

تأثیر نرخ شیوع بیماری در جمعیت مرجع بر عملکرد روش های جنگل تصادفی و بیز آستانه‌ای A

• یوسف نادری (نویسنده مسئول)

استادیار، گروه علوم دامی، باشگاه پژوهشگران جوان و نخبگان، دانشگاه آزاد اسلامی، واحد آستارا، آستارا، ایران

تاریخ دریافت: مرداد ۱۳۹۷ تاریخ پذیرش: آذر ۱۳۹۷

شماره تماس نویسنده مسئول: ۰۹۱۲۲۴۸۱۳۸۱

Email: y.naderi@iau-astara.ac.ir

شناسه دیجیتال (DOI): 10.22092/asj.2018.123060.1754

چکیده

هدف از تحقیق حاضر، بررسی نقش نرخ شیوع بیماری در جمعیت مرجع و معماری‌های مختلف ژنومی بر عملکرد روش جنگل تصادفی و بیز A در صفات آستانه‌ای بود. بدین منظور، جمعیت‌های ژنومی برای سطوح متفاوت وراثت‌پذیری (۰/۰۵ و ۰/۲۵)، سطوح مختلف LD (پایین و بالا) و تعداد متفاوت جایگاه‌های صفات کمی (۱۵۰ و ۶۰۰) بر روی ۳۰ کروموزم شبیه‌سازی شدند. جهت ایجاد فنوتیپ دودویی با نسبت‌های مختلف شیوع بیماری، ابتدا برای ۵ درصد حیوانات جمعیت مرجع که کم‌ترین میانگین فنوتیپی را داشتند کد یک (یا بیمار) و ۹۵ درصد دیگر کد صفر (یا سالم) در نظر گرفته شد. این روند با افزایش یک نرخ ۵ درصدی تا زمانی که ۵۰ درصد افراد جمعیت مرجع کد یک داشتند ادامه یافت. صحت ژنومی در هر دو روش جنگل تصادفی و بیز A، با افزایش نرخ شیوع بیماری از ۵ به ۲۰ درصد، افزایش و پس از آن تا رسیدن به ۵۰ درصد، کاهش یافت. اثر منفی سطوح بالای شیوع بیماری بر صحت ژنومی بیش‌تر از سطوح پایین آن بود. در مجموع روش جنگل تصادفی نسبت به بیز A نوسانات بیشتری به تغییرات معماری ژنومی و نرخ شیوع بیماری داشت. با وجود صحت ژنومی بالاتر بیز آستانه‌ای A در جمعیت‌های مختلف، هنگامی که صفات با وراثت‌پذیری بالا توسط تعداد زیادی QTL کنترل شدند، روش جنگل تصادفی عملکرد بهتری داشت. علی‌رغم نقش مهم ساختار ژنتیکی جمعیت مورد آنالیز، بهترین روش پیش‌بینی ارزش‌های اصلاحی ژنومی صفات آستانه‌ای به نرخ شیوع بیماری در جمعیت مرجع وابسته بود.

واژه‌های کلیدی: ارزش اصلاحی، صحت ژنومی، عدم تعادل پیوستگی، وراثت‌پذیری

Animal Science Journal (Pajouhesh & Sazandegi) No 124 pp: 131-146

The impact of disease prevalence rate in training set on performance of random forest and threshold Bayes A methods

By: Yousef Naderi

Assistant Professor, Department of Animal Science, Young Researchers and Elite Club, Astara Branch, Islamic Azad University, Astara, Iran.

Corresponding author. E-mail address: y.naderi@iau-astara.ac.ir

Received: August 2018**Accepted: December 2018**

The objective of current study was to investigate the role of disease prevalence rate of training set and genomic architecture on performance of random forest (RF) and threshold Bayes A (BTA) in threshold traits. For this purpose, genomic population were simulated to reflect variations in heritability (0.05 and 0.25), number of QTL (150 and 600) and linkage disequilibrium (low and high) for 30 chromosomes. To create binary phenotype with different disease prevalence rate, at first, 5 percent of training set animals which had the lowest phenotype average defined code 1 (or diseased) and 95 percent of others defined code 0 (or healthy). This process continued with a 5% increase rate until 50 percent of animals had code 1 in training set. In both random forest and Bayes A methods, genomic accuracy with increase in disease prevalence rate 5 to 20 percent was increased, and afterwards to achieve of 50 percent was decreased. The negative effect of high levels of disease prevalence rate on genomic accuracy was higher than low levels of it. Overall, RF was fluctuation to variations of genetic architecture and disease prevalence rate. Despite the higher accuracy of TBA at different scenarios, RF showed a better performance when high-heritability traits were controlled by a large number of QTLs. Despite the important role of genetic basis of the population analyzed, the best method to predict genomic breeding value of threshold traits depend on disease prevalence rate.

Key words: breeding value, genomic accuracy, linkage disequilibrium, heritability.**مقدمه**

اصلاحی شده است (Piccoli و همکاران، ۲۰۱۴). در این راستا انتخاب ژنومی با استفاده از تعداد زیادی از نشانگرهای ژنتیکی در سراسر ژنوم، بر تعدادی از محدودیت‌های انتخاب کلاسیک غلبه کرد (Meuwissen و همکاران، ۲۰۰۱) و اجازه انتخاب حیوانات جوان و کاهش فاصله نسلی را در پی داشت (Hayes و همکاران، ۲۰۰۹).

تحقیقات زیادی موفقیت انتخاب ژنومی را در راستای بهبود پیشرفت ژنتیکی در گاو شیری گزارش کردند که اکثر این تحقیقات توجه خود را بر صفات پیوسته معطوف نموده‌اند

بهبود ژنتیکی از طریق گزینش مصنوعی در طول سال‌های اخیر نقش بسیار مهمی در افزایش بهره‌وری حیوانات اهلی داشته است (Hayes, ۲۰۰۷). روش‌های سنتی بر آورد ارزش‌های اصلاحی با استفاده از اطلاعات شجره‌ای گرچه منجر به افزایش بهبود ژنتیکی در سالیان اخیر شده است، اما نیازمند یک شرایط ایده‌آلی است که مؤثر بودن انتخاب فنوتیپی را محدود می‌کند (Dekkers, ۲۰۱۰). پیشرفت‌های اخیر در تکنولوژی DNA، منجر به توالی ژنوم چندین گونه از جمله گاو شیری و توسعه تکنولوژی‌های جدید ژنوم به منظور ارزیابی دقیق ارزش‌های

(۲۰۱۱) جهت برآورد صحت پیش‌بینی ژنومی استفاده شد. علاوه بر تأثیر مدل‌های آماری و نوع صفت مورد مطالعه عوامل مختلفی می‌توانند صحت ارزش‌های اصلاحی ژنومی و ارزیابی ژنومی را تحت تأثیر قرار دهد، این عوامل شامل تعداد QTL (González-Recio و Forni، ۲۰۱۱)، توزیع اثرات QTL (Ghafouri-Kesbi و همکاران، ۲۰۱۷)، مقدار عدم تعادل پیوستگی (Yin و همکاران، ۲۰۱۴)، تراکم نشانگرها (Badke و همکاران، ۲۰۱۴)، وراثت‌پذیری، تعداد داده‌های فنوتیپی در جمعیت مرجع (Meuwissen و همکاران، ۲۰۰۱) و فاصله زمانی (تعداد نسل) بین جمعیت مرجع و جمعیت تأیید (GorganiFirozjah و همکاران، ۲۰۱۴) می‌باشند. با توجه به این که در انتخاب ژنومی ارزش‌های اصلاحی حیوانات جمعیت کاندیدا (دارای ژنوتیپ) از طریق برآورد میزان اثر هر کدام از نشانگرها بر صفت در جمعیت مرجع (دارای ژنوتیپ و فنوتیپ) برآورد می‌شود، با این حال برای صفات آستانه‌ای علاوه بر عوامل فوق‌الذکر، نسبت فنوتیپی جمعیت مرجع (نسبت شیوع بیماری) یکی از عوامل تأثیرگذار در برآورد ارزش‌های اصلاحی حیوانات جمعیت کاندیدا است (Pimentel و همکاران، ۲۰۱۳). تحقیقات در این زمینه نشان داد که نسبت شیوع بیماری در جمعیت مرجع از عوامل مؤثر بر صحت ژنومی می‌باشد (Naderi و همکاران، ۲۰۱۶). همچنین در تحقیقات دیگر افزایش نسبت فنوتیپی حیوانات ماده نسبت به نرها در جمعیت مرجع از عوامل مؤثر بر صحت ژنومی عنوان شد (Buch و همکاران، ۲۰۱۲). امروزه تلاش پژوهشگران بر آن است تا با شناسایی عوامل مؤثر بر صحت برآوردهای ارزش‌های اصلاحی ژنومی بهترین بهره‌برداری از داده‌های موجود انجام شود. از آنجا که توالی یابی کل ژنوم بر تعداد بسیار زیادی از افراد بسیار سخت و هزینه‌بر است و از سوی دیگر نیازمند به در اختیار داشتن ارزش ژنتیکی افراد است، در

(Hayes و همکاران، ۲۰۰۹؛ Wiggans و همکاران، ۲۰۱۱). با این حال از دیدگاه اصلاح‌گران و همچنین پرورش‌دهندگان در برنامه‌های مدرن اصلاح نژادی، پیشرفت‌های قابل توجه در بهبود گاو شیری و افزایش سود اقتصادی، مستلزم ورود مستقیم به صفات رفتاری، سلامت حیوان و مقاومت به بیماری‌ها است که دارای ماهیتی آستانه‌ای هستند (Egger-Danner و همکاران، ۲۰۱۵). در نتیجه به خاطر عدم رابطه خطی بین متغیرها و صفت وابسته، عدم توزیع نرمال متغیر وابسته و توزیع متفاوت واریانس این صفات بدیهی است که روش‌های انتخاب ژنومی برای صفات پیوسته نمی‌تواند راه حل مناسبی برای دستیابی به پیشرفت ژنتیکی در این گونه صفات باشد. بنابراین، استفاده بهتر از انتخاب ژنومی و کاربرد آن در اصلاح نژاد دام هم‌گام با توسعه روش‌های مولکولی در این صفات، مستلزم به کارگیری مدل‌های آماری آستانه‌ای به منظور ارزیابی دقیق‌تر این صفات است (Wang و همکاران، ۲۰۱۷).

