

اثر تعداد و توزیع اثرات QTL بر صحت پیش‌بینی‌های ژنومی برخی روش‌های آماری در یک صفت آستانه‌ای

صاحب فروتنی فر^{*۱}

تاریخ دریافت: ۱۳۹۵/۰۴/۱۱

تاریخ پذیرش: ۱۳۹۵/۰۸/۱۱

چکیده

هدف این مطالعه استفاده از شبیه‌سازی تصادفی برای محاسبه ارزش‌های اصلاحی ژنومی در سناریوهایی با تعداد و توزیع مختلف آماری برای اثرات QTL در یک صفت آستانه‌ای با استفاده از روش‌های بی‌زی A، B، C، L و R بود. برای شبیه‌سازی سناریوهای مختلف، ابتدا یک نسل پایه با تعداد ۱۰۰ حیوان ایجاد شد که به مدت ۵۰ نسل با هم آمیزش تصادفی داشتند. جمعیت مرجعی با اندازه ۱۰۰۰ حیوان از تلاقی تصادفی افراد نسل ۵۰ ایجاد شد و برای این حیوانات یک صفت آستانه‌ای شبیه‌سازی شد. از تلاقی تصادفی افراد جمعیت مرجع، جمعیت تأیید ایجاد شد که فاقد اطلاعات فنوتیپی بودند. برای هر حیوان، ژنومی به طول ۱۰ مورگان شبیه‌سازی شد و ۱۰۰۰۰ نشانگر با فواصل یکسان بر روی آن قرار گرفت. در هر سناریو QTLهایی با توزیع مختلف برای اثرات آن به صورت تصادفی بر روی ژنوم پخش شد. اثرات نشانگرها در جمعیت مرجع به وسیله روش‌های مختلف بی‌زی برآورد گردید و ارزش‌های اصلاحی ژنومی حیوانات فاقد فنوتیپ جمعیت تأیید با استفاده از این اثرات محاسبه شد. صحت ارزش‌های اصلاحی مبنای مقایسه سناریوهای مختلف این تحقیق قرار گرفت. در روش‌های بی‌زی مورد بررسی به استثنای روش بی‌زی R و L که در همه سناریوها صحت مشابهی داشتند، در بین سایر روش‌های بی‌زی، بی‌زی B با تعداد ده QTL و توزیع گاما بالاترین صحت ارزش‌های اصلاحی را داشت. با افزایش تعداد QTLها صحت ارزش‌های اصلاحی برآوردی با استفاده از روش‌های بی‌زی A، B و C کمتر شد، به گونه‌ای که برای تعداد ۱۵۰ QTL و بالاتر صحت ارزش‌های اصلاحی آنها تقریباً برابر با روش بی‌زی R و L شد. به طور کلی نتایج این پژوهش نشان داد که روش‌های آماری مقایسه‌شده در معماری متفاوت صفت نتایج متفاوتی ارائه می‌دهند، ولی می‌توان برای صفات آستانه‌ای که با تعداد کمی ژن کنترل می‌شوند یا دارای یک یا چند ژن با اثر عمده هستند، از روش‌هایی مانند بی‌زی B استفاده کرد.

واژه‌های کلیدی: انتخاب ژنومی، روش‌های بی‌زی، صفت آستانه‌ای، QTL.

مقدمه

BLUP، روش‌های بی‌زی و روش‌های ناپارامتری گزارش شدند. در این بین، روش‌های بی‌زی به دلیل اینکه به خوبی می‌توانند خود را با ساختار داده‌ها که در آن تعداد نشانگرها از تعداد داده‌ها بسیار بیشتر است وفق دهند و اینکه این روش‌ها یک روش طبیعی برای در نظر گرفتن عدم قطعیت در مورد همه نانشناخته‌های موجود در مدل هستند، توجه بیشتری به سمت خود جلب کرده‌اند. همچنین وقتی روش‌های بی‌زی با قدرت و انعطاف پذیری زنجیره مارکف مونت کارلو (MCMC^۲) همراه شوند می‌توان آنها را تقریباً در همه مدل‌های آماری پارامتری استفاده کرد (۱۰).

تفاوت اساسی روش‌های آماری در فرض‌های آنها و همچنین نحوه استفاده آنها در انتخاب ژنومی می‌باشد. روش‌های BLUP، بی‌زی R، بی‌زی A و بی‌زی L فرض می‌کنند که نشانگرها به ترتیب دارای توزیع

مطالعات شبیه‌سازی و تجربی برتر بودن صحت ارزیابی‌های ژنومی نسبت به ارزیابی‌های سنتی بر مبنای اطلاعات شجره را تأیید می‌کنند. علاوه بر عوامل متعددی مانند اندازه و ساختار جمعیت مرجع، درجه خویشاوندی کاندیداهای انتخاب با افراد جمعیت مرجع و معماری ژنتیکی صفات، روش مورد استفاده برای برآورد آثار نشانگرها بر برتری صحت ارزیابی‌های ژنومی نسبت به روش‌های سنتی موثر است (۴). برای برآورد اثرات نشانگرها روش‌های آماری متعددی مانند

۱- استادیار ژنتیک و اصلاح دام، گروه علوم دامی، دانشکده کشاورزی، دانشگاه رازی کرمانشاه.

