دانشگاه علوم کشاورزی و منابع طبیعی ساری
پژوهشهای تولیدات دامی

"مقاله پژوهشی"

تغییرات صحت ارزش‌های اصلاحی ژنومی تحت تاثیر روش‌های آماری برآورد اثر نشانگری، خصوصیات جمعیت و معماری ژنتیکی صفت

عباس عاطفی[۱]، عبدالاحد شادپرور[۲] و نوید قوی حسین‌زاده[۳]

۱- دانش‌آموخته دکتری، گروه علوم دامی، دانشگاه گیلان
۲- استاد گروه علوم دامی، دانشگاه گیلان (نویسنده مسوول: shad@guilan.ac.ir)
۳- استاد گروه علوم دامی، دانشگاه گیلان

تاریخ دریافت: ۱۳۹۸/۱۲/۷ تاریخ پذیرش: ۱۳۹۹/۸/۱۹
صفحه: ۱۰۱ تا ۱۰۸

چکیده

هدف تحقیق حاضر بررسی اثر روش برآورد اثرات نشانگرها، توزیع اثرات QTLها، تعداد QTL، اندازه مؤثر جمعیت و وراثت‌پذیری صفت بر صحت ارزیابی‌های ژنومی بود. جمعیت پایه با دو اندازه مؤثر ۱۰۰ و ۵۰۰ فرد در نظر گرفته شده و به‌وسیله نرم‌افزار QMSim شبیه‌سازی شد. برای انجام این تحقیق، ژنومی متشکل از یک کروموزوم به طول ۱۰۰ سانتی‌مورگان با تعداد ۵۰۰ نشانگر SNP در طول ژنوم و با تعداد متفاوت QTL (۵۰ و ۲۰۰) که آن‌ها نیز به‌طور تصادفی روی کروموزوم پراکنده شده‌اند شبیه‌سازی شد. مقادیر وراثت‌پذیری بررسی شده در تحقیق حاضر نماینده صفات با وراثت‌پذیری پایین (۰/۱)، متوسط (۰/۳) و بالا (۰/۵) بودند. ارزش‌های اصلاحی ژنومی به‌وسیله روش‌های رگرسیون ریج بیزی (BRR)، بیزA (BayesA)، بیز B (BayesB)، بیز C(BayesC)، بیز لزو (BayesL)، روش نیمه پارامتری بر پایه کرنل (RKHS) و شبکه‌های عصبی (NN) پیش‌بینی شدند. بر اساس نتایج به‌دست آمده، عوامل وراثت‌پذیری، اندازه مؤثر جمعیت، روش برآورد اثرات نشانگری، توزیع QTL و تعداد QTL بر صحت ارزیابی ژنومی تأثیر محسوسی داشتند. بین روش‌های برآورد اثرات نشانگری، روش‌های بیز A و B بیشترین و روش شبکه‌های عصبی کمترین صحت را داشتند. در تحقیق حاضر با افزایش وراثت‌پذیری و تعداد QTL، صحت پیش بینی‌های ژنومی افزایش و برعکس با کاهش اندازه مؤثر جمعیت، کاهش یافت. بیشترین صحت زمانی به‌دست آمد که QTLها دارای توزیع نرمال بودند.

واژه‌های کلیدی: ارزیابی ژنومی، روش پارامتری، روش ناپارامتری، معماری ژنتیکی صفت

مقدمه

با توسعه فناوری‌های توالی‌یابی DNA با سرعت بالاتر و هزینه کمتر این امکان فراهم شده است که اطلاعات صدها هزار جایگاه نشانگر تک نوکلئوتیدی به‌صورت تراشه‌های SNP توسط شرکت‌هایی مانند ایلومینیا و افیمتیریکس در دسترس محققین قرار گیرد (۱۲). با فراهم‌شدن تراشه‌های SNP با تراکم بالا و به‌منظور غلبه بر مشکلات استفاده از MAS، انتخاب ژنومی پیشنهاد شد (۱۶). فرض اساسی در انتخاب ژنومی این است که هاپلوتیپ‌ها با آلل‌های QTL مؤثر بر صفات مورد انتخاب در حالت عدم تعادل پیوستگی هستند. انتخاب ژنومی به‌عنوان شکلی از انتخاب بر اساس نشانگر است که در آن از نشانگرهای ژنتیکی که سطح کل ژنوم را پوشانده‌اند، استفاده می‌شود به‌طوری‌که همه جایگاه‌های صفات کمّی با حداقل یک نشانگر در حالت عدم تعادل پیوستگی ژنی باشند (۱۶).

روش‌های متعددی برای برآورد آثار نشانگر توسط محققین مختلف پیشنهاد شده است. روش‌های حداقل مربعات معمولی و حداکثر درست‌نمایی زمانی که تعداد اثرات نشانگرها نسبت به تعداد مشاهدات خیلی بیشتر است ممکن است منجر به چالش فزونی تعداد پارامترها (اثرات نشانگرهای متراکم، p) نسبت به تعداد مشاهدات (n) ، یا پدیده موسوم به بلای ابعاد[1] می‌شود (۷) برای مواجهه با این مشکل، روش‌های جریمه‌ای، روش‌های جریمه‌ای همراه با انتخاب متغیر، روش‌های ناپارامتری و نیمه پارامتری پیشنهاد شده است.

روش‌های مختلف بیزی از نظر نوع توزیع پیشین مفروض برای اثرات ژنومی متفاوتند. در روش رگرسیون ریج توزیع اثرات نشانگرها به‌صورت توزیع نرمال و اثر نشانگرها غیر صفر و جزئی فرض شده است. در روش Bayes A فرض پیشین به این صورت است که تعداد زیادی از جایگاه‌ها اثرات جزئی و تعداد کمی از جایگاه ها اثرات عمده دارند و توزیع شرطی در نظر گرفته شده برای اثرات نشانگری توزیع t می‌باشد. توزیع پیشین اثرات نشانگرها در Bayes B، یک توزیع ترکیبی است که در آن، تعداد زیادی از نشانگرها دارای اثر صفر و بقیه نشانگرها اثر با توزیع t دارند. همچنین روش بیز C مشابه بیز B می‌باشد با این تفاوت که به‌جای توزیع پیشین t از توزیع نرمال برای اثرات نشانگرها استفاده می‌کند. در روش Bayes LASSO توزیع اثرات نشانگرها به‌صورت توزیع نمایی دوگانه هستند.

