapping and have required the expertise of BioSS in other areas: for example the analysis of sensory data from a tasting panel assessing different potatoes, and the analysis of colour data using image analysis techniques. A variety of statistical approaches can be used to look at modelling developmental processes: principal coordinates were used to summarise a series of repeated visual assessments of raspberry ripening, and analysed to identify QTLs affecting different stages of ripening. Another role of the preliminary analysis is to check for spatial trends across field trials, or temporal trends related to the ordering of samples through laboratory testing, which can conceal or bias QTL effects.

Mixed models are being increasingly used in QTL analyses. Their flexible correlation structure means that they can be used to represent the relationship between, say, yields measured in different locations or under different treatment regimes in a factorial experiment, so enabling tests for QTL by environment interactions. These interactions can, in turn, be modelled in a mixed model framework as functions of environmental variables such as temperature or rainfall. Mixed models can also be used to combine trait values from related populations, taking into account the degree of relatedness: the scale of this can range from combining two or three large mapping populations that share a common parent through to association mapping of germplasm collections.

In the future, the challenges are most likely to come from the increasing dimensions of phenotypic data as ‘omics technologies lead to measurements of metabolites or gene expressions, or image analysis methods such as LemnaTec give detailed pictures of crop development. Dealing with these large, potentially multivariate data sets will be challenging both statistically and computationally. We anticipate further developments in the types and quantities of marker data, and changes in the target traits as issues of sustainability and food security in a changing climate become increasingly important.