

بهبودسازی جمعیت مرجع در پویش ژنومی و ارزیابی ژنومیک

• مرتضی مهدوی

دانش آموخته دکتری ژنتیک و اصلاح نژاد دام و طیور، گروه علوم دامی، دانشکده کشاورزی، دانشگاه زابل، زابل
محقق، شعبه اراک، موسسه تحقیقات واکسن و سرم‌سازی رازی، سازمان تحقیقات، آموزش و ترویج کشاورزی، اراک، ایران.

• غلامرضا داشاب (نویسنده مسئول)

دانشیار گروه علوم دامی، دانشکده کشاورزی، دانشگاه زابل، زابل.

• مهدی وفانی واله

دانشیار گروه علوم دامی، دانشکده کشاورزی، دانشگاه زابل، زابل.

• محمد رکوعی

دانشیار گروه علوم دامی و گروه بیوانفورماتیک، دانشکده کشاورزی، دانشگاه زابل، زابل. تاریخ دریافت: آذر ۱۳۹۷ تاریخ پذیرش: اردیبهشت ۱۳۹۸

شماره تماس نویسنده مسئول: ۰۹۱۵۵۴۴۱۴۷۹

• مهدی سرگلزنی

دانشیار گروه پاتولوژی، دانشگاه گولف و HiggsGene Solutions Inc، گولف، کانادا. Email: dashab@uoz.ac.ir

شناسه دیجیتال (DOI): 10.22092/asj.2019.123551.1838

چکیده

بهبودسازی جمعیت مرجع در ارزیابی‌های ژنومی به دلیل تأثیر آن بر صحت برآورد اثرات نشانگرها و ارزش‌های ارثی ژنومی نقش کلیدی در اصلاح حیوانات اهلی دارد. در مطالعه حاضر هفت روش متفاوت انتخاب جمعیت مرجع شامل انتخاب همه، انتخاب بیشترین و کمترین عملکردها، انتخاب تصادفی، انتخاب افراد با بیشترین و کمترین شباهت نشانگر و ژنگاه مورد ارزیابی قرار گرفت. در مطالعات پویش ژنومی روش انتخاب همه به عنوان جمعیت مرجع، نشانگرهای با فراوانی متوسط و بالا با اثر بزرگ بر روی صفت را تعیین نمود، اما استفاده از روش بیشترین و کمترین عملکردها، نشانگرهای با فراوانی نادر با اثر بزرگ بر روی صفت را گزارش نمود. استفاده هم-زمان از تراکم بالای نشانگری و مرجع انتخابی در مقایسه با تراکم پایین و مرجع کامل، موجب کاهش صحت ارزیابی‌ها گردید، ولی رتبه‌بندی حیوانات را تغییر نداد. بین روش‌های انتخاب جمعیت مرجع با نسل ($P \leq 0.0134$) و همچنین با عدم تعادل لینکاژ ($P \leq 2e-16$) اثر متقابل وجود داشت. به طور کلی انتخاب همه حیوانات به عنوان جمعیت مرجع منتج به صحت‌های بالاتری در مقایسه با روش‌های انتخابی گردید. بین روش‌های انتخاب، در جمعیت‌های با اندازه مؤثر کم ($r^2 = 0.255$, $N_e = 100$) اختلافی وجود نداشت، اما در جمعیت‌های با اندازه مؤثر زیاد ($r^2 = 0.086$, $N_e = 400$) روش‌های انتخاب جمعیت مرجع با صحت‌های متفاوتی ارزش-های اصلاحی ژنومی را پیش‌بینی کردند. بیشترین و کمترین صحت پیش‌بینی ارزش اصلاحی ژنومی به ترتیب متعلق به روش‌های انتخاب دام‌ها بر اساس حداکثر شباهت ژنگاه و نشانگری بود ($P \leq 0.0231$).

واژه‌های کلیدی: نشانگر، ژنتیک، انتخاب، ژنوتایپ، ژنگاه

Animal Science Journal (Pajouhesh & Sazandegi) No 126 pp: 191-204

Optimization of Training Set in Genome Wide Association Study and Genomic EvaluationBy: M. Mahdavi^{1,2}, G. R. Dashab^{*3}, M. Vafaye Valleh⁴, M. Rokouei⁵ and M. Sargolzaei^{6,7}¹Ph.D of Animal Breeding and Genetic, Department of Animal Science, Faculty of Agriculture, University of Zabol, Zabol.² Researcher, Arak Branch, Razi Vaccine and Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Arak, Iran³Associate Professor of Animal Breeding and Genetic, Department of Animal Science, Faculty of Agriculture, University of Zabol, Zabol.⁴Associate Professor of Animal Breeding and Genetic, Department of Animal Science, Faculty of Agriculture, University of Zabol, Zabol.⁵Associate Professor of Animal Breeding and Genetic, Department of Animal Science and Bioinformatics, Faculty of Agriculture, University of Zabol, Zabol.⁶Associate Professor of Animal Breeding and Genetic, Department of Pathobiology, University of Guelph, Guelph, Canada⁷Associate Professor of Animal Breeding and Genetic, HiggsGene Solutions Inc., Guelph, Canada**Received: November 2018****Accepted: April 2019**

The optimization of the reference population in genomic evaluation plays an important role in livestock breeding, because of its potential impact on the accuracy of estimating the marker effects and genomic breeding values. In the present study, seven different train set selection methods including selection of all, selection of the highest and lowest performances, random selection, selection of individuals with the most and least marker and QTL similarity were evaluated. In genome wide association study selection of all as train set detected common SNPs which make a high variation on the trait. However selective train set was just reported rare SNPs with a major effect on the trait. In genomic selection simultaneous use of high-density markers and selective train set in comparison with low-density and selection of all as train set reduced accuracy, but did not change the ranking of animals. There was also an interaction between train set selection method and generation ($P \leq 0.0134$) as well as the linkage disequilibrium ($P \leq 2e-16$). In general, selection of all animals as a train set resulted in higher accuracy compared to six selective train set methods. There were no differences between the methods of selecting train set in populations with a low effective size ($r^2 = 0.255$, $N_e = 100$), but in populations with a high effective size ($r^2 = 0.086$, $N_e = 400$) methods, with different accuracy predicted genomic breeding values. The highest and lowest accuracy were respectively belonged to most QTL and marker similarity methods.

Key words: Marker, Genetic, Selection, Genotyping, QTL**مقدمه**

است (VanRaden و همکاران، ۲۰۰۹؛ Hayes و همکاران، ۲۰۰۹).

پیشنهادها و فرضیات متعددی در ارتباط با خصوصیات جمعیت مرجع و چگونگی بهینه‌سازی آن به خصوص در جمعیت‌های کوچک و صفات با وراثت‌پذیری پایین مطرح شده است (VanRaden و همکاران، ۲۰۰۹). در بسیاری از کشورها و در مورد گاوهای شیرده حیوانات نر به عنوان جمعیت مرجع ژنوتایپ می‌گردند (Loberg and Durr، ۲۰۰۹)، اما تعداد محدود نرها به عنوان جمعیت مرجع ممکن است برای دستیابی به یک ارزیابی ژنومی مناسب کافی نباشد. روش‌های جایگزین متعددی برای رفع

انتخاب بهینه جمعیت مرجع در ارزیابی ژنومی و پوشش ژنومی به دلیل تأثیر آن بر صحت برآورد اثرات نشانگرها و متعاقب آن پیش‌بینی ارزش‌های ارثی ژنومی نقش کلیدی در اصلاح حیوانات اهلی دارد. در دسترس بودن پنل‌های متراکم نشانگری^۱، انتخاب ژنومی را به عنوان یک ابزار نوظهور و نوید بخش در افزایش نرخ پیشرفت ژنتیکی در حیوانات اهلی و نباتات امکان‌پذیر نموده است (Weigel و همکاران، ۲۰۱۰). ژنوتایپ کردن مرتب جمعیت مرجع و تأیید، فرآیند هزینه‌بری است. طراحی یک جمعیت مرجع که اقتصادی و از طرف دیگر حداکثر صحت پیش‌بینی ارزش اصلاحی ژنومی را فراهم نماید، اساس برنامه‌های ارزیابی ژنومی

