BLUPF90 suite of programs for animal breeding with focus on genomics

I. Aguilar

INIA, Uruguay

S. Tsuruta, Y. Masuda, D.A.L. Lourenco, I. Misztal

University of Georgia, USA

A. Legarra

INRA, France

BLUPF90 suite of programs

- Collection of software for mixed-model computations
- Focus on breeding and genetics applications
- Features
  - Solving of mixed model equations
  - Variance component estimation
- Supports
  - general multiple trait
  - multiple effect
  - different model design per trait
  - correlated random effects

Birth

Developed to support a Fortran 90/95 programming course:

"Computational techniques in animal breeding"

University of Georgia, 1999 by I. Misztal

- BLUP program
  - solutions of mixed model equations
- Coded in Fortran 90/95 (F90)

`BLUPF90`

Computational techniques in animal breeding

by
Ignacy Misztal
University of Georgia
Athens, USA

Original BLUP program evolved

- estimation of variance components (REML, Gibbs sampler)
- support for threshold models
- large scale genetic evaluations
  - computations of solutions
  - approximations of accuracy

- Instead of one big program, several programs are available

BLUPF90 family programs

BLUPF90 and related programs (BGF90)

I. Misztal, S. Tsuruta, T. Strabel, B. Auvray, T. Druet & D.H. Lee. 2002

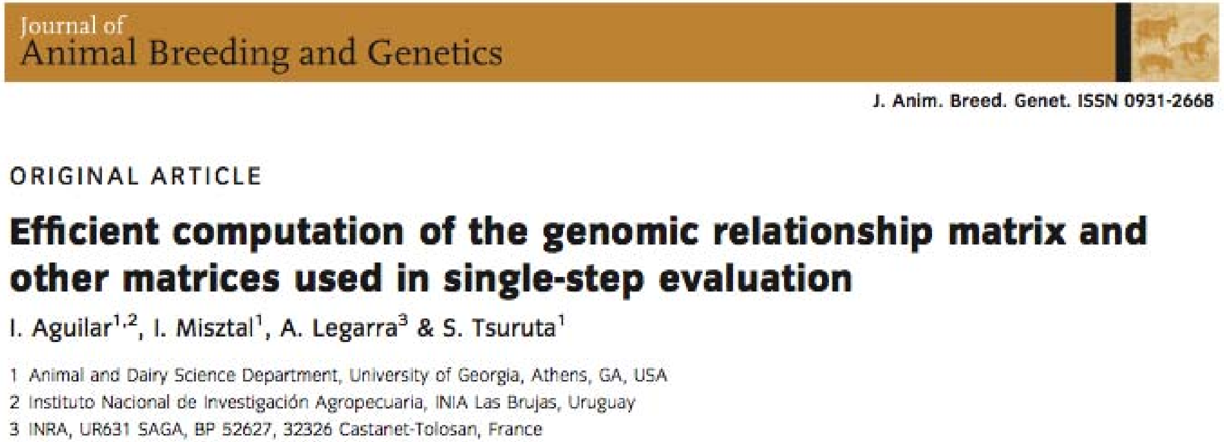
7th World Congress on Genetics Applied to Livestock Production, Montpellier, France

Scholar Google citations: 632

Genomic Era: Single-Step genomic evaluation

- Unified approach with pedigree, phenotypic and genomic information
- Pedigree-based relationships augmented by genomic relationship matrix (Misztal et al. 2009)

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \alpha H^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$



