

تأثیر کاهش تراکم نشانگرها بر صحت پیش‌بینی ژنومی روش‌های پارامتری

• خبات خیرآبادی (نویسنده مسئول)

دانشجوی دکتری گروه علوم دامی دانشگاه کشاورزی و منابع طبیعی خوزستان

• جمال فیاضی

دانشیار گروه علوم دامی دانشگاه کشاورزی و منابع طبیعی خوزستان

• هدایت‌اله روشنفکر

استاد گروه علوم دامی دانشگاه کشاورزی و منابع طبیعی خوزستان

• رستم عبداللهی آرپناهی

استادیار گروه علوم دام و طیور پردیس ابوریحان دانشگاه تهران

تاریخ دریافت: شهریور ۱۳۹۶ تاریخ پذیرش: فروردین ۱۳۹۷

شماره تماس نویسنده مسئول: ۰۹۱۸۲۴۷۴۵۰۱

Email: Kheirabadikh@gmail.com

شناسه دیجیتال (DOI): 10.22092/asj.2017.115459.1536

چکیده

افزایش تعداد نشانگرها (p) به تعداد مشاهدات (n)، نخستین چالش انتخاب ژنومی است (مزاحمت ابعاد؛ $n \gg p$). لذا پژوهش حاضر با هدف امکان کاهش مزاحمت ابعاد، به بررسی تأثیر کاهش تعداد نشانگرها بر صحت پیش‌بینی ارزش‌های اصلاحی ژنومیک روش‌های پارامتری پرداخته است. بدین منظور، برای هر فرد ژنومی متشکل از ۱۰۰۰۰ نشانگر تک‌نوکلئوتیدی دو آلی (SNP) با فواصل یکسان روی ۱۰ کروموزوم (هریک به طول ۱۰۰ سانتی‌مورگان) شبیه‌سازی شد. در این پژوهش توزیع‌های متفاوت تأثیرات ژنی (یکنواخت، نرمال یا گاما)، سطوح مختلف وراثت‌پذیری صفت (۰/۰۵، ۰/۲۵، ۰/۴۵ یا ۰/۶۵) و تعداد متفاوت جایگاه‌های ژنی (QTL؛ ۱۰۰ یا ۵۰۰) به عنوان فرضیات شبیه‌سازی در نظر گرفته شد. سپس در یک سناریوی انتخاب ۱۰ یا ۲۰ درصد نشانگرها به‌طور تصادفی انتخاب شدند؛ به‌طوری‌که برای هر یک از جمعیت‌های موجود سه ماتریس نشانگری با ابعاد مختلف (همه نشانگرها، ۱۰ یا ۲۰ درصد نشانگرها) تعریف گردید. مطابق یافته‌های پژوهش حاضر، به‌طور کلی صحت ارزش‌های اصلاحی ژنومیک به ترتیب تحت اثر میزان وراثت‌پذیری صفت، فاصله نسل از جمعیت مرجع، تراکم نشانگرها، توزیع و تعداد QTL‌ها و مدل آماری قرار داشت ($p < 0.001$). به‌طوری‌که در هر سه توزیع آثار ژنی و با هر فرم ساختار نشانگری (کاهش یا عدم کاهش تراکم نشانگرها) مشاهده شد که در نتیجه کاهش وراثت‌پذیری صفت و یا افزایش فاصله از جمعیت مرجع میزان صحت پیش‌بینی‌ها به‌طور چشم‌گیری تقلیل یافت ($p < 0.05$). بررسی اثر کاهش تراکم نشانگرها بر صحت پیش‌بینی ژنومی صفات با معماری ژنتیکی متفاوت نشان داد که جز در مورد صفات با وراثت‌پذیری پائین (۰/۰۵) و تعداد زیاد QTL (۵۰۰) با توزیع ژنی گاما، به‌طور کلی کاهش تراکم نشانگری به‌طور معنی‌داری منجر به افت صحت پیش‌بینی‌های ژنومیک روش‌های پارامتری خواهد شد ($p < 0.05$).

واژه‌های کلیدی: چالش ابعاد، کاهش نشانگرها، صحت ارزیابی‌های ژنومیک، معماری ژنتیک صفت

Animal Science Journal (Pajouhesh & Sazandegi) No 122 pp: 3-16

Impact of reducing the markers density on the genomic evaluation of parametric methodsBy: Khabat Kheirabadi^{1*}, Jamal Fayazi², Hedayatollah Roshanfekar³ and Rostam Abdollahi-Arpanahi⁴

1, 2, 3, Ph.D. student, Associate Professor and Professor, Department of Animal Science, Khuzestan Ramin Agricultural & Natural Resources University

4, Assistant Professor, Department of Animal and Poultry Science, College of Aburairhan, University of Tehran

*Corresponding author: Kheirabadikh@gmail.com Phone: +98-918-247-4501

Received: September 2017**Accepted: April 2018**

The higher number of markers (p) to the number of observations (n) is the first challenge (i.e., course of dimensionality; $p \gg n$) in genomic selection. Therefore with aim to reduce the course of dimensionality, the effect of reduce markers density on the accuracy of genomic prediction breeding values of parametric methods investigated. In this order a genome consisted of 10000 bi-allelic single nucleotide polymorphism (SNP) over 10 chromosomes, with 100 cM length each, was simulated. In this research, the different gene distributions (i.e., uniform, Gaussian and gamma), different levels of heritability (0.05, 0.25, 0.45 or 0.65) and number of quantitative trait loci (QTL; 100 or 500) were considered as assumption of simulation. Then in a selection scenario only 10% of the markers and in the other scenario 20% of the markers were selected (randomly), so that for each of the population three different marker matrices with different dimension (all, 10% and 20% of markers) were defined. According to the finding of this research, in order of preference the accuracy of genomic breeding values is influenced by the trait inheritance, the distance with the reference population, the marker density, the distribution and the number of QTLs and statistical models ($p < 0.001$). So that in all three genetic effects distribution and with each form of the marker structure (selection or without selection of markers) it was observed that the accuracy of predictions declined significantly ($p < 0.05$) as a result of reduction of heritability or with increasing distance from the reference population. The effect of reduce markers density on the genomic prediction accuracy of traits with different genetic architecture showed that except for traits with low heritability (0.05) and high number of QTL (500) and allele effects drawn from a gamma distribution, in general reduction of marker density resulted to decrease accuracy of genomic predictions ($p < 0.05$).

