lme4breeding: enabling genetic evaluation in the era of genomic data.

Giovanny Covarrubias-Pazaran[1]
[1] International Rice Research Institute, Pili Drive, Los Baños, Laguna 4031, Philippines

Abstract

Mixed models are a cornerstone in quantitative genetics to study the genetics of complex traits. A standard quantitative genetic model assumes that the effects of some random effects (e.g., individuals) are correlated based on their identity by descent and state. In addition, other relationships arise in the genotype by environment interactions (i.e., covariance structures). Open-source mixed model routines are available but do not account for complex covariance structures and are able to fit big genomic models. The lme4breeding R package was developed as an extension of the lme4 package and allows correlated random effects and complex covariance structures to be fitted for Gaussian, binary, and count responses. The correlation between levels of the random effect (e.g., individuals) is accounted for by post-multiplying the incidence matrix of the levels of this random factor by the Cholesky factor of the corresponding (co)variance matrix (e.g., the genomic relationship matrix). To enable big genomic models the eigen decomposition of relationship matrices is enabled. Maximum likelihood and REML estimation are available in lme4breeding. This note describes the type of models that can be fitted using lme4breeding and presents some examples.

Key words: quantitative genetics, R package, covariance structure, generalized linear mixed model.

Introduction

Mixed models are the cornerstone to estimate genetic parameters and predict surrogates of merit (i.e., breeding values, general combining ability, or total genetic value) associated with Gaussian traits (Henderson et al., 1959; Henderson, 1963, 1973). In linear mixed models with Gaussian responses, the marginal likelihood has a closed form, and maximum likelihood or REML estimation can be performed, whereas in nonlinear models the likelihood does not have a closed form and Laplacian approximation and others are needed. A more general class of mixed models is represented by the generalized mixed model, which is appropriate for analysis of data from the exponential family of distributions (Tempelman, 1998).

The development of mixed model packages in the R environment (R Development Core Team, 2008) has increased in the last decade due to a quick learning curve and easiness to code complex algorithms and sophisticated graphics. R is an open-source system written by volunteers and is extended via packages (www.r-project.org). The lme4 package (Bates and Maechler, 2008) is the gold-standard when it comes to fitting linear models and GLMM to data. The program handles any number of random effects, nested or cross-classified, and uses a combination of sparse and dense matrix representations to process large data sets at high speed. The use of lme4 for genetic analysis was limited because it does not allow covariances between levels of a random effect. To close that gap Vazquez et al. (2010) developed a package called pedigreemm to expand lme4 capabilities to allow correlations between levels of random effects, specifically pedigrees and made it available on CRAN. With the advent of genomic information, Ziyatdinov et al. (2018) expanded the lme4 to include genomic relationship matrices inspired by pedigreemm with a package called lme4qtl but was neither published on CRAN nor properly documented. Camel et al. (2021) published a package to do the same as

lme4qtl called lme4GS but was again neither published on CRAN nor properly documented. These two packages are slow to fit big genomic relationship matrices due to the dense equations provided to the lme4 machinery.

In the present note, we discuss the approach used by lme4breeding for GLMM to include covariances between levels of random effects (any Cholesky factor), coupled with the eigen (UDU') decomposition of the relationship matrix to speed up operations. In addition, customized random effects (e.g., overlay models for half diallel or indirect genetic effects) can also be provided, and the ability to model complex covariance structures like the ones found in the genotype by environment problem (e.g., factor analytic structures). We show the use of the package with two examples using data included in the package. The package is available on CRAN (https://CRAN.R-project.org/package=lme4breeding).

Methods

The general formulation of the Generalized Linear Mixed Model (GLMM) is as follows:

$$g(\mu_{y|U}) = \mathbf{Z}\mathrm{u} + \mathbf{X}\beta,$$

where y is the random variable representing the response; g(.) is a function that links the response with a model that is linear in the explanatory variables; $\mu_{y|U} = E[y|U = u]$ is the expectation of the response conditional to the random effects; β is the vector of fixed effects; **X** is the incidence matrix relating fixed effects to $g(\mu_{y|U})$; u is the vector of random effects, which may row-bind multiple factors, where each factor of random effect is distributed as $u \sim N(0, \mathbf{I}\sigma^2_u)$, where N(.,.) represents the normal distribution with mean and variance indicated in the parentheses; σ^2_u is a variance-covariance matrix; **I** is an identity matrix; and **Z** is the incidence matrix relating u to $g(\mu_{y|U})$. Some possible families for the conditional distribution of the data, given the random effects in GLMM, are the binomial, Gaussian, gamma, inverse Gaussian, and Poisson distributions, among others, which are derived from the exponential family of distributions. Parameter estimates are obtained by minimizing the Laplacian approximation to the deviance function according to Bates (2009). In lme4breeding, canonical links are the default [other links can be checked by typing help(family) in the R console].

