Adaptation of BLUPF90 package for genomic computations

Ignacy Misztal, Ignacio Aguilar, Shogo Tsuruta, Andres Legarra

Genomic best linear unbiased prediction by simulated annealing

Esquivelzeta, C. and Casellas, J., Universitat Autònoma de Barcelona, Facultat de Veterinària, Departament de Ciència Animal i dels Aliments, Campus Universitat Autònoma S/N, 08029 Cerdanyola del Vallès, Spain; ceciliaer@gmail.com

Genomic best linear unbiased prediction (gBLUP) models predict genomic breeding values by accounting for the information of thousands of single nucleotide polymorphisms (SNP) summarised into the genomic relationship matrix (G). Although current parametrizations use all SNP to compute G, it is highly questionable to believe that all SNP contribute reliable and relevant information when computing genomic relationship coefficients conditional to a given phenotype. Within this context, we simulated genomic data and performed serial gBLUP analyses by simulated annealing in order to identify which SNP contributed relevant information and which SNP must be removed from the calculation of G. We used as reference the mean square error (MSE) between simulated and predicted data, and the analytical process started with all SNP included in the calculation of G. For each new iteration, a SNP was selected at random and the MSE was calculated after changing its state (i.e., removed or included to the list of SNP for the calculation of G). This change was accepted if MSE reduced from previous iteration. The simulated annealing process stopped after 1,000 iterations without changes in the list of used/discarded SNP. The simulation process involved 1,000 preliminary generations (Ne=100) and five more generations (Ne=200) contributing phenotypic data (h^2=0.5). Genomes had a unique 100 cM chromosome with 5,000 SNP and 500 quantitative trait loci with mutation probabilities of 10-3 and 10-5, respectively. Ten independent data sets were generated and analysed. On average, the full model with all SNP reached a MSE of 1.370±0.004, whereas this parameter reduced until 1.293±0.005 when dropping off between 35% and 47% of the SNP. These results involved a ~6% reduction of MSE when using appropriate SNP for G and suggested a very appealing way to improve the statistical performance of gBLUP models when analysing massive genomic data.

Adaptation of BLUPF90 package for genomic computations

Misztal, I.[1], Aguilar, I.[2], Tsuruta, S.[1] and Legarra, A.[3], [1]University of Georgia, 425 River Rd, Athens GA 30602, Georgia, [2]INIA, Las Brujas, 90200 Canelones, Uruguay, [3]INRA, BP 52627, 32326 Castanet-Tolosan, France; ignacy@uga.edu

The original BLUPF90 package contains programs for renumbering, BLUP, parameter estimation, accuracy approximation and for sample visualization. A renumbering program (RENUMF90) supports national data sets. BLUP programs are for equations in memory (BLUPF90) and iteration on data (BLUP90IOD). Parameter estimation is via REML (REMLF90 and AIREMLF90) and Bayesian programs (GIBBS*F90), which uses optimized algorithms able to support large number of traits (20+). Samples from GIBBS* programs can be analyzed by POSTGIBBSF90, and accuracies of predictions can be approximated by ACCF90. Specific programs are available for threshold-linear models. Nearly all programs have been updated to support the genomic information and several new programs were added. Program PreGSF90 analyzes the SNP information, provides basic quality control, creates a genomic relationship matrix using a large variety of options, and combines pedigree and genomic relationship matrices for a single-step methodology. Preparing matrices for 30k animals with 50k SNP takes about 1 h. PostGSF90 converts GEBV to SNP effects, displays Manhattan plots possibly using moving averages, and estimates weights of SNP effects. Program PredF90 predicts GEBV based only on estimates of SNP effects obtained from PostGSF90. Most of the programs are available online at nce.ads.uga.edu. The package can be used for genomic predictions (including national data sets), parameter estimation (including GBLUP and G-REML), and GWAS. Unequal variances for SNP effects similar to those in BayesX and subsequently 'Manhattan' plots can be computed by iterating on postGSF90, preGSF90, and one of BLUP programs. These operations do not require deregression and are fast. Classical GWAS can be carried out with BLUPF90 fitting one SNP at a time as fixed regression and an animal effect with a genomic relationship matrix. The updated package simplifies genomic analyses in breeding applications.

Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science

Book of abstracts No. 18 (2012)
Bratislava, Slovakia
27 - 31 August 2012