Key words: dimensionality problem, marker selection, accuracy of genomic evaluations, genetic architecture.**مقدمه**

جهت تحقیق در باره فاکتورهای ژنتیکی مؤثر بر صفات کمی (پیچیده) نخستین بار در سال ۱۹۲۳ مطرح شد (Sax, 1923)، اما استفاده از این اطلاعات ژنومی جهت انجام انتخاب عملاً به اوایل قرن بیست و یک میلادی موقوف و تحت عنوان انتخاب ژنومی ارائه گردید (Meuwissen و همکاران، 2001). انتخاب ژنومی که به تصمیم‌گیری انتخاب براساس ارزش اصلاحی ژنومیک

به رغم تلاش‌های علمی جوامع پیشرفته، بشر امروزی توانسته به نقشه‌های نشانگری بسیاری از گونه‌های حیوانی و گیاهی دست یابد (Wang و همکاران، 2005). با وجود نوظهور بودن این منابع اطلاعاتی، گزارشات مختلفی برای استفاده از آنها در حوزه‌های مختلف علمی وجود داشته و شمار آنها به طور چشم‌گیری رو به افزایش است. گرچه مفهوم استفاده از نشانگرها

به‌طور کلی تئوری انتخاب روش مذکور در استفاده از نشانگرهای چندشکلی تک‌نوکلئوتیدی^۱ (SNP)، در دو مرحله انجام و بر مبنای وجود عدم تعادل پیوستگی^۲ (LD) بین نشانگر و جایگاه صفت کمی^۳ (QTL) مجاور استوار است (Meuwissen و همکاران، 2001؛ Goddard and Hayes, 2007). در مرحله اول با تابع کردن مستقیم نشانگرها از رکوردهای مشاهده‌ای افراد جمعیت مرچ (شامل مجموعه‌ای از حیوانات تعیین ژنوتیپ شده و دارای اطلاعات مشاهده‌ای) یا با استفاده از ماتریس روابط ژنومی، اثر تمام نشانگرها به طور همزمان برآورد می‌گردد (De Los Campos و همکاران، 2009). سپس ارزش اصلاحی ژنومیک^۴ (GBV) هریک از افراد جمعیت کاندید (متشکل از حیوانات جوان فاقد اطلاعات مشاهده‌ای) تنها براساس ژنوتیپ نشانگری آنها و با استفاده از ضرایب رگرسیونی از قبل برآورد شده پیش‌بینی می‌گردد (Meuwissen و همکاران، 2001؛ Goddard and Hayes, 2007؛ Hayes و همکاران، 2009). در حقیقت میزان موفقیت انتخاب ژنومی، در گرو صحت برآورد اثرات نشانگری است (Goddard and Hayes, 2009). به‌طوریکه با افزایش صحت برآورد این اثرات، ارزش‌های اصلاحی پیش‌بینی شده برای حیوانات جمعیت کاندید به توان ژنتیکی حقیقی آنها نزدیک‌تر و در نتیجه رتبه‌بندی آنها به شکل صحیح‌تری صورت خواهد گرفت (Meuwissen و همکاران، 2001؛ Goddard and Hayes, 2007؛ Hayes و همکاران، 2009).

به‌طور کلی فاکتورهای زیادی موفقیت انتخاب ژنومی را تحت تأثیر قرار می‌دهند (Meuwissen و همکاران، 2001؛ Goddard and Hayes, 2007؛ Hayes و همکاران، 2009)، که برخی از آنها مانند وراثت‌پذیری صفت یا توزیع آثار جایگزینی QTLها غیرقابل کنترل‌اند. در مورد تأثیرپذیری صحت ارزیابی‌های ژنومی از تعداد نشانگرها، نتایج متفاوتی وجود دارد؛ در برخی از مطالعات گزارش شده که با افزایش تراکم نشانگری می‌توان سهم بیشتری از واریانس ژنتیکی را به دام انداخت (Brito و همکاران، 2011)، درحالی‌که برخی محققین نشان

تعریف شده (Hayes و همکاران، 2009)، عبارت است از بکارگیری تنوع ژنتیکی موجود در سرتاسر ژنوم نمونه‌ای از افراد یک جمعیت جهت تعیین پتانسیل ژنتیکی همه افراد جمعیت. این روش با بررسی دقیق‌تر صفات پیچیده‌ای که تحت کنترل تعداد بسیار زیادی ژن (با اثرات جزئی) قرار دارند (Goddard, 2007؛ Hayes and Yang) توجه محققین جوامع مختلف انسانی (Yang و همکاران، 2010)، گیاهی (Jannink و همکاران، 2010) و حیوانی (Meuwissen و همکاران، 2001؛ Hayes و همکاران، 2009) را به خود جلب کرده است. نتایج شبیه‌سازی بکارگیری این روش در ارزیابی گاوهای شیری نشان داده که با هزینه‌ای معادل ۸ درصد روش مرسوم (روش آزمون نتاج) و با تعیین ارزش اصلاحی و امکان انتخاب گاو نر مستعد قبل از تولد و استفاده از آن به عنوان گاو نر تأیید شده در سن ۲ سالگی (در مقایسه با سن ۵ سالگی، به عنوان حداقل زمان لازم برای تأیید گاو نر در روش آزمون نتاج)، فاصله بین نسل‌ها کاهش و سرعت پیشرفت ژنتیکی حدود ۲ برابر افزایش می‌یابد (Schaeffer, 2006). علاوه بر کاهش هزینه‌های ارزیابی و بهبود سرعت پیشرفت ژنتیکی، امکان برآورد واریانس نمونه‌گیری مندلی از دیگر مزایای روش انتخاب ژنومی است (Visscher و همکاران، 2006؛ Hayes و همکاران، 2009). به‌طوریکه در روش رایج انتخاب (مبتنی بر اطلاعات شجره‌ای)، ضریب خویشاوندی بین تمام افراد تنی (حاصل از والدین غیرخویشاوند) ثابت و برابر ۰/۵ در نظر گرفته می‌شود. این در حالی است که با وجود نمونه‌گیری‌های فاز میوز میزان تشابه بین کروموزوم‌های دو فرد تنی ممکن است بیشتر یا کمتر از ۵۰ درصد باشد (Hayes and Goddard, 2008). در واقع با تفکیک سهم نمونه‌گیری مندلی از ارزش اصلاحی افراد، پتانسیل ژنتیکی متفاوتی برای هریک از کاندیدهای انتخاب (افراد فاقد رکورد) پیش‌بینی می‌گردد. بنابراین، امکان انتخاب همزمان یکسری افراد تنی صرفاً به دلیل داشتن والدین مستعد وجود نداشته که نتیجه نهایی آن حفظ تنوع ژنتیکی (Daetwyler و همکاران، 2007)، و تداوم بیشتر پیشرفت ژنتیکی است (Hayes و همکاران، 2009).

ژن‌ها، میانگین آماره عدم تعادل پیوستگی r^2 (Hayes و همکاران، 2009)، سهمی از واریانس QTL که به وسیله نشانگرها قابل توجیه باشد (Goddard and Hayes, 2007)، برای نسل ۱۰۰ام به حدود ۰/۲۰ برسد (نتایج نشان داده نشده است). سپس افراد نسل ۱۰۱ و نسل‌های ۱۰۲ تا ۱۰۵ با نمونه‌گیری تصادفی از نسل قبل خود به ترتیب به تعداد ۱۰۰۰ و ۲۰۰۰ فرد گسترش یافتند. افراد نسل ۱۰۱ که همه اطلاعات ژنوتیپی و فنوتیپی را داشتند به عنوان جمعیت مرجع و نسل‌های بعدی که فقط دارای اطلاعات ژنوتیپی بودند به عنوان جمعیت‌های تأیید در نظر گرفته شدند. جهت اطمینان از عدم تصادفی بودن نتایج، برای تمام سناریوهای این پژوهش ۱۰ بار تکرار در نظر گرفته شد. در هر یک از این تکرارها و قبل از برآورد ضرایب رگرسیونی، نشانگرهایی با فراوانی آلی نامطلوب (کمتر از ۲٪ یا بیشتر از ۹۸٪) حذف شدند. به‌طوریکه برای صفتی با توزیع یکنواخت ۱۰۰ یا ۵۰۰ QTL، نهایی نشانگرها (p) به ترتیب به ۹۶۷۷ و ۹۳۴۹ عدد کاهش یافت (نتایج مربوط به تکرار ۵ام شبیه‌سازی است).