¹ High-density SNP chip

نمود. بنابراین، از سیستم مرکز اصلاح نژاد دام کشور استفاده شد. برای صرفه‌جویی در زمان، محاسبات به صورت موازی^۲ انجام شدند. خلاصه‌ایی از مراحل مختلف شبیه‌سازی و پارامترهای مورد استفاده در شبیه‌سازی در جدول ۱ ارائه شده است. در اولین مرحله یک جمعیت تاریخی برای ایجاد عدم تعادل لینکاژ مطلوب شبیه‌سازی شد. جمعیت تاریخی برای ۴۰۰۰ نسل و جفتگیری تصادفی و با کاهش تدریجی اندازه مؤثر جمعیت به ۱۰۰ و ۴۰۰ شبیه‌سازی شدند. ژنوم شبیه‌سازی شده شامل ۲۶ کروموزوم غیرجنسی با ۵۰۰۰۰۰ نشانگر (که به طور تصادفی در طول ژنوم جانمایی شدند) و ۸۰۰ جایگاه ژنی دو آللی بود. در فرآیند شبیه‌سازی از داده‌های بانک اطلاعاتی ژنگاه گوسفند^۳ و پارامترهای ژنتیکی برآورد شده استفاده شد تا جمعیت شبیه‌سازی شده تا حد ممکن به شرایط واقعی در گوسفند نزدیک باشد (Hu و همکاران، ۲۰۱۳ Safari؛ Fogarty and، ۲۰۰۳). هموزایگوسیتی نشانگرها و ژنگاه‌ها به منظور حصول اطمینان از ایجاد تعادل جهش-رانس و نیز فراوانی آلل‌های نشانگر و ژنگاه برای کنترل داده‌های ژنوم مورد پایش قرار گرفتند. نشانگرها و ژنگاه‌های تثبیت شده^۴ و همچنین با فراوانی کمتر^۵ از ۰/۰۱ در نسل آخر جمعیت تاریخی از داده‌های ژنوم خارج شدند. در نهایت ۴۰۰۰۰۰ نشانگر و ۶۳۸ ژنگاه در مجموعه ژنوم جهت آنالیزهای بعدی مورد استفاده قرار گرفتند. نوع جهش بازگشتی و نرخ آن $10^{-4} \times 2/5$ تنظیم گردید. میانگین آماره عدم تعادل پیوستگی (لینکاژ) برآورد شده (r^2) با استفاده از پنل‌های ۵۰k در نژادهای گوسفند با اندازه‌های مؤثر مختلف از ۰/۰۸ تا ۰/۲۲ گزارش شده است (Rupp و همکاران، ۲۰۱۶). در این مطالعه تعداد نسل‌ها و اندازه مؤثر جمعیت تاریخی به نحوی انتخاب شدند که در نسل آخر جمعیت تاریخی، عدم تعادل لینکاژ مشابه جمعیت‌هایی با اندازه مؤثر ۱۰۰ و ۴۰۰ رأس حاصل گردد.

در مرحله دوم با بسط جمعیت تاریخی و در نظر گرفتن سیستم جفتگیری به منظور دستیابی به حداکثر همخونی، عدم تعادل لینکاژ در فواصل زیاد^۶ نشانگر و ژنگاه در یک جمعیت معاصر که "Pop1" نامگذاری شد، ایجاد گردید. دو جمعیت "Pop1"

این محدودیت پیشنهاد شده است. به اشتراک‌گذاری داده‌های جمعیت مرجع بین کشورها و جمعیت‌های مختلف، یک روش مؤثر در افزایش اندازه جمعیت مرجع در جمعیت‌های انسانی و حیوانات اهلی است که منجر به بهبود صحت پیش‌بینی ارزش اصلاحی ژنومی می‌گردد (Wiggans و همکاران، ۲۰۰۸؛ Cromie و همکاران، ۲۰۱۰؛ Lund و همکاران، ۲۰۱۰؛ Su و همکاران، ۲۰۱۶). در برخی برنامه‌های ارزیابی ژنومی پیشنهاد شده است که برای افزایش اندازه جمعیت مرجع، ماده‌های برتر (کاندید برای مادران گاو نر) نیز ژنوتایپ گردند (Loberg and Durr، ۲۰۰۹). یکی دیگر از راه‌حل‌ها در مواردی که تعداد کاندیداها زیاد هستند، ژنوتایپ حیوانات با پنل‌های کم تراکم نشانگری است؛ که می‌تواند به طور قابل توجهی هزینه ژنوتایپ کردن را کاهش دهد (Blonk و همکاران، ۲۰۱۰). ژنوتایپ کردن نسبت‌های متفاوتی از یک جمعیت، یکی از متداول‌ترین روش‌هایی است که به منظور انتخاب جمعیت مرجع کارآمد پیشنهاد شده است (Zhao و همکاران، ۲۰۱۲). مطالعات انجام شده تأکید بر انتخاب افراد بر اساس توزیع داده‌های فنوتیپی (توزیع نرمال داده‌ها) دارند که شامل انتخاب یک طرفه از یک طرف منحنی (بیشترین یا کمترین) و یا انتخاب دو طرفه از دو طرف توزیع نرمال داده‌ها می‌باشد (Boligon و همکاران، ۲۰۱۲؛ Nishio و همکاران، ۲۰۱۴).

بنابراین در مطالعه حاضر عملکرد معیارهای مختلف در انتخاب جمعیت مرجع، تراکم نشانگری و اثرات متقابل ساختار جمعیت با معیارهای انتخاب جمعیت مرجع بر صحت پیش‌بینی ارزش‌های ارثی ژنومیک مورد ارزیابی قرار می‌گیرند.

مواد و روش‌ها

شبیه‌سازی یک جمعیت گوسفند با استفاده از نرم‌افزار QMSim نسخه لینوکس (Sargolzaei and Schenkel، ۲۰۰۹) در سه مرحله صورت گرفت. به دلیل حجم بالای اطلاعات نشانگری و نیاز به حافظه موقت زیاد و تعداد زیاد پردازنده محاسباتی، انجام شبیه‌سازی و تجزیه و تحلیل‌ها با سیستم‌های معمولی امکان‌پذیر

² Parallel analysis

³ Sheep quantitative trait locus (QTL) database

⁴ Select segregating loci

⁵ Minor allele frequencies

⁶ Long range LD

روابط ژنگاه معیارهایی بودند که برای انتخاب افراد جهت وارد شدن به جمعیت مرجع استفاده شدند. با استفاده از جمعیت مرجع در هر سناریو و پنل‌های ۴۰۰k، ۲۰۰k، ۱۰۰k و ۵۰k آثار نشانگرها در سه سطح آزمون معنی‌داری (۱٪، ۵٪ و ۱۰٪) برآورد شدند.

در جمعیت‌های شبیه‌سازی شده، نمونه‌های جمعیت ۱ (Pop1) که دارای اطلاعات ژنوتیپی و فنوتیپی بودند به عنوان جمعیت مرجع و جمعیت ۲ (Pop2)؛ افراد جوان فاقد اطلاعات فنوتیپی) به عنوان جمعیت تائید در نظر گرفته شد. لذا ارزش‌های اصلاحی ژنومی نمونه‌های جمعیت ۲ با توجه به اطلاعات ژنومی آن‌ها و بر مبنای ضرایب رگرسیونی اثرات نشانگری (برآورد شده بر مبنای اطلاعات ژنومی و فنوتیپی نمونه‌های موجود در جمعیت ۱) پیش بینی گردید. بدین منظور در یک سناریو اثرات نشانگرها از کل جمعیت شماره ۱ برآورد گردید و در سناریوهای بعدی یک بخش انتخابی (۲۰۰۰ حیوان) از جمعیت شماره یک بر اساس معیارهای انتخابی که به آن اشاره شد به عنوان جمعیت مرجع جهت برآورد اثرات نشانگرها مورد استفاده قرار گرفت. در نهایت ارزش‌های اصلاحی ژنومی و صحت پیش‌بینی‌ها برای جمعیت هدف (جمعیت ۲) در چهار وضعیت متفاوت یعنی نسل‌های جداگانه، میانگین نسل‌های ۱، ۲، ۳، میانگین نسل‌های ۴، ۵، ۶ و میانگین نسل‌های ۷، ۸، ۹ و ۱۰ برآورد شدند. میانگین صحت پیش‌بینی ارزش‌های ارثی ژنومی در سناریوهای مختلف با استفاده از آزمون توکی تصحیح شده^{۱۵} مقایسه شد. SNP1101 قابلیت اطمینان^{۱۶}، مربع همبستگی بین ارزش‌های اصلاحی ژنومی پیش‌بینی شده و واقعی (R^2)، را به منظور ارزیابی صحت پیش‌بینی‌های ژنومی با استفاده از روش Sargolzae و همکاران (۲۰۱۴) محاسبه و گزارش می‌نماید. در این روش قابلیت اطمینان با استفاده از معادله شماره (۲) و واریانس خطای پیش‌بینی تصحیح شده برای زیر مجموعه‌ای از افراد و یا زیر مجموعه‌ای از نشانگرها به جای استفاده از کل افراد و کل نشانگرها محاسبه می‌گردد.

