



Impact of different priors of Bayesian LASSO regularization parameter on genomic prediction of traits with different genetic architectures

Sayed Mahdi Hosseini-Vardanjani 

Corresponding author, Department of Animal sciences, Animal breeding and genetic branch, Faculty of Agriculture, Lorestan University, Khorramabad, Iran. Email: hosseini.sm@lu.ac.ir

| Article Info | ABSTRACT |
|---|--|
| Article type: Research Article | <p>Bayesian LASSO is a penalized regression method with the potential of variable selection using the regularization parameter λ. In a fully Bayesian framework, it is possible to treat λ as random parameter and different priors can be assigned to it. The effect of assigning three different priors including gamma, beta, and fixed distributions on the predictability and estimation of genetic parameters was investigated using mice genome and seven traits with different genetic architecture. The results showed that estimation of variance components and heritability are strongly affected by the assigned prior. Highest genetic variance and heritability and the lowest residual variance were estimated by fixed λ model for different traits. The average accuracy obtained for each trait using different priors was relatively similar, but standard deviation of the accuracies by the fixed λ model was always greater than the other two priors. Comparison of the minimum and maximum values of the models accuracies via fivefold cross validation revealed that the highest accuracy values were obtained using the assignment of gamma and beta distributions to the regularization parameter. By increasing the heritability and the number of background QTL of the traits, the accuracy of the models increased and the difference between the three different priors decreased. The difference between the models was greater regarding to prediction bias, but it generally followed the same trend of the accuracy, i.e. the models with the highest accuracy also had the lowest prediction bias. Based on the present results, proper prior assignment to λ is more necessary for traits under small number of QTL and low heritability.</p> |
| Article history: Received: 11 June 2025 Received in revised form: 13 September 2025 Accepted: 21 September 2025 Published online: Summer 2026 | |
| Keywords: <i>Bayesian approach, Genomic prediction, Predictability, Proper prior, Variable selection.</i> | |

Cite this article: Hosseini-Vardanjani, S. M. (2026). Impact of different priors of Bayesian LASSO regularization parameter on genomic prediction of traits with different genetic architectures. *Iranian Journal of Animal Science*, 57 (2), 195-211. DOI: <https://doi.org/10.22059/ijas.2025.397606.654084>



© The Author(s).

DOI: <https://doi.org/10.22059/ijas.2025.397606.654084>

Publisher: The University of Tehran Press.

Extended Abstract

Introduction

When large amount of genomic markers became available in the last decade, livestock breeding programs have evolved rapidly through genomic selection (GS). GS refers to selection of individuals based on the aggregate genetic merit of mentioned markers. Briefly, GS involves application of a genomic statistical method in a reference population which has both phenotypic and genotypic data, and predicting genomic breeding value (GBV) in a candidate population which has only genotypic data. In context of GS in livestock, several statistical methods have been proposed to predict marker effects which are consequently used to estimate GBV of individuals. However, as the number of markers is very large relative to the phenotypes, genomic selection is faced to the challenge of the “curse of dimensionality”. Bayesian methods are usually proposed to handle these large sets. Bayesian LASSO is a penalized regression model with the potential of variable selection using the regularization parameter λ , which has a more prominent role than the other Bayesian methods. In a fully Bayesian framework, it is possible to treat λ as a random parameter and therefore different priors can be assigned to it. Despite the theoretical foundations of developing different priors and hyper parameters in Bayesian LASSO, the potential of applying different priors to this model has been less evaluated in genomic prediction studies. Therefore, the aim of the present study was to evaluate the application of different priors and hyper parameters to Bayesian LASSO regularization parameter on the genomic prediction of traits with different genetic architectures using real and simulated data of a mouse population. Variance component estimation and predictability of the models measured by prediction accuracy and bias were targeted.

Material and methods

In general, two datasets were analyzed in the present study. First, a real dataset consisted of phenotypic and genotypic data from a mice study. Second, a simulated dataset using mentioned real mice genome. The used phenotypes and genotypes and how they were collected have been fully described in previous studies, and these datasets are freely available (<http://gscan.well.ox.ac.uk>) and can be found in popular libraries in R such as BGLR.

Bayesian Least Absolute Shrinkage and Selection Operator (LASSO) was used to investigate its predictability under different priors for regularization parameter. Bayesian LASSO is a hierarchical model in which a double exponential distribution is considered as the marginal distribution of marker effect. This distribution is implemented as mixture of scaled normal densities by assigning independent normal densities to marker effects in the first level of the hierarchy. Then, IID exponential densities were assigned to the marker-specific scale parameters at the second hierarchy. Different priors were used for rate parameter at the final hierarchy. The effect of assigning three different priors including gamma, beta, and fixed distributions was investigated using three real traits of mice data and four simulated traits based on the real mice genome. These traits are selected to represent a range of different genetic architectures. Markov chain Monte Carlo (MCMC) was used to estimate the posterior distribution of variances component and the model parameters. The capability of the different models was compared by considering the correlation between the observed phenotypes and predicted GBV, and the regression slope of observed phenotypes onto predicted GBV from a fivefold cross validation. Estimated variance components and heritability of traits were also compared.

Results

Estimation of variance components and heritability are strongly affected by the assigned prior. The fixed λ model estimated the highest genetic variance and heritability and the lowest residual variance for different real traits. Regardless of the type of simulated trait, three priors underestimated the variance component and the heritability of traits. The average accuracy obtained for each trait using different priors was relatively similar, but standard deviation of the accuracies by the fixed λ model was always greater than the other two priors. Comparison of the minimum and maximum values of the models accuracies via fivefold cross validation revealed that the highest accuracy values were obtained using the assignment of gamma and beta distributions to the regularization parameter. By increasing the heritability and the number of background QTL of the traits, the accuracy of the models increased and the difference between the three different priors decreased. The difference between the models was greater in terms of prediction bias, but it generally followed the same trend of accuracy, i.e. the models with the highest accuracy also had the lowest prediction bias.

Conclusion

Based on the present results, the effect of prior assignment to λ is more critical for traits that are influenced by a small number of QTL or have low heritability.

Data Availability Statement

Analyzed data in the present work are freely available at (<http://gscan.well.ox.ac.uk>) and can be found in popular libraries in R such as BGLR.

Ethical considerations

This study did not involve data collection, and pre-existing data were used. The author avoided data fabrication, falsification, plagiarism, and misconduct.

Competing interests

The authors declare that they have no competing interests.

تأثیر پیشین‌های مختلف پارامتر تنظیم بیز LASSO بر پیش‌بینی‌های ژنومی صفات با معماری‌های ژنتیکی مختلف

سیدمهدی حسینی وردنجانی ✉

گروه علوم دامی، دانشکده کشاورزی، دانشگاه لرستان، خرم‌آباد ایران. رایانامه: hosseini.sm@lu.ac.ir

| اطلاعات مقاله | چکیده |
|--|--|
| نوع مقاله: مقاله پژوهشی | بیز LASSO یک روش رگرسیون جریمه‌ای و دارای پتانسیل انتخاب متغیر با استفاده از پارامتر تنظیم λ است. در چارچوب کاملاً بیزی امکان تصادفی تلقی کردن λ وجود دارد و می‌توان به آن پیشین‌های مختلف اختصاص داد. در این مطالعه اثر اختصاص سه نوع پیشین مختلف شامل توزیع گاما، بتا و ثابت بر قابلیت پیش‌بینی و برآورد پارامترهای ژنتیکی با استفاده از داده‌های ژنوم موش و هفت صفت با معماری‌های ژنتیکی مختلف بررسی شد. نتایج نشان داد که برآورد مؤلفه‌های واریانس و وراثت‌پذیری به شدت تحت تأثیر پیشین انتخابی قرار می‌گیرند. مدل λ ی ثابت بیشترین واریانس ژنتیکی و وراثت‌پذیری و کمترین میزان واریانس باقیمانده را برای صفات مختلف برآورد کرد. میانگین صحت بدست آمده برای هر صفت با استفاده از پیشین‌های مختلف نسبتاً مشابه بود ولی همواره صحت مدل λ ی ثابت دارای انحراف استاندارد بالاتر در مقایسه با دو پیشین دیگر بود. مقایسه مقادیر کمینه و بیشینه صحت مدل‌ها در قالب اعتبارسنجی متقابل پنج طرفه نشان داد که بالاترین مقادیر صحت با تخصیص توزیع‌های گاما و بتا به پارامتر تنظیم بدست می‌آید. با افزایش مقدار وراثت‌پذیری و تعداد QTL پس زمینه‌ای صفت، صحت مدل‌ها افزایش و تفاوت بین سه پیشین مختلف کاهش یافت. تفاوت مدل‌ها در رابطه با اریب پیش‌بینی‌ها بیشتر بود ولی عموماً همان روند صحت را دنبال می‌کرد، یعنی مدل‌هایی که بالاترین صحت را داشتند کمترین اریب پیش‌بینی را نیز داشتند. بر اساس نتایج حاضر، تخصیص پیشین مناسب به λ برای صفاتی که تحت تأثیر تعداد کمی ژن هستند و یا وراثت‌پذیری پایین‌تری دارند ضروری‌تر است. |
| تاریخ دریافت: ۱۴۰۴/۰۴/۰۷ تاریخ بازنگری: ۱۴۰۴/۰۶/۱۶ تاریخ پذیرش: ۱۴۰۴/۰۸/۰۷ تاریخ انتشار: تابستان ۱۴۰۵ | |
| کلیدواژه‌ها: رویکرد بیزی، پیش‌بینی ژنومی، قابلیت پیش‌بینی، پیشین مناسب، انتخاب متغیر | |

استناد: حسینی وردنجانی، سیدمهدی (۱۴۰۵). تأثیر پیشین‌های مختلف پارامتر تنظیم بیز LASSO بر پیش‌بینی‌های ژنومی صفات با معماری‌های ژنتیکی مختلف. نشریه علوم دامی ایران، ۵۷ (۲)، ۱۹۵-۲۱۱. DOI: <https://doi.org/10.22059/ijas.2025.397606.654084>



© نویسندگان.

