



تولیات دامی

دوره ۱۵ ■ شماره ۱ ■ بهار و تابستان ۱۳۹۲

صفحه‌های ۶۵-۷۷

مقایسه روش‌های گوناگون ارزیابی ژنومیک در صفاتی با معماری ژنتیکی متفاوت

رستم عبداللهی آریناهی^۱، عباس پاکدل^{۲*}، اردشیر نجاتی جوارمی^۳، محمد مرادی شهر بابک^۴

۱. دانشجوی دکتری گروه علوم دامی، دانشکده علوم زراعی و دامی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج-ایران.
۲. دانشیار گروه علوم دامی، دانشکده علوم زراعی و دامی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج-ایران
۳. دانشیار گروه علوم دامی، دانشکده علوم زراعی و دامی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج-ایران
۴. استاد گروه علوم دامی، دانشکده علوم زراعی و دامی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج-ایران

تاریخ پذیرش مقاله: ۱۳۹۲/۰۶/۲۷

تاریخ وصول مقاله: ۱۳۹۱/۰۴/۰۶

چکیده

هدف از انجام تحقیق حاضر، مقایسه روش‌های گوناگون آماری در پیش‌بینی ارزش‌های اصلاحی ژنومیک برای صفاتی با معماری ژنتیکی متفاوت از نظر توزیع تأثیرات ژنی و نیز تعداد متفاوت جایگاه‌های صفت کمی (QTLs) با استفاده از شبیه‌سازی کامپیوتری است. بدین منظور، ژنومی حاوی ۵۰۰ نشانگر تک‌نوکلئوتیدی دوآلی (SNP) روی کروموزومی به طول ۱۰۰ سانتی‌مورگان شبیه‌سازی، و توزیع‌های متفاوت تأثیرات ژنی (یکنواخت، نرمال، و گاما) و نیز سه تعداد QTL (۵۰، ۱۰۰ و ۲۰۰) به صورت فرضیه‌های شبیه‌سازی صفت روی آن در نظر گرفته شد. بدین ترتیب نه صفت با معماری ژنتیکی متفاوت ایجاد شد. به منظور پیش‌بینی ارزش‌های اصلاحی افراد موجود در جمعیت‌های مرجع و تأیید، از شش روش بهترین پیش‌بینی نااریب خطی ژنومیک^۱ (GBLUP)، رگرسیون ریدج^۲ (RRBLUP)، بیز A (BayesA)، بیز B (BayesB)، بیز Cπ (BayesCπ)، و بیز L (BayesL) استفاده گردید. با افزایش فاصله بین نسل مرجع و نسل تأیید ناشی از برهم‌خوردن فاز لینکاژ، صحت ارزش‌های اصلاحی ژنومیک در تمام روش‌ها کاهش معنی‌داری نشان داد (P<۰/۰۵). همچنین زمانی که توزیع تأثیر ژنی گاما بود روش‌های بیزی در مقایسه با روش GBLUP و RRBLUP برتری آشکاری نشان دادند. هنگامی که توزیع تأثیرات ژنی نرمال بود، بیشترین صحت ارزش‌های اصلاحی به دست آمد. با افزایش تعداد QTLها به سمت ۲۰۰، صحت ارزش‌های اصلاحی ژنومیک کاهش یافت. نتایج این تحقیق نشان داد که در مجموع روش‌های بیزی و GBLUP از نظر صحت ارزش‌های اصلاحی پیش‌بینی‌شده در مقایسه با روش RRBLUP عملکرد بهتری دارند. نتایج تحقیق حاضر نشان می‌دهد هنگامی که معماری صفات بررسی‌شده از مدل تعداد زیاد جایگاه ژنی پیروی نکند، معمولاً روش‌های بیزی بر روش‌های GBLUP و RRBLUP ارجحیت دارند.

کلیدواژه‌ها: ارزش اصلاحی، روش‌های بیزی، ژنوم، معماری ژنتیکی.

1. Genomic-Best Linear Unbiased Prediction
2. Ridge Regression

Email: abpakdel@ut.ac.ir

* نویسنده مسئول

مقدمه

باتوجه به اینکه دسترسی به اطلاعات واقعی ژنومیک در کشور محدود است، شبیه‌سازی ابزار مناسبی است تا محققان همپای دانشمندان ژنتیک جهان حرکت نمایند. در زمینه پیشرفت‌های ژنتیک با پیشرفت فناوری‌های توالی‌یابی DNA و دردسترس‌بودن نشانگرهای متراکم گسترده در سرتاسر ژنوم، گزارش‌های زیادی از جایگزین شدن انتخاب به‌کمک کل ژنوم^۱ به‌جای انتخاب کلاسیک^۲ و انتخاب به‌کمک نشانگرها^۳ ارائه شده است (۱ و ۵). گرچه انتخاب به‌کمک کل ژنوم در حال پیشرفت است، اما پیش‌بینی شده است که این روش ارزیابی، روشی مرجع در ژنتیک گیاهی، حیوانی، و پزشکی شود. روش انتخاب ژنومیک در حدود ۹۲ درصد در هزینه‌های اصلاح نژادی صرفه‌جویی کرده است و فاصله نسل را نیز به‌طور چشم‌گیری در صنایعی نظیر صنعت پرورش گاو شیری می‌کاهد و بدین ترتیب به پیشرفت ژنتیکی به میزان دو برابر می‌انجامد (۲۰).

از عوامل مؤثر بر صحت پیش‌بینی ارزش‌های اصلاحی ژنومیک می‌توان به وراثت‌پذیری صفت، تعداد افراد و نسل‌های جمعیت مرجع^۴ و تأیید^۵، نوع و تراکم نشانگرها، معماری ژنتیکی صفت کمی (نظیر توزیع تأثیرات ژنی، توزیع فراوانی آلی، و تعداد QTLها)، و روش استفاده‌شده در پیش‌بینی ارزش‌های اصلاحی ژنومیک اشاره کرد (۸).