$$R_i^2 = 1 - \frac{PEV_i}{g_{ii}} \quad \text{معادله شماره (۲)}$$

و "Pop2" به صورت جداگانه و با در نظر گرفتن پارامترهای یکسان و طی ۱۰ نسل با استفاده از آخرین نسل "Ld_Pop" ایجاد شدند. داده‌های ژنوم مرتبط با سناریوهایی که در آنها تراکم پائین تری از نشانگرها فرض شده بود با استفاده از کاهش تصادفی^۷ تراکم فایل‌های ۴۰۰k به ۲۰۰k، ۱۰۰k و ۵۰k بدست آمد. به منظور بررسی کیفیت داده‌های ژنوم تولید شده و تأیید اینکه شبیه‌سازی عدم تعادل لینکاژ مورد انتظار را تولید نموده است، LD بین ژنگاه و نشانگرها با استفاده از آماره r^2 (Hill and Robertson, ۱۹۶۸) اندازه‌گیری شد. برای تعیین خصوصیات کاهش LD با افزایش فواصل، متوسط r^2 به عنوان تابعی از فاصله بین نشانگرها مورد استفاده قرار گرفت. جفت نشانگرهای مجاور بر اساس فاصله در دامنه بین ۰ تا ۳۰۰ مگا جفت باز گروه‌بندی شدند. متوسط r^2 برای جفت‌ها در هر فاصله به عنوان میانگین r^2 در آن فاصله منظور گردید.

آثار نشانگری و ارزش‌های اصلاحی ژنومی در هر سناریو با استفاده از روش R-BLUP؛ و مدل زیر با استفاده از نرم‌افزار SNP1101 (Sargolzaei, ۲۰۱۴) برآورد گردید.

معادله شماره (۱)

$$y_i = \mu + \sum_{j=1}^m x_{ij}\beta_j + e_i$$

y_i عملکرد فرد i (تصحیح شده برای اثرات ثابت)، μ میانگین و β_j ضریب رگرسیون (اثر جایگزینی ژن) برای ژنوم SNP و e_i باقیمانده می‌باشد. شکل استاندارد ماتریس نشانگری افراد است. سناریوها طوری طراحی شدند که قادر به تشخیص اختلاف بین معیارهای مختلف انتخاب افراد برای جمعیت مرجع باشند. انتخاب همه حیوانات^۸، انتخاب حیوانات با بیشترین و کمترین عملکردها (انتخاب افراد برای جمعیت مرجع از دو طرف منحنی نرمال داده‌های عملکردی انتخاب شدند^۹)، انتخاب تصادفی حیوانات^{۱۰}، انتخاب افراد با بیشترین^{۱۱} و کمترین^{۱۲} روابط خویشاوندی نشانگر، انتخاب حیوانات با بیشترین^{۱۳} و کمترین^{۱۴}

⁷ Randomly masking

⁸ Full selection

⁹ Extreme selection

¹⁰ rnd

¹¹ Max_rel_mrk

¹² Min_rel_mrk

¹³ Max_rel_qtl

¹⁴ Min_rel_qtl

¹⁵ Adjusted Tukey's test

¹⁶ Reliability

جدول ۱ - خلاصه‌ایی از فرآیندهای شبیه‌سازی جمعیت‌ها

عنوان	ارزش‌ها/توضیحات
تعداد کروموزوم	۲۶ غیر جنسی
مجموع طول کروموزوم‌ها	۲۴۵۱
تعداد نشانگرها	۵۰۰۰۰۰
تعداد ژنگاه	۷۹۸
حداقل مجاز فراوانی آللی	۰/۰۱
نوع جهش/نرخ جهش (نشانگر)	بازگشتی/ 10^{-4} $2/5 \times$
نوع جهش/نرخ جهش (ژنگاه)	بازگشتی/ 10^{-4} $2/5 \times$
وراثت پذیری	۰/۲
وراثت پذیری ژنگاه	۰/۲
واریانس فوتیپی	۱
مرحله اول: جمعیت تاریخی	
شروع نسل از تا تسل	از نسل ۰ تا ۴۰۰۰
اندازه موثر جمعیت	۱۰۰/۴۰۰
مرحله دوم: تولید جمعیت با LD در فواصل زیاد	
تعداد نسل/سیستم جفتگیری	۵ نسل / حداکثر نمودن همخونی
مرحله: سوم	Pop1/pop2
تعداد نسل	10
سیستم جفتگیری	تصادفی
منشاء پایه نرها (ماده‌ها: تعداد / منشاء/نسل نرها: تعداد/منشاء/نسل)	$50(ld_pop)(5)/30(ld_pop)(4)$
منشاء پایه نرها (ماده‌ها: تعداد / منشاء/نسل نرها: تعداد/منشاء/نسل)	$300(ld_pop)(5)/200(ld_pop)(5)$
نرخ جایگزینی نرها	٪۵۰
نرخ جایگزینی ماده‌ها	٪۳۰
اساس طراحی سیستم انتخاب	ارزش اصلاحی برآورد شده (EBV)
اساس حذف افراد	سن

نتایج و بحث

ارزیابی کیفیت داده‌های شبیه‌سازی شده

تغییر نمود. سرعت این تغییرات در مورد ژنگاه‌هایی که دارای اثر بزرگتری بودند بیشتر بود که می‌تواند به دلیل سیستم انتخاب دام‌ها بر اساس ارزش‌های اصلاحی در جمعیت حاضر باشد. نمودار کاهش متوسط LD در فواصل مختلف نشانگری در آخرین نسل جمعیت ld_pop در شکل ۲-الف نمایش داده شده است، که ایجاد LD در فواصل زیاد را تأیید می‌نماید. همچنین دامنه تغییرات LD در فواصل مختلف نشانگرهای مجاور در جمعیت ۱ و ۲ بر روی کروموزوم یک با نرم افزار SNP1101 محاسبه گردید که در شکل ۲-ب و ۲-ج نشان داده شده است بر

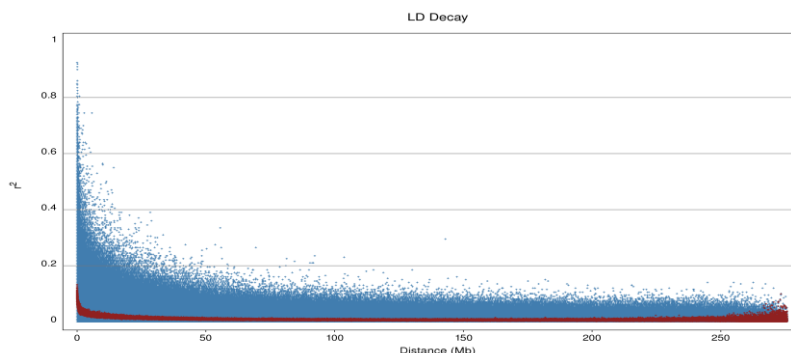
جمعیت تاریخی در نشانگر و ژنگاه در نسل ۴۰۰۰ به تعادل جهش-رانس رسید. متوسط هموزایگوسیتی در نسل ۹ و ۱۰ در هر دو جمعیت مرجع و هدف ۰/۶۶۷ محاسبه شد. بر اساس خروجی‌های QMSim میانگین همخونی در فواصل ۱ سانتی مورگان (فواصل بدون هم‌پوشانی)^{۱۷} در جمعیت ۱ در نسل ۹ و ۱۰ به ترتیب ۰/۰۴۰۷ و ۰/۰۲۰۷ بودند. بررسی الگوهای هموزایگوسیتی در داده‌ها نشان داد که فراوانی آلل‌های نشانگرهایی که در نزدیکی QTLها جانمایی شده بودند و همچنین هموزایگوسیتی در نزدیکی ژنگاه‌ها در هر نسل به سرعت

¹⁷ Non_overlapping windows

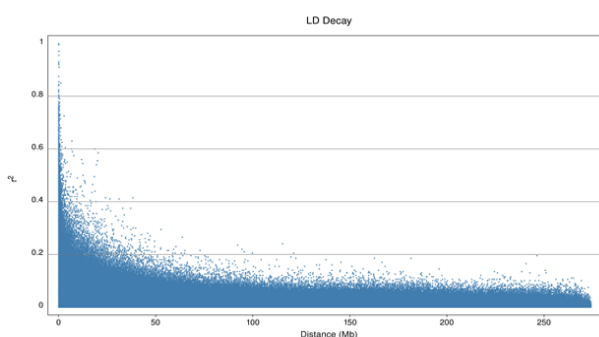
LD در جمعیت شماره ۲ نتایج مشابهی بدنبال داشت و میزان LD در فواصل ۱۵۰ Mb تا ۲۷۲ Mb بین ۰/۰۵ تا ۰/۱۹ متغیر بودند.

