



صحت ارزیابی ژنومی با در نظر گرفتن اثر متقابل بین روش آماری برآورد آثار نشانگری، ساختار جمعیت و معماری ژنتیکی صفت

عباس عاطفی^{*}، عبدالاحد شادپرور^۲، نوید قوی حسین زاده^۲

۱- دانش‌آموخته دکتری ژنتیک و اصلاح نژاد دام، گروه علوم دامی، دانشکده علوم کشاورزی، دانشگاه گیلان
۲- استاد، گروه علوم دامی، دانشکده علوم کشاورزی، دانشگاه گیلان

(تاریخ دریافت: ۹۹/۰۱/۲۳ - تاریخ پذیرش: ۹۹/۰۴/۱۲)

چکیده

هدف تحقیق حاضر بررسی اثر متقابل بین روش‌های برآورد آثار نشانگری، ساختار جمعیت و معماری ژنتیکی صفت بر صحت ارزیابی ژنومی بود. جمعیت پایه‌ای با دو اندازه موثر ۱۰۰ و ۵۰۰ فرد به وسیله نرم افزار QMSim شبیه‌سازی شد. تعداد ۵۰۰ نشانگر و دو تعداد متفاوت جایگاه‌های کنترل کننده صفات کمی یا QTL (۵۰ و ۲۰۰) روی ژنوم متشکل از یک کروموزوم به طول ۱۰۰ سانتی‌مورگان به‌طور تصادفی پراکنده شدند. در پژوهش حاضر، سه صفت با وراثت پذیری ۰/۱، ۰/۳ و ۰/۵ شبیه‌سازی شد. ارزش‌های اصلاحی ژنومی به وسیله روش‌های رگرسیون ریج بیزی، بیز A، بیز B، بیز C، بیز LASSO، روش نیمه پارامتری بر پایه کرنل و شبکه‌های عصبی پیش‌بینی شدند. در هر سه میزان وراثت پذیری، با افزایش اندازه موثر جمعیت، صحت برآوردهای ژنومی کاهش یافت، ولی شدت آن برای روش‌های آماری یکسان نبود. در وراثت پذیری پایین، با افزایش تعداد QTL، میزان صحت ارزش‌های اصلاحی ژنومی افزایش یافت. بر عکس، میزان صحت ارزش‌های اصلاحی در وراثت پذیری متوسط و بالا کاهش یافت. در صفات با توزیع یکنواخت QTLها، با افزایش تعداد QTL، مقدار صحت ارزیابی ژنومی کاهش، ولی افزایش تعداد QTL در صفاتی که QTLهای آنها دارای توزیع نرمال و گاما بود، منجر به افزایش صحت شد. در هر سه نوع توزیع یکنواخت، نرمال و گاما، میزان صحت با افزایش اندازه موثر جمعیت کاهش یافت. نتایج پژوهش حاضر آثار متقابل بین روش‌های آماری، ساختار جمعیت و معماری ژنتیکی صفت را بر صحت ارزیابی ژنومی به وضوح نشان می‌دهد.

واژه‌های کلیدی: اثر متقابل، انتخاب ژنومی، ساختار جمعیت، صحت پیش‌بینی، معماری ژنتیکی صفت

* نویسنده مسئول: abbas.atefi@gmail.com

مقدمه

مواد و روش‌ها

ارزش‌های اصلاحی برآورد شده با استفاده از نشانگرهای توزیع شده روی ژنوم (GEBV) در اصلاح نژاد گاو شیری مورد توجه قرار گرفته و تلاش‌ها برای استفاده از این فناوری در گونه‌های حیوانی دیگر و گیاهان آغاز شده است (Gonzalez-Recio *et al.*, 2007). مزیت‌های انتخاب بر اساس GEBV نسبت به روش‌های انتخاب سنتی، که طی آنها برآورد ارزش‌های اصلاحی فقط بر اساس اطلاعات فنوتیپی و شجره صورت می‌گیرد، کاهش فاصله نسل و همچنین میزان همخونی کمتر به دلیل قدرت تمایز بهتر بین نتاج است (Muir, 2007).

در روش انتخاب ژنومی، کروموزوم‌ها بر اساس نشانگرهای بسیار مترکم به قطعاتی با طول بسیار اندک، به صورت کسری از یک سانتی مورگان، تقسیم شده و سپس آثار این قطعات با مدل آماری مناسب برآورد می‌شود. قطعات مذکور به صورت یک یا ترکیبی از دو یا چند نشانگر تعریف می‌شوند. در حالت تک نشانگری، هر یک از مکان‌های نشانگر به صورت مستقل از یکدیگر فرض می‌شوند (Meuwissen *et al.*, 2001). این آثار را می‌توان به منظور پیش‌بینی ارزش‌های اصلاحی سایر حیوانات آن جمعیت بدون حضور رکوردهای فنوتیپی بکار گرفت.

معمولاً فرآیند انتخاب ژنومی در دو مرحله انجام می‌شود. در مرحله اول، آثار نشانگرها (در مواردی هاپلوتایپ‌ها) در گروهی از افراد که دارای ژنوتیپ معلوم و رکورد فنوتیپی هستند (جمعیت مرجع) برآورد می‌شوند و سپس ارزش اصلاحی ژنومی افراد بدون رکورد فنوتیپی ولی دارای ژنوتیپ (جمعیت هدف یا تایید)، فقط بر اساس اطلاعات ژنوتیپی نشانگرهای آنها محاسبه می‌شود (Hayes, 2007).

مطالعات با استفاده از داده‌های واقعی و شبیه‌سازی شده نشان داده است که برای تعیین بهترین مدل آماری باید ترکیب معماری ژنتیکی صفت و ساختار جمعیت نیز در نظر گرفته شود (Daetweler *et al.*, 2010; Valente *et al.*, 2018). بنابراین هدف از تحقیق حاضر، بررسی صحت ارزش‌های اصلاحی ژنومی با در نظر گرفتن آثار متقابل بین روش‌های برآورد اثر نشانگری، معماری ژنتیکی صفت و ساختار جمعیت بود.