روش‌های آماری مختلفی برای تخمین اثرات نشانگرها ارائه شده است. تفاوت عمده این روش‌ها علاوه بر معماری ژنتیکی صفت مورد مطالعه، به فرضیات در نظر گرفته شده برای مدل ژنتیکی پشت‌صحنه آن‌ها وابسته است (Chen و همکاران، ۲۰۱۴). روش‌های بیزی و یادگیری ماشین از جمله روش‌های آماری برای غلبه بر مشکل $p < n$ در انتخاب ژنومی هستند. استفاده از روش بیز A در انتخاب ژنومی اولین بار توسط Meuwissen و همکاران (۲۰۰۱) ارائه و بعدها توسط González-Recio و Forni (۲۰۱۱) مدل آستانه‌ای آن گسترش یافت. از دیگر روش‌های مورد استفاده در انتخاب ژنومی، روش جنگل تصادفی است که یکی از الگوریتم‌های یادگیری ماشینی می‌باشد که اولین بار توسط Breiman (۲۰۰۱) پیشنهاد شد و بعدها از آن برای آنالیز ژنومی صفات آستانه‌ای (Forni و González-Recio،

دیگر جمعیت تکثیر شد (نسل ۱۲۱۰). در طراحی جمعیت نهایی، افراد آخرین نسل (نسل ۱۲۱۰) به عنوان جمعیت تایید (۲۰۰۰ راس) که این افراد اطلاعات ژنوتیپی داشته، اما فاقد اطلاعات فنوتیپی بودند. هم چنین افراد ۵ نسل ما قبل جمعیت تایید (نسل ۱۲۰۵ تا ۱۲۰۹) در گروه جمعیت های مرجع (۱۰۰۰۰ راس) که این افراد هم اطلاعات ژنوتیپی داشته و هم ارزش های اصلاحی ژنومی آنها مشخص می باشد طبقه بندی شدند. شانس تلاقی در همه ی حیوانات برابر (در هر دو جنس) و یک فرزند برای هر زایش در نظر گرفته شد. درصد جایگزینی برای نر و ماده به ترتیب ۵۰ و ۲۰ درصد در نظر گرفته شد. انتخاب حیوانات برتر برای نسل بعد براساس ارزش اصلاحی بالا صورت گرفت. دو سطح مختلف QTL (۱۵۰ و ۶۰۰) شبیه سازی شد که به صورت تصادفی در طول کروموزم ها توزیع شدند. نرخ جهش برای نشانگرها و QTL ها در هر جایگاه و در هر نسل $2/5 \times 10^{-5}$ فرض شد (Solberg و همکاران، ۲۰۱۸).

با توجه به این که از دیدگاه آماری توزیع احتمال QTL های صفات مهم اقتصادی توسط شمار اندکی ژن ها دارای اثر عمده و درصد بالایی از ژن ها کوچک اثر هستند و این فرضیه به توزیع گاما نزدیک تر است (Goddard و Hayes، ۲۰۰۱). در نتیجه توزیع احتمال QTL ها، گاما فرض شد. هم چنین فراوانی آللی اولیه برای نشانگرها ۰/۵ در نظر گرفته شد. در هر نسل و هر جایگاه کل میزان واریانس افزایشی توسط QTL توجیه شد. دو سطح مختلف وراثت پذیری (۰/۰۵ و ۰/۲۵) برای هر جمعیت شبیه سازی شده در نظر گرفته شد. خلاصه جمعیت شبیه سازی شده و پارامترهای به کار رفته در جدول ۱ نشان داده شده است.

نتیجه مطالعات شبیه سازی ابزاری ارزشمند برای ارزیابی و سنجش اعتبار روش های پیشنهادی در انتخاب ژنومی با هزینه ای خیلی کم می باشد که امکان پیش بینی تغییرات پارامترهای ژنتیکی در آینده را نیز فراهم می آورد. لذا پژوهش حاضر با هدف بررسی نقش نسبت های فنوتیپی جمعیت مرجع (نرخ شیوع بیماری) و معماری های مختلف ژنوم (تعداد متفاوت QTL، سطوح متفاوت وراثت پذیری و LD) بر صحت ژنومی صفات آستانه ای دودویی با استفاده از روش های یادگیری ماشینی (جنگل تصادفی و بوستینگ) و بیز آستانه ای A در داده های شبیه سازی شده انجام شده است.

مواد و روش ها

شبیه سازی جمعیت

با استفاده از نرم افزار QMSim (Sargolzaei و Schenkel، ۲۰۰۹) ژنومی متشکل از ۳۰ کروموزوم هر یک به طول ۱ مورگان شبیه سازی شدند. به ازای هر کروموزوم ۱۸۰۰ نشانگر شبیه سازی شد. در نتیجه ۵۴۰۰۰ نشانگر برای تراشه های ۵۴K شبیه سازی شدند. برای تولید جمعیتی با LD پایین، یک جمعیت پایه ۲۰۰۰ راسی طی ۱۲۰۰ نسل شبیه سازی شد. برای تولید جمعیتی با LD بالا، یک جمعیت پایه با ۲۰۰۰ راسی طی ۱۰۰۰ نسل شبیه سازی شد. سپس تعداد افراد جمعیت از طریق ایجاد یک گلوگاه ژنتیکی^۱ به ۲۰۰ راس در نسل ۱۱۰۰ کاهش یافت. در گام بعدی، افراد این جمعیت برای ۹۰ نسل دیگر تلاقی یافتند (تا نسل ۱۲۰۰) تا به تعداد اولیه خود (۲۰۰۰ راس) برسند.

برای ایجاد جمعیت مرجع و تایید، همه افراد (۲۰۰۰ راس) آخرین نسل جمعیت پایه برای تولید مثل در جمعیت حاضر مورد استفاده قرار گرفتند که در این بین ۵۰ راس نر و اندازه مؤثر جمعیت ۱۹۵ در نظر گرفته شد. نوع سیستم تلاقی تصادفی بود و برای ۱۰ نسل

¹Bottleneck

جدول ۱: پارامترهای فرآیند شبیه سازی

ساختار جمعیت	عدم تعادل پیوستگی بالا	عدم تعادل پیوستگی پایین
جمعیت اولیه		
فاز اول تعداد نسل (تعداد افراد)	۱۰۰۰(۲۰۰۰)	۱۰۰۰(۲۰۰۰)
گلوگاه	بله	خیر
فاز دوم تعداد نسل (تعداد افراد)	۱۱۰۰(۲۰۰)	۱۱۰۰(۲۰۰۰)
فاز سوم تعداد نسل (تعداد افراد)	۱۲۰۰(۲۰۰۰)	۱۲۰۰(۲۰۰۰)
تعداد حیوانات در نسل آخر	۲۰۰۰	
جمعیت اخیر		
تعداد نرهای در نسل اخیر	۵۰	
تعداد ماده‌ها در نسل اخیر	۱۹۵۰	
تعداد تکثیری جمعیت اخیر بعد از نسل ۱۲۰۰	۱۰	
جمعیت مرجع		افراد نسل ۱۲۰۵ تا ۱۲۰۹ (۱۰۰۰۰ فرد)
جمعیت تأیید		افراد نسل ۱۲۱۰ (۲۰۰۰ فرد)
تعداد نتایج به ازای هر زایش	۱	
احتمال نر بودن نتاج	۰/۵	
انتخاب و طرح آمیزش	تصادفی	
نرخ جایگزینی برای نرها	٪۵۰	
نرخ جایگزینی برای ماده‌ها	٪۲۰	
معیار حذف	سن بالا/ ارزش اصلاحی برآوردی	
ژنوم		
تعداد کروموزوم	۳۰	
طول هر کروموزوم (مورگان)	۱	
تعداد QTL به ازای هر کروموزوم	۵ یا ۲۰	
اثر آلل‌های QTL	گاما (۰/۴)	
تعداد نشانگر به ازای هر کروموزوم	۱۸۰۰	
نرخ جهش در نشانگر و QTLها	$2/5 \times 10^{-5}$	
وراثت‌پذیری	۰/۲۵ و ۰/۰۵	

عدم تعادل پیوستگی

نشانگرهای مختلف در ژنوم همه حیوانات موجود در آخرین نسل مورد استفاده قرار گرفت.