در بررسی‌های صورت گرفته توسط داده‌های واقعی و شبیه‌سازی شده اثر ساختار جمعیت (اندازه جمعیت مرجع، اندازه جمعیت مؤثر، وراثت‌پذیری صفت، ارتباط بین جمعیت مرجع و جمعیت تأیید) و معماری ژنتیکی صفت (تعداد QTL، توزیع QTL، فراوانی آلل نادر QTL) و روش آماری برآورد اثرات نشانگری و همچنین تراکم نشانگرها بر صحت ارزیابی‌های ژنومی تأیید شده است (۲۷،۱۵،۲،۳).

در بسیاری از کشورها انتخاب ژنومی تبدیل به روش اصلی ارزیابی گاوهای شیری شده است. به‌دلیل عدم وجود جمعیت مرجع مناسب گاوهای شیری در ایران به گونه‌ای که فقط

1- Curse of dimensionality

[Downloaded from rap.sanru.ac.ir on 2026-11-06]
[DOR: 20.1001.1.22518622.1399.11.30.7.9]
[DOI: 10.52547/rap.11.30.101]

تعداد ۸۰ گاو نر پروف شده یا جوان در ایران وجود دارد (۲۳) هنوز قابلیت اجرای روش انتخاب ژنومی در کشور وجود ندارد. بنابراین بررسی جنبه‌های مختلف روش انتخاب ژنومی توسط مطالعات شبیه‌سازی ضروری می‌باشد.

هدف از تحقیق حاضر بررسی تأثیر خصوصیات جمعیت و معماری ژنتیکی صفت بر صحت ارزش‌های اصلاحی ژنومی و تعیین بهترین روش برآورد اثرات نشانگری از بین هفت روش پارامتری و نیمه پارامتری بود.

مواد و روش‌ها

جمعیت‌ها توسط نرم‌افزار QMSim (۲۲) بر پایه فرآیند پیش رونده شبیه‌سازی شدند. در تحقیق حاضر جمعیت پایه با دو اندازه مؤثر ۱۰۰ و ۵۰۰ فرد در نظر گرفته شدند (۲). به‌منظور ایجاد عدم تعادل پیوستگی بین نشانگر و QTL، بین افراد نسل پایه، ۵۰ نسل آمیزش تصادفی انجام گرفت و در نسل ۵۱ اثر جایگزینی ژن‌ها برای هر یک از QTLها با استفاده از سه نوع توزیع مختلف شامل توزیع یکنواخت، نرمال استاندارد و گاما (با پارامتر شکل=۰/۴ و پارامتر مقیاس ۱/۶۶) برآورد شد.

بعد از ۵۰ نسل آمیزش تصادفی، افراد در نسل ۵۱ با نمونه‌گیری تصادفی از والدین نسل قبل ازدیاد یافته و به تعداد ۱۰۵۰ فرد رسیدند که شامل ۵۰ فرد نر و ۱۰۰۰ فرد ماده بود و این ساختار تا نسل ۵۳ ادامه یافت. نسبت نتاج نر در هر نسل برابر ۰/۵ بود. انتخاب والدین نسل بعد به‌صورت تصادفی انجام گرفته و سیستم آمیزشی نیز از نوع آمیزش تصادفی بود. افراد نسل ۵۱ به‌عنوان جمعیت مرجع در نظر گرفته شده که دارای رکورد فنوتیپی و هم ژنوتیپی بوده و افراد نسل‌های ۵۲ تا ۵۴ نیز به‌عنوان جمعیت تأیید محسوب شدند که فقط دارای اطلاعات ژنتیکی بودند.

برای انجام این تحقیق، ژنومی متشکل از یک کروموزوم به طول ۱۰۰ سانتی‌مورگان با تعداد ۵۰۰ نشانگر SNP در طول ژنوم و با تعداد متفاوت QTL (۵۰ و ۲۰۰) که آن‌ها نیز به‌طور تصادفی روی کروموزوم پراکنده شده‌اند شبیه‌سازی شد. نشانگرها و QTLها به‌صورت دو آللی با فراوانی اولیه ۰/۵ بودند. نرخ جهش برای نشانگرها و QTLها 2.5×10^{-5} به ازای هر جایگاه ژنی در هر نسل در نظر گرفته شد.

مقادیر وراثت‌پذیری بررسی شده در تحقیق حاضر نماینده صفات با وراثت‌پذیری پایین (۰/۱)، متوسط (۰/۳) و بالا (۰/۵) بودند. QTLها تمامی واریانس ژنتیکی صفت را در بر گرفتند و ارزش اصلاحی واقعی هر فرد برابر مجموع اثرات جایگزینی آللی QTLها و تنها با فرض اثرات افزایشی به‌دست آمد. فنوت‌یپ‌ها توسط اضافه کردن باقیمانده‌ها که از توزیع نرمال با میانگین صفر نمونه‌گیری شده بودند، به TBVها به‌دست آمدند. ساختار عمومی مدل‌ها به‌شکل خطی (رابطه ۱) است.

$$y = \mu + \sum_{j=1}^{p} X_j g_j + e \qquad \text{(رابطه ۱)}$$

که y بردار رکوردهای فنوتیپی، μ میانگین کل، Σ_j نشان‌دهنده جمع برای همه نشانگرها است، g_j ضریب نشانگر j و نشان‌دهنده اثر جایگزینی آللی است، X_j ماتریس طرح برای کدهای ژنوتیپی نشانگر مربوطه و e بردار باقیمانده‌ها است.

از هفت روش آماری مختلف شامل رگرسیون ریدج بیزی[1] (BRR)، بیزA، بیزB، بیزC، بیز لزو[2]، روش نیمه پارامتری بر پایه کرنل (RKHS)[3] و شبکه‌های عصبی (NN)[4] برای برآورد اثرات نشانگرها از روی فنوتیپ جمعیت مرجع استفاده شد.

رگرسیون ریدج بیزی

در روش رگرسیون ریدج، برآورد اثرات نشانگری به یک اندازه جریمه می‌شوند و این امر ممکن است مطلوب نباشد چون بعضی نشانگرها در مکان‌هایی قرار دارند که با واریانس ژنتیکی مرتبط نیستند این در حالی است که بقیه با QTLها مرتبط هستند (۱۱) برای غلبه بر این محدودیت، روش‌هایی که شامل انتخاب متغیر و کاهش هستند (LASSO) (۲۴) یا روش‌های بیزی با استفاده از کاهش اثرات مختص هر نشانگر (BayesA , BL) پیشنهاد شد.