در تحقیق حاضر با استفاده از نرم افزار QMSim (Sargolzaei and Schenkel, 2009)، جمعیت پایه با دو اندازه موثر ۱۰۰ و ۵۰۰ فرد شبیه‌سازی شد. در وهله اول، 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 فراسنجه شکل برابر با ۰/۴ و فراسنجه مقیاس برابر با ۱/۶۶) برآورد شد.

به منظور ایجاد عدم تعادل پیوستگی، 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. انتخاب والدین نسل بعد به صورت تصادفی انجام گرفته و سیستم آمیزشی نیز از نوع آمیزش تصادفی بود. افراد نسل ۵۱ به عنوان جمعیت مرجع در نظر گرفته شدند که دارای فنوتیپ و ژنوتیپ نشانگری معلوم بوده و افراد نسل‌های ۵۲ تا ۵۴ نیز به عنوان جمعیت تایید محسوب شدند که فقط دارای اطلاعات ژنتیکی بودند.

ژنوم شبیه‌سازی شده متشکل از یک کروموزوم ۱۰۰ سانتی‌مورگانی بود که تعداد ۵۰۰ نشانگر تک نوکلئوتیدی یا SNP و تعداد متفاوت جایگاه‌های کنترل کننده صفات کمی یا QTL (۵۰ و ۲۰۰) به‌طور تصادفی روی کروموزوم پراکنده شده بودند. نشانگرها و QTLها به صورت دو آلی با فراوانی اولیه ۰/۵ بودند. نرخ جهش برای نشانگرها و QTLها برابر با $10^{-5} \times 2/5$ به ازای هر جایگاه ژنی در هر نسل در نظر گرفته شد.

در این پژوهش، سه صفت با وراثت پذیری کم، متوسط و زیاد (به ترتیب ۰/۱، ۰/۳ و ۰/۵) شبیه‌سازی شدند. تمام واریانس ژنتیکی صفت حاصل از QTLها بوده و ارزش اصلاحی واقعی هر فرد برابر مجموع آثار جایگزینی آلی QTLها و فقط با فرض آثار افزایشی بدست آمد. فنوتیپ‌ها از جمع باقیمانده‌ها، که از توزیع نرمال با میانگین صفر نمونه‌گیری شده بودند، و ارزش‌های اصلاحی واقعی یا TBV بدست آمدند.

ساختار عمومی مدل‌ها به شکل خطی به صورت زیر است:

مواجه هستیم و بنابراین امید ریاضی شرطی به صورت
 $E(y_i|X_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$ حاصل می‌شود.

تجزیه‌های آماری به وسیله بسته‌های BGLR و BRNN نرم افزار R انجام شد. به منظور کاهش پیچیدگی‌ها و در نتیجه کاهش زمان محاسبات، تعداد نورون‌های شبکه‌های عصبی استفاده شده در این تحقیق برابر با سه بود.

نتایج و بحث

اثر متقابل اندازه موثر جمعیت و وراثت پذیری: در تحقیق حاضر در هر سه سطح وراثت پذیری پایین، متوسط و بالا (به ترتیب ۰/۱، ۰/۳ و ۰/۵)، با افزایش اندازه موثر جمعیت، صحت برآوردهای ژنومی کاهش یافت (شکل ۱). البته مقدار کاهش در وراثت پذیری‌های مورد بررسی، متفاوت بود و بیشترین کاهش صحت ارزیابی‌ها (۰/۱۲)، برای صفات با وراثت پذیری پایین و کمترین کاهش (۰/۰۶)، مربوط به صفات با وراثت پذیری بالا بود. این امر مبین آن است که وراثت پذیری بالا، تا حدودی می‌تواند بر مشکل کاهش عدم تعادل پیوستگی یا LD به واسطه افزایش اندازه موثر جمعیت فائق آید. یادآوری می‌شود که در گونه‌های با اندازه موثر جمعیت بالا، برای داشتن ارزش‌های اصلاحی ژنومی با صحت مناسب، تراکم نشانگری باید افزایش یابد، ولی با توجه به این که در مطالعه حاضر، تراکم نشانگری ثابت فرض شده بود، با افزایش اندازه موثر جمعیت، کاهش صحت ارزش‌های اصلاحی قابل انتظار بود.

در تحقیقی که هدف آن بررسی تاثیر انتخاب ژنومی در گاوهای گوشتی بود (Weber, 2012)، دو سطح اندازه موثر جمعیت (۲۰۰ و ۳۰۰) و سه سطح وراثت پذیری (۰/۱، ۰/۳ و ۰/۶) در نظر گرفته شده بود. نتایج این مطالعه نشان داد که برای صفات با وراثت پذیری پایین، متوسط و بالا، افزایش اندازه موثر جمعیت منجر به کاهش صحت ارزیابی‌های ژنومی می‌شود که مقدار این کاهش برای صفات با وراثت پذیری پایین، کمترین و برای صفات با وراثت پذیری بالا، بیشترین مقدار بود (Weber, 2012). نتایج مشابهی در بررسی اثر ساختارهای جمعیتی متفاوت دو صفت (رشد و تولید دانه) بر صحت پیش‌بینی‌های ژنومی ذرت مشاهده شد (Valente et al., 2016).

$$y = \mu + \sum_{j=1}^p X_j g_j + e$$

که y بردار رکوردهای فنوتیپی، μ میانگین کل، Σ نشان-دهنده جمع برای همه نشانگرها، g_j ضریب نشانگر j و نشان‌دهنده اثر جایگزینی آلی، X_j ماتریس طرح برای کدهای ژنوتیپی نشانگر مربوطه و e بردار باقیمانده‌ها است. هفت روش آماری مختلف شامل رگرسیون ریبج بیزی (Hoerl and Kennard, 1970)، بیز A و بیز B (Meuwissen et al., 2001)، بیز C (Habier et al., 2011)، بیز LASSO (Park and Casella., 2008)، روش نیمه پارامتری بر پایه کرنل (RKHS) و روش ناپارامتری شبکه‌های عصبی (Gianola et al., 2006) برای برآورد آثار نشانگرها مورد استفاده قرار گرفتند که روش‌های نیمه پارامتری و ناپارامتری ذکر شده به اختصار به شرح زیر توضیح داده می‌شود.
RKHS: در این روش، تابع رگرسیون به شکل زیر تبدیل می‌شود:

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

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

شبکه‌های عصبی: در مدل‌های رگرسیون خطی، امید ریاضی شرطی به صورت مجموع وزنی متغیرهای ورودی $E(y_i|X_i) = \sum_{j=1}^p x_{ij} \beta_j$ نشان داده می‌شود. تعداد زیادی الگوهای غیرخطی را می‌توان با انتخاب توابع پایه مناسب به صورت خطی به شکل زیر نشان داد:

$E(y_i|X_i) = \sum_{m=0}^M \phi(X_i) w_m$ که $\{\phi(X_i)\}_{m=1}^M$ توابع پایه‌ای هستند که از متغیرهای ورودی به خط واقعی ترسیم می‌شوند. یک مثال از این‌ها، توابع پایه چند جمله‌ای $\phi = \{\phi_m(x_i) = x_i^m\}_{m=0}^M$ است. برای مثال، اگر $M=2$ ، با توابع پایه چندجمله‌ای درجه دوم $\phi = \{1, x_i, x_i^2\}$

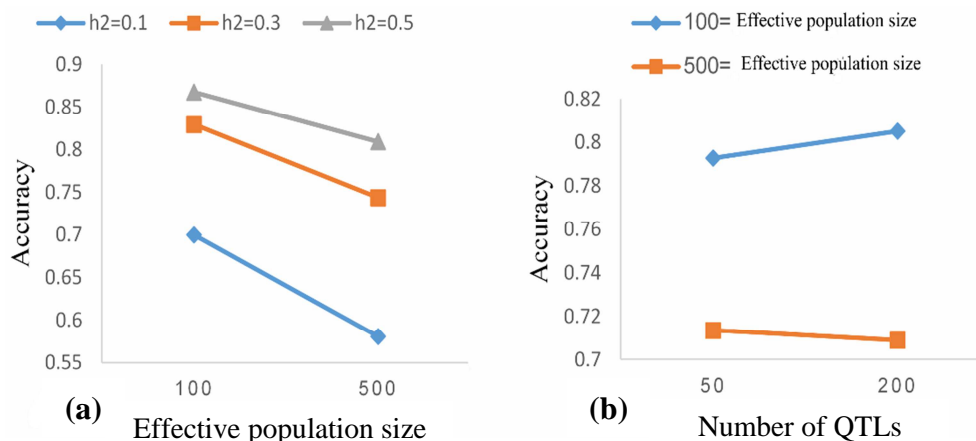


Fig. 1. Changes in accuracies of GEBVs for traits with different heritability estimates due to increment in effective population size (a), and for traits with different number of QTLs due to increment in effective population size (b)

شکل ۱- تغییر صحت ارزش‌های اصلاحی ژنومی برای صفات با وراثت پذیری متفاوت بر اثر افزایش اندازه موثر جمعیت (a) و برای صفات با تعداد متفاوت QTL بر اثر افزایش اندازه موثر جمعیت (b)

۰/۴۰۵، ۰/۴۵۰، ۰/۴۲۹، ۰/۴۱۴، ۰/۴۴۴، ۰/۴۱۶ و ۰/۳۹۸ بود. ولی در حالت کلی مشاهده شد که صحت ارزش‌های اصلاحی ژنومی در $N_e = 200$ و $N_e = 2000$ کاهش و در $N_e = 1000$ افزایش یافت، که این امر به وضوح به دلیل اثر متقابل بین تعداد QTL و اندازه موثر جمعیت است.

اثر متقابل روش برآورد آثار نشانگری و وراثت پذیری: در شکل ۲ اثر متقابل روش برآورد آثار نشانگری و وراثت پذیری از دو منظر مورد بررسی قرار گرفته است. همان‌گونه که مشاهده می‌شود در همه روش‌های پارامتری و ناپارامتری، با افزایش وراثت پذیری، صحت ارزیابی‌های ژنومی افزایش می‌یابد، ولی به وضوح مشخص است شیب افزایش صحت برای روش‌های بیزی و بین وراثت پذیری-های ۰/۱ و ۰/۳ بیشتر از این شیب بین وراثت پذیری‌های ۰/۳ و ۰/۵ است، که برعکس این مورد برای روش‌های نیمه پارامتری صادق است. این امر نشان‌دهنده بهبود بیشتر روش‌های ناپارامتری در وراثت پذیری‌های بالا است. قدرت پیش‌بینی روش RKHS در وراثت پذیری ۰/۱ و ۰/۵ تقریباً مشابه روش‌های بیزی است ولی در وراثت پذیری ۰/۳، این توانایی بسیار کمتر از آنها است. با این حال، روش شبکه‌های عصبی دارای قدرت پیش‌بینی کمتری نسبت به روش‌های بیزی است.

اثر متقابل تعداد QTL و اندازه موثر جمعیت: در صفاتی که تعداد QTL زیاد است، برای کسب یک سطح معین از صحت یا باید تراکم نشانگری بیشتری در نظر گرفت تا واریانس ژنتیکی مربوط به QTLها را در اختیار گرفت، یا در صورتی که تراکم نشانگری ثابت بود، به LD قوی‌تری بین نشانگرها و QTLها دسترسی داشت. از آنجا که بین اندازه موثر جمعیت و میزان LD رابطه منفی برقرار است، در تحقیق حاضر ملاحظه می‌شود که میزان صحت در صفات با تعداد QTL بالا کمتر از صفات با تعداد QTL پایین است (شکل ۱).