سطح LD برای جمعیت‌های مختلف شبیه سازی شده با استفاده از محاسبه‌ی توان دوم ضریب همبستگی (r^2) بین همه‌ی جفت نشانگرهای مورد ارزیابی قرار گرفت. نرم افزار PLINK (Purcell و همکاران، ۲۰۰۷) برای برآورد LD بین جفت

ایجاد نسبت های مختلف فنوتیپ دودویی

در مجموع ۴ جمعیت (جمعیت اول: وراثت پذیری صفت ۰/۲۵ - تعداد QTL ۶۰۰ - سطح پایین LD؛ جمعیت دوم: وراثت پذیری صفت ۰/۲۵ - تعداد QTL ۱۵۰ - سطح پایین LD؛ جمعیت سوم: وراثت پذیری صفت ۰/۰۵ - تعداد QTL ۱۵۰ - سطح پایین LD؛ جمعیت چهارم: وراثت پذیری صفت ۰/۰۵ - تعداد QTL ۱۵۰ - سطح بالای LD) در تحقیق حاضر شبیه سازی شد. برای ایجاد نسبت های مختلف فنوتیپ آستانه ای دودویی تغییراتی در فایل خروجی QMSim ایجاد شد. بدین صورت که فنوتیپ پیوسته افراد از بیشترین به کمترین مرتب شد و با توجه به شیوع بیماری مورد نظر، فنوتیپ پیوسته به فنوتیپ آستانه ای تبدیل شد. به طوری که در گروه اول: ابتدا ۵ درصد حیوانات جمعیت مرجع که کمترین میانگین فنوتیپی را داشتند کد یک (یا بیمار) و ۹۵ درصد حیوانات دیگر کد صفر (یا سالم) در نظر گرفته شدند. در گروه دوم: ۱۰ درصد افراد جمعیت مرجع را افراد بیمار و ۹۰ درصد دیگر را افراد سالم تشکیل دادند. این روند در گروه های دیگر با افزایش ۵ درصدی افراد بیمار تا زمانی ادامه یافت که ۵۰ درصد حیوانات جمعیت مرجع بیمار (کد صفر) و ۵۰ درصد حیوانات دیگر سالم (کد یک) ادامه پیدا کرد. در نتیجه ۱۰ گروه متفاوت نسبت فنوتیپی برای جمعیت مرجع در نظر گرفته شد. هم چنین نشانگرهای با فراوانی آللی کمیاب (MAF) کمتر از ۰/۰۳ حذف شدند. برای ارزیابی مدل ۱۰ تکرار از هر جمعیت در نظر گرفته شد.

روش آماری

در این تحقیق از یک روش بیزی (بیز آستانه ای A) و یک روش یادگیری ماشینی (جنگل تصادفی) برای ارزیابی صحت ژنومی جمعیت های مورد مطالعه استفاده شد. مدل کلی بیز آستانه ای A به صورت زیر می باشد.

$$\lambda = \mu 1 + Xb + e$$

در این جا λ بردار ستونی با ابعاد $n \times 1$ برای متغیر y می باشد. μ میانگین جمعیت، 1 بردار ستونی با ابعاد $n \times 1$ با درایه های یک

می باشد. b برداری برای برآوردهای ضرایب رگرسیونی اثرات p نشانگر با فرض توزیع نرمال $(N(0, \sigma_j^2))$ و واریانس متفاوت و مستقل برای هر نشانگر می باشد. σ_j^2 به صورت واریانس ناشناخته در ارتباط با نشانگرها مفروض شده است و دارای توزیع کای مربع مقیاس دار معکوس $(\nu_j \chi_{\nu_j}^{-2} s_j^2)$ با $\nu_j = 4$ و $s_j^2 = 0.002$ بود. عناصر ماتریس X به ابعاد $n \times p$ در برگرفته اثرات افزایشی می باشد. در بیز آستانه ای A باقی مانده ها (e) با فرض میانگین صفر و واریانس یک در نظر گرفته شدند. این روش از طریق نمونه گیری گیبس انجام گرفت. بیز آستانه ای A از طریق بسته ی BGLR در نرم افزار R مورد آنالیز قرار گرفت.

جنگل تصادفی یک روش غیر پارامتری باز نمونه گیری است که خلاف روش های بیزی نیازمند فرضیات کم تری درباره توزیع داده ها دارد (Goldstein و همکاران، ۲۰۱۰). این الگوریتم درختان تصمیم زیادی در نمونه های بوت استرپ^۲ ایجاد می کند و از طریق میانگین هر درخت، پیش بینی های نهایی را محاسبه می کند. به طور کلی جنگل تصادفی از طریق استراتژی بگینگ^۳ (Breiman، ۲۰۰۱) خطای پیش بینی را کاهش و با استفاده از انتخاب ویژگی^۴، گزینش متغیر تصادفی جهت ایجاد هر درخت ایجاد می کند. پارامترهای کلیدی برای مدل جنگل تصادفی، تعداد درختان و تعداد متغیرهای پیشگو می باشند (Breiman، ۲۰۰۱). سه پارامتر مهمی که در جنگل تصادفی در مورد کلاسه بندی بایستی تنظیم شود عبارت اند از mtry، تعداد SNP یا کواریت هایی نمونه برداری شده در هر بار نمونه گیری تصادفی، ntree یا تعداد بوت استرپ و یا تعداد درختانی که بایستی رشد کنند و معیاری برای انتخاب بهترین SNP برای تقسیم شدن هر گره است. node (تعداد گره) یا وزن دهی است که نشان دهنده ی تعداد مشاهدات در هر خوشه درخت می باشد. مدل کلی جنگل تصادفی به صورت زیر است.

$$y = \mu + \sum_{t=1}^T c_t h_t(y; X)$$

در این جا برای هر مشاهده y از طریق میانگین پیش بینی های هر

²Bootstrap

³Bagging

⁴Feature Selection

تصادفی نشان داد. بیشترین میزان صحت ژنومی برای شیوع بیماری ۲۰ درصد در جمعیت اول با استفاده از جنگل تصادفی و کمترین میزان صحت ژنومی برای شیوع بیماری ۵۰ درصد در جمعیت سوم با استفاده از جنگل تصادفی بود. به طور کلی در تمامی جمعیت‌ها کمترین میزان صحت ژنومی برای درصد‌های پایین (۵ درصد) و بالای (۴۵ و ۵۰) شیوع بیماری مشاهده شد. با این حال نتایج نشان داد که اثر منفی درصد بالای شیوع بیماری بر صحت ژنومی بیشتر از درصد پایین شیوع بیماری است.

Forni و González-Recio (۲۰۱۱) یک جمعیت ژنومی با ۲۵۰۰ حیوان برای یک صفت گسسته دودویی با شیوع بیماری ۵۰ درصد در جمعیت مرجع را شبیه‌سازی کرد. نتایج آن‌ها نشان داد که صحت ژنومی برای جنگل تصادفی و بیز آستانه‌ای A به ترتیب ۰/۳۶ و ۰/۲۶ بود. Ogutu و همکاران (۲۰۱۱) یک جمعیت ژنومی با نرخ شیوع بیماری ۵۰ درصد برای یک تراشه $10K$ شبیه‌سازی کرد. نتایج آن‌ها نشان داد که صحت ژنومی روش جنگل تصادفی ۰/۴۸۳ بود. همچنین Naderi و همکاران (۲۰۱۶) جمعیت‌های مختلف ژنومی را برای صفات گسسته با نسبت‌های مختلف شیوع بیماری در جمعیت مرجع شبیه‌سازی کردند. آن‌ها گزارش کردند که صحت ژنومی با افزایش نسبت شیوع بیماری در جمعیت مرجع تا ۲۰ درصد افزایش یافته و پس از آن روندی ثابت به خود دارد. هم‌چنین برای تراشه‌های کم تراکم همواره عملکرد GBLUP بیشتر از جنگل تصادفی بود. بانه و همکاران (۱۳۹۶) ارزیابی ژنومی را با استفاده از توزیع پروبیت برای شیوع صفت آستانه‌ای در جمعیت مرجع ۱۰۰۰ راسی با استفاده از روش‌های بیزی انجام دادند و نشان داد که تفاوت معنی‌داری بین صحت روش‌های بیزی مشاهده نشده است. Wang و همکاران (۲۰۱۳) یک جمعیت ژنومی با نرخ شیوع ۳۰ درصد در جمعیت مرجع را شبیه‌سازی کرد و نشان داد که صحت ژنومی برای بیز A حدود ۰/۲۶۳ بود که نسبت به جمعیت مشابه آن در تحقیق حاضر صحت به مراتب کمتری بود که دلیل این امر می‌تواند به خاطر تعداد بسیار پایین افراد جمعیت مرجع عنوان کرد. Melchinger و Technow (۲۰۱۳) یک جمعیت ژنومی برای