اساس نتایج بدست آمده LD در فواصل ۲۰۰ Mb تا ۲۶۸ Mb در جمعیت ۱ در دامنه بین ۰/۰۸ تا ۰/۲ قرار داشت که بیانگر وسعت LD و وجود آن در فواصل نسبتاً زیاد است. همچنین تغییرات

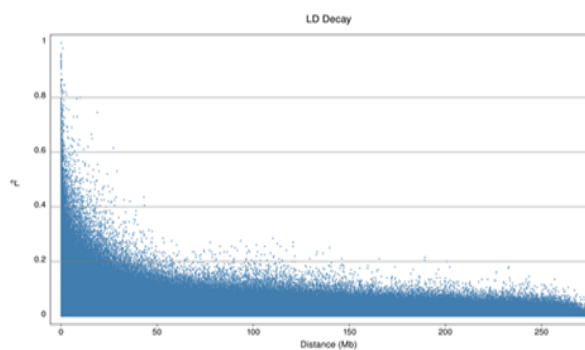
الف :



ج :



ب :



شکل ۲- نمودار LD در فواصل زیاد (الف) و کاهش LD در کروموزوم یک جمعیت مرجع (ب) و هدف (ج) با افزایش فاصله بین نشانگرهای مجاور هم.

نتایج ارزیابی‌های ژنومیک

پیش‌بینی ارزش‌های اصلاحی ژنومی در طی نسل‌های مختلف جمعیت تایید و همچنین میانگین تغییرات صحت ارزیابی در کوتاه مدت، میان مدت و بلند مدت در جدول شماره ۲ ارائه شده است. نتایج همان طوری که در ارزیابی‌های ژنومی مورد انتظار است، نشان داد که صحت ارزش‌های اصلاحی ژنومی با افزایش نسل در هر دو حالت انتخاب یا استفاده از کل نمونه‌های به عنوان جمعیت مرجع کاهش یافت، اما سرعت کاهش در سناریوی جمعیت مرجع انتخابی بیشتر بود. دلیل کاهش صحت ارزیابی‌ها می‌تواند به واسطه افزایش فاصله نسلی بین جمعیت ۱ (جمعیت مرجع) و جمعیت ۲ (تایید) باشد و این افزایش فاصله صحت پیش‌بینی

ارزش‌های اصلاحی ژنومی برای دام‌ها در جمعیت تائید در طی نسل‌های مختلف با نرم‌افزار SNP1101 و با استفاده از اثرات نشانگری برآورد شده از جمعیت مرجع در دو وضعیت شامل کل جمعیت شماره ۱ (سناریوی مرجع کامل) یا انتخاب ۱۰۰۰ حیوان با بالاترین و ۱۰۰۰ حیوان با کمترین عملکرد در جمعیت شماره ۱ (مرجع انتخابی) به عنوان جمعیت مرجع پیش‌بینی گردید. ارزش-های اصلاحی ژنومیکی با استفاده از دو تراکم نشانگری ۵۰k و ۱۰۰k پیش‌بینی گردید.

نتایج تأثیر اندازه جمعیت مرجع شامل انتخاب همه یا انتخاب بیشترین و کمترین عملکردها به عنوان جمعیت مرجع بر صحت

نظر می‌رسد که این سودمندی در یک موازنه‌ای بین کمیت و کیفیت افراد در جمعیت مرجع و تراکم نشانگری مورد استفاده جهت برآورد اثرات نشانگری باشد.

در سناریوی مرجع انتخابی، چگونگی انتخاب حیوانات یکی دیگر از عوامل مؤثر بر صحت پیش‌بینی‌ها است. به منظور تعیین مؤثرترین زیرمجموعه از حیوانات علاوه بر انتخاب حیوانات بر اساس عملکرد، وجود اشتراک نشانگری و نیز ژنگاهی بین حیوانات در یک جمعیت مرجع به عنوان یک عامل بالقوه و تأثیرگذار بر صحت پیش‌بینی ارزیابی‌ها مورد بررسی قرار گرفت. میانگین ده نسل صحت برآورد ارزش‌های اصلاحی ژنومی و خطای پیش‌بینی آن در روش‌های مختلف انتخاب دام‌ها در سناریوی مرجع انتخابی و مقایسه آن با مرجع کامل در جدول ۳ ارائه شده است. نتایج نشان داد که بین روش‌های انتخاب افراد در جمعیت مرجع انتخابی با نسل ($P \leq 0.0134$) و همچنین با عدم تعادل لینکاژ ($P \leq 2e-16$) اثر متقابل وجود داشت.

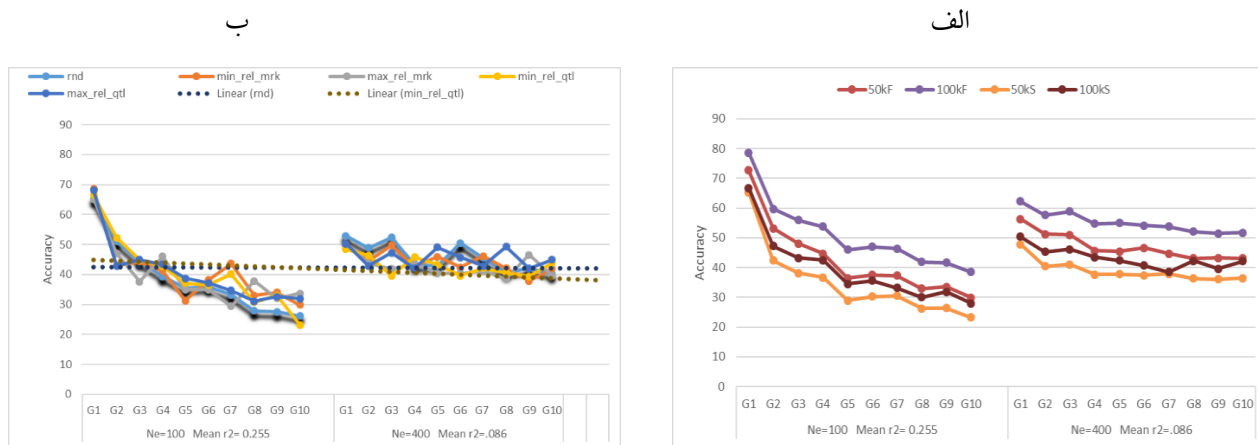
روش انتخاب همه حیوانات به عنوان جمعیت مرجع در هر دو حالت، $r^2 = 0.255$ و $r^2 = 0.086$ در مقایسه با روش‌های مختلف مرجع انتخابی ارزش‌های ارثی را با صحت بالاتری پیش‌بینی نمود. اما مقایسه روش‌های مختلف انتخاب در سناریوی مرجع انتخابی نشان داد که بین روش‌های انتخاب، زمانی که میزان LD بالا است اختلافی وجود ندارد ($r^2 = 0.255$)، اما زمانی که LD کاهش یافت ($r^2 = 0.086$) روش‌های انتخاب جمعیت مرجع با صحت‌های متفاوتی ارزش‌های اصلاحی ژنومی را پیش‌بینی کردند. زمانی که LD پایین است ($r^2 = 0.086$) انتخاب دام‌ها بر اساس حداکثر روابط خویشاوندی ژنگاهی و حداکثر روابط خویشاوندی نشانگری، به ترتیب بیشترین و کمترین صحت پیش‌بینی ارزش اصلاحی ژنومی را داشته است ($P \leq 0.0231$). تفاوت عملکرد روش‌های انتخاب جمعیت مرجع در سطوح متفاوت LD ممکن است به دلیل ساختار متفاوت جمعیت‌ها باشد که به دلیل اندازه مؤثر متفاوت در زمان شبیه‌سازی ایجاد شده است. همچنین، نتایج نشان داد که شباهت ژنتیکی بیشتر جمعیت‌ها، منجر به افزایش صحت پیش‌بینی‌ها شد که می‌تواند به دلیل اختلاف میزان

ارزش‌های اصلاحی را در هر دو سناریو کاهش داد (شکل ۳- الف). البته با توجه به این که هر دو جمعیت مرجع و تایید از یک جمعیت تاریخی منشأ گرفته‌اند، لذا دلیل اصلی این کاهش صحت، می‌تواند شکسته شدن فاز LD بین جایگاه ژنی و نشانگرها در طول زمان باشد. جمعیت‌های مذکور در نسل‌های ابتدایی (۱، ۲ و ۳) دارای روابط خویشاوندی و LD بالاتری در مقایسه با نسل‌های آخری (۴ تا ۱۰) بودند. نتایج نشان داد که با افزایش فاصله جمعیت‌ها از جمعیت پایه "Id_Pop" درجه خویشاوندی بین جمعیت مرجع و تایید کاهش یافت. انتخاب برخی از نمونه‌های موجود در جمعیت مرجع و استفاده از آن‌ها به جای تمام نمونه‌های جمعیت مرجع، موجب کاهش تعداد مشاهدات به تعداد مجهولات می‌گردد که در حوزه آمار تحت عنوان چالش ابعاد مطرح شده است. این چالش احتمالاً دلیل تسریع کاهش صحت پیش‌بینی‌های ژنومی در سناریوی مرجع انتخابی است.