ارزیابی‌های ژنتیکی به‌کمک کل ژنوم، با ترکیب روش‌های آماری و داده‌های ژنومیک ارزش‌های ژنتیکی صفات کمی را پیش‌بینی می‌کنند. هنوز هیچ روش آماری

مناسبی در این خصوص معرفی نشده است و چالش‌های اساسی مانند افزایش تعداد پارامترهای برآورد شده (تأثیر نشانگرهای متراکم، p) در مقایسه با تعداد مشاهدات (n) در این حوزه از دانش وجود دارد. برای مثال، با یک تراشه ۶۰۰۰۰ نشانگری و تعداد ۲۰۰۰ رکورد فنوتیپی در جمعیت مرجع، به‌فرض وجود فقط تأثیر افزایشی، تعداد تأثیرهای نشانگری که بایستی برآورد شوند ۳۰ برابر تعداد مشاهدات است. از این‌رو در حل معادلات چندمجهولی به‌علت زیاد شدن تعداد مجهولات و محدود بودن تعداد معادلات، مشکل معکوس کردن ماتریس‌های ضرایب در روش‌های آماری مانند روش حداقل مربعات معمولی^۶ پیش می‌آید، به‌طوری‌که معکوس کردن مستقیم^۷ این ماتریس‌ها امکان‌پذیر نیستند. به‌منظور فائق‌آمدن بر این مشکل، استفاده از روش‌های جریمه‌ای^۸ مانند روش بهترین پیش‌بینی ناریب خطی ژنومیک، GBLUP (۱۷ و ۲۴)، ریج‌های رگرسیون، RRBLUP (۱۴)، BayesA (۱۵)، و روش‌های جریمه‌ای همراه با انتخاب متغیر^۹ مانند BayesB (۱۵)، BayesL (بی‌بی‌اس) (LASSO) (۱۸)، و روش‌های ناپارامتری (۷) و نیمه پارامتری (۱۰) در این خصوص پیشنهاد شده است. پژوهشگران با مرور مطالعات شبیه‌سازی انجام‌شده در زمینه انتخاب ژنومیک گزارش کردند برتری روش بی‌ز B بر RRBLUP وقتی نسبت تعداد QTL به تعداد افراد موجود در جمعیت مرجع کمتر از یک باشد، بیشتر خواهد شد اما با افزایش این شاخص به بیشتر از یک، روش RRBLUP ارجحیت پیدا می‌کند (۲۵). محققان با مرور نتایج مطالعات انتخاب ژنومیک انجام‌شده با داده‌های واقعی، نشان دادند برای

1. Genome Assisted Selection or Genome-enabled Prediction
2. Traditional Selection
3. Marker Assisted Selection
4. Training set
5. Validation set

6. Ordinary Least Square
7. Direct Inverse
8. Penalized or Shrinkage Methods
9. Variable or Subset Selection

تولیدات دامی

عدم تعادل پیوستگی بین نشانگر و QTL، بین افراد نسل پایه، ۵۰ نسل آمیزش تصادفی انجام شد. هاپلوتیپ‌های پدری و مادری هرکدام از والدین به‌طور تصادفی برای ایجاد نتایج نمونه‌گیری شد. برای ایجاد نوترکیبی از تابع نوترکیبی هالدین استفاده شد. طی ۵۰ نسل بررسی‌شده و بافرض براینکه از هر والد دو فرزند گرفته خواهد شد، اندازه جمعیت در هر نسل ثابت و معادل با ۱۰۰ فرد باقی ماند و در این حالت رانش و نوترکیبی باعث شد که میانگین آمار عدم تعادل پیوستگی (r^2) در نسل ۵۰ به ۰/۱۶ برسد. در نسل ۵۱ اثر جایگزینی برای هریک از QTLها برحسب سناریوی مطالعه‌شده از توزیع یکنواخت (حداقل صفر و حداکثر یک)، توزیع نرمال استاندارد (میانگین صفر و واریانس یک)، و توزیع گاما (با پارامتر شکل = ۰/۴ و پارامتر مقیاس = ۰/۶۰) به‌طور تصادفی گرفته شد. تمامی واریانس ژنتیکی صفت به‌وسیله QTLها پوشش داده شد و ارزش اصلاحی واقعی هریک از افراد باتوجه به ژنوتیپ هر فرد، از مجموع اثر QTLها محاسبه شد.

بعد از ۵۰ نسل آمیزش تصادفی، افراد در نسل ۵۱ با نمونه‌گیری تصادفی از والدین نسل قبل گسترش‌یافته، به‌تعداد ۱۰۰۰ رسید و این ساختار تا نسل ۵۴ ادامه یافت. افراد نسل ۵۱ به‌عنوان جمعیت مرجع (دارای اطلاعات فنوتیپی و ژنوتیپی) و افراد نسل‌های ۵۱ تا ۵۴ نیز به‌عنوان جمعیت تأیید (فقط دارای اطلاعات ژنوتیپی) در نظر گرفته شدند. همچنین از نرم‌افزار hypred برای شبیه‌سازی ساختار جمعیت‌ها استفاده شد (۲۳).

مدل‌های آماری

از شش روش آماری متفاوت که به‌ترتیب در ادامه توضیح داده خواهد شد برای برآورد تأثیر نشانگرها استفاده شد.

۱. روش بهترین پیش‌بینی نااریب خطی ژنومیک

(GBLUP)

صفاتی مانند چربی شیر که چندین ژن عمده مؤثر بر آن شناسایی شده است، روش‌های بی‌ز A و بی‌ز B بهتر از GBLUP عمل می‌کنند، اما برای صفاتی مانند پروتئین شیر که اکثر واریانس صفت با ژن‌هایی با اثر جزئی کنترل می‌شود و معماری ژنتیکی آن از مدل بی‌نهایت جایگاه ژنی^۱ (IFM) تبعیت می‌کند، عملکرد روش‌های بی‌زین و GBLUP تقریباً مشابه است (۸). بنابر مطالعات انجام‌شده، پیشنهاد شده است که به‌دلیل متفاوت بودن معماری ژنتیکی صفات کمی، انتخاب مناسب‌ترین روش ارزیابی بایستی بر مبنای معماری ژنتیکی صفت بررسی شده صورت گیرد (۸ و ۲۵). از این‌رو هدف از انجام تحقیق حاضر، مقایسه شش روش آماری GBLUP، RRBLUP، BayesA، BayesB، و BayesL در پیش‌بینی ارزش‌های اصلاحی ژنومیک برای صفاتی با معماری ژنتیکی متفاوت از نظر توزیع تأثیرات ژنی و تعداد QTLها با استفاده از شبیه‌سازی کامپیوتری است.

مواد و روش‌ها

شبیه‌سازی جمعیت پایه و ایجاد عدم تعادل پیوستگی^۲ (LD)

برای انجام مطالعه حاضر، ژنومی متشکل از کروموزومی به‌طول ۱۰۰ سانتی‌مورگان با ۵۰۰ عدد نشانگر SNP با فواصل نشانگری یکسان از یکدیگر در طول ژنوم شبیه‌سازی شد. بر مبنای سناریوی شبیه‌سازی، ۵۰، ۱۰۰ و ۲۰۰ QTL به‌طور تصادفی روی کروموزوم پراکنده شدند. برای نشانگرها و QTLهای دوآلی فراوانی آلی اولیه ۰/۵ منظور شد. در این مطالعه، جمعیت پایه با اندازه مؤثر ۱۰۰ فرد (۵۰ نر و ۵۰ ماده) در نظر گرفته شد. به‌منظور ایجاد

