

Experiences in bioinformatics

Luc L. Janss

Center for Quantitative Genetics and Genomics

Aarhus University

Denmark

Knowledge from bioinformatics research can in principle be used to improve genomic predictions. Examples are the use of databases with collected QTL mapping results, the use of genomic feature annotations, and pathway or Gene Ontology (GO) data. There are, however, several hurdles to appropriately use this information and include it in genomic models for analysis of livestock data. A limitation in the use of QTL mapping results is that results from classical Linkage Analysis studies have wide confidence intervals such that for any trait a large part of the genome will be tagged as 'QTL region'. Despite this limitation SNPs in QTL regions can be found to explain more variance than those outside QTL regions, which is for instance shown for milk and fat yield in dairy cattle, but not for protein yield. A limitation of genomic feature annotations is that this information is not covering the whole genome and is not directly related to traits, or, for pathway or GO data, is mostly categorized in fundamental biological processes. This makes it difficult to attach genuine prior information to this data. The common approach to use this kind of bioinformatics data is to simply try which features or GO groups explain more variance or predict better. This leads to results such as that SNPs in/near genes (but not necessarily the coding parts of genes) explain more variance in phenotypes. A final significant hurdle to use this information is that our animal populations are highly structured and include relatives. This leads to long-range LD and LD between chromosomes, which effectively spreads QTL effects over the whole genome. This creates a polygenic image which matches the assumptions of GBLUP that all SNPs contribute equally, and makes genomic relationships the main driver for genomic prediction within animal populations. Better modeling, notably a separation of effects of linkage and LD in genomic prediction, allows more meaningful inferences from bioinformatics data, and potentially allows to improve genomic predictions where relationships are weak, e.g. across breed.

Author background

Luc Janss is all-round developer of statistical tools for gene mapping, genomic prediction and general quantitative genetics applications. His tools are often based on Bayesian approaches and he is the main developer of the Bayesian analysis package Bayz. Research work includes Bayesian approaches for QTL and gene mapping, disease genetics, non-linear models, use of gene expression data, genotyping by sequencing and use of bioinformatics data, with applications in humans, animals and plants.