DOI: <https://doi.org/10.22059/ijas.2025.397606.654084>

ناشر: مؤسسه انتشارات دانشگاه تهران.

مقدمه

انتخاب ژنومی (GS) به تصمیم‌گیری برای انتخاب افراد بر اساس ارزش‌های اصلاحی برآورد شده از اطلاعات ژنومی اشاره دارد (Meuwissen *et al.*, 2001, Nejati-Javaremi *et al.*, 1997). برای محاسبه ارزش‌های اصلاحی ژنومی (GBV)، ابتدا یک مدل آماری پیش‌بینی بر اساس تعداد زیادی از نشانگرهای مولکولی (معمولاً چند شکلی‌های تک نوکلئوتیدی (SNP)) برازش داده می‌شود. اثرات این نشانگرها در یک جمعیت مرجع که در آن افراد دارای رکوردهای فنوتیپی و ژنوتیپی هستند برآورد می‌شوند. در جمعیت یا در نسل‌های بعدی می‌توان حیوانات را برای نشانگرهای پراکنده شده در کل ژنوم تعیین ژنوتیپ کرد و با استفاده از برآوردهای مرحله قبل، GBV آنها را پیش‌بینی کرد. پس از معرفی انتخاب ژنومی، بسیاری از محققان این فن‌آوری نوین را در گونه‌های مختلف حیوانات مزرعه و یا محصولات زراعی با استفاده از داده‌های حقیقی و یا داده‌های شبیه‌سازی ارزیابی نموده‌اند. تاکنون صحت ارزش‌های اصلاحی پیش‌بینی‌شده با استفاده از اطلاعات ژنومی در بسیاری از گونه‌های حیوانات مزرعه گزارش شده است (Yuan *et al.*, 2025, Alemu *et al.*, 2025, Ros-Freixedes *et al.*, 2022, Calus *et al.*, 2014, Hosseini- (Vardanjani *et al.*, 2019).

پیشینه پژوهش

با معرفی انتخاب ژنومی، یک حوزه تحقیقاتی بسیار فراگیر در زمینه توسعه مدل‌های آماری کارآمد و مبتنی بر پس‌زمینه ژنتیکی صفات یا به اصطلاح معماری ژنتیکی صفات شکل گرفت تا انواع اطلاعات دیگر را نیز در پیش‌بینی ژنومی وارد کند (Habier *et al.*, 2011). شواهد تجربی حاکی از آن است که در شرکت‌های تجاری و تحقیقاتی اصلاح‌نژادی و همچنین صنعت دام و طیور، مدل‌های مختلفی برای پیش‌بینی ژنومی توسعه و بکارگیری شده‌اند. در بسیاری از این مراکز، صفات مختلفی از جمله صفات کمی رکوردبرداری می‌شوند و ارتباط بین آنها و ویژگی‌های ژنومی خاص بررسی می‌شود. به طور معمول ارتباط بین متغیر وابسته با متغیر یا متغیرهای مستقل (SNPها) با مدل‌های خطی برازش داده می‌شود. چنین رویکردی امکان برآورد اثرات نشانگر را بر اساس میانگین شرطی متغیر وابسته یعنی $E(y|SNPs)$ فراهم می‌کند که y متغیر پاسخ و SNPها متغیرهای مستقل هستند. از آنجایی که SNPها موجود در بیشتر حیوانات مزرعه، برای اهداف تحقیقاتی و کاربردی برنامه‌های اصلاح‌نژاد و انتخاب تجاری، نسبت به فنوتیپ‌ها به طور قابل توجهی افزایش یافته است، روش‌های بیزی برای مدیریت این مجموعه‌های بسیار بزرگ از متغیرهای توضیحی با تعداد محدودی از مشاهدات استفاده می‌شوند. روش‌های بیزی به طور معمول با در نظر گرفتن یک توزیع نرمال (به عنوان مثال بیز ریج (BR)) و یا سایر توزیع‌ها در روش‌های انقباضی دیگر (مثلاً بیز A یا بیز B) (Meuwissen *et al.*, 2001, Gianola, 2013) برای ضرایب تصادفی نشانگرها، میزان تأثیر هر SNP بر واریانس کل صفت را تشریح می‌کنند. این رویکردها قادر به ارائه اطلاعاتی راجع به مناطقی از ژنوم که بر صفت مورد نظر اثر دارند، هستند. با این حال، مطالعات پویش گسترده ژنوم، نشان دادند که احتمالاً بسیاری از نشانگرها در مناطقی از ژنوم هستند که در تشریح واریانس صفت نقشی ندارند. در مقابل برخی از نشانگرها در عدم تعادل پیوستگی با برخی مناطق یا QTLهایی در ژنوم هستند که بخش قابل توجهی از واریانس صفت را تشریح می‌کنند. این نشان می‌دهد که انقباض افتراقی اثرات نشانگر باید جزء ویژگی‌های مدل مورد استفاده در پیش‌بینی ژنومی باشد.

جذابیت و موفقیت مدل‌های جریمه‌ای در بسیاری از حوزه‌های کاربردی، از جمله انتخاب ژنومی، وابسته به استفاده آنها از پارامترهای جریمه‌ای است که برازش مدل‌ها با پیش‌بینی‌کننده‌هایی را که به هزاران عدد می‌رسند را با کارایی و دقت پیش‌بینی بالا تسهیل می‌کنند. ممکن است بسیاری از این پیش‌بینی‌کننده‌ها حتی به متغیر پاسخ مربوط نباشند، همچنین بسیار بیشتر از اندازه نمونه باشند یا همبستگی بالایی بین آنها وجود داشته باشد. در این راستا، استفاده از مجموع مقادیر مطلق ضرایب رگرسیون

1. Genomic selection
2. Genomic breeding value
3. Single Nucleotide Polymorphisms
4. Bayesian ridge regression
5. Shrinkage methods

نرم (L1) به عنوان عامل جریمه در مدل‌های رگرسیونی توسط Tibshirani (1996) پیشنهاد شد که پتانسیل همزمان انتخاب متغیر و انقباض ضرایب را دارد و آن را عملگر انتخاب و انقباض کمترین قدر مطلق (LASSO) نامید. Park and Casella (2008)، رویکرد بیزی LASSO را به صورت $DE(\beta_j|\lambda^2, \sigma_\varepsilon^2) = \int N(\beta_j|0, \sigma_\varepsilon^2 \tau_j^2) \text{Exp}\left(\tau_j^2 \left|\frac{\lambda^2}{2}\right.\right) \partial \sigma_{\beta_j}^2$ در واقع، LASSO یک روش رگرسیون جریمه‌شده است و شدت جریمه‌سازی توسط مقادیر پارامتر تنظیم مشخص می‌شود که به طور گسترده با λ شناخته می‌شوند. در چارچوب LASSO کاملاً بیزی، برای هر یک از ناشناخته‌ها، توزیع پیشینی اختصاص می‌یابد که در de los Campos *et al.* (2009) به طور مشروح مورد بحث قرار گرفته است. اگر چه پارامتر تنظیم (λ) نیز می‌تواند به عنوان ناشناخته تلقی شود ولی مشخص نیست که نتایج چقدر ممکن است نسبت به انتخاب پیشین و فرآیندهای توزیع پیشین مربوطه حساس باشند. با وجود مبانی نظری توسعه‌ی پیشین‌ها و فرآیندهای مختلف در بیز LASSO، به ارزیابی پتانسیل بکارگیری پیشین‌های مختلف بر قابلیت پیش‌بینی‌های ژنومی با استفاده از این مدل، کمتر پرداخته شده است. بنابراین هدف مطالعه حاضر ارزیابی بکارگیری انواع مختلف پیشین‌ها و فرآیندها در بیز LASSO بر قابلیت پیش‌بینی ژنومی صفات با معماری‌های ژنتیکی مختلف با استفاده از داده‌های واقعی و شبیه‌سازی شده بر اساس ژنوم واقعی موش بود.