مدل‌های آماری

در این پژوهش، از یک روش پارامتری خطی (G-BLUP) و چهار روش پارامتری غیرخطی مبتنی بر نمونه‌گیری گیبس (BayesA, BayesB, BayesC & Bayes LASSO) جهت برآورد اثرات نشانگری استفاده شد. در مورد G-BLUP، که مبتنی بر روش بهترین پیش‌بینی ناآریب خطی (Henderson, 1975) است، فرض بر نرمال بودن توزیع اثرات نشانگری (یکسان بودن سهم همه نشانگرها در توجیه کل واریانس ژنتیکی) است (Meuwissen و همکاران، 2001). اما در مورد روش‌های مبتنی بر نمونه‌گیری گیبس، علیرغم مشترک بودن همه آن‌ها برای فرض ناهمگنی واریانس اثرات نشانگری، فرضیات متفاوتی در برآورد ضرایب رگرسیونی دارند. به‌طوریکه در مورد بیز A و B، توزیع شرطی در نظر گرفته شده برای تأثیرات نشانگری از نوع t اما در مورد بیز C (Habier و همکاران، 2011) و LASSO (Park and Casella, 2008) به ترتیب توزیع نرمال و نمائی دوگانه^۷ است.

داده‌اند که به دلیل وجود هم‌راستایی بین اثرات نشانگری افزایش تراکم آن‌ها به کاهش صحت ارزیابی‌های ژنومیک خواهد انجامید (Muir, 2007). در شماری دیگر از پژوهش‌های پیشین نیز گزارش شده که کاهش تراکم نشانگری تأثیر زیادی روی صحت پیش‌بینی‌های ژنومیک نداشته است (Meuwissen و همکاران، 2001؛ Hayes و همکاران، 2003؛ Luan و همکاران، 2009). مفهوم کلی کاهش تراکم نشانگری با حفظ صحت پیش‌بینی‌ها، کم‌رنج کردن چالش‌های انتخاب ژنومی از قبیل هزینه توالی‌یابی، هزینه‌های محاسباتی و چالش ابعاد^۵ است. لذا پژوهش حاضر با در نظر گرفتن معماری‌های متفاوت ژنتیکی، به شبیه‌سازی تغییرات صحت پیش‌بینی ارزش‌های اصلاحی ژنومیک روش‌های پارامتری در پاسخ به کاهش تراکم نشانگرها می‌پردازد.

مواد و روش‌ها

شبیه‌سازی جمعیت

در پژوهش حاضر با استفاده از بسته نرم افزاری *hypred* (Technow, 2013) ژنومی متشکل از ۱۰ کروموزوم با طول یکسان ۱۰۰ سانتی مورگان شبیه‌سازی شد. سپس روی هر یک از این کروموزوم‌ها ۱۰۰۰ نشانگر SNP با فراوانی اولیه ۰/۵ و با فواصل نشانگری یکسان توزیع گردید. به منظور بررسی صفات با معماری ژنتیکی متفاوت، در مورد اثر جایگزینی QTL‌ها از سه فرم یکنواخت (حداقل صفر و حداکثر یک)، گاما (با پارامتر شکل ۰/۴۰ و پارامتر مقیاس ۰/۶۰) و نرمال استاندارد (میانگین صفر و واریانس یک) استفاده شد و برای تعداد QTL‌ها و مقدار وراثت‌پذیری صفت به ترتیب دو (۱۰۰ و ۵۰۰) و چهار (۰/۰۵، ۰/۲۵، ۰/۴۵ و ۰/۶۵) سطح مختلف تعریف گردید. در مورد تمام سناریوهای این پژوهش و به منظور ایجاد عدم تعادل پیوستگی بین نشانگر و QTL، به یک جمعیت پایه با اندازه موثر ۱۰۰ فرد (۵۰ نر و ۵۰ ماده) اجازه داده شد تا در آن برای ۱۰۰ نسل متوالی آمیزش تصادفی صورت گیرد. با فرض اینکه از هر جفت والد تنها دو فرزند حاصل گردد، اندازه جمعیت در هر نسل ثابت و معادل ۱۰۰ فرد باقی ماند. ثابت ماندن اندازه جمعیت و تصادفی بودن آمیزش‌ها سبب شد تا در مورد هر سه فرم توزیع آثار جایگزینی

ارائه شده است (شکل‌های ۱ تا ۳). مطابق یافته‌های پژوهش حاضر، به‌طور کلی صحت ارزش‌های اصلاحی ژنومیک به ترتیب تحت اثر میزان وراثت‌پذیری صفت، فاصله نسل از جمعیت مرجع، تراکم نشانگرها، توزیع و تعداد QTLها و مدل آماری قرار دارد (بر مبنای مقادیر F جدول آنالیز واریانس؛ نتایج ارائه نشده است). لذا به دلیل حداقل تأثیرپذیری صحت مذکور از تعداد QTLها، تنها نتایج مربوط به صفت با تعداد ۱۰۰ جایگاه ژنی ارائه شده است (شکل‌های ۱ تا ۳). همانطور که مشاهده می‌شود، در هر سه توزیع آثار ژنی و با هر فرم ساختار نشانگری (استفاده از تمام نشانگرها یا نسبتی از آنها) در نتیجه کاهش وراثت‌پذیری و یا افزایش فاصله از جمعیت مرجع میزان صحت پیش‌بینی‌ها به طور چشم‌گیری تقلیل می‌یابد ($p < 0.05$). در واقع کاهش صحت ارزیابی‌های ژنومیک با کاهش تنوع ژنتیکی صفت (وراثت‌پذیری)، ناشی از افزایش واریانس برآورد اثرات نشانگری با افزایش واریانس فنوتیپی (افزایش واریانس محیطی و باقی‌مانده‌ها) است که با ازدیاد تعداد نمونه‌های جمعیت مرجع می‌توان سهمی از این کاهش صحت را جبران کرد (Meuwissen و همکاران، 2001؛ Goddard and Hayes، 2009؛ Hayes و همکاران، 2009). تأثیر افزایش فاصله نسلی بین جمعیت‌های مرجع و تأیید بر صحت پیش‌بینی‌های ژنومیک نیز ناشی از ظهور نوترکیبی بین نشانگرها و جایگاه‌های ژنی (یعنی برهم خوردن فاز LD؛ Habier و همکاران، 2007) و در نتیجه تضعیف قدرت نشانگرها برای به دام انداختن آثار ژنی است. به‌رحال در تأیید نتایج پژوهش‌های پیشین (Habier و همکاران، 2007)، مشاهده می‌شود (شکل‌های ۱ تا ۳) که رابطه میزان کاهش صحت ارزیابی‌های ژنومیک با افزایش فاصله بین نسل‌ها به صورت غیرخطی بوده و بیشترین میزان کاهش (نسبت به جمعیت قبلی) مربوط به نسل نخست تأیید می‌باشد که با افزایش وراثت‌پذیری شیب آن کاهش یافت.