در یک مطالعه شبیه‌سازی، که طی آن صحت ارزش‌های اصلاحی ژنومی محاسبه شده با سه روش مورد بررسی قرار گرفته بود، در تمام سطوح اندازه موثر جمعیت، با افزایش تعداد QTLها، صحت ارزش‌های اصلاحی ژنومی کاهش یافت (Alanoshahr et al., 2018). در یک مطالعه دیگر (Daetwyler et al., 2010)، به منظور بررسی صحت ارزیابی ژنومی در سه سطح اندازه موثر جمعیت یا N_e (۲۰۰، ۱۰۰۰ و ۲۰۰۰) و هفت سطح QTL (0.03 Me, 0.05 Me, 0.15 Me, 0.3 Me, 0.5 Me, 0.75 Me, 1 Me) که برابر تعداد قطعات کروموزومی مستقل بود، مشاهده شد که در هر سه سطح N_e ، با افزایش تعداد QTL، صحت ارزش‌های اصلاحی ژنومی به صورت نامنظمی تغییر یافت. برای نمونه در $N_e = 200$ مقدار صحت برای سطوح QTL ذکر شده به ترتیب برابر با

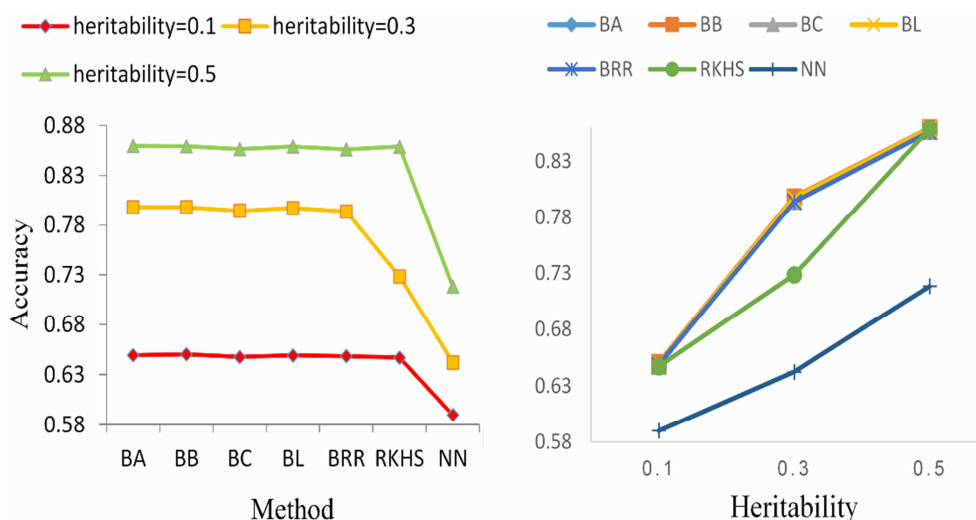


Fig. 2. Interaction effect of heritability and markers effect estimation methods on the accuracy.

BA: Bayes A, BB: Bayes B, BC: Bayes C, BL: Bayesian LASSO, BRR: Bayesian Ridge Regression, RKHS: Reproducing kernel Hilbert spaces, NN: Neural Networks

شکل ۲- اثر متقابل میان وراثت پذیری صفت و روش برآورد آثار نشانگری بر صحت ارزیابی‌های ژنومی

نظر می‌رسد این آثار متقابل علاوه بر میزان وراثت پذیری صفات به عوامل دیگری نیز بستگی داشته باشند. اثر متقابل تعداد *QTL* و وراثت پذیری: در شکل ۳ به وضوح مشاهده می‌شود که در وراثت پذیری پایین (۰/۱)، میزان صحت ارزش‌های اصلاحی ژنومی با افزایش تعداد *QTL* به میزان اندکی (۰/۰۱۸) افزایش یافته است، ولی میزان صحت ارزش‌های اصلاحی در وراثت پذیری متوسط و بالا به میزان مساوی (۰/۰۰۳) کاهش یافته است که نشان‌دهنده اثر متقابل بین این دو عامل است. شاید دلیل این موضوع این باشد که وقتی تعداد *QTL* افزایش می‌یابد به دلیل ثابت بودن تراکم نشانگری، سهم کمتری از واریانس مربوط به آن‌ها به تصرف در می‌آید. این موضوع وقتی اثر منفی خود را آشکارتر می‌کند که اصولاً سهم بیشتری از تنوع فنوتیپی صفت تحت تاثیر *QTL*ها قرار داشته باشد.

در یک مطالعه شبیه‌سازی که تاثیر سه سطح وراثت پذیری (۰/۱۰، ۰/۲۵ و ۰/۴۰)، دو سطح *QTL* (۴۰۰ و ۶۰۰) و سه تراکم نشانگری (۲۰۰۰، ۴۰۰۰ و ۶۰۰۰) بر میزان صحت ارزیابی‌های ژنومی را مورد بررسی قرار داد (Gorgani Firozjah *et al.*, 2014) تقریباً در همه سناریوها با افزایش تعداد *QTL*، میزان صحت ارزیابی‌ها افزایش یافت، با این تفاوت که بیشترین مقدار افزایش برای وراثت پذیری پایین (۰/۰۱) و کمترین مقدار افزایش برای صفات با وراثت پذیری بالا رخ داد. این در حالی است که در یک

بر اساس نتایج ارائه شده در شکل ۲ مشاهده می‌شود که برای صفات با وراثت پذیری پایین، متوسط و بالا، روش‌های پارامتری بهتر از روش‌های ناپارامتری عمل نموده‌اند، هر چند همان‌گونه که قبلاً ذکر شد روش RKHS در صفات با وراثت پذیری پایین و بالا دارای عملکرد مشابهی با روش‌های پارامتری مورد بررسی است. پیش از این بیان شده بود که کم بودن صحت روش‌های ناپارامتری عمدتاً به دلیل بیش برآورد کردن مدل در روش‌های ناپارامتری است (Howard *et al.*, 2014).