*- نویسنده مسئول: (Email: s.foroutanifar@razi.ac.ir)

DOI: 10.22067/ijasr.v4i1.57206

طول یکسان (یک مورگان) شبیه سازی شد و تعداد ۱۰۰۰ نشانگر SNP در فواصل مساوی بر روی هر کروموزوم قرار گرفت. پنج حالت مختلف ۱۰، ۵۰، ۱۵۰، ۵۰۰ و ۲۰۰۰ برای تعداد QTL‌های مؤثر بر صفات در نظر گرفته شد و در هر سناریو QTL‌ها به صورت تصادفی بر روی کل سطح ژنوم قرار گرفتند. اثر جایگزینی QTL‌ها بسته به سناریوی مورد مطالعه از یکی از پنج توزیع نرمال، یکنواخت، t ، گاما و لاپلاس با میانگین صفر و واریانس یک گرفته شد. فرض شد که QTL‌های در نظر گرفته شده تمام واریانس ژنتیکی صفت را در بر می‌گیرند و ارزش‌های اصلاحی واقعی افراد از مجموع آثار این QTL‌ها با توجه به ژنوتیپ آنها محاسبه شد. در این مطالعه اثرات غیر افزایشی در نظر گرفته نشد و واریانس ژنتیکی کل برابر با واریانس ژنتیکی افزایشی بود. برای همه حیوانات جمعیت مرجع یک ارزش فنوتیپی شبیه سازی شد. برای شبیه سازی فنوتیپ یک اثر باقی‌مانده از توزیع نرمال با میانگین صفر و واریانس باقی‌مانده چهار به منظور رسیدن به توارث پذیری ۰/۲، به ارزش اصلاحی واقعی افراد اضافه شد. به منظور ایجاد فنوتیپ‌های آستانه‌ای (صفر و یک) برای صفت مورد نظر، یک چارک بالایی فنوتیپ‌ها برابر یک و سه چارک پایینی برابر با صفر قرار داده شد. ژنوتیپ‌ها و ارزش‌های اصلاحی واقعی برای جمعیت تأیید و جمعیت مرجع شبیه سازی شد ولی فنوتیپ‌های آستانه‌ای فقط برای حیوانات نسل مرجع برای برآورد آثار نشانگرها در نظر گرفته شد.

برای برآورد اثرات نشانگرها با استفاده از روش‌های بیزی بسته BGLR (۱۲) نرم افزار R مورد استفاده قرار گرفت. برای صفات آستانه‌ای فرض می‌شود که این صفات داری یک متغیر پشت صحنه پیوسته هستند. مدل کلی مورد استفاده برای برآورد اثرات نشانگرها در جمعیت مرجع به صورت زیر بود.

$$l_i = \mu + \sum_{j=1}^n x_{ij}\beta_j + e_i \quad (1)$$

که در این مدل l_i متغیر پشت صحنه فرد i در جمعیت مرجع، μ میانگین کل، x_{ij} ژنوتیپ نشانگر j برای فرد i ، اثر جایگزینی آلی نشانگر j و e_i اثر باقی‌مانده برای حیوان i بود. در این بسته برای ارتباط متغیر پشت صحنه با فنوتیپ‌های دو دسته‌ای از تابع زیر استفاده می‌شود.

$$p(y_i = 1) = \Phi \left(\mu + \sum_{j=1}^n x_{ij}\beta_j + e_i \right) \quad (2)$$

که در اینجا $\Phi(\cdot)$ تابع توزیع تجمعی استاندارد یا تابع ربط پروبیت است (۱۳).

پنج روش مختلف بیزی بر اساس این مدل کلی برای برآورد

نرمال، نرمال، t و لاپلاس هستند (۳). توزیع اثرات QTL‌های مؤثر بر صفات هنوز ناشناخته است. در پژوهشی توزیع گاما برای اثرات QTL‌ها در حیوانات اهلی گزارش شد (۷) و این توزیع در بیشتر مطالعات شبیه سازی برای مدل کردن آثار QTL‌ها استفاده شده است. نتایج حاصل از پیش‌بینی‌های ژنومی با استفاده از داده‌های واقعی در حیوانات اهلی توزیع نرمال را برای اثرات QTL‌ها پیشنهاد می‌کنند (۹). با این حال تعداد و توزیع اثرات QTL‌ها و به طور کلی معماری ژنتیکی صفات متفاوت است. بخش عمده مطالعات انجام شده در انتخاب ژنومی بیشتر بر صفات پیوسته متمرکز بوده‌اند، با این حال بسیاری از صفات دارای اهمیت اقتصادی در حیوانات اهلی مانند چند قلو زایی، سخت زایی، زنده مانی و صفات تیپ دارای توزیع ناپیوسته هستند (۵). بنابراین هدف این مطالعه استفاده از شبیه سازی تصادفی برای در نظر گرفتن تعداد متفاوت QTL و توزیع‌های مختلف گاما، نرمال، یکنواخت، t و لاپلاس برای اثرات QTL در یک صفت آستانه‌ای و استفاده از روش‌های بیزی A، B، C، L و R برای محاسبه ارزش‌های اصلاحی ژنومی و مقایسه حساسیت روش‌ها به برقرار نبودن فرض‌های آنها برای توزیع اثرات نشانگرها در پیش‌بینی‌های ژنومی بود.

مواد و روش‌ها

صفتی که در این مطالعه در نظر گرفته شد، صفتی آستانه‌ای با توارث پذیری ۰/۲ بود. در سناریوهای مختلف برای این صفت پنج تعداد ۱۰، ۵۰، ۱۵۰، ۵۰۰ و ۲۰۰۰ QTL و پنج توزیع آماری نرمال، یکنواخت، t ، گاما و لاپلاس برای اثرات QTL‌ها شبیه سازی شد. اثرات QTL‌ها با استفاده از پنج روش بیزی A، B، C، L و R برآورد گردید. تأثیر تعداد و توزیع اثرات QTL‌های مؤثر بر صفت بر برآورد ارزش‌های اصلاحی با استفاده از این روش‌ها مورد ارزیابی قرار گرفت. در مجموع در تحقیق حاضر ۱۲۵ سناریوی مختلف مورد ارزیابی قرار گرفت و برای بررسی این سناریوها از شبیه سازی تصادفی در محیط نرم افزار آماری R نسخه ۳.۱.۱ استفاده شد. با توجه به اینکه شبیه سازی به صورت مدل تصادفی بود، هر سناریو ۴۰ بار تکرار شد و میانگین آنها به عنوان نتایج گزارش شد.

