



رشد ژنتیکی مورد انتظار و همبستگی بین ارزش‌های اصلاحی بر اساس مقیاس‌های مختلف ماتریس روابط خویشاوندی ژنومی

سید مهدی حسینی وردنجانی^۱، محمد مهدی شریعتی^{۲*}

۱. دانشجوی دکتری ژنتیک و اصلاح نژاد دام دانشگاه فردوسی مشهد

۲. هیئت علمی گروه علوم دامی دانشگاه فردوسی مشهد

ایمیل نویسنده مسئول: shariati52@gmail.com

چکیده

بهترین پیش بینی نا اریب خطی ژنومی به عنوان یک روش آماری از روابط خویشاوندی محاسبه شده از نشانگرهای تک نوکلئوتیدی جهت توضیح کواریانس ژنتیکی بین حیوانات در جایگاه‌های صفات کمی استفاده می‌کند. هدف مطالعه حاضر بررسی پتانسیل مقیاس‌های مختلف ماتریس روابط خویشاوندی ژنومی (G) در افزایش رشد ژنتیکی و همبستگی بین ارزش‌های اصلاحی واقعی (TBV) و پیش بینی شده (GBV) بود. شش نسبت انتخاب مختلف ۵، ۱۰، ۲۰، ۴۰، ۸۰ و ۱۰۰ درصد بررسی و نتایج نشان داد که با افزایش نسبت انتخاب همبستگی پیرسون و رتبه‌ای بین TBV و GBV افزایش می‌یابد. پایین‌ترین همبستگی پیرسون و رتبه‌ای (به ترتیب ۰/۲۶ و ۰/۲۴) با فراوانی آلی ۰/۵ به عنوان مقیاس ماتریس G در نسبت انتخاب ۵٪ و بالاترین آنها (به ترتیب ۰/۷۳ و ۰/۷۲) در نسبت انتخاب ۱۰۰٪ با استفاده از ماتریس روابط خویشاوندی شجره‌ای (A) برای مقیاس کردن ماتریس G بدست آمد. تفاوت‌ها، در یک نسبت انتخاب برابر بین استراتژی‌های مختلف مقیاس کوچک بودند. استفاده از ماتریس A برای مقیاس کردن ماتریس G در مقایسه با سایر روش‌های مقیاس رشد ژنتیکی را در تمامی سطوح نسبت انتخاب افزایش داد ولی پایین‌ترین رشد ژنتیکی بسته به سطح نسبت انتخاب، مختلف بود. انتظار می‌رود که با استفاده از ماتریس A برای مقیاس کردن ماتریس G رشد ژنتیکی و همبستگی‌ها افزایش یابد اگر چه که ممکن است تفاوت‌ها کوچک باشند.

کلمات کلیدی: شدت انتخاب- همبستگی رتبه‌ای- فراوانی آلی- ارزش اصلاحی ژنومی

مقدمه

بهترین پیش بینی نا اریب خطی ژنومی (GBLUP) از یک ماتریس روابط خویشاوندی ژنومی که در آن خویشاوندی بین حیوانات بر مبنای ژنوتیپ‌های چند شکلی‌های تک نوکلئوتیدی (SNPs) محاسبه می‌شود [۶-۷] به جای استفاده از ماتریس روابط خویشاوندی شجره‌ای [۴] استفاده می‌کند. در این مدل صحت به صورت همبستگی بین ارزش‌های اصلاحی واقعی و پیش بینی شده تعریف می‌شود و به همراه شدت انتخاب که مربوط به نسبت افرادی است که به عنوان والدین نسل بعد انتخاب می‌شوند از مهمترین عوامل در تعیین میزان بهبود ژنتیکی است [۲].



تشکیل ماتریس واریانس کواریانس ژنومی بین حیوانات و نوع مقیاس کردن آن در مقایسه با ماتریس روابط خویشاوندی مبتنی بر شجره که نسبت به یک جامعه پایه بدون انتخاب با میانگین ارزش اصلاحی صفر بود بسیار حایز اهمیت است. از لحاظ تئوری برای تشکیل ماتریس روابط خویشاوندی ژنومی باید از فراوانی آلی حیوانات بنیانگذار در شجره استفاده شود [۷] ولی حیوانات پایه به ندرت تعیین ژنوتیپ می‌شوند و عملاً چنین فراوانی آلی در دست نخواهد بود. بنابراین نیاز به روش‌های جایگزین برای تشکیل این ماتریس است. هدف مطالعه حاضر نیز بررسی استفاده از مقیاس‌های مختلف برای تشکیل ماتریس روابط خویشاوندی ژنومی و تأثیر آنها بر رشد ژنتیکی و همبستگی بین ارزش‌های اصلاحی واقعی و پیش بینی شده در نسبت‌های مختلف انتخاب بود.