Bayes A

روش بیزA توزیع پیشین مشابهی برای همه واریانس‌های موقعیت‌های نشانگری در نظر می‌گیرد. توزیع احتمال کای اسکوئر معکوس مقیاس‌دار $\chi^{-2}(df, s^2)$ با درجه آزادی df و پارامتر مقیاس s^2 می‌تواند به‌عنوان توزیع پیشین در نظر گرفته شود. این یک انتخاب مطلوب است چون یک پرایور توام می‌باشد که در آن توزیع پسین در خانواده‌ای مشابه با توزیع پیشین قرار دارد. بنابراین توزیع پسین نیز یک توزیع کای اسکوئر معکوس مقیاس‌دار $\chi^{-2}\left(df + n_j, s^2 + \beta_j'\beta_j\right)$ است که n_j تعداد اثرات هاپلوتیپی در موقعیت نشانگری j می‌باشد.

Bayes B

این روش از یک پرایور استفاده می‌کند که دارای چگالی زیاد، π، در $\sigma^2_{\beta i} = 0$ بوده و یک توزیع کای‌مربع مقیاس‌دار معکوس برای $\sigma^2_{\beta i} > 0$ دارد. در این روش توزیع پیشین به این صورت است (رابطه ۲ و ۳).

$$\sigma^2_{\beta i} = 0 \text{ with probability } \pi, \qquad \text{(رابطه ۲)}$$

$$\sigma^2_{\beta i} \sim \chi^{-2}(df, s^2) \text{ with probability } (1-\pi) \qquad \text{(رابطه ۳)}$$

Bayes C

اگر در روش بیز B به‌جای چگالی t استاندارد شده از چگالی گوسی استفاده شود، روش بیز C ایجاد می‌شود.

Bayesian LASSO

پارک و کاسلا (۱۹) روش لزو بیزی (BL) را برای برآورد ضرایب رگرسیونی معرفی نمودند. آن‌ها از ایده تیبشیرانی (۲۴) برای ارتباط بین روش لزو و آنالیز بیزی بهره بردند. تیبشیرانی (۲۴) بیان کرد که برآوردهای لزو برای ضرایب رگرسیون را می‌توان به‌عنوان برآوردهای مد پسین با فرض این که ضرایب رگرسیون دارای توزیع پیشین دو نمایی باشند، در نظر گرفت. لزوی بیزی همچنین با استفاده از مدل آشیانه‌ای با تابع درست‌نمایی (رابطه ۴) در انتخاب ژنومی مورد استفاده قرار گرفت (۸).

$$f(y|\mu, X, m, \sigma^2) \sim N(\mu + Xm, \sigma^2 I), \qquad \text{(رابطه ۴)}$$

1- Bayesian Ridge Regression 2- Bayesian LASSO 3- Reproducing kernel Hilbert spaces regression 4- Neural Networks

[Downloaded from rap.sanru.ac.ir on 2026-11-06]
[DOR: 20.1001.1.22518622.1399.11.30.7.9]
[DOI: 10.52547/rap.11.30.101]

که y یک بردار n*1، μ بردار میانگین کل، m بردار اثرات نشانگری و X ماتریس طرحی است که m را به y ارتباط می‌دهد. $N(\mu + Xm, \sigma^2 I)$ نشان‌دهنده چگالی نرمال با میانگین μ + Xm و واریانس $\sigma^2 I$ است که I یک ماتریس یکه n×n می‌باشد.

RKHS

در این روش، تابع رگرسیون به‌صورت رابطه ۵ تبدیل می‌شود.

$$f(x_i) = \mu + \sum_{i'}^{n} \alpha_{i'} k(x_i, x_{i'}) \quad \text{(رابطه ۵)}$$

که $x_i = (x_{i1},...,x_{ip})$ و $x_{i'} = (x_{i'1},...,x_{i'p})$ بردارهای ورودی ژنوتیپ‌های نشانگری در افراد i، i' و $\alpha_{i'}$ ضرایب رگرسیون و $k(x_i, x_{i'}) = \exp(-h\|x_i - x_{i'}\|^2$ نشان‌دهنده Reproducing kernel می‌باشد که در تحقیق حاضر به دلخواه تعریف شده است و h پارامتر عرض باند و $\|x_i - x_{i'}\|$ نرم اقلیدسی بین هر جفت از بردارهای ورودی می‌باشد.

شبکه‌های عصبی

در مدل‌های رگرسیون خطی، امید ریاضی شرطی به‌صورت مجموع وزنی متغیرهای ورودی $E(y_i|X_i) = \sum_{j=1}^{p} x_{ij}\beta_j$ نشان داده می‌شود. تعداد زیادی الگوهای غیر خطی را می‌توان با انتخاب توابع پایه مناسب به‌صورت خطی نشان داد: $E(y_i|X_i) = \sum_{m=0}^{M} \phi(X_i) w_m$ که $\{\phi(X_i)\}_{m=1}^{M}$ توابع پایه‌های هستند که از متغیرهای ورودی به خط واقعی ترسیم می‌شوند.

انجام تجزیه‌های آماری مربوط به شش روش شامل روش‌های رگرسیون ریدج بیزی، بیز A، بیز B، بیز C، بیز LASSO و RKHS توسط بسته نرم‌افزاری BGLR و برای به‌کارگیری شبکه‌های عصبی در ارزیابی‌های ژنومی از بسته نرم‌افزاری BRNN استفاده شد. تعداد نورون‌های در نظر گرفته شده در این تحقیق برابر سه بود.

نتایج و بحث

در تحقیق حاضر صحت ارزیابی‌های ژنومی توسط وراثت‌پذیری، اندازه مؤثر جمعیت، روش برآورد اثرات نشانگری، توزیع QTL و تعداد QTL تحت تأثیر قرار گرفت. با افزایش وراثت‌پذیری و تعداد QTL، صحت ارزش‌های اصلاحی ژنومی افزایش یافت این در حالی بود که کاهش اندازه مؤثر جمعیت منجر به افزایش صحت ارزش‌های اصلاحی ژنومی گردید. میزان صحت بین روش‌های مختلف برآورد اثرات نشانگری و توزیع‌های QTL ها متفاوت بود.

