

Genetics and population analysis

bWGR: Bayesian whole-genome regression

Alencar Xavier^{1,2,*}, William M. Muir² and Katy M. Rainey ^{2,*}

¹Corteva Agriscience, Johnston, IA 50131, USA and ²Purdue University, West Lafayette, IN 47907, USA

*To whom correspondence should be addressed.

Associate Editor: Russell Schwartz

Received on June 13, 2019; revised on October 14, 2019; editorial decision on October 15, 2019; accepted on October 23, 2019

Abstract

Motivation: Whole-genome regressions methods represent a key framework for genome-wide prediction, cross-validation studies and association analysis. The bWGR offers a compendium of Bayesian methods with various priors available, allowing users to predict complex traits with different genetic architectures.

Results: Here we introduce bWGR, an R package that enables users to efficient fit and cross-validate Bayesian and likelihood whole-genome regression methods. It implements a series of methods referred to as the Bayesian alphabet under the traditional Gibbs sampling and optimized expectation-maximization. The package also enables fitting efficient multivariate models and complex hierarchical models. The package is user-friendly and computational efficient.

Availability and implementation: bWGR is an R package available in the CRAN repository. It can be installed in R by typing: `install.packages('bWGR')`.

Contact: alencar.xavier@corteva.com or krainey@purdue.edu

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Genome-wide markers have been increasingly deployed for the prediction of complex traits since the concept of genomic prediction was introduced (Meuwissen *et al.*, 2001). Whole-genome regression (WGR) methods predict traits as a linear combination of marker effects that capture quantitative trait loci and the relationship among individuals (Habier *et al.*, 2007). A large variety of models exist, each with different prior assumptions that are optimized for a specific genetic architecture (de los Campos *et al.*, 2013). Evaluating the parameterizations of genomic information in prediction models to suit different genetic architectures can enhance prediction accuracy.

Few statistical packages enable genome-wide prediction, including rrBLUP, BGLR and VIGoR (Endelman, 2011; Onogi and Iwata, 2016; Pérez and de los Campos, 2014). Genome-wide models are sensitive to the algorithm implementation, such that two implementations of the same model often lead to reasonably different results (Gianola *et al.*, 2009; Lehermeier *et al.*, 2013). In a user-friendly framework, the bWGR package implements a compendium of likelihood and Bayesian methods, via expectation-maximization (EM) and Markov Chain Monte Carlo (MCMC), at univariate and multivariate level (see [Supplementary Material](#) for details). It also implements a mixed model solver that enables modeling replicated observations, computing marker effects using link functions and accounting for nuisance parameters (Fig. 1).

2 MCMC methods

MCMC methods constitute the most popular set of WGR (Gianola, 2013). These include Bayesian Ridge Regression, BayesA, BayesB

(Meuwissen *et al.*, 2001), BayesC, BayesCpi, BayesDpi (Habier *et al.*, 2011), Bayesian LASSO (Park and Casella, 2008) and Reproducing Kernel Hilbert Spaces (RKHS) regression (de los Campos *et al.*, 2010). The variable selection of BayesB and BayesC was implemented through Gibbs Sampling unconditional prior (Kuo and Mallick, 1998) and Metropolis-Hasting for BayesCpi and BayesDpi. In our models, the prior specifications are similar but not identical to the BGLR package (Pérez and de los Campos, 2014). We kept the models less hierarchical like those originally proposed by Meuwissen *et al.* (2001), with restricted Bayesian learning (Lehermeier *et al.*, 2013) to avoid under- and over-regularization. These methods can be performed either from bWGR's generalized function 'wgr' or by their standalone implementation written entirely in C++ (Eddelbuettel *et al.*, 2011). The generalized function 'wgr' enable users to combine a WGR with a kernel method, such as combining BayesB and RKHS. It also has an exclusive feature as it enables the subsampling of Markov chains to save time and computational power (Xavier *et al.*, 2017).

3 EM methods

EM methods provide an elegant and efficient way to reduce the computation time due to MCMC iterations (Shepherd *et al.*, 2010). Iterative procedures may replace Gibbs sampling by updating parameters with the expectation as opposed to sampling and averaging the posteriors. This algorithmic variation of the traditional MCMC solver of the Bayesian methods was proposed by Meuwissen *et al.* (2009). These EM Bayesian methods can calibrate WGR without loss in accuracy (Lopez *et al.*, 2019). From the Bayesian alphabet implemented via EM, the package provides

- Endelman, J.B. (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*, **4**, 250–255.
- Gianola, D. (2013) Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics*, **194**, 573–596.
- Gianola, D. *et al.* (2009) Additive genetic variability and the Bayesian alphabet. *Genetics*, **183**, 347–363.
- Gilmour, A.R. *et al.* (1995) Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, **51**, 1440–1450.
- Habier, D. *et al.* (2011) Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*, **12**, 186.
- Habier, D. *et al.* (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, **177**, 2389–2397.
- Hadfield, J.D. (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J. Stat. Softw.*, **33**, 1–22.
- Kuo, L. and Mallick, B. (1998) Variable selection for regression models. *Sankhyā Indian J. Stat. B*, **60**, 65–81.
- Lado, B. *et al.* (2013) Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. *G3*, **3**, 2105–2114.
- Legarra, A. and Misztal, I. (2008) Computing strategies in genome-wide selection. *J. Dairy Sci.*, **91**, 360–366.
- Lehermeier, C. *et al.* (2013) Sensitivity to prior specification in Bayesian genome-based prediction models. *Stat. Appl. Genet. Mol. Biol.*, **12**, 375–391.
- Lopez, M.A. *et al.* (2019) Phenotypic variation and genetic architecture for photosynthesis and water use efficiency in soybean (*Glycine max* L. Merr). *Front. Plant Sci.*, **10**.
- Meuwissen, T.H. *et al.* (2009) A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. *Genet. Sel. Evol.*, **41**, 2.
- Meuwissen, T.H.E. *et al.* (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, **157**, 1819–1829.
- Montesinos-López, O.A. *et al.* (2019) An R package for Bayesian analysis of multi-environment and multi-trait multi-environment data for genome-based prediction. *G3*, **9**, 1355–1369.
- Muir, W.M. (2005) Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics*, **170**, 1247–1259.
- Onogi, A. and Iwata, H. (2016) VIGoR: variational Bayesian inference for genome-wide regression. *J. Open Res. Softw.*, **4**, e11.
- Park, T. and Casella, G. (2008) The Bayesian lasso. *J. Am. Stat. Assoc.*, **103**, 681–686.
- Pérez, P. and de los Campos, G. (2014) Genome-wide regression & prediction with the BGLR statistical package. *Genetics*, **198**, 483–495.
- Schaeffer, L.R. (1986) Pseudo expectation approach to variance component estimation. *J. Dairy Sci.*, **69**, 2884–2889.
- Shepherd, R.K. *et al.* (2010) Genomic selection and complex trait prediction using a fast EM algorithm applied to genome-wide markers. *BMC Bioinformatics*, **11**, 529.
- VanRaden, P.M. (2008) Efficient methods to compute genomic predictions. *J. Dairy Sci.*, **91**, 4414–4423.
- Xavier, A. (2019) Efficient Estimation of Marker Effects in Plant Breeding. *G3*, **9**, 1–12.
- Xavier, A. *et al.* (2017) Genomic prediction using subsampling. *BMC Bioinformatics*, **18**, 191.