مواد و روش‌ها

صفتی با وراثت پذیری ۰/۲۵ بر روی ژنومی متشکل از سه کروموزوم، هر یک به طول ۱۰۰ سانتی مورگان شبیه سازی شد. تعداد ۳۰۰۰ نشانگر تک نوکلئوتیدی و ۱۰۵ QTL چند آلی با موقعیت تصادفی بر روی این ژنوم توزیع شد. اثرات QTLها از یک توزیع گاما با پارامتر شکل ۰/۴ و پارامتر نرخ ۱/۶ نمونه گیری شدند. نرخ جهش برای نشانگرها 2.5×10^{-3} و برای QTLها 2.5×10^{-5} در نظر گرفته شد. برای ایجاد تعادل جهش-رانس ژنتیکی، جمعیتی متشکل از ۴۰۰ حیوان ماده و ۲۰ حیوان نر برای ۱۰۰ نسل آمیزش تصادفی داده شدند که در طی آن به ۱۰۰۰ حیوان افزایش یافتند. آمیزش تصادفی برای ۴۰۰ نسل دیگر ادامه یافت و تعداد حیوانات نر در نسل آخر به ۷۰ رسانده شد. به طور تصادفی ۴۵۵ حیوان ماده و ۳۵ حیوان نر از نسل ۵۰۰ به عنوان حیوانات پایه گذار (نسل صفر) انتخاب شدند. انتخاب حیوانات از نسل صفر به بعد بر اساس ارزش‌های اصلاحی آنها بود به طوری که حیوانات مورد نیاز برای نسل بعدی از آمیزش حیوانات دارای بالاترین ارزش اصلاحی در نسل قبل انتخاب می‌شدند. شبیه سازی برای ۹ نسل دیگر ادامه یافت. شجره برای تمامی افراد و نشانگرها و QTLهای شبیه سازی شده برای تمام افراد نسل صفر، ۸ و ۹ ذخیره شدند. برای هر دو جنس نر و ماده نیز یک ارزش فنوتیپی با اضافه کردن یک انحراف محیطی دارای توزیع نرمال با میانگین صفر و واریانس یک به مجموع ارزش QTLهای آن فرد شبیه سازی شد. حیوانات نسل ۸ و ۹ به عنوان جامعه آزمون استفاده شدند. شبیه سازی ژنوم و جمعیت ۱۰ بار تکرار شد.

برای آنالیز داده‌ها از مدل آماری $y = Xb + Zg + e$ استفاده شد که در آن بردار ارزش‌های فنوتیپی، b بردار اثر ثابت جنس، g بردار ارزش‌های اصلاحی ژنومی و e نیز بردار باقیمانده هستند. ماتریس‌های X و Z نیز به ترتیب رکوردها را به اثرات ثابت و حیوانات مرتبط می‌نمایند. مؤلفه‌های واریانس و سایر پارامترهای مدل به روش بیز و با استفاده از نمونه‌گیری گیبس در محیط نرم افزار BGLR به دست آمدند. به این صورت که برای اثرات ثابت یک توزیع یکنواخت در نظر گرفته شد. به ارزش‌های اصلاحی و واریانس ژنتیکی به ترتیب یک توزیع نرمال و یک توزیع کای اسکویر معکوس با مشخصات $p(g, \sigma_g^2) = N(g|0, G\sigma_g^2)\chi^{-2}(\sigma_g^2|S_g, df_g)$ نیز ماتریس روابط خویشاوندی ژنومی است. توزیع پیشین بردار باقیمانده به صورت $p(e|\sigma_e^2) = N(e|0, I\sigma_e^2)$ بود که واریانس باقیمانده نیز از یک توزیع کای اسکویر معکوس با مشخصات $p(\sigma_e^2) = \chi^{-2}(\sigma_e^2|S_e, df_e)$ نمونه‌گیری می‌شد. در هر اجرا از یک زنجیره به طول ۱۵۰۰۰ نمونه استفاده شد که ۱۵۰۰ نمونه اول به عنوان دوره سوخته حذف می‌شدند.