روش‌شناسی پژوهش

رکوردهای فنوتیپی و ژنوتیپی

در این پژوهش، دو مجموعه داده‌ی واقعی و شبیه‌سازی بررسی شد. مجموعه داده‌ی واقعی شامل داده‌های فنوتیپی و ژنوتیپی مربوط به مطالعه‌ای در موش بودند. فنوتیپ‌ها و ژنوتیپ‌های مورد استفاده و نحوه‌ی جمع‌آوری آنها در مطالعات قبلی به طور مثال Solberg *et al.* (2006) و Valdar *et al.* (2006b) به طور کامل شرح داده شده است. این داده‌ها به صورت آزاد (<http://gscan.well.ox.ac.uk>) در اختیار عموم است و در کتابخانه‌های معروفی در نرم افزار R نظیر BGLR یافت می‌شوند. مجموعه داده‌ی واقعی شامل رکوردهای فنوتیپی برخی صفات مورفولوژیکی و برخی صفات بیوشیمی مربوط به حدود ۱۹۰۰ موش است. علاوه بر این صفات، متغیرهای کمکی نظیر جنس، سن، شناسه قفس و غیره نیز برای موش‌ها در دسترس است. داده‌های ژنوتیپی نیز شامل ۱۰۹۴۶ نشانگر پراکنده شده در ژنوم هستند. در این مطالعه داده‌های ویرایش شده مربوط به ۱۸۱۴ موش و ۱۰۳۴۶ نشانگر استفاده شدند. از میان صفات مورفولوژیکی دو صفت وزن بدن و طول بدن و از میان صفات بیوشیمی صفت پروتئین کل خون انتخاب شدند. انتخاب این صفات به این دلیل بود که معرف دامنه‌ای از معماری‌های ژنتیکی مختلف شامل صفات تحت تأثیر تعداد زیادی ژن (به اصطلاح چندژنی) و در مقابل صفات تحت تأثیر تعداد ژن‌های بسیار کمتر باشند. همچنین دو صفت چندژنی اول به ترتیب دارای وراثت‌پذیری در حدود ۰/۷ و ۰/۲۵ هستند، یعنی در گروه صفات چندژنی با وراثت‌پذیری بالا و متوسط قرار می‌گیرند. همچنین صفت سوم نیز وراثت‌پذیری در حدود ۰/۱ دارد و به احتمال قوی تعداد ژن‌های بسیار کمی در تنوع آن دخالت دارند. جهت آگاهی بیشتر از نحوه‌ی معماری ژنتیکی دقیق‌تر این صفات می‌توان به مطالعات Valdar *et al.* (2006b) و Valdar *et al.* (2006a) که برای شناسایی و تعیین محل QTL مؤثر بر صفات پیچیده انجام شده است، رجوع کرد. مجموعه داده‌ی شبیه‌سازی شده، شامل چهار صفت بود. در فرآیند شبیه‌سازی، فنوتیپ صفات بر اساس رکوردهای ژنوتیپی واقعی موش تولید شدند، این یعنی صفات شبیه‌سازی شده نیز به جای یک ژنوم شبیه‌سازی شده، از همان ویژگی‌های ساختار واقعی ژنوم موش مثلاً همان عدم تعادل پیوستگی پیروی می‌کنند. در این بخش سعی شد صفات طوری شبیه‌سازی شوند که دقیقاً معماری مقابل صفات واقعی داشته باشند. به این شکل که متقابل با دو صفت اول که تحت تأثیر تعداد زیادی ژن بودند، صفاتی شبیه‌سازی شد که از همان ساختار ژنوم واقعی و همان مقدار وراثت‌پذیری تبعیت می‌کردند ولی تحت تأثیر تعداد کمی QTL، در اینجا ۱۰ عدد، قرار می‌گرفتند. صفت شبیه‌سازی شده سوم نیز وراثت‌پذیری ۰/۱ را داشت ولی تحت تأثیر تعداد زیادی QTL قرار می‌گرفت، در اینجا ۱۰۰۰ عدد، در نظر گرفته شد. پس از تجزیه و تحلیل ۶ صفت ذکر شده و مشاهده روند پیش‌بینی‌ها،

1. Least absolute shrinkage and selection operator
2. Regularization
3. Hyperparameter

شبیه‌سازی صفت چهارم نیز انجام گرفت. به این شکل که تصور شد که نیاز است صفتی با وراثت‌پذیری پایین (یعنی همان ۰/۱) و تحت تأثیر سناریوی تعداد کمی ژن (یعنی ۱۰ عدد QTL) شبیه‌سازی شود. برای شبیه‌سازی QTLها، از نشانگرهای موجود بدون اعمال هیچ معیاری و به طور تصادفی به تعداد QTLهای مورد نظر برای هر صفت انتخاب و به عنوان QTL تلقی شدند (واضح است که این QTLها در تجزیه و تحلیل‌های بعدی کنار گذاشته شدند). اثرات QTLها نیز از یک توزیع گاما نمونه‌گیری شدند. ارزش اصلاحی واقعی (TBV) فرد از مجموع اثرات QTLهایی که آن فرد حمل می‌کرد محاسبه شد. برای ساخت فنوتیپ فرد نیز ابتدا یک اثر باقیمانده از یک توزیع نرمال نمونه‌گیری شد. میانگین این توزیع نرمال صفر بود و واریانس آن طوری انتخاب شد که با توجه به واریانس ارزش‌های اصلاحی واقعی، وراثت‌پذیری مورد نظر را تأمین کند. مجموع ارزش اصلاحی واقعی و اثر باقیمانده، فنوتیپ فرد را تشکیل می‌داد. فرآیند شبیه‌سازی، ۱۵ بار تکرار شد و پارامترهای مورد ارزیابی به صورت میانگین به همراه انحراف استاندارد گزارش شده‌اند. بنابراین در کل هفت صفت (۳ صفت واقعی و ۴ صفت شبیه‌سازی شده) با معماری‌های ژنتیکی مختلف بررسی شد که صفات واقعی با نام آنها و صفات شبیه‌سازی شده به ترتیب با شماره‌های ۱ تا ۴ مشخص شدند. سعی شد در مجموع ۷ صفت مورد بررسی، صفاتی با وراثت‌پذیری بالا تا پایین و تحت تأثیر تعداد زیاد تا کم QTL بررسی شوند. از طرفی استفاده همزمان از صفات واقعی و صفات شبیه‌سازی شده‌ای که آنها نیز با استفاده از ژنوم واقعی تولید شده بودند امکان بررسی و اظهار نظر و بسط به داده‌های واقعی را بهتر میسر می‌کند. در جدول یک خلاصه‌ای از صفات مورد بررسی به تفکیک گزارش شده است.

جدول ۱. معماری ژنتیکی صفات مورد بررسی با توجه به مقدار وراثت‌پذیری و تعداد QTL مؤثر

Table 1. Genetic architecture of investigated traits with respect to heritability and number of effective QTL

| وراثت‌پذیری | | | | |
|-------------|-------|-------------------|-------------------|-------------------|
| بالا | | متوسط | پایین | |
| تعداد QTL | بالا | طول بدن | شبیه‌سازی شده (۳) | |
| | پایین | شبیه‌سازی شده (۲) | پروتئین کل خون | شبیه‌سازی شده (۴) |

مدل آماری و توزیع‌های پیشین

تمام تجزیه و تحلیل‌ها در چارچوب بیزی و با استفاده از نمونه‌گیری گیبس در قالب روش بیز LASSO انجام گرفت. در کل ۲۰۰۰۰ چرخه گیبس اجرا شد که ۳۰۰۰ دور اول آن به عنوان دوره سوخته در نظر گرفته شدند. همگرایی نامعلوم‌های مدل با بررسی گراف‌های نمونه‌های پسین بررسی شد. مدل خطی عمومی زیر برای برآورد اثرات ثابت و تصادفی صفات واقعی و شبیه‌سازی شده استفاده شد:

$$y = Xb + Zg + e \quad (1)$$

که در آن y بردار ارزش‌های فنوتیپی (واقعی یا شبیه‌سازی شده)، b بردار اثرات ثابت، g بردار اثرات نشانگرها و e نیز بردار باقیمانده هستند. ماتریس‌های X و Z نیز به ترتیب رکوردها را به اثرات ثابت و نشانگرها مرتبط می‌نمایند. مطالعات قبلی بر روی این مجموعه داده گزارش کردند که متغیرهای جنس، اندازه زایش و قفسی که موش‌ها در آن پرورش می‌یابند، بر روی صفات اندازه‌گیری شده اثر معنی‌دار دارند. بنابراین زمانی که داده‌های واقعی واکاوی شدند، علاوه بر اثر میانگین، این سه متغیر نیز به عنوان عوامل ثابت در مدل وارد شدند.

در چارچوب بیزی برای متغیرهای ثابت میانگین، جنس و اندازه زایش یک توزیع پیشین یکنواخت در نظر گرفته شد. به اثر قفس یک توزیع پیشین نرمال $N(\text{Cage}|0, \sigma_{\text{cage}}^2)$ و به پارامتر واریانس آن نیز یک توزیع مربع کای معکوس مقیاس شده $\chi^{-2}(\sigma_{\text{cage}}^2 | df_{\text{Cage}}, S_{\text{Cage}})$ اختصاص داده شد. در بیز LASSO، توزیع حاشیه‌ای اثرات نشانگرها نمایی مضاعف است و به صورت مخلوطی از چگالی‌های نرمال مقیاس شده اجرا می‌گردد. در اولین سطح مرتبه‌بندی این روش بر طبق Pérez and de los Campos (2014)، به اثرات نشانگر یک توزیع نرمال مستقل با میانگین صفر و واریانس خاص نشانگر $\tau_{jk}^2 \times \sigma_e^2$ اختصاص