Journal of
Animal Breeding and Genetics

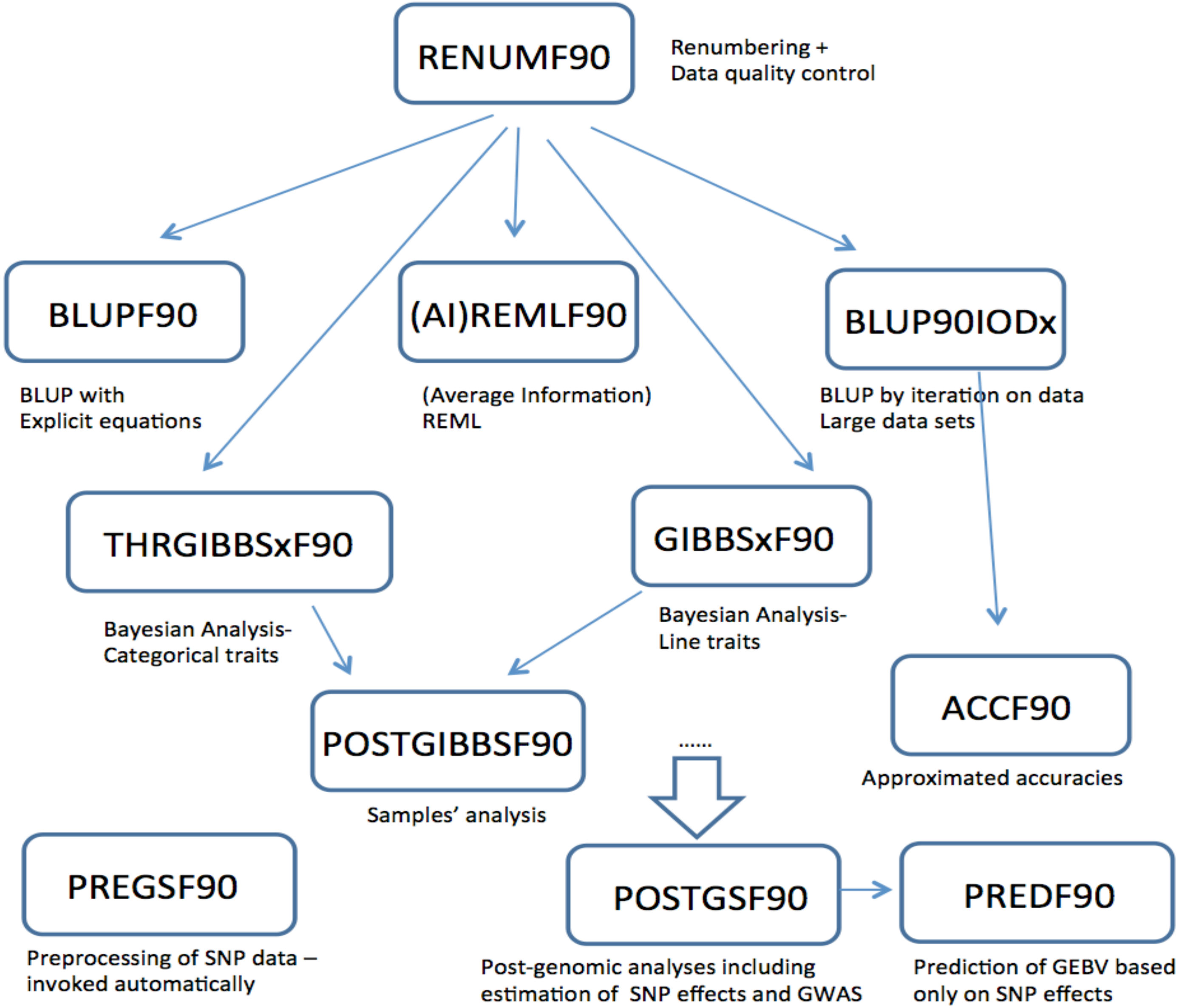
J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation

I. Aguilar[1,2], I. Misztal[1], A. Legarra[3] & S. Tsuruta[1]

1 Animal and Dairy Science Department, University of Georgia, Athens, GA, USA
2 Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, Uruguay
3 INRA, UR631 SAGA, BP 52627, 32326 Castanet-Tolosan, France