درخت محاسبه می‌شود. μ میانگین جمعیت و y بردار $n \times 1$ برای مشاهدات فنوتیپی گسسته، $X = \{x_i\}$ که در این جا x_i بردار $p \times I$ ژنوتیپ هر حیوان برای p نشانگر که T درخت تصمیم ساخته شده است. هر درخت $(h_t(y; X))$ یک نمونه تصادفی با جایگزینی از n نمونه می‌باشد و هر گره یک نمونه کوچکی تصادفی از SNP و C_t فاکتور انقباضی میانگین درختان است. سایر حیواناتی که جز این نمونه‌گیری نیستند به عنوان خارج از مجموعه شناخته شده و در اعتبار سنجی هر درخت گزینش می‌شوند. در این تحقیق ترکیب بهینه و ایده آل پارامترها جنگل تصادفی (از طریق خطای خارج از کیسه) برای تراشه $54k$ به ترتیب $mtry=10000$ ، $ntree=5000$ و $node=5$ بود. داده‌های ژنومی از طریق بسته‌ی RanFoG (González-Recio و Forni، ۲۰۱۱) و نرم افزار R مورد آنالیز قرار گرفتند. همبستگی بین ارزش‌های اصلاحی ژنومی واقعی و برآورد شده‌ی روش‌های بیز A و جنگل تصادفی (در جمعیت تایید) به عنوان معیار ارزیابی صحت مورد استفاده قرار گرفت.

نتایج و بحث

اثر شیوع بیماری بر صحت ژنومی

جدول ۲ میانگین صحت ژنومی روش‌های بیز آستانه‌ای A و جنگل تصادفی را برای نسبت‌های مختلف شیوع بیماری در جمعیت مرجع و هر یک از جمعیت‌های شبیه‌سازی شده را نشان می‌دهد. به طور کلی در همه‌ی جمعیت‌ها با افزایش شیوع بیماری در جمعیت مرجع تا ۲۰ درصد صحت ژنومی روندی صعودی داشت و بعد از آن با افزایش درصد شیوع بیماری تا ۵۰ درصد، صحت ژنومی کاهش یافت. برای جمعیت یک و بیز آستانه‌ای A با افزایش شیوع بیماری تا ۲۰ درصد، صحت ژنومی افزایشی ۴۹/۱۳ درصدی داشت. این میزان افزایش در صحت برای جمعیت یک و جنگل تصادفی ۶۸/۹۲ درصد بود. همچنین در این جمعیت عملکرد جنگل تصادفی نسبت به بیز A هنگامی که ۱۰ تا ۴۵ درصد شیوع بیماری در جمعیت مرجع وجود داشت بهتر بود. برای سایر جمعیت‌ها همواره بیز A عملکرد بهتری نسبت به جنگل

مورد نظر نیست، برخلاف داده های واقعی، در شبیه سازی فنوتیپ آستانه ای صفت از طریق فنوتیپ پیوسته حیوان برآورد و تخمین زده می شود. بنابراین اگر درصد بالا و پایینی از حیوانات جمعیت مرجع آستانه ای بیمار داشته باشند در نتیجه برخی حیوانات دارای شایستگی ژنوتیپی بالا و پایین به اشتباه از روی فنوتیپشان در گروه های بیمار و یا سالم قرار گرفته و این برآورد اشتباه باعث تخمین اشتباه در صحت ژنومی افراد تایید و کاهش صحت پیش بینی ژنومی را به همراه خواهد داشت.

گندم با استفاده از توزیع پروبیت در شیوع صفت آستانه ای شبیه سازی کردند. نتایج آن ها نشان داد که صحت ژنومی برای بیز B (۰/۵۱۱) نسبت به GBLUP (۰/۵۰۵) بیشتر بود. صادقی و همکاران (۱۳۹۶) جمعیت های مختلف ژنومی را برای نرخ شیوع ۵۰ درصد بیماری در جمعیت مرجع برای صفات آستانه ای شبیه سازی کردند. نتایج آن ها نشان داد که به طور کلی عملکرد بیز لاسو و A نسبت به روش جنگل تصادفی بهتر بود. با توجه به این که همیشه فنوتیپ بیانگر ژنوتیپ و توانایی حیوان در بروز صفت

جدول ۲: میانگین (انحراف معیار) صحت پیش بینی ژنومی روش های بیز آستانه ای A و جنگل تصادفی در هر یک از جمعیت های شبیه سازی شده با توجه به نسبت های مختلف شوع بیماری در جمعیت مرجع

درصد شیوع بیمار در جمعیت مرجع	بیز آستانه ای A				جنگل تصادفی			
	جمعیت ۱	جمعیت ۲	جمعیت ۳	جمعیت ۴	جمعیت ۱	جمعیت ۲	جمعیت ۳	جمعیت ۴
۵	۰/۳۴۶	۰/۳۷۹	۰/۲۴۴	۰/۲۶۰	۰/۳۴۱	۰/۲۶۳	۰/۱۸۴	۰/۲۱۹
	(۰/۰۹)	(۰/۰۷)	(۰/۰۸)	(۰/۱۰)	(۰/۰۶)	(۰/۰۳)	(۰/۰۶)	(۰/۰۳)
۱۰	۰/۴۲۲	۰/۴۴۴	۰/۳۱۸	۰/۳۵۲	۰/۴۵۸	۰/۳۵۶	۰/۲۲۷	۰/۲۸۳
	(۰/۰۸)	(۰/۰۸)	(۰/۰۹)	(۰/۰۸)	(۰/۰۴)	(۰/۰۷)	(۰/۰۷)	(۰/۰۲)
۱۵	۰/۴۶۳	۰/۴۸۱	۰/۳۶۲	۰/۴۰۲	۰/۵۳۵	۰/۴۵۴	۰/۳۱۵	۰/۳۸۱
	(۰/۱۰)	(۰/۰۸)	(۰/۱۰)	(۰/۰۸)	(۰/۰۴)	(۰/۰۲)	(۰/۰۴)	(۰/۰۱)
۲۰	۰/۵۱۶	۰/۵۳۲	۰/۴۱۴	۰/۴۶۵	۰/۵۷۶	۰/۴۷۳	۰/۳۳۲	۰/۴۲۱
	(۰/۰۸)	(۰/۰۹)	(۰/۱۰)	(۰/۰۸)	(۰/۰۳)	(۰/۰۳)	(۰/۰۳)	(۰/۰۴)
۲۵	۰/۴۷۳	۰/۵۲۸	۰/۴۲۴	۰/۴۶۶	۰/۵۴۲	۰/۴۵۸	۰/۳۲۶	۰/۳۹۸
	(۰/۰۹)	(۰/۱۰)	(۰/۰۸)	(۰/۰۹)	(۰/۰۵)	(۰/۰۱)	(۰/۰۵)	(۰/۰۲)
۳۰	۰/۴۵۱	۰/۵۱۴	۰/۴۰۷	۰/۴۴۲	۰/۵۰۱	۰/۴۲۵	۰/۲۹۴	۰/۳۴۱
	(۰/۱۱)	(۰/۱۰)	(۰/۰۷)	(۰/۰۹)	(۰/۰۴)	(۰/۰۲)	(۰/۰۶)	(۰/۰۱)
۳۵	۰/۴۱۱	۰/۴۷۶	۰/۳۵۸	۰/۳۹۲	۰/۴۶۰	۰/۳۹۱	۰/۲۶۲	۰/۲۹۸
	(۰/۰۷)	(۰/۰۸)	(۰/۰۹)	(۰/۰۹)	(۰/۰۳)	(۰/۰۷)	(۰/۰۳)	(۰/۰۲)
۴۰	۰/۳۶۱	۰/۴۳۱	۰/۳۰۹	۰/۳۳۷	۰/۳۹۵	۰/۳۳۲	۰/۲۲۶	۰/۲۵۱
	(۰/۰۸)	(۰/۰۸)	(۰/۰۹)	(۰/۰۸)	(۰/۰۴)	(۰/۰۶)	(۰/۰۲)	(۰/۰۲)
۴۵	۰/۳۱۳	۰/۳۷۸	۰/۲۳۸	۰/۲۶۲	۰/۳۲۱	۰/۲۷۰	۰/۱۸۱	۰/۱۹۸
	(۰/۰۹)	(۰/۰۸)	(۰/۰۹)	(۰/۱۱)	(۰/۰۶)	(۰/۰۳)	(۰/۰۵)	(۰/۰۴)
۵۰	۰/۲۵۴	۰/۳۳۱	۰/۱۸۳	۰/۲۰۱	۰/۲۳۸	۰/۲۰۳	۰/۱۴۹	۰/۱۶۵
	(۰/۱۰)	(۰/۰۹)	(۰/۰۷)	(۰/۰۸)	(۰/۰۵)	(۰/۰۴)	(۰/۰۵)	(۰/۰۳)

جمعیت اول: وراثت پذیری صفت ۰/۲۵ - تعداد QTL ۶۰۰ - سطح پایین LD: جمعیت دوم: وراثت پذیری صفت ۰/۲۵ - تعداد QTL ۱۵۰ - سطح پایین LD: جمعیت سوم: وراثت پذیری صفت ۰/۰۵ - تعداد QTL ۱۵۰ - سطح پایین LD: جمعیت چهارم: وراثت پذیری صفت ۰/۰۵ - تعداد QTL ۱۵۰ - سطح بالای LD