In the Gaussian linear model, the variance-covariance matrix of the marginal distribution of the records is:

$$\mathrm{V(y)} = \mathbf{Z}(\mathbf{A}\sigma^2_{\mathrm{u}})\mathbf{Z}' + \mathbf{R},$$

where V(y) is the variance of the response y; σ^2_u is the variance of a given random factor, $\mathbf{A}\sigma^2_u$ is the covariance matrix of the levels of the random effects (u); and **R** is the covariance matrix of the random residuals, often assumed to follow a normal distribution, independently and identically distributed. **Z** is again the incidence matrix mentioned before. N structures from random effects can be added in V(y). Different options for modeling **R** are available and are independently and identically distributed (i.e., $\mathbf{R} = \mathbf{I}\,\sigma^2_e$ is the default). In quantitative genetics, the variance-covariance matrix of individuals is modeled through additive and non-additive relationships based on the identical-by-descent and identical-by-state theory (Mrode, 2005).

The methodology described by Harville and Callanan (1989) consists of post-multiplying the model matrix **Z** by the Cholesky decomposition of the numerator relationship matrix (**A**). Note

that **A** is a positive-definite matrix (unless identical twins or clones are in the pedigree, in which case it would be positive semi-definite). Let **A** = **LL′**, where **L** is the Cholesky factor. The matrix **L** can be written directly from the pedigree information (Henderson, 1976). Subsequently, let **Z*** = **ZL**, then

$$\begin{aligned} V(\mathbf{Z}u) &= \mathbf{ZAZ}'\sigma_u^2 \\ &= \mathbf{ZLL}'\mathbf{Z}\sigma_u^2 \\ &= (\mathbf{Z}^*)(\mathbf{Z}^*)\sigma_u^2 \end{aligned}$$

Define u* = $\mathbf{L}^{-1}$u, and rewrite g($m_{y|u}$) as:

$$\mathbf{ZL}(\mathbf{L}^{-1}u) + \mathbf{X}\beta = \mathbf{Z}^*u^* + \mathbf{X}\beta$$

Assuming u ~ N(0, $\mathbf{A}\sigma^2_u$), the distribution of u* is N[0,$\mathbf{I}\sigma^2_u$] because

$$\begin{aligned} \mathbf{L}^{-1}\mathrm{Var}(u)(\mathbf{L}'^{-1}) &= \mathbf{L}^{-1}\mathbf{A}(\mathbf{L}'^{-1})\sigma_u^2 \\ &= \mathbf{I}\sigma_u^2 \end{aligned}$$

so that the elements of u* are mutually independent. The lme4 procedure can then be applied to g($\mu_{y|U}$) = **Z***u* + **X**β because the random effects are now independent.

To speed up the genomic models we applied the **HDH**' decomposition proposed by Lee and Van der Werf (2016) to be applied at the level of the mixed model equations (p x p problem), instead of the number of records as they did (n x n problem). Lee and Van der Werf proposed the parameterization:

$$\mathbf{H}y = \mathbf{HX}\beta + \mathbf{HZ}u + \mathbf{H}e$$

$$V(\mathbf{H}y) = \mathbf{ZDZ}' + \mathbf{R}$$

Where **H** and **D** are the eigen vectors and eigen values respectively of any relationship matrix (e.g., genomic relationship matrix). **H** can be derived as:

$$\mathbf{H} = \mathbf{Z}_u\mathbf{A}\,\mathbf{Z}_u' * \mathbf{Z}_r\mathbf{Z}_r'$$

Where $\mathbf{Z}_u$ is the incidence matrix for the individuals, $\mathbf{Z}_r$ is the incidence matrix for the records for the individuals, and * is the element-wise multiplication of two objects (Schur product).

Examples

All programs and data sets used are available with the package at the R CRAN archive (http://cran.r-project.org/).