RENUMF90

Renumbering +
Data quality control

BLUPF90

BLUP with
Explicit equations

(AI)REMLF90

(Average Information)
REML

BLUP90IODx

BLUP by iteration on data
Large data sets

THRGIBBSxF90

Bayesian Analysis-
Categorical traits

GIBBSxF90

Bayesian Analysis-
Line traits

ACCF90

Approximated accuracies

POSTGIBBSF90

Samples' analysis

……

PREGSF90

Preprocessing of SNP data –
invoked automatically

POSTGSF90

Post-genomic analyses including
estimation of SNP effects and GWAS

PREDF90

Prediction of GEBV based
only on SNP effects

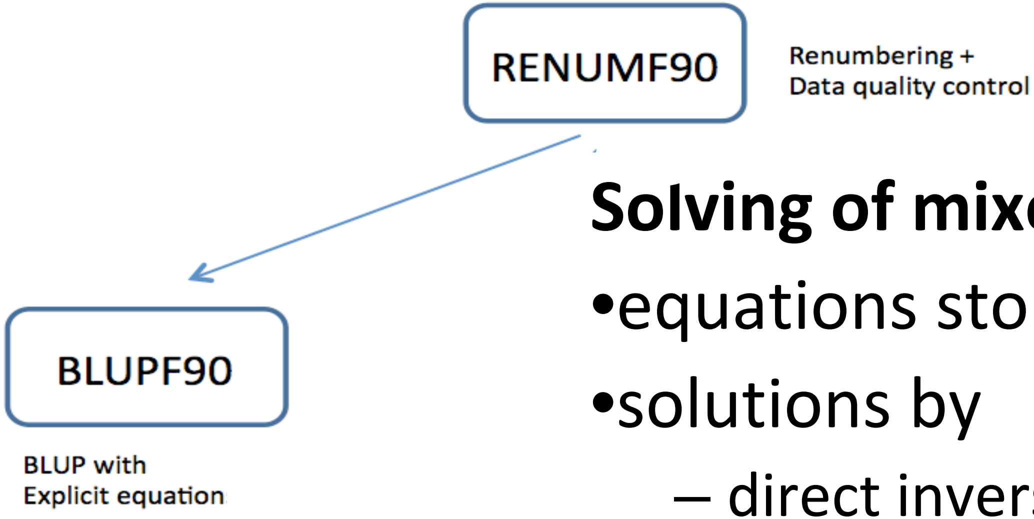
RENUMF90

Renumbering +
Data quality control

Data Preparation

- Reads alphanumeric information
  - Phenotypes, Pedigree, Markers
- Prepares
  - renumbered files
  - parameter file

  to be used in all programs
- Features
  - Trackback and extract pedigrees by generations
  - Inclusion of inbreeding in A^{-1}
  - Unknown parent groups
  - Merging effects (e.g. herd-year-season)

