



مقایسه روش‌های چند مرحله‌ای و تک مرحله‌ای بیزی برای برآورد ارزش‌های اصلاحی ژنومی در حیوانات ژنوتیپ شده و نشده - مطالعه شبیه‌سازی

مصطفی مدد جیرنده^۱، جلیل شجاع^۲، صادق علیجانی^۳، سید عباس رافت^۴ و جک سی. ام دکرز^۵

۱- دانشجوی دکتری، گروه علوم دامی، دانشگاه تبریز، تبریز، ایران

۲- استاد، گروه علوم دامی، دانشگاه تبریز، تبریز، ایران

۳- دانشیار، گروه علوم دامی، دانشگاه تبریز، تبریز، ایران (نویسنده مسوول: sad-ali@tabrizu.ac.ir)

۴- استاد، گروه علوم دامی، دانشگاه ایالتی آیووا، ایماز، ایالات متحده آمریکا

تاریخ دریافت: ۹۸/۴/۱۱ تاریخ پذیرش: ۹۸/۶/۲۰

صفحه: ۱۲۲ تا ۱۳۱

چکیده

هدف از این مطالعه، مقایسه صحت ارزیابی ژنومی روش‌های چندمرحله‌ای BayesA، BayesB، BayesC، BayesL و روش‌های تک‌مرحله‌ای بیزی SSBR-C و SSBR-A در مقادیر متفاوت π برای برآورد ارزش‌های اصلاحی ژنومی حیوانات تعیین ژنوتیپ شده و نشده بود. ژنومی حاوی ۴۰۰۰۰ نشانگر تکنوکلوتیدی دوالی پراکنده شده روی ۲۰ کروموزوم هرکدام به طول ۱۰۰ سانتی‌مورگان شبیه‌سازی شد. مقادیر بهینه π در روش BayesC به ترتیب ۰/۹۸۰ و ۰/۹۹۵ در دو توزیع نرمال و گامای اثرات ژنی برآورد شد و در روش SSBR-C نیز مورد استفاده قرار گرفت. صحت پیش‌بینی ژنومی در روش SSBR-C ($\pi=0.995$) نسبت به سایر روش‌ها از ۰/۰۲ تا ۰/۰۹ در توزیع گامای اثرات ژنی بیش‌تر برآورد شد. بنابراین، روش SSBR-C ($\pi=0.995$) با در نظر گرفتن توزیع مزدوج و استفاده همزمان از همه اطلاعات شجره‌ای، فنوتیپی و ژنومی توانست در حالت توزیع گامای اثرات ژنی عملکرد بهتری را از خود نشان داده و انتخاب مناسب‌تری به‌شمار رود. کلیه روش‌های تک‌مرحله‌ای و چندمرحله‌ای بیزی عملکرد تقریباً مشابهی را در حالت توزیع نرمال اثرات ژنی از خود نشان دادند. فلذا، در حالت توزیع نرمال اثرات ژنی توصیه می‌شود تا از روش SSBR-C ($\pi=0$) با توجه به ضریب تابعیت پیش‌بینی ژنومی نزدیک به یک استفاده شود. همچنین، افت صحت پیش‌بینی ژنومی با افزایش فاصله نسلی بین جمعیت مرجع و تایید برای افراد ژنوتیپ شده در مقایسه با افراد ژنوتیپ نشده از حساسیت کمتری برخوردار بود.

واژه‌های کلیدی: انتخاب ژنومی، بیزی، روش چندمرحله‌ای، شبیه‌سازی، صحت ژنومی

مقدمه

ارزیابی ژنومی با استفاده از روش‌های چند مرحله‌ای بر مبنای داده‌های فنوتیپی و ژنوتیپی صورت گرفته و اثرات نشانگرها برآورد می‌شود. در مدل آنالیز برای مرحله دوم، تنها اثرات در نظر گرفته شده، اثرات نشانگرها خواهد بود و پیش‌بینی ارزش‌های اصلاحی ژنومی^۱ برای دام‌های تعیین ژنوتیپ شده جوان با استفاده از این اثرات صورت خواهد گرفت (۳۳). به دلیل معایب روش‌های چندمرحله‌ای^۲ در توسعه مدل‌ها، روش تک‌مرحله‌ای پیشنهاد شد که در این روش به‌طور همزمان اطلاعات فنوتیپی، ژنومی و شجره‌ای در مدل قرار داده می‌شوند. در این روش برای حیوانات تعیین ژنوتیپ شده و نشده به‌طور همزمان ارزش اصلاحی برآورد می‌شود (۳، ۱۱، ۲۲).

روش‌های SSBR^۴ با تلفیق ماتریس روابط خویشاوندی ژنومی و شجره‌ای ماتریس H را به‌وجود می‌آورند. در این روش از زنجیره مارکوف مونت کارلو^۵ برای برآورد مولفه‌های واریانس استفاده شده تا بتوان از توزیع‌های اثرات ژنی متنوع‌تری بهره‌مند شد (۱۱). بنابر نتایج پژوهش انجام شده، صحت پیش‌بینی تمام روش‌های چندمرحله‌ای بیزی در توزیع نرمال اثرات ژنی به هم نزدیک بود. همچنین زمانی که توزیع اثرات ژنی به‌صورت گاما در نظر گرفته شد، صحت برآوردها با استفاده از روش‌های بیزی بالاتر از روش‌های غیر بیزی گزارش شد (۲).

توزیع نرمال یکی از مهم‌ترین توزیع‌های احتمالی پیوسته است. تابع چگالی این توزیع دارای دو پارامتر است که یکی

تعیین‌کننده میانگین (μ) و دیگری تعیین‌کننده واریانس (σ) توزیع هستند. توزیع گاما نیز از توزیع‌های احتمالی پیوسته به‌شمار می‌رود و دارای دو پارامتر مقیاس β و شکل α است. به دلیل پارامترهای شکل و مقیاس، این توزیع انعطاف‌پذیری زیادی برای تحلیل هر مجموعه داده حقیقی مثبت را داراست (۹، ۲۴).