به‌طور کلی فرم ماتریسی مدل استفاده شده برای روش‌های مذکور، به صورت زیر نمایش داده می‌شود.

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

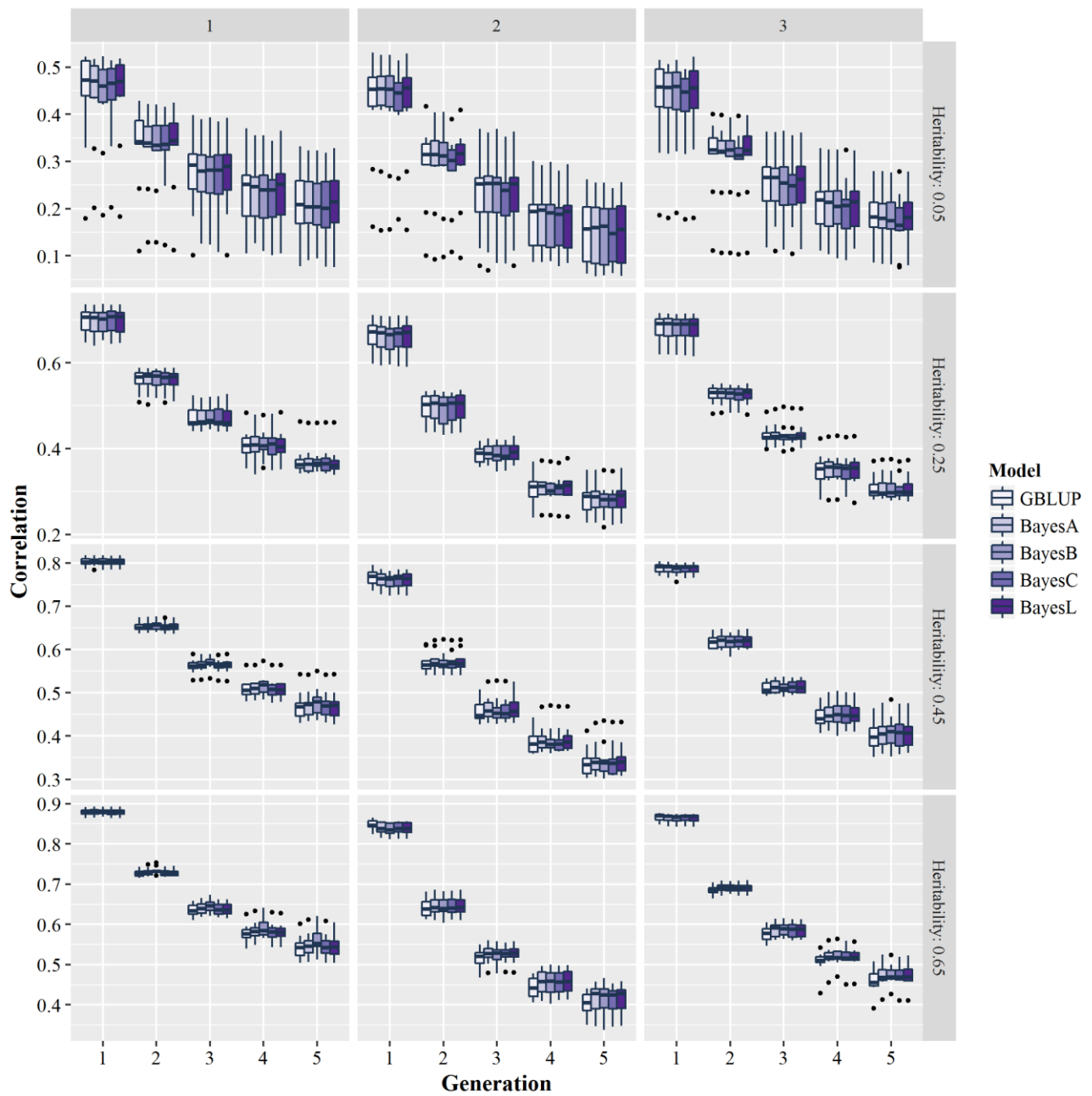
در این رابطه y بردار مشاهدات، μ میانگین کل، X ماتریس ژنوتیپ‌های نشانگری با ابعاد $(n \times p)$ ، g بردار اثرات نشانگری با ابعاد $(p \times 1)$ ، 1 بردار واحد و e بردار مربوط به اثرات باقی‌مانده است. عناصر ماتریس X در حالتی که ژنوتیپ فرد برای یک نشانگر خاص هموزیگوت مغلوب (۱۱)، هتروزیگوت (۱۲) یا هموزیگوت غالب (۲۲) باشد به ترتیب به صورت صفر، یک و دو گذاری می‌گردد (Hayes و همکاران، ۲۰۰۹).