همانطوریکه در جدول شماره ۲ نشان داده شده است افزایش تراکم نشانگرها از پنل ۵۰k به ۱۰۰k در هر دو سناریو (مرجع کامل و مرجع انتخابی) و کلیه نسل‌ها موجب افزایش صحت پیش‌بینی ارزش‌های اصلاحی ژنومی شد ($P \leq 0.01$). تغییرات صحت پیش‌بینی ارزش‌های ارثی با افزایش نسل در هر دو حالت $Ne = 100$ و $Ne = 400$ برای سناریوی مرجع انتخابی با تراکم نشانگری ۵۰k و ۱۰۰k (۵۰kS و ۱۰۰kS) و سناریوی مرجع کامل با تراکم نشانگری ۵۰k و ۱۰۰k (۵۰kF و ۱۰۰kSF) در شکل ۳- الف نشان داده شده است. استفاده همزمان از تراکم بالا (۱۰۰k) و سناریوی مرجع انتخابی (بیشترین و کمترین عملکردها) در مقایسه با تراکم پایین (۵۰k) و مرجع کامل، اگرچه موجب کاهش صحت ارزیابی‌ها گردید، ولی رتبه بندی حیوانات را تغییر نداد. همچنین تفاضل صحت ارزش‌های اصلاحی ژنومی پیش‌بینی شده توسط سناریو ۵۰kF از سناریوی ۱۰۰kS طی نسل‌های اول تا دهم کاهش یافت. بنابراین، استفاده از استراتژی جمعیت مرجع انتخابی در شرایطی که تراکم نشانگرها بالا است ممکن است به اندازه سناریوی جمعیت مرجع کامل و تراکم پایین نشانگری در پیش‌بینی ارزش‌های اصلاحی ژنومی مفید باشند. به

در مقابل در جمعیت‌هایی با LD پایین صحت پیش‌بینی ارزش-های اصلاحی از پایداری بیشتری در طول نسل برخوردار بود (شکل ۳-ب).

LD در جمعیت‌های با درجات متفاوت خویشاوندی بین افراد باشد. صحت پیش‌بینی ارزش اصلاحی ژنومی در طول نسل‌ها در جمعیت‌های با LD بالا پایدار نبود و با افزایش نسل تنزل یافت. اما



شکل ۳- صحت پیش‌بینی ارزش‌های اصلاحی ژنومی با استفاده از استراتژی جمعیت مرجع کامل (الف) و مرجع انتخابی (ب).

تعداد بیشتری از خانواده‌ها و زیرجمعیت‌ها را در ارزیابی ژنتیکی فراهم می‌نماید. بنابراین به نظر می‌رسد استفاده از پنل‌های با تراکم کم از نشانگرهای خاص، برای ژنوتایپ کردن کل افراد جمعیت استراتژی خوبی باشد. این بررسی اولیه از کل جمعیت، امکان شناسایی افراد مؤثر و انتخاب آنها به عنوان جمعیت مرجع را فراهم می‌نماید.

روش انتخاب جمعیت مرجع بر اساس بیشترین و کمترین عملکردها، در مطالعات قبلی (Boligon و همکاران، ۲۰۱۲؛ Jimenez-Montero و همکاران، ۲۰۱۲) به عنوان یک روش مناسب برای انتخاب مرجع توصیه شده است. مقایسه روش‌های انتخاب جمعیت مرجع در این مطالعه نشان داد که استفاده از روش انتخاب بیشترین و کمترین عملکرد علی‌رغم اینکه در حالت $r^2 = 0.255$ منجر به بهبود جزئی در صحت ارزش‌های ارثی ژنومی گردید، اما در حالت $r^2 = 0.086$ ارزش‌های ارثی ژنومی را با صحت کمتری در مقایسه با روش انتخاب حداکثر روابط ژنگاهی پیش‌بینی نمود.

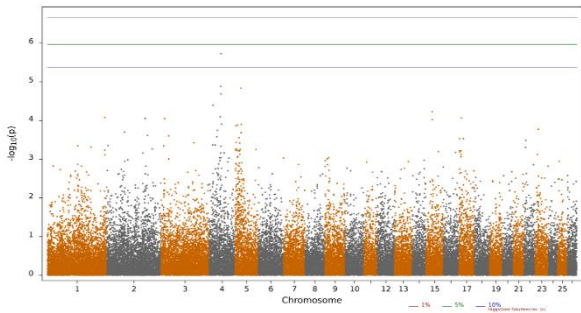
نتایج پویش ژنومی: استفاده از پنل‌های با تراکم‌های مختلف (تراکم-های بسیار بالا تا تراکم‌های کم) در مطالعات پویش ژنومی و مقایسه نتایج حاصل، ابزار مناسبی برای تعیین نشانگرهای حاوی اطلاعات مفید^{۱۸} و تولید پنل‌های تجاری تقلیل تراکم یافته^{۱۹} است. این پنل‌ها شامل نشانگرهای است که حاوی مفیدترین اطلاعات از نظر فراوانی، اثر و لینکاژ با جایگاه‌های ژنی کنترل‌کننده صفت است. در این مطالعه با استفاده از جمعیت مرجع کامل و پنل ۴۰۰k آثار نشانگرها برآورد گردید و نشانگرهای معنی‌دار در سه نرخ معنی‌داری کاذب نشانگرها (۱٪، ۵٪ و ۱۰٪) تعیین شدند. اثرات نشانگرها همچنین با کاهش تراکم پنل ۴۰۰k به پنل‌های ۲۰۰k، ۱۰۰k و ۵۰k و با استفاده از جمعیت مرجع کامل و مرجع انتخابی مجدداً برآورد شدند.

با توجه به نتایج بدست آمده بهترین معیار برای انتخاب جمعیت مرجع بستگی به معماری صفت، ساختار و اندازه جمعیت و اثرات متقابل آنها دارد. نتایج این مطالعه بیشتر بر تأثیر روابط خویشاوندی ژنتیکی تأکید دارد، درحالی‌که که Jin و همکاران (۲۰۰۴) و Jannink (۲۰۰۵) گزارش کردند وجود اطلاعات زمینه‌ای از نشانگرهایی که در عدم تعادل با ژنگاه‌های کنترل‌کننده صفت هستند امکان تمرکز بر مناطق ژنتیکی خاص را فراهم و موجب افزایش کارایی استراتژی مرجع انتخابی می‌گردد. همچنین نتیجه‌گیری کردند که یک مجموعه ۳۰۰ تا ۱۰۰۰ نشانگری می‌تواند بخش عمده‌ای از دستاوردهای مورد نظر در پیش‌بینی صحت ارزیابی‌ها را تامین نماید. استفاده از پنل‌های نشانگری با تراکم کم و هزینه پایین امکان بررسی