جیانولا و همکاران (۱۲) با مقایسه روش‌های BayesA و BayesB نشان دادند که می‌توان مقادیری از پیش تعیین شده را برای اثرات نشانگرهای تکنوکلوتیدی در نظر گرفت. این مقدار از پیش تعیین شده می‌تواند با احتمال π اثرات نشانگری را صفر در نظر گرفته و واریانس نشانگرها با احتمال $1-\pi$ توجیه شود. در روش BayesA مقدار π با صفر برابری می‌کند؛ به‌طوری‌که در این روش همه اثرات نشانگری در نظر گرفته می‌شود، درحالی‌که در روش‌های BayesB و BayesC مقدار π بزرگتر از صفر بوده و به‌عنوان یک عامل ناشناخته می‌تواند در ارزیابی روش‌های بیزی اعم از چندمرحله‌ای و تک‌مرحله‌ای حائز اهمیت باشد (۱۶). برخی از محققین به مروری کلی بر استفاده از روش‌های بیزی در انتخاب ژنومی پرداختند؛ اما روش‌های تک‌مرحله‌ای مبتنی بر بیزی (SSBR) را مورد مطالعه قرار ندادند (۱۰، ۳۰). از آنجایی‌که توالی‌یابی کل ژنوم روی تعداد بسیار زیادی از افراد، سخت و هزینه‌بر است و از سوی دیگر نیازمند در اختیار داشتن ارزش ژنتیکی افراد است، انجام مطالعات با استفاده از شبیه‌سازی، روشی منطقی برای چنین امری خواهد بود (۳۲).

1- Genomic Estimated Breeding Value

2- Multiple Step

3- Single Step

4- Single-Step Bayesian Regression

5- Markov chain Monte-Carlo

نسل بعدی (نسل های ۱۰۰۱ تا ۲۰۲۰) برای ایجاد عدم تعادل پیوستگی اندازه جمعیت به صورت تدریجی از ۱۰۰۰ به ۴۰۰ کاهش یافت. سپس ۱۵۰ نر و ۱۵۰ ماده از آخرین نسل از جمعیت پایه گزینش شدند. در این گام، جمعیت با استفاده از آمیزش تصادفی برای ۱۰ نسل دیگر شبیه‌سازی شد. در گام دوم شبیه‌سازی، ۴۰ نر و ۱۰۰۰ ماده به طور تصادفی آمیزش داده شدند و ۱۰ نسل دیگر شبیه‌سازی شد. برای تشکیل جمعیت مرجع، ۴۰۰۰ حیوان از نسل‌های ششم و هفتم انتخاب شدند، به طوری که ۲۰۸۰ فرد انتخاب شده از پدرها و مادرهای نسل ششم و هفتم بودند و ۱۹۲۰ حیوان نیز به طور تصادفی از نسل‌های ششم و هفتم انتخاب شدند. لازم به ذکر است که جمعیت مرجع دارای هر دو رکورد فنوتیپی و ژنوتیپی بودند و افراد نسل‌های هشتم، نهم و دهم نیز به عنوان جمعیت کاندیدا در نظر گرفته شدند به طوری که ۵۰ درصد از افراد در هر نسل از جمعیت تایید به طور تصادفی به عنوان حیوانات دارای ژنوتیپ و افراد باقی‌مانده به عنوان حیوانات ژنوتیپ نشده انتخاب شدند. خلاصه جمعیت شبیه‌سازی شده و پارامترهای به کار رفته در جدول ۱ نشان داده شده است.

پژوهش حاضر با هدف بررسی صحت ارزیابی ژنومی روش‌های چند مرحله‌ای بیزی BayesB, BayesA, BayesC, BayesL و روش‌های تک‌مرحله‌ای بیزی SSBR-A و SSBR-C در مقادیر متفاوت π برای برآورد ارزش‌های اصلاحی ژنومی حیوانات ژنوتیپ شده و نشده در دو توزیع متفاوت اثرات ژنی نرمال و گاما طی سه نسل متوالی انجام شد.

مواد و روش‌ها شبیه‌سازی ژنوم

جمعیت‌ها با استفاده از نرم‌افزار QMSim شبیه‌سازی شدند (۳۱). ابتدا ژنومی حاوی ۴۰۰۰۰ نشانگر تک‌نوکلئوتیدی دوآلی پراکنده شده روی ۲۰ کروموزوم هر کدام به طول ۱۰۰ سانتی‌مورگان شبیه‌سازی شد و توزیع‌های متفاوت اثرات ژنی (اعم از نرمال و گاما) به صورت فرضیات شبیه‌سازی صفت در نظر گرفته شدند و در نتیجه دو صفت با معماری ژنتیکی متفاوت ایجاد شد. در مرحله اول شبیه‌سازی، یک جمعیت پایه به اندازه موثر ۱۰۰۰ حیوان غیرخویشاوند (۵۰۰ نر و ۵۰۰ ماده) در نظر گرفته شد. این جمعیت در ۱۰۰۰ نسل اول آمیزش تصادفی داشت و اندازه آن ثابت باقی ماند، و در ۱۰۲۰

جدول ۱- پارامترهای فرایند شبیه‌سازی

Table 1. Parameters of the simulation process

جمعیت اول/دوم	ساختار جمعیت
۱۰۰۰(۱۰۰۰)	جمعیت اولیه
۲۰۲۰(۴۰۰)	فاز اول تعداد نسل (تعداد افراد)
۱۵۰	فاز دوم تعداد نسل (تعداد افراد)
۱۵۰	جمعیت بسط داده شده
۱۰	تعداد نرها از جمعیت پایه
	تعداد ماده‌های از جمعیت پایه
	تعداد نسل
	جمعیت نهایی
۴۰	تعداد نرها از جمعیت بسط داده شده
۱۰۰۰	تعداد ماده‌های از جمعیت بسط داده شده
۱۰	تعداد نسل
۰/۵۰	احتمال نر بودن نتاج
تصادفی	گزینش و طرح آمیزش
ارزش اصلاحی برآوردی	معیار حذف
۲۰	ژنوم
۱۰۰	تعداد کروموزوم
۴۰۰۰۰	طول هر کروموزوم (سانتی مورگان)
۲۰۰	تعداد نشانگرهای تک‌نوکلئوتیدی
	تعداد QTL ها
	اثر الل‌های QTL
نرمال / گاما (با پارامتر شکل = ۰/۴)	نرخ جهش در نشانگر و QTL ها
$2/5 \times 10^{-5}$	وراثت‌پذیری
۰/۳۵	

نشانگری و e اثر باقیمانده هستند. توزیع چگالی احتمال توام پیشین مجهولات رابطه فوق به صورت زیر است:

$$P(\mu, b, \sigma^2 | df, s, \omega) \sim \left\{ \prod_{j=1}^m P(b_j | \sigma^2) P(\theta_{bj} | \omega) \right\} X^{-2}(\sigma^2 | df, s)$$