شبیه سازی با یک جمعیت پایه با تعداد ۱۰۰ حیوان شروع شد و این حیوانات جهت ایجاد عدم تعادل پیوستگی بین نشانگرها، ۵۰ نسل با هم آمیزش تصادفی داشتند. نسل‌ها به صورت نامتداخل بودند و تعداد افراد در جمعیت پایه و نسل‌های بعدی ثابت بود. از آمیزش تصادفی افراد نسل ۵۰، جمعیت مرجع با تعداد ۱۰۰۰ حیوان ایجاد شد و از آمیزش تصادفی افراد جمعیت مرجع، جمعیت تأیید با تعداد ۱۰۰۰ حیوان ایجاد شد.

برای هر حیوان ژنومی با طول ۱۰ مورگان و تعداد ۱۰ کروموزوم با

روش بیزی C: در این روش π درصد نشانگرها دارای اثر صفر و $1 - \pi$ درصد نشانگرها مانند روش بیزی R دارای توزیع یکسان هستند و از توزیع نرمال گرفته می‌شوند. در این روش مانند روش بیزی B میزان π برابر با ۰/۹۵ در نظر گرفته شد. روش بیزی L: در این روش فرض می‌شود که آثار نشانگرها یکسان و دارای توزیع لاپلاس هستند. برای راحتی کار در این روش نشانگرها از یک توزیع نرمال $N(0, \tau_j^2)$ گرفته می‌شوند، که τ_j^2 دارای توزیع دوجمله‌ای با پارامتر λ است که در این مطالعه مقدار λ با استفاده از فرمول $\sqrt{2 \frac{1-h^2}{h^2} MSx}$ محاسبه شد (۳).

در همه روش‌های بیزی جهت به دست آوردن نمونه‌های توزیع پسین پارامترها از یک نمونه‌گیری گیبس با تعداد تکرار ۳۰۰۰۰ استفاده شد که ۱۰۰۰۰ نمونه اول به‌عنوان دوره گرم کردن^۱ در نظر گرفته شد. پس از برآورد اثر نشانگرها معادله $GEBV_{ij} = Z_i \hat{m}_j$ برای پیش‌بینی ارزش‌های اصلاحی ژنومی در جمعیت تأیید استفاده شد. در این فرمول $GEBV_{ij}$ ارزش اصلاحی ژنومی حیوان i برای صفت j ، Z_i ردیف مربوط به حیوان i در ماتریس ژنوتیپ نشانگرها و \hat{m}_j اثر نشانگرها برای صفت j است (۱۱).

برای اندازه‌گیری میزان عدم تعادل پیوستگی بین نشانگرها مجاور از معیار r^2 و معادله $r^2 = \frac{D^2}{p_1 p_2 q_1 q_2}$ استفاده شد. در این معادله D ضریب عدم تعادل پیوستگی بین نشانگرها و p_1, p_2, q_1 و q_2 فراوانی‌های آلی در نشانگرهای مجاور هستند.

پس از محاسبه ارزش‌های اصلاحی ژنومی، برای مقایسه حساسیت روش‌های بیزی مورد استفاده به تغییر تعداد و توزیع اثرات QTL، سناریوهای مختلف با استفاده از صحت ارزش‌های اصلاحی ژنومی، که همبستگی بین ارزش اصلاحی ژنومی برآورد شده و ارزش اصلاحی واقعی حیوانات است، با هم مقایسه شدند. برای آزمون معنی‌داری تفاوت بین میانگین سناریوهای مختلف از آزمون t استفاده شد.

نتایج

آمیزش تصادفی افراد نسل پایه به مدت ۵۰ نسل به دلیل اندازه مؤثر کوچک جمعیت و تحت تأثیر رانش ژنتیکی منجر به عدم تعادل پیوستگی بین نشانگرها و QTLها (آماره r^2) به میزان ۰/۲۱ شد. صحت ارزش‌های اصلاحی حاصل از برآورد روش‌های مختلف بیزی در تعداد و توزیع‌های متفاوت QTLها در جدول ۱ آمده است.

روش بیزی A: در این روش بالاترین صحت ارزش‌های اصلاحی ژنومی مربوط به توزیع گاما، t و لاپلاس و تعداد QTL ۱۰ به میزان ۰/۵۸ و کمترین صحت مربوط به توزیع آماری یکنواخت به میزان

اثرات نشانگرها استفاده شده است. تفاوت اصلی این روش‌ها در توزیع پیشین اثرات نشانگرها است. طبق تئوری بیز (معادله ۳)، مدل‌های بیزی به توزیع پیشین و توزیع داده‌ها جهت تشکیل توزیع پسین نیاز دارند که میانگین توزیع پسین به‌عنوان برآوردی از پارامترهای موجود در مدل در نظر گرفته شد.

$$p(\theta|y) = \frac{p(y|\theta) \cdot p(\theta)}{p(y)} \propto p(y|\theta) \cdot p(\theta) \quad (3)$$

در رابطه فوق $p(y|\theta)$ توزیع داده‌ها یا تابع درست‌نمایی به شرط پارامترهای موجود در مدل، $p(\theta)$ توزیع پیشین پارامترها و $p(\theta|y)$ توزیع پسین θ به شرط بردار مشاهدات است. بردار θ شامل اثر میانگین، بردار اثر نشانگرها، واریانس نشانگرها و واریانس خطا است. توزیع پسین بردار پارامترها با استفاده از نمونه‌گیری گیبس به دست می‌آید. در روش‌های بیزی همه پارامترهای موجود در مدل به صورت تصادفی با یک توزیع پیشین در نظر گرفته می‌شوند. برای همه روش‌ها توزیع پیشین برای μ توزیع نرمال با میانگین صفر و واریانس $1 \cdot 10^{10}$ بود و توزیع پیشین برای واریانس باقی‌مانده از توزیع مربع کای معکوس با درجه آزادی ۵ انتخاب شدند.