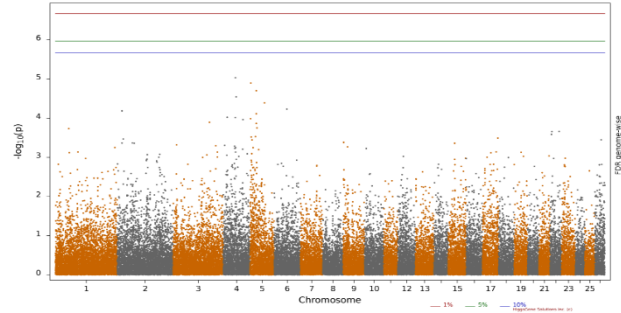
¹⁸ Most informative markers

¹⁹ Reduced Panel

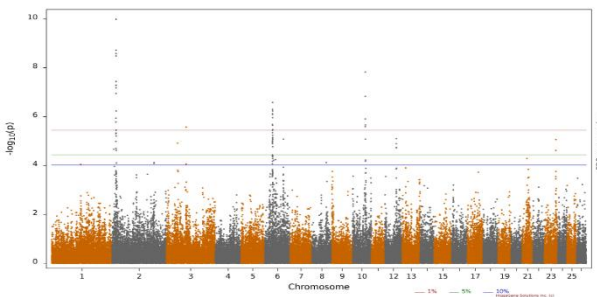
الف-۱



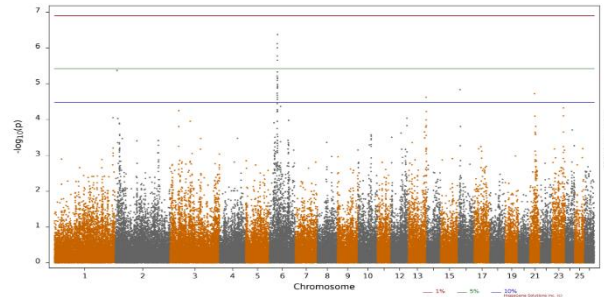
ب-۱



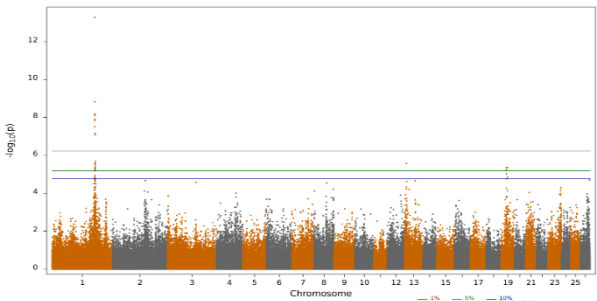
الف-۲



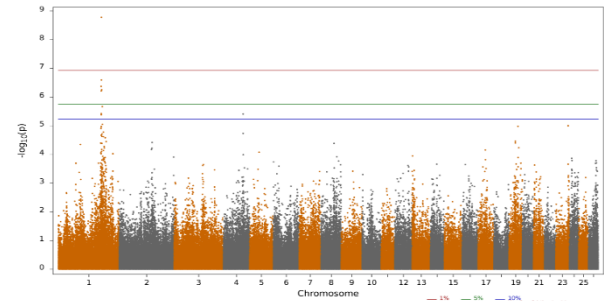
ب-۲



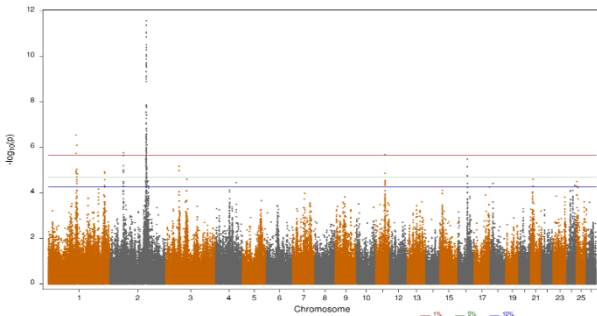
الف-۳



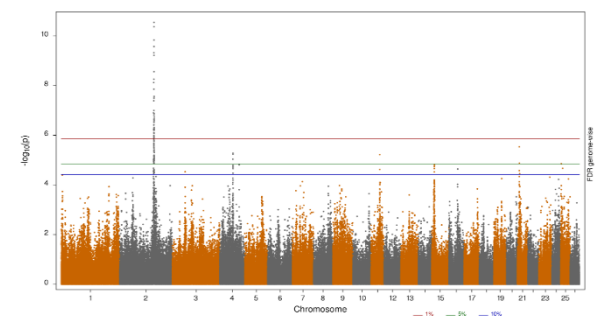
ب-۳



الف-۴



ب-۴



شکل ۴- نمودار $-\log(p\text{-value})$ اثرات نشانگرها با استفاده از جمعیت مرجع کامل (الف) و جمعیت مرجع انتخابی (ب)

جدول ۲- صحت پیش‌بینی ارزش‌های اصلاحی ژنومیکی در کوتاه مدت، میان مدت و درازمدت
با استفاده از سناریوهای مرجع کامل و مرجع انتخابی^۱ و سناریوی تراکم نشانگرها

سناریو		مرجع کامل				مرجع انتخابی			
تراکم نشانگر	تراکم نشانگری ۵۰k	تراکم نشانگری ۱۰۰k	تراکم نشانگری ۵۰k	تراکم نشانگری ۱۰۰k	تراکم نشانگری ۱۰۰k	تراکم نشانگری ۵۰k	تراکم نشانگری ۱۰۰k	تراکم نشانگری ۵۰k	تراکم نشانگری ۱۰۰k
نسل	صحت	تعداد حیوانات	صحت	تعداد حیوانات	صحت	تعداد حیوانات	صحت	تعداد حیوانات	تعداد حیوانات
۱	۷۲/۷±۰/۵۵	۵۵۹	۶۵/۳±۰/۹۷	۵۷۴	۶۶/۸±۰/۹۳	۵۷۴	۶۶/۸±۰/۹۳	۵۷۴	۵۷۴
۲	۵۳/۲±۰/۴۵	۵۷۵	۴۲/۴±۰/۴۹	۵۶۱	۴۷/۲±۰/۶۵	۵۶۱	۴۷/۲±۰/۶۵	۵۶۱	۵۶۱
۳	۴۸±۰/۴۳	۵۶۸	۳۸/۲±۰/۳۶	۵۵۹	۴۳/۲±۰/۴۸	۵۵۹	۴۳/۲±۰/۴۸	۵۵۹	۵۵۹
۴	۴۴/۷±۰/۳۳	۵۷۰	۳۶/۷±۰/۳۱	۵۶۴	۴۲/۵±۰/۴۵	۵۶۴	۴۲/۵±۰/۴۵	۵۶۴	۵۶۴
۵	۳۶/۵±۰/۲۴	۵۵۹	۲۸/۹±۰/۲۵	۵۸۵	۳۴/۵±۰/۲۷	۵۸۵	۳۴/۵±۰/۲۷	۵۸۵	۵۸۵
۶	۳۷/۵±۰/۲	۵۶۳	۳۰/۳±۰/۲۳	۵۷۷	۳۵/۵±۰/۲۷	۵۷۷	۳۵/۵±۰/۲۷	۵۷۷	۵۷۷
۷	۳۷/۳±۰/۱۸	۵۷۰	۳۰/۵±۰/۲	۵۶۴	۳۳/۲±۰/۲	۵۶۴	۳۳/۲±۰/۲	۵۶۴	۵۶۴
۸	۳۲/۹±۰/۱۴	۵۶۶	۲۶/۲±۰/۱۶	۵۵۷	۳۰/۱±۰/۱۸	۵۵۷	۳۰/۱±۰/۱۸	۵۵۷	۵۵۷
۹	۳۳/۵±۰/۱۳	۵۹۴	۲۶/۳±۰/۱۴	۵۷۰	۳۱/۸±۰/۱۷	۵۷۰	۳۱/۸±۰/۱۷	۵۷۰	۵۷۰
۱۰	۲۹/۹±۰/۱۱	۵۷۱	۲۳/۳±۰/۱۱	۵۵۵	۲۷/۹±۰/۱۴	۵۵۵	۲۷/۹±۰/۱۴	۵۵۵	۵۵۵
نسل ۳،۲،۱	۵۸±۰/۶۸	۱۸۰۵	۴۸/۶±۰/۷۵	۱۸۰۲	۵۲/۴±۰/۸	۱۸۰۲	۵۲/۴±۰/۸	۱۸۰۲	۱۸۰۲
نسل ۶،۵،۴	۵۴/۶±۰/۳۵	۲۲۶۰	۴۵/۶±۰/۳۷	۲۲۹۰	۴۹/۹±۰/۴۴	۲۲۹۰	۴۹/۹±۰/۴۴	۲۲۹۰	۲۲۹۰
نسل ۱۰،۹،۸،۷	۴۵/۶±۰/۱۹	۲۳۰۰	۳۶/۵±۰/۲۱	۲۲۴۵	۴۱/۹±۰/۲۳	۲۲۴۵	۴۱/۹±۰/۲۳	۲۲۴۵	۲۲۴۵

۱. انتخاب جمعیت مرجع بر اساس روش بیشترین و کمترین عملکردها

استفاده از جمعیت مرجع کامل جهت برآورد اثرات نشانگرها فقط نشانگرهای با فراوانی متوسط به بالا که سهم قابل توجهی در واریانس صفت (بزرگ اثر) داشتند را شناسایی نمود، اما در مقابل پویش‌هایی که از جمعیت مرجع انتخابی استفاده کردند توانستند نشانگرهای با فراوانی کم، اما اثر بزرگ بر روی صفت را نمایان سازند. نشانگرهای با اثرات کوچک در حالت اخیر شناسایی نشدند. بیشتر نشانگرهای معنی‌دار شناسایی شده در پویش ژنومی با استفاده از مرجع انتخابی مربوط به سطح احتمال ۱۰٪ بودند. تعداد و موقعیت نشانگرهای معنی‌دار در تنوع صفت در هر دو پویش ژنومی مرجع کامل (FDR=۱٪) تراکم نشانگری ۱۰۰k و ۲۰۰k (و مرجع انتخابی (FDR=۱۰٪) و تراکم نشانگری ۱۰۰k و ۲۰۰k) با اختلاف اندک یکسان بودند که بیانگر این مسئله است که مرجع انتخابی زمانی می‌تواند به اندازه مرجع کامل کارآیی داشته باشد که نرخ معنی‌داری کاذب نشانگرها برای تعیین حد آستانه بالا نباشد.