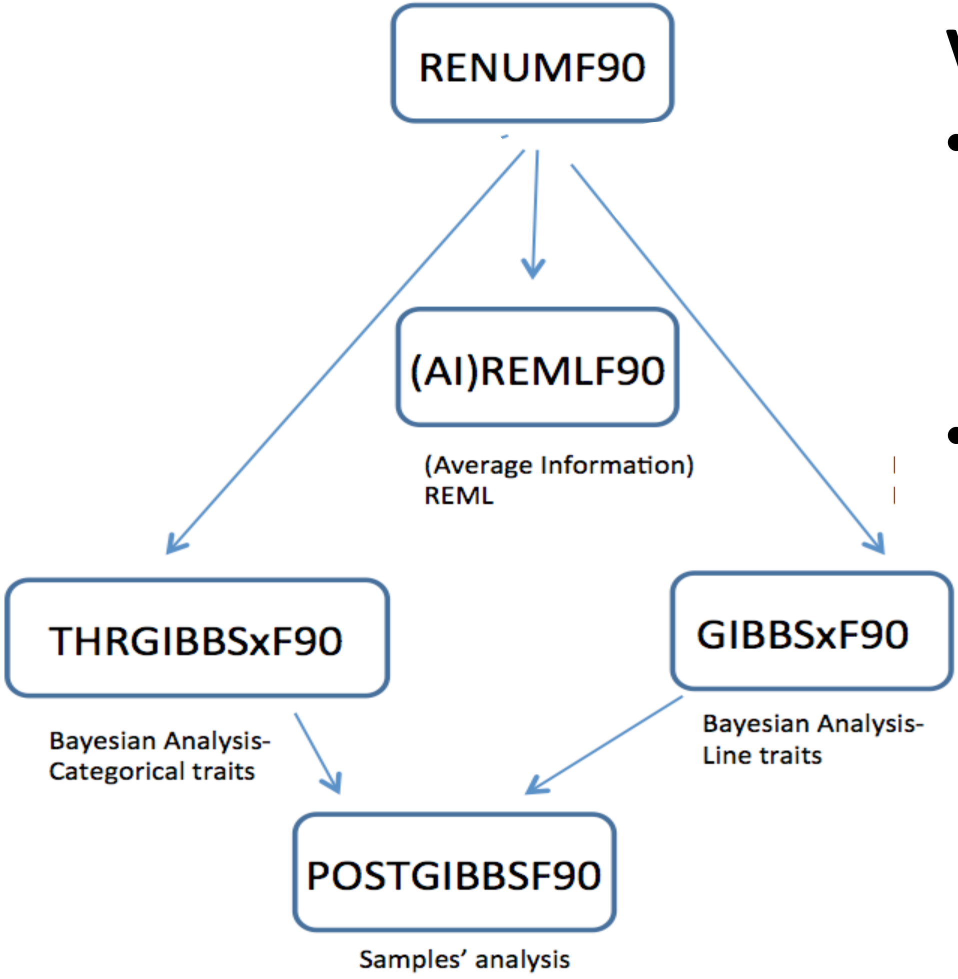


Solving of mixed model equations

- equations stored in memory
- solutions by
  - direct inversion
  - iterative methods (PCG)
- prediction error (co)variances from the sparse inverse to derive accuracy
  - single/multiple trait animal models
  - random regression models
- supports heterogeneous residual variances

Random effects supported

- pedigree additive relationships (animal or sire model)
  - with or without unknown parent groups
  - with or without inbreeding
  - combined with genomic information ssGBLUP
- parental dominance
- metafounders
- relationships derived for honey bee production
- user defined matrices


RENUMF90

(AI)REMLF90

(Average Information) REML

THRGIBBSxF90

Bayesian Analysis- Categorical traits

GIBBSxF90

Bayesian Analysis- Line traits

POSTGIBBSF90

Samples' analysis

Variance components programs

- REML
  - EM-REML – **REMLF90**
  - Average Information REML – **AIREMLF90**
    - standard errors of any function of variance components
- Gibbs sampling
  - highly optimized for storage of mixed model equations and block sampling:
  - for multiple traits **GIBBS1F90**
  - multiple traits and random correlated effects **GIBBS2F90**
  - and heterogeneous residual variances **GIBBS3F90**
  - and multiple trait threshold-linear traits **THRGIBBS1F90**
  - with heterogeneous residual variance **THRGIBBS3F90**

Sparse-dense matrix efficient methods

- key feature of the BLUF90 programs
  sparse matrix module with **efficient programming of sparse matrix computations**
  FSPAKF90 (Misztal & Perez-Enciso 1998) interface to FSPAK (Perez-Enciso *et al.,* 1994)

- With single-step GBLUP
  - large blocks of dense matrices deteriorate the performance of the FSPAK subroutines

- Module YAMS (Masuda et al 2014)
  - detects dense blocks in MME
  - rearrange computations using dense operations with optimized and parallelized subroutines.
- Drastically reduction of computing time for variance component (REML), and exact accuracies by inversion.

Genomic information

- Single-step GBLUP included in **all programs**
- **PREGSF90**
  - interface to process the genomic information
  - set of quality control on genotypes
  - provides several outputs to detect possible errors with genotypes
  - creation and inversion of matrices (e.g. G, G^{-1},A_{22}, A_{22}^{-1})
- **POSTGSF90**
  - solutions from ssGBLUP used to backsolve estimates for SNP effects
  - can be used to predict interim DGVs for newly genotyped individuals using **PREDF90**
- **SEEKPARENTF90**
  - Validation and discovery of paternity using genomic information