در رابطه (۲)، توزیع پیشین یکنواخت برای میانگین، $X^2(\sigma^2 | df, s)$ توزیع کای اسکور معکوس مقیاس‌دار برای

روش‌های چندمرحله‌ای بیزی

مدل آماری در تمامی روش‌های چندمرحله‌ای مورد مطالعه، به صورت ذیل بود:

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

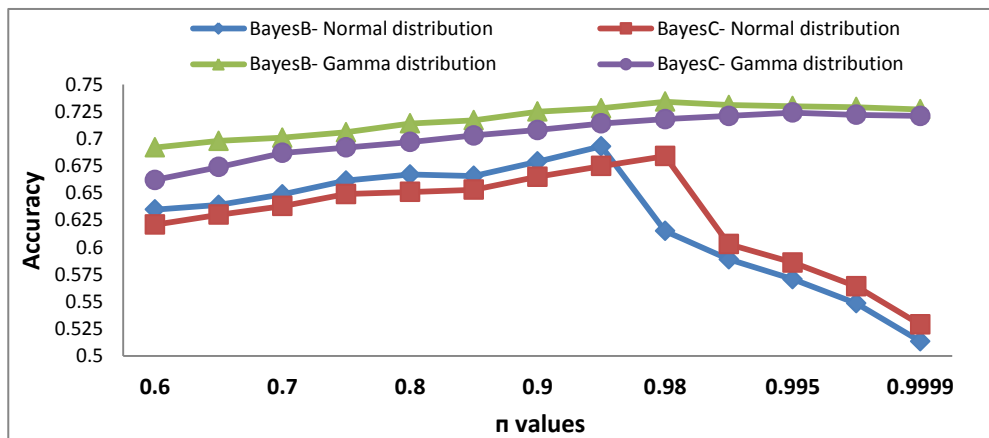
در معادله (۱)، Y بردار مشاهدات شامل افراد تعیین ژنوتیپ شده، μ عرض از مبدا، X_{ij} ژنوتیپ‌های نشانگری، b_j اثر

نتایج و بحث

بررسی تاثیر مقادیر متفاوت π در پیش‌بینی ژنومی

به‌طوری‌که صحت‌های پیش‌بینی شده با افزایش مقادیر π تا مقدار بهینه افزایش و سپس روندی تقریباً ثابت داشت و همواره صحت‌های ژنومی با استفاده از BayesB بالاتر از BayesC تخمین زده شد. نتایج نشان داد که صفات با توزیع گامای اثرات ژنی تحت تاثیر تعداد کمی QTL قرار دارند، به‌طوری‌که می‌توان بخش قابل توجهی از واریانس ژنتیکی را به وسیله آنها توجیه کرد. همچنین، روش BayesB در مقایسه با BayesC، QTL‌های بیشتری با اثرات کوچک را به صفر منقبض کرد، که در نتیجه آن موجب بدام انداختن QTL‌های با اثرات بزرگتر و افزایش صحت‌های پیش‌بینی ژنومی را در توزیع گامای اثرات ژنی به همراه داشت (۳۴). صحت‌های پیش‌بینی شده مشابه روش‌های BayesB و BayesC در مقادیر π بالا نشان داد که این دو روش در این شرایط می‌توانند در هر دو توزیع اثرات ژنی، تعداد QTL‌های کم با اثرات بزرگ را توجیه کنند.

صحت پیش‌بینی ارزش‌های اصلاحی ژنومی برای دو توزیع متفاوت اثرات ژنی مربوط به BayesC و BayesB در مقادیر مختلف π در شکل ۱ نشان داده شده است. مقادیر بهینه π که بالاترین صحت پیش‌بینی ژنومی را با استفاده از روش BayesB دارا بودند به ترتیب ۰/۹۵۰ و ۰/۹۸۰ در دو توزیع نرمال و گامای اثرات ژنی تخمین زده شد. مقادیر بهینه در نظر گرفته شده π در روش BayesC به ترتیب ۰/۹۸۰ و ۰/۹۹۵ در دو توزیع نرمال و گامای اثرات ژنی برآورد گردید و در روش SSBR-C نیز مورد استفاده قرار گرفت. برای توزیع نرمال اثرات ژنی، صحت‌های پیش‌بینی شده با روش‌های BayesB و BayesC مشابه بودند، اما صحت پیش‌بینی شده با افزایش مقادیر π از مقدار بهینه به بعد کاهش یافت. در توزیع گامای اثرات ژنی، الگوی متفاوتی مشاهده گردید،



شکل ۱- میانگین صحت‌های پیش‌بینی ژنومی با استفاده از مقادیر مختلف π در توزیع‌های متفاوت اثرات ژنی نرمال و گاما
Figure 1. The mean prediction accuracies using different π values in the normal and gamma distributions of QTL

مقایسه روش‌های چند مرحله‌ای و تک مرحله‌ای بیزی میانگین صحت پیش‌بینی ارزیابی ژنومی حیوانات ژنوتیپ شده روش‌های مورد مطالعه در جدول ۲ و ۳ نشان داده شد. میانگین صحت پیش‌بینی ژنومی روش‌های چند مرحله‌ای بیزی شامل BayesB (π), BayesB ($\pi=0.98$), BayesA (π Estimation), BayesC ($\pi=0$), BayesC ($\pi=0.995$), BayesC (π Estimation), BayesL و BayesC به ترتیب ۰/۷۱۲، ۰/۷۳۴، ۰/۷۳۳، ۰/۷۲۴، ۰/۷۲۱، ۰/۷۱۶ و ۰/۶۸۷ در حالت توزیع گامای اثرات ژنی و در نسل اول برآورد شد. به‌طوری‌که میانگین پیش‌بینی ژنومی روش‌های تک‌مرحله‌ای بیزی شامل

SSBR-A، SSBR-C ($\pi=0.995$)، SSBR-C ($\pi=0$) و SSBR-C (π Estimation) در شرایطی مشابه به ترتیب ۰/۷۱۸، ۰/۷۷۳، ۰/۷۰۶ و ۰/۷۵۶ به دست آمد. صحت پیش‌بینی ژنومی در روش چندمرحله‌ای BayesB ($\pi=0.98$) نسبت به سایر روش‌های چندمرحله‌ای بیزی (BayesA، BayesC (π Estimation)، BayesB ($\pi=0.995$)) از ۰/۰۱ تا ۰/۰۵ در حالت توزیع گامای اثرات ژنی بیشتر برآورد شد.