1. Infinitesimal Model
2. Linkage Disequilibrium

تولیدات دامی

در این رابطه، b بردار تأثیر نشانگری و سایر نمادها مشابه با رابطه ۱ است. برای برآورد تأثیر نشانگری معادلات ماتریسی به صورت رابطه ۵ هستند:

$$[X'X + \lambda D] \hat{b} = X'y \quad (5)$$

در این رابطه، λ مقدار ثابت است که در برخی منابع به آن ثابت تنظیم‌کننده^۱ هم می‌گویند و D نیز ماتریس قطری است که به جز عدد صفر که در عنصر قطری اول آن قرار می‌گیرد ($d_1=0$) تا از جریمه‌کردن مقدار میانگین یا عرض از مبدأ صرف‌نظر شود) سایر عناصر قطری آن یک هستند. هنگامی که λ برابر با صفر است جواب معادلات مذکور مساوی جواب معادلات حداقل مربعات معمولی است، اما اضافه‌کردن مقدار ثابت (λ) به عناصر قطری ماتریس ضرایب $[X'X + \lambda D]$ باعث می‌شود که ماتریس رتبه کامل شود و ضرایب رگرسیون به استثناء عرض از مبدأ به سمت صفر کاهش یابند. انتخاب λ هنوز هم چالش است، اما معمولاً مقدار وراثت‌پذیری صفت و تعداد پارامترهای مجهول مدل در آن تأثیر دارند. این تغییر شکل در معادلات باعث معرفی اریب اما کاهش دادن واریانس برآوردها خواهد شد. در صورت وجود مشکلاتی مانند تعداد زیاد پارامتر (p) و تعداد کم مشاهدات (n)، ($p > n$) این روش، مقدار میانگین مربعات خطا را کاهش می‌دهد و ممکن است پیش‌بینی‌های دقیق‌تری به دست آید.

روش‌های بیزین

مدل آماری در تمام روش‌های بیزی که در ادامه به‌طور مختصر توضیح داده خواهد شد به صورت رابطه ۶ است:

$$y_i = \mu + \sum_{j=1}^m x_{ij} b_j + e_i \quad (6)$$

در رابطه ۶، y_i مقدار مشاهده‌آم، μ عرض از مبدأ، x_{ij}

مدل آماری استفاده‌شده در روش GBLUP (۲۴) به صورت رابطه ۱ است:

$$y = 1\mu + Xu + e \quad (1)$$

در رابطه ۱، y فنوتیپ تصحیح‌شده، μ میانگین کل، u بردار دربرگیرنده ارزش‌های اصلاحی ژنومیک، e بردار تأثیر تصادفی باقی‌مانده، X ماتریس ضرایب ارتباط‌دهنده مشاهدات به بردارهای ارزش‌های اصلاحی، و 1 نیز بردار واحد است. عناصر ماتریس X شامل اعداد صفر، یک، و دو است که نشان‌دهنده تعداد آل‌های خاص هریک از مکان‌های نشانگری باتوجه به ژنوتیپ هر فرد است. در این مدل اثر نشانگرها به صورت تصادفی با واریانس یکسان در نظر گرفته شد. برای برآورد اثر نشانگرها معادلات مختلط هندرسون به صورت رابطه ۲ تشکیل شد:

$$\hat{u} = [X'X + G^{-1}k]^{-1} [X'y] \quad (2)$$

در رابطه ۲، k برابر است با:

$$k = \left(\frac{1-h^2}{h^2} \right)$$

در این رابطه، h^2 وراثت‌پذیری صفت است. هر کدام از عناصر ماتریس G نیز با استفاده از رابطه ۳ محاسبه شد:

$$G_{ij} = \frac{1}{m} \sum_{i=1}^m \frac{(x_{ij} - 2p_i)(x_{ij} - 2p_i)}{2p_i(1-p_i)} \quad (3)$$

در این رابطه، x_{ij} تعداد کپی آل مرجع برای ژنوم SNP در ژنوم فرد، p_i نیز فراوانی آل مرجع، و m تعداد نشانگر است. در پایان تأثیرات هر SNP با استفاده از ارزش‌های اصلاحی ژنومیک برآورد شده محاسبه شدند.

۲. روش بهترین پیش‌بینی نااریب خطی رگرسیون ریدج (RRBLUP)

مدل آماری در روش رگرسیون ریدج (۱۴) که در برخی منابع SNP BLUP هم نامیده شده است، به صورت رابطه ۴ است:

$$y = 1\mu + Xb + e \quad (4)$$

1. Regularization Constant

$$p(b_j | \theta_{b_j}, \sigma^2) = t(b_j | df_b, S_b) = \int N(b_j | 0, \sigma_{b_j}^2) x^{-2} (\sigma_{b_j}^2 | df_b, S_b) d\sigma_{b_j}^2 \quad (9)$$

در این رابطه، df_b و S_b به ترتیب درجات آزادی و پارامتر مقیاس و $x^{-2} (\sigma_{b_j}^2 | df_b, S_b)$ توزیع کای اسکور مقیاس دار معکوس است.

۴. بیز B (BayesB)

در روش بیز B (۱۵) تأثیرهای نشانگری با احتمال π برابر با صفر و با احتمال $(1-\pi)$ از توزیع t به روشی که در بیز A شرح داده شد به دست می‌آیند، به این صورت که:

$$b_j | df_b, S_b, \pi \sim \text{IID} \begin{cases} 0 \text{ with probability } \pi \\ t(b_j | df_b, S_b) \text{ with probability } 1-\pi \end{cases} \quad (10)$$

۵. بیز C δ (BayesC π)

بیز C شکلی از بیز B است (۱۲) و در برخی از مقالات به عنوان بیز C π شناخته می‌شود، فقط به جای اینکه از توزیع پیشین t برای تأثیرات نشانگری استفاده کند از توزیع نرمال استفاده می‌شود و از این رو توزیع پسین هم مزدوج و نرمال است.

۶. بیز L (BayesL)

در روش بیز L که بیز LASSO^۱ هم نامیده می‌شود (۱۸)، تابع چگالی احتمال پیشین نمایی دوگانه^۲ به تأثیرات نشانگری اختصاص داده شده است و شکل آن در رابطه ۱۱ است:

ژنوتیپ‌های نشانگری است که دارای عناصر صفر، یک، و دو است، b اثر نشانگری و e اثر باقی‌مانده است. توزیع چگالی احتمال توأم پیشین مجهولات رابطه ۶ به صورت رابطه ۷ است:

$$p(\mu, b, \sigma^2 | df, S, \omega) \propto \left\{ \prod_{j=1}^m p(b_j | \theta_{b_j}, \sigma^2)_p(\theta_{b_j} | \omega) \right\} x^{-2} (\sigma^2 | df, S)$$