Large scale genetic evaluations programs

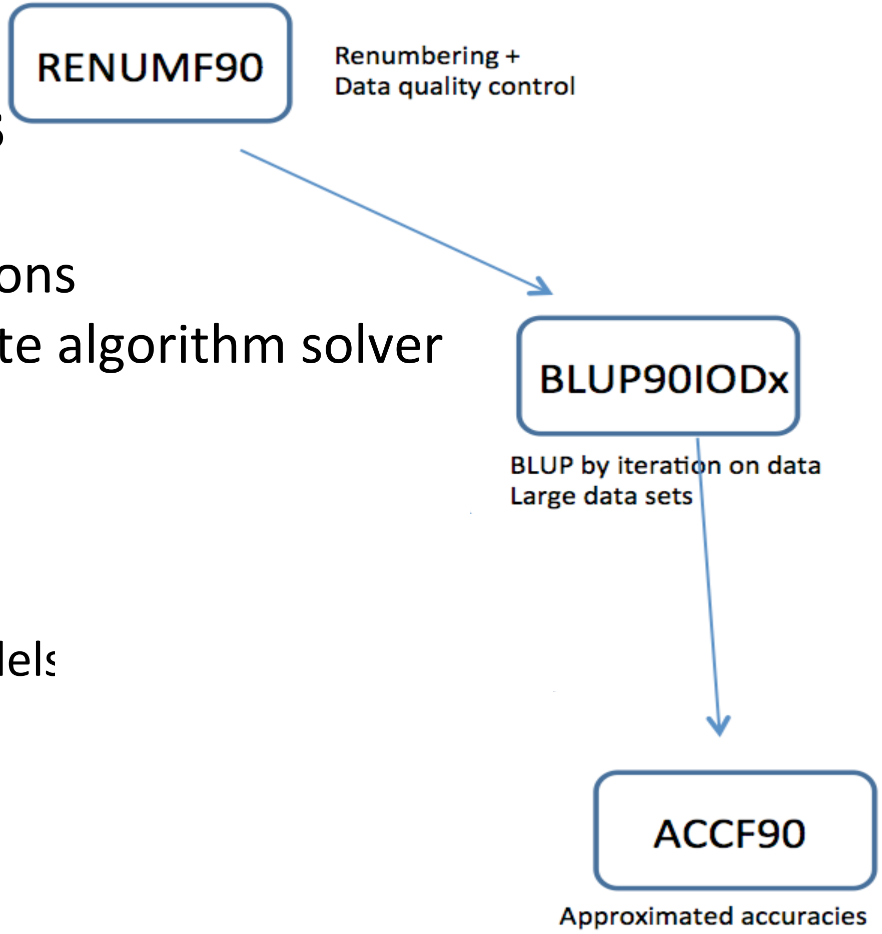
- **BLUP90IOD**

  compute solutions for large scale genetic evaluations

  iteration on data with the preconditioner conjugate algorithm solver

- Modified versions provide support
  - heterogeneous residual variance
  - multiple breed evaluation
  - optimized preconditioners for random regression models
  - threshold-linear models.

- Single-step genomic evaluations
  - full storage of $\mathbf{G}^{-1}$ - $\mathbf{A}_{22}^{-1}$
  - APY method with sparse inverse of $\mathbf{G}^{-1}$
  - efficient sparse implementation of $\mathbf{A}_{22}^{-1}$
  - unknown parent groups in ssGBLUP