داده شد که به واریانس باقیمانده یک توزیع مربع کای معکوس مقیاس شده $\chi^{-2}(\sigma_e^2 | S_e, df_e)$ اختصاص داده شد. در دومین سطح به پارامتر مقیاس خاص نشانگر، τ_{jk}^2 ، چگالی نمایی مستقل و یکسان $\left\{ \tau_{jk}^2 \left| \frac{\lambda^2}{2} \right. \right\}$ با پارامتر نرخ $\frac{\lambda^2}{2}$ اختصاص داده می‌شود. در نهایت در آخرین سطح مرتبه‌بندی، λ^2 به صورت ثابت، یا دارای توزیع گاما $G(\lambda^2 | r, s)$ با پارامترهای نرخ و شکل λ و s یا دارای توزیع بتا، $B\left(\frac{\lambda}{\max} | p_0, \pi_0\right)$ در نظر گرفته شد. توزیع توأم اثرات نشانگرها و فرآپارامترهای هر یک از سه مدل شرح داده شده به ترتیب در زیر آورده شده است:

$$p(g_j, \tau_j^2 | \sigma_e^2, \lambda) = \left\{ \Pi_k N(g_{jk} | 0, \tau_{jk}^2 \times \sigma_e^2) \text{Exp} \left\{ \tau_{jk}^2 \left| \frac{\lambda^2}{2} \right. \right\} \right\} \quad (2)$$

$$p(g_j, \tau_j^2, \lambda^2 | \sigma_e^2) = \left\{ \Pi_k N(g_{jk} | 0, \tau_{jk}^2 \times \sigma_e^2) \text{Exp} \left\{ \tau_{jk}^2 \left| \frac{\lambda^2}{2} \right. \right\} \right\} \times G(\lambda^2 | r, s) \quad (3)$$

$$p(g_j, \tau_j^2, \lambda | \sigma_e^2, \max) = \left\{ \Pi_k N(g_{jk} | 0, \tau_{jk}^2 \times \sigma_e^2) \text{Exp} \left\{ \tau_{jk}^2 \left| \frac{\lambda^2}{2} \right. \right\} \right\} \times B\left(\frac{\lambda}{\max} | p_0, \pi_0\right) \quad (4)$$

اعتبارسنجی مدل‌ها

برای ارزیابی مدل‌ها از روش اعتبارسنجی متقابل ۵ طرفه استفاده شد. در این روش مجموعاً حیوانات جامعه مرجع به ۵ گروه تقسیم شدند و در هر دور ۴ گروه از آنها به عنوان گروه آموزش و گروه پنجم به عنوان آزمون در نظر گرفته شد. در گروه‌های آموزش پارامترهای مدل برآورد و در گروه پنجم اعتبارسنجی شد. از دو معیار همبستگی و اریب برای مقایسه مدل‌ها استفاده شد. این دو معیار در هر دور اندازه‌گیری شدند و میانگین ۵ دور گزارش گردید. برای سنجش قابلیت پیش‌بینی مدل‌ها از همبستگی بین فنوتیپ‌ها و ارزش‌های ژنتیکی پیش‌بینی شده به عنوان صحت هر مدل استفاده شد. برای سنجش اریبی پیش‌بینی‌ها نیز از تابعیت فنوتیپ‌ها بر روی ارزش‌های ژنتیکی پیش‌بینی شده استفاده شد. در این شیوه ضریب رگرسیون برابر با یک نشان دهنده عدم اریبی پیش‌بینی‌ها و ضریب رگرسیون کمتر از یک و بیشتر از یک نیز به ترتیب نشان دهنده بیش برآورد و کم برآورد ارزش‌های ژنتیکی پیش‌بینی شده در مقایسه با فنوتیپ مشاهده شده است. مؤلفه‌های واریانس ژنتیکی و باقیمانده و همچنین مقدار وراثت‌پذیری هر صفت نیز با استفاده از پیشین‌های مختلف برآورد و مورد مقایسه قرار گرفتند.

یافته‌های پژوهش

برآورد مؤلفه‌های واریانس ژنتیکی، باقیمانده و وراثت‌پذیری برای سه صفت واقعی مورد بررسی شامل وزن بدن، طول بدن و پروتئین کل خون در جدول ۲ نشان داده شده است. وراثت‌پذیری صفات با استفاده از پیشین‌های مختلف به طور قابل توجهی تفاوت داشت. این تفاوت در رابطه با وزن بدن در حدود ۰/۱۱ بود (۰/۳۱ تا ۰/۴۲). در مقایسه با وراثت‌پذیری، مقادیر مطلق مؤلفه‌های واریانس کمتر متفاوت بودند. مؤلفه‌های واریانس ژنتیکی و باقیمانده برآورد شده برای هر یک از سه صفت با استفاده از اختصاص پیشین‌های بتا و گاما شباهت بیشتری به هم داشت نسبت به آنچه با ثابت تلقی کردن λ حاصل می‌شد. به طور مثال در رابطه با وزن بدن دو پیشین نام‌برده شده به ترتیب مقادیر ۰/۰۹۱ را برای واریانس ژنتیکی و مقادیر ۰/۲۰۲ و ۰/۲۰۱ را برای واریانس باقیمانده برآورد کردند. در حالی که با λ ثابت این دو پارامتر به ترتیب ۰/۱۲۴ و ۰/۱۷۳ برآورد شدند.

1. Rate
2. Shape
3. Beta
4. Fivefold cross validation

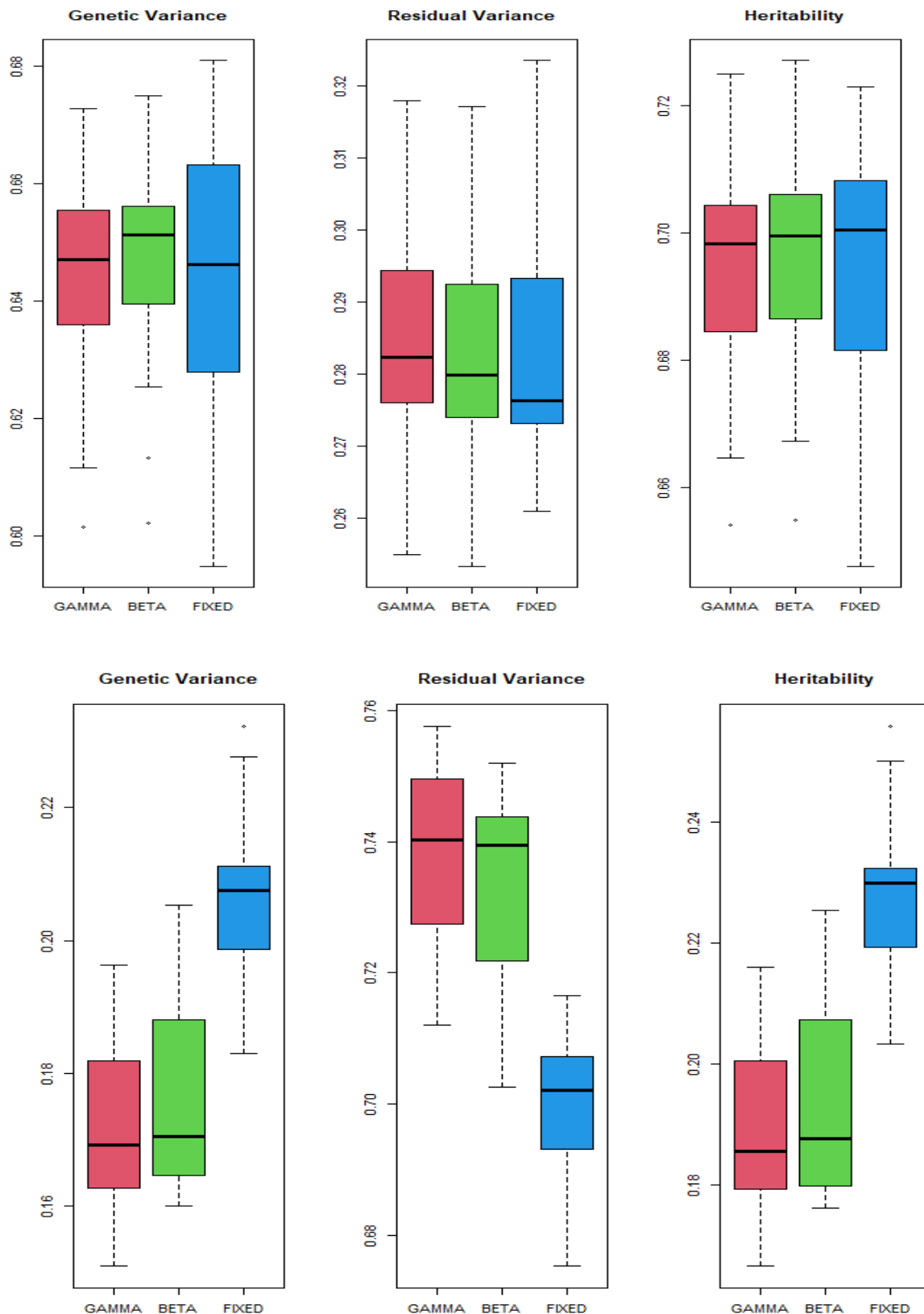
جدول ۲. برآورد مؤلفه‌های واریانس و وراثت‌پذیری صفات واقعی با استفاده از بیز LASSO با پیشین‌های مختلف

Table 2. Estimation of variance components and heritability of real traits using Bayesian LASSO with different priors