روش بیزی R: در این روش از توزیع نرمال با میانگین صفر و واریانس یکسان برای همه نشانگرها، σ_m^2 استفاده شده است، که σ_m^2 دارای توزیع کای مربع معکوس با پارامترهای v و s^2 بود. درجه آزادی در این مدل برابر با ۴/۲ در نظر گرفته شد و s^2 با استفاده از فرمول $\frac{\sigma_a^2}{MSx}$ محاسبه شد که در این فرمول σ_a^2 واریانس افزایشی صفت، $MSx = \frac{\sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2}{n}$ و \bar{x}_j میانگین ژنوتیپ در نشانگر j است (۳).

روش بیزی A: در این روش فرض می‌شود که همه نشانگرها دارای اثر هستند ولی واریانس هر نشانگر در کل ژنوم با یکدیگر متفاوت است. در این مدل توزیع پیشین حاشیه‌ای آثار نشانگرها توزیع t است و برای ساده‌سازی محاسباتی اثر هر نشانگر از توزیع نرمال $N(0, \sigma_m^2)$ گرفته می‌شود. که در اینجا σ_m^2 واریانس ناشناخته برای هر نشانگر است و توزیع پیشین آن از توزیع کای مربع معکوس با پارامترهای v و s^2 انتخاب شدند. این روش آشیانه‌ای منجر به توزیع t برای آثار نشانگرها می‌شود. مقدار اولیه برای v ۴/۲ در نظر گرفته شد و پارامتر s^2 با استفاده از فرمول $\frac{(v-2) \sigma_a^2}{v MSx}$ محاسبه شد (۳).

روش بیزی B: در این روش π درصد نشانگرها دارای اثر صفر است و $1 - \pi$ درصد نشانگرها مانند روش بیزی A دارای توزیع پیشین t هستند و اثر آنها مانند روش بیزی A برآورد می‌شود. در این روش پارامتر s^2 با استفاده از فرمول $\frac{1}{1-\pi} \frac{(v-2) \sigma_a^2}{v MSx}$ محاسبه شد. در این مطالعه π برابر با ۰/۹۵ در نظر گرفته شد (۳).

و ۵۰ QTL محسوس تر بود. در این روش نیز مانند روش بیزی A در همه توزیع‌های آماری بالاترین صحت مربوط به تعداد ۱۰ QTL بود ($p < 0.01$)، با افزایش تعداد QTL‌ها به ۵۰ صحت ارزش‌های اصلاحی کاهش یافت و تفاوتی بین صحت ارزش‌های اصلاحی (دامنه ۰/۴۶ تا ۰/۴۸) در تعداد ۱۵۰، ۵۰۰ و ۲۰۰۰ در توزیع‌های مختلف مشاهده نشد ($p > 0.05$). در تعداد ۱۰ و ۵۰ QTL کمترین صحت پیش‌بینی‌ها همانند روش بیزی A مربوط به توزیع یکنواخت بود ($p < 0.01$).

۰/۴۹ بود. در همه توزیع‌ها بالاترین صحت مربوط به تعداد ۱۰ QTL بود ($p < 0.01$). برای تعداد QTL ۵۰ و ۱۵۰ بالاترین صحت مربوط به توزیع گاما به ترتیب به میزان ۰/۵۲ و ۰/۴۹ نسبت به سایر توزیع‌ها بود که صحت‌های در دامنه ۰/۴۸-۰/۴۶ داشتند. برای تعداد ۵۰۰ و ۲۰۰۰ QTL بین توزیع‌های مختلف آماری تفاوت معنی‌داری وجود نداشت ($p > 0.05$). روش بیزی B: صحت ارزش‌های اصلاحی در این روش در همه سناریوها برابر و یا بیشتر از بقیه روش‌ها بود و این برتری در تعداد ۱۰

جدول ۱- میانگین \pm انحراف معیار صحت ارزش‌های اصلاحی ژنومی پیش‌بینی شده با استفاده از پنج روش بیزی برای تعداد و توزیع‌های مختلف آثار QTL‌ها در صفت آستانه‌ای با توارث پذیری ۰/۲

Table 1- Mean \pm SD of genomic breeding values accuracy predicted by five Bayesian methods for different numbers and distributions of a categorical trait with heritability=0.2