در رابطه ۷، توزیع پیشین یکنواخت برای μ ، $x^{-2} (\sigma^2 | df, S)$ توزیع کای اسکور مقیاس دار برای واریانس باقی‌مانده با درجه آزادی df و S پارامتر مقیاس است. $p(b_j | \theta_{b_j}, \sigma^2)$ توزیع پیشین ژامین نشانگر را نشان می‌دهد که در آن θ_{b_j} برداری از پارامترهاست که نوع توزیع پیشین در نظر گرفته شده را برای تأثیرات نشانگری مشخص می‌کند. $p(\theta_{b_j} | \omega)$ توزیع پیشین در نظر گرفته شده برای θ_{b_j} و ω پارامتری است که این نوع توزیع را مشخص می‌کند. توزیع پسین ناشناخته‌های مدل به صورت رابطه ۸ است:

$$p(\mu, b, \sigma^2 | y, df, S, \omega) \propto \prod_{i=1}^n N(y_i | \mu + \sum_{j=1}^m x_{ij} b_j, \sigma^2) \times \left\{ \prod_{j=1}^m p(b_j | \theta_{b_j}, \sigma^2) p(\theta_{b_j} | \omega) \right\} x^{-2} (\sigma^2 | df, S)$$

تفاوت اصلی روش‌های بیزین در انتخاب توزیع پیشین برای تأثیرات نشانگری است. در ادامه برای هر روش شکل توزیع پیشین شرح داده خواهد شد.

۳. بیز A (BayesA)

در روش بیز A (۱۵) توزیع شرطی در نظر گرفته شده برای تأثیرات نشانگری توزیع t و به صورت رابطه ۹ است:

1. Bayesian Least Absolute Shrinkage and Selection Operator (Bayesian LASSO)
2. Double-Exponential

(۱۱)

صحت برآورد ارزش‌های اصلاحی ژنومیک از مربع همبستگی بین ارزش‌های اصلاحی واقعی و ارزش‌های اصلاحی ژنومیک برآورده شده به دست آمد. تغییرات صحت GEBV برای نسل مرجع و سه نسل گروه تأیید بررسی شد. هر سناریوی شبیه‌سازی که دربرگیرنده توزیع تأثیر QTL و تعداد QTL بود ۱۰ بار تکرار شد و میانگین تکرارها همراه با خطای استاندارد میانگین‌ها ذخیره شد.

$$DE(b_j | \lambda, \sigma_e^2) = \int N(b_j | 0, \sigma_e^2 \tau_j^2) \text{Exp}\left(\tau_j^2 | \frac{\lambda^2}{2}\right) d\sigma_{b_j}^2$$

در معادله فوق، λ ناشناخته فرض شده است و به آن توزیع گاما با پارامتر شکل λ و مقیاس τ اختصاص داده می‌شود.

اجرای روش‌های آماری بیزی

از نرم‌افزار BGLR (۹) به منظور پیاده‌سازی روش‌های بیزی استفاده گردید. در اجرای روش‌های بیزی از الگوریتم نمونه‌گیری گیبس برای نمونه‌گیری توزیع پسین شرطی تأثیر نشانگری استفاده شد. در هر تکرار و سناریو، زنجیره‌ای به طول ۵۰۰۰۰ نمونه شامل ۱۰ هزار نمونه برای گرم‌کردن^۱ و ۴۰۰۰۰ نمونه بعدی برای استنباط‌های توزیع پسین استفاده شد. برای اطمینان از همگرایی نمونه‌گیری‌های گیبس، از رسم پلات‌ها و خودهمبستگی بین نمونه‌ها استفاده شد.

نتایج و بحث

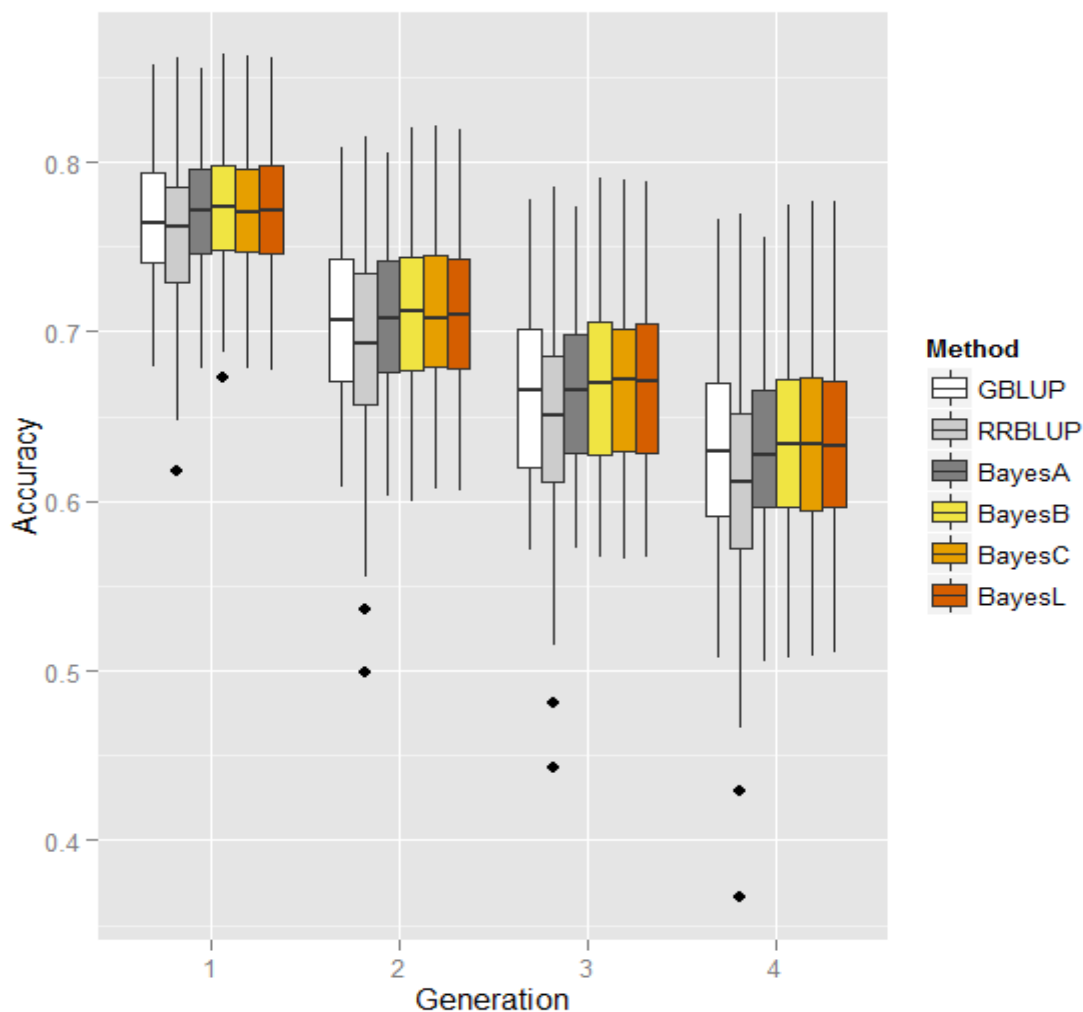
به علت تنوع در جواب‌های به دست آمده از تکرارهای متفاوت، نتایج به صورت نمودار جعبه‌ای در شکل‌های ۱، ۲ و ۳ ارائه شده است. نتایج حاصل از مقایسه شش روش آماری BayesL، BayesC π ، BayesB، BayesA، RRBLUP، GBLUP از نظر صحت ارزش‌های اصلاحی ژنومیک در طی چهار نسل (نسل اول افراد جمعیت مرجع و نسل دوم تا چهارم جمعیت تأیید را تشکیل می‌دهند) در شکل ۱ ارائه شده است. سه روش اول GBLUP، RRBLUP و BayesA از جمله روش‌های جریمه‌ای هستند که با اضافه کردن مقادیری ثابت به عناصر قطری ماتریس ضرایب ژنومیک، مقدار برآوردی تأثیرات نشانگری را به سمت صفر تنزیل می‌دهند. لیکن سه روش بعدی یعنی روش‌های BayesB، BayesC π و BayesL از جمله روش‌های جریمه‌ای توأم با انتخاب متغیر هستند، چراکه در این روش‌ها توزیع‌های پیشینی برای تأثیر نشانگرها تعریف می‌شود که به اجبار تأثیر درصدی از نشانگرها را صفر فرض می‌کنند. این درحالی است که در روش‌های فقط جریمه‌ای، تأثیر تمام نشانگرها به نسبت یکسانی به سمت صفر کاهش داده می‌شوند.