اثر تعداد QTL بر صحت ژنومی جمعیت‌های مختلف

با شیوع متفاوت بیماری

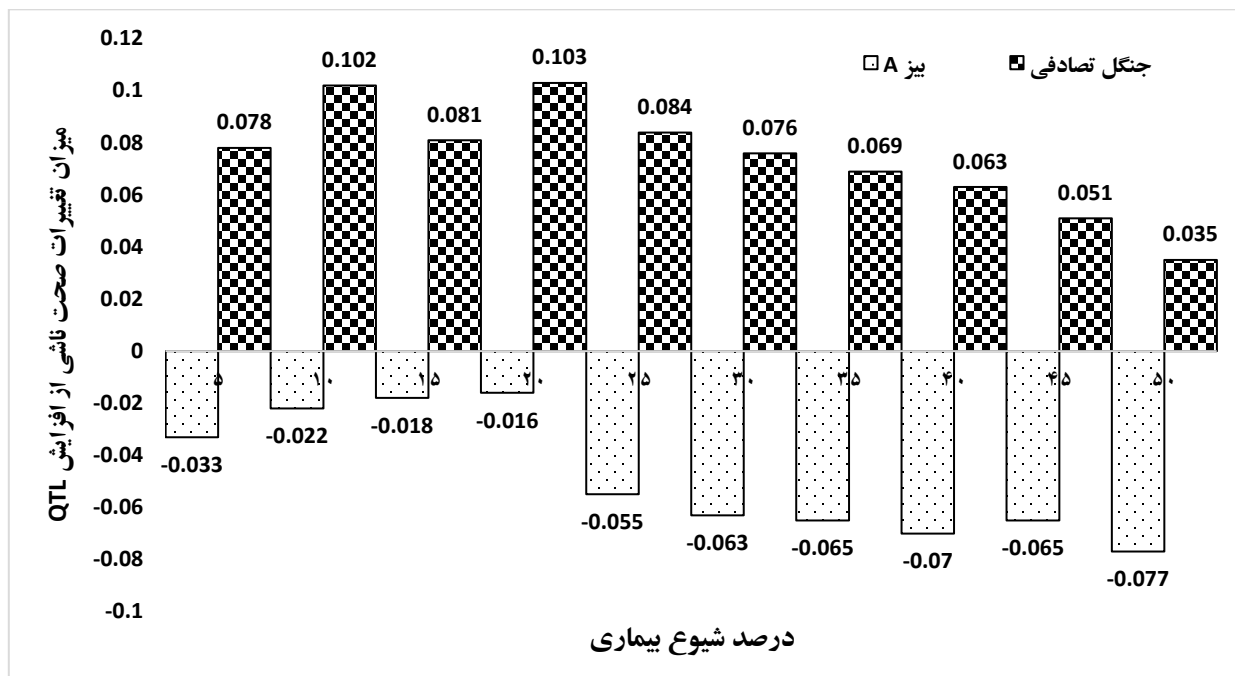
افزایش یافت. نتایج نشان داد که اثر افزایش و کاهش بیش از حد شیوع بیماری برای تعداد کم QTL بازخورد منفی تری نسبت به تعداد بالای QTL دارد. این درحالی بود که، با افزایش درصد شیوع بیماری تا ۲۰ درصد، صحت ژنومی جنگل تصادفی نوسان بیش تری به تغییرات QTL نشان داد و بعداز آن این میزان نوسان کم تر شد. در نتیجه برای مقادیر بهینه شیوع بیماری، اثر مثبت تعداد زیاد QTL بر صحت ژنومی روش جنگل تصادفی نسبت به تعداد کم QTL بیشتر بود.

تحقیقات در زمینه اثر تعداد QTL بر صحت ژنومی نشان داد که صحت روش های آماری به شدت متاثر از تعداد QTL قرار می گیرد در این میان Shirali و همکاران (۲۰۱۵) با شبیه سازی تعداد مختلف QTL در جمعیت های مختلف به این نتیجه رسیدند که روش های بیزی نسبت به GBLUP صحت های بالایی را در حالتی که تعداد کم QTL بود از خود نشان دادند. Zhang و همکاران (۲۰۱۱) بیان کردند که صحت ژنومی به آرامی با افزایش تعداد QTL افزایش می یابد. هم چنین GorganiFirozjah و همکاران (۲۰۱۴) نشان دادند که با افزایش تعداد QTL از ۴۰۰ به ۶۰۰ میانگین صحت ژنومی از ۴۲/۷ به ۴۳/۳ افزایش می یابد. بانه و همکاران (۱۳۹۶) نوسانات غیر معنی دار و غیرمنظمی در صحت روش های بیز آستانه ای برای تعداد مختلف QTL مشاهده کردند. Naderi و همکاران (۲۰۱۶) به بررسی اثر سطوح مختلف QTL بر صحت ژنومی گروه های مختلف بیمار در جمعیت مرجع پرداخت و نشان دادند که در سطوح مختلف شیوع بیماری، با افزایش تعداد QTL عملکرد جنگل تصادفی افزایش می یابد. Forni و González-Recio (۲۰۱۱) به بررسی تاثیر تعداد QTL بر صحت ژنومی با نرخ شیوع ۵۰ درصد جمعیت مرجع در تراشه های کم تراکم پرداختند و بر خلاف نتایج این تحقیق نشان داد که صحت بیز آستانه ای A و جنگل تصادفی با افزایش تعداد QTL به ترتیب افزایش و کاهش می یابد. صادق و همکاران (۱۳۹۶) به بررسی اثر تعداد QTL بر صحت ژنومی

برای ارزیابی اثر تعداد QTL بر صحت ژنومی، جمعیت ۱ (QTL۱۵۰) و ۲ (QTL ۶۰۰) با هم مقایسه شدند (جدول ۲). شکل ۱ اثر تغییرات QTL بر صحت ژنومی بیز A و جنگل تصادفی را نشان می دهد. بیش ترین و کم ترین میزان صحت ژنومی برای جمعیت ۱ با شیوع ۲۰ درصد و جمعیت ۲ با شیوع ۵۰ درصد با استفاده از جنگل تصادفی بود. صحت پیش بینی ژنومی ناشی از روش بیز آستانه ای A در تمامی سطوح شیوع بیماری با افزایش تعداد QTL کاهش یافت. زیرا با افزایش تعداد QTL ها سهم هر QTL در ارزش ژنتیکی کل کاهش یافته و به نوعی قدرت مدل ها در پیش بینی این اثرات کوچک کاهش خواهد یافت. هم چنین به خاطر این که روش بیزی A از برخی خصوصیات نظیر انتخاب متغیر سود می برند و مفروضات آن با تعداد QTL کم سازگارتر است در نتیجه در تعداد اندک QTL، روش بیز آستانه ای A بهتر عمل کرد. Habier و همکاران ۲۰۰۹ گزارش کردند هنگامی که تعداد QTL افزایش می یابد به تعداد بیشتری نشانگر جهت به دام انداختن اثرات همه QTL ها احتیاج خواهد بود. در نتیجه افزایش تعداد QTL می تواند منجر به افزایش صحت ژنومی شود هنگامی که به طور موازی با افزایش تعداد QTL ها، تعداد نشانگرها نیز افزایش یابد. هم چنین در روش های باز نمونه گیری از جمله جنگل تصادفی، افزایش تعداد QTL در تراشه های با تراکم بالا، منجر به تولید عدم پیوستگی قوی بین برخی نشانگرها با QTL های کنترل کننده ی صفت، نزدیک تر شدن فاصله نشانگرها با QTL ها و افزایش شانس نمونه گیری شده، در نتیجه افزایش صحت ژنومی را به همراه خواهد داشت (Naderi و همکاران، ۲۰۱۶)، که این اثر مثبت در مورد روش جنگل تصادفی در نتایج تحقیق حاضر صادق بود. بطور کلی نوسانات روش جنگل تصادفی نسبت به روش بیز A نسبت به تغییرات QTL بیشتر بود. با افزایش درصد شیوع بیماری تا ۲۰ درصد، صحت ژنومی بیز A نوسانات کم تری به تغییرات QTL نشان داد و بعداز آن این میزان نوسانات به طور صعودی

تغییرات QTL نسبت به روش‌های بیزی بیشتر بود. هم‌چنین افزایش تعداد نشانگرها همراه با تعداد QTL نسبت به تعداد کم QTL منجر به بهبود صحت ژنومی در روش جنگل تصادفی شد.