توزیع QTL‌ها Distribution of QTLs	تعداد QTL‌ها Number of QTLs	روش برآورد آثار نشانگرها Method for estimation of markers effects				
		بیزی A Bayes A	بیزی B Bayes B	بیزی C Bayes C	بیزی L Bayes L	بیزی R Bayes R
		Normal	10	0.52 \pm 0.1	0.59 \pm 0.1	0.51 \pm 0.07
گاما Gamma	50	0.46 \pm 0.06	0.48 \pm 0.05	0.49 \pm 0.06	0.47 \pm 0.06	0.46 \pm 0.06
	150	0.47 \pm 0.06	0.47 \pm 0.06	0.47 \pm 0.04	0.47 \pm 0.06	0.47 \pm 0.06
	500	0.47 \pm 0.06	0.46 \pm 0.06	0.46 \pm 0.06	0.46 \pm 0.05	0.46 \pm 0.05
	2000	0.47 \pm 0.05	0.46 \pm 0.05	0.47 \pm 0.05	0.46 \pm 0.06	0.46 \pm 0.04
	یکنواخت Uniform	10	0.58 \pm 0.1	0.68 \pm 0.1	0.56 \pm 0.09	0.46 \pm 0.08
t-student	50	0.52 \pm 0.1	0.56 \pm 0.1	0.51 \pm 0.08	0.48 \pm 0.06	0.46 \pm 0.06
	150	0.49 \pm 0.07	0.48 \pm 0.07	0.52 \pm 0.06	0.47 \pm 0.05	0.46 \pm 0.06
	500	0.47 \pm 0.05	0.46 \pm 0.08	0.47 \pm 0.05	0.47 \pm 0.06	0.46 \pm 0.06
	2000	0.46 \pm 0.05	0.46 \pm 0.06	0.46 \pm 0.04	0.47 \pm 0.06	0.48 \pm 0.05
	لاپلاس Laplace	10	0.49 \pm 0.06	0.53 \pm 0.08	0.51 \pm 0.06	0.48 \pm 0.06
50		0.46 \pm 0.06	0.48 \pm 0.06	0.47 \pm 0.06	0.46 \pm 0.06	0.48 \pm 0.06
150		0.46 \pm 0.06	0.46 \pm 0.05	0.46 \pm 0.06	0.46 \pm 0.06	0.48 \pm 0.05
500		0.46 \pm 0.06	0.46 \pm 0.06	0.46 \pm 0.05	0.46 \pm 0.06	0.47 \pm 0.06
2000		0.46 \pm 0.07	0.46 \pm 0.06	0.46 \pm 0.06	0.46 \pm 0.05	0.46 \pm 0.05
لاپلاس Laplace	10	0.58 \pm 0.1	0.64 \pm 0.1	0.55 \pm 0.09	0.46 \pm 0.06	0.46 \pm 0.05
	50	0.5 \pm 0.08	0.51 \pm 0.06	0.48 \pm 0.06	0.47 \pm 0.06	0.47 \pm 0.04
	150	0.47 \pm 0.07	0.51 \pm 0.1	0.48 \pm 0.06	0.46 \pm 0.06	0.47 \pm 0.06
	500	0.46 \pm 0.05	0.48 \pm 0.07	0.48 \pm 0.05	0.47 \pm 0.06	0.47 \pm 0.05
	2000	0.47 \pm 0.05	0.46 \pm 0.06	0.46 \pm 0.05	0.46 \pm 0.05	0.48 \pm 0.06
لاپلاس Laplace	10	0.58 \pm 0.1	0.63 \pm 0.1	0.54 \pm 0.07	0.47 \pm 0.06	0.48 \pm 0.06
	50	0.48 \pm 0.06	0.52 \pm 0.07	0.48 \pm 0.05	0.47 \pm 0.06	0.47 \pm 0.06
	150	0.47 \pm 0.05	0.48 \pm 0.06	0.47 \pm 0.07	0.46 \pm 0.06	0.48 \pm 0.05
	500	0.46 \pm 0.06	0.48 \pm 0.06	0.46 \pm 0.05	0.46 \pm 0.06	0.47 \pm 0.06
	2000	0.47 \pm 0.05	0.46 \pm 0.05	0.46 \pm 0.05	0.48 \pm 0.06	0.47 \pm 0.06

توزیع‌های صحتی به در دامنه ۰/۴۶ تا ۰/۴۸ داشتند. برتری روش‌های بیزی A، بیزی B و بیزی C نسبت به بیزی L و بیزی R در تعداد پایین QTL بود ($p < 0.01$) و با افزایش تعداد QTL‌ها به ۱۵۰ و بیشتر این برتری کاملاً از بین رفت. به گونه‌ای که برای تعداد

روش بیزی C: در روش بیز C هم بالاترین صحت مربوط به تعداد 10 QTL و توزیع‌های گاما به میزان ۰/۵۶ بود که این صحت کمتر از روش بیزی B به میزان ۰/۶۸ بود ($p < 0.01$). در تعداد QTL ۵۰ هم بالاترین صحت مربوط به توزیع گاما بود ($p < 0.01$) و سایر

A و بی‌زی C کم‌رنگ شده و در نهایت از بین می‌رود. میزان حساسیت روش‌های مختلف نسبت به توزیع آثار QTL‌ها از تعداد QTL‌ها کمتر بود و داشتن توزیع اثرات QTL مشابه با فرض روش‌ها برای آثار نشانگرها، مزیتی برای روش‌های مختلف نیست. به‌عنوان مثال داشتن توزیع‌های نرمال و t برای آثار QTL‌ها منجر به افزایش صحت برای به ترتیب روش‌های بی‌زی C و بی‌زی A نسبت به سایر توزیع‌ها نشد و جالب آنکه صحت روش بی‌زی C در توزیع t و تعداد QTL ۱۰ بیشتر از توزیع نرمال با تعداد QTL ۱۰ بود. به نظر می‌رسد آنچه برای روش‌های مختلف بی‌زی مهم است ابتدا تعداد QTL‌ها و بعد از آن میزان و متفاوت بودن اثر QTL‌ها است. هر چه تعداد QTL‌ها کمتر باشد و در بین آنها QTL‌هایی با اثر زیاد وجود داشته باشد منجر به برتری روش‌هایی مانند بی‌زی B می‌شود که هر نشانگر واریانس مخصوص خود را دارد و با شدت بیشتری اثرات برآوردی کوچک را به سمت صفر منقبض می‌کند. به‌عنوان مثال توزیع یکنواخت که در آن سهم QTL‌ها در واریانس صفت یکسان است نسبت به سایر توزیع‌ها در تعداد یکسان QTL منجر به صحت‌های کمتری شده است. در مقابل توزیع گاما که نسبت به سایر توزیع‌ها تراکم بیشتری در نزدیکی صفر دارد، گرفتن اثرات QTL‌ها از این توزیع منجر به ایجاد تعداد کمتری QTL با اثر متوسط و زیاد و تعداد بیشتری QTL با اثر کم شده و به دنبال آن باعث افزایش صحت ارزش‌های اصلاحی در روش بی‌زی B (به‌خصوص اگر تعداد QTL‌ها کم باشد) می‌شود.