حد آستانه^{۲۰} p-value برای تعیین و گزارش نشانگرهای معنی‌دار با استفاده از روش کنترل نرخ معنی‌داری کاذب نشانگرها^{۲۱} (Benjamini and Yosef Hochberg، ۱۹۹۵) مشخص و منفی لگاریتم مقدار پی (-log(p-value)) محاسبه شده برای اثرات کلیه نشانگرها به عنوان تابعی از جایگاه کروموزومی نشانگرها در قالب نمودار منهن به تصویر کشیده شده است. این نمودار برای تراکم‌های مختلف و در سناریوهای مختلف انتخاب جمعیت مرجع در شکل (۳) نشان داده شده است. خطوط آبی، سبز و قرمز، در نمودار به ترتیب حد آستانه در نرخ معنی‌داری کاذب نشانگرها ۱٪، ۵٪ و ۱۰٪ را نشان می‌دهد. تعداد نشانگرهای معنی‌دار و همچنین نشانگرهای معنی‌دار با فراوانی کمتر از ۰/۲ نیز در سناریوها و تراکم نشانگری مختلف در جدول (۴) ارائه شده است. نتایج نشان داد که عملکرد پویش ژنومی در تعیین و گزارش نشانگرهای معنی‌دار مؤثر بر صفت در جمعیت مرجع انتخابی در مقایسه با جمعیت مرجع کامل کاهش می‌یابد. پویش ژنومی با

²⁰ Cut of Point

²¹ False discovery rate (FDR)

همکاران، ۲۰۱۰) و انتخاب ژنومی (Jiménez-Montero و همکاران، ۲۰۱۲) صرف نظر از روش‌های که برای انتخاب حیوانات در ایجاد جمعیت مرجع، استفاده نموده‌اند، همگی بر مؤثر بودن و امکان‌پذیری استفاده از این استراتژی در برنامه‌های اصلاح نژاد دام تأکید دارند. نتایج حاصل از این مطالعه و مقایسه آن با تلاش‌های قبلی در این زمینه، نقش استفاده از داده‌های ژنومی و مدل‌های بهینه را به عنوان ابزاری مناسب برای یافتن مؤثرترین افراد و نشانگرها در کاربرد استراتژی مرجع انتخابی برجسته نموده است.

نتایج حاصل از این پویس ژنومی با نتایج Chao Xing و همکاران (۲۰۰۹)، که گزارش کردند استراتژی مرجع انتخابی برای تعیین نشانگرهای متداول در مطالعات بر پایه ژن مناسبترین استراتژی است، شباهت دارد. همچنین Yunbi و همکاران (۲۰۰۸) پیشنهاد کردند که استراتژی مرجع انتخابی می‌تواند برای نقشه‌یابی مؤثر در مقیاس وسیع مورد استفاده قرار گیرد. بیشتر مطالعاتی که در ارتباط با استفاده از استراتژی مرجع انتخابی انجام شده‌اند از جمله مطالعات انتخاب به کمک نشانگرها (Ansari-Mahyari و همکاران، ۲۰۰۸)، مطالعات پویس ژنومی (Xing و همکاران، ۲۰۰۹)، نقشه‌یابی ژنگاه (Sun و

جدول ۳- صحت^۱ پیش‌بینی ارزش‌های ارثی با استفاده از روش‌های مختلف انتخاب افراد برای جمعیت مرجع در جمعیت‌هایی با اندازه مؤثر جمعیت متفاوت

اندازه مؤثر جمعیت = ۱۰۰			اندازه مؤثر جمعیت = ۴۰۰			ساختار جمعیت
خطای معیار	انحراف معیار	میانگین	خطای معیار	انحراف معیار	میانگین ^۲	روش انتخاب افراد برای جمعیت مرجع
۰/۰۱۳۹	۰/۱۵۸	۰/۴۸۹ ^a	۰/۰۱۳۱	۰/۱۴۸	۰/۵۵۱ ^a	انتخاب همه
۰/۰۱۳۲	۰/۱۶۲	۰/۴۰۹ ^b	۰/۰۱۱۶	۰/۱۲۵	۰/۴۶۲ ^b	حداکثر شباهت ژنگاه
۰/۰۱۴۰	۰/۱۷۲	۰/۳۹۱ ^b	۰/۰۱۲۴	۰/۱۱۶	۰/۴۵۸ ^{bc}	تصادفی
۰/۰۱۲۱	۰/۱۴۸	۰/۴۰۹ ^b	۰/۰۱۴۲	۰/۱۴۲	۰/۴۴۶ ^{bc}	حداقل شباهت نشانگری
۰/۰۱۳۶	۰/۱۴۱	۰/۴۱۹ ^b	۰/۰۱۱۹	۰/۱۲۹	۰/۴۳۵ ^{cd}	بیشترین و کمترین عملکردها
۰/۰۱۲۹	۰/۱۵۸	۰/۴۱۲ ^b	۰/۰۱۳۹	۰/۱۳۹	۰/۴۳۱ ^d	حداقل شباهت ژنگاه
۰/۰۱۳۱	۰/۱۶۱	۰/۴۰۳ ^b	۰/۰۱۲۵	۰/۱۲۳	۰/۴۲۴ ^d	حداکثر شباهت نشانگری

۱. میانگین صحت پیش‌بینی ارزش‌های ارثی ۱۰ نسل با استفاده از روش‌های مختلف انتخاب جمعیت مرجع ارائه شده است.

۲. حروف متفاوت (d, c, b, a) در بالای میانگین بیانگر اختلاف آماری معنی‌دار بین میانگین‌ها در سطح ۱٪ است ($P \leq 0.01$)

جدول ۴ - تعداد نشانگرهای معنی‌دار بعد از تصحیح بر اساس نرخ معنی‌داری کاذب نشانگرها

مرجع کامل			مرجع انتخابی ^۱			عنوان
٪۱۰	٪۵	٪۱	٪۱۰	٪۵	٪۱	تراکم نشانگر / سطح معنی‌داری
۱	۰	۰	۰	۰	۰	تعداد کل نشانگرها
۰	۰	۰	۰	۰	۰	نشانگرها با فراوانی کمتر از ۰/۲
۷۴	۵۶	۲۷	۲۵	۵	۰	تعداد کل نشانگرها
۹	۷	۵	۲	۰	۰	نشانگرها با فراوانی کمتر از ۰/۲
۷۸	۵۷	۲۷	۲۶	۵	۱	تعداد کل نشانگرها
۹	۸	۵	۳	۱	۰	نشانگرها با فراوانی کمتر از ۰/۲
۱۶۸	۱۱۹	۶۷	۱۱۹	۸۹	۴۳	تعداد کل نشانگرها
۴	۰	۰	۶	۰	۰	نشانگرها با فراوانی کمتر از ۰/۲