صفات آستانه‌ای دودویی (با نرخ شیوع ۵۰ درصد) با استفاده از روش‌های بیزی و یادگیری ماشین پرداختند. مطابق با نتایج تحقیق حاضر نتایج نشان داد که نوسانات روش‌های جنگل تصادفی به



شکل ۱: اثر افزایش تعداد QTL (از ۱۵۰ به ۶۰۰) بر صحت ژنومی روش‌های بیز آستانه‌ای A و جنگل تصادفی برای نسبت‌های مختلف شیوع بیماری در جمعیت مرجع

اثر وراثت‌پذیری بر صحت ژنومی جمعیت‌های مختلف با شیوع متفاوت بیماری

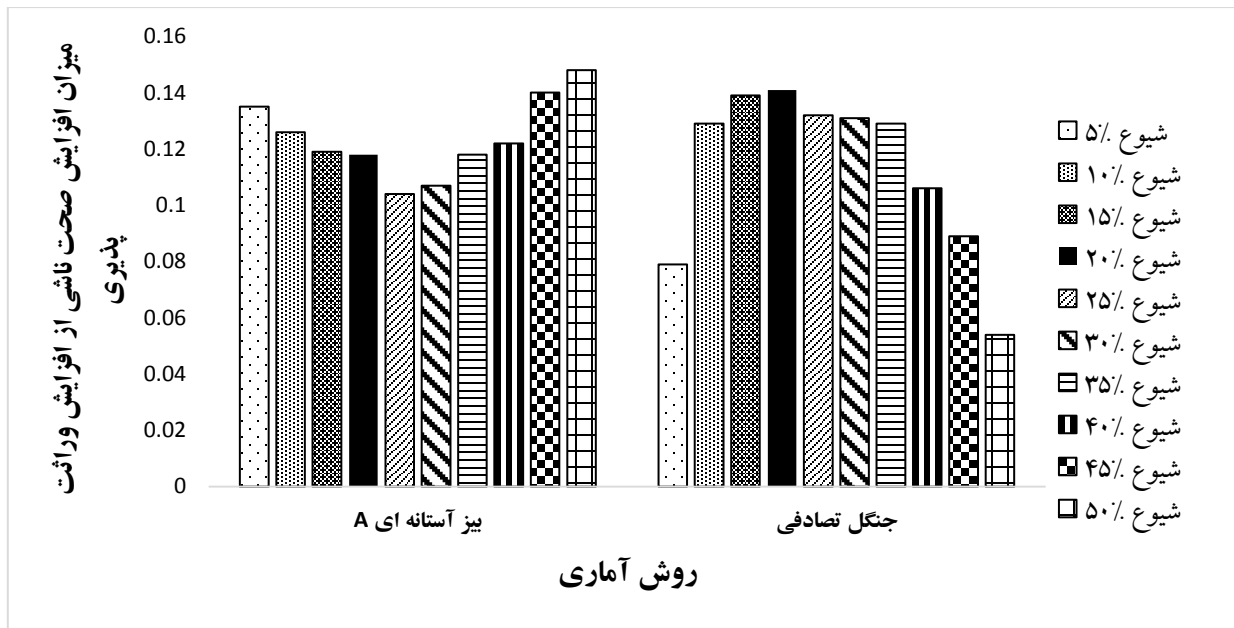
باشند. صحت پیش‌بینی ژنومی ناشی از روش بیز آستانه‌ای A و جنگل تصادفی در تمامی سطوح شیوع بیماری با افزایش وراثت‌پذیری افزایش یافت. با این حال این میزان افزایش برای بیز آستانه‌ای A هنگامی که میزان شیوع بیماری حالت بهینه داشت (۰/۲۵ شیوع) نسبت به مقادیر پایین و بالای شیوع بیماری کم‌تر بود. در روش‌های رایج انتخاب، روند بهبود صفات با وراثت‌پذیری پایین به آرامی صورت می‌گیرد، به دلیل این که صحت در این روش‌ها به میزان زیادی به وراثت‌پذیری صفت بستگی دارد. در حالی که در انتخاب ژنومی خصوصاً در حالت تراکم‌های بالای نشانگرها، صحت برای صفات با توارث پایین اثر

برای بررسی اثر وراثت‌پذیری بر صحت ژنومی جمعیت ۲ ($h^2=0/25$) و ۳ ($h^2=0/05$) مورد ارزیابی قرار گرفت. هم‌چنین شکل ۲ اثر افزایش وراثت‌پذیری بر صحت ژنومی روش‌های بیز آستانه‌ای A و جنگل تصادفی برای نسبت‌های مختلف شیوع بیماری در جمعیت مرجع را نشان می‌دهد. بیش‌ترین میزان صحت ژنومی برای جمعیت با وراثت‌پذیری پایین (جمعیت ۳) و بالا (جمعیت ۲) در روش بیز آستانه‌ای A هنگامی بود که به ترتیب ۲۵ و ۲۰ درصد افراد جمعیت بیمار باشند. کم‌ترین میزان صحت ژنومی برای جمعیت ۳ در روش جنگل تصادفی هنگامی بود که به ترتیب ۵۰ درصد افراد جمعیت مرجع بیمار

ارتباط مستقیم بین وراثت‌پذیری و صحت پیش بینی ارزش‌های اصلاحی ژنومی بود. در این تحقیق به طور کلی نوسانات روش بیزی به تغییرات وراثت‌پذیری بیش از جنگل تصادفی بود. مطالعات جهت بررسی صحت ژنومی صفات آستانه‌ای با نرخ شیوع بیماری ۵۰ درصد در جمعیت مرجع نشان داد که با افزایش وراثت‌پذیری (از ۰/۱ به ۰/۳)، روش بیزی (با افزایش ۰/۱۷ در صحت) نسبت به روش جنگل تصادفی (با افزایش ۰/۱۵) از نوسانات بیشتری برخوردار است (صادقی ۱۳۹۶). مطالعات اعلاء نوشهر (۱۳۹۵) در جمعیت‌های شبیه سازی نشان داد که با افزایش وراثت‌پذیری از ۰/۱ به ۰/۵ روش بیزی صحت بالاتری نسبت به روش GBLUP داشت و صحت پیش بینی ژنومی برای مدل‌های بیزی نسبت به سایر روش‌ها به شدت به تغییرات وراثت‌پذیری واکنش نشان داد. هم چنین Naderi و همکاران (۲۰۱۶) به بررسی اثر وراثت‌پذیری بر سطح زیر منحنی مشخصه عملکرد (AUROC) با روش‌های GBLUP و جنگل تصادفی پرداختند و این نتیجه رسیدند که نوسانات جنگل تصادفی به تغییرات وراثت‌پذیری کم تراز روش GBLUP است. بهرام و همکاران (۱۳۹۲) به بررسی اثر وراثت‌پذیری بر صحت ارزش‌های اصلاحی و پیشرفت ژنتیکی صفات آستانه با ارزیابی کلاسیک و ژنومی پرداخت. نتایج نشان داد که افزایش وراثت‌پذیری می‌تواند عاملی تأثیر گذار بر صحت ارزش اصلاحی و پیشرفت ژنتیکی در صفات آستانه باشد، و مقادیر صحت ارزش‌های اصلاحی ژنومی را برای صفات با وراثت‌پذیری ۰/۰۵ و ۰/۳ به ترتیب در دامنه‌ای از ۰/۲۲-۰/۴۵ و ۰/۲۷-۰/۶۱ گزارش کرد. در چندین مطالعه (Wang و همکاران، ۲۰۱۷) اثر مطلوب افزایش وراثت‌پذیری بر صحت پیش بینی ارزش‌های اصلاحی ژنومی ناشی از مدل‌های بیزی به اثبات رسیده است. این تأثیر مثبت و مطلوب وراثت‌پذیری منجر به تغییرات بالای ژنتیکی و در نتیجه کمک به پیش بینی بهتر اثرات نشانگری شد.

مثبت تری نسبت به صفات با وراثت‌پذیری بالا دارد (Muir، ۲۰۰۷) در نتیجه سود حاصل از تراکم بالای نشانگرها (در حالت بهینه شیوع بیماری) بر صفات با توارت‌پذیری پایین بالا بوده و اختلاف صحت کاهش یافته است. این در حالی بود که برای جنگل تصادفی این وضعیت به طور معکوسی رخ داد و بیش‌ترین افزایش صحت به خاطر افزایش وراثت‌پذیری در شیوع ۲۰ درصد رخ داد. این امر نشان می‌دهد که اثر بهینه شیوع بیماری در صفات با وراثت‌پذیری بالا نسبت به صفات با وراثت‌پذیری پایین اثر مثبت تری بر صحت ژنومی روش جنگل تصادفی دارد. به طور کلی دلیل افزایش صحت ژنومی با افزایش وراثت‌پذیری را می‌توان این‌گونه عنوان کرد که بالا بودن وراثت‌پذیری یعنی بیشتر بودن نسبت واریانس ژنتیکی به واریانس فنوتیپی و به معنای بیشتر بودن نقش ژن‌های با بیان افزایشی در ایجاد پراکنش در صفت است که باعث برآورد صحیح‌تر تأثیرات نشانگرها می‌شود. در نتیجه هرچه وراثت‌پذیری صفت بیشتر باشد، فنوتیپ فرد به ارزش ژنتیکی فرد نزدیک‌تر بوده و در نتیجه اثر نشانگرها و به دنبال آن ارزش‌های اصلاحی ژنومی افراد به طور صحیح‌تر پیش بینی می‌شود (Goddard و Hayes، ۲۰۰۹). هم چنین افزایش واریانس نمونه‌گیری برآورد اثرات نشانگرها به دلیل افزایش واریانس محیطی از مهم‌ترین دلایل کاهش صحت ژنومی در صفات با توارت‌پذیری پایین می‌باشند.