همان‌طور که نتایج نشان دادند روش‌های مختلف نمی‌توانند در همه شرایط برتری خود نسبت به سایر روش‌ها را حفظ کنند و بنابراین نمی‌توان روش‌های مختلف را بدون در نظر تعداد و توزیع اثرات QTL‌ها با هم مقایسه کرد. به‌عنوان نمونه روش بی‌زی B که بالاترین صحت را برای صفتی با تعداد ۱۰ QTL مؤثر بر صفت داشت وقتی تعداد QTL‌ها افزایش یافت برتری خود نسبت به سایر روش‌ها را از دست داد. مطالعات شبیه‌سازی متعددی برتری روش‌های بی‌زی نسبت به روش BLUP گزارش کردند (۱، ۱۱ و ۱۶)، که به نظر می‌رسد یکی از دلایل این برتری در نظر گرفتن تعداد محدود QTL در شبیه‌سازی صفات در این مطالعات بود. در تأیید این نتایج وانگ و همکاران (۱۸) با مقایسه روش‌های مختلف بی‌زی برای صفات آستانه‌ای مشاهده کردند که روش‌های بی‌زی B و بی‌زی C نسبت به تعداد QTL حساس بودند و با افزایش تعداد QTL‌ها از ۲۰ به ۵۰۰ صحت ارزیابی‌ها به شدت کاهش یافت. همچنین دات‌وایلر و همکاران (۲) با مقایسه روش‌های بی‌زی B و BLUP مشاهده کردند که صحت ارزش‌های اصلاحی روش BLUP در تعداد متفاوت QTL ثابت است ولی روش بی‌زی B در حالتی که تعداد اندکی QTL صفت را تحت تأثیر قرار می‌دهند، صحت‌های بالاتری داشت. آنها همچنین با مقایسه توزیع‌های لاپلاس و نرمال برای اثر QTL‌ها، گزارش

QTL ۱۵۰ تا ۲۰۰۰ همه توزیع‌ها صحتی در دامنه ۰/۴۶ تا ۰/۴۹ داشتند.

روش‌های بی‌زی L و بی‌زی R: در این روش‌ها همان‌طور که در جدول ۱ مشخص است در همه سناریوها صرف نظر از تعداد QTL‌ها و توزیع آثار QTL‌ها صحت‌های تقریباً یکسان و به میزان ۰/۴۸-۰/۴۶ مشاهده شد و این روش‌ها حساسیتی به تغییر تعداد و توزیع اثرات QTL‌ها نداشتند ($p > 0.05$).

بحث

در روش انتخاب ژنومی معمولاً تعداد نشانگرها خیلی بیشتر از تعداد افراد است و در نتیجه در مدل‌های رگرسیونی مورد استفاده برای برآورد اثرات نشانگرها تعداد متغیرهای مستقل خیلی بیشتر از تعداد مشاهدات است. برای حل این مشکل معمولاً برای روش‌های مختلف انتخاب نشانگرها پیش می‌آید، که این انتخاب یا با کنار گذاشتن نشانگرهای غیر مهم و یا با منقبض کردن اثرات آنها به سمت صفر اتفاق می‌افتد. در روش‌های بی‌زی مکانیسم القای انقباضی، به وسیله تعیین کردن توزیع مناسب پیشین برای ضرایب رگرسیونی در مدل وارد می‌شود و تفاوت روش‌های مختلف بی‌زی در همین توزیع پیشین نشانگرها و میزان انقباض اثرات نشانگرها به سمت صفر است (۳). در روش بی‌زی R و L که به ترتیب دارای توزیع پیشین نرمال و لاپلاس برای آثار نشانگرها هستند، میزان انقباض اثرات نشانگرها به سمت صفر در بین همه نشانگرها یکسان است. می‌توان ثابت کرد که برآوردهای حاصل از روش بی‌زی R بهترین پیش‌بینی ناریب خطی (BLUP) اثرات نشانگرها می‌باشد. این روش وقتی برخی نشانگرها در نزدیکی QTL بوده و یا دارای اثر زیاد هستند و برخی نشانگرها در نزدیکی آنها QTL نباشد و یا اثر آنها کوچک باشد، مناسب نیست. توزیع t (توزیع پیشین برای روش بی‌زی A) نسبت به توزیع نرمال دارای تراکم بیشتر در نقطه صفر که منجر به انقباض قویتر اثرات برآوردی نشانگرهای با اثر کوچک به سمت صفر و دنباله ضخیم‌تر است که باعث انقباض کمتر برآورد نشانگرهای با اثر زیاد می‌شود. در روش‌های بی‌زی B و بی‌زی C، π درصد از نشانگرها دارای اثر صفر هستند و $1-\pi$ درصد نشانگرها به ترتیب دارای توزیع پیشین t و نرمال هستند. در واقع اگر π برابر صفر قرار گیرد، این روش‌ها به ترتیب بی‌زی A و بی‌زی R خواهند بود. وقتی تعداد QTL‌ها کم باشد، هر QTL بخش بیشتری از واریانس صفت را به خود اختصاص می‌دهد و در نتیجه همان‌طور که نتایج این پژوهش نشان داد این شرایط به سود روش‌هایی مانند بی‌زی B، بی‌زی A و بی‌زی C است که اثرات نشانگرها را با شدت بیشتر و به صورت غیر یکسان به سمت صفر منقبض می‌کنند. هر چه تعداد QTL‌ها افزایش می‌یابد سهم آنها از واریانس صفت کم شده و در نتیجه برتری روش‌های بی‌زی B، بی‌زی