اجرای روش‌های آماری بیزی

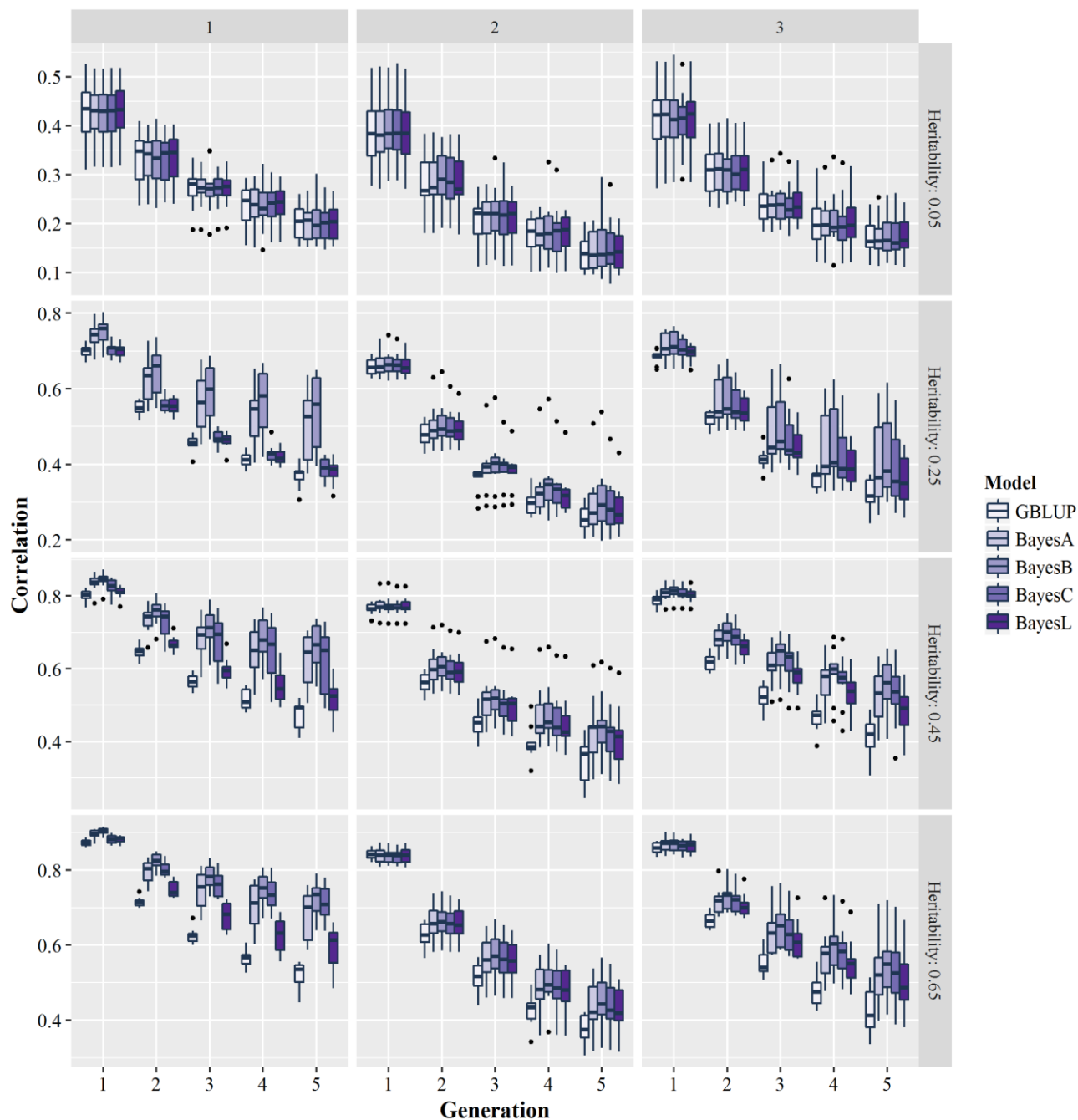
با به‌کارگیری بسته BGLR (Rodriguez, 2016) نرم افزار آماری R، در هر یک از سناریوهای این پژوهش زنجیره‌ای به طول ۱۲۰۰۰ نمونه تعریف گردید. با بررسی وضعیت همگرایی نمونه‌گیری‌های گیس (با رسم پلات برای توزیع واریانس مانده‌ها)، مشخص شد که بعد از حذف ۲۰۰۰ نمونه ابتدایی (دوره سوخته^۱) استنباط‌های توزیع پسین نمونه‌های باقی‌مانده (۱۰۰۰۰ نمونه) قابل اعتماد خواهند بود. پس از برآورد اثرات نشانگری جمعیت مرجع، ارزش اصلاحی ژنومیک^۲ (GEBV) افراد گروه‌های تأیید با توجه به ژنوتیپ آنها به صورت $GEBV = X\hat{g}$ پیش‌بینی گردید (Hayes و همکاران، 2009). در این پژوهش جهت بررسی میزان صحت پیش‌بینی‌ها، از معیار همبستگی بین ارزش‌های اصلاحی واقعی و ارزش‌های اصلاحی ژنومیک و برای مقایسه روش‌های آماری از آزمون t استفاده شد.

نتایج و بحث

به دلیل تنوع ناشی از تکرار ارزیابی‌ها، نتایج پیش‌بینی‌های ژنومیک به صورت نمودار جعبه‌ای و به تفکیک توزیع آثار ژنی



شکل ۱- صحت پیش‌بینی‌های ژنومی روش‌های پارامتری در جمعیت‌های مرجع (۱) و تائید (۲ تا ۵)، با فرض وجود تنها ۱۰۰ QTL مؤثر با اثر جایگزینی یکنواخت در سناریوهای بکارگیری نسبت‌های متفاوت نشانگری (۱: تمام نشانگرها، ۲: استفاده از ۱۰٪ نشانگرها، ۳: استفاده از ۲۰٪ نشانگرها). * داده‌های پرت.



شکل ۲- صحت پیش‌بینی‌های ژنومی روش‌های پارامتری در جمعیت‌های مرجع (۱) و تائید (۲ تا ۵)، با فرض وجود تنها ۱۰۰ QTL مؤثر با اثر جایگزینی گاما در سناریوهای بکارگیری نسبت‌های متفاوت نشانگری (۱: تمام نشانگرها، ۲: استفاده از ۱۰٪ نشانگرها، ۳: استفاده از ۲۰٪ نشانگرها). •: داده‌های پرت.

مختلف نیز ناپدید گردید (گاما (۰/۴۷۶) < نرمال استاندارد (۰/۴۷۵) < یکنواخت (۰/۴۷۳)؛ $p > 0.05$). بنابراین در ادامه جهت مقایسه سناریوهای مختلف تراکم نشانگری، تنها نتایج پیش‌بینی ژنومی بهترین و بدترین مدل آماری (از حیث عددی) برای صفت با توزیع ژنی گاما (دارای بیشترین مقدار صحت پیش‌بینی‌های ژنومی) در نسل نخست تأیید ارائه شده است (جدول ۱). براساس این نتایج، و به‌طور موافق با نتایج پژوهش پیشین (Zhang و همکاران، ۲۰۱۰)، مشاهده می‌شود که به‌طور کلی با افزایش تعداد جایگاه‌های ژنی (از ۱۰۰ به ۵۰۰) صحت پیش‌بینی‌های ژنومی روش GBLUP برخلاف روش‌های بیزی افزایش یافته است. به‌رحال عبداللهی و همکاران (۱۳۹۲) گزارش کردند که با افزایش تعداد QTL از ۵۰ به ۱۰۰ صحت پیش‌بینی ارزش‌های اصلاحی ژنومی روش‌های بیزی (بجز بیزی A) بهبود یافته و با افزایش تعداد QTLها از ۱۰۰ به ۲۰۰ صحت پیش‌بینی‌های ژنومی تمام روش‌های پارامتری (GBLUP و روش‌های مختلف بیزی) کاهش یافته است؛ که با نتایج پژوهش حاضر هم‌خوانی ندارد.