به منظور برآورد تأثیرات نشانگری در جمعیت مرجع از فنوتیپ تصحیح شده برای اثر ثابت جنس استفاده شد و سپس از تأثیرات نشانگری برآورد شده در جمعیت مرجع برای پیش‌بینی ارزش‌های اصلاحی ژنومیک افراد جمعیت تأیید استفاده گردید. پس از حل معادلات مذکور و برآورد اثر هر یک از نشانگرها، ارزش اصلاحی ژنومیک^۲ (GEBV) افراد گروه تأیید از مجموع اثر نشانگرها و باتوجه به ژنوتیپ آن‌ها به صورت رابطه ۱۲ محاسبه شد:

$$GEBV_i = \sum_j^m x_{ij} \hat{b}_j \quad (12)$$

1. Burn in
2. Genomic Estimated Breeding Value

تولیدات دامی



شکل ۱. صحت ارزش‌های اصلاحی ژنومیک با استفاده از شش روش ارزیابی بهترین پیش‌بینی ناریب خطی ژنومیک (GBLUP). ریدج رگرسیون (RRBLUP)، بیز A (BayesA)، بیز B (BayesB)، بیز $C\pi$ (BayesC π)، و بیز L (BayesL) در طول چهار نسل. نسل اول جمعیت مرجع و نسل‌های ۲ تا ۴ جمعیت‌های تأیید یا آزمون هستند. نقاط سیاه داده‌های پرت را نشان می‌دهند.

مقیاس‌بندی نشدند و همچنین مقدار λ یا ثابت تنظیم‌کننده در هر دو روش یکسان نبود نتایج متفاوتی حاصل شدند (۱۱). از مزایای GBLUP بر RRBLUP می‌توان به کاهش دادن ابعاد معادلات مختلط به تعداد افراد جمعیت مرجع، محاسبه صحت و خطای پیش‌بینی^۱ ارزش‌های

نتایج این تحقیق نشان داد که روش GBLUP و چهار روش بیزی (BayesA، BayesB، BayesC π ، و BayesL) در هر چهار نسل در مقایسه با روش RRBLUP از نظر صحت پیش‌بینی ارزش‌های اصلاحی ژنومیک بهتر عمل می‌کنند. تحقیقات نشان داد که GBLUP و RRBLUP مدل‌های مشابهی هستند، اما در این پژوهش به دلیل اینکه عناصر ماتریس طرح یا X به‌طور مشابهی در دو روش

1. Prediction Error Variance

شده است. در کل روش‌ها، صحت ارزش‌های اصلاحی در حالت توزیع نرمال تأثیر ژنی بهتر از گاما و یکنواخت بود اما این برتری معنی‌دار نبود. روش RRBLUP در حالت توزیع نرمال تأثیر ژنی عملکرد مشابهی با سایر روش‌ها از خود نشان داد اما در حالت توزیع یکنواخت و گاما RRBLUP نتوانست با روش‌های بیزی و GBLUP رقابت کند. هنگامی‌که توزیع تأثیر ژنی گاما است، برتری روش‌های بیزی در مقایسه با روش‌های غیر بیزی به‌ویژه در مقایسه با روش GBLUP کاملاً مشهود است، اما هنگامی که توزیع تأثیر ژنی نرمال در نظر گرفته شد، عملکرد روش‌های بیزی و GBLUP و RRBLUP مشابه بود.

در بین روش‌های بیزی، روش بیز B و بیز L در حالت توزیع گاما، تأثیرات ژنی بهترین عملکرد را نشان دادند، اما تفاوت آن‌ها از نظر آماری معنی‌دار نبود ($P > 0.05$). هنگامی‌که توزیع تأثیر ژنی گاما است، برخی از ژن‌ها داری اثر عمده و درصد بالایی از ژن‌ها تأثیراتی نزدیک به صفر دارند. از این رو روش‌های جریمه‌ای همراه انتخاب متغیر بهتر از روش‌های فقط جریمه‌ای عمل می‌کنند. با استفاده از داده‌های واقعی صفت تولید شیر و نرخ آبستنی در گاوهای هلشتاین فرانسوی، دلیل مشابه بودن نتایج حاصل از روش‌های بیزی و GBLUP در پیش‌بینی ارزش‌های اصلاحی ژنومیک، معماری ژنتیکی پلی‌ژنی بودن و توزیع نرمال تأثیر ژنی پیشنهاد شده است (۴). در این مطالعه نیز در حالت توزیع نرمال تأثیر ژنی، تفاوتی بین روش‌های فراوانی‌گرا (GBLUP و RRBLUP) با روش‌های بیزی مشاهده نشد. یکی از ضعف‌های روش‌های RRBLUP و GBLUP این است که برای همه نشانگرها سهم یکسانی در پیش‌بینی ارزش‌اصلاحی در نظر می‌گیرند. این در حالی است که در روش‌های بیزی و به‌استثنای روش بیز A بر حسب توزیع پیشین، وزن‌های متفاوتی به SNP‌ها اختصاص داده می‌شود.