در یک مطالعه، تعداد ۱۳۶۸ گاو سیاه ژاپنی برای ۵۹۲۰۳۴ SNP با استفاده از کیت Illumina BovineHD BeadChip تعیین ژنوتیپ شده و فنوتیپ آن‌ها بر اساس عوامل مورد بررسی در سناریوهای مختلف شبیه‌سازی شد. عوامل مورد بررسی در این تحقیق شامل فراوانی آلل نادر QTL، وراثت‌پذیری QTL، تعداد QTL، توزیع QTL، مدل پیش‌بینی، اندازه جمعیت مرجع و جمعیت تأیید بود (۲۲). بر اساس نتایج حاصله از این تحقیق، صحت ارزیابی‌های ژنومی تحت تأثیر عوامل فراوانی آلل نادر QTL، وراثت‌پذیری QTL، تعداد QTL، توزیع QTL و اندازه جمعیت مرجع قرار گرفت.

برای بررسی رفتار متفاوت روش‌های ارزیابی ژنومی با معماری‌های ژنتیکی صفات کمّی و ژنوم، یک مطالعه شبیه‌سازی طراحی شد که طی آن دو روش GBLUP و BayesB تحت سه اندازه مؤثر جمعیت (۲۰۰، ۱۰۰۰ و ۲۰۰۰)، دامنه وسیعی از تعداد جایگاه صفت کمّی (Me ۰/۰۳، Me ۰/۰۵، Me ۰/۱۵، Me ۰/۳، Me ۰/۵، Me ۰/۷۵ و Me ۱ که Me نشان‌دهنده تعداد قطعات کروموزومی مستقل است) و تعداد متفاوت نشانگر ژنتیکی مورد مقایسه قرار گرفتند. نتایج به‌دست آمده مبین آن بود که صحت ارزیابی‌های ژنومی تحت تأثیر عوامل مورد بررسی شامل روش برآورد اثرات نشانگری، تعداد جایگاه صفت کمّی، اندازه مؤثر جمعیت و تعداد نشانگرهای ژنتیکی پخش‌شده در سطح ژنوم قرار داشت (۶).

وراثت‌پذیری

مقادیر وراثت‌پذیری بررسی شده در تحقیق حاضر نماینده صفات با وراثت‌پذیری پایین (۰/۱)، متوسط (۰/۳) و بالا (۰/۵) بودند. وراثت‌پذیری نشان‌دهنده میزان مشارکت ژن‌ها در واریانس فنوتیپی است و در صفات با وراثت‌پذیری بالا بخش بیشتری از واریانس فنوتیپی، به‌وسیله واریانس ژنتیکی افزایشی تبیین شده و ارتباط بین ژنوتیپ و فنوتیپ بالا است که در نتیجه این امر صفات با وراثت‌پذیری بالا پاسخ مناسب‌تری به انتخاب نشان می‌دهند.

صفات با وراثت‌پذیری کمتر از قبیل صفات باروری و مقاومت در برابر بیماری‌ها اغلب به‌دلیل فقدان رکوردهای فنوتیپی معتبر و پاسخ نامناسب به انتخاب مستقیم، از تابع هدف اصلاح نژادی کنار گذاشته می‌شوند. از آنجا که روش انتخاب ژنومی توانایی پیش بینی ارزش ارثی برای حیوانات جمعیت تأیید بدون رکورد را تنها با استفاده از اثرات نشانگری برآورد شده دارد، می‌توان بسیاری از صفات ذکر شده را با این روش، تحت فرآیند انتخاب قرار داد.

با توجه به موارد ذکر شده انتظار می‌رود با افزایش وراثت‌پذیری، به دام انداختن اثرات ژن‌ها با استفاده از رکوردهای فنوتیپی به‌صورت مؤثرتری انجام گیرد. شکل *1a* نشان می‌دهد که با افزایش وراثت‌پذیری صفت از ۰/۱۰ به ۰/۵۰، میزان صحت ارزیابی‌های ژنومی افزایش یافت. البته شیب افزایش صحت ارزیابی‌های ژنومی بین وراثت‌پذیری ۰/۱ و ۰/۳۰ بیشتر از میزان آن بین ۰/۳۰ به ۰/۵۰ است که مبین کاهش شیب صحت با افزایش میزان وراثت‌پذیری است.

قابل ذکر است در پلات جعبه‌ای زیر (*1a*)، باکس بالایی، باکس پایینی، خط میانی و ستاره به‌ترتیب نشان‌دهنده، صدک ۷۵، صدک ۲۵، میانه و میانگین است. شکل نشان می‌دهد که پراکنش صحت‌های به‌دست آمده در وراثت‌پذیری ۰/۱۰ بیشترین و در وراثت‌پذیری ۰/۵۰ کمترین است.

تحقیقات زیادی (۱۵،۶،۲) صحت بالا برای وراثت‌پذیری‌های بالا را مورد تأیید قرار داده اند. برای صفات با وراثت‌پذیری بالا (مانند خصوصیات لاشه) اثرات ژنی سهم

[Downloaded from rap.sanru.ac.ir on 2026-11-06]
[DOR: 20.1001.1.22518622.1399.11.30.7.9]
[DOI: 10.52547/rap.11.30.101]

بزرگی در واریانس فنوتیپی داشته و بنابراین ارزیابی‌های دقیق‌تری ایجاد می‌کنند. بنابراین اجرای برنامه‌های انتخاب منجر به پیشرفت ژنتیکی بالاتر در این نوع صفات می‌شود.

اندازه مؤثر جمعیت

در تحقیق حاضر اثر دو اندازه مؤثر جمعیت (۱۰۰ و ۵۰۰) مورد بررسی قرار گرفت. شکل *1b* به‌وضوح مقدار بالاتر صحت در حالت N_e=100 نسبت به N_e=500 را نشان می‌دهد. فرآیند انتخاب طی نسل‌های متمادی، منجر به افزایش رابطه خویشاوندی افراد جامعه، کاهش اندازه مؤثر جمعیت و در نتیجه افزایش LD شده است. بنابراین در جوامع امروزی که اندازه مؤثر جمعیت آن‌ها کمتر است، به‌دلیل این که قطعات کروموزومی مشترک که از جد مشترک به افراد به ارث رسیده است دارای اندازه بزرگتری هستند، LD بالاتر و بالتبع صحت ارزیابی‌های بالاتری نیز دارند (۵). در تحقیق حاضر نیز میزان LD برای N_e =۱۰۰ و N_e =۵۰۰ به‌ترتیب برابر ۰/۱۹ و ۰/۰۹ بود که به‌وضوح رابطه عکس بین اندازه مؤثر جمعیت و میزان عدم تعادل پیوستگی را نشان می‌دهد. N_e معیاری برای تعیین مشابهت آلل‌های یک جایگاه ژنی در نتیجه هم‌تباری بوده و مشخص کننده سطح هم‌خوانی و میزان واریانس ژنتیکی کاهش یافته به‌دلیل رانش تصادفی ژن است. مطالعات مختلف نشان داده است به‌دلیل این که بسیاری از افراد سهم اندک و یا در برخی موارد هیچ سهمی در خزانه ژنی نسل‌های آینده ندارند مقدار N_e غالباً از تعداد واقعی مولدها کمتر است (۱۸). در دام‌ها اغلب علیرغم افزایش افراد جمعیت طی سال‌های متمادی مقدار N_e کاهش یافته که این موضوع بیانگر این واقعیت است که که به موازات افزایش تعداد افراد، متوسط رابطه خویشاوندی بین مولدها در نتیجه آمیزش‌های فامیلی نیز افزایش یافته است.