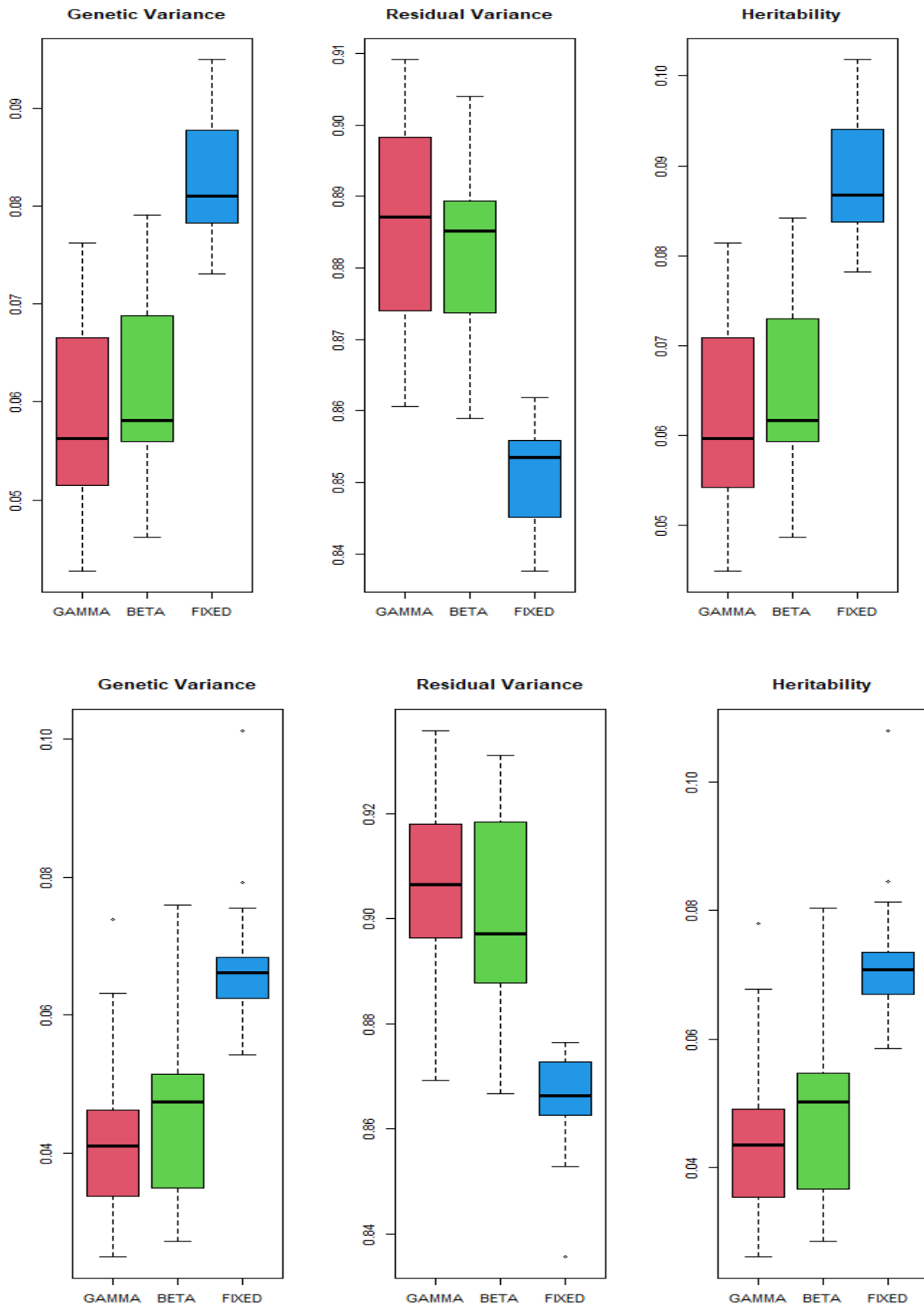
| | Priors | Genetic Variance | Residual Variance | Heritability |
|---------------------|--------|------------------|-------------------|--------------|
| Body weight | GAMMA | 0.091 | 0.202 | 0.311 |
| | BETA | 0.091 | 0.201 | 0.312 |
| | FIXED | 0.124 | 0.173 | 0.419 |
| Body length | GAMMA | 0.075 | 0.503 | 0.129 |
| | BETA | 0.080 | 0.499 | 0.139 |
| | FIXED | 0.100 | 0.482 | 0.172 |
| Total blood protein | GAMMA | 0.011 | 0.617 | 0.017 |
| | BETA | 0.011 | 0.617 | 0.018 |
| | FIXED | 0.032 | 0.586 | 0.051 |

مطابق با جدول ۲، نمودار جعبه‌ای برآورد مؤلفه‌های واریانس و وراثت‌پذیری برای صفات شبیه‌سازی شده حاصل از تکرارهای مختلف در شکل‌های ۱ و ۲ آورده شده است. فارغ از صفت شبیه‌سازی شده، تمامی پیشین‌ها پارامترهای مورد بررسی را کمتر از مقدار شبیه‌سازی شده برآورد کردند. به طور مثال واریانس ژنتیکی شبیه‌سازی شده برای دو صفت اول به ترتیب ۰/۷ و ۰/۲۵ بود (وراثت‌پذیری‌ها به ترتیب ۰/۷ و ۰/۲۵) ولی به طور آشکار در شکل یک دیده می‌شود که هر سه مدل این پارامتر را کمتر برآورد کردند. همچنین هر چه مقدار وراثت‌پذیری کاهش یافت این انحراف بیشتر شد. در یک وراثت‌پذیری مشابه نیز، انحراف از مقدار واقعی پارامترها برای صفات تحت تأثیر تعداد کمتر QTL بیشتر بود (شکل ۲). در تطابق با نتایج حاصل از صفات واقعی، هر چه وراثت‌پذیری صفت کاهش یافت میزان اختلاف بین برآوردهای حاصل از ۳ پیشین مختلف اختصاص داده شده به λ افزایش یافت. به طوری که حتی دامنه تغییرات حاصل از ۳ پیشین در وراثت‌پذیری ۰/۱ بر هم منطبق نبودند. هر چه وراثت‌پذیری صفات و QTL‌های زمینه‌ای آنها کمتر شد تفوق انتخاب یک میزان بالا برای λ یعنی آنچه در مدل ثابت رخ می‌داد آشکارتر می‌شد. برآوردهای حاصل از اختصاص توزیع‌های گاما و بتا به پارامتر تنظیم مشابهت بیشتری به همدیگر داشتند. اگر چه همین دو پیشین نیز همان روند تمایل به افتراق را در وراثت‌پذیری‌های پایین نشان می‌دادند. دامنه تغییرات هر یک از این مؤلفه‌ها نیز جالب به نظر می‌رسد. هر چقدر وراثت‌پذیری و تعداد QTL زمینه‌ای کاهش می‌یابد، تکرار پذیری نتایج بالاتر می‌رود (جعبه‌ها فشرده‌تر می‌گردند). این مورد در مدل λ ثابت نسبت به دو پیشین دیگر به طور آشکارتر صدق می‌کند.

نکته‌ی دیگری که در اینجا باید اشاره شود توجه همزمان به مؤلفه‌های برآورد شده در مورد پروتئین کل خون (به عنوان صفت واقعی) و صفات شبیه‌سازی شده‌ی ۳ و ۴ (به عنوان صفات شبیه‌سازی جفت آن) است. توجه کنید که در این سه صفت وراثت‌پذیری تقریباً یکسان است (در شبیه‌سازی شده‌ها دقیقاً ۰/۱ و در صفت واقعی حدود ۰/۱ گزارش شده است) فقط تعداد QTL زمینه‌ای تفاوت دارد. با توجه به مقدار مؤلفه‌های برآورد شده و روند آنها این استنباط متصور است که احتمالاً صفت پروتئین کل خون باید تعداد QTL زمینه‌ای کمتر از ۱۰ عدد را داشته باشد. این روند و روندی که در مورد صحت پیش‌بینی برای این صفت خواهید دید به عنوان دلیلی برای شبیه‌سازی صفت ۴ بود. هنوز این استنباط نیاز به بررسی بیشتر و اضافه شدن اطلاعات مولکولی و بیولوژیکی بیشتر دارد.



شکل ۱. نمودار جعبه‌ای برآورد مؤلفه‌های واریانس و وراثت‌پذیری صفات ۱ (بالا) و ۲ (پایین) شبیه‌سازی شده با استفاده از بیز LASSO با پیشین‌های مختلف
 Figure 1. Boxplot of variance components and heritability of simulated trait 1 (upper) and 2 (down) using Bayesian LASSO with different priors



شکل ۲. نمودار جعبه‌ای برآورد مؤلفه‌های واریانس و وراثت‌پذیری صفات ۳ (بالا) و ۴ (پایین) شبیه‌سازی شده با استفاده از بیز LASSO با پیشین‌های مختلف
Figure 2. Boxplot of variance components and heritability of simulated trait 3 (upper) and 4 (down) using Bayesian LASSO with different priors

کمینه، بیشینه و میانگین (\pm انحراف استاندارد) مقادیر صحت و اریب پیش‌بینی‌ها برای سه صفت واقعی و چهار صفت شبیه‌سازی شده به ترتیب در جدول ۳ و ۴ آورده شده است. در جدول ۳ ستونی با نام صحت مقیاس شده آورده شده که از تقسیم میانگین صحت بر جذر وراثت‌پذیری حاصل می‌شود. این ستون امکان مقایسه مدل‌ها برای صفات مختلف علاوه بر هر صفت به تنهایی را نیز فراهم می‌کند. در کل میانگین صحت هر یک از مدل‌ها در صفات شبیه‌سازی شده بالاتر از صفات واقعی بود. پایین‌ترین مقادیر صحت بدون توجه به پیشین‌های مختلف اختصاص داده شده، در رابطه با صفت پروتئین کل خون بدست آمد. مقایسه مقادیر صحت این صفت نیز نشان داد که اختصاص مقدار ثابت به λ به طور قابل توجهی منجر به افزایش صحت ارزیابی‌ها در مقایسه با دو پیشین دیگر می‌شود. در رابطه با دو صفت دیگر تفاوت چشم‌گیری بین سه پیشین اختصاص داده شد دیده نشد، اگر چه اختصاص توزیع بتا به صحت بالاتری در رابطه با طول بدن منجر گردید. اریب پیش‌بینی‌ها همان روند صحت را نشان دادند. به این صورت که مدل‌های دارای صحت بالاتر، اریب کمتر و مدل‌های با صحت پایین‌تر، اریب بیشتر داشتند. همچنین میانگین اریب پیش‌بینی‌ها در صفات شبیه‌سازی شده همواره کمتر از صفات واقعی بود. با توجه به میانگین اریب‌ها و داشتن ضریب بالاتر از یک استنباط می‌گردد که هر سه مدل به اتفاق ارزش‌های اصلاحی حیوانات را کمتر از حد برآورد کرده‌اند. این میزان کم برآوردی برای مدل λ ثابت کمتر از دو مدل دیگر بود.