برای تشکیل ماتریس G از روش ارائه شده توسط VanRaden (۲۰۰۸) به صورت $G = \frac{ZZ^T}{2\sum p_i(1-p_i)}$ استفاده شد. در این رابطه p_i فراوانی آلل دوم در هر جایگاه و $Z = M - P$ است که M ماتریس ژنوتیپ‌های افراد با کدهای ۰، ۱ و ۲ بر اساس تعداد کپی آلل



دوم و P ماتریس حاوی دو برابر فراوانی آلل دوم ($2p_i$) در هر جایگاه است. مقیاس‌های مختلف برای ماتریس روابط خویشاوندی ژنومی با استفاده از p_i ‌های مختلف بدست آمد. از سه فراوانی آللی مشاهده شده در حیوانات جمعیت پایه گذار، جمعیت مرجع و به طور فرضی ۰/۵ به ترتیب برای ساختن GOF ، GOR و $G05$ استفاده شد. ماتریس $GNORM$ نیز با استفاده از فراوانی آللی مشاهده شده در جمعیت مرجع ولی با این تفاوت که مخرج با $\frac{trase(ZZ^T)}{n}$ جایگزین شد نیز بدست آمد [۳]. مقیاس پنجم نیز با استفاده از معادله $GA = \beta GOR + \alpha$ بدست آمد که ضرایب α و β با استفاده از معادلات زیر بدست آمدند [۱]:

$$\text{mean}(\text{diag}(\mathbf{G}))\beta + \alpha = \text{mean}(\text{diag}(\mathbf{A}))$$

$$\text{mean}(\mathbf{G})\beta + \alpha = \text{mean}(\mathbf{A})$$

نتایج و بحث

جدول یک معیارهای مختلفی که معمولاً برای مقایسه مدل‌ها به کار گرفته می‌شوند را نشان می‌دهد. بر مبنای معیار DIC بهترین مدل، مدلی است که برای تشریح کواریانس ژنتیکی بین حیوانات از GA استفاده می‌کند. بر اساس معیار pD این مدل تعداد پارامتر بیشتری نیز داشته و بنابراین پیچیدگی بالاتری دارد.

جدول ۱- معیارهای اندازه‌گیری نکویی برآزش و پیچیدگی مدل‌های مختلف

Criteria	<i>GOF</i>	<i>GOR</i>	<i>G05</i>	<i>GNORM</i>	<i>GA</i>
logLikAtPostMean	-۱۱۸۰/۷۳	-۱۱۸۲/۲۹	-۱۱۷۹	-۱۱۸۱/۲۶	-۱۱۷۳/۹۴
postMeanLogLik	-۱۲۲۰/۵۹	-۱۲۲۱/۴۱	-۱۲۱۹/۷۳	-۱۲۲۰/۸۲	-۱۲۱۶/۴۵
pD	۷۹/۷۱۸۶	۷۸/۲۴۲۲۱	۸۱/۴۶۹۸۸	۷۹/۱۱۸۴	۸۵/۰۳۶۴۷
DIC	۵۲۲۰/۹۰۶	۲۵۲۱/۰۵۵	۲۵۲۰/۹۳۴	۲۵۲۰/۷۶۴	۲۵۱۷/۹۴۶

مقادیر همبستگی پیرسون و رتبه‌ای میان ارزش‌های اصلاحی واقعی و پیش بینی شده با استفاده از G ‌های مختلف به ترتیب در جدول ۲ و ۳ آورده شده‌اند. دامنه‌ی همبستگی پیرسون بین ۰/۷۳-۰/۲۶ متغیر است. همواره پایین‌ترین ضرایب همبستگی در نسبت انتخاب بالاتر بدست آمدند به طور مثال زمانی که شدت انتخاب ۲/۰۶، معادل انتخاب ۵٪ بهترین حیوانات است ضریب همبستگی در مدل‌های مختلف از ۰/۲۹۳ بالاتر نیست. این در حالی است که در شدت انتخاب صفر (معادل انتخاب تمامی حیوانات) این ضریب به ۰/۷۳ افزایش می‌یابد. روند کاهشی ضرایب همبستگی رتبه‌ای با افزایش نسبت انتخاب همانند روند همبستگی پیرسون بود ولی به لحاظ عددی از متناظرهای خود کوچکتر بودند.

