Multivariate meta-analysis of QTL mapping studies

Giustino Gaspa[1], Davide Perini[2], Rosanna Marino[2],
Raffaele Mazza[2], Silvia Sorbolini[1], Roberto Steri[1]

[1]Dipartimento di Scienze Zootecniche, Università di Sassari, Italy

[2]Istituto di Zootecnica, Università Cattolica del Sacro Cuore, Piacenza, Italy

Corresponding author: Giustino Gaspa. Dipartimento di Scienze Zootecniche, Facoltà di Agraria, Università di Sassari. Via E. De Nicola 9, 07100 Sassari, Italy – Tel. +39 079 229308 – Fax: +39 079229302 – Email: gigaspa@uniss.it

ABSTRACT

A large number of quantitative trait loci (QTLs) for milk production and quality traits in dairy cattle has been reported in literature. The large amount of information available could be exploited by meta-analyses to draw more general conclusions from results obtained in different experimental conditions (animals, statistical methodologies). QTL meta-analyses have been carried out to estimate the distribution of QTL effects in livestock and to find consensus on QTL position. In this study, multivariate dimension reduction techniques are used to analyse a database of dairy cattle QTL published results, in order to extract latent variables able to characterise the research. A total of 92 papers by 72 authors were found on 25 scientific Journals for the period January 1995-February 2008. More than thirty parameters were picked up from the articles. To overcome the problem of different map location, the flanking markers were mapped on release 4.1 of the *Bos taurus* genome sequence (www.ensembl.org). Their position was retrieved from public databases and, when absent, was calculated *in silico* by blasting (http://blast.wustl.edu/) the markers' nucleotide sequence against the genomic sequence. Records were discarded if flanking markers or P-values were not available. After these edits, the final archive consisted of 1,162 records. Seven selected variables were analysed both with the Factor Analysis (FA), combined with the varimax rotation technique, and Principal Component Analysis (PCA). FA was able to explain 68% of the original variability with 3 latent factors: the first factor extracted was highly associated (factor loading of 0.98) to marker location along the chromosome and could be considered as a marker map index; the second factor showed factor loadings of 0.74 and 0.84 related to the variable number of animals involved and year of the experiment, respectively, and it can be regarded as an indicator of the dimension of the study; the third factor was correlated to the significance level of the statistical test (0.78), number of families (0.63), and, negatively, to the marker density (-0.43). It can be named as index of power of the experiment. Same patterns can be observed in the eigenvectors of PCA. Four PCs were able to explain about 80% of the original variance. The first two PCs basically underlined accurately the same structure found with the first two factors in FA, whereas PC3 and PC4 summarized the structure of F3. The score that each QTL gets on each Factor or PC could be useful to classify the original QTL records and make them more comparable once that the redundancy of information has been removed.