جدول ۳. صحت و اریب پیش‌بینی‌ها (کمینه، بیشینه، میانگین \pm انحراف استاندارد، ضریب تغییرات) در صفات واقعی با استفاده از بیز LASSO با پیشین‌های مختلف

Table 3. Accuracy and bias of predictions (Min, Max, Mean \pm SD, CV) in real traits using Bayesian LASSO with different priors

| | Prior | Accuracy | | | | | Bias | | |
|---------------------|-------|----------|-------|------------------|--------|--------|-------|-------|------------------|
| | | Min | Max | Mean \pm SD | CV | Scaled | Min | Max | Mean \pm SD |
| Body weight | GAMMA | 0.328 | 0.388 | 0.350 \pm 0.02 | 6.592 | 0.418 | 1.121 | 1.383 | 1.252 \pm 0.10 |
| | BETA | 0.199 | 0.400 | 0.339 \pm 0.08 | 24.951 | 0.405 | 0.710 | 1.451 | 1.217 \pm 0.30 |
| | FIXED | 0.289 | 0.395 | 0.352 \pm 0.04 | 12.285 | 0.421 | 0.901 | 1.158 | 1.053 \pm 0.10 |
| Body length | GAMMA | 0.332 | 0.402 | 0.353 \pm 0.03 | 8.000 | 0.706 | 1.292 | 1.560 | 1.401 \pm 0.10 |
| | BETA | 0.271 | 0.435 | 0.370 \pm 0.06 | 16.563 | 0.740 | 0.933 | 2.108 | 1.474 \pm 0.40 |
| | FIXED | 0.249 | 0.439 | 0.364 \pm 0.08 | 20.759 | 0.728 | 0.814 | 1.655 | 1.264 \pm 0.30 |
| Total blood protein | GAMMA | 0.125 | 0.181 | 0.149 \pm 0.03 | 18.864 | 0.471 | 0.901 | 2.251 | 1.709 \pm 0.60 |
| | BETA | 0.076 | 0.197 | 0.148 \pm 0.05 | 33.164 | 0.468 | 0.523 | 3.022 | 1.765 \pm 0.90 |
| | FIXED | 0.033 | 0.302 | 0.179 \pm 0.10 | 53.996 | 0.566 | 0.186 | 1.989 | 1.127 \pm 0.60 |

جدول ۴. صحت و اریب پیش‌بینی‌ها (کمینه، بیشینه، میانگین \pm انحراف استاندارد، ضریب تغییرات) در صفات شبیه‌سازی شده با استفاده از بیز LASSO با پیشین‌های مختلف

Table 4. Accuracy and bias of predictions (Min, Max, Mean \pm SD, CV) in simulated traits using Bayesian LASSO with different priors

| | Prior | Accuracy | | | | | Bias | | |
|---------|-------|----------|-------|------------------|--------|--------|-------|-------|------------------|
| | | Min | Max | Mean \pm SD | CV | Scaled | Min | Max | Mean \pm SD |
| Trait 1 | GAMMA | 0.694 | 0.828 | 0.780 \pm 0.02 | 3.107 | 0.932 | 0.869 | 1.093 | 1.003 \pm 0.05 |
| | BETA | 0.692 | 0.833 | 0.780 \pm 0.03 | 3.647 | 0.932 | 0.878 | 1.113 | 1.001 \pm 0.05 |
| | FIXED | 0.705 | 0.821 | 0.781 \pm 0.02 | 3.146 | 0.933 | 0.910 | 1.134 | 1.012 \pm 0.05 |
| Trait 2 | GAMMA | 0.271 | 0.515 | 0.386 \pm 0.04 | 10.472 | 0.772 | 0.722 | 1.443 | 1.012 \pm 0.10 |
| | BETA | 0.316 | 0.464 | 0.387 \pm 0.04 | 10.651 | 0.774 | 0.679 | 1.361 | 1.003 \pm 0.10 |
| | FIXED | 0.287 | 0.471 | 0.387 \pm 0.05 | 12.440 | 0.774 | 0.636 | 1.317 | 0.921 \pm 0.10 |
| Trait 3 | GAMMA | 0.113 | 0.387 | 0.235 \pm 0.05 | 22.125 | 0.743 | 0.475 | 1.977 | 1.055 \pm 0.30 |
| | BETA | 0.131 | 0.342 | 0.230 \pm 0.05 | 19.995 | 0.727 | 0.521 | 1.710 | 0.999 \pm 0.20 |
| | FIXED | 0.090 | 0.389 | 0.229 \pm 0.05 | 21.908 | 0.724 | 0.299 | 1.533 | 0.852 \pm 0.20 |
| Trait 4 | GAMMA | 0.067 | 0.323 | 0.186 \pm 0.06 | 30.236 | 0.588 | 0.387 | 2.338 | 1.061 \pm 0.40 |
| | BETA | 0.063 | 0.313 | 0.192 \pm 0.05 | 26.162 | 0.607 | 0.272 | 2.091 | 1.040 \pm 0.30 |
| | FIXED | 0.081 | 0.293 | 0.193 \pm 0.05 | 26.188 | 0.610 | 0.335 | 1.281 | 0.810 \pm 0.20 |

بحث

هدف مطالعه حاضر، بررسی توانایی پیش‌بینی بیز LASSO وقتی پیشین‌های مختلف برای پارامتر تنظیم λ استفاده می‌شود، بود. اولین بخش استفاده از داده‌های واقعی، با در نظر گرفتن اطلاعات ژنومی و برخی صفات از مطالعه‌ای در موش بود. نتایج این بخش با استفاده از داده‌های شبیه‌سازی شده‌ی دیگری بر مبنای ژنوم واقعی موش تکمیل گردید تا به یک استنباط نسبتاً قابل استناد که تا حد ممکن دامنه‌ای از صفات با معماری‌های ژنتیکی مختلف را در نظر گرفته باشد منجر گردد.

مؤلفه‌های واریانس و وراثت‌پذیری به عنوان پارامترهای کلیدی در برنامه‌های اصلاح‌نژادی، به شدت تحت تأثیر پیشین انتخابی قرار گرفتند (جدول ۲). اگر چه اطلاعاتی وجود ندارد که مقادیر واقعی این پارامترها در جمعیت چقدر هستند تا درستی پیش‌بینی این مدل‌ها در داده‌های واقعی بحث شود ولی مقایسه نسبی بین این پیشین‌ها امکان‌پذیر است. در رابطه با مقدار پارامترها، Legarra *et al.* (2008) پتانسیل انتخاب ژنومی را در برخی صفات مورفولوژیکی موش با استفاده از مدل‌هایی که فقط شامل اثرات نشانگرها و اثر نشانگر و پلی‌ژنی باهم بودند، ارزیابی کردند. آنها گزارش کردند که واریانس ژنتیکی ناشی از نشانگرها برای وزن بدن در ۶ هفتهگی با استفاده از دو مدل ذکر شده در حدود ۰/۰۴ و ۰/۰۳ است. همین مؤلفه در رابطه با طول بدن به ترتیب ۰/۰۹ و ۰/۰۸ گزارش شده است که هر دو به طور قابل توجهی از مطالعه‌ی حاضر پایین‌تر هستند. قابل ذکر است که مدل آماری تجزیه و تحلیل و اثرات کمکی داخل مدل در این دو مطالعه متفاوت بود. همچنین داده‌های مورد تجزیه و تحلیل در این مطالعه بخشی ویرایش شده از داده‌های اولیه مورد استفاده در مطالعه Legarra *et al.* (2008) بود. برخلاف صفات واقعی که از مقدار حقیقی این مؤلفه‌ها اطلاعاتی وجود ندارد، صفات شبیه‌سازی شده این امکان را می‌دهند تا قابلیت پیش‌بینی مقادیر واقعی این پارامترها با استفاده از پیشین‌های مختلف مشخص گردد. نتایج ارائه شده در شکل‌های ۳ و ۴ تأیید کردند که در داده‌های شبیه‌سازی شده نیز اختلاف بین سه پیشین به وضوح مشاهده می‌شود. همچنین در این بخش برخلاف داده‌های واقعی که مقایسه با مقدار حقیقی ممکن نبود، هر سه پیشین به طور مشترک کم‌برآوردهایی از مقادیر واقعی پارامترهای شبیه‌سازی شده را ارائه کردند و بنابراین در این معیار تفاوتی بین مدل‌ها دیده نشد. ولی در رابطه با میزان انحراف از مقدار واقعی برتری نسبی بین پیشین‌ها دیده شد و اینکه هر چه قدر صفات تحت تأثیر تعداد کمتری QTL و وراثت‌پذیری پایین باشند، انتخاب پیشین اهمیت بیشتری می‌یابد. این موارد از یک طرف ضرورت انتخاب پیشین مناسب برای صفاتی با این معماری ژنتیکی را نشان می‌دهد و از طرف دیگر پتانسیل بالقوه انتخاب متغیر بیز LASSO را نشان می‌دهد. احتمالاً صفاتی که از چنین معماری ژنتیکی پیروی نمی‌کنند از بیز LASSO بهره‌ی زیادی نمی‌برند. البته هدف این مطالعه مقایسه پیش‌بینی روش‌های دارای پتانسیل انتخاب متغیر مانند بیز LASSO در مقابل دیگر روش‌های رایج در انتخاب ژنومی نبود. ولی مطالعاتی وجود دارد که به بررسی این مسئله پرداخته‌اند و می‌توانند این استنتاج را تأیید کنند (Colombani *et al.*, 2013, Hosseini-Vardanjani *et al.*, 2019, Azevedo *et al.*, 2015).