یانگ و سیکورا (۲۸) اندازه مؤثر جمعیت برای گاوهای هلشتاین آمریکا را ۱۰۰ برآورد کردند که طی بررسی‌های بعدی عدد مشابهی برای گاوهای هلشتاین استرالیایی به‌دست آمد (۹). طی تحقیقات صورت گرفته، اندازه مؤثر جمعیت در گاوهای جرزی کمتر بوده و برابر ۳۰ برآورد شد. بنابراین برای این نژاد می‌توان از چیپ‌های با تراکم کمتر استفاده نمود.

در بررسی اندازه جمعیت مرجع و اندازه مؤثر جمعیت در صحت ارزیابی صفات رشد و لاشه جوجه‌های چینی، دلیل بالاتر بودن صحت ارزیابی ژنومی در جامعه‌ای با اندازه مؤثر جمعیت کمتر بدین صورت بیان شد که زمانی که اندازه مؤثر جمعیت کوچک است، افراد قطعات کروموزومی بزرگتری را به اشتراک گذاشته و در نتیجه پیش بینی‌های ژنومی با صحت بالا به‌دست خواهد آمد (۱۳).

توزیع اثرات QTL

در این مطالعه، تفاوت صحت برای صفات دارای توزیع اثرات ژنی متفاوت، معنی‌دار بود و بیشترین صحت متعلق به توزیع نرمال بود، ولی بین توزیع گاما و یکنواخت تفاوت معنی‌داری مشاهده نشد (شکل *1c*).

عبدالهی و همکاران (۱) طی یک مطالعه شبیه‌سازی گزارش کردند که علیرغم این که صحت ارزش‌های اصلاحی ژنومی در حالت توزیع نرمال بهتر از گاما و یکنواخت بود اما اثرات معنی‌داری بین صحت توزیع‌های اثرات ژنی مشاهده نشد. این محققین بیان کردند که در تعداد QTL پایین، احتمال شکل‌گیری توزیع اثرات ژنی کم بوده و توزیع آماری مورد نظر با تعداد ژن‌های بزرگ اثر و کوچک اثر به‌خوبی بیان و نمایان نمی‌شود که می‌تواند محتمل‌ترین دلیل برای عدم تفاوت معنی‌داری بین این توزیع‌ها باشد. گودارد (۱۰) گزارش نمود روش‌های بیزی برای موقعی که اثرات ژن‌ها با توزیع گاما مطابقت داشته باشند به صحت بیشتری منجر می‌شوند.

در یک مطالعه شبیه‌سازی که صحت پیش بینی دو روش بیزی Ridge وLasso تحت معماری‌های مختلف ژنتیکی مورد بررسی قرار گرفت، از سه توزیع مختلف گاما ۱ (shape=0.4 , scale=1.66)، گاما ۲ (shape=0.4 , scale=1) و توزیع بتا (shape2=1.16 , shape1=3.11) برای اثرات جایگزینی ژن‌ها استفاده شد و مشاهده گردید در حالت کلی هنگامی که توزیع QTL از تابع گاما پیروی می‌کند، روش‌های بیزی صحیح‌تر از بقیه می‌توانند پیش بینی کنند (۱۴) محققین دلیل این برتری را به دو عامل نسبت دادند. اول این‌که توزیع پیش فرض دو روش بیزی ذکر شده با توزیع گامای اثرات و واریانس ژن‌های عمده اثر هماهنگی بیشتری نسبت به توزیع بتا دارد که باعث برآورد صحیح‌تر اثرات چند شکلی‌های تک نوکلئوتیدی می‌گردد. دلیل دوم این است که توزیع گاما دارای تعداد بیشتری ژن‌های عمده اثر با اثرات بزرگتر است.

تعداد QTL

در تحقیق حاضر اثر معنی‌دار تعداد QTL بر صحت ارزیابی‌های ژنومی مشهود بوده و با افزایش تعداد جایگاه ژنی صفات کمّی از ۵۰ به ۲۰۰، صحت پیش بینی ژنومی افزایش یافت (شکل 1d). در بررسی تحقیقات انجام گرفته توسط سایر محققین الگوی خاصی برای روند صحت ارزیابی‌های ژنومی در نتیجه افزایش تعداد QTLها مشاهده نشد.