RENUMF90

Renumbering +
Data quality control

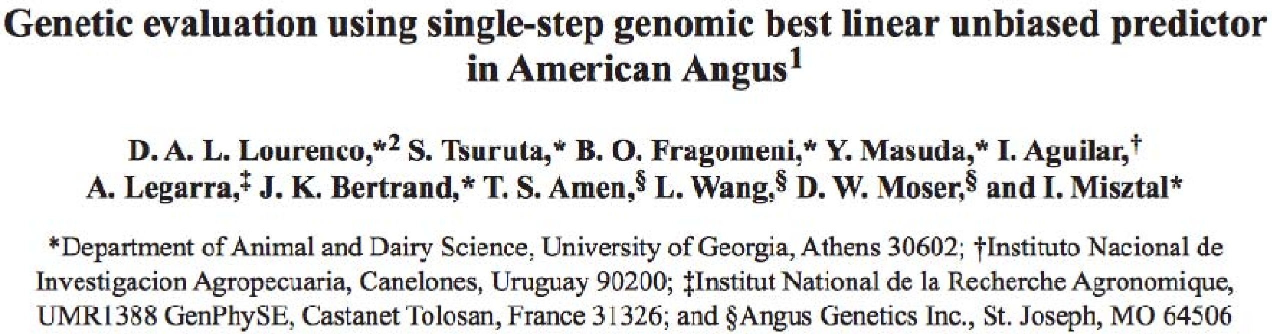
BLUP90IODx

BLUP by iteration on data
Large data sets

ACCF90

Approximated accuracies

Single-step GBLUP in large scale evaluations

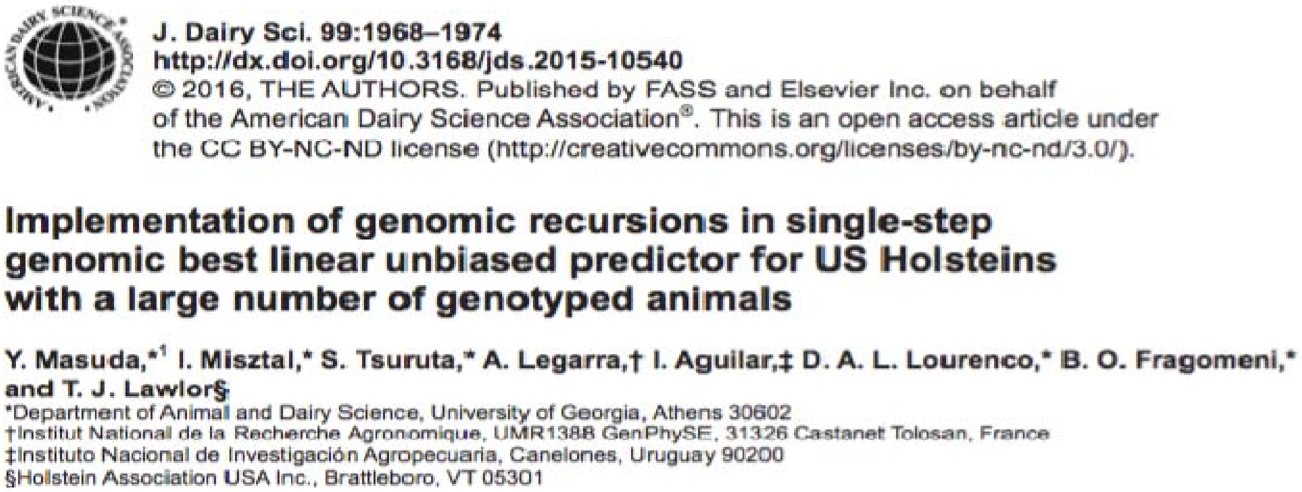


Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus[1]

D. A. L. Lourenco,*[2] S. Tsuruta,* B. O. Fragomeni,* Y. Masuda,* I. Aguilar,† A. Legarra,‡ J. K. Bertrand,* T. S. Amen,§ L. Wang,§ D. W. Moser,§ and I. Misztal*

*Department of Animal and Dairy Science, University of Georgia, Athens 30602; †Instituto Nacional de Investigacion Agropecuaria, Canelones, Uruguay 90200; ‡Institut National de la Recherche Agronomique, UMR1388 GenPhySE, Castanet Tolosan, France 31326; and §Angus Genetics Inc., St. Joseph, MO 64506

- **Performance data**
  - BW & WW: 6 M records
  - PWG: 3.4 M records
- **Pedigree data:**
  - 8.2 M animals
- **Genotype data**
  - 81 K animals



J. Dairy Sci. 99:1968–1974
http://dx.doi.org/10.3168/jds.2015-10540

Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals

Y. Masuda,*[1] I. Misztal,* S. Tsuruta,* A. Legarra,† I. Aguilar,‡ D. A. L. Lourenco,* B. O. Fragomeni,* and T. J. Lawlor§
*Department of Animal and Dairy Science, University of Georgia, Athens 30602
†Institut National de la Recherche Agronomique, UMR1388 GenPhySE, 31326 Castanet Tolosan, France
‡Instituto Nacional de Investigación Agropecuaria, Canelones, Uruguay 90200
§Holstein Association USA Inc., Brattleboro, VT 05301

- **Performance data**
  - Final score: 11 M records
- **Pedigree data:**
  - 10 M animals
- **Genotype data**
  - 570 K animals

ANIMAL BREEDING AND GENETICS GROUP

UNIVERSITY OF GEORGIA

Software

Download

You can find the latest release of BLUPF90 family of programs at the following page.

- Repository for BLUPF90 programs
  (64 bit Linux versions updated on October 10, 2017)

Documentation & Support

The BLUPF90 manual and the other documentation on the programs are available at the *dokuwiki*. For help, questions and bug reports, please visit the Yahoo Group.

- Dokuwiki for BLUPF90 Family of Programs for documentation
- Blupf90 discussion group at yahoo.com for support

http://nce.ads.uga.edu/software/

Summary

- BLUPF90 suite is a collection of software for mixed-model analysis with focus on breeding and genetics applications

- Great flexibility of models including multiple-trait, random correlated effects and linear-threshold models

- Genomic analyses using single-step GBLUP are fully integrated in all programs with efficient optimizations for large scale genetic evaluations.

Summer Course 2018

Programming and computer algorithms in animal breeding with a focus on single-step GBLUP and reality of genomic selection

May 7th to May 25th, 2018

University of Georgia, Athens, GA