۱. انتخاب جمعیت مرجع بر اساس بیشترین و کمترین عملکردها

نتیجه‌گیری

استفاده از یک روش مناسب جهت انتخاب افراد برای ایجاد جمعیت مرجع عامل مهم و تأثیرگذار بر کارایی و اثربخشی انتخاب ژنومی است. در این مطالعه هفت روش برای انتخاب زیرمجموعه‌ای از جمعیت به عنوان جمعیت مرجع به منظور برآورد اثرات نشانگرها و پیش‌بینی ارزش‌های ارثی در مطالعات پویش ژنومی و انتخاب ژنومیک مورد بررسی قرار گرفت. روش انتخاب همه حیوانات به عنوان جمعیت مرجع در هر دو حالت، $r^2=0/255$ و $r^2=0/086$ در مقایسه با روش‌های مختلف مرجع انتخابی ارزش‌های ارثی را با صحت بالاتری پیش‌بینی نمود. اما مقایسه روش‌های مختلف انتخاب در سناریوی‌های مرجع انتخابی نشان داد که بین روش‌های انتخاب، زمانی که میزان LD بالا است اختلافی وجود ندارد ($r^2=0/255$)، اما زمانی که LD کاهش یافت ($r^2=0/086$) روش‌های انتخاب جمعیت مرجع با صحت‌های متفاوتی ارزش‌های اصلاحی ژنومی را پیش‌بینی کردند. زمانی که LD پایین است ($r^2=0/086$) انتخاب دام‌ها بر اساس حداکثر روابط خویشاوندی ژنگاهی و حداکثر روابط خویشاوندی نشانگری، به ترتیب بیشترین و کمترین صحت پیش‌بینی ارزش اصلاحی ژنومی را بدنبال داشت. با توجه به عدم تغییر رتبه بندی حیوانات در سناریوی مرجع انتخابی (بیشترین و کمترین عملکردها) و تراکم بالا (100k) در مقایسه با تراکم پایین (50k) و مرجع کامل، استفاده ترکیبی از انتخاب افراد با بیشترین و کمترین عملکرد و تراکم نشانگری بالاتر می‌تواند به عنوان استراتژی تشکیل جمعیت مرجع در نظر گرفته شود. اگر چه صحت پیش‌بینی ارزش‌های ارثی ژنومی در سناریوی مرجع انتخابی در مقایسه با مرجع کامل کاهش یافت ولی با توجه به اختلاف نسبتاً کم صحت ارزیابی‌ها در دو سناریو و در صورت بررسی هزینه و فایده هر واحد کاهش و یا افزایش در صحت، احتمالاً استراتژی مرجع انتخابی پتانسیل کاهش هزینه‌های تشکیل جمعیت مرجع را دارد. البته سودمندی اقتصادی اتخاذ این استراتژی‌ها برای تشکیل جمعیت مرجع خصوصاً در نشخوارکنندگان کوچک بایستی در مطالعه دیگری بررسی گردد.

منابع

- Ansari-Mahyari, S., Sørensen, A. C., Lund, M. S., Thomsen, H., & Berg, P. (2008). Across-Family Marker-Assisted Selection Using Selective Genotyping Strategies in Dairy Cattle Breeding Schemes. *Journal of Dairy Science*, 91(4), 1628–1639. <https://doi.org/10.3168/jds.2007-0613>
- Blonk, R.J.W., Komen, J. and van Arendonk, J.A.M., (2010). Minimizing genotyping in breeding programs with natural mating. World Congress on Genetic Applied to Livestock Production, Leipzig, Germany, 195, 2–7.
- Boligon, A. A., Long, N., Albuquerque, L. G., Weigel, K. A., Gianola, D., & Osa, G. J. M. (2012). Comparison of selective genotyping strategies for prediction of breeding values in a population undergoing selection. *Journal of Animal Science*, 90(13), 4716–4722. <https://doi.org/10.2527/jas.2012-4857>
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2), 433–443. <https://doi.org/10.3168/jds.2008-1646>
- Hill, W. G., & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, 38(6), 226–231. <https://doi.org/10.1007/BF01245622>
- Hu, Z. L., Park, C. A., Wu, X. L., & Reecy, J. M. (2013). Animal QTLdb: An improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Research*, 41(D1), 871–879. <https://doi.org/10.1093/nar/gks1150>
- In: <https://www.animalgenome.org/cgi-bin/QTLdb/OA/index>
- Jannink, J. L. (2005). Selective phenotyping to accurately map quantitative trait loci. *Crop Science*, 45(3), 901–908. <https://doi.org/10.2135/cropsci2004.0278>
- Jiménez-Montero, J. A., González-Recio, O., & Alenda, R. (2012). Genotyping strategies for genomic selection in small dairy cattle populations. *Animal*, 6(8), 1216–1224. <https://doi.org/10.1017/S1751731112000341>

- Jin, C., Lan, H., Attie, A. D., Churchill, G. A., Bulutuglo, D., & Yandell, B. S. (2004). Selective phenotyping for increased efficiency in genetic mapping studies. *Genetics*, 168(4), 2285–2293.
<https://doi.org/10.1534/genetics.104.027524>
- Loberg, A. and Durr, J.W. (2009). Interbull survey on the use of genomic information. In: Proceedings of the Interbull technical workshop, Uppsala, Sweden, 39, 3–1
- Lund, M. S., de Roos, A. P. W., de Vries, A. G., Druet, T., Ducrocq, V., Guillaume, F., Liu, Z., Schrooten, C., Su, G. (2010). Improving genomic prediction by EuroGenomics collaboration. In: 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany (p. 150).
<https://doi.org/10.1007/s10530-015-1047-4>
- Nishio, M., & Satoh, M. (2014). Impacts of genotyping strategies on long-term genetic response in genomic selection. *Animal Science Journal*, 85(5), 511–516.
<https://doi.org/10.1111/asj.12184>
- Rupp, R., Mucha, S., Larroque, H., McEwan, J., & Conington, J. (2016). Genomic application in sheep and goat breeding. *Animal Frontiers*, 6(1), 39. <https://doi.org/10.2527/af.2016-0006>
- Safari, A., & Fogarty, N. M. (2003). Genetic parameters for sheep production traits: estimates from the literature. *Technical Bulletin, NSW Agriculture, Australia*, 49, 13–21. Retrieved from http://www.sheepcrc.org.au/files/pages/articles/publications--genetics/Genetic_Parameters_entire_report.pdf
- Sargolzaei, M. (2014). SNP1101 User's Guide. Version 1.0. HiggsGene Solutions Inc.
- Sargolzaei, M., & Schenkel, F. S. (2009). QMSim: A large-scale genome simulator for livestock. *Bioinformatics*, 25(5), 680–681.
<https://doi.org/10.1093/bioinformatics/btp045>
- Sargolzaei, M., Schaeffer, L. R., Wiggans, G. R., & Schenkel, F. S. (2014). Approximation of Reliability of Direct Genomic Breeding. In: 10th World Congress on Genetics Applied to Livestock Production, Vancouver, Canada, August 17-22.
- Su, G., Ma, P., Nielsen, U. S., Aamand, G. P., Wiggans, G., Guldbbrandtsen, B., & Lund, M. S. (2016). Sharing reference data and including cows in the reference population improve genomic predictions in Danish Jersey. *Animal*, 10(6), 1067–1075.
<https://doi.org/10.1017/S1751731115001792>
- Sun, Y., Wang, J., Crouch, J. H., & Xu, Y. (2010). Efficiency of selective genotyping for genetic analysis of complex traits and potential applications in crop improvement. *Molecular Breeding*, 26(3), 493–511.
<https://doi.org/10.1007/s11032-010-9390-8>
- VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited Review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92(1), 16–24.
<https://doi.org/10.3168/jds.2008-1514>
- Weigel, K. A., Van Tassell, C. P., O'Connell, J. R., VanRaden, P. M., & Wiggans, G. R. (2010). Prediction of unobserved single nucleotide polymorphism genotypes of Jersey cattle using reference panels and population-based imputation algorithms. *Journal of Dairy Science*, 93(5), 2229–2238.
<https://doi.org/10.3168/jds.2009-2849>
- Wiggans, G.R., Sonstegard, T.S., VanRaden, P.M., Matukumalli, L.K., Schnabel, R.D., Taylor, J.F., Chesnais, J.P., Schenkel, F. and Van, Tassell, C.P. (2008). Genomic evaluations in the United States and Canada: collaboration. In: Proceedings of International Committee of Animal Recording, Niagara Falls, NY, 6pp
- Xing, C. and Xing, G. (2009). Power of selective genotyping in genome-wide association studies of quantitative traits. *BMC Proceedings*, 3(Suppl 7): S23. <https://doi:10.1186/1753-6561-3-S7-S23>
- Yoav, B. and Yosef, H. (1995). Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*. 57(1), 289-300.
- Zhao, Y., Gowda, M., Longin, F. H., Würschum, T., Ranc, N., & Reif, J. C. (2012). Impact of selective genotyping in the training population on accuracy and bias of genomic selection. *Theoretical and Applied Genetics*, 125(4), 707–713.