جدول ۲- میانگین صحت (خطای استاندارد) پیش‌بینی ژنومی برای حیوانات ژنوتیپ شده در توزیع گامای اثرات ژن‌ها طی سه نسل متوالی
 Table 2. Meanof genomic prediction accuracies and standard errors (in narenthesis) for the genotyped animals in the gamma distribution of QTL effects during three consecutive generations

نسل			روش‌ها
۱۰	۹	۸	
۰/۶۲۸ (۰/۰۰۹)	۰/۶۶۴ (۰/۰۰۹)	۰/۷۱۲ (۰/۰۰۸)	BayesA
۰/۶۶۴ (۰/۰۰۱)	۰/۶۹۱ (۰/۰۰۱)	۰/۷۳۴ (۰/۰۰۹)	BayesB ($\pi=0.98$)
۰/۶۵۷ (۰/۰۰۷)	۰/۶۸۱ (۰/۰۰۷)	۰/۷۲۳ (۰/۰۰۶)	BayesB (π Estimation)
۰/۶۴۵ (۰/۰۰۱)	۰/۶۷۶ (۰/۰۰۱)	۰/۷۲۴ (۰/۰۰۹)	BayesC ($\pi=0.995$)
۰/۶۳۲ (۰/۰۰۲)	۰/۶۶۳ (۰/۰۰۲)	۰/۷۲۱ (۰/۰۰۹)	BayesC ($\pi=0$)
۰/۶۲۵ (۰/۰۰۸)	۰/۶۴۹ (۰/۰۰۷)	۰/۷۱۶ (۰/۰۰۶)	BayesC (π Estimation)
۰/۵۹۳ (۰/۰۰۱)	۰/۶۳۵ (۰/۰۰۱)	۰/۶۸۷ (۰/۰۰۸)	BayesL
۰/۶۰۳ (۰/۰۰۲)	۰/۶۷۴ (۰/۰۰۹)	۰/۷۱۸ (۰/۰۰۸)	SSBR-A
۰/۷۰۱ (۰/۰۰۱)	۰/۷۱۸ (۰/۰۰۱)	۰/۷۷۳ (۰/۰۰۶)	SSBR-C ($\pi=0.995$)
۰/۶۱۶ (۰/۰۰۱)	۰/۶۴۷ (۰/۰۰۱)	۰/۷۰۶ (۰/۰۰۶)	SSBR-C ($\pi=0$)
۰/۶۷۸ (۰/۰۰۲)	۰/۶۹۱ (۰/۰۰۲)	۰/۷۵۶ (۰/۰۰۸)	SSBR-C (π Estimation)

جدول ۳- میانگین صحت (خطای استاندارد) پیش‌بینی ژنومی برای حیوانات ژنوتیپ شده در توزیع نرمال اثرات ژن‌ها طی سه نسل متوالی
 Table 3. Meanof genomic prediction accuracies and standard errors (in narenthesis) for the genotyped animals in the normal distribution of QTL effects during three consecutive generations

نسل			روش‌ها
۱۰	۹	۸	
۰/۶۰۳ (۰/۰۰۹)	۰/۶۲۹ (۰/۰۰۷)	۰/۶۹۱ (۰/۰۰۶)	BayesA
۰/۶۲۱ (۰/۰۰۸)	۰/۶۳۷ (۰/۰۰۸)	۰/۶۹۳ (۰/۰۰۶)	BayesB ($\pi=0.95$)
۰/۶۱۹ (۰/۰۰۹)	۰/۶۳۴ (۰/۰۰۹)	۰/۶۸۶ (۰/۰۰۸)	BayesB (π Estimation)
۰/۶۰۷ (۰/۰۰۸)	۰/۶۲۵ (۰/۰۰۸)	۰/۶۸۴ (۰/۰۰۷)	BayesC ($\pi=0.98$)
۰/۵۹۸ (۰/۰۰۹)	۰/۶۳۱ (۰/۰۰۷)	۰/۶۸۹ (۰/۰۰۳)	BayesC ($\pi=0$)
۰/۵۸۲ (۰/۰۰۱)	۰/۶۲۶ (۰/۰۰۹)	۰/۶۷۶ (۰/۰۰۶)	BayesC (π Estimation)
۰/۶۱۴ (۰/۰۰۹)	۰/۶۳۲ (۰/۰۰۷)	۰/۶۷۶ (۰/۰۰۶)	BayesL
۰/۶۰۷ (۰/۰۰۱)	۰/۶۳۳ (۰/۰۰۹)	۰/۷۰۴ (۰/۰۰۸)	SSBR-A
۰/۶۳۵ (۰/۰۰۱)	۰/۶۶۴ (۰/۰۰۱)	۰/۷۰۶ (۰/۰۰۹)	SSBR-C ($\pi=0.98$)
۰/۶۲۶ (۰/۰۰۸)	۰/۶۵۱ (۰/۰۰۸)	۰/۶۹۸ (۰/۰۰۶)	SSBR-C ($\pi=0$)
۰/۶۲۴ (۰/۰۰۱)	۰/۶۵۷ (۰/۰۰۱)	۰/۷۰۲ (۰/۰۰۸)	SSBR-C (π Estimation)

از رکورد حیوانات تعیین ژنوتیپ شده و نشده به‌طور همزمان استفاده شد، بنابراین، در این روش تعداد مشاهدات بیشتری نسبت به روش‌های چندمرحله‌ای بیزی استفاده می‌شود (۲۱). نادری (۲۶) نیز نشان داد که صحت پیش‌بینی با افزایش درصد حیوانات رکورداری می‌تواند روند صعودی داشته باشد. همانطور که در جداول ۲ و ۳ مشاهده می‌شود، روش SSBR-C ($\pi=0.995$) در حالت توزیع گامای اثرات ژنی (۰/۷۷۳) عملکرد مناسب‌تری نسبت به روش SSBR-C ($\pi=0.98$) در حالت توزیع نرمال (۰/۷۰۶) از خود نشان داد. هنگام در نظر گرفتن توزیع گاما برای اثرات ژنی، برخی از ژن‌ها دارای اثر عمده و درصد بالایی از ژن‌ها اثرات نزدیک به صفر دارند، لذا روش‌هایی که دارای توزیع مزدوج باشند، در حالت توزیع گامای اثرات ژنی، عملکرد بهتری خواهند داشت (۱۱،۲۱). مشابه بودن نتایج حاصل از روش‌های چندمرحله‌ای در پیش‌بینی ارزش‌های اصلاحی ژنومی را می‌توان به معماری ژنتیکی پلی‌ژنی و توزیع نرمال اثرات ژنی مرتبط دانست (۱،۷). در این مطالعه نیز به دلیل عدم تفاوت بین صحت روش‌های چندمرحله‌ای و تک‌مرحله‌ای بیزی در