جدول ۲- همبستگی پیرسون بین ارزش‌های اصلاحی واقعی و پیش بینی شده با ماتریس‌های مختلف روابط خویشاوندی ژنومی

نسبت انتخاب	<i>GOF</i>	<i>GOR</i>	<i>G05</i>	<i>GNORM</i>	<i>GA</i>
۵٪	۰/۲۸۴±۰/۰۶	۰/۲۹۰±۰/۰۶	۰/۲۶۴±۰/۰۶	۰/۲۷۶±۰/۰۶	۰/۲۹۳±۰/۰۶
۱۰٪	۰/۳۶۱±۰/۰۵	۰/۳۶۰±۰/۰۵	۰/۳۶۶±۰/۰۶	۰/۳۵۵±۰/۰۶	۰/۳۵۳±۰/۰۶
۲۰٪	۰/۴۸۳±۰/۰۴	۰/۴۸۸±۰/۰۴	۰/۴۹۰±۰/۰۴	۰/۴۸۴±۰/۰۴	۰/۴۸۳±۰/۰۴
۴۰٪	۰/۵۴۰±۰/۰۴	۰/۵۴۱±۰/۰۴	۰/۵۴۰±۰/۰۴	۰/۵۳۵±۰/۰۴	۰/۵۴۵±۰/۰۴
۸۰٪	۰/۶۲۸±۰/۰۳	۰/۶۳۷±۰/۰۳	۰/۶۳۶±۰/۰۳	۰/۶۳۳±۰/۰۳	۰/۶۴۶±۰/۰۳
۱۰۰٪	۰/۷۱۸±۰/۰۳	۰/۷۱۷±۰/۰۳	۰/۷۱۹±۰/۰۳	۰/۷۱۷±۰/۰۳	۰/۷۳۰±۰/۰۳

± اشتباه استاندارد



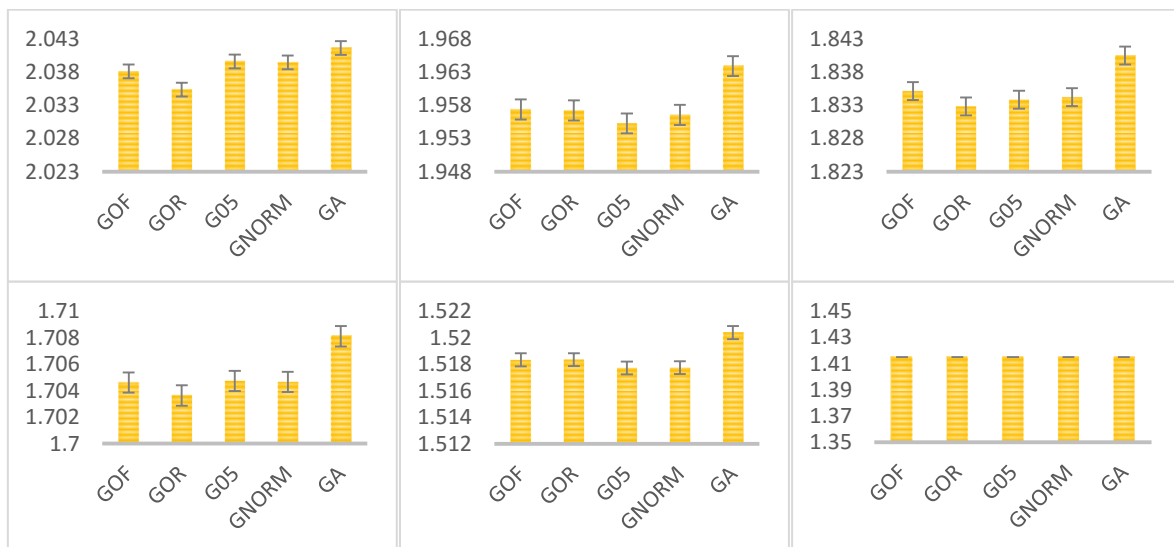
جدول ۳- همبستگی رتبه‌های بین ارزش‌های اصلاحی واقعی و پیش بینی شده با ماتریس‌های مختلف روابط خویشاوندی ژنومی

نسبت انتخاب	GOF	GOR	G05	GNORM	GA
۰.۵٪	۰.۲۷	۰.۲۸	۰.۲۴	۰.۲۵	۰.۲۷
۱.۰٪	۰.۳۶	۰.۳۶	۰.۳۷	۰.۳۵	۰.۳۵
۲.۰٪	۰.۴۹	۰.۴۹	۰.۵۰	۰.۴۹	۰.۴۸
۴.۰٪	۰.۵۲	۰.۵۲	۰.۵۲	۰.۵۱	۰.۵۲
۸.۰٪	۰.۶۱	۰.۶۲	۰.۶۲	۰.۶۲	۰.۶۳
۱۰.۰٪	۰.۷۱	۰.۷۱	۰.۷۱	۰.۷۱	۰.۷۲

تمامی همبستگی‌ها در سطح $P < 0.0001$ معنی دار هستند.