در یک تحقیق با استفاده از داده‌های واقعی جوجه‌های چینی، صحت ارزیابی ژنومی چهار روش BLUP، GBLUP، Bیز LASSO و BayesMix4 برای صفات وزن بدن در شش هفتگی (BW6)، وزن بدن در ۱۲ هفتگی (BW12)، درصد *epviscerating* (EP)، درصد عضله سینه (BMP) و درصد عضله پا (LMP) مورد ارزیابی قرار گرفت (Liu *et al.*, 2014). نتایج بررسی نشان داد که روش‌های مختلف در صفات بررسی شده، از نظر صحت دارای رتبه-بندی متفاوت هستند، به گونه‌ای که بیشترین صحت در صفات BW12 و BW6 (به ترتیب با وراثت پذیری ۰/۲۶ و ۰/۱۳) متعلق به روش GBLUP، در صفت EP و BMP (به ترتیب با وراثت پذیری ۰/۴۴ و ۰/۳۹) متعلق به روش BLUP و در صفت LMP (با وراثت پذیری ۰/۳۹) متعلق به روش BayesMix4 بود (Liu *et al.*, 2014). البته به

در یک مطالعه شبیه‌سازی، به منظور مطالعه توزیع آماری آثار QTL بر صحت ارزش‌های اصلاحی ژنومی، سناریوهای شبیه‌سازی شده با استفاده از سه نوع توزیع آماری گاما ۱ ($shape=0.4, scale=1$) و گاما ۲ ($shape=0.4, scale=1.66$) و توزیع بتا ($shape_1=3.11, shape_2=1.16$) برای ژن‌های عمده اثر، دو سطح وراثت پذیری (۰/۵ و ۰/۲) و تعداد متفاوت QTL (۱۰، ۲۰، ۵۰ و ۱۰۰) طراحی شد (محمودی و همکاران، ۱۳۹۴). نتایج این تحقیق نشان داد که در هر دو وراثت پذیری مورد بررسی، توزیع گاما ۱ دارای کمترین صحت ارزیابی ژنومی بود، ولی ترتیب دو توزیع گاما ۲ و بتا در وراثت پذیری‌های مورد بررسی تغییر نمود، به گونه‌ای که توزیع بتا در وراثت پذیری ۰/۲ و توزیع گاما ۲ در وراثت پذیری ۰/۵ دارای بیشترین صحت ارزیابی ژنومی بودند.

اثر متقابل تعداد و توزیع QTL در پژوهش حاضر مشاهده شد که وقتی QTL‌های صفت دارای توزیع یکنواخت باشند، با افزایش تعداد QTL از ۵۰ تا ۲۰۰، مقدار صحت ارزیابی‌های ژنومی کاهش یافت. برای صفاتی که QTL‌های آنها دارای توزیع نرمال و گاما بودند، مقدار صحت ارزیابی‌های ژنومی با افزایش تعداد QTL افزایش یافت که شیب این افزایش برای توزیع گاما بیشتر بود (شکل ۴).

مطالعه شبیه‌سازی دیگر، در هر سه سطح وراثت پذیری (۰/۱، ۰/۳ و ۰/۵) با افزایش تعداد QTL (۱۰۰، ۲۰۰ و ۳۰۰)، صحت ارزش‌های اصلاحی ژنومی کاهش یافت (Alanoshahr et al., 2017).

طی یک مطالعه شبیه‌سازی، سه سطح وراثت پذیری (۰/۰۵، ۰/۱۰ و ۰/۲۵) و تعداد متفاوت QTL (۴، ۱۰، ۲۰ و ۴۰) برای ارزیابی صحت ارزش‌های اصلاحی ژنومی در نظر گرفته شد. نتایج این تحقیق نشان داد که در وراثت پذیری ۰/۰۵ و ۰/۲۵، با افزایش تعداد QTL از ۴ به ۱۰ و سپس به ۲۰، میزان صحت، افزایش و سپس کاهش یافت، ولی در وراثت پذیری ۰/۱۰، با افزایش تعداد QTL از ۴ به ۱۰، میزان صحت با شیب زیادی افزایش و سپس تا ۴۰ QTL کاهش یافت (Honarvar and Rostami, 2014).

اثر متقابل توزیع QTL و وراثت پذیری: بر اساس شکل ۳ ملاحظه می‌شود که وقتی ژن‌های عمده اثر دارای توزیع نرمال بوده و وراثت پذیری ۰/۵۰ دارند، بیشترین مقدار صحت بدست می‌آید. در وراثت پذیری ۰/۱۰ و ۰/۵۰ مشاهده شد که صحت ارزیابی‌های ژنومی برای توزیع یکنواخت کمتر از توزیع نرمال بود، ولی صحت ارزیابی‌های ژنومی دو توزیع گاما و نرمال تقریباً یکسان بود. در وراثت پذیری ۰/۳۰، دو توزیع نرمال و یکنواخت دارای صحت ارزیابی ژنومی تقریباً برابری بودند و توزیع گاما دارای صحت کمتری بود.

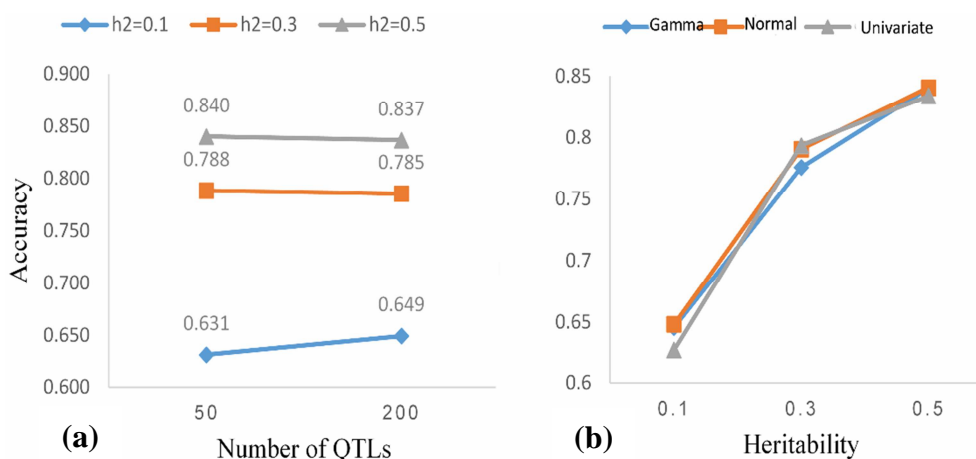


Fig. 3. Changes in accuracies of GEBVs for traits with different heritability estimates due to increment in the number of QTLs (a) and for traits with different distribution of QTLs due to increment in heritability (b)