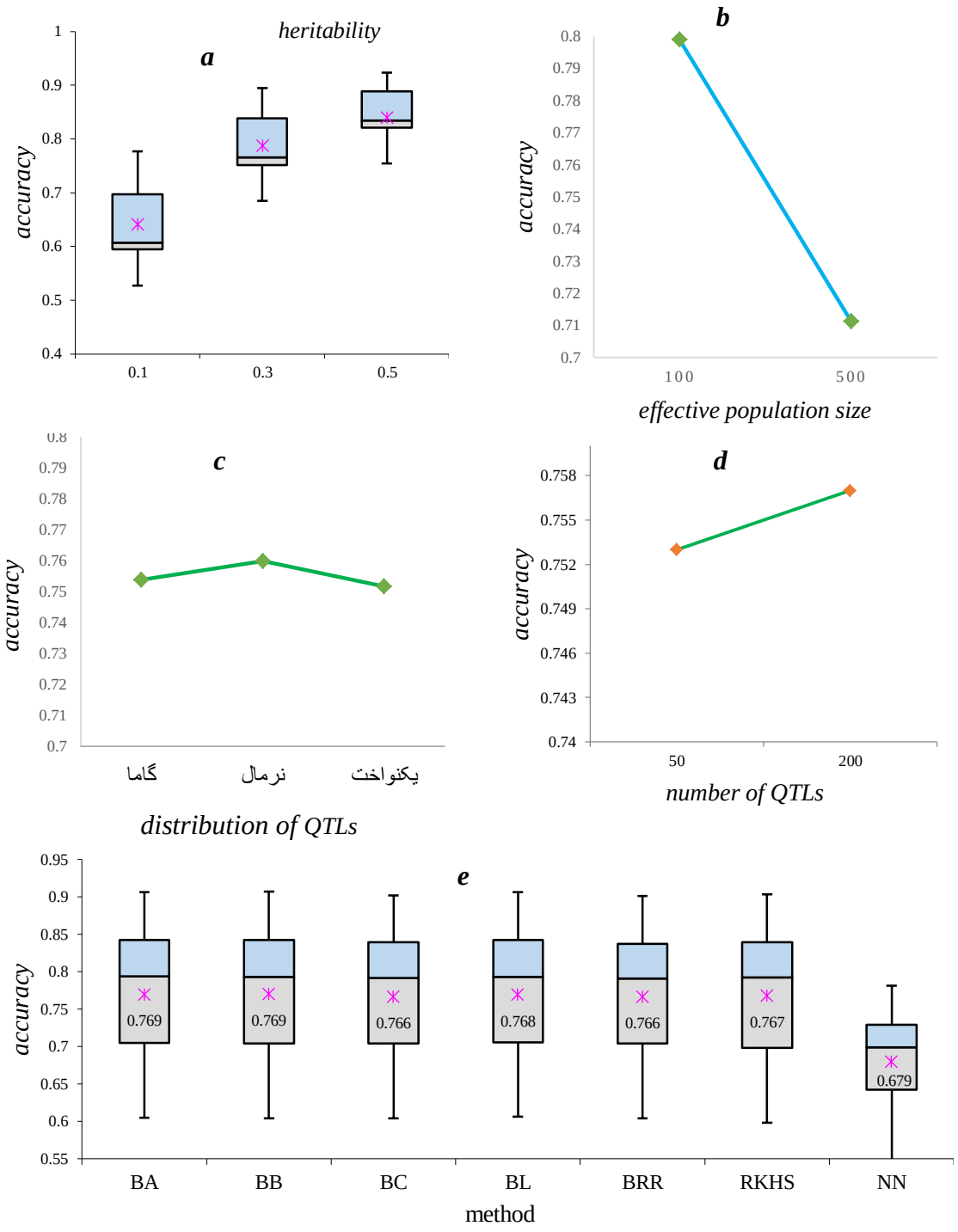
در یک تحقیق که صحت پیش‌بینی ژنومی برای تعداد متفاوت QTL (۵۰، ۱۰۰، ۳۰۰، ۵۰۰، ۱۰۰۰ و ۲۰۰۰) مورد بررسی قرار گرفت، در عمده سناریوها با افزایش تعداد QTL از ۵۰ به ۳۰۰ صحت افزایش از ۳۰۰ تا ۵۰۰ صحت کاهش، از ۵۰۰ تا ۱۰۰۰ صحت بدون تغییر و پس از آن با شیب اندکی کاهش یافت (۶) که با نتایج تحقیق حاضر مطابقت دارد. در یک مطالعه شبیه‌سازی که صفاتی با ۳، ۳۰، ۳۰۰ و ۳۰۰۰ QTL شبیه‌سازی شدند، در همه سناریوها با افزایش تعداد QTL صحت ارزیابی‌های ژنومی کاهش یافت (۲۶).

در ارزیابی ژنومی صفات آستانه ای با معماری‌های ژنتیکی متفاوت با استفاده از روش‌های بیزی، تعداد ۲۰، ۱۰۰ و ۲۰۰ QTL شبیه سازی شدند (۵) و گزارش گردید که صحت روش‌ها بین سناریو های دارای تعداد مختلف QTL معنی دار نبود و با افزایش شمار QTL صحت برآورد ها به میزان بسیار ناچیز (و غیرمعنی‌دار) در توزیع گاما، کاهش، در توزیع آماری نرمال، ثابت، و در توزیع یکنواخت، افزایش یافته است.

افزایش تعداد QTLها در صورتی که با افزایش تعداد نشانگرها همراه گردد، می‌تواند صحت ارزیابی‌های ژنومی را افزایش دهد. افزایش تعداد نشانگرها باعث می‌شود اثرات همه

[Downloaded from rap.sanru.ac.ir on 2026-11-06]
[DOR: 20.1001.1.22518622.1399.11.30.7.9]
[DOI: 10.52547/rap.11.30.101]

QTLها در نظر گرفته شود. در صورتی‌که افزایش تعداد QTL با افزایش تعداد نشانگرها همراه نباشد، صحت ارزش‌های اصلاحی ژنومی کاهش می‌یابد، به این دلیل که به‌دلیل محدود بودن مقدار واریانس ژنتیکی در مقابل تعداد زیادی QTL که سهم هر QTL در ارزش ژنتیکی کل کاهش می‌یابد و قدرت مدل‌ها در برآورد تأثیرات کم خواهد شد (۲۰).



شکل ۱- اثر عوامل وراثت‌پذیری (*a*)، اندازه موثر جمعیت (*b*)، توزیع QTL (*c*)، تعداد QTL (*d*) و روش برآورد اثرات نشانگری (*e*) بر صحت ارزش های اصلاحی ژنومی

Figure 1. Effect of Heritability (*a*), Effective population size (*b*), Distribution of QTLs (*c*), Number of QTLs (*d*) and Marker effect estimation methods (*e*) on the accuracy of genomic breeding values

[Downloaded from rap.sanru.ac.ir on 2026-11-06]
[DOR: 20.1001.1.22518622.1399.11.30.7.9]
[DOI: 10.52547/rap.11.30.101]

روش برآورد اثرات نشانگری

در تحقیق حاضر صحت ارزیابی ژنومی به‌دست آمده توسط هفت روش متشکل از پنج روش بیزی (BA, BB, BC, BL, BRR) و دو روش ناپارامتری (RKHS, NN) مورد مقایسه قرار گرفت. بیشترین مقدار صحت پیش‌بینی ژنومی (۰/۹۰۷) مربوط به‌روش BB و کمترین آن (۰/۵۱۳) متعلق به‌روش NN بود. قابل ذکر است که علاوه بر توانایی متفاوت روش‌های مختلف، مقدار صحت در روش NN به عواملی چون تعداد نورون‌ها، در RKHS به کرنل انتخابی و در سایر روش‌ها به تعیین هایپر پارامترها بستگی دارد.