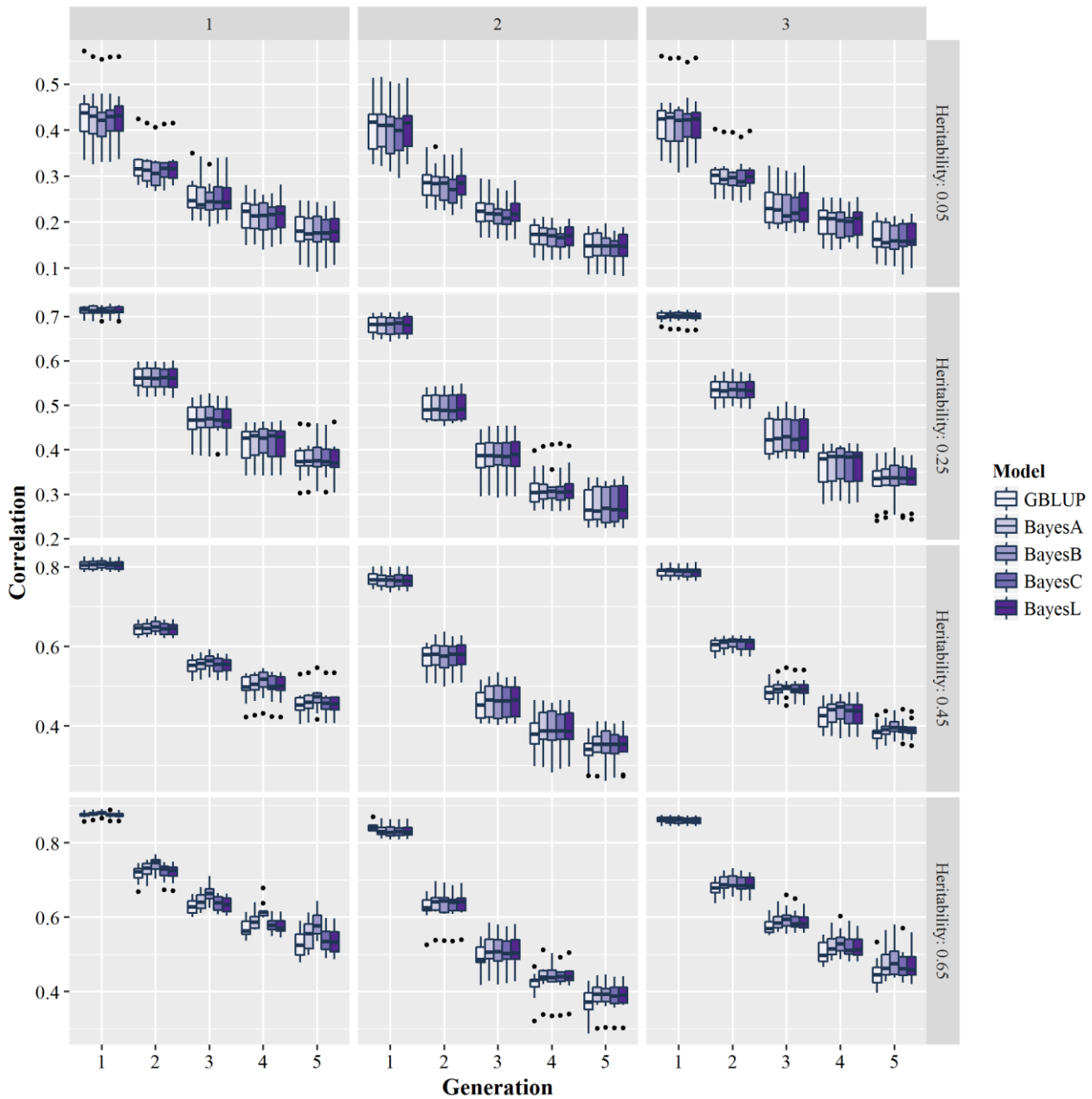
بررسی اثر کاهش تراکم نشانگرها بر صحت پیش‌بینی ژنومی صفات با معماری ژنتیکی متفاوت (با فرض گاما بودن توزیع اثرات جایگزینی ژن‌های کنترل‌کننده صفت)، نشان داد که جز در مورد صفات با تنوع ژنتیکی پائین (وراثت‌پذیری ۰/۰۵) و تحت کنترل تعداد زیاد جایگاه‌های ژنی (۵۰۰ QTL) به‌طور کلی کاهش تراکم نشانگری به‌طور معنی‌دار ($p < 0.05$) منجر به کاهش صحت پیش‌بینی‌های ژنومیک روش‌های پارامتری خواهد شد (جدول ۱). براساس نتایج این پژوهش، دامنه کاهش صحت پیش‌بینی‌های ژنومی روش‌های پارامتری برای استفاده از حداقل تعداد نشانگرها (۱۰ درصد SNPها) نسبت به تمام نشانگرها از ۰/۰۴ (برای صفت با وراثت‌پذیری ۰/۰۵) تا ۰/۱۱ (برای صفت با وراثت‌پذیری ۰/۶۵) برای سناریوی وجود ۵۰۰ QTL و از ۰/۰۴ (برای صفت با وراثت‌پذیری ۰/۰۵) تا ۰/۱۶ (برای صفت با وراثت‌پذیری ۰/۶۵) برای سناریوی وجود ۱۰۰ QTL متغیر است. مشاهده می‌شود که در مورد هر دو رقم تعداد QTLها، میزان

مطابق یافته‌های پژوهش حاضر و در تأیید نتایج تحقیقات پیشین، صحت پیش‌بینی‌های ژنومی روش‌های پارامتری برحسب معماری ژنتیکی صفت متفاوت بود (Hayes و همکاران، ۲۰۰۹؛ VanRaden و همکاران، ۲۰۰۹). به‌طوریکه در مورد صفات با توزیع تأثیرات ژنی یکنواخت علیرغم برتری روش‌های بیزی، در بسیاری از سناریوها (به ویژه برای صفات با وراثت‌پذیری پائین تا متوسط) تفاوت آماری معنی‌داری بین روش‌های پارامتری وجود نداشته است اما در مورد صفات با تأثیر ژنی گاما جز در حالت حداقل تنوع ژنتیکی همیشه روش‌های بیزی (به‌طور ویژه BayesB) شایستگی بهتری نسبت به G-BLUP نشان دادند (شکل ۲). لذا به دلیل متفاوت بودن فرض روش‌های پارامتری برای توزیع اثرات جایگزینی ژن‌ها، در مطالعه داده‌های ژنومی حقیقی با مقایسه روش‌های پارامتری مختلف علاوه بر تعیین مدل سازگار می‌توان اطلاعات مناسبی از معماری ژنتیکی نامعلوم صفات نیز به دست آورد. در پژوهش انجام شده روی داده‌های ژنومی گاوهای شیری نیز گزارش شده که در مورد بسیاری از صفات اقتصادی صحت پیش‌بینی ژنومی روش‌های پارامتری خطی و غیرخطی مشابه بوده و همبستگی بین آن‌ها بیشتر از ۹۹٪ است (VanRaden و همکاران، ۲۰۰۹). به‌رحال نتایج یک مطالعه شبیه‌سازی نشان داده که چنانچه نسبت QTLها به اندازه مؤثر جمعیت کمتر باشد، روش‌های جریمه‌ای توأم با انتخاب متغیر (مانند BayesB) به روش‌های صرفاً جریمه‌ای (مانند G-BLUP) ارجحیت می‌یابند (Daetwyler و همکاران، ۲۰۱۰).