گودارد (۱۳) گزارش کرد که توزیع پیش فرض روش‌های مبتنی بر BayesB با توزیع گامای ژن‌های عمده اثر هماهنگی دارد که باعث برآورد صحیح‌تر اثرات چندشکلی‌های تک‌نوکلئوتیدی می‌شود، به‌طوری‌که توزیع پیش فرض گاما در روش BayesB برآورد صحیح‌تری از این اثرات ایجاد می‌کند. علاوه بر این، با افزایش فاصله نسل بین جمعیت مرجع و جمعیت تایید صحت ارزش‌های اصلاحی ژنومی به‌شدت کاهش یافت (۱۷،۴). صحت‌های پیش‌بینی با استفاده از روش SSBR-C ($\pi=0.995$) در حالت توزیع گامای اثرات ژنی بالاتر از روش‌های چندمرحله‌ای بیزی و سایر روش‌های تک‌مرحله‌ای بیزی تخمین زده شد، به‌طوری‌که مقدار این برتری از ۰/۰۲ تا ۰/۰۹ برآورد گردید. برتری روش تک‌مرحله‌ای SSBR-C ($\pi=0.995$) بیانگر این واقعیت است که در حالت توزیع گامای اثرات ژنی، تعداد کمی از QTLها بخش عمده‌ای از واریانس ژنتیکی را به خود اختصاص می‌دهد. بنابراین، روش‌های دارای توزیع مزدوج، می‌توانند عملکرد مناسب‌تری را در این شرایط از خود نشان دهند. دلیل دیگر این امر، این است که در روش SSBR-C ($\pi=0.995$)

تاثیر روابط خویشاوندی شجره‌ای و ژنومی بین جمعیت مرجع و تایید

صحت پیش‌بینی ارزش‌های اصلاحی ژنومی برای حیوانات ژنوتیپ نشده در نسل‌های ۸، ۹ و ۱۰ با استفاده از روش‌های تک‌مرحله‌ای بیزی در دو حالت توزیع اثرات ژنی گاما و نرمال در شکل‌های ۲ و ۳ نشان داده شده است. با افزایش فاصله بین جمعیت مرجع و تایید، وضعیت LD بین QTLها و نشانگرها به هم خورده و اثرات QTLها نمی‌توانند به‌وسیله نشانگرها به خوبی توجیه شوند. این نتایج با یافته‌های پیشین نیز تطابق دارد (۱۴،۱۸). نکته دیگر این است که با افزایش فاصله بین نسل مرجع و نسل هشتم و نهم، خطای استاندارد صحت‌های پیش‌بینی ارزش‌های اصلاحی نیز افزایش یافته است. مویر (۲۵) نیز نشان داد که با گذشت چند نسل پس از برآورد اثر نشانگرها، صحت پیش‌بینی ارزش‌های اصلاحی ژنومی کاهش می‌یابد و از کارایی روش‌های برآورد اثرات نشانگری کاسته شده و لازم است اثرات نشانگرها مجدداً برآورد شود.

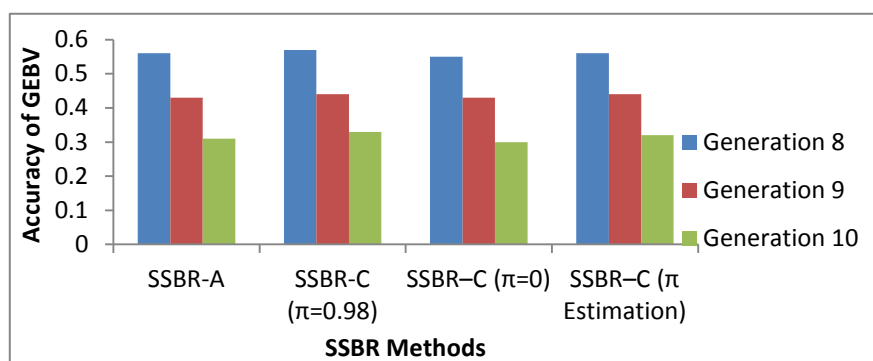
مقایسه روش‌های تک‌مرحله‌ای بیزی

صحت پیش‌بینی ارزش‌های ژنومی چهار روش تک‌مرحله‌ای بیزی SSBR-A، SSBR-C ($\pi=0.995$)، SSBR-C ($\pi=0$) و SSBR-C (π Estimation) در حالت توزیع گاما و نرمال اثرات ژنی در جداول ۲ و ۳ نشان داده شده است. در حالت توزیع گامای اثرات ژنی، روش‌های SSBR-C (π Estimation) و SSBR-C ($\pi=0.995$) عملکرد مناسب‌تری را نسبت به SSBR-A از خود نشان دادند به طوری که این برتری به ترتیب ۰/۰۴ و ۰/۰۶ تخمین زده شد. برتری SSBR-C نسبت به SSBR-A در این مطالعه ناشی از توزیع مزدوجی است که در روش SSBR-C مورد استفاده قرار می‌گیرد و باعث تغییر ساختار نشانگری شده به طوری که در این ساختار اثرات نشانگری به QTLها نزدیک‌تر و دامنه اثرات نشانگری بیشتر تخمین زده می‌شود (۳۵). همچنین، برتری ۰/۰۷ صحت پیش‌بینی ارزش‌های اصلاحی ژنومی توسط روش SSBR-C ($\pi=0.995$) نسبت به روش SSBR-C ($\pi=0$) در حالت توزیع گامای اثرات ژنی نشان می‌دهد که در نظر گرفتن توزیع مزدوج در این شرایط می‌تواند قدرت مدل را در توجیه واریانس ژنتیکی افزایشی و به دام انداختن مقدار بیشتر LD بین نشانگرها و QTLها بهبود بخشد. علاوه بر این، با توجه به در نظر گرفتن تعداد ۲۰۰ QTL در سناریوهای شبیه‌سازی شده و وجود معماری ژنتیکی پلی‌ژنی در حالت توزیع نرمال اثرات ژنی، روش‌های تک‌مرحله‌ای بیزی عملکرد مشابهی داشتند. کارمن و همکاران (۲۰) و کانگ و همکاران (۱۹) نشان دادند که بازدهی روش‌های تک‌مرحله‌ای مبتنی بر BLUP و بیزی تحت توزیع نرمال اثرات ژنی مشابه بود که تاییدی بر نتایج تحقیق حاضر است.