در یک مطالعه شبیه‌سازی که صحت ارزش‌های اصلاحی ژنومی به‌دست آمده توسط روش‌های random forest، support vector machine، RKHS، رگرسیون ریج و Bayes A مورد مقایسه گرفت، بیشترین صحت ارزش‌های اصلاحی ژنومی توسط روش Bayes A حاصل شد و در حالت کلی بیان گردید که صحت ارزش‌های اصلاحی ژنومی با استفاده از روش‌های پارامتری بر روش‌های نیمه پارامتری و ناپارامتری برتری دارند (۲۱).

علت پایین‌بودن صحت روش‌های ناپارامتری خصوصاً روش NN بیش برازش‌شدن مدل در این روش است. یعنی در این نوع روش‌های غیر پارامتری مدل آنچنان روی مشاهدات مورد استفاده برازش می‌شود که قابلیت پیش بینی خود را برای نمونه‌های مستقل دیگر از دست می‌دهد. این مشکل زمانی پیش می‌آید که تعداد مشاهدات از تعداد مجهولات خیلی کم است و سعی داشته باشیم که بیشترین برازش را هم به‌دست بیاوریم. همان‌گونه که قبلاً بیان شد، در این تحقیق بالاترین صحت متعلق به‌روش BB بود که احتمالاً به‌دلیل چگالی پیشین واقع‌گرایانه‌تر آن برای برآورد اثرات نشانگری است.

در تحقیقی که اخیراً جهت ارزیابی صحت پیش‌بینی ارزش‌های اصلاحی ژنومی صورت گرفته، چهار روش شامل GBLUP، Bayes B، RKHS و NN برای صفات تولید شیر، تولید چربی، تولید پروتئین و شمارش سلول‌های سوماتیک گاوهای هلشتاین ایران مورد بررسی قرار گرفته است که مشاهده شد در حالت کلی روش بیز B دارای بیشترین و روش شبکه عصبی دارای کمترین صحت بوده‌اند. البته برای صفت شمارش سلول‌های سوماتیک دو روش GBLUP و بیز B دارای صحت برابر با ۰/۴۹ بودند. در تحقیق ذکر شده بیشترین صحت مربوط به صفت تولید چربی بود که احتمالاً به این دلیل است که این صفت تحت تأثیر تعداد محدودی ژن‌های بزرگ اثر است که این ژن‌ها نسبت زیادی از واریانس ژنتیکی صفت را توجیه می‌کنند. یکی از این ژن‌ها ژن بزرگ اثر DGAT1 می‌باشد. در توصیف این ویژگی‌ها گزارش شده که مدل‌هایی با توزیع پیشین انتها کشیده متأثر از نشانگر نظیر بیز A و روش‌های انتخاب متغیر نظیر بیز Cπ و بیز LASSO در مقایسه با GBLUP با دقت بیشتری تأثیرات نشانگرها را برآورد می‌کنند (۱۷).

یک روش برآورد اثرات نشانگری بهینه باید بیشترین صحت ممکن را داشته، از برآورد نامناسب داده‌ها ممانعت به‌عمل آورد و تا حد امکان به‌جای روابط خویشاوندی شجره‌ای، بر پایه LD بین نشانگر و QTL باشد. به‌علاوه چنین روش‌هایی باید به آسانی قابلیت استفاده داشته باشند، در دامنه وسیعی از صفات دارای نتیجه پایدار بوده و از نظر محاسباتی آسان باشند (۵).

نتایج این تحقیق مشخص کرد که بین روش‌های مورد بررسی، روش‌های بیزی پیشنهاد شده توسط موویسن و همکاران (۱۶) یعنی روش‌های بیز B و A بیشترین صحت را داشتند بنابراین می‌توان خصوصاً روش بیز B را به‌دلیل چگالی پیشین واقع‌گرایانه‌تر آن برای برآورد اثرات نشانگری توصیه نمود. بین دو روش ناپارامتری مورد بررسی روش RKHS دارای عملکرد بهتری نسبت به NN بوده و در بعضی حالات در حد روش‌های بیزی بود ولی در هر حال تا مادامی که می‌توان از روش‌های بیز استفاده کرد، استفاده از این روش‌ها منطقی به‌نظر نمی‌رسد. با افزایش وراثت‌پذیری، صحت ارزش‌های اصلاحی ژنومی افزایش یافت. افزایش اندازه مؤثر جمعیت، میزان صحت ارزش‌های اصلاحی ژنومی کاهش یافت. تعداد و توزیع QTL بر صحت ارزیابی‌های ژنومی مؤثر بودند. با افزایش تعداد QTL، میزان صحت ارزیابی‌های ژنومی افزایش یافت. بین توزیع‌های QTL، صفاتی که دارای ژن‌های عمده با توزیع نرمال بودند، دارای بیش‌ترین صحت ارزیابی ژنومی بودند.

منابع

1. Abdollahi-Arpanahi, R., A. Pakdel, A. Nejati-Javaremi and M. Moradi Shahrebabak. 2014. Comparison of genomic evaluation methods in complex traits with different genetic architecture. Journal of Animal Production, 15(1): 65-77 (In Persian).
2. Atefi, A., A.A. Shadparvar and N. Ghavi Hossein-Zadeh. 2016. Comparison of whole genome prediction accuracy across generations using parametric and semi parametric methods. Acta Scientiarum. Animal Sciences, 38(4): 447-453.
3. Atefi, A., A.A. Shadparvar and N. Ghavi Hossein-Zadeh. 2018. Accuracy of genomic prediction under different genetic architectures and estimation Mmethods. Iraninan Journal of Applied Animal Science, 8(1): 43-52.
4. Baneh, H., A. Nejati Javaremi, G. Rahimi-Mianji and M. Honarvar.2015. Genomic Evaluation of Threshold Traits with Different Genetic Architecture using Bayesian Approaches. Research on Animal Production, 8 (14): 149-154 (In Persian).
5. Daetwyler, H.D., R. Pong-Wong, B. Villanueva and J.A. Woolliams. 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics, 185(3): 1021-1031.