برتری نسبی پیشین‌ها ظاهراً تا حد زیادی به معماری ژنتیکی صفت وابسته است. به طور مثال تفاوت بین سه پیشین مختلف در رابطه با صفت پروتئین کل خون نسبت به دو صفت دیگر مشهودتر است. این تفاوت در داده‌های شبیه‌سازی شده نیز دیده می‌شود و برای صفت چهارم مشهودتر است. دلیل این تفاوت را می‌توان این گونه بیان کرد. می‌دانیم که انتخاب ژنومی از عدم تعادل پیوستگی بین نشانگرها و QTL‌ها برای پیش‌بینی استفاده می‌کند. بنابراین در صفاتی که تحت تأثیر تعداد کمی ژن قرار دارند پراکنش نشانگرها به شدت باعث تشریح مقادیر متفاوت واریانس توسط هر نشانگر می‌شود. از طرفی قابلیت بیز LASSO در مقایسه با برخی روش‌های بیزی دیگر، توانایی انتخاب متغیر است. در LASSO، حل مسئله بهینه‌سازی محدود شده‌ی $\min_{\beta} \{ \sum (y_i - x_i \beta)^2 \}$ منوط به $\sum |\beta_j| \leq t$ برآوردها را ارائه می‌دهد که x_i همان بردار نشانگرها، β بردار ضرایب رگرسیون و t یک ثابت مثبت دلخواه است. اگر $\lambda(t) \geq 0$ در نظر گرفته شود، این مسئله بهینه‌سازی معادل عبارت $\min_{\beta} \{ \sum (y_i - x_i \beta)^2 + \lambda(t) \sum |\beta_j| \}$ است که آشکارا منجر به صفر شدن برخی عناصر β خواهد شد. اکنون چنانچه صفتی تحت تأثیر تعداد کمی ژن باشد انتخاب پیشین اهمیت می‌یابد. واضح است که انتخاب یک مقدار بالا برای پارامتر تنظیم λ منجر به صفر شدن تعداد بیشتری از ضرایب نشانگرها می‌شود. قابلیت انتخاب متغیر بیز LASSO رویکردی ارزشمند برای تجزیه و تحلیل داده‌های ژنومی است که با کاهش مجموعه متغیرهای پیش‌بینی کننده و حفظ SNP‌های مهم مؤثر بر فنوتیپ‌ها قابلیت پیش‌بینی‌ها را بهبود می‌بخشد (Gianola and Fernando, 2020).

از دیدگاه آمار بیزی، بیز LASSO شامل یک احتمال درست‌نمایی $p(y|\mu, \beta, \sigma_\varepsilon^2) = \prod_{i=1}^n N(y_i|\mu + \sum_{j=1}^p x_{ij} \beta_j, \sigma_\varepsilon^2)$ و یک پیشین از β حاصل ضرب p توزیع نمایی مستقل با میانگین صفر $p(\beta|\lambda) = \prod_{j=1}^p (\lambda/2) \exp(-\lambda|\beta_j|)$ است. توزیع نمایی مضاعف نسبت به توزیع نرمال تراکم بیشتری حول صفر دارد و دم‌های ضخیم‌تری دارد. این یعنی امکان انقباض قوی‌تری برای ضرایب نشانگر که نزدیک به صفر هستند (که احتمالاً از QTL‌های مؤثر بر صفت بسیار دور هستند) فراهم می‌شود و انقباض کمتری نسبت به ضرایب بزرگتر ایجاد می‌کند. بنابراین پارامتر تنظیم λ نقش اصلی را دارد و با نزدیک شدن به صفر، به کمترین مربعات معمولی تمایل دارد، در حالی که مقادیر بزرگ λ نرم L_1 ضرایب β ، $\sum_j \beta_j$ را به شدت جریمه می‌کند. در واقع، λ پیشین β را کنترل می‌کند، که مقادیر بزرگتر این پارامتر پیشین‌های اطلاع دهنده‌تر هستند. این قابلیت مدل LASSO در مورد صفات تحت تأثیر تعداد زیادی ژن احتمالاً از دست خواهد رفت. از نظر تئوری، انتظار می‌رود که هر زمان که ناهمگونی قابل ملاحظه‌ای در مورد اثرات نشانگرها در تراشه‌ی استفاده شده وجود داشته باشد، پیشین‌های مختلف تفاوت بارزتری را نشان دهند در حالی که زمانی که تمام نشانگرها اثرات کوچک و مشابهی دارند، یکسان رفتار کنند. زمانی که تعداد زیادی ژن صفت را تحت تأثیر قرار می‌دهند و طبعاً اثر هر کدام نیز کوچک است احتمالاً اختصاص یک واریانس برابر به تمام نشانگرها (به طور مثال مانند بیز ریج) اطلاع دهنده‌تر باشد. در بیز ریج درست‌نمایی همان است که در بیز LASSO وجود دارد ولی پیشین β حاصل ضرب p چگالی نرمال مستقل است، یعنی $p(\beta_j, \sigma_{\beta_j}^2) = \prod_{j=1}^p N(\beta_j|0, \sigma_{\beta_j}^2)$ که در آن $\sigma_{\beta_j}^2$ یک پارامتر واریانس مشترک برای تمام ضرایب نشانگرها است و یک $\chi^2(\sigma_{\beta_j}^2|df_{\beta}, S_{\beta})$ به آن اختصاص داده می‌شود. این یعنی هیچ برتری بین نشانگرها وجود ندارد و در چنین حالتی روش‌های انتخاب متغیر عملاً هیچ منفعتی نسبت به روش‌های دیگر بدون انتخاب متغیر نخواهند داشت.

بررسی مقادیر کمینه و بیشینه‌ی مدل‌ها علاوه بر مقایسه میانگین صحت مدل‌ها نیز جالب به نظر می‌رسد. اگر چه در این مطالعه بر اساس یک اعتبارسنجی متقابل پنج طرفه میانگین صحت ارزیابی‌ها نسبتاً مشابه بود اما وقتی مقادیر کمینه و بیشینه‌ی پارامترها را بررسی می‌کنیم، در می‌یابیم که پتانسل این سه پیشین در ارزیابی‌های ژنومی به طور قابل توجهی متفاوت است. تجلی این نکته در مقادیر انحراف استاندارد و سایر آماره‌های توصیفی گزارش شده مشهود است (این نکته در مورد برآورد مؤلفه‌های واریانس نیز صادق بود). اختصاص پیشین گاما بدون استثنا منجر به ارائه مقادیر صحتی گردید که در بین جمعیت‌های آزمون مختلف نسبتاً پایدار بودند. در مقابل اختصاص مقدار ثابت به λ منجر به ناپایداری مقادیر صحت در بین جمعیت‌های آزمون مختلف می‌شود. این نکته از یک طرف اثر گزینش پیشین مناسب برای پارامتر تنظیم λ و از طرف دیگر لزوم ارزیابی‌هایی چندطرفه بالاتر را بارزتر می‌کند. اگر چه در مطالعات انتخاب ژنومی واقعی تعیین جمعیت ارزیابی انتخابی نیست. به طور مثال معمول است که فرزندان گاوهای نر به عنوان جمعیت آزمون و گاوهای نر آزمون شده به عنوان جمعیت مرجع محسوب می‌شوند.

مقایسه‌ی یافته‌های این تحقیق با تحقیقات دیگر از جنبه‌های آماری و قابلیت پیش‌بینی‌ها مدل‌ها، مطابقت با نتایج به دست آمده توسط Pérez *et al.* (2010) وجود دارد. آنها گزارش کردند که تجزیه و تحلیل کاملاً بیزی با استفاده از میانگین پسین λ به همان خوبی λ به دست آمده از یک اعتبارسنجی متقابل و جستجوی شبکه‌ای عمل می‌کند. زیرا این میانگین پسین در مجاورت مقادیر بهینه یافت شده از جستجوی شبکه‌ای قرار داشت. Cleveland *et al.* (2010) نیز در رابطه با توانایی پیش‌بینی بیز لاسو با استفاده از یک مجموعه داده شبیه‌سازی شده نشان دادند که بیز LASSO با λ تصادفی، مقادیر صحت بالا (تقریباً ۰/۹۲) را زمانی که تمام افراد در تجزیه و تحلیل استفاده شدند، تولید کرد. از سوی دیگر، Legarra *et al.* (2010) از بیز لاسو با λ تصادفی در تجزیه و تحلیل داده‌های واقعی گاوهای نر هولشتاین استفاده کردند و مقادیر صحت را از ۰/۳ برای عملکرد پروتئین تا ۰/۷۳ برای درصد چربی گزارش کردند. کلیت این گزارش‌ها نشان می‌دهند که مقادیر صحت در صفات واقعی همواره کمتر از صفات شبیه‌سازی شده هستند که با نتایج ارائه شده در مطالعه حاضر در مورد صفات واقعی و شبیه‌سازی شده، مطابقت دارد.