مطالعات نشان داده‌اند که وراثت‌پذیری صفت مورد نظر به عنوان یک فاکتور مهم در معماری ژنتیکی صفت، می‌تواند بر صحت پیش بینی ژنومی اثرگذار باشد. در وراثت‌پذیری بالاتر، ارزش‌های اصلاحی افراد سهم بیشتری در بروز رکوردهای فنوتیپ داشته و لذا در برآورد اثرات نشانگرها با استفاده از اطلاعات فنوتیپی صحت بیشتری مشاهده خواهد شد (Meuwissen و همکاران، ۲۰۰۱). نتایج این تحقیق مطابق با تئوری Bo و همکاران (۲۰۱۷) در مورد



شکل ۲: اثر افزایش وراثت پذیری (از ۰/۰۵ به ۰/۲۵) بر صحت ژنومی روش های بیز آستانه ای A و جنگل تصادفی برای نسبت های مختلف شیوع بیماری در جمعیت مرجع

اثر عدم تعادل پیوستگی بر صحت ژنومی جمعیت های

مختلف با شیوع متفاوت بیماری

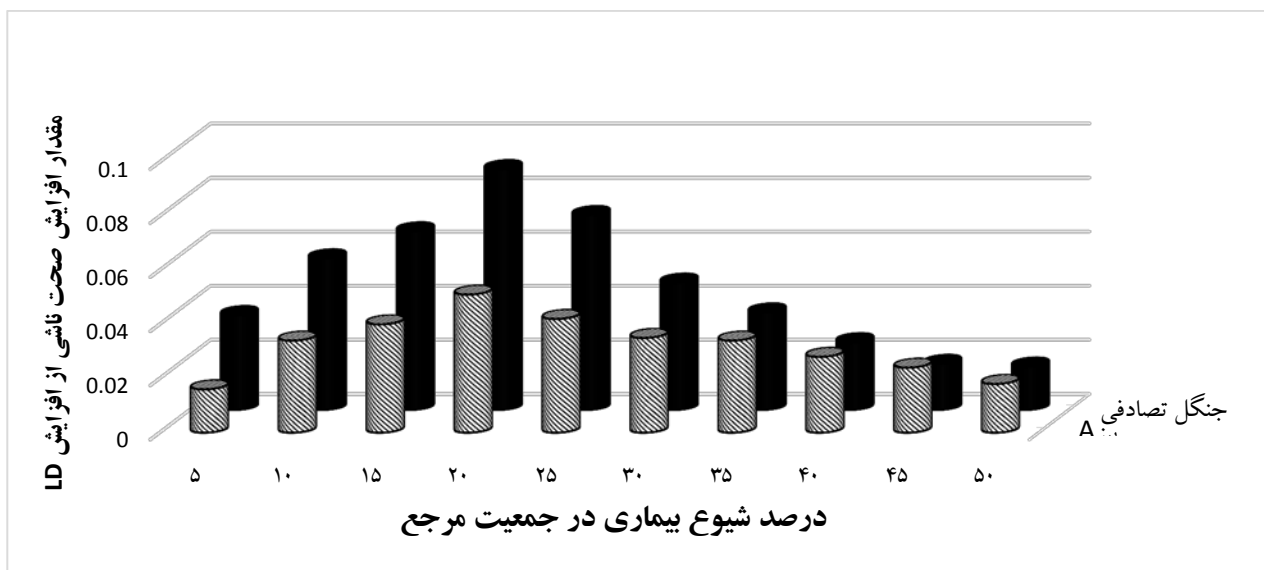
فواصل خیلی نزدیک به هم قرار گرفته اند دیده می شود) از طرف دیگر، باعث می شود روش جنگل تصادفی که مبتنی بر نمونه گیری می باشد شانس قرار گرفتن نشانگرهای با LD بالا را در کنار هم بیشتر کرده در نتیجه این امر منجر به عملکرد مثبت جنگل تصادفی شود. بیشترین اثر مثبت LD بر صحت ژنومی دو روش جنگل تصادفی و بیز A هنگامی بود که ۲۰ درصد افراد در جمعیت مرجع آستانه بیمار داشتند و با افزایش و کاهش این نسبت شیوع، نوسانات این دو مدل به تغییرات LD کاهش یافت. این امر نشان می دهد هنگامی که مقدار بهینه ی شیوع بیماری در جمعیت مرجع وجود داشته باشد اثر مثبت LD بر صحت ژنومی بهتر نمایان می شود. در این تحقیق صحت پیش بینی ارزش های اصلاحی ژنومی مدل جنگل تصادفی هم عرض با افزایش LD افزایش یافت که این مطابق با نتایج شبیه سازی Naderi و همکاران (۲۰۱۶) بود.

به عنوان یک اصل کلی، وجود LD بین نشانگر و QTL منبع اصلی اطلاعات است و نقش عمده ای در صحت پیش بینی ارزش های اصلاحی ژنومی ایفا می کنند (Sun و همکاران،

برای ارزیابی اثر LD بر صحت ژنومی، جمعیت های ۳ (LD پایین) و ۴ (LD بالا) باهم مقایسه شدند (جدول ۲). میانگین LD برای جمعیت های ۳ و ۴ در فاصله ۰/۰۵ سانتیمورگان به ترتیب ۰/۲۳۵ و ۰/۴۴۱ بود. شکل ۳ اثر تغییرات LD بر صحت ژنومی هر یک از روش های بیز A و جنگل تصادفی را نشان می دهد. بیشترین و کمترین میزان صحت ژنومی ناشی از دو روش بیز A و جنگل تصادفی در جمعیت ۳ با شیوع ۲۰ درصد بیماری و جمعیت ۴ با شیوع ۵۰ درصد بیماری بود. صحت پیش بینی ژنومی ناشی از روش بیز آستانه ای A و جنگل تصادفی در تمامی سطوح شیوع بیماری با افزایش LD افزایش یافت و به طور کلی نوسانات روش جنگل تصادفی نسبت به تغییرات LD بیشتر از بیز A بود. به دلیل این که سطح بالای LD میان QTLها و نشانگرها نشان می دهد که نشانگرها با تراکم بالا سهم بالایی از واریانس ژنتیکی را به خود اختصاص می دهند (Goddard، ۲۰۰۹) از یک طرف و استفاده از تراشه متراکم در تحقیق حاضر (با توجه به اینکه در تراشه های متراکم LD تابعی از فواصل بین جفت های SNP بوده و بیشترین مقادیر LD برای SNPهایی که در

ژنومی افزایش خواهد یافت. آن‌ها گزارش کردند که اگر سطح عدم تعادل پیوستگی بین نشانگرها از ۱۰ درصد به ۲۶ درصد افزایش یابد، صحت برآورد ارزش‌های اصلاحی ژنومی از ۰/۶۸ به ۰/۸۲ افزایش خواهد یافت. Naderi و همکاران (۲۰۱۶) به بررسی اثر LD بر صحت ژنومی (سطح زیر منحنی مشخصه عملکرد) پرداختند و نشان داد که برای سطوح مختلف شیوع بیماری، نوسانات جنگل تصادفی به تغییرات LD نسبت به GBLUP بیشتر است.

Jonas و همکاران (۲۰۱۷) وجود LD در بین نشانگرها را عاملی تأثیرگذار در بهبود صحت پیش‌بینی ارزش‌های اصلاحی ژنومی عنوان کردند. همچنین Ke و همکاران (۲۰۰۴) وجود LD قوی بین نشانگرها مجاور در بررسی مطالعات ژنوم انسان را اساسی عنوان کردند. GorganiFirozjah و همکاران (۲۰۱۴) گزارش کردند که با افزایش اندازه‌ی مؤثر جمعیت، معیار و در نتیجه LD کاهش می‌یابد و این کاهش به نوبه خود کاهش صحت ژنومی را در پی خواهد داشت. Calus و همکاران (۲۰۰۸) نشان دادند که با افزایش LD صحت برآورد ارزش‌های اصلاحی



شکل ۳: اثر افزایش LD (از ۰/۲۳۵ به ۰/۴۴۱) بر صحت ژنومی روش‌های بیز آستانه‌ای A و جنگل تصادفی برای نسبت‌های مختلف شیوع بیماری در جمعیت مرجع

نتیجه‌گیری

ژنومی بود. در این راستا اثر بخشی وراثت پذیری نسبت به تعداد QTL و میزان LD بر صحت ژنومی بیشتر بود. در مجموع نوسانات روش جنگل تصادفی نسبت به بیز A به تغییرات معماری ژنومی و نرخ شیوع بیماری بیشتر بود. با وجود صحت ژنومی بالاتر بیز آستانه‌ای A در معماری‌های مختلف، هنگامی که صفات با وراثت پذیری بالا توسط تعداد زیادی QTL کنترل می‌شوند روش جنگل تصادفی بالاترین صحت ژنومی را به دست آورد.