صحت پیش‌بینی ارزش‌های اصلاحی ژنومی برای حیوانات ژنوتیپ نشده در نسل‌های ۸، ۹ و ۱۰ با استفاده از روش‌های تک‌مرحله‌ای بیزی در دو حالت توزیع اثرات ژنی گاما و نرمال در شکل‌های ۲ و ۳ نشان داده شده است. با افزایش فاصله بین جمعیت مرجع و تایید، وضعیت LD بین QTLها و نشانگرها به هم خورده و اثرات QTLها نمی‌توانند به‌وسیله نشانگرها به خوبی توجیه شوند. این نتایج با یافته‌های پیشین نیز تطابق دارد (۱۴،۱۸). نکته دیگر این است که با افزایش فاصله بین نسل مرجع و نسل هشتم و نهم، خطای استاندارد صحت‌های پیش‌بینی ارزش‌های اصلاحی نیز افزایش یافته است. مویر (۲۵) نیز نشان داد که با گذشت چند نسل پس از برآورد اثر نشانگرها، صحت پیش‌بینی ارزش‌های اصلاحی ژنومی کاهش می‌یابد و از کارایی روش‌های برآورد اثرات نشانگری کاسته شده و لازم است اثرات نشانگرها مجدداً برآورد شود.

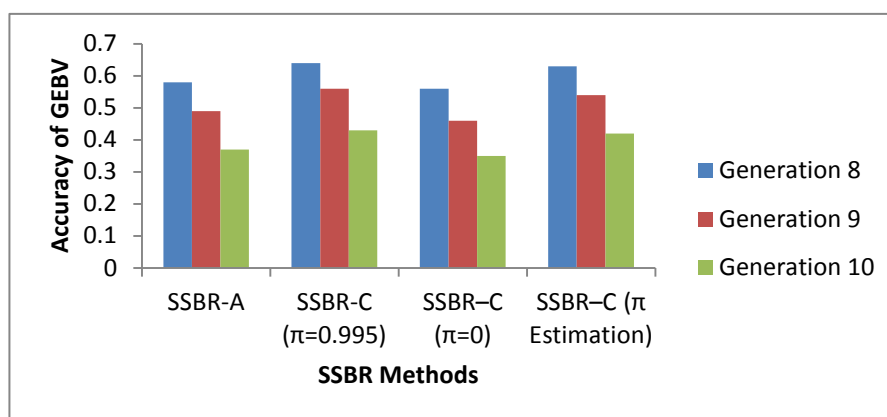
مقایسه روش‌های تک‌مرحله‌ای بیزی

صحت پیش‌بینی ارزش‌های ژنومی چهار روش تک‌مرحله‌ای بیزی SSBR-A، SSBR-C ($\pi=0.995$)، SSBR-C ($\pi=0$) و SSBR-C (π Estimation) در حالت توزیع گاما و نرمال اثرات ژنی در جداول ۲ و ۳ نشان داده شده است. در حالت توزیع گامای اثرات ژنی، روش‌های SSBR-C (π Estimation) و SSBR-C ($\pi=0.995$) عملکرد مناسب‌تری را نسبت به SSBR-A از خود نشان دادند به طوری که این برتری به ترتیب ۰/۰۴ و ۰/۰۶ تخمین زده شد. برتری SSBR-C نسبت به SSBR-A در این مطالعه ناشی از توزیع مزدوجی است که در روش SSBR-C مورد استفاده قرار می‌گیرد و باعث تغییر ساختار نشانگری شده به طوری که در این ساختار اثرات نشانگری به QTLها نزدیک‌تر و دامنه اثرات نشانگری بیشتر تخمین زده می‌شود (۳۵). همچنین، برتری ۰/۰۷ صحت پیش‌بینی ارزش‌های اصلاحی ژنومی توسط روش SSBR-C ($\pi=0.995$) نسبت به روش SSBR-C ($\pi=0$) در حالت توزیع گامای اثرات ژنی نشان می‌دهد که در نظر گرفتن توزیع مزدوج در این شرایط می‌تواند قدرت مدل را در توجیه واریانس ژنتیکی افزایشی و به دام انداختن مقدار بیشتر LD بین نشانگرها و QTLها بهبود بخشد. علاوه بر این، با توجه به در نظر گرفتن تعداد ۲۰۰ QTL در سناریوهای شبیه‌سازی شده و وجود معماری ژنتیکی پلی‌ژنی در حالت توزیع نرمال اثرات ژنی، روش‌های تک‌مرحله‌ای بیزی عملکرد مشابهی داشتند. کارمن و همکاران (۲۰) و کانگ و همکاران (۱۹) نشان دادند که بازدهی روش‌های تک‌مرحله‌ای مبتنی بر BLUP و بیزی تحت توزیع نرمال اثرات ژنی مشابه بود که تاییدی بر نتایج تحقیق حاضر است.



شکل ۲- میانگین صحت پیش‌بینی ارزش‌های اصلاحی ژنومی با استفاده از روش‌های تک‌مرحله‌ای بیزی برای حیوانات ژنوتیپ نشده مربوط به هر نسل از جمعیت تایید در توزیع نرمال اثرات ژن‌ها

Figure 2. Mean of genomic prediction accuracies using single step Bayesian regression for the nongenotyped animals for each validation set in the normal distribution



شکل ۳- میانگین صحت پیش‌بینی ارزش‌های اصلاحی ژنومی با استفاده از روش‌های تک‌مرحله‌ای بیزی برای حیوانات ژنوتیپ نشده مربوط به هر نسل از جمعیت تایید در توزیع گامای اثرات ژن‌ها

Figure 3. Mean of genomic prediction accuracies using single step Bayesian regression for the nongenotyped animals for each validation set in the gamma distribution

احتمالی این مسئله این است که نشانگرها در LD کامل با ژن‌های عامل نبوده و لذا واریانس ژنتیکی کل به‌طور کامل محاسبه نمی‌شود. با توجه به مقادیر برآورد شده، می‌توان گفت که ارزش اصلاحی ژنومی به‌دست آمده توسط روش SSBR-C ($\pi=0$) نسبت به سایر روش‌ها ناریب‌تر بوده است؛ زیرا که ضریب تابعیت آن در حالت توزیع گاما و نرمال اثرات ژنی به‌ترتیب مساوی و نزدیک به یک تخمین زده شد.