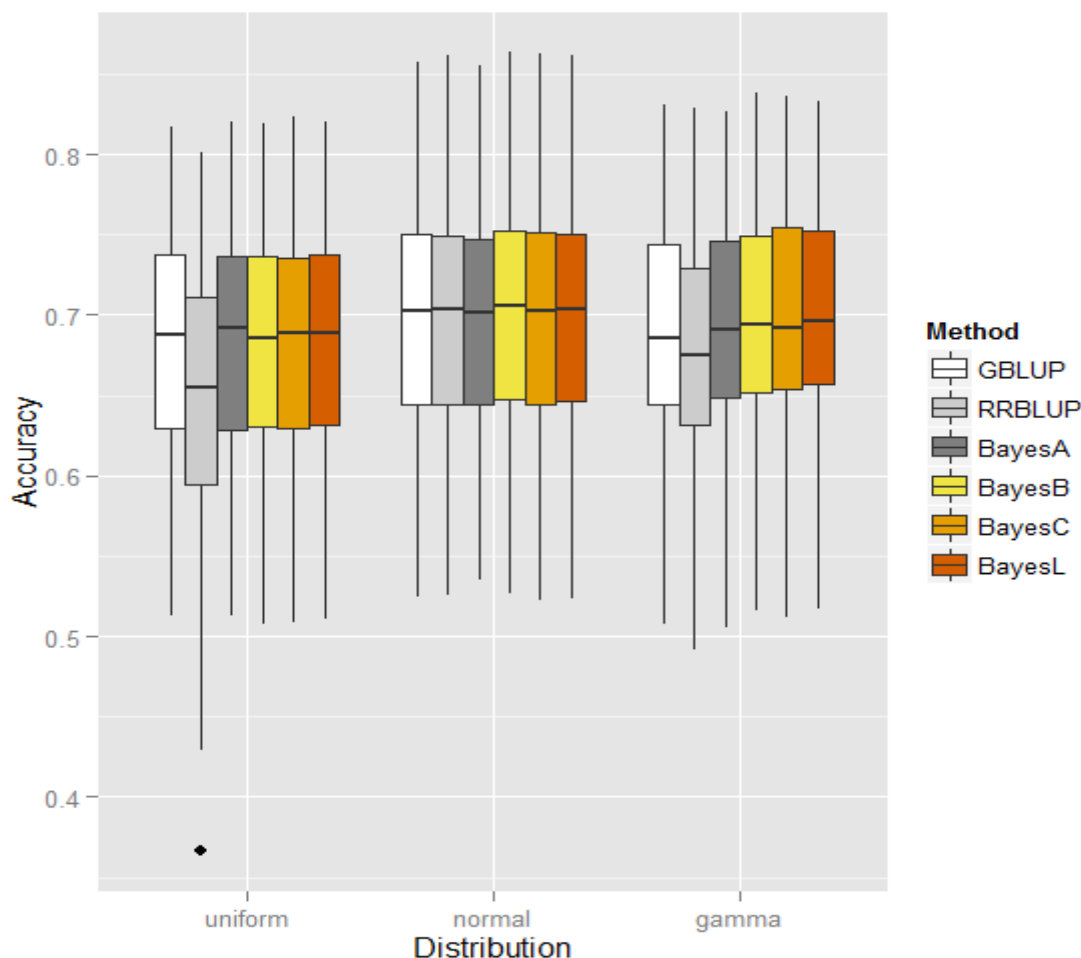
اصلاحی به‌شیوه مرسوم استفاده‌شده در روش‌های شجره‌ای، و ترکیب‌کردن اطلاعات افراد ژنوتیپ‌شده و نشده به‌طور همزمان در معادلات مختلط اشاره کرد. آخرین مزیت GBLUP بر RRBLUP به‌نام GBLUP یک‌مرحله‌ای^۱ نیز نامیده می‌شود.

با افزایش فاصله نسل بین جمعیت مرجع و جمعیت تأیید صحت ارزش‌های اصلاحی ژنومیک به‌شدت کاهش می‌یابد (شکل ۱). این روند انتظار می‌رفت، چراکه با افزایش فاصله بین جمعیت مرجع و جمعیت تأیید، فاز LD بین QTL‌ها و نشانگرها به هم می‌خورد و تأثیرهای QTL‌ها به‌وسیله نشانگرها به‌خوبی بیان نمی‌شود، که این نتایج با یافته‌های پیشین مطابقت دارد (۲ و ۱۱). نکته دیگر این است که با افزایش فاصله بین نسل مرجع و نسل چهارم، خطای استاندارد صحت پیش‌بینی ارزش‌های اصلاحی نیز، افزایش یافته است. با گذشت چند نسل پس از برآورد اثر نشانگرها، صحت برآورد ارزش‌های اصلاحی ژنومیک کاهش می‌یابد و از کارایی روش‌های برآورد تأثیرات نشانگری کاسته می‌شود و لازم است تأثیرهای نشانگرها مجدداً برآورد شود (۱۶). این نتایج حاکی از آن است که روش‌های بیزی در مقایسه با GBLUP و RRBLUP حساسیت کمتری به تغییر فاز پیوستگی بین جمعیت مرجع و جمعیت تأیید نشان می‌دهد. دلیل این تغییر می‌تواند این باشد که اکثر روش‌های بیزی علاوه بر ویژگی جریمه‌کردن از ویژگی انتخاب متغیر هم بهره می‌برند، در حالی‌که روش‌های GBLUP و RRBLUP این خصوصیت را ندارند.

مقایسه صحت پیش‌بینی ارزش‌های اصلاحی ژنومیک تحت تأثیر سه توزیع متفاوت تأثیر ژنی (یکنواخت، نرمال، و گاما) و با استفاده از شش روش آماری در شکل ۲ ارائه

1. Single Step Genomic BLUP

تولیدات دامی



شکل ۲. صحت ارزش‌های اصلاحی ژنومیک با استفاده از شش روش ارزیابی بهترین پیش‌بینی ناریب خطی ژنومیک (GBLUP)، ریدج رگرسیون (RRBLUP)، بیز A (BayesA)، بیز B (BayesB)، بیز C (BayesC)، و بیز L (BayesL) در سه حالت توزیع تأثیر ژنی یکنواخت (Uniform)، نرمال (Normal)، و گاما (Gamma). نقاط سیاه داده‌های پرت را نشان می‌دهند.