نتیجه‌گیری و پیشنهاد

نتایج مطالعه حاضر نشان داد که انتخاب مقادیر ثابت برای λ منجر به افزایش میانگین صحت پیش‌بینی‌ها در برخی صفات که به طور معمول تحت تأثیر تعداد کمی ژن هستند می‌شود ولی همواره مقادیر بیشینه با استفاده از پیشین‌های دیگر بدست آمدند.

همچنین دامنه‌ی صحت پیش‌بینی‌ها نشان داد که انتخاب یک λ ثابت منجر به ناپایداری پارامترها بین گروه‌های مورد ارزیابی می‌شود. از طرفی انتخاب یک مقدار ثابت مناسب نیاز به برخی معلومات بیولوژیکی دارد که ممکن است کمتر در دسترس باشند چرا که دانش مبانی ژنتیکی صفات پیچیده محدود است و به اندازه کافی گسترده نیست که بتواند نسخه‌های پیشین را برای هر صفت یا موقعیت خاصی تهیه کند. همچنین تعداد، مکان، اثرات جهش‌های علی، ساختار عدم تعادل پیوستگی و نحوه عملکرد QTLها تا حد زیادی ناشناخته است. همچنین اجرای یک جستجوی شبکه‌ای برای یافتن مقدار λ هزینه‌های محاسباتی بالایی را درخواست می‌کند زیرا برای هر مقدار λ ثابت یک اجرای نمونه‌گیر گیبس نیاز است، در حالی که تصادفی تلقی کردن λ به استفاده از یک نمونه‌گیر گیبس نیاز دارد که سریع‌تر است و نشان می‌دهد تصادفی تلقی کردن λ جذاب‌تر است. در صورت تصادفی تلقی کردن λ اطلاعات حاصل از داده‌ها و پیشین نسبتاً اطلاع دهنده اختصاص یافته می‌تواند مقادیر پایدارتری از پیش‌بینی‌ها را برای گروه‌های مختلف جمعیت فراهم نماید. نتایج حاصل از این مطالعه باید به سایر مطالعات قابل بسط دادن باشد در صورتی که ویژگی‌های صفات مورد بررسی منطبق بر صفات مطالعه حاضر باشند. این نتایج مستقیماً از ساختار روش ناشی می‌شوند بدون اینکه لزوماً به یک مسئله استنتاجی یا پیش‌بینی واقعی مرتبط باشد. در صورتی که صفات دیگری مورد مطالعه باشند، برای تعیین نظری قابلیت پیش‌بینی این پیشین‌ها، مطالعات بیشتری با استفاده از شبیه‌سازی‌های رایانه‌ای را می‌توان به راحتی برای ایجاد سناریوهای مرتبط تنظیم کرد.

تعارض منافع

هیچ گونه تعارض منافع بین نویسندگان وجود ندارد.

منابع

حسینی وردنجانی، س. م. شریعتی، م. م. مرادی شهربابک، ح و طهمورث پور، م (۱۳۹۸). صحت پیش‌بینی‌های ژنومی برای صفات مرتبط با شیر در گاو نژاد نجدی. نشریه علوم دامی، ۱۲۲، ۱۰۴-۹۳. شناسه دیجیتال: 10.22092/asj.2018.115893.1549

REFERENCES

- Alemu, S. W., Lopdell, T. J., Trevarton, A. J., Snell, R. G., Littlejohn, M. D., & Garrick, D. (2025). Comparison of genomic prediction accuracies in dairy cattle lactation traits using five classes of functional variants versus generic SNP. *Genetics Selection Evolution*, 57(1), 20. <https://doi.org/10.1186/s12711-025-00966-2>.
- Azevedo, C. F., de Resende, M. D. V., e Silva, F. F., Viana, J. M. S., Valente, M. S. F., Resende, M. F. R., & Muñoz, P. (2015). Ridge, Lasso and Bayesian additive-dominance genomic models. *BMC genetics*, 16, 1-13. <https://doi.org/10.1186/s12863-015-0264-2>.
- Calus, M. P., Huang, H., Vereijken, A., Visscher, J., Ten Napel, J., & Windig, J. (2014). Genomic prediction based on data from three layer lines: a comparison between linear methods. *Genetics Selection Evolution*, 46, 1-13. <https://doi.org/10.1186/s12711-014-0057-5>.
- Cleveland, M. A., Forni, S., Deeb, N., & Maltecca, C. (2010). Genomic breeding value prediction using three Bayesian methods and application to reduced density marker panels. Paper presented at the BMC proceedings. <https://doi.org/10.1186/1753-6561-4-S1-S6>.
- Colombani, C., Legarra, A., Fritz, S., Guillaume, F., Croiseau, P., Ducrocq, V., & Robert-Granié, C. (2013). Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesC π methods for genomic selection in French Holstein and Montbéliarde breeds. *Journal of Dairy Science*, 96(1), 575-591. <https://doi.org/10.3168/jds.2011-5225>.
- de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, 182(1), 375-385. <https://doi.org/10.1534/genetics.109.101501>.
- Gianola, D., & Fernando, R. (2020). A multiple-trait Bayesian Lasso for genome-enabled analysis and prediction of complex traits. *Genetics*, 214(2), 305-331. <https://doi.org/10.25386/genetics.10689380>.
- Gianola, D. (2013). Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics*, 194(3), 573-596. <https://doi.org/10.1534/genetics.113.151753>
- Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC bioinformatics*, 12, 1-12. <https://doi.org/10.1186/1471-2105-12-186>.
- Hosseini-Vardanjani, S. M., Shariati, M. M., Moradi Shahrehabak, H., & Tahmoorespur, M. (2019). The accuracy of genomic predictions for milk related traits in Najdi cattle breed. *Research Journal of Livestock Science*, 32(122), 93-104. DOI:10.22092/asj.2018.115893.1549. (In Persian).
- Legarra, A., Robert-Granié, C., Croiseau, P., Guillaume, F., Fritz, S., & Ducrocq, V. (2010). Aptitude of Bayesian lasso for genomic selection. Paper presented at the 9. World Congress on Genetics Applied to Livestock Production. Leipzig, DEU (2010-08-01-2010-08-06).
- Legarra, A., Robert-Granié, C., Manfredi, E., & Elsen, J. (2008). Performance of genomic selection in mice. *Genetics*, 180(1), 611-618. <https://doi.org/10.1534/genetics.108.088575>.
- Meuwissen, T. H., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819-1829. <https://doi.org/10.1093/genetics/157.4.1819>.
- Nejati-Javaremi, A., Smith, C., & Gibson, J. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of animal science*, 75(7), 1738-1745. <https://doi.org/10.2527/1997.7571738x>.
- Park, T., & Casella, G. (2008). The bayesian lasso. *Journal of the american statistical association*, 103(482), 681-686. <https://doi.org/10.1198/016214508000000337>.
- Pérez, P., de Los Campos, G., Crossa, J., & Gianola, D. J. T. p. g. (2010). Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. *The plant genome*, 3(2). <https://doi.org/10.3835/plantgenome2010.04.0005>.
- Pérez, P., & de Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198(2), 483-495. <https://doi.org/10.1534/genetics.114.164442/>

/DC1.

- Ros-Freixedes, R., Johnsson, M., Whalen, A., Chen, C.-Y., Valente, B. D., Herring, W. O. Hickey, J. (2022). Genomic prediction with whole-genome sequence data in intensely selected pig lines. *Genetics Selection Evolution*, 54(1), 65. <https://doi.org/10.1186/s12711-022-00756-0>.
- Solberg, L. C., Valdar, W., Gauguier, D., Nunez, G., Taylor, A., Burnett, S. Burns, P. (2006). A protocol for high-throughput phenotyping, suitable for quantitative trait analysis in mice. *Mammalian Genome*, 17, 129-146. <https://doi.org/10.1007/s00335-005-0112-1>.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58(1), 267-288. <https://www.jstor.org/stable/2346178>.
- Valdar, W., Solberg, L. C., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W. O. Flint, J. (2006). Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature genetics*, 38(8), 879-887. <https://doi.org/10.1038/ng1840>.
- Valdar, W., Solberg, L. C., Gauguier, D., Cookson, W. O., Rawlins, J. N. P., Mott, R., & Flint, J. (2006). Genetic and environmental effects on complex traits in mice. *Genetics*, 174(2), 959-984. <https://doi.org/10.1534/genetics.106.060004>.
- Yuan, C., Gillon, A., Gualdrón Duarte, J. L., Takeda, H., Coppieters, W., Georges, M., & Druet, T. (2025). Evaluation of genomic selection models using whole genome sequence data and functional annotation in Belgian Blue cattle. *Genetics Selection Evolution*, 57(1), 10. <https://doi.org/10.1186/s12711-025-00955-5>.