رشد ژنتیکی مورد انتظار با استفاده از **G**‌های مختلف به صورت میانگین ارزش‌های اصلاحی واقعی حیوانات انتخاب شده بر اساس ارزش‌های اصلاحی ژنومی پیش بینی شده با هر یک از مدل‌ها، تعریف و نتایج در شکل یک خلاصه شدند. از لحاظ تئوری با افزایش تعداد افراد انتخاب شده رشد ژنتیکی کمتر خواهد شد. همانطور که در شکل دیده می‌شود با انتخاب ۰.۵٪ حیوانات برتر رشد ژنتیکی در تمام مدل‌ها از ۲/۰۳ بالاتر است ولی با افزایش حیوانات این رشد کمتر شده و در سطح انتخاب تمامی حیوانات به کمتر از ۱/۴۲ کاهش می‌یابد که برای تمامی مدل‌ها هم برابر است. بدون استثنا در نسبت‌های مختلف انتخاب بیشترین رشد ژنتیکی با استفاده از ماتریس **GA** بدست آمد و **GOR** نیز در انتخاب ۵، ۲۰ و ۴۰ درصد حیوانات پایین‌ترین رشد ژنتیکی را نشان داد. هر چند اختلافات بسیار اندک هستند که بخشی از آن ناشی از دامنه کم تغییرات ارزش‌های اصلاحی واقعی است. در مطالعه‌ای شبیه سازی شده رشد ژنتیکی با استفاده از اطلاعات ژنومی در مقایسه با اطلاعات شجره و همچنین رشد ژنتیکی در مدل‌های مختلف با استفاده از اطلاعات ژنومی بسیار اندک بود ولی با افزایش وراثت پذیری تفاوت‌ها مشهودتر می‌شود [۵].

شکل ۱- رشد ژنتیکی مورد انتظار با استفاده از ماتریس‌های مختلف روابط خویشاوندی ژنومی در نسبت‌های مختلف انتخاب



نمودارهای بالا از سمت چپ: انتخاب ۰.۵٪، ۱.۰٪ و ۲.۰٪ حیوانات برتر؛ نمودارهای پایین از سمت چپ: انتخاب ۴.۰٪، ۸.۰٪ و ۱۰.۰٪ حیوانات برتر؛ محور **Y** میانگین ارزش‌های اصلاحی واقعی در حیوانات انتخاب شده را نشان می‌دهد.



- 1- Christensen, O.F., Madsen, P., Nielsen, B., Ostersen, T. & Su, G. (2012). Single-step methods for genomic evaluation in pigs. *Animal* 6, 1565-1571.
- 2- Dekkers, J. (2012). Application of Genomics Tools to Animal Breeding. *Current Genomics*. 13, 207-212
- 3- Forni, S., Aguilar, I. & Misztal, I. (2011). Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution* 43.
- 4- Henderson, C. (1984). *Applications of linear models in animal breeding* (University of Guelph, Guelph, ON, Canada).
- 5- Ødegård, J. & Meuwissen, T. (2015). Identity-by-descent genomic selection using selective and sparse genotyping for binary traits. *Genetics Selection Evolution* 47: 8.
- 6- 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, 1738-1745.
- 7- VanRaden, P. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science* 91, 4414-4423.

Expected genetic gain and correlation between breeding values under different scaling genomic relationship matrix

Abstract

Genomic best linear unbiased prediction (GBLUP) as a statistical method uses relationships calculated from single nucleotide polymorphisms to explain genetic covariance between animals at quantitative trait loci. The aim of present study was to evaluate the potential of different scaling Genomic relationship matrix (**G**) on increment genetic gain and correlation among true (TBV) and predicted breeding values (GBV). Six different proportion of selection as 5, 10, 20, 40, 80 and 100 percent were evaluated and results showed that Rank and Pearson correlations among TBV and GBV increased as the proportion of selection increased. Lowest Pearson and Rank correlations (0.26 and 0.24, respectively) were obtained by using allele frequency 0.5 as the scale of **G** matrix at %5 proportion of selection, and highest of those (0.73 and 0.72, respectively) by using pedigree relationship matrix (**A**) as the scale of **G** matrix at %100 proportion of selection. Differences were small between scaling strategies at the same proportion of selection. Using **A** matrix for scaling **G** matrix increased genetic gain compared to others scaling methods at all levels of proportion of selection, but lowest genetic gain depends on proportion of selection. It is expected that genetic gain and correlations increased by using **A** matrix for scaling **G** matrix although differences may be small.

Keywords: Selection intensity- Rank correlation- Allele frequency- Genomic breeding value