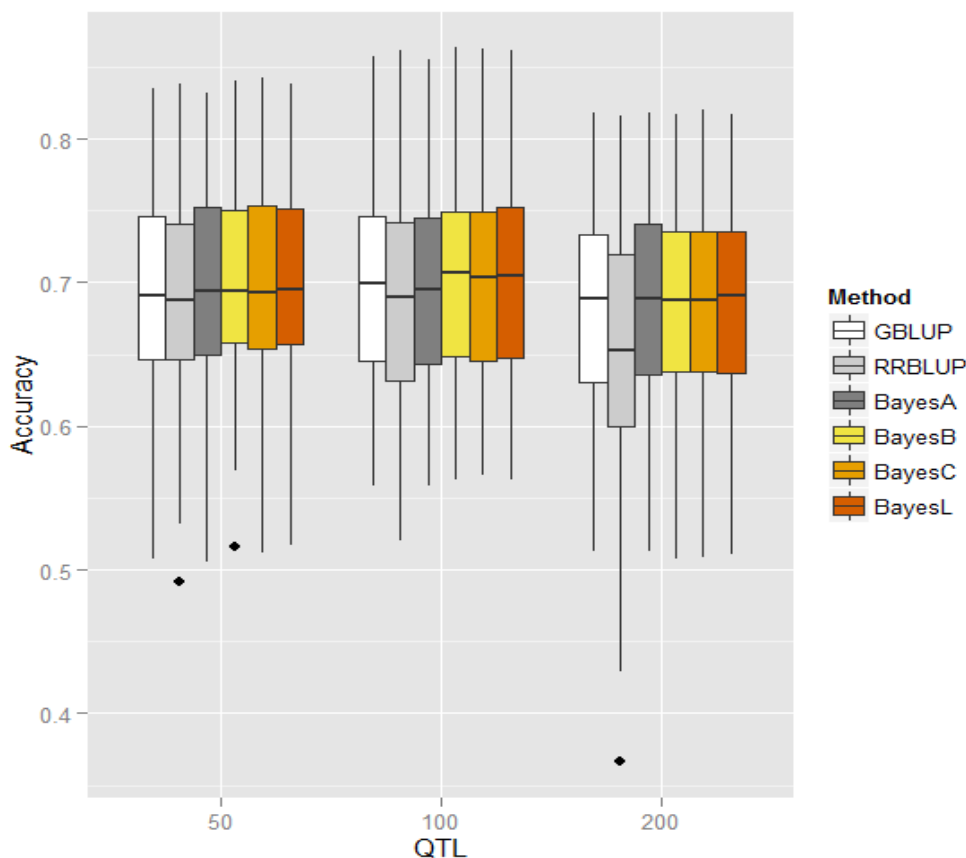
صفت، صحت برآوردها در روش‌های بیزی کاهش می‌یابد، مطابقت دارد (۶). در تعداد کم QTL، یعنی ۵۰، برتری روش‌های بیزی در مقایسه با روش‌های غیربیزی کاملاً مشهود است و روش RRBLUP ضعیف‌ترین عملکرد را داشته است، اما با افزایش تعداد QTL به سمت ۲۰۰ تفاوت بین شش روش آماری کاهش یافته است. نکته جالب این بود که صحت ارزش‌های اصلاحی در دو حالت تعداد QTL ۵۰ و ۱۰۰ برای سه روش جریمه‌ای (RRBLUP، GBLUP، BayesA) و مشابه بود اما در روش‌های جریمه‌ای توأم با

نتایج حاصل از مقایسه روش‌های گوناگون آماری در مقابل تعداد متفاوت QTL ها در شکل ۳ نشان داده شده است. در تمام روش‌ها با افزایش تعداد QTL‌ها از ۱۰۰ به ۲۰۰ صحت ارزش‌های اصلاحی ژنومیک کاهش یافته است. دلیل این امر را می‌توان به توزیع مقدار محدودی واریانس ژنتیکی بر تعداد زیادی QTL دانست که در نتیجه سهم هر QTL در ارزش ژنتیکی کل کاهش یافته است و قدرت مدل‌ها در برآورد تأثیرات کم خواهد شد. این نتایج با دیگر یافته‌های گزارش شده که با افزایش تعداد ژن‌های مؤثر بر

تولیدات دائمی

افزایش یابد تفاوت معنی‌داری بین این دو روش وجود ندارد (۲۱). در مطالعه مذکور نیز ملاحظه شد که در تعداد کم QTL (۱۰ تا ۳۰) احتمال اینکه توزیع تأثیرات ژنی شکل بگیرد، بسیار پایین است. از این رو مقایسه نتایج مطالعه حاضر با این گزارش منطقی به نظر نمی‌رسد، چراکه صفاتی با معماری‌های ژنتیکی متفاوت در هر دو مطالعه، بررسی شده است. با استفاده از داده‌های واقعی تولید تخم‌مرغ و کیفیت تخم‌مرغ در مرغان تخم‌گذار نشان داده شد که تفاوت معنی‌داری بین روش‌های بیز A، بیز C π و GBLUP از نظر صحت و تداوم صحت ارزش‌های اصلاحی ژنومیک وجود ندارد (۲۶).

انتخاب متغیر (BayesB, BayesC π , BayesL) صحت ارزش‌های اصلاحی وقتی ۱۰۰ QTL صفت را کنترل می‌کرد بهتر از حالتی بود که ۵۰ QTL بیان صفت کنترل می‌کنند. برتری روش BayesA و BayesC π بر RRBLUP برای صفات مرتبط با بیماری‌ها در درختان کاج که با تعداد کمی ژن کنترل می‌شوند نیز گزارش شده است (۱۹). انتخاب بهترین روش ارزیابی ژنومیک به معماری ژنتیکی صفت، اندازه جمعیت مرجع، میزان رابطه خویشاوندی بین افراد جمعیت مرجع و تأیید بستگی دارد (۸). نشان داده شده است که روش بیز C π در مقایسه با روش GBLUP در تعداد کم QTL بهتر عمل می‌کند، اما وقتی تعداد QTL



شکل ۳. صحت ارزش‌های اصلاحی ژنومیک با استفاده از شش روش ارزیابی بهترین پیش‌بینی ناریب خطی ژنومیک (GBLUP)، ریدج رگرسیون (RRBLUP)، بیز A (BayesA)، بیز B (BayesB)، بیز C π (BayesC π)، و بیز L (BayesL) در سه حالت تعداد QTL (۵۰، ۱۰۰، و ۲۰۰). نقاط سیاه داده‌های پرت را نشان می‌دهند.

تولیدات دائمی

دوره ۱۵ ■ شماره ۱ ■ بهار و تابستان ۱۳۹۲

اریبی‌هایی مصنوعی است و این تفاوت‌ها می‌تواند باعث متفاوت شدن نتایج مطالعات شبیه‌سازی در مقایسه با واقعیت شود.

باتوجه به نتایج، صحت پیش‌بینی ارزش‌های اصلاحی در انتخاب به‌کمک کل ژنوم به معماری ژنتیکی صفت مشخص نظیر توزیع تأثیر ژن‌ها و تعداد ژن‌ها بستگی دارد. درحالتی که توزیع تأثیر ژنی به سمت ژن‌های بزرگ‌اثر یا کوچک‌اثر چولگی دارد، معمولاً روش‌های جریمه‌ای توأم با انتخاب متغیر (BayesL, BayesCπ, BayesB) برتر از روش‌های فقط جریمه‌ای (RRBLUP و GBLUP) عمل می‌کنند. همچنین با زیاد شدن تعداد جایگاه‌های ژنی مؤثر بر صفت تفاوت بین روش‌های بی‌زی و GBLUP در پیش‌بینی ارزش‌های اصلاحی ژنومیک کاهش می‌یابد. از طرفی با مقایسه روش‌های گوناگون در پیش‌بینی ارزش‌های ژنتیکی، می‌توان اطلاعاتی درباره معماری ژنتیکی پشت صحنه صفت بررسی شده به‌دست آورد.