در مورد توزیع آثار جایگزینی QTLها مشاهده شد که به‌طور کلی برای صفات تحت کنترل تعداد کم QTL (در این پژوهش ۱۰۰ جایگاه ژنی) بیشترین و کم‌ترین میانگین صحت پیش‌بینی‌ها (میانگین صحت پیش‌بینی همه روش‌های پارامتری بین چهار جمعیت تأیید، بدون توجه به وراثت‌پذیری صفت و در سناریوی استفاده از تمام نشانگرها) به ترتیب مربوط به توزیع گاما (۰/۵۲۵) و نرمال استاندارد (۰/۴۶۵) است ($p < 0.05$) اما با افزایش تعداد QTLها به سمت ۵۰۰ تفاوت آماری بین توزیع‌های

استفاده از تمام نشانگرها و یا ۲۰ درصد از آن‌ها به ترتیب ۰/۵۶۱ و ۰/۵۳۳ بوده است (اُفت حدود ۳٪ صحت پیش‌بینی‌های ژنومی روش‌های پارامتری؛ جدول ۱) که نزدیک به مقادیر گزارش شده برای داده‌های ژنومی گاوهای شیری نروژ است (به ترتیب ۰/۵۹۱ و ۰/۵۵۴ صحت پیش‌بینی ژنومی برای استفاده از ۱۰۰ یا ۲۵ درصد نشانگرها؛ میزان اُفت حدود ۴٪). همچنین محققین این پژوهش نشان دادند که در همه حالات کاهش تراکم نشانگری، روش‌های جریمه‌ای توأم با انتخاب متغیر (نتایج مربوط به BayesB) دچار اُفت صحت بیشتری هستند، که با نتایج پژوهش حاضر هم‌خوانی دارد. در واقع تأثیرپذیری بیشتر صحت روش‌های جریمه‌ای توأم با انتخاب متغیر از کاهش تراکم نشانگری، ناشی از فرض آنها برای توزیع پیشین اثرات نشانگری است؛ به‌طوری‌که اجباراً تأثیر درصدی از نشانگرها را صفر فرض کرده که نتیجه آن تشدید کاهش تعداد نشانگرها برای روش‌های مذکور خواهد بود.

کاهش صحت ارزیابی‌های ژنومیک برای حالت استفاده از ۲۰ درصد نشانگرها همواره کمتر و حدود نصف مقادیر مشابه برای استفاده ۱۰ درصد از نشانگرها است (جدول ۱). در مطالعه انجام شده روی اطلاعات ژنومی گاوهای شیری نروژ (Luan و همکاران، 2009) میزان صحت پیش‌بینی‌های ژنومیک روش-G BLUP برای صفت شیر (با وراثت‌پذیری حدود ۲۸٪) در حالت‌های مختلف استفاده از ۱۰۰، ۷۵، ۵۰ یا ۲۵ درصد نشانگرهای موجود (با تعداد نشانگر SNPs) به ترتیب ۰/۵۹۱، ۰/۵۸۷، ۰/۵۸۱ و ۰/۵۵۴ گزارش شده که نتایج پژوهش حاضر را از حیث عددی و نیز میزان اُفت صحت پیش‌بینی در نتیجه کاهش تعداد نشانگرها تأیید می‌کند. به‌طوری‌که برای صفات با وراثت‌پذیری ۲۵ درصد (نزدیک به وراثت‌پذیری ۲۸ درصد برای صفت شیر) و تحت کنترل ۵۰۰ جایگاه با توزیع آثار ژنی گاما، مشاهده می‌شود که صحت پیش‌بینی‌های ژنومی برای سناریوهای



شکل ۳- صحت پیش‌بینی‌های ژنومی روش‌های پارامتری در جمعیت‌های مرجع (۱) و تائید (۲ تا ۵)، با فرض وجود تنها ۱۰۰ QTL مؤثر با اثر جایگزینی نرمال استاندارد در سناریوهای بکارگیری نسبت‌های متفاوت نشانگری (۱: تمام نشانگرها، ۲: استفاده از ۱۰٪ نشانگرها، ۳: استفاده از ۲۰٪ نشانگرها).
داده‌های پرت.

جدول ۱- صحت پیش‌بینی روش‌های پارامتری (نخستین نسل تأیید)، برای صفات با تعداد متفاوت QTL (۱۰۰ یا ۵۰۰)

با توزیع آثار ژنی گاما.

صحت [†] (انحراف معیار)		مدل آماری		تراکم نشانگری	وراثت‌پذیری
QTL = ۵۰۰	QTL = ۱۰۰	QTL = ۵۰۰	QTL = ۱۰۰		
۰/۳۳۲ ^{a,a} (۰/۰۳۹)	۰/۳۳۳ ^{a,a} (۰/۰۵۱)	G-BLUP ¹	BayesL ¹	همه نشانگرها	
۰/۳۱۹ ^{a,a} (۰/۰۳۴)	۰/۳۳۱ ^{a,a} (۰/۰۵۶)	BayesB ⁵	BayesB ⁵		
۰/۲۹۴ ^{a,b} (۰/۰۳۳)	۰/۲۹۲ ^{a,c} (۰/۰۵۷)	BayesL	BayesB	۱۰ درصد نشانگرها	۰/۰۵
۰/۲۸۲ ^{a,b} (۰/۰۳۷)	۰/۲۸۲ ^{a,c} (۰/۰۵۹)	BayesC	G-BLUP		
۰/۳۱۸ ^{a,a} (۰/۰۴۲)	۰/۳۰۹ ^{a,b} (۰/۰۵۵)	G-BLUP	BayesA	۲۰ درصد نشانگرها	
۰/۳۰۷ ^{a,a} (۰/۰۳۹)	۰/۳۰۷ ^{a,b} (۰/۰۵۰)	BayesC	BayesC		
۰/۵۶۱ ^{b,a} (۰/۰۳۲)	۰/۶۴۴ ^{a,a} (۰/۰۶۲)	BayesB	BayesB	همه نشانگرها	
۰/۵۵۷ ^{a,ab} (۰/۰۳۰)	۰/۵۵۰ ^{a,bc} (۰/۰۱۸)	BayesC	G-BLUP		
۰/۴۹۶ ^{a,d} (۰/۰۳۶)	۰/۵۰۵ ^{a,d} (۰/۰۶۱)	BayesA	BayesB	۱۰ درصد نشانگرها	۰/۲۵
۰/۴۹۲ ^{a,d} (۰/۰۳۲)	۰/۴۷۸ ^{a,d} (۰/۰۳۲)	G-BLUP	G-BLUP		
۰/۵۳۳ ^{b,bc} (۰/۰۲۸)	۰/۵۷۵ ^{a,b} (۰/۰۶۵)	BayesA	BayesB	۲۰ درصد نشانگرها	
۰/۵۳۰ ^{a,c} (۰/۰۲۷)	۰/۵۱۹ ^{a,c} (۰/۰۲۳)	G-BLUP	G-BLUP		
۰/۶۶۲ ^{b,a} (۰/۰۳۵)	۰/۷۵۶ ^{a,a} (۰/۰۳۵)	BayesB	BayesB	همه نشانگرها	
۰/۶۵۵ ^{a,a} (۰/۰۱۸)	۰/۶۴۶ ^{a,c} (۰/۰۱۸)	G-BLUP	G-BLUP		
۰/۵۸۷ ^{a,c} (۰/۰۱۸)	۰/۶۱۰ ^{a,d} (۰/۰۵۱)	BayesL	BayesB	۱۰ درصد نشانگرها	۰/۴۵
۰/۵۸۰ ^{a,c} (۰/۰۱۸)	۰/۵۶۲ ^{a,e} (۰/۰۲۸)	G-BLUP	G-BLUP		
۰/۶۲۹ ^{b,b} (۰/۰۲۳)	۰/۶۹۵ ^{a,b} (۰/۰۴۱)	BayesB	BayesB	۲۰ درصد نشانگرها	
۰/۶۲۱ ^{a,b} (۰/۰۲۳)	۰/۶۱۷ ^{a,cd} (۰/۰۲۲)	G-BLUP	G-BLUP		
۰/۷۴۵ ^{b,a} (۰/۰۳۰)	۰/۸۲۲ ^{a,a} (۰/۰۲۳)	BayesB	BayesB	همه نشانگرها	
۰/۷۲۷ ^{a,a} (۰/۰۲۷)	۰/۷۱۵ ^{a,b} (۰/۰۱۲)	G-BLUP	G-BLUP		
۰/۶۳۷ ^{a,c} (۰/۰۲۹)	۰/۶۶۴ ^{a,c} (۰/۰۴۴)	BayesL	BayesB	۱۰ درصد نشانگرها	۰/۶۵
۰/۶۲۸ ^{a,c} (۰/۰۲۵)	۰/۶۲۸ ^{a,d} (۰/۰۳۴)	G-BLUP	G-BLUP		
۰/۶۹۴ ^{b,b} (۰/۰۲۸)	۰/۷۲۷ ^{a,b} (۰/۰۳۵)	BayesB	BayesB	۲۰ درصد نشانگرها	
۰/۶۸۶ ^{a,b} (۰/۰۲۵)	۰/۶۶۷ ^{a,c} (۰/۰۲۲)	G-BLUP	G-BLUP		