میانگین ضریب تابعیت ارزش‌های اصلاحی واقعی به ارزش‌های اصلاحی تخمین زده شده، در جدول ۴ نشان داده شده است. ضریب تابعیت پیش‌بینی ژنومی به‌ترتیب برای روش‌های BayesA و SSBR-C ($\pi=$ chosen) و $1/0.2$ و $0/99$ در حالت توزیع گامای اثرات ژنی و $1/0.4$ و $0/95$ در حالت توزیع نرمال اثرات ژنی تخمین زده شد. ضرایب تابعیت پیش‌بینی ژنومی کمتر از یک نشان می‌دهد که واریانس ارزش‌های اصلاحی ژنومی بیش از حد برآورد شده است. دلیل

جدول ۴- ضریب تابعیت (خطای استاندارد) ارزش‌های اصلاحی واقعی به ارزش‌های اصلاحی تخمین زده شده در توزیع اثرات ژنی نرمال و گاما
Table 4. Regression and standard error (in parenthesis) of true on estimated breeding value in the distribution of normal and gamma QTL effects

توزیع اثرات ژنی		روش‌ها
گاما	نرمال	
۱/۰۲ (۰/۰۳)	۱/۰۴ (۰/۰۲)	BayesA
۱/۰۲ (۰/۰۲)	۱/۰۳ (۰/۰۲)	BayesB (π =chosen) ^a
۱/۰۳ (۰/۰۲)	۱/۰۵ (۰/۰۲)	BayesB (π Estimation)
۱/۰۳ (۰/۰۳)	۱/۰۴ (۰/۰۳)	BayesC (π =chosen) ^b
۱/۰۰ (۰/۰۳)	۱/۰۳ (۰/۰۲)	BayesC (π =0)
۱/۰۲ (۰/۰۱)	۱/۰۵ (۰/۰۱)	BayesC (π Estimation)
۱/۰۳ (۰/۰۳)	۱/۰۵ (۰/۰۲)	BayesL
۱/۰۶ (۰/۰۱)	۱/۰۳ (۰/۰۲)	SSBR-A
۰/۹۹ (۰/۰۲)	۰/۹۵ (۰/۰۳)	SSBR-C (π =chosen)
۱/۰۰ (۰/۰۲)	۰/۹۹ (۰/۰۲)	SSBR-C (π =0)
۰/۹۶ (۰/۰۲)	۰/۹۸ (۰/۰۱)	SSBR-C (π Estimation)

a مقدار π انتخاب شده برای روش BayesB به ترتیب ۰/۹۵۰ و ۰/۹۸۰ در دو توزیع نرمال و گامای اثرات ژنی در نظر گرفته شد.
b مقدار π انتخاب شده برای روش‌های BayesC و SSBR-C به ترتیب ۰/۹۸۰ و ۰/۹۹۵ در دو توزیع نرمال و گامای اثرات ژنی در نظر گرفته شد.

توزیع اثرات ژنی گاما و نرمال در روش SSBR-C ($\pi=0$) نزدیک به یک تخمین زده شد، بنابراین، نسبت به سایر روش‌ها ناریب‌تر بوده و پیشنهاد می‌شود که در توزیع نرمال اثرات ژنی از این روش استفاده شود. علاوه بر این، نتایج این پژوهش نشان داد که افت صحت پیش‌بینی ژنومی با افزایش فاصله نسلی برای افراد ژنوتیپ شده در مقایسه با افراد ژنوتیپ نشده از حساسیت کمتری برخوردار بود.

تشکر و قدردانی

از جناب آقای دکتر مهدی ساعتچی به دلیل اظهار نظرهای ارزشمند ایشان سپاسگزاری می‌شود.

روش‌های تک‌مرحله‌ای بیزی که از مزیت استفاده همزمان از اطلاعات فنوتیپی، ژنوتیپی و شجره بهره‌مند هستند، صحت‌های مشابه و یا بالاتری را نسبت به روش‌های چند مرحله‌ای بیزی از خود نشان دادند. در بین روش‌های چند مرحله‌ای، روش BayesB ($\pi=0.98$) بهترین عملکرد را در حالت توزیع گامای اثر ژن‌ها از خود نشان داد. علاوه بر این، استفاده از روش‌های SSBR-C ($\pi=0.98$) و SSBR-C ($\pi=0$) در حالت توزیع نرمال اثرات ژنی دارای برتری نبود. با توجه به نتایج حاضر، هنگامی که معماری صفات مورد بررسی از مدل تعداد زیاد جایگاه‌های ژنی تبعیت نکند، روش SSBR-C ($\pi=0.995$) و BayesB ($\pi=0.98$) نسبت به روش‌های دیگر ارجحیت دارد. همچنین، ضریب تابعیت در

منابع

1. Abdollahi-Arpanahi, R. 2013. The Impact of allelic architecture of complex traits on genetic evaluations and evolutionary genetics. Ph.D. Dissertation University of Tehran, Tehran.
2. Abdollahi-Arpanahi, R., A. Pakdel, A. Nejati-Javaremi and M.M. Shahrbabak. Comparison of different methods of genomic evolution in traits with different genetic architecture. Journal of Animal Production, 15: 65-77 (In Persian).
3. Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra and S. Tsuruta. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science, 93: 743-752.
4. Calus, M.P. 2009. Genomic breeding value prediction: methods and procedures. Animal, 4(2): 157-164.
5. Chen, L., C. Li, M. Sargolzaei and F. Schenkel. 2014. Impact of genotype imputation on the performance of GBLUP and Bayesian methods for genomic prediction. PLoS ONE, 9(7): e101544.
6. Cheng, H., D.J. Garrick and R.L. Fernando. 2016. JWAS: Julia implementation of whole-genome analyses software using univariate and multivariate Bayesian mixed effects model. Retrieved June 8, 2019 from <http://QTL.rocks>.
7. Colombani, C., A. Legarra, S. Fritz, F. Guillaume, P. Croiseau, V. Ducrocq and C. Robert-Granié. 2012. Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesCp methods for genomic selection in French Holstein and Montbéliarde breeds. Journal of Dairy Science, 96: 575-91.
8. Daetwyler, H.D., K.E. Kemper, J.H. Vander Werf and B.J. Hayes. 2012. Components of the accuracy of genomic prediction in a multi-breed sheep population, Journal of Animal Science, 90(10): 3375-3384.
9. 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.
10. De los Campos, G., A.I. Vazquez, R. Fernando, Y.C. Klimentidis and D. Sorensen. 2013. Prediction of complex human traits using the genomic best linear unbiased predictor. PLoS Genetics, 9: e1003608.