Data description

Dataset 1. Information from a collection of 599 historical CIMMYT wheat lines was used. The wheat data set is from CIMMYT's Global Wheat Program. Historically, this program has conducted numerous international trials across a wide variety of wheat-producing environments. The environments represented in these trials were grouped into four basic target sets of

environments comprising four main agroclimatic regions previously defined and widely used by CIMMYT's Global Wheat Breeding Program. The phenotypic trait considered here was the average grain yield (GY) of the 599 wheat lines evaluated in each of these four mega-environments. Lines were genotyped using 1447 Diversity Array Technology (DArT) generated by Triticarte Pty. Ltd. (Canberra, Australia; http://www.triticarte.com.au). The DArT markers may take on two values, denoted by their presence or absence. Markers with a minor allele frequency lower than 0.05 were removed, and missing genotypes were imputed with samples from the marginal distribution of marker genotypes. The dataset and script can be accessed by typing "?DT_wheat" in the terminal.

Dataset 2. Information from a collection of 413 rice hybrids was used to show the advantages of a multi-trait fit. The data set belongs to the Rice Diversity Organization Program. The lines are genotyped with 36,901 SNP markers and phenotyped for more than 30 traits (Zhao et al., 2011). The dataset and script can be accessed by typing "?DT_rice" in the terminal.

Fitting the models

Example 1. The wheat yield lines (dataset 1) were analyzed with the classical multi-environment GBLUP model:

$$y_{ijk} = L\beta_i + u_j + e_{ijk}$$

Where y_{ijk} is the kth observation (yield performance); β_i is the effect of location i (i=1,….4), u_j is the breeding value of individual j (j=1, …, 599), and e_{ijk} is the random residual associated to k^{th} observation. The following distributions are assumed for the vector of breeding values and residual effects.

$$\begin{bmatrix} \mathrm{u} \\ \mathrm{e} \end{bmatrix} \sim \mathrm{N}\left[0\,,\ \begin{pmatrix} \mathbf{A}_{599} \otimes \mathbf{\Sigma} & 0 \\ 0 & \mathbf{I}_{2396} \otimes \mathbf{\Phi} \end{pmatrix}\right]$$

Where u is the vector of breeding values and e is the vector of residual effects. A represents the matrix of additive relationships between lines, whereas $\otimes$ denotes the Kronecker product of two matrices. The variance-covariance of breeding values Σ is:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2_{u.1} & \sigma_{u.1,u.2} & \sigma_{u.1,u.3} & \sigma_{u.1,u.4} \\ . & \sigma^2_{u.2} & \sigma_{u.2,u.3} & \sigma_{u.2,u.4} \\ . & . & \sigma^2_{u.3} & \sigma_{u.3,u.4} \\ . & . & . & \sigma^2_{u.4} \end{bmatrix}$$

Representing the genetic covariance between the four locations and Φ represents a one by one matrix of residual variance. This model was fitted in R as follows:

```
system.time(
 mix <- lmebreed(pheno ~ (0 + e1 + e2 + e3 + e4 | id),
                  relmat = list(id=K),
                  rotation = TRUE,
                  data=DT)
)
   user  system elapsed
  2.265   0.134   2.480
```

The “mix” object contains the fitted model and is of class *lmebreed*. The left-hand side of the formula denotes the response variable and the right-hand side the explanatory variable. Terms in parentheses represent the random effects [e.g. (intercept|slope)] expressed as random regressions (slopes) over numerical covariates (intercepts) as normally done in lme4. In this model the columns e1 to e4 are columns coded with a 1 if the record belongs to the ith environment and a 0 when it doesn’t. These random intercepts are written on the left side of the | symbol. The term “id” represents the random factor being regressed over the environments, in this case the wheat line evaluated. The parameter *relmat* is a named list containing the relationship matrices. The parameter *rotation* controls the activation of the eigen decomposition. The function `system.time()` is used to measure how long the procedure takes to run the analysis. As the result shows, it took 2.480s for this example, run on a Mac-OS system with an M1 chip and 8 GB of RAM memory.