شیری دارد. برای این حالت خاص مطالعات متعددی برتری روش‌هایی مانند بیزی A را در ارزیابی صفت چربی شیر نسبت به سایر روش‌ها گزارش کردند. برای سایر صفات مانند تولید شیر تفاوت چندانی بین روش‌های مختلف مشاهده نشد که بیانگر این موضوع است که این صفات تحت تأثیر تعداد زیادی QTL با اثر کوچک هستند (۸ و ۱۴).

نتیجه گیری کلی

نتایج این پژوهش نشان داد که روش‌های بیزی A، بیزی B و بیزی C نسبت به تغییر در تعداد و توزیع اثرات QTL حساس هستند ولی روش‌های بیزی R حساسیتی به این تغییرات ندارد. صحت پیش‌بینی روش‌های بیزی A، بیزی B و بیزی C نسبت به بیزی L و بیزی R در تعداد پایین QTL بالاتر بود ولی با افزایش تعداد QTL‌ها به ۱۵۰ و بیشتر این برتری کاملاً از بین رفت. به طور کلی نتایج این پژوهش نشان داد که روش‌های مختلف مقایسه شده در صفات با معماری متفاوت، نتایج متفاوتی ارائه می‌کنند و نمی‌توان روشی را پیشنهاد داد که در همه شرایط و برای همه صفات برترین روش باشد. به هر حال در صفاتی که تعداد زیادی ژن آنها را کنترل می‌کنند می‌توان از روش‌هایی مانند بیزی R یا BLUP استفاده کرد ولی در صفاتی که تعداد کمی ژن آنها را کنترل می‌کنند یا دارای یک یا چند ژن با اثر عمده هستند که بخش زیادی از صفت را کنترل می‌کنند، می‌توان از روش‌هایی مانند بیزی B استفاده کرد.

کردند که صحت ارزش‌های اصلاحی GBLUP برای هر دو توزیع یکسان بود. در تحقیقی دیگر گودارد (۶) با مقایسه صحت ارزش‌های اصلاحی روش بیزی با استفاده از دو توزیع نرمال و گاما برای QTL‌ها مشاهده کردند که در حالت استفاده از توزیع گاما صحت پیش‌بینی‌های روش بیزی بهتر از توزیع نرمال بود که نتایج این تحقیق را تأیید می‌کنند. شیرعلی و همکاران (۱۵) با مقایسه روش BLUP و روش بیزی C تحت شرایط تعداد متفاوت QTL و وجود سه توزیع نرمال، گاما و یکنواخت مشاهده کردند که روش BLUP تحت تأثیر تعداد و توزیع آثار QTL‌ها قرار نگیرد ولی روش بیزی C در حالت وجود ۵ QTL و توزیع گاما برای اثر QTL‌ها صحت بالاتری داشتند.

نتایج این تحقیق و مطالعات شبیه سازی نشان داد که برای صفاتی که تعداد محدودی QTL با اثر زیاد آنها را کنترل می‌کند و توزیع QTL‌ها هم توزیع غیریکنواخت و دارای چولگی بیشتری است، روش‌هایی مانند بیزی B که آثار نشانگرها را با شدت بیشتری به سمت صفر منقبض می‌کنند، نسبت به سایر روش‌ها برتری دارند. اما مطالعات انجام شده بر اساس داده‌های واقعی تفاوت بسیار کمی بین روش‌های مختلف در خیلی از صفات گزارش کردند (۹ و ۱۷). در مطالعات تجربی معماری صفات کمتر شناخته شده است و نمی‌توان صفات مختلف را بر اساس آن دسته بندی کرد. در حیوانات اهلی مثال‌های بسیار کمی از صفاتی وجود دارد که یک یا چند ژن عمده بخش قابل توجهی از واریانس صفت را توجیه کند. یکی از این مثال‌ها ژن DGAT1 است که تأثیر زیادی بر درصد چربی گاوهای

منابع

- Calus, M. P. L., and R. F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. *Genetics Selection Evolution*, 43(1): 26-39.
- 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.
- de Los Campos, G., J. M. Hickey, R. Pong-Wong., H. D. Daetwyler, and M. P. Calus. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193(2): 327-345.
- Foroutanifar, S. 2015. Genomic Selection. *Genetics in the 3rd Millennium*, 12(4): 3680-3691. (In Persian).
- Gianola, D. 1982. Theory and analysis of threshold characters. *Journal of Animal Science*, 54(5): 1079-1096.
- Goddard, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, 136(2): 245-257.
- Hayes, B., and M. E. Goddard. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution*, 33(3): 209-229.
- Hayes, B. J., P. J. Bowman., A. C. Chamberlain., K. Verbyla, and M. E. Goddard. 2009. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution*, 41(1): 51-59.
- Hayes, B. J., P. J. Bowman., A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science*, 92(2):433-443.
- Karkkainen, H. P., and M. J. Sillanpaa. 2012. Back to basics for Bayesian model building in genomic selection. *Genetics*, 191(3): 969-987.
- Meuwissen, T. H., B. Hayes, and M. Goddard. 2001. Prediction of total genetic value using genome-wide dense