شکل ۳- تغییر صحت ارزش‌های اصلاحی ژنومی برای صفات با وراثت پذیری متفاوت بر اثر افزایش تعداد QTL (a) و برای صفات با توزیع متفاوت QTL بر اثر افزایش وراثت پذیری (b)

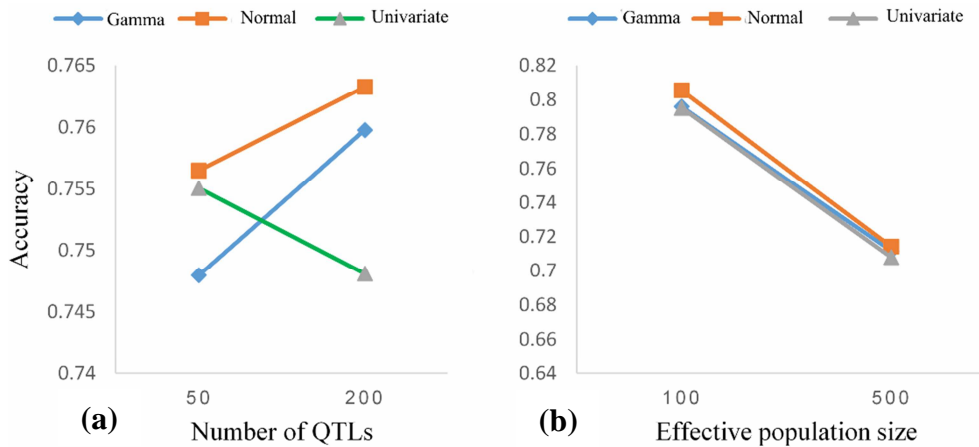


Fig. 4. Changes in accuracies of GEBVs for traits with different distribution of QTLs due to increment in the number of QTLs (a) and or due to increment in the effective population size (b)

شکل ۴- تغییر صحت ارزش‌های اصلاحی ژنومی برای صفات با توزیع متفاوت QTL بر اثر افزایش تعداد QTL (a) و یا بر اثر افزایش اندازه موثر جمعیت (b)

صحت با افزایش اندازه جمعیت مرجع کاهش یافت. در حالی که در هر دو اندازه موثر جمعیت (۱۰۰ و ۵۰۰)، بیشترین صحت برای توزیع نرمال مشاهده شد، ولی در عین حال بیشترین کاهش در میزان صحت ارزیابی‌های ژنومی به دلیل افزایش اندازه موثر جمعیت در توزیع نرمال مشاهده شد. بعد از توزیع نرمال، توزیع گاما بیشترین کاهش در میزان صحت ارزیابی‌های ژنومی به دلیل افزایش اندازه موثر جمعیت را متحمل شد (شکل ۴).

اثر متقابل روش برآورد آثار نشانگری و اندازه موثر جمعیت: شکل ۵ اثر متقابل بین دو عامل روش برآورد آثار نشانگری و اندازه موثر جمعیت را نشان می‌دهد. با توجه به این که با افزایش اندازه موثر جمعیت، طول قطعات کروموزومی که مشترکاً از جد مشترک به نتاج نسل‌های بعد به ارث می‌رسد کوتاه‌تر است (Daetwyler *et al.*, 2008)، انتظار می‌رود که با افزایش اندازه موثر جمعیت، در تراکم نشانگری یکسان، میزان صحت ارزیابی‌های ژنومی کاهش یابد، ولی نتایج ارائه شده در شکل ۵ نشان داد که میزان این کاهش در روش‌های مختلف، یکسان نبوده و با هم متفاوت بودند. در این تحقیق، کمترین کاهش در اثر افزایش اندازه موثر جمعیت مربوط به روش شبکه‌های عصبی و بیشترین کاهش متعلق به روش بیز B بود.

بیشترین صحت ارزیابی‌های ژنومی در $Ne = 100$ ، متعلق به روش RKHS و در $Ne = 500$ ، متعلق به روش

در یک مطالعه شبیه‌سازی (Uemoto *et al.*, 2015)، اثر دو نوع توزیع (گاما و واریانس برابر که طی آن فرض می‌شود همه QTLها در واریانس QTL سهم یکسانی دارند) و تعداد متفاوت QTL (۵۰، ۱۰۰، ۳۰۰، ۵۰۰، ۱۰۰۰ و ۲۰۰۰) بر صحت ارزیابی‌های ژنومی مورد بررسی قرار گرفت. نتایج حاصل وجود اثر متقابل بین دو عامل تعداد و توزیع QTL را نشان داد. وقتی ژن‌های عمده اثر از توزیع گاما تبعیت می‌کردند، با افزایش تعداد QTL از ۵۰ به ۱۰۰، صحت افزایش یافت، با افزایش تعداد QTL به ۳۰۰، صحت کاهش، با افزایش به ۵۰۰ و ۱۰۰۰ دوباره صحت افزایش و پس از آن با شیب اندکی کاهش یافت، ولی وقتی QTLها دارای توزیع واریانس برابر بودند، با افزایش تعداد QTL از ۵۰ به ۱۰۰، صحت با شیب تندی کاهش یافت، با افزایش تعداد QTL به ۳۰۰ و سپس ۵۰۰، صحت افزایش، سپس تا ۱۰۰۰ QTL تقریباً بدون تغییر و در نهایت با شیب اندکی افزایش یافت. در حالت کلی، توزیع گاما دارای صحت بیشتری بود (Uemoto *et al.*, 2015).