[†] همبستگی پیرسون بین ارزش‌های اصلاحی واقعی و ارزش‌های اصلاحی ژنومی پیش‌بینی شده برای صفات با تعداد متفاوت ۱۰۰ یا ۵۰۰ QTL. ^۱ بهترین مدل پارامتری برای صفت مورد بررسی، ^۵ بدترین مدل پارامتری برای صفت مورد بررسی. حروف غیرمشترک، تفاوت معنی‌دار بین حالات مختلف در سطح ۵ درصد را نشان می‌دهند (از سمت چپ؛ حرف اول بیانگر تفاوت بین ستون و حرف دوم تفاوت درون ستون است).

نتیجه گیری

براساس یافته‌های پژوهش حاضر، عملکرد روش‌های پارامتری در پاسخ به کاهش تعداد نشانگرها بسته به معماری ژنتیکی صفت متفاوت بوده و در برخی از سناریوها میزان کاهش صحت پیش-بینی‌های ژنومی ناچیز و گاهاً بی‌معنی است (در حالت استفاده از ۲۰ درصد نشانگرها برای صفات با وراثت‌پذیری کم و تحت کنترل تعداد زیاد جایگاه‌های ژنی با توزیع آثار ژنی گاما). بنابراین بنظر می‌رسد که در مطالعه داده‌های ژنومی حقیقی بررسی امکان کاهش تعداد نشانگرها و استفاده از تراشه‌هایی با تراکم نشانگری پائین بسته به معماری ژنتیکی صفت وجود داشته باشد. به‌هرحال جهت افزایش اعتبار نتایج، شبیه‌سازی مطالعات تکمیلی با در نظر گرفتن دیگر فاکتورهای مؤثر بر صحت ارزیابی‌های ژنومیک (از قبیل تعداد افراد مؤثر در جمعیت پایه، تعداد متفاوت افراد جمعیت مرجع، تعداد و نسبت‌های متفاوت نشانگر به جایگاه‌های ژنی) و نیز استفاده از دیگر رویکردهای کاهش تعداد نشانگرها (انتخاب براساس ارزش وزنی آن‌ها) در جهت بهبود نتایج ضروری است.

پانویس

- 1- Single nucleotide polymorphism
- 2- Linkage disequilibrium
- 3- Quantitative trait loci
- 4- Genomic breeding value
- 5- Course of dimensionality
- 6- Single marker regression
- 7- Double exponential
- 8- Burn in
- 9- Genomic Estimated Breeding Value

منابع

عبداللهی آرپناهی، ر.، پاکدل، ع.، نجاتی جوارمی، الف. و مرادی شهربابک، م. (۱۳۹۲). مقایسه روش‌های گوناگون ارزیابی ژنومیک در صفاتی با معماری ژنتیکی متفاوت. تولیدات دامی. شماره ۱، صص ۶۵-۷۷.

Brito, F.V., Neto, J.B., Sargolzaei, M., Cobuci, J.A. and Schenkel, F.S. (2011). Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. *BMC genetics*. 12:80.

Daetwyler, H.D., Pong-Wong, R., Villanueva, B. and Woolliams, J.A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics*. 185:1021-1031.

Daetwyler, H.D., Villanueva, B., Bijma, P. and Woolliams, J.A. (2007). Inbreeding in genome-wide selection. *Journal of Animal Breeding and Genetics*. 124:369-376.

de Los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E. et al. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*. 182:375-385.

de los Campos, G. and Perez Rodriguez, P. (2016). BGLR: Bayesian Generalized Linear Regression. R package version 1.0.5. <https://CRAN.R-project.org/package=BGLR>.

Goddard, M.E. and Hayes, B. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*. 124:323-330.

Goddard, M.E. and Hayes, B.J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature reviews. Genetics*. 10:381.

- Habier, D., Fernando, R.L. and Dekkers, J.C.M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*. 177:2389-2397.
- 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.J., Bowman, P.J., Chamberlain, A.J. and Goddard, M.E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of dairy science*. 92:433-443.
- Hayes, B., Carrick, M., Bowman, P. and Goddard, M. (2003). Genotype \times environment interaction for milk production of daughters of Australian dairy sires from test-day records. *Journal of dairy science*. 86:3736-3744.
- Hayes, B.J. and Goddard, M. (2008). Technical note: Prediction of breeding values using marker-derived relationship matrices. *Journal of animal science*. 86:2089-2092.
- Hayes, B. and Goddard, M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome*. 53:876-883.
- Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*. 423-447.
- Jannink, J.L., Lorenz, A.J. and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in functional genomics*. 9:166-177.
- Luan, T., Woolliams, J.A., Lien, S., Kent, M., Svendsen, M. and Meuwissen, T.H. (2009). The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. *Genetics*. 183:1119-1126.
- Meuwissen, T., Hayes, B. and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157:1819-1829.
- Muir, W. (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*. 103:681-686.
- Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics*. 8:552.
- Schaeffer, L. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*. 123:218-223.
- Technow, F. (2013). R Package hypred: Simulation of Genomic Data in Applied Genetics. *University of Hohenheim*.
- VanRaden, P., Van Tassell, C., Wiggans, G., Sonstegard, T., Schnabel, R., Taylor, J. and Schenkel, F. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of dairy science*. 92:16-24.
- Visscher, P.M., Medland, S.E., Ferreira, M.A., Morley, K.I., Zhu, G. and Cornes, B.K. et al. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS genetics*. 2:e41.
- Wang, W.Y., Barratt, B.J., Clayton, D.G. and Todd, J.A. (2005). Genome-wide association studies: theoretical and practical concerns. *Nature reviews. Genetics*. 6:109.
- Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R. et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature genetics*. 42:565-569.

Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D.J., Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding

values using a trait-specific marker-derived relationship matrix. *PLoS ONE*. 5:e12648.

▪ ▪ ▪ ▪ ▪ ▪ ▪ ▪ ▪ ▪ ▪