Example 2. This illustration uses a multi-trait fitting model. The rice (dataset 2) has been modeled previously with univariate models for genome-wide association studies (Zhao et al., 2011). The multi-trait GBLUP model can be defined as:

$$y_{itk} = \beta_0 + u_j + e_{ik}$$

Where y_{ijk} is the kth observation of yield performance; β_0 is the effect of an overall intercept, u_t is the breeding value of hybrid j (j=1, …, 413), and e_{itk} is the random residual associated to k^{th} observation. Trait observations are arranged in the long format to suit the lme4breeding machinery. The following distributions are assumed for the vector of breeding values and residual effects.

$$\begin{bmatrix} \mathrm{u} \\ \mathrm{e} \end{bmatrix} \sim \mathrm{N}\left[0\,,\ \begin{pmatrix} \mathbf{A}_{413} \otimes \mathbf{\Sigma} & 0 \\ 0 & \mathbf{I}_{4130} \otimes \mathbf{\Phi} \end{pmatrix}\right]$$

Where u is the vector of breeding values and e is the vector of residual effects. A represents the matrix of additive relationships between the rice hybrids, whereas $\otimes$ denotes the Kronecker product of two matrices. The variance-covariance Σ for the multi-trait fit is:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2_{u.1} & \sigma_{u.1,u.2} & \cdots & \sigma_{u.1,u.10} \\ . & \sigma^2_{u.2} & \cdots & \sigma_{u.2,u.10} \\ . & . & \ddots & \vdots \\ . & . & . & \sigma^2_{u.10} \end{bmatrix}$$

Representing the genetic covariance between the ten traits and Φ represents a one-by-one matrix of residual variance. This model was fitted in R as follows:

```
system.time(
  mix <- lmebreed(valueS ~ (0+trait|geno),
                  relmat = list(geno=A),
                  rotation = TRUE,
                  data=DT)
)
   user  system elapsed
 26.217   0.703  27.082
```

In this model the column “trait” is coded as factor so *lme4breeding* creates internally ten indicator columns, one for each trait that associate observations to traits. The random

regression of genotypes proceeds over traits with an unstructured covariance. The term “geno” represents the random factor being regressed over the traits, in this case the rice hybrid evaluated. The parameter *relmat* is the additive relationship matrix between hybrids. The parameter *rotation* controls the activation of the eigen decomposition. The function `system.time()` was used to measure how long the procedure took. As the result shows, it took 27.082s for this example, run on a Mac-OS system with an M1 chip and 8 GB of RAM memory.

Model output

This section illustrates a few properties and uses of the *lmebreed* objects to obtain the information of interest for the scientist.

Variance components

The variance components associated to the random effects can be observed calling the `summary()` function but they can be retrieved more easily using the `varCorr()` function as follows:

```
vc <- VarCorr(mix); print(vc)
Groups   Name Variance Cov
 id       e1   0.43701
          e2   0.82435  -0.022
          e3   0.86499  -0.073  0.841
          e4   0.59358  -0.095  0.540  0.573
 Residual      0.42511
```

This function already returns the variance covariance matrices for each random effect in a ready-to-use format where each element of the list is the variance-covariance matrix for each random effect. The one above is the unstructured variance covariance matrix of the multi-environment model. The residual variance can be retrieved in the same way than lme4 since it is stored in the `varCorr()` object as the “sc” attribute:

```
attr(vc, "sc")^2
[1] 0.4251133
```

Random and fixed effects

Effects for the fixed and random factors are of special interest for geneticists to estimate merit of individuals, allelic effects, and test other evolutionary theories. These effects can be retrieved from the *lmebreed* objects using the `fixef()` and `ranef()` functions. The fixed effects can be retrieved for the dataset 1 as:

```
fixef(mix)
 (Intercept)
4.615603e-16
```

The random effects retrieved by the ranef() function are returned in their original scale (*u*). To calculate *u*, the transformed predicted effects inside the model [$u^* = (\mathbf{L}^{-1})u$] are premultiplied/transformed by **L**, the Cholesky factor of the relationship matrix which are stored in the *lmebreed* object. If the eigen decomposition (**HDH**’) is activated using the *rotation* argument,

an additional rotation/transformation is required [$u^{\#} = \mathbf{H}'(\mathbf{L}^{-1})u$] by premultiplying **L**u by **H** (u=**HL**u*), the eigen vectors of the relationship matrix. The function is used as follows:

```
head(ranef(mix)$id)
                e1         e2         e3         e4
103122  -0.7883656 -2.0593553 -2.0137375 -1.3759669
1047493  0.1890526  0.1899264  0.1898657  0.4486322
1058137 -0.7394242 -2.1852791 -2.1435186 -1.3686411
1064007  0.5100866  0.8277841  0.7895387  0.5677032
1066175  0.4328772 -0.2691902 -0.3239864 -0.2532966
1066279  0.2528579  1.2818465  1.2676190  0.6493584
```

Where rows are the levels of the random effect and columns are the estimate at each intercept level in the random regression (e.g., environment in the wheat example and traits in the rice example).