در یک مطالعه شبیه‌سازی دیگر، اثر متقابل بین این دو عامل به وضوح مورد تایید قرار گرفت (محمودی و همکاران، ۱۳۹۴). همچنین گزارش شده است که با افزایش تعداد QTL از ۱۰۰ به ۳۰۰ در هر دو توزیع یکنواخت و گاما، صحت GEBVها کاهش می‌یابد (Alanoshahr *et al.*, 2018).

اثر متقابل توزیع QTL و اندازه موثر جمعیت: در تحقیق حاضر در هر سه نوع توزیع یکنواخت، نرمال و گاما، میزان

نتیجه‌گیری کلی

وجود اثر متقابل بین عوامل بررسی شده در این تحقیق به وضوح موید این امر است که روش‌های آماری متفاوت حسب معماری‌های ژنتیکی مختلف صفات و ساختارهای متفاوت جمعیت، عملکرد یکسانی نداشته و برای شرایط مختلف باید از روش آماری مناسب استفاده شود. با افزایش وراثت پذیری صفات، بهبود عملکرد روش‌های نیمه پارامتری نسبت به روش‌های بیزی بیشتر بوده و همچنین با افزایش اندازه موثر جمعیت، کاهش صحت روش‌های نیمه پارامتری نسبت به روش‌های بیزی کمتر بود. در تحقیق حاضر در هر سه سطح وراثت پذیری پایین، متوسط و بالا، با افزایش اندازه موثر جمعیت، صحت برآوردهای ژنومی کاهش یافت. البته بیشترین کاهش صحت ارزیابی‌ها برای صفات با وراثت پذیری پایین و کمترین کاهش مربوط به صفات با وراثت پذیری بالا بود. این امر مبین آن است که وراثت پذیری بالا، تا حدودی می‌تواند بر مشکل کاهش LD به واسطه افزایش اندازه موثر جمعیت فائق آید.

بود که بر اساس نتایج تحقیق حاضر، نشان‌دهنده عملکرد بهتر این روش نیمه پارامتری در LD پایین است. در بررسی صحت سه روش بیز A، بیز B و بیز LASSO، با افزایش اندازه موثر جمعیت از ۵۰ به ۲۰۰، صحت ارزش‌های اصلاحی ژنومی افزایش یافت و بیشترین صحت در سناریویی حاصل شد که شامل کمترین تعداد QTL، بیشترین اندازه موثر جمعیت و روش بیز B بود (Alanoshahr *et al.*, 2018). در یک مطالعه شبیه‌سازی که صحت ارزیابی ژنومی دو روش بیز B و GBLUP را با در نظر گرفتن چند عامل به ویژه اندازه موثر جمعیت (۲۰۰، ۱۰۰۰ و ۲۰۰۰) مورد بررسی قرار داده بود، مشاهده شد که در حالت کلی، میزان صحت حاصل از روش بیز B بالاتر از روش GBLUP بود. در اندازه موثر جمعیت ۲۰۰، ۱۰۰۰ و ۲۰۰۰، میزان برتری روش بیز B به GBLUP به ترتیب ۰/۰۵۷، ۰/۰۸۵ و ۰/۰۶۹ بود که مبین وجود اثر متقابل بین دو عامل روش برآورد آثار نشانگری و اندازه جمعیت مرجع است (Daetwyler *et al.*, 2010).

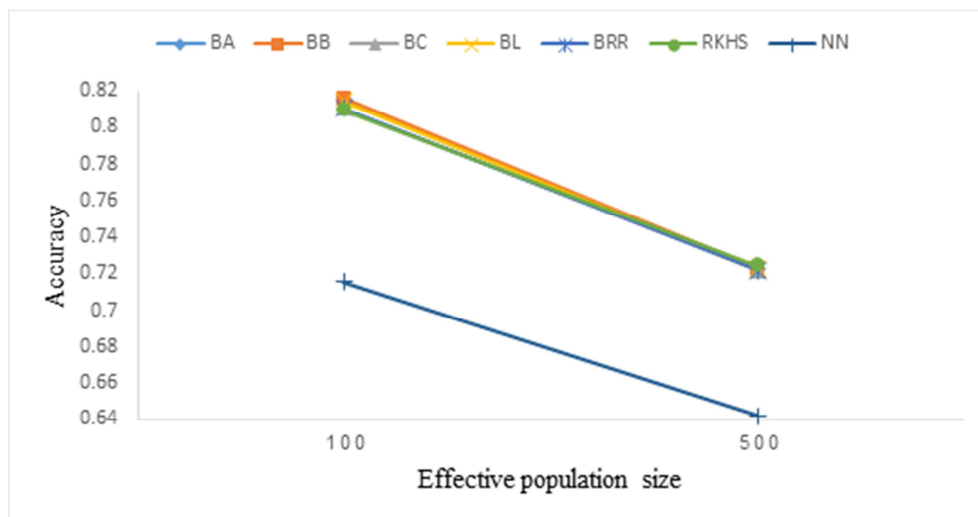


Fig. 5. Interaction effect of effective population size and estimation method of marker effects on the accuracy of genomic evaluations. BA: Bayes A, BB: Bayes B, BC: Bayes C, BL: Bayesian LASSO, BRR: Bayesian Ridge Regression, RKHS: Reproducing kernel Hilbert spaces, NN: Neural Networks

شکل ۵- اثر متقابل بین اندازه موثر جمعیت و روش برآورد آثار نشانگری بر صحت ارزیابی‌های ژنومی