11. Fernando, R.L., J.C. Dekkers and D.J. Garrick. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genetics Selection Evolution*, 46: 50.
12. Gianola, D., G. DeLos Campos, W.G. Hill, E. Manfredi and R. Fernando. 2009. Additive genetic variability and the Bayesian alphabet. *Genetics*, 183: 347-363.
13. Goddard, M. 2009. Genomic selection: prediction of accuracy and maximization of long term response. *Genetica*, 136: 245-257.
14. Habier, D., R.L. Fernando, J.C. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177(4): 2389-2397.
15. Habier, D., J. Tetens, F.R. Seefried, P. Lichtner and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution*, 42(1): 5.
16. Habier, D., R.L. Fernando, K. Kizilkaya and D.J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*, 2: 186-194.
17. 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: 433-443.
18. Hickey, J., S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B. Prasanna, M. Grondona, A. Zambelli, V. Windhausen, K. Mathews and G. Gorjanc. 2014. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Science*, 54: 1476-1488.
19. Kang, H., L. Zhou, R. Mrode, Q. Zhang and J.F. Liu. 2016. Incorporating single-step strategy into random regression model to enhance genomic prediction of longitudinal trait. *Heredity*, 119: 459-467.
20. Karaman, E., H. Cheng, M.Z. Firat, D.J. Garrick and R.L. Fernando. 2016. An upper bound for accuracy of prediction using GBLUP. *PLoS ONE*, 11(8): e0161054.
21. Lee, J., H. Cheng, D. Garrick, B. Golden, J.C. Dekkers, K. Park, D. Lee and R. Fernando. 2017. Comparison of alternative approaches to single - trait genomic prediction using genotyped and non - genotyped Hanwoo beef cattle. *Genetics Selection Evolution*, 49: 2.
22. Legarra, A., I. Aguilar and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92: 4656-63.
23. Meuwissen, T., B. Hayes and M. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157: 1819-29.
24. Moschopoulos, P.G. 1985. The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics*, 37(1): 541-544.
25. 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.
26. Naderi, Y. 2018. The importance of genetic relationships and phenotypic record on genomic accuracy of simulated imputation data via animal models in presence of genotype \times environment interactions. *Research on Animal Production*, 22: 119-130 (In Persian).
27. Nejati-Javaremi, A., C. Smith and J. Gibson. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of animal science*, 75: 1738-45.
28. Park, T. and G. Casella. 2008. The Bayesian lasso. *Journal of the American Statistical Association*, 103: 681-6.
29. Pérez, P. and G. Delos Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198: 483-495.
30. Saheb Alam, H., M. Gholizadeh, H. Hafezian and A. Farhadi. 2017. Comparison of Bayesian methods in the genomic evaluation with different genetic architecture. *Research on Animal Production*, 18: 177-186 (In Persian).
31. Sargolzaei, M. and F.S. Schenkel. 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics*, 25: 680.
32. Toghiani, S., S.E. Aggrey and R. Rekaya. 2016. Multi-generational imputation of single nucleotide polymorphism marker genotypes and accuracy of genomic selection. *Animal*, 10: 1077-85.
33. VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor and F.S. Schenkel. 2009. Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92: 16-24.
34. Wolc, A., J. Arango, P. Settar, J.E. Fulton, N.P. O'Sullivan and J.C. Dekkers. 2016. Mixture models detect large effect QTL better than GBLUP and result in more accurate and persistent predictions. *Journal of Animal Science*, 7: 7.
35. Zhou, L., R. Mrode, S. Zhang, Q. Zhang, B. Li and J. Liu. 2018. Factors affecting GEBV accuracy with single-step Bayesian models. *Heredity*, 120: 100-109.

Comparison of Single and Multi-Step Bayesian Methods for Predicting Genomic Breeding Values in Genotyped and Non-Genotyped Animals-A Simulation Study

Mostafa Madad Jirandeh¹, Jalil Shodja², SadeghAlijani³, Seyed Abbas Rafat² and Jack C.M. Dekkers⁴

1- Ph.D. Student, Department of Animal Sciences, University of Tabriz, Tabriz, Iran

2- Professor, Department of Animal Sciences, University of Tabriz, Tabriz, Iran.

3- Associate Professor, Department of Animal Sciences, University of Tabriz, Tabriz, Iran

(Corresponding author: sad-ali@tabrizu.ac.ir)

4- Professor, Department of Animal Sciences, Iowa state university, Ames, United States of America

Received: July 2, 2019

Accepted: September 11, 2019

Abstract

The purpose of this study was to compare the accuracy of genomic evaluation for Bayes A, Bayes B, Bayes C and Bayes L multi-step methods and SSBR-C and SSBR-A single-step methods in the different values of π for predicting genomic breeding values of the genotyped and non-genotyped animals. A genome with 40000 SNPs on the 20 chromosom was simulated with the same distance (100cM). The π values that maximized the prediction accuracies in BayesC were 0.980 and 0.995 for the normal and gamma distributions of QTL, respectively, and were also used in SSBR-C method. Genomic prediction accuracy in the SSBR-C ($\pi = 0.99$) method was higher than multi step methods from 0.02 to 0.09 for gamma distribution. Results showed that considering mixture distribution and use of phenotype, genotype and pedigree information simultaneously, the SSBR-C ($\pi = 0.99$) method had higher accuracy than other methods and is considered a better choice in this scenario. Moreover, both single and multi-step methods showed similar prediction accuracy when the genetic architecture appeared to approach the normal distribution. Furthermore, SSBR-C ($\pi = 0$) method appeared to be more reliable choice that was due to regressions of true breeding value on estimated breeding value close to one in normal distribution. Generally, GEBV accuracy decreased as the distance increased between validations and training set, which was more sensitive for non-genotyped individuals compared to genotyped individuals.

Keywords: Bayesian, Genomic Accuracy, Genomic Selection, Multi-Step Methods, Simulation