- marker maps. *Genetics*, 157(4): 1819-1829.
- 12- Perez, P, and G. de los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198(2): 483-495.
- 13- Pérez, P, and G. de Los Campos. 2014. Genome-wide regression & prediction with the BGLR statistical package. *Genetics*, 198: 483-495.
- 14- Resende, M. F. R., P. Munoz., M. D. V. Resende., D. J. Garrick., R. L. Fernando., J. M. Davis., E. J. Jokela., T. A. Martin., G. F. Peter, and M. Kirst. 2012. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics*, 190(4): 1503-1510.
- 15- Shirali, M., S. R. Miraei-Ashtiani., A. Pakdel., C. Haley., P. Navarro, and R. Pong-Wong. 2015. A Comparison of the sensitivity of the bayes C and genomic best linear unbiased prediction (GBLUP) methods of estimating genomic breeding values under different quantitative trait locus (QTL) model assumptions. *Iranian Journal of Applied Animal Science*, 5(1): 41-46.
- 16- Solberg, T., A. Sonesson., J. Woolliams, and T. Meuwissen. 2008. Genomic selection using different marker types and densities. *Journal of Animal Science*, 86(10): 2447-2454.
- 17- VanRaden, P. M., M. E. Tooker., J. R. Wright., C. Sun, and J. L. Hutchison. 2014. Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. *Journal of Dairy Science*, 97(12): 7952-7962.
- 18- Wang, C. L., X. D. Ding., J. Y. Wang., J. F. Liu., W. X. Fu., Z. Zhang., Z. J. Yin, and Q. Zhang. 2013. Bayesian methods for estimating GEBVs of threshold traits. *Heredity*, 110(3): 213-219.



Effect of QTL Number and Distribution Effects on Some Statistical Methods Genomic Prediction of a Threshold Trait

Saheb Foroutanifar^{1*}

Received: 01-07-2016

Accepted: 01-11-2016

Introduction Advances in high-throughput assays for genotyping single nucleotide polymorphisms (SNP) led to using these dense markers for predicting genomic breeding values, called genomic selection, and have been revolutionized both animal and crop breeding programs. The accuracy of genomic predictions is determined by the size of the reference population, extent of relationships between selection candidates and the reference population, linkage disequilibrium among markers and QTLs, the method used to estimate marker effects, and genetic architecture of the trait. Most of studies on this topic focus on continuous traits, but some of economical important traits in livestock are threshold. Therefore, a stochastic simulation was used to compare 125 different scenarios for a threshold trait based on five 10, 50, 150, 500 and 2000 underlying QTL numbers, five normal, uniform, t, gamma and Laplace distributions for QTL effects and five BayesA, BayesB, BayesC, BayesL and BayesR methods for estimation of marker effects.

Materials and Methods In order to compare the different methods, R software were used to simulate datasets. Simulation started with a base population of 100 animals, including 50 male and 50 female, which randomly mated for subsequent 50 generations. Generations were discrete and number of animals for each generation fixed at 100 animals. Thereafter, by randomly mating of animals in generation 50, reference population generation with 1000 animals were obtained and a threshold trait was simulated for each of them. Finally, Next validation population generation was produced by randomly mating of animals from the previous generation. Simulated genome for each animal consisted of 10 chromosomes with equal 1 Morgan lengths, each having 1000 evenly spaced SNPs. In each scenario QTLs were randomly distributed on genome and their substitution effects were drawn from one of five normal, uniform, t, gamma and Laplace distributions. Marker effects were estimated in reference population using five different Bayesian methods that differ with respect to assumptions regarding distribution of marker effects, including: Bayes A, Bayes B, Bayes C, Bayes L and Bayes R. These estimated markers effects were used for genomic breeding value predictions of animals in validation population which did not have any phenotype. The accuracy of the different methods was calculated as correlation between true and estimated genomic breeding values.

Results and Discussion Results of this study showed that Bayes A, Bayes B and Bayes C predictions was affected by QTL numbers and distributions, while Bayes L and Bayes R had almost same accuracies for all scenarios. Scenario with gamma distribution for QTL effects and number of 10 QTLs had highest accuracy for Bayes A, Bayes B and Bayes C methods. Bayes C method was the best method when number of QTL was low. Increases in number of QTL from 10 to 150 were decreased accuracy of Bayes A, Bayes B and Bayes C methods and after that accuracy of different methods was constant and same. When distributions of QTL effects was uniform lower accuracy achieved rather than other distributions for same QTL number, instead distributions with unequal QTL variance like gamma increased accuracy of Bayes A, Bayes B, and Bayes C methods. These results were because of different extent of shrinkage of estimates of effects for different methods. Methods like Bayes B induced differential shrinkage of estimates relative to methods like Bayes R and Bayes L lead to higher accuracy for scenarios with low QTL number and unequal QTL variance.

Conclusion in a nutshell, different studied methods had diverse results for various QTL number and distribution, and its difficult to suggest a method that best in all situations. However, it is better to use methods like Bayes B when the number of QTLs was low and each QTL had not equal contribution to the trait of interest.

Keywords: Bayesian methods, Categorical trait, Genomic selection, QTL.

1- Assistant Professor of Animal Breeding and Genetics, Department of Animal Science, Faculty of Agriculture, Razi University, Kermanshah, Iran.

(* - Corresponding Author Email: s.foroutanifar@razi.ac.ir)