فهرست منابع

- محمودی ن.، آیت الهی مهرجردی ا.، هنرور م.، و اسماعیلی زاده کشکوئیه، ع. ۱۳۹۴. مطالعه توزیع آماری آثار QTL بر صحت ارزش‌های اصلاحی ژنومی برآورد شده با روش Bayesian. پژوهش‌های علوم دامی ایران، ۷(۳): ۳۵۶-۳۶۳.
- Alanoshahr F., Rafat S. A., Imany-Nabiyyi R., Alijani S. and Robert Granie C. 2017. Effects of marker density, number of quantitative trait loci and heritability of trait on genomic selection accuracy. *Iranian Journal of Applied Animal Science*, 7(4): 595-601.
- Alanoshahr F., Rafat S. A., Imany-Nabiyyi R., Alijani S. and Robert Granie C. 2018. The impact of different genetic architectures on accuracy of genomic selection using three Bayesian methods. *Iranian Journal of Applied Animal Science*, 8(1): 53-59.
- Atefi A., Shadparvar A. A. and Ghavi Hossein-Zadeh N. 2018. Accuracy of genomic prediction under different genetic architectures and estimation methods. *Iranian Journal of Applied Animal Science*, 8(1): 43-52.
- Daetwyler H. D., Pong-Wong R., Villanueva B. and Woolliams J. A. 2010. The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, 185(3): 1021-1031.
- Daetwyler H., Villanueva B. and Woolliams J. A. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PloS One*, 3(10): e3395.
- Gianola D., Fernando R. L. and Stella A. 2006. Genomic assisted prediction of genetic value with semi-parametric procedures. *Genetics*, 173(3): 1761-1776.
- Gonzalez-Recio O., Gianola D., Long N., Weigel K. A., Rosa G. J. M. and Avendano S. 2008. Nonparametric methods for incorporating genomic information into genetic evaluations: an application to mortality in broilers. *Genetics*, 178: 2305-2313.
- Gorgani Firozjah N., Atashi H., Dadpasand M. and Zamiri M. 2014. Effect of marker density and trait heritability on the accuracy of genomic prediction over three generations. *Journal of Livestock Science and Technologies*, 2(2): 53:58.
- Habier D., Fernando R. L., Kizilkaya K. and Garrick D. J. 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12: 186.
- Hayes B. 2007. QTL mapping, mas, and genomic selection. A short-course. Department of Animal Science, Iowa State University. <https://www.ans.iastate.edu/files/page/files/notes.pdf>.
- Hoerl A. E. and Kennard R. W. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1): 55-67.
- Honarvar M. and Rostami M. 2013. Accuracy of genomic prediction using rr-blup and bayesian lasso. *European Journal of Experimental Biology*, 3(3): 42-47.
- Howard R., Carriquiry A. L. and Beavis W. D. 2014. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes, Genomes, Genetics*, 4(6): 1027-1046.
- Liu T., Qu H., Luo C., Shu D., Wang J., Lund M. S. and Su G. 2014. Accuracy of genomic prediction for growth and carcass traits in Chinese triple-yellow chickens. *BMC genetics*, 15(1): 110.
- Meuwissen M. H. E., Hayes B. J. and Goddard M. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4): 1819-1829.
- Muir W. M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *Journal of Animal Breeding and Genetics*, 124: 342-355.
- Park T. and Casella G. 2008. The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482): 681-686.
- Sargolzaei M. and Schenkel F. S. 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics*, 25(5): 680-681.
- Uemoto Y., Sasaki S., Kojima T., Sugimoto Y. and Watanabe T. 2015. Impact of QTL minor allele frequency on genomic evaluation using real genotype data and simulated phenotypes in Japanese Black cattle. *BMC Genetics*, 16: 134.
- Valente M. S. F., Viana J. M. S., Resende M. D. V., Silva F. F. and Lopes M. T. G. 2016. Genomic selection for plant breeding with different population structures. *Pesquisa Agropecuária Brasileira*, 51(11): 1857-1867.
- Weber K., Thallman R., Keele J., Snelling W., Bennett G., Smith T., McDanel T., Allan M., Van Eenennaam A. and Kuehn L. 2012. Accuracy of genomic breeding values in multibreed beef cattle populations derived from deregressed breeding values and phenotypes. *Journal of Animal Science*, 90(12): 4177-4190.



Research paper

Accuracy of genomic evaluation considering the interaction effect between estimation method of marker effects, population structure, and genetic architecture of the trait

A. Atefi^{1*}, A. A. Shadparvar², N. Ghavi Hossein-Zadeh²

1. Former Ph.D. Student in Animal Breeding and Genetics, Department of Animal Science, Faculty of Agricultural Sciences, University of Guilan, Rasht, Iran

2. Professor, Department of Animal Science, Faculty of Agricultural Sciences, University of Guilan, Rasht, Iran

(Received: 11-04-2020 – Accepted: 02-07-2020)

Abstract

This study aimed to investigate the interaction effects between marker effect estimation methods, population structure, and genetic architecture of the trait on the accuracy of genomic evaluations. A reference population with two different effective population sizes (100 and 500) was simulated using the QMSim software. 500 markers and two different numbers of quantitative trait loci or QTLs (50 and 200) were distributed randomly through the genome including a 100 cM chromosome. In this study, three traits with different heritabilities (0.1, 0.3, and 0.5) were simulated. The genomic breeding values were predicted using Bayesian ridge regression, Bayes A, Bayes B, Bayes C, Bayesian LASSO, Reproducing kernel Hilbert space, and neural networks methods. Through the three heritabilities, as the effective population size increased, the accuracy of genomic evaluation decreased with different trends. As the number of QTLs increased, the accuracy of low heritability trait increased, but the accuracy of medium and high heritability traits decreased. Similarly, as the number of QTLs increased, the accuracy of the trait with normal distributed QTLs increased, but the accuracy of traits with gamma and univariate distributed QTLs decreased. For all types of QTL distributions, the increment of effective population size decreased the accuracy of genomic evaluations. The results of this study clearly showed the interaction effects between markers effect estimation methods, population structure, and genetic architecture of the trait on the accuracy of genomic evaluations.

Keywords: Interaction effect, Genomic selection, Population structure, Prediction accuracy, Trait genetic architecture

*Corresponding author: abbas.atefi@gmail.com