The standard errors of the random effects are of special interest for scientist to carry on comparisons and hypothesis testing. Just like with random effects, the recovery of accurate standard errors requires the to pre and post multiply the inverse of the coefficient matrix by the eigen vectors when the rotation is used but they don't require to be multiplied by the Cholesky factor **L**. To retrieve the standard errors the argument *condVar* in the `ranef()` function can be set to TRUE as follows in the example dataset 1:

```
attr( ranef(mix, condVar=TRUE)$id, which="postVar")[,,1]
          [,1]      [,2]      [,3]     [,4]
[1,] 0.8434209 0.6847328 0.8400128 1.111775
[2,] 0.6301186 1.0355293 0.5728308 1.022975
[3,] 1.1606441 0.5928033 0.8965903 1.111910
[4,] 0.5999681 0.8396017 0.8697833 1.167705
```

Where the attribute *postVar* retrieves the standard errors for each level of the random effect and each slice corresponds to one level for the covariance structure specified (unstructured between four environments).

Time comparisons against other popular software

In order to create an intuition of the time gains using lme4breeding compared to popular software, we run the example dataset 1 that includes 599 lines genotyped and tested in 4 environments modeled with a diagonal covariance matrix using the following libraries; lme4qtl (0.2.2), sommer (4.3.4), and Asreml (4.1.0.182) using the default parameters on a Mac-OS machine with an M1 chip and 8 GB of RAM memory. The results were the following:

Table 1. Variance component comparison between libraries and lme4breeding

Package/Var	σ^2_{e1}	σ^2_{e2}	σ^2_{e3}	σ^2_{e4}	σ^2_{r}
lme4breeding with rotation	0.2775	0.2566	0.2554	0.2480	0.5849
lme4breeding	0.2775	0.2566	0.2554	0.2480	0.5849
lme4qtl	0.2775	0.2566	0.2554	0.2480	0.5849
sommer	0.2772	0.2564	0.2551	0.2478	0.5850
Asreml	0.2775	0.2567	0.2554	0.2480	0.5850

As it can be seen in Table 1 the five variance-covariance components were almost identical among the libraries. This shows that the accuracy of lme4breeding to estimate variance components is very similar to other popular software.

Table 2. Time comparison (average across 10 samples) between different libraries for a genotype (599) by environment (4) problem fitted with a diagonal variance-covariance structure in wheat dataset.

Package	User time	System time	Elapsed time	X times*
lme4breeding (with rotation)	1.491	0.130	1.696	1 x
lme4breeding (without rotation)	165.830	1.031	166.879	98 x
lme4qtl#	150.431	0.894	151.325	89 x
Sommer-mmec	267.677	3.408	272.363	160 x
Asreml	7.535	0.094	7.786	4 x

* With respect to lme4breeding using the eigen decomposition (using the rotation argument).
# The lme4 symbols | and || to specify the formulation for random effects is not respected as declared. We had to trick the machinery.

In addition, the comparison of times with the different software shows the great advantage of lme4breeding over popular software, ranging from 4x to 160x for a diagonal model. The gain is even higher for a fully unstructured model (using the same dataset) where the gain of lme4breeding with rotation against Asreml was 1:1350, 1:1261 against lme4qtl when the problem becomes denser. This increase in speed is thanks to the coupling of the eigen decomposition proposed by Lee and Van der Werf (2016) with the Cholesky decomposition of Harville and Callanan (1989) to create a totally sparse formulation of the genetic evaluation problem and using the lme4 machinery that was design for sparse problems (Bates, 2009). These two methods together achieve great speeds for genetic evaluation difficult to match with more classical approaches. The only drawback is that only one relationship matrix can be eigen-decomposed, and the rest of the random effects can only have a diagonal relationship matrix (Lee and Van der Werf, 2016). Fortunately, most of the genetic evaluation problems only have one dense relationship matrix fitted.