[Downloaded from rap.sanru.ac.ir on 2026-11-06]
[DOR: 20.1001.1.22518622.1399.11.30.7.9]
[DOI: 10.52547/rap.11.30.101]

6. De los Campos, G., D. Gianola, G.J. Rosa, K.A. Weigel and J. Crossa. 2010. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genetics research, 92(40): 295-308.
7. De Los Campos, G., H. Naya, D .Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel and J.M. Cotes. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics, 182(1): 375-385.
8. De Roos, A.P.W., B.J. Hayes, R.J. Spelman and M.E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. Genetics, 179: 1503-1512.
9. Goddard, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica, 136(2): 245-257.
10. Goddard, M.E. and B.J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics, 10(6): 381-391.
11. Habier, D., R. Fernando and J. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics, 177(4): 2389-2397.
12. LaFramboise, T. 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Research, 37(13): 4181-4193.
13. Liu, T., H. Qu, C. Luo, D. Shu, J. Wang, M.S. Lund and G. Su. 2014. Accuracy of genomic prediction for growth and carcass traits in Chinese triple-yellow chickens. BMC Genetics, 15(1): 110.
14. Mahmoudi, N., A.A. Mehrgardi, M. Honarvar and A. Esmailizadeh kashkooeiyeh.2015. Study of QTL Effects Distribution on Accuracy of Genomic Breeding values Estimated Using Bayesian Method. 2015. Iranian Journal of Animal Science Research, 7(3): 356-363 (In Persian).
15. Martinez, R., W. Burgos-Paz, D. Bejarano, P. Reyes and J.F. Rocha. 2018. Genomic predictions and accuracy of weight traits in a breeding program for Colombian Brahman, Proceedings of the World Congress on Genetics Applied to Livestock Production, Volume Genetic Gain- Strategies for Local Breeds 2, 644, Auckland, New Zealand.
16. Meuwissen, M.H.E., B.J. Hayes and M. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157(4): 1819-1829.
17. Mohammadi, Y., M.M. Shariati, S. Zerehdaran, M. Razmkabir, M.B. Sayyadnejad and M.B. Zandi. 2016. The accuracy of genomic breeding value for production trait in Iranian Holstein dairy cattle using parametric and non-parametric methods. Journal of Animal Production, 18(1): 1-11 (In Persian).
18. Norberg, E. and A.C. Sorensen. 2007. Inbreeding trend and inbreeding depression in the Danish populations of Texel, Shropshire, and Oxford Down. Journal of animal science, 85(2): 299-304.
19. Park, T. and G. Casella. 2008. The bayesian lasso. Journal of the American Statistical Association, 103(482): 681-686.
20. Saheb Alam, H., M. Gholizadeh, H. Hafezian and A. Farhadi. 2018. Comparison of Bayesian methods in the genomic evaluation with different genetic architecture. Research on Animal Production 8(18): 177-186 (In Persian).
21. Saheb Alam, H., M. Gholizadeh, H. Hafezian and A. Farhadi. 2019. Comparison of parametric, semiparametric and nonparametric methods in genomic evaluation. Journal of Genetics 98, 102.
22. Sargolzaei, M. and F. S. Schenkel. 2009. QMSim: a large-scale genome simulator for livestock. Bioinformatics, 25(5): 680-681.
23. Teimuria, M., M.M. Shariati and A.A. Aslaminejad. 2016. Comparison of methods for the implementation of genomic selection in Holstein. Research on Animal Production, 7(14): 198-203 (In Persian).
24. Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.
25. Uemoto, Y., S. Sasaki, T. Kojima, Y. Sugimoto and T. Watanabe. 2015. Impact of QTL minor allele frequency on genomic evaluation using real genotype data and simulated phenotypes in Japanese Black cattle. BMC Genetics, 16(1): 1.
26. Van den Berg, S., M.P. Calus and Y. Wientjes. 2015. Across population genomic prediction scenarios in which Bayesian variable selection outperforms GBLUP. BMC Genetics, 16(1): 146.
27. Wang, X., J. Miao, T. Chang, J. Xia, B. An, Y. Li, L. Xu, L. Zhang, X. Gao, J. Li and H. Gao. 2019. Evaluation of GBLUP, BayesB and elastic net for genomic prediction in Chinese Simmental beef cattle. PLoSONE 14(2): e0210442.
28. Young, C. and A. Seykora. 1996. Estimates of inbreeding and relationship among registered Holstein females in the United States. Journal of Dairy Science, 79(3): 502-505.

[Downloaded from rap.sanru.ac.ir on 2026-11-06]
[DOR: 20.1001.1.22518622.1399.11.30.7.9]
[DOI: 10.52547/rap.11.30.101]

Effect of Markers Effect Estimation Methods, Population Structure and Trait Architercture on the Accuracy of Genomic Breeding Values

Abbas Atefi[1], Abdol Ahad Shadparvar[2] and Navid Ghavi Hossein-Zadeh[3]

1- Graduated PhD Student, Department of Animal Science, University of Guilan
2- Professor, Department of Animal Science, University of Guilan, (Crossponding author: shad@guilan.ac.ir)
3- Professor, Department of Animal Science, University of Guilan
Recived: 26 February 2019 Accepted: 10 October 2020

Abstract

This study aimed to investigate the effect of the method of estimating the effects of markers , QTLs distribution, number of QTLs, effective population size and trait heritability on the accuracy of genomic predictions. Two effective population sizes, 100 and 500 individuals, were simulated by QMSim software. A 100 cM genome including one chromosome was simulated where 500 SNPs and two different numbers of QTLs (50 and 200) were distributed on it randomly. In this study three levels of heritability (0.1, 0.3 and 0.5) were considered. Genomic breeding values were predicted using Bayesian ridge regression, BayesA, BayesB, BayesC, Bayesian LASSO, Reproducing kernel Hilbert space and neural networks. In this research, the accuracy of genomic breeding values were affected by trait heritability, effective population size, markers effect estimation methods, QTLs distribution and number of QTLs. The Bayes A and B had the highest accuracy while accuracy of neural networks method was the lowest. The accuracy of genomic breeding values were increased as the heritability of trait and number of QTLs increased while the accuracy was decreased as the effective population decreased. Considering the QTLs distribution, the highest accuracy was achieved when the QTLs distributed normally.

Keywords: Genomic Evaluation, Genetic architecture, Non-parameteric method, Parameteric method

[Downloaded from rap.sanru.ac.ir on 2026-11-06]

[DOR: 20.1001.1.22518622.1399.11.30.7.9]
[DOI: 10.52547/rap.11.30.101]