نرخ شیوع فنوتیپ در جمعیت مرجع یکی از مهمترین فاکتورهای مؤثر بر صحت ژنومی صفات آستانه‌ای با استفاده از روش‌های بیز آستانه‌ای A و جنگل تصادفی بود. صحت پیش‌بینی ژنومی هنگام بیشینه خواهد بود که بهینه شیوع بیماری در جمعیت مرجع برقرار باشد. به طور کلی بیش‌ترین صحت ژنومی دو روش بیز آستانه‌ای A و جنگل تصادفی هنگامی بود که ۲۰ درصد افراد جمعیت مرجع فنوتیپ بیمار را داشته باشند و با افزایش و کاهش نرخ بهینه شیوع بیماری، صحت ژنومی به شدت کاهش یافت. علاوه بر نرخ شیوع، نوع معماری ژنومی از فاکتورهای مهم و مؤثر بر صحت

منابع

- 9(7): e101544.
- Dekkers, J.C. (2010). Animal genomics and genomic selection. Paper presented at the Adapting animal production to changes for a growing human population. Proceedings of International Conference, Lleida, Spain.
- Egger-Danner, C., Cole, J., Pryce, J., Gengler, N., Heringstad, B., Bradley, A., *et al.* (2015). Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. *Animal*. 9(2): 191-207.
- Ghafari-Kesbi, F., Rahimi-Mianji, G., Honarvar, M., and Nejati-Javaremi, A. (2017). Predictive ability of Random Forests, Boosting, Support Vector Machines and Genomic Best Linear Unbiased Prediction in different scenarios of genomic evaluation. *Animal Production Science*. 57(2): 229-236.
- Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*. 136(2): 245-257.
- Goldstein, B.A., Hubbard, A.E., Cutler, A., and Barcellos, L.F. (2010). An application of Random Forests to a genome-wide association dataset: methodological considerations & new findings. *BMC genetics*. 11(1): 49.
- González-Recio, O., and Forni, S. (2011). Genome-wide prediction of discrete traits using Bayesian regressions and machine learning. *Genetics Selection Evolution*. 43(1): 7.
- Gorgani Firozjah, N., Atashi, H., Dadpasand, M., and Zamiri, M. (2014). Effect of marker density and trait heritability on the accuracy of genomic prediction over three generations. *Journal of Livestock Science and Technologies*. 2(2): 53-58.
- Habier, D., Fernando, R.L., and Dekkers, J.C. (2009). Genomic selection using low-density marker panels. *Genetics*. 182(1): 343-353.
- Hayes, B. (2007). QTL mapping, MAS, and genomic selection. A short-course. Animal Breeding & Genetics Department of Animal Science. Iowa State University. 1(1): 3-4.
- اعلاء نوشهر، ف. (۱۳۹۵). به کارگیری مدل های آماری انتخاب ژنومی و آنالیز QTL در برنامه های اصلاح نژادی گوسفند. پایان نامه دکترا، دانشکده کشاورزی دانشگاه تبریز.
- بانسه، ح.، نجاتی جوارمی، ا.، رحیمی میانجی، ق. و هنرور، م. (۱۳۹۶). ارزیابی ژنومی صفات آستانه ای با معماری های ژنتیکی متفاوت با استفاده از روشهای بیژی. *پژوهشهای تولیدات دامی*. ۱۵، ص.ص. ۱۵۴-۱۴۹.
- بهمرام، ر. (۱۳۹۲). مطالعه شبیه سازی صحت ارزش اصلاحی و پیشرفت ژنتیکی در صفات آستانه با ارزیابی کلاسیک و ژنومی. پایان نامه دکترا، دانشکده کشاورزی دانشگاه فردوسی مشهد.
- صادقی، س. (۱۳۹۶). ارزیابی معماری ژنومی صفات گسسته با جهانی داده های ژنومی گوسفند توسط روش های آستانه ای بیژی و یادگیری ماشین. پایان نامه دکترا، دانشکده کشاورزی دانشگاه تبریز.
- Badke, Y.M., Bates, R.O., Ernst, C.W., Fix, J., and Steibel, J.P. (2014). Accuracy of estimation of genomic breeding values in pigs using low-density genotypes and imputation. *G3: Genes/ Genomes/ Genetics*. 4(4): 623-631.
- Bo, Z., ZHANG, J.-j., Hong, N., Long, G., Peng, G., XU, L.-y., *et al.* (2017). Effects of marker density and minor allele frequency on genomic prediction for growth traits in Chinese Simmental beef cattle. *Journal of Integrative Agriculture*. 16(4): 911-920.
- Breiman, L. (2001). Random forests. *Machine learning*. 45(1): 5-32.
- Buch, L.H., Kargo, M., Berg, P., Lassen, J., and Sørensen, A.C. (2012). The value of cows in reference populations for genomic selection of new functional traits. *Animal*. 6(6): 880-886.
- Calus, M., De Roos, A., and Veerkamp, R. (2008). Accuracy of genomic selection using different methods to define haplotypes. *Genetics*. 178(1): 553-561.
- Chen, L., Li, C., Sargolzaei, M., and Schenkel, F. (2014). Impact of genotype imputation on the performance of GBLUP and Bayesian methods for genomic prediction. *PLoS one*.

- Hayes, B., and Goddard, M.E. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution*. 33(3): 209.
- Hayes, B.J., Bowman, P.J., Chamberlain, A., and Goddard, M. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*. 92(2): 433-443.
- Jónás, D., Ducrocq, V., and Croiseau, P. (2017). The combined use of linkage disequilibrium-based haploblocks and allele frequency-based haplotype selection methods enhances genomic evaluation accuracy in dairy cattle. *Journal of Dairy Science*. 100(4): 2905-2908.
- Ke, X., Hunt, S., Tapper, W., Lawrence, R., Stavrides, G., Ghorji, J., et al. (2004). The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Human Molecular Genetics*. 13(6): 577-588.
- Meuwissen, T., Hayes, B., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157(4): 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(6): 342-355.
- Naderi, S., Yin, T., and König, S. (2016). Random forest estimation of genomic breeding values for disease susceptibility over different disease incidences and genomic architectures in simulated cow calibration groups. *Journal of Dairy Science*. 99(9): 7261-7273.
- Ogutu, J.O., Piepho, H.-P., and Schulz-Streeck, T. (2011). A comparison of random forests, boosting and support vector machines for genomic selection. Paper presented at the BMC proceedings.
- Piccoli, M.L., Braccini, J., Cardoso, F.F., Sargolzaei, M., Larmer, S.G., and Schenkel, F.S. (2014). Accuracy of genome-wide imputation in Braford and Hereford beef cattle. *BMC genetics*. 15(1): 157.
- Pimentel, E.C., Wensch-Dorendorf, M., König, S., and Swalve, H.H. (2013). Enlarging a training set for genomic selection by imputation of un-genotyped animals in populations of varying genetic architecture. *Genetics Selection Evolution*. 45(1): 12.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. and Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*. 81: 559-575.
- Sargolzaei, M., and Schenkel, F.S. (2009). QMSim: a large-scale genome simulator for livestock. *Bioinformatics*. 25(5): 680-681.
- Shirali, M., Miraei-Ashtiani, S. R., Pakdel, A., Haley, C., Navarro, P. and Pong-Wong, R., (2015). A comparison of the sensitivity of the bayesC and genomic best linear unbiased prediction (GBLUP) methods of estimating genomic breeding values under different quantitative trait locus (QTL) model assumptions. *Iranian Journal of Applied Animal Science*. 5: 41-46.
- Solberg, T., Sonesson, A., and Woolliams, J. (2008). Genomic selection using different marker types and densities. *Journal of animal science*. 86(10): 2447-2454.
- Sun, X., Fernando, R., and Dekkers, J. (2016). Contributions of linkage disequilibrium and co-segregation information to the accuracy of genomic prediction. *Genetics Selection Evolution*. 48(1): 77.
- Technow, F., and Melchinger, A.E. (2013). Genomic prediction of dichotomous traits with Bayesian logistic models. *Theoretical and applied genetics*. 126(4): 1133-1143.
- Wang, C., Ding, X., Wang, J., Liu, J., Fu, W., Zhang, Z., et al. (2013). Bayesian methods for estimating GEBVs of threshold traits. *Heredity*. 110(3): 213-219.

Wang, C., Li, X., Qian, R., Su, G., Zhang, Q., and Ding, X. (2017). Bayesian methods for jointly estimating genomic breeding values of one continuous and one threshold trait. *PloS one*. 12(4): e0175448.

Wiggans, G., VanRaden, P., and Cooper, T. (2011). The genomic evaluation system in the United States: Past, present, future. *Journal of Dairy Science*. 94(6): 3202-3211.

Yin, T., Pimentel, E., Borstel, U.K.v., and König, S. (2014). Strategy for the simulation and analysis of longitudinal phenotypic and genomic data in the context of a temperature× humidity-dependent covariate. *Journal of Dairy Science*. 97(4): 2444-2454.

Zhang, Z., Zhang, Q., and Ding, X. (2011). Advances in genomic selection in domestic animals. *Chinese science bulletin*. 56(25): 2655-2663.

♦ ♦ ♦ ♦ ♦ ♦ ♦ ♦ ♦ ♦