Concluding remarks

The strategy used on *lme4breeding* uses the eigen decomposition of the relationship matrix coupled with the Cholesky decomposition of the resulting diagonal (co)variance structure of the random effects to speed up the REML computation, especially in multi-trait scenarios. This approach expands the powerful lme4 capabilities to fit random regressions and sparse problems to solve many of the plant and animal breeding problems of genetic evaluation. As shown, thousands of genotyped individuals tested in dozens of environments can be fitted in seconds due to the sparse parameterization of the problem and surpasses the capabilities of similar software. The tool is available in R (https://CRAN.R-project.org/package=lme4breeding), which is an open-source and powerful statistical software. Additionally, a few functions for computing relationship matrices and geospatial kernels are included.

Acknowledgments

Special thanks to Douglas Bates, Ben Bolker and the lme4 development team for making available the lme4 engine for these types of extensions and answering requests and questions actively. I also want to thank Johan Aparicio-Arce and Chris Gaynor for letting me bounce some ideas with them to improve the software.

Literature review

1. Bates, D. M. 2009. Index of http://lme4.r-forge.r-project.org/ slides/2009-07-21-Seewiesen/ Section: 4Precision. Accessed Aug. 5, 2009.
2. Bates, D., and M. Maechler. 2008. The Comprehensive R Archive Network. http://cran.r-project.org/ Accessed Jan. 2009.
3. Caamal-Pat, D., Pérez-Rodríguez, P., Crossa, J., Velasco-Cruz, C., Pérez-Elizalde, S., & Vázquez-Peña, M. (2021). lme4GS: an R-package for genomic selection. Frontiers in Genetics, 12, 680569.
4. Covarrubias-Pazaran, G. (2016). Genome-assisted prediction of quantitative traits using the R package sommer. *PloS one*, *11*(6), e0156744.
5. Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S., & Thompson, R. (2015). ASReml user guide release 4.1 structural specification. *Hemel hempstead: VSN international ltd*.
6. Harville, D. A., and T. P. Callanan. 1989. Computational aspects of likelihood-based inference for variance components. Pages 136–176 in Advances in Statistical Methods for Genetic Improvement of Livestock. D. Gianola and K. Hammond, ed. Springer-Verlag, Berlin, Germany.
7. Henderson, C. R. 1963. Selection index and expected genetic advance. Pages 141–163 in Statistical Genetics and Plant Breeding. W. D. Handson and H. F. Robinson, ed. Natl. Acad. Sci. and Natl. Res. Counc., Washington, DC.
8. Henderson, C. R. 1973. Sire evaluation and genetic trends. Pages 10–41 in Proc. Anim. Breed. Genet. Symp., Blacksburg, VA. Am. Soc. Anim. Sci., Champaign, IL.
9. Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69–83.
10. Henderson, C. R., O. Kempthorne, S. R. Searle, and C. N. VonK- rosig. 1959. Estimation of environmental and genetic trends from records subject to culling. Biometrics 15:192–218.
11. Lee, S. H., & Van der Werf, J. H. (2016). MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. *Bioinformatics*, *32*(9), 1420-1422.
12. Mrode, R. 2005. Linear Models for the Prediction of Animal Breeding Values. 2nd ed. CAB Int., New York, NY.
13. R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
14. Tempelman, R. J. 1998. Generalized linear mixed models in dairy cattle breeding. J. Dairy Sci. 81:1428–1444.
15. Vazquez, A. I. 2007. Analysis of number of episodes of clinical mas- titis in Norwegian Red and Holstein cows with Poisson and categorical data mixed models. MS Thesis. Univ. Wisconsin, Madison.
16. Vazquez, A. I., D. Gianola, D. Bates, K. A. Weigel, and B. Heringstad. 2009. Assessment of Poisson, logit and linear models for genetic analysis of clinical mastitis in Norwegian Red Cows. J. Dairy Sci. 92:739–748.
17. Vazquez, A. I., Bates, D. M., Rosa, G. J. M., Gianola, D., & Weigel, K. A. (2010). an R package for fitting generalized linear mixed models in animal breeding. Journal of animal science, 88(2), 497-504.

18. Zhao, K., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., ... & McCouch, S. R. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. *Nature communications*, *2*(1), 467.
19. Ziyatdinov, A., Vázquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H., & Soria, J. M. (2018). lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. BMC bioinformatics, 19, 1-5.