منابع

1. Abdollahi-Arpanahi R, Pakdel A and Zandi-Baghchehmaryam MB (2012) From infinitesimal model to Genomic Selection. *Modern Genetics*. 17(2): 105-114 (in Persian).
2. Bostan A, Nejati Javaremi A, Moradi Shahr Babak M and Saatchi M (2012) Using the dense molecular markers near the specific trait major genes for genomic prediction. *Iranian Journal of Animal Science*. 44(1): 53-60. (in Persian)
3. Cole JB, VanRaden PM, O'Connell JR, Van Tassel CP, Sonstegard TS, Schnabel RD, Taylor JF and Wiggans GR (2009) Distribution and location of genetic effects for dairy traits. *Dairy Science*. 92(6): 2931-46.

روش‌های فراوانی‌گرا^۱ (RRBLUP و GBLUP) از نظر سرعت محاسبات و تخصیص حافظه کامپیوتری در مقایسه با روش‌های بی‌زی در حالتی که تراکم نشانگری در حد متوسط (۳۰۰۰ تا ۶۰۰۰۰) و تعداد مشاهدات فنوتیپی در جمعیت مرجع به‌اندازه کافی باشد (بیشتر از ۱۰۰۰) روش‌های مناسبی هستند. مطالعات در همین زمینه نشان داده است که وقتی تعداد QTL‌های مؤثر بر صفت زیاد باشد، روش GBLUP مشابه و یا حتی بهتر از روش‌های بی‌زی عمل می‌کند (۳، ۶ و ۲۵). با وجود این، به دلیل اینکه روش‌های بی‌زی از برخی خصوصیات نظیر "انتخاب متغیر" سود می‌برند، در تعداد اندک QTL برتر عمل می‌کنند. شایان ذکر است که می‌توان نه سناریوی توزیع تأثیر ژنی و تعداد QTL را به‌عنوان نه صفت متفاوت در نظر گرفت که دارای معماری ژنتیکی متفاوتی هستند. از این رو به نظر می‌رسد برای صفاتی با تعداد QTL زیاد، روش GBLUP، روشی مطمئن و بهینه برای استفاده در ارزیابی‌های ژنتیکی است.

از جمله محدودیت‌های این تحقیق می‌توان به محدود بودن صفات فرضی مطالعه‌شده با وراثت‌پذیری ۰/۳ و شبیه‌سازی فقط تأثیر افزایشی و در نظر نگرفتن تأثیر غیرافزایشی و نیز در نظر گرفتن تعداد QTL‌ها در دامنه ۵۰ تا ۲۰۰ روی ژنوم یک مورگانی اشاره کرد. مطالعات آینده قطعاً می‌توانند خلأ موارد مذکور را پر کنند و به‌صورت جامع‌تری به مقایسه روش‌های ارزیابی ژنومیک بپردازند. از طرفی مطالعات ژنومیکی که از داده‌های واقعی استفاده می‌کنند، در معرض اریبی‌هایی مانند خطاهای ژنوتیپی و نمونه‌گیری^۲ هستند، لیکن مطالعات شبیه‌سازی از چنین

1. Frequentist
2. Ascertainment Bias

4. Colombani C, Legarra A, Fritz S, Guillaume F, Croiseau P, Ducrocq V and Robert-Granié C (2012) Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesCp methods for genomic selection in French Holstein and Montbéliarde breeds. *Dairy Science*. 96: 575-591.
5. Daetwyler HD, Calus MP, Pong-Wong R, De Los Campos G and Hickey JM (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics*. 193: 347-65.
6. Daetwyler HD, Pong-Wong R, Villanueva B and Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. *Genetics*. 185(3): 1021-31.
7. De los Campos G, Gianola D and Rosa GJM (2009) Reproducing kernel Hilbert spaces regression: A general framework for genetic evaluation. *Animal Science*. 87: 1883-1887.
8. De Los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD and Calus MP (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*. 193: 327-45.
9. 9 . De los Campos G and Perez PR (2012) BGLR: Bayesian Generalized Linear Regression. R Package. <http://bglr.r-forge.r-project.org/>
10. Gianola D, Fernando RL and Stella A (2006) Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*. 173(3): 1761-76.
11. Habier D, Fernando RL and Dekkers JC (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics*. 177(4): 2389-2397.
12. Habier D, Fernando RL, Kizilkaya K and Garrick DJ (2011) Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*. 12(1): 186.
13. Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ and Goddard ME (2010) Genetic architecture of complex traits and accuracy of genomic prediction: coat color, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLOS Genetics*. 6(9): 1001139.
14. Hoerl AE and Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 12(1): 55-67.
15. Meuwissen TH, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157(4): 1819-1829.
16. Muir WM (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *Animal Breeding and Genetics*. 124: 342-355.
17. Nejati-Javaremi A, Smith C and Gibson JP (1997) Effect of total allelic relationship on accuracy of evaluation and response to selection. *Animal Science*. 7(5): 1738-1745.
18. Park T and Casella G (2008) The bayesian lasso. *American Statistical Association*. 103(482): 681-686.
19. Resende MFR, Muñoz P, Resende MD, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF and Kirst M (2012) Accuracy of

- Genomic Selection Methods in a Standard Data Set of Loblolly Pine (*Pinus taeda* L.). *Genetics*. 190: 1503-1510.
20. Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*. 123: 218-223.
21. Shirali M, Miraei-Ashtiani SR, Pakdel A, Haley C and Pong-Wong R (2012) Comparison between Bayesc and GBLUP in Estimating Genomic Breeding Values under Different QTL Variance Distributions, in Abstract from ICQG2012 - 4th International Conference on Quantitative Genetics, . Edinburgh, United Kingdom. 261-268.
22. Swami M (2010) Complex traits: Using genetic architecture to improve predictions. *Nature Review Genetics*. 11(11): 748.
23. Technow FR (2011) Package hypred: Simulation of Genomic Data in Applied Genetics. University of Hohenheim.
24. VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. *Dairy Science*. 91: 4414-4423.
25. Wimmer V, Lehermeier C, Albrecht T, Auinger H-J, Wang Y, Schön C-C (2013) Genome-Wide Prediction of Traits with Different Genetic Architecture Through Efficient Variable Selection. *Genetics*. 195: 573-587.
26. Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Preisinger R, Habier D, Fernando R, Garrick DJ and Dekkers JC (2011) Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genetics